ARTIFICIAL INTELLIGENCE APPLICATIONS IN SPECIALTY CROPS

EDITED BY: Yiannis Ampatzidis, Reza Ehsani, Spyros Fountas and Panos M. Pardalos

PUBLISHED IN: Frontiers in Plant Science and Frontiers in Artificial Intelligence

frontiers Research Topics

frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88974-557-9
DOI 10.3389/978-2-88974-557-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ARTIFICIAL INTELLIGENCE APPLICATIONS IN SPECIALTY CROPS

Topic Editors:
Yiannis Ampatzidis, University of Florida, United States
Reza Ehsani, University of California, Merced, United States
Spyros Fountas, Agricultural University of Athens, Greece
Panos M. Pardalos, University of Florida, United States

Citation: Ampatzidis, Y., Ehsani, R., Fountas, S., Pardalos, P. M., eds. (2022). Artificial Intelligence Applications in Specialty Crops. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-557-9

Table of Contents

06 ***Editorial: Artificial Intelligence Applications in Specialty Crops***
Sri Charan Kakarla, Zhicheng Zhu, Yiannis Ampatzidis, Spyros Fountas, Reza Ehsani and Panos Pardalos

10 ***A Monitoring System for the Segmentation and Grading of Broccoli Head Based on Deep Learning and Neural Networks***
Chengquan Zhou, Jun Hu, Zhifu Xu, Jibo Yue, Hongbao Ye and Guijun Yang

23 ***Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest***
Kasper Johansen, Mitchell J. L. Morton, Yoann Malbeteau, Bruno Aragon, Samer Al-Mashharawi, Matteo G. Ziliani, Yoseline Angel, Gabriele Fiene, Sónia Negrão, Magdi A. A. Mousa, Mark A. Tester and Matthew F. McCabe

44 ***Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network***
Jun Liu and Xuewei Wang

56 ***A Cloud-Based Environment for Generating Yield Estimation Maps From Apple Orchards Using UAV Imagery and a Deep Learning Technique***
Orly Enrique Apolo-Apolo, Manuel Pérez-Ruiz, Jorge Martínez-Guanter and João Valente

71 ***Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning***
Zhe Lin and Wenxuan Guo

84 ***Wheat Kernel Variety Identification Based on a Large Near-Infrared Spectral Dataset and a Novel Deep Learning-Based Feature Selection Method***
Lei Zhou, Chu Zhang, Mohamed Farag Taha, Xinhua Wei, Yong He, Zhengjun Qiu and Yufei Liu

96 ***Investigation on Data Fusion of Multisource Spectral Data for Rice Leaf Diseases Identification Using Machine Learning Methods***
Lei Feng, Baohua Wu, Susu Zhu, Junmin Wang, Zhenzhu Su, Fei Liu, Yong He and Chu Zhang

113 ***Automatic Detection of Flavescence Dorée Symptoms Across White Grapevine Varieties Using Deep Learning***
Justine Boulent, Pierre-Luc St-Charles, Samuel Foucher and Jérome Théau

127 ***Machine Learning Unmasked Nutritional Imbalances on the Medicinal Plant* Bryophyllum *sp. Cultured* in vitro**
Pascual García-Pérez, Eva Lozano-Milo, Mariana Landin and Pedro Pablo Gallego

141 ***Machine Learning Techniques for Soybean Charcoal Rot Disease Prediction***
Elham Khalili, Samaneh Kouchaki, Shahin Ramazi and Faezeh Ghanati

153 ***Attention-Based Recurrent Neural Network for Plant Disease Classification***
Sue Han Lee, Hervé Goëau, Pierre Bonnet and Alexis Joly

161 ***Application of Laser-Induced Breakdown Spectroscopy in Detection of Cadmium Content in Rice Stems***
Wei Wang, Wenwen Kong, Tingting Shen, Zun Man, Wenjing Zhu, Yong He, Fei Liu and Yufei Liu

172 ***Accuracy of a Smartphone-Based Object Detection Model, PlantVillage Nuru, in Identifying the Foliar Symptoms of the Viral Diseases of Cassava—CMD and CBSD***
Latifa M. Mrisho, Neema A. Mbilinyi, Mathias Ndalahwa, Amanda M. Ramcharan, Annalyse K. Kehs, Peter C. McCloskey, Harun Murithi, David P. Hughes and James P. Legg

186 ***Fine-Grained Image Classification for Crop Disease Based on Attention Mechanism***
Guofeng Yang, Yong He, Yong Yang and Beibei Xu

201 ***Modeling and Optimizing Culture Medium Mineral Composition for* in vitro *Propagation of* Actinidia arguta**
Radhia Hameg, Tomás A. Arteta, Mariana Landin, Pedro P. Gallego and M. Esther Barreal

220 ***Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean***
Mohsen Yoosefzadeh-Najafabadi, Hugh J. Earl, Dan Tulpan, John Sulik and Milad Eskandari

234 ***Fusion of Deep Convolution and Shallow Features to Recognize the Severity of Wheat* Fusarium *Head Blight***
Chunyan Gu, Daoyong Wang, Huihui Zhang, Jian Zhang, Dongyan Zhang and Dong Liang

248 ***Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation***
Jennifer Hobbs, Prajwal Prakash, Robert Paull, Harutyun Hovhannisyan, Bernard Markowicz and Greg Rose

262 ***Corrigendum: Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation***
Jennifer Hobbs, Prajwal Prakash, Robert Paull, Harutyun Hovhannisyan, Bernard Markowicz and Greg Rose

264 ***Combining Multi-Dimensional Convolutional Neural Network (CNN) With Visualization Method for Detection of* Aphis gossypii *Glover Infection in Cotton Leaves Using Hyperspectral Imaging***
Tianying Yan, Wei Xu, Jiao Lin, Long Duan, Pan Gao, Chu Zhang and Xin Lv

279 ***DAM: Hierarchical Adaptive Feature Selection Using Convolution Encoder Decoder Network for Strawberry Segmentation***
Talha Ilyas, Muhammad Umraiz, Abbas Khan and Hyongsuk Kim

295 ***Deep Segmentation of Point Clouds of Wheat***
Morteza Ghahremani, Kevin Williams, Fiona M. K. Corke, Bernard Tiddeman, Yonghuai Liu and John H. Doonan

310 ***Tomato Anomalies Detection in Greenhouse Scenarios Based on YOLO-Dense***
Xuewei Wang and Jun Liu

321 ***Biotic Yield Losses in the Southern Amazon, Brazil: Making Use of Smartphone-Assisted Plant Disease Diagnosis Data***
Anna C. Hampf, Claas Nendel, Simone Strey and Robert Strey

337 ***Multiscale Parallel Algorithm for Early Detection of Tomato Gray Mold in a Complex Natural Environment***
Xuewei Wang and Jun Liu

349 ***Leaf Counting: Fusing Network Components for Improved Accuracy***
Guy Farjon, Yotam Itzhaky, Faina Khoroshevsky and Aharon Bar-Hillel

361 ***Occlusion Robust Wheat Ear Counting Algorithm Based on Deep Learning***
Yiding Wang, Yuxin Qin and Jiali Cui

375 ***Investigating a Selection of Methods for the Prediction of Total Soluble Solids Among Wine Grape Quality Characteristics Using Normalized Difference Vegetation Index Data From Proximal and Remote Sensing***
Aikaterini Kasimati, Borja Espejo-Garcia, Eleanna Vali, Ioannis Malounas and Spyros Fountas

385 ***Grape Leaf Black Rot Detection Based on Super-Resolution Image Enhancement and Deep Learning***
Jiajun Zhu, Man Cheng, Qifan Wang, Hongbo Yuan and Zhenjiang Cai

401 ***TheLNet270v1 – A Novel Deep-Network Architecture for the Automatic Classification of Thermal Images for Greenhouse Plants***
Md. Parvez Islam, Yuka Nakano, Unseok Lee, Keinichi Tokuda and Nobuo Kochi

416 ***Predictive Chromatography of Leaf Extracts Through Encoded Environmental Forcing on Phytochemical Synthesis***
Junelle Rey C. Bacong and Drandreb Earl O. Juanico

429 ***Hyperspectral Imaging Combined With Deep Transfer Learning for Rice Disease Detection***
Lei Feng, Baohua Wu, Yong He and Chu Zhang

frontiers | Frontiers in Plant Science

EDITORIAL
published: 21 April 2022
doi: 10.3389/fpls.2022.866724

Editorial: Artificial Intelligence Applications in Specialty Crops

Sri Charan Kakarla[1], *Zhicheng Zhu*[1], *Yiannis Ampatzidis*[1]*, *Spyros Fountas*[2], *Reza Ehsani*[3] *and Panos Pardalos*[4]

[1] Department of Agricultural and Biological Engineering, Southwest Florida Research and Education Center, University of Florida, IFAS, Immokalee, FL, United States, [2] Department of Natural Resources Management and Agricultural Engineering, Agricultural University of Athens, Athens, Greece, [3] School of Engineering, University of California, Merced, Merced, CA, United States, [4] Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, United States

Keywords: machine learning and AI, unmanned aerial vehicle (UAV), remote sensing, automation and robotics, smart farming

Editorial on the Research Topic

Artificial Intelligence Applications in Specialty Crops

The rapid development of remote sensing systems, mechatronics, robotics, big data analytics, and artificial intelligence (AI) has revolutionized the specialty crop industry by allowing more precise, efficient, and cost-effective management. This Research Topic "*Artificial Intelligence Applications in Specialty Crops*" has attracted 31 high quality articles that cover the state-of-the-art applications and technical development of artificial intelligence in crop production, such as nutrient management, pest and disease management, phenotyping, and yield prediction.

Understanding the cause-effect of nutrients and the crop growth is of vital importance to the efficiency of the plant tissue culture. Conventional method using complex multivariate design of experiments (DoE) may become time consuming and lead to uninterpretable results as the number of variables increases. Garcia-Perez et al. combined artificial neural networks (ANNs) and fuzzy logic to determine the factors such as mineral nutrition affecting *Bryophyllum in vitro* growth. The result showed that the proposed model can unmask the concealed interactions and generate insights in understanding the factors involved in a certain response. Hameg et al. compared performance of different methods such as DoE, artificial neural network, fuzzy logic, and genetic algorithm (GA) in identifying and optimizing the mineral combination for hardy kiwi *in vitro* micropropagation. They demonstrated that DoE can be more efficient, and ANNs combined with fuzzy logic can understand the cause-effect of the factors and responses, and ANNs-GA can predict new mineral media formulations.

Disease diagnosis by visual observation of symptoms can get complex due to the similarity of symptoms between various diseases. This complex nature can even confuse experienced personnel to misidentify a disease which can lead to further problems. This Research Topic includes the state-of-the-art machine learning techniques that address this emerging challenges. These methods can be divided into two categories. The first type of methods developed new features and image sets that may not be visible to naked eyes for classification. For example, Feng, Wu, Zhu, et al. fused three different types of spectral datasets with visible/near-infrared hyperspectral imaging (HSI), mid-infrared spectroscopy (MIR), and laser-induced breakdown spectroscopy (LIBS) to detect three different rice diseases namely leaf blight, rice blast, and rice sheath blight. Classification models were developed using support vector machine (SVM), logistic regression (LR), and convolutional neural network (CNN) models. Best results were achieved with the models based on HSI with accuracy over 93%. Feng, Wu, He, et al. used HSI combined with deep transfer learning methods to detect rice leaf diseases in four different varieties of rice. The results suggested that this method is a promising possibility for the efficient and cost-saving field detection of rice diseases among different rice varieties. Yan et al.

OPEN ACCESS

Edited and Reviewed by:
Julie A. Dickerson,
Iowa State University, United States

***Correspondence:**
Yiannis Ampatzidis
i.ampatzidis@ufl.edu

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *31 January 2022*
Accepted: *28 March 2022*
Published: *21 April 2022*

Citation:
Kakarla SC, Zhu Z, Ampatzidis Y, Fountas S, Ehsani R and Pardalos P (2022) Editorial: Artificial Intelligence Applications in Specialty Crops. Front. Plant Sci. 13:866724. doi: 10.3389/fpls.2022.866724

also used HSI data of both healthy and infected cotton plants to detect *Aphis gossypii* Glover infection. The spectra, RGB images, and hyperspectral images were used for one-dimensional, two-dimensional, and three-dimensional analysis. This study developed an initial assessment for pest detection in cotton and provides new and alternative ideas for future researchers. Khalili et al. developed and evaluated several machine learning methods to predict charcoal rot disease in soybean. And the result show the gradient tree boosting (GBT) was found to be the best performing method, which obtained 96.25 and 97.33% in terms of sensitivity and specificity. The second type of methods improved existing CNN models to achieve a better performance. For example, you only look once (YOLO) is a state-of-the-art real-time object detection algorithm, and Liu and Wang developed an improved YOLO V3 algorithm to detect tomato diseases and insect pests such as early blight, gray mold, late blight, leaf mold, leaf miner, and whitefly. The experimental results showed that the proposed algorithm has a high detection accuracy of 92.39% and a shorter detection time compared with other widely used methods such as single shot detector (SSD), fast region-based convolutional neural network (Faster R-CNN), and the original Yolo V3. Boulent et al. developed a CNN algorithm to automatically detect Flavescence Dorée symptoms across various white grapevine varieties but the true positive rates varied from 98.3 to 8.3% for different varieties which underlined the need for multi-varietal training dataset to capture the diversity of Flavescence Dorée symptoms. Mrisho et al. developed a smartphone-based object detection model called Nuru to detect symptoms of cassava mosaic disease, cassava brown streak disease, and the damage caused by cassava green mites. Results showed that Nuru can achieve a higher accuracy rate (65%) than cassava experts (researchers trained on cassava pests and diseases), agricultural extension officers, and farmers. Wang and Liu developed a multiscale parallel algorithm MP-YOLOv3 based on the MobileNetv2-YOLOv3 model. The parallel detection algorithm was used to effectively improve the detection performance of multiscale tomato gray mold lesions while ensuring the real-time performance of the algorithm. The experimental results showed that the proposed algorithm can accurately and real-time detect multiscale tomato gray mold lesions in a real natural environment. Zhu et al. utilized low resolution images of disease spots on grape leaves and used super-resolution image enhancement, which were fed into a YOLOv3-SPP network developed by them for grape leaf rot detection. The results showed that the developed method has better detection accuracy than a base CNN model, and it can be effectively used for detection of grape leaf black rot. Future works can use similar techniques for detection of small targets. Gu et al. developed a random forest model to identify different severity levels of wheat Fusarium head blight disease by fusing deep convolution features extracted from AlexNet and shallow features such as the color and the texture of wheat ears. Results showed that the proposed model has a higher accuracy than that of using either of the features alone and remains robust in terms of lights, angles, image resolutions, and crop growth periods. Yang et al. developed a model based on NASNetLarge classification model and an attention mechanism that emphasizes a target region of the image for the fine-grained classification of crop disease images. The proposed model achieved the best classification performance compared with other widely used machine learning models on different datasets. Lee et al. developed an attention-based recurrent neural network to automatically locate the infect regions for plant disease classification. The proposed model led to a better overall accuracy compared with two CNN models.

Phenotyping is an important process in the grand scheme of agricultural decision making. This process involves quantification of several plant physiological factors like plant height, plant nutrients, plant architecture, crop biomass, etc. To detect and count these factors, it requires a lot of manual effort and time. Several ground-based or remote sensing technologies combined with CNN models were used to save time and efforts required from the farmers and increase the accuracy of the detection and counting. Lin and Guo used unmanned aerial vehicles (UAVs) to acquire high resolution imagery of sorghum crop and developed a U-Net CNN model to detect and count sorghum panicles. The algorithm performed the task with an accuracy of 95.5% and a root mean square error of 2.5. Farjon et al. used two different deep learning techniques: one was a multiple scale regression and the other was a regression model that counts the leaves after locating leaf center points and aggregating them. They found that both the methods outperform the existing models in a leaf counting challenge. Hobbs et al. developed a density-estimation deep learning model based on a U-net backbone to detect flowering pineapple plants in a field. This deep learning technique runs on RGB images collected using an UAV. This allowed farmers to rapidly detect and count over 1.6 million flowering plants in a field. Hampf et al. demonstrated that a high-level accuracy in detecting pathogens and animal pests using CNN models can be achieved when it is paired with a citizen-science approach. More than 78,000 geo referenced images collected using Plantix app helped them identify yield gaps and advance the field of crop loss research. Islam et al. used a deep neural network called TheLNet270v1 to automatically classify thermal images to continuously monitor the canopy surface temperature inside a greenhouse. Liu and Wang developed an optimized YOLO-Dense algorithm based on DenseNet and YOLO V3 to detect anomalies in greenhouse tomatoes. Compared with SSD, Faster R-CNN, and the original YOLOv3 network, and the proposed model achieved the best performance in tomato anomaly detection under a complex natural environment. Apart from commonly used RGB images, there are studies that use of hyperspectral and laser spectral data combined with deep learning techniques to detect and count various factors. Zhou et al. used hyperspectral imaging of 30 different wheat varieties and developed a convolutional neural network with attention (CNN-ATT) model for wheat kernel variety identification. This model was evaluated and compared with other popular machine learning models such as SVM and partial least square discrimination analysis. The result showed that the CNN-ATT model achieved a higher accuracy than other models and the combination of hyperspectral imaging and the proposed CNN-ATT model has great potential for automatic non-destructive classification of wheat kernels. Wang W. et al. showed that the combination of laser induced breakdown spectroscopy and extreme learning machine can

TABLE 1 | Summary.

Tasks	Methodologies	Insights	Articles
Understanding the cause-effect of nutrients and the crop growth	Design of experiments (DoE), artificial neural network (ANN), genetic algorithm (GA), fuzzy logic	ANN with fuzzy logic can better unmask the interaction between nutrient and crop growth	Garcia-Perez et al., Hameg et al.
Disease diagnosis	New features combined with different classifiers, such as support vector machine (SVM), logistic regression (LR), and convolutional neural network (CNN)	The images of visible/near-infrared hyperspectral imaging (HSI), mid-infrared spectroscopy (MIR), and laser-induced breakdown spectroscopy (LIBS) can help to improve the performance	Feng, Wu, Zuh, et al., Feng, We, Hu, et al., Khalili et al., Yan et al.
	CNN	YOLO network attracts most attention, and CNN combined attention mechanism shows promising result for fine-grained classification	Boulent et al., Gu et al., Lee et al., Liu and Wang, Mrisho et al., Wang and Liu, Yang et al., Zhu et al.
Phenotype detection and counting	CNN	Most popular classification, detection and segmentation algorithms are studied, such as GoogleNet, VggNet, ResNet, YOLO, and U-Net	Bacong et al., Farjon et al., Ghahremani et al.,Hampf et al., Hobbs et al., Ilyas et al., Islam et al., Liu and Wang, Lin and Guo, Wang W. et al., Zhou L. et al., Zhou C. et al.,
Yield prediction	Random forest, SVM, neural network, CNN	An essential task is to develop a cost-effective way to collect data, and most research adopts UAVs	Apolo-Apolo et al., Johansen et al., Kasimati et al., Wang Y. et al., Yoosefzadeh-Najafabadi et al.

quickly detect the presence of cadmium in rice stems, and can accurately distinguish different degree of cadmium pollution. Bacong et al. applied CNN models to predict the entire chromatographic profile of sambong leaves under different time-varying environmental and laboratory variables parameters, which reduced the labor efforts in conventional method to evaluate herbal phytochemical composition.

Another type of technology used for phenotyping detection is segmentation. Segmentation is usually performed based on various visual factors of the crops involved to assess health, predict quality, maturity, or yield of the crop. These segmentation tasks can often be time consuming if done manually. Zhou C. et al. developed an imaging processing method, Improved ResNet, for segmenting and grading broccoli head in field conditions based on a deep learning architecture and color information. They compared this method with three other approaches, GoogleNet, VggNet, and ResNet, and illustrated that the Improved ResNet algorithm yields better results. They also showed that this algorithm can be used to obtain other data such as above ground biomass, yield, and biologic rhythm. Similarly, Ilyas et al. developed a segmentation network, named Straw-Net, for the segmentation of strawberries into four classes depending upon the ripeness of the fruit. In addition, they developed a real-time attention mechanism by integrating local and global semantic features efficiently. They compared the results with other existing segmentation models and their model demonstrated enhanced performance. For segmentation of point clouds of wheat, Ghahremani et al. developed Pattern-Net that can detect, classify, and measure features directly in the 3D point clouds with sufficient accuracy compared with manual phenotyping.

Crop yield prediction is an essential task for rapid decision-making. An accurate crop yield prediction model can help farmers to decide on what to grow and when to grow. They are also key to organizing harvesting operations. Data obtained using remote sensing technologies like RGB imagery, multispectral imagery, hyperspectral imagery combined with conventional machine learning methods can be used to predict crop yields. Johansen et al. collected UAV-based RGB and multispectral images during regular intervals prior to the harvest of tomato crops and used a random forest machine learning approach to predict end of season biomass and yield. Several shape features such as plant area, border length, width, and length were used as input variables. Both the multispectral and RGB imagery produced similar results with small differences in explained variance. Apolo-Apolo et al. used UAV-based RGB imagery to generate Orthomosaic and a region-convolutional neural network was trained to detect and count the number of apple fruit on individual trees. The results showed that an R^2 value of 0.86 is obtained when compared with an agrotechnician's numbers. With these results, they further used Google Colab to generate yield maps, which provide a visualization of the data collected. Yoosefzadeh-Najafabadi et al. collected hyperspectral imagery of soybean in two different growth stages for yield prediction. Three common machine learning algorithms, multilayer perceptron, SVM, and random forest, were evaluated for predicting soybean seed yield using hyperspectral reflectance. The results showed that the random forest algorithm combined with ensemble–stacking method had achieved the highest performance. Kasimati et al. investigated an alternative approach to predict wine grape quality by combining UAV-based multispectral images, satellite imagery, normalized difference vegetation index, and machine learning methods. This study investigated the combination of a selection of methods required for the strongest correlation with sugar content. Wang Y. et al. developed an improved EfficientDet-D0 object detection model for wheat ear counting to evaluate crop yield. To address the challenges due to the occlusion and overlap in wheat images, they used the Random-Cutout method

for image augmentation and a convolutional block attention module to refine the features. Experiments showed that the proposed model can improve the accuracy of EffientDet-D0 model by 2%.

Table 1 provides a summary of this Research Topic.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We would like to acknowledge the help of the reviewers and thank the authors for their contributions.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kakarla, Zhu, Ampatzidis, Fountas, Ehsani and Pardalos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

METHODS
published: 15 April 2020
doi: 10.3389/fpls.2020.00402

A Monitoring System for the Segmentation and Grading of Broccoli Head Based on Deep Learning and Neural Networks

Chengquan Zhou[1], Jun Hu[1], Zhifu Xu[1], Jibo Yue[2], Hongbao Ye[1] and Guijun Yang[3,4]**

[1] Institute of Agricultural Equipment, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2] International Institute for Earth System Science, Nanjing University, Nanjing, China, [3] Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture P. R. China, Beijing Research Center for Information Technology in Agriculture, Beijing, China, [4] National Engineering Research Center for Information Technology in Agriculture, Beijing, China

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Dimitrios Moshou, Aristotle University of Thessaloniki, Greece
Hao Gan, The University of Tennessee, Knoxville, United States
Grzegorz Cielniak, University of Lincoln, United Kingdom

***Correspondence:**
Hongbao Ye
yhb2008@zaas.ac.cn
Guijun Yang
guijun.yang@163.com

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *11 December 2019*
Accepted: *20 March 2020*
Published: *15 April 2020*

Citation:
Zhou C, Hu J, Xu Z, Yue J, Ye H and Yang G (2020) A Monitoring System for the Segmentation and Grading of Broccoli Head Based on Deep Learning and Neural Networks. Front. Plant Sci. 11:402. doi: 10.3389/fpls.2020.00402

Achieving the non-contact and non-destructive observation of broccoli head is the key step to realize the acquisition of high-throughput phenotyping information of broccoli. However, the rapid segmentation and grading of broccoli head remains difficult in many parts of the world due to low equipment development level. In this paper, we combined an advanced computer vision technique with a deep learning architecture to allow the acquisition of real-time and accurate information about broccoli head. By constructing a private image dataset with 100s of broccoli-head images (acquired using a self-developed imaging system) under controlled conditions, a deep convolutional neural network named "Improved ResNet" was trained to extract the broccoli pixels from the background. Then, a yield estimation model was built based on the number of extracted pixels and the corresponding pixel weight value. Additionally, the Particle Swarm Optimization Algorithm (PSOA) and the Otsu method were applied to grade the quality of each broccoli head according to our new standard. The trained model achieved an Accuracy of 0.896 on the test set for broccoli head segmentation, demonstrating the feasibility of this approach. When testing the model on a set of images with different light intensities or with some noise, the model still achieved satisfactory results. Overall, our approach of training a deep learning model using low-cost imaging devices represents a means to improve broccoli breeding and vegetable trade.

Keywords: growth monitoring, deep learning, improved ResNet, freshness grading, ground-based imaging system

INTRODUCTION

Broccoli (*Brassica oleracea* L. var. *italica*), which is belongs to the genus *Brassica* in the family Cruciferae, is considered as an important global vegetable crop. The present broccoli cultivation area in China is about 140,000 hectares. Broccoli is an important export vegetable in China, especially in Zhejiang Province (Kalisz et al., 2015). The broccoli head is an important agronomic component, which is not only used for yield estimation but also to assess plant quality and analyze

plant resistance to biotic and abiotic stresses (Guo et al., 2017). A variety of techniques have been used for the quantitative measurement of broccoli head, including the destructive measurement of geometric parameters and dry weight, mass spectrometric analysis, and techniques using non-contact sensors. Destructive techniques are not suitable for the measurement of broccoli head under a controlled environment due to its low-throughput and unsustainability. Mass spectrometric analysis has been applied to measure the quality and freshness of vegetation, however, the wide application of this method is restricted by its high cost (Cho et al., 2018). The sustainable monitoring of broccoli head can be achieved through various technological innovations such as non-contact sensors. Non-contact sensors, which are principally based on digital RGB cameras, are suitable for use in agriculture due to their high resolution, low cost, and small size. The use of RGB cameras can provide a non-invasive and high-throughput approach to collect morphological information about broccoli head and analyze its health status (Dell' Aquila, 2009). Changes in soil reflection and weather conditions, as well as the transition between growth stages, will all cause differences in the reflectivity, size, shape, and color of a broccoli canopy. The existing segmentation methods based on RGB images are of two main types: (1) those solely based on color information; and (2) those based on multi-features and a trained model (Hamuda et al., 2016). For example, for the first type, Ji et al. (2007) presented a real-time segmentation algorithm for plant images under natural outdoor conditions by using a threshold-based method. Their experimental results demonstrated that segmentation was generally of good quality in the case of bare soil background. Furthermore, Wang et al. (2013) established relationships between image feature parameters and several plant indexes by setting threshold values based on magnitude. The high correlation coefficients (over 0.9) which were achieved between the segmented canopy cover and the selected plant indexes indicate that this technique could be used to estimate the nitrogen status of vegetation. However, due to the complexity of field environments, color information will be seriously affected by the illumination intensity or shadows. In the second type of segmentation method, regions of interest are generated by extracting multi-features and training a classifier. Islam et al. (2017) used a support vector machine (SVM) classifier to realize the automated detection of potato diseases with an accuracy of 95%, which presented a path toward the automated diagnosis of plant diseases at a very large scale. Additionally, Yang et al. (2016) proposed a novel plant-inspired optimization algorithm which essentially mimics iterative root foraging behaviors, named the "hybrid artificial root foraging optimizer," in order to determine root density. Cates et al. (2004) developed a novel leaf-segmentation tool by combining *a priori* information with local images showing the orientations of leave. The approach of Cates et al. (2004) achieved higher accuracy compared to three other state-of-the-art segmentation techniques. Moreover, Chantal et al. (2014) used an improved method based on a non-destructive and high-throughput machine learning method to accomplish non-contact analysis in order to measure root architecture. Previous methods for the segmentation of plant utilized handcrafted features such as shape, color, and texture to quantify the pixel character of plants. Extracting such features often requires some theoretical knowledge of botany and a computationally expensive preprocessing step in order to enhance differences between plants and background, i.e., an image binarization step (Wang and Xu, 2018). To allow simple and effective segmentation, most studies based on botanical theories use images captured in a controlled environment with a uniform background.

Furthermore, in recent years, convolutional neural networks (CNNs) have matured and have revolutionized computer vision. Currently, CNNs achieve superior results in the identification and segmentation of plants compared with state-of-the-art traditional methods. CNNs have been used to improve the performance of the approach for identifying and counting plant species, to quantitatively phenotype plants grown in controlled environments, and to provide detailed quantitative characterization of fruits and leaves. For example, Jeon et al. (2008) proposed an image segmentation method to detect individual weeds using color space transformation, threshold calculation, and the training of an artificial neural network (ANN). Additionally, Yamamoto et al. (2014) developed a method for the accurate detection of individual intact tomato fruits by using a conventional RGB digital camera in conjunction with machine learning approaches. Moreover, Xiong et al. (2017) used a rice panicle segmentation algorithm, called Panicle-SEG, to segment and calculate of rice panicle, and performed phenotyping of rice panicle to assist rice breeding. The overall accuracy of their model (more than 0.89) demonstrated the practical utility of their model for the estimation of field yield. Furthermore, Zhang and Xu (2018) introduced an unsupervised field image segmentation algorithm, called Unsupervised Learning Conditional Random Field (ULCRF), to accelerate the unsupervised segmentation of greenhouse plant images at the organ level. Kusumam et al. (2017) describe a 3D vision system for robotic harvesting of broccoli using low-cost RGB-D sensors. Ramirez (2006) develop a computer vision algorithm to locate the broccoli head within an image of an entire broccoli plant and has the ability to distinguish between mature and immature broccoli heads. However, despite the improvements which have been achieved in deep learning, the accurate segmentation and grading of broccoli requires a large amount of training data and depends on the quality of the images used. For field-based imaging and analysis systems, it is important to overcome the lack of training data and low processor capacity to achieve a high-throughput phenotype acquisition process (Lee et al., 2018).

Therefore, the aim of this study was to develop an automatic method to segment near-ground RGB images of broccoli field in order to extract broccoli head and construct a high-throughput grading standard. Some major advantages of this approach are that it requires only a few images and reduces the data volume and memory requirements for the image processing. This approach also allows the use of deep learning technology which is not specific to agriculture and plant phenotyping. Compared with three other approaches, the evaluation results showed better performance regarding the segmentation and grading accuracy.

MATERIALS AND METHODS

Experimental Setup

All experiments were conducted at the Zhejiang Academy of Agricultural Sciences (ZAAS) Yangdu Scientific Research Innovation Base, Haining County, Zhejiang Province, China (latitude 30°27′ N, longitude 120°25′ E). Soil samples were collected from a depth of 0–20 cm with a pH range from 4.5 to 6.5 and organic matter content of more than 30 g kg^{-1}. Other soil chemical properties of the experimental area are as follows: the available phosphorous content was 20 mg kg^{-1}, the content of rapidly available potassium was 300 mg kg^{-1}, and alkali-hydrolyzable nitrogen was 300.2 mg kg^{-1}. The broccoli samples consisted of three varieties, named *Zheqing95*, *Tailv1*, and *Tailv2* which are the main broccoli planting varieties in Zhejiang Province, and each variety was planted in a separate plot. All plants were directly artificially seeded on 15 September, 2018. The three experimental plots consisted of raised beds with a length of 20 m and a width of 1 m with the seed lines separated by 0.3 m. The inter-row distance was 0.8 m and the inter-plant distance was 0.35 m.

Remote Monitoring System and Image Acquisition

The vision system was composed of an Canon EOS 90D camera (Canon, Inc., Tokyo, Japan) with a resolution of 32.5 megapixels (6960 × 4640 pixels) and a 22.3 mm × 14.8 mm CMOS sensor, a uniform LED surface light source, and a Surface Laptop 2 computer (Microsoft, Corp., Redmond, WA, United States) with an Intel Core i7-8650U processor and 8 GB Samsung DDR4 memory. All of these devices were mounted on a semi-automatic field self-propelled platform with dimensions of 1.5 m × 1.5 m × 1 m. The component elements of the surface light source were 120 white LEDs (3000 K) arranged in a circular pattern. The camera was controlled by an electronic shutter connected to the computer by a USB 3.0 interface. The computer installed with a visualization software named "pylon Camera Software Suite" (BASLER, Inc., Ahrensburg, Germany) was used to monitor the quality of the images in real time and guarantee the data quality. The camera was located 1.5 m from the ground level with the focal length of 18 mm and the exposure time of 1/500 s. The platform moved with a speed of 1 m/s, and the field of view was about 0.2 m^2, which generated an image sequence with an average overlap of more than 70%. Two data acquisitions were conducted for each of the three plots. In total, 506 images containing grown broccoli plants were captured, and these were used to build our original image datasets. Among these images, 300 images were obtained during our first data acquisition, which was called "T1," while the rest of the images were captured during the second acquisition, which was called "T2." In this way, we can make our dataset contains plant images with different flower-ball shape, color distribution and hollow degree (**Figure 1**).

Image Preprocessing

The plant images captured by the self-developed monitoring system were pre-processed to denoise the background and enhance the images. This pre-processing was performed via several python-based scripts (**Figure 2**).

Background Denoising

Field images typically contain various sources of noise, which will affect the final training results. In order to remove high-frequency noise from the images, a Robert detection operator was applied to extract the edge of the broccoli image (Chaudhuri and Chanda, 1984), followed by a median filter with a size of 3*3 pixels to remove the noise from the images (according to the size of broccoli head and flower bud displayed in the images) (Zheng et al., 2017).

Data Augmentation

A sufficiently large training dataset is essential to improve the final accuracy of deep-learning projects [19]. However, in our study of broccoli heads, it was not possible to capture enough images due to the limitations of the indoor environment. Therefore, in order to improve the quantity and quality of the training images, data enhancement methods were adopted to expand the dataset by 12 times (show in **Table 1**).

Random cropping

First, the original images were resized to 1440 × 1080 pixels due to the memory limitation of the GPU. Then, three sub-images with a size of 480 × 360 pixels were randomly cropped from each image. Thus, the number of training images was tripled.

Rotation

In order to further increase the number of training images, the cropped images were rotated by 90, 180, and 270 degrees, respectively, thus increasing the size of the dataset by a further four times.

Fancy principal component analysis

The last step in the data enhancement procedure involved applying a fancy principal component analysis (FPCA) algorithm to change the intensity of the RGB channels in order to enhance the contrast between the broccoli heads and the background (Morais et al., 2019). The feature vectors were generated by the FPCA algorithm, and then a weight factor was added to the corresponding channel according to the extracted feature vector. The use of FPCA can transform the illumination intensity and color of the image without affecting the object to be recognized.

Data Analysis

The purpose of this study was to develop a general system for the automatic detection and grading of broccoli head. This system requires the input of orthophoto images of the field canopy and outputs the segmentation results generated by a novel "Improved ResNet" model and grading results determined by a pixel clustering method. The overall flow of the method is shown in **Figure 3**. Our approach comprises three steps: (1) detect the positions of the broccoli heads and calculate the orthophoto projection area; (2) establish a yield estimation model based on the correlation between flower-ball area and weighing results; and (3) conduct flower-ball grading based on a pixel clustering algorithm.

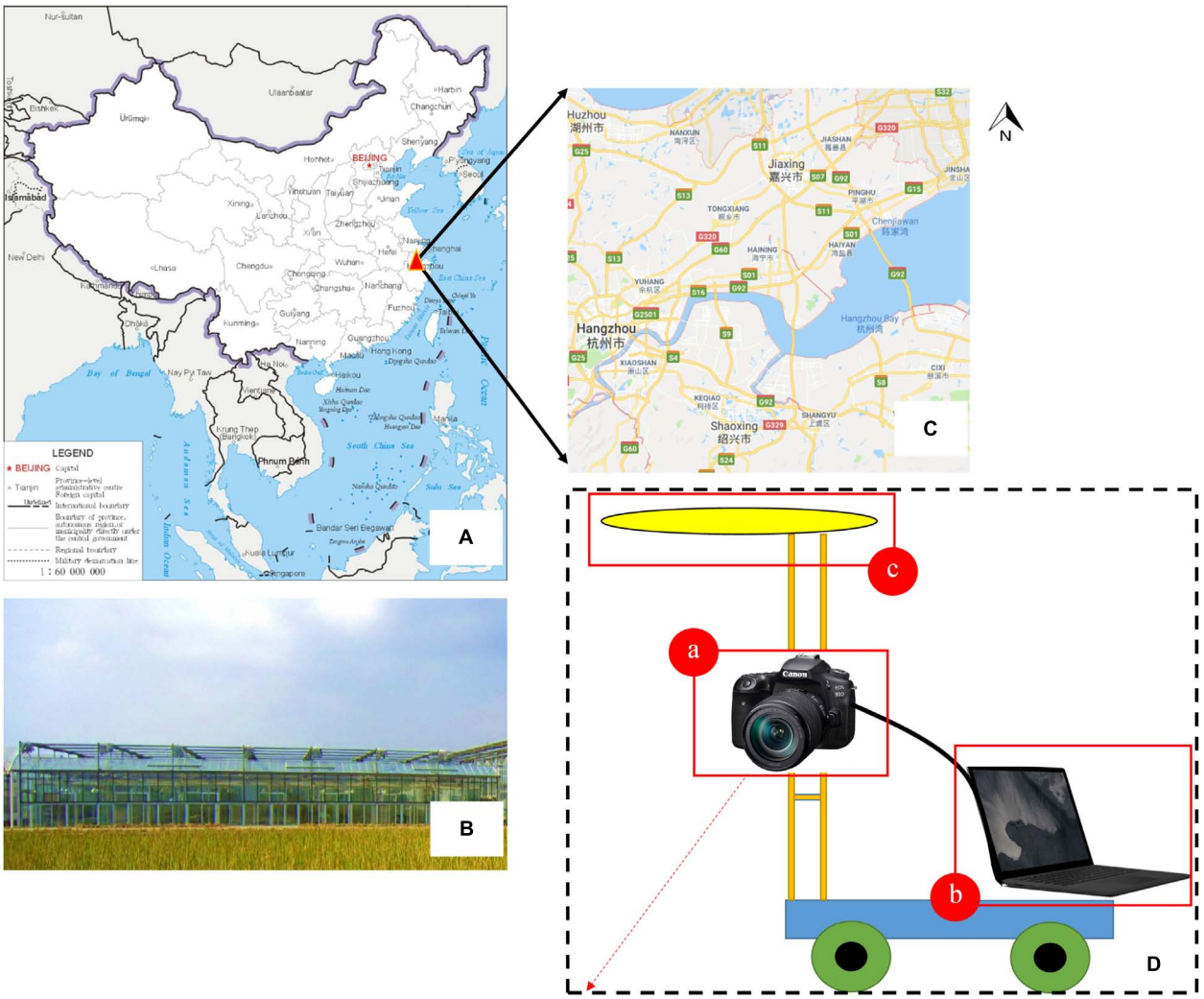

FIGURE 1 | Part of the trial design. **(A–C)** Test site location and the appearance of the greenhouse; **(D)** field imaging system and integrated sensors—**(a)** Canon EOS 90D digital camera, **(b)** Surface Laptop 2 computer, **(c)** uniform LED surface light source.

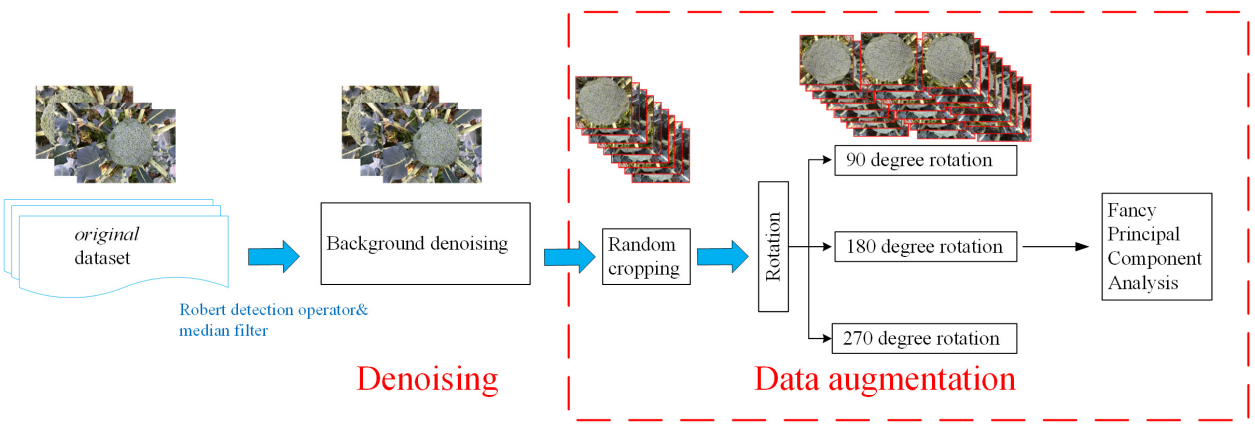

FIGURE 2 | Flowchart of the image preprocessing.

TABLE 1 | Dataset configuration.

Variety	Original	Augmentation	#Train	#Val/Test
Zheqing95	165	2000	1500	250/250
Tailv1	171	2000	1500	250/250
Tailv2	170	2000	1500	250/250

Architecture of the Training Models

The growing broccoli heads were detected using a CNN named "Improved ResNet." In order to verify the accuracy of the model, three classical deep learning architectures were used for comparative experiments, namely GoogleNet, VGG16, and ResNet 50. The open-source codes of these approaches were implemented under the TensorFlow framework.

GoogleNet

GoogleNet reduces the number of filters and training parameters which are required compared with the traditional Inception Structure (Tang et al., 2016). Additionally, it maximizes the depth and width of the network and divides the multidimensional convolution layer in the Inception Module into several smaller one-dimensional convolution layers by decomposition factor, which not only reduces the number of parameters in the model, but also effectively avoids over-fitting.

VGGNet

The VGG network constructs a CNN with a depth of 16/19 layers by repeatedly stacking small convolution cores with a size of 3*3 and maximum pooling layers with a size of 2*2 (Mehdipour Ghazi et al., 2017). In this study, the VGG16 network model was adopted; the number 16 denotes the number of layers using convolution layer besides pooling layer. VGG16 uses convolution blocks consisting of 2–3 convolution layers so that the network has more receptive fields and fewer network parameters, and also uses a Rectified Linear Unit (ReLU) activation function to perform numerous linear transformations to achieve greater learning ability.

ResNet

ResNet is a complete network formed by the repeated accumulation of residual learning modules (He et al., 2016). The introduction of residual modules solves the problem of gradient dispersion and enhances the feature-learning ability and recognition performance. The structure of the residual modules is shown in **Figure 4**. Set x as the input and F (x, W_1, W_2) as

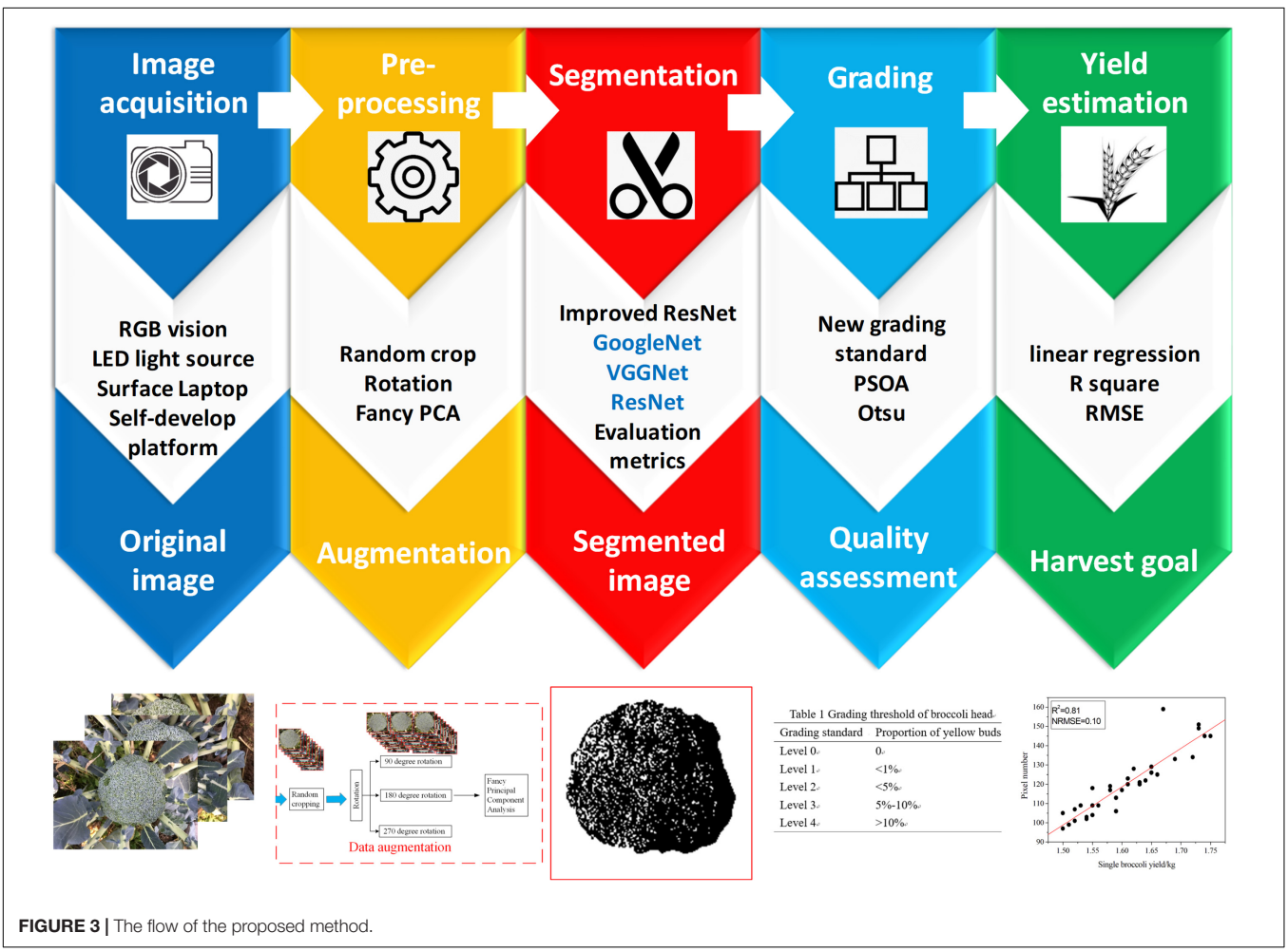

FIGURE 3 | The flow of the proposed method.

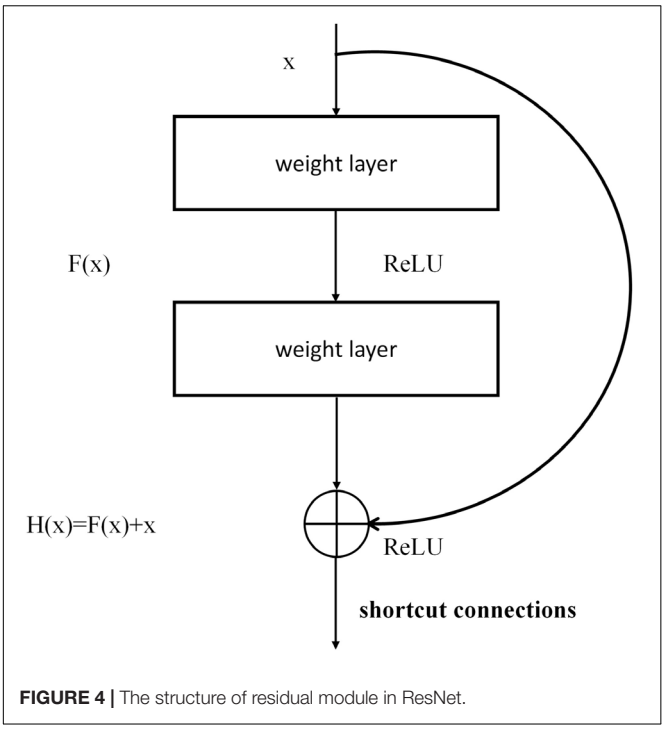

FIGURE 4 | The structure of residual module in ResNet.

the output after the convolution of W_1 and W_2 (the weighting parameters to be learned). The activation function is set as ReLU, so the final output of the residual module unit y can be expressed as follows:

$$y = F(x, W_1, W_2) + W_s x \quad (1)$$

where W_1 and W_2 represent the weighting parameters to be learned, and W_s represents a square matrix that transforms x from the input residual module dimension to the output dimension.

Improved ResNet

In supervised learning mode, a large amount of data is needed to train the residual network model. However, at present, few broccoli images are available with labels, which cannot meet the needs of training deep network models. Therefore, in order to improve the accuracy and generalization ability of the ResNet-50 model, a feature- based transfer learning was adopted which combines transfer learning and deep learning. First, ImageNet was used to pre-train the ResNet-50 network to allow it to extract image features, and the trained network parameters were used as network models. Then, broccoli-head images were precisely segmented by adjusting the parameters of the ResNet-50 network. A three-layer adaptive network was used to replace the full connection layers and the classification layers of the ResNet-50 model, and the LReLUSoftplus was adopted as the activation function of the architecture. The formulas for ReLU and LReLUSoftplus are given as follows.

$$f(x) = \max(0, x) \quad (2)$$

$$f(x) = \begin{cases} \ln(e^x + 1) - \ln 2, x \geq 0 \\ \text{ax}, x < 0 \end{cases} \quad (3)$$

where x indicates the input value, and a is set as 0.01. A general scheme of the proposed method is shown in **Figure 5**. M_1, M_2, M_3, and M_4 are the four residual blocks in the ResNet-50 model, while N_1, N_2, and N_3 are the three components of the adaptive network. In order to enrich the extracted features, all residual blocks were deconvoluted to get the corresponding features before the residual block convolution, then the deconvolution features were fused by using weighted fusion method.

The hyper-parameters for all experiments were as follows: the loss function was set to dice loss due to this function's good performance in dichotomous problems; the base learning rate was 0.001 in the first 3000 iterations and was changed to 0.0005 in the subsequent 2000 iterations; the value of momentum and dropout were 0.9 and 0.8, respectively; and the number of epochs was 200 and the batch size was 64. We set the ratio of training images, validation images and test images to 6:1:1 in order to ensure the credibility of the training results.

Yield Estimation

After the canopy area of broccoli-head had been measured using the deep learning approach, a yield estimation model was developed by calculating the relationship between the weight of head balls (mean fresh matter of the broccoli heads within 2 h after harvest) and the segmentation area using

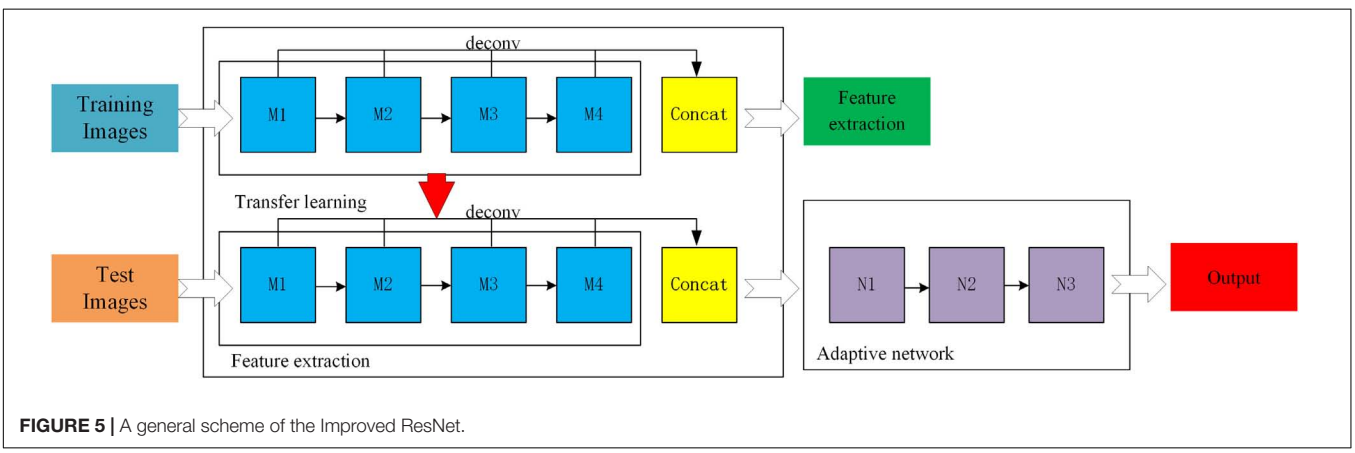

FIGURE 5 | A general scheme of the Improved ResNet.

regression analysis. Here, we treat the mean fresh weights of the broccoli heads and the mean value of the projected area of the corresponding variety as the independent and dependent variables of the formulas, respectively.

Grading System

An artificial grading standard for broccoli based on the sensory yellowness of the flower-ball was adopted from Tu et al. (2007). The details of the grading standard are shown as follows:

Level 0: unable to detect yellow flower buds;
Level 1: 1–3 yellow flower buds detected;
Level 3: detect 5% yellow area;
Level 5: 50% yellow area detected;
Level 7: yellowness degree between 50 and 75%;
Level 9: all the detection area displayed yellow color;

The yellowness degree of broccoli head can be calculated by Eq. 4:

$$Y = \sum M_1 * N_1 / M_2 * N_2 \tag{4}$$

where Y represents the degree of sensory yellowness, M_1 represents the level of each broccoli under our grading system, N_1 is number of the broccoli in the corresponding level, M_2 is the highest level, and N_2 is the total number of observed broccoli. Since it was only necessary to distinguish two types of color in our research (green and yellow) and these two parameters were completely different, it was convenient to separate them by using a threshold segmentation technology. In this paper, an optimized Otsu method was used to transform the grayscale images of the head ball into two parts and calculate the number of black and white pixels respectively to determine the proportion of yellow area. In our practical experience, it was not necessary to achieve such precise grading since broccoli heads with more than 10% yellowness area has no commercial value to any consumers. We merged Level 4 (10% yellow area detected) to Level 9 and then divided all the broccolis into five levels according to the new standard (**Table 2**).

Data Annotation

For the supervised learning algorithm, the quality of ground truth determines the accuracy of the final results. In the field of image segmentation, in addition to some cases where open datasets are available, many application scenarios require specialized datasets for migration learning or end-to-end training. The methods for constructing ground truth datasets can be separated into three categories: manual labeling, automatic labeling, and outsourcing labeling. Among them, automatic labeling usually requires a second review to avoid program errors, and outsourcing labeling introduces the risk of data leakage and loss. Meanwhile, manual labeling is usually time consuming, although the results are relatively reliable. In our study, the manual works were conducted by four people using the Labelme tool to draw curved lines to precisely segment the broccoli heads. The code of this tool[1] is open source so that it can be used by anyone to build a labeled training dataset.

TABLE 2 | Grading threshold for broccoli head.

Grading standard	Proportion of yellow buds
Level 0	0
Level 1	<1%
Level 2	<5%
Level 3	5–10%
Level 4	>10%

Evaluation Index

The performance of our segmentation model was evaluated using five different metrics: (1) Accuracy; (2) Precision; (3) Intersection over Union (IOU); and (4) Recall and (5) F-Measure. "Accuracy" is the proportion of correctly extracted flower ball pixels to the total number of pixels. The higher the value (approach to 1), the more accurate the segmentation is. IOU is applied to describe the degree of ratio of intersection and union of real and predicted values. Precision and Recall can be used to reveal accuracy and the completeness of the segmented region. These two indexes interact with each other, and the F-measure was used to balance them. The computational formulas of these five evaluation indexes are shown in Eqs 5–9:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$IOU = \frac{Predicted \cap GroundTruth}{Predicted \cup GroundTruth} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

where, TP, TN, FP, and FN in Eqs 5–7 represent the numbers of true positives, true negatives, false positives, and false negatives, respectively. Among them, the true positives represent the extracted pixels and the corresponding ground truth which both belong to the flower ball region. True negatives are when extracted pixels and the corresponding ground truth are both background pixels. False positives are when the pixels are classified as flower ball pixels but the ground truth results display them as background pixels. False negatives represent the background pixels that are not correctly discriminated. The "*Predicted*" in Eq. 6 represents the prediction results achieved by these segmentation algorithms.

RESULTS AND DISCUSSION

The experiment was conducted on images of growing broccoli which were captured by a camera mounted on a self-developed near-ground imaging system equipped with a series of auxiliary imaging devices. Using our post-processing system (Microsoft

[1]https://github.com/wkentaro/labelme

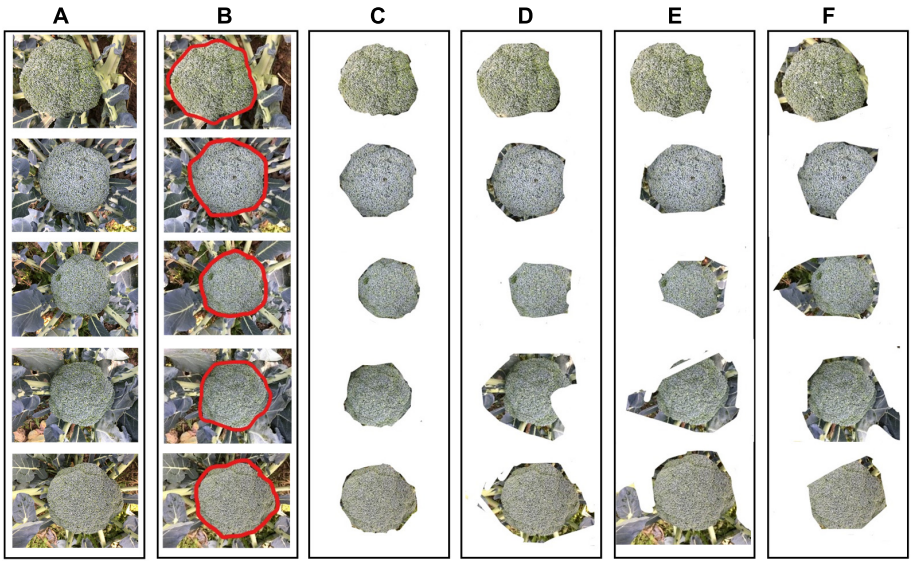

FIGURE 6 | The results of broccoli head segmentation using different approaches. **(A)** Original images, **(B)** annotation results **(C)** segmented by Improved ResNet, **(D)** segmented by GoogleNet, **(E)** segmented by VggNet, and **(F)** segmented by ResNet.

Windows 10 Professional operating environment with a 12-core Intel Core i7-8700K CPU, 16 GB of memory, and an NVIDA GTX 1080Ti video card), the segmentation process for one test image with a resolution of 1440*1080 only takes 1.5 s. Moreover, by applying compute unified device architecture (CUDA) parallel acceleration model, the average processing time of a single image could be increased to 0.7 s. The performance of the developed deep learning method was evaluated using the evaluation metrics mentioned in Section "Evaluation Index," by comparison to manual ground truth results. All of the models presented in this paper are based on Python.

Broccoli Head Segmentation

The broccoli segmentation method was tested using the whole test set. Some of the results are presented in **Figure 6**. In **Figure 6**, three representative testing images were chosen to show the segmentation results obtained using different approaches. For each broccoli, the original image is shown in Line A and the segmentation results for each compared model are shown in Lines C–F. To demonstrate the accuracy and robustness of deep learning for plant monitoring, the evaluation metrics of Accuracy, Precision, IOU, Recall, and F-Measure were analyzed, as shown in **Figure 7**. In this figure, the color columns represent the mean value of the evaluation metrics. The color differences between these columns represent the various indicators. All the statistical analysis was performed using the SPSS 19.0 software (IBM, Inc., Armonk, NY, United States) (**Figures 6**, **7**).

As shown in **Figures 6**, **7**, for Improved ResNet, the average Accuracy and Precision for the test images are about 0.896 and 0.897, which are higher than the values obtained using the three other contrast algorithms. Additionally, the proposed method can achieve better consistency with labeling results. Furthermore, the high IOU of 0.901 of the proposed method shows a high overlapping rate between the candidate region and the labeling area. The IOU values of the other CNN models were as follows: GoogleNet, 0.801; VggNet, 0.799; and traditional ResNet, 0.832. A mean value of Recall of 0.879 can be achieved by using the Improved ResNet, compared with 0.721, 0.744, and 0.813 for GoogleNet, VggNet, and ResNet, respectively. Moreover, as illustrated in the figure, a higher F-measure can be obtained using Improved ResNet (F-measure is a comprehensive indicator that accounts for Recall). Compared with the other approaches—which achieved relatively lower F-measure values, with values of 0.751, 0.758, and 0.838 achieved using GoogleNet, VggNet, and ResNet, respectively—the Improved ResNet had a higher mean F-measure (0.899) with a lower standard deviation. This shows that the proposed algorithm was able to accurately distinguish broccoli head regions from the background region and guarantee the integrity of flower-ball structural information.

Yield Estimation Results

All of the broccolis were harvested on 6 January 2019, and the quality of each plant was recorded by skilled workers. The test set contained a total of 100 broccolis, which were arranged and numbered in order of weight and then split into an odd group and an even group, which were used for modeling and verification, respectively. In this study, linear regression was used to realize non-destructive production estimation. The number of pixels occupied by the head of broccoli as obtained by the segmentation algorithm was used as the independent variable, and the quality of a single broccoli was treated as the dependent variable. To quantitatively assess the performance of the linear regression algorithm, various popular elevation metrics were computed, namely the determination coefficient (R^2) and the normal root-mean-squared error (NRMSE). These indexes have been widely used to estimate the predicative power of regression models.

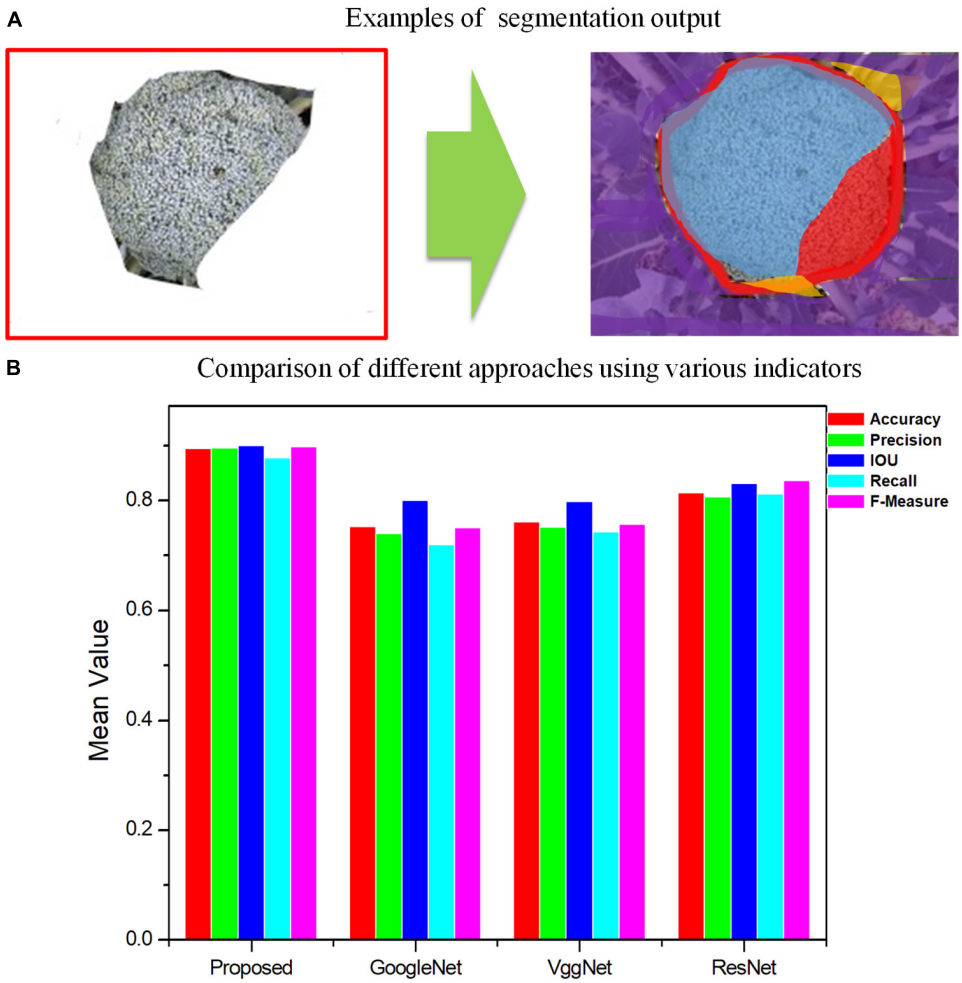

FIGURE 7 | Comparison of different approaches by segmentation quality for GoogleNet, VggNet, ResNet, and the proposed method, Improved ResNet. **(A)** Examples of segmentation output. Blue region-TP, Red region-FN, Orange region-FP, Purple region-TN**. (B)** Comparison of different approaches using various indicators.

A larger R^2 indicates a better fitting of the model, while a smaller NRMSE indicates a better estimation. We defined R^2 and NRMSE as follows:

$$R^2 = \frac{1 - SSE}{SST} \tag{10}$$

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}[Y(i) - Y'(i)]^2}}{\bar{Y}} \tag{11}$$

relating the obtained number of pixels with the quality of each broccoli. In Eq. 10, *SSE* represents the *Sum of Squares for Error* and *SST* represents the *error sum of squares*. In Eq. 11, N and $Y'(i)$ correspond to the number of samples and the actual value of sample i, respectively; $Y'(i)$ represents the estimated value of sample i; and $\bar{Y}$ is the average actual value of the sample. For regression analysis, the number of pixels identified for the three species of broccoli was selected as the independent variable, and the corresponding individual weight was used as the dependent variable. The regression model conducted with our test set is shown in **Figure 8**.

As illustrated in **Figure 8A**, for *Zheqing95*, the number of pixels was strongly correlated with manual measurements of plant quality (adjusted R^2 = 0.90), which is consistent with the result for NRMSE (adjusted NRMSE = 0.07). As shown in **Figure 8C**, the lowest R^2 occurs with the largest NRMSE of *Tailv2*. These differences seem to be strongly species-specific, even in the same growth stage. Compared with the accurate quality results, the image-based yield estimation had two additive error sources: (1) the head region was covered by leaves or insects; and (2) misclassification of head pixels and background pixels. In the first situation, the estimated production may be higher than the actual production due to the loss of pixel statistics. This is a technical limitation, and the RGB camera cannot remove the occluded objects from a certain view. A possible solution is to use opening and closing operations to remove these small holes in the head; however, this may not suitable for broccolis with large flower buds. An alternative solution is to recognize the leaves,

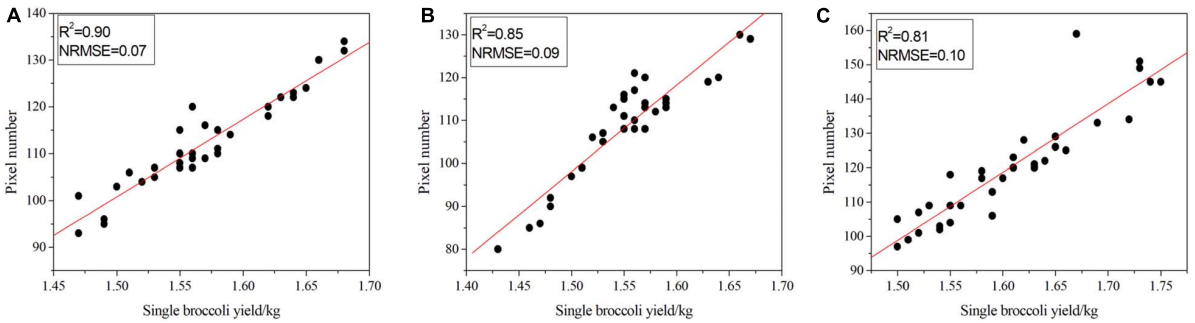

FIGURE 8 | Correlation between pixel number and the quality of broccoli. **(A)** Correlation results for *Zheqing95*, **(B)** correlation results for *Tailv1* and **(C)** correlation results for *Tailv2*.

etc., independently, which will greatly increase the complexity of the model. In the second situation, the number of pixels that presented a certain plant is larger than the real value. This issue can be solved by providing more training samples and fine-tuning the hyper-parameters to improve the accuracy of the pixel classification process.

Grading Results Based on Image Analysis

After applying grayscale transformation and the Otsu algorithm, the fine classification of the broccoli-head pixels was carried out. In this study, the optimal threshold for each image was determined by using a traversal algorithm within a small range. The Particle Swarm Optimization Algorithm (PSOA) was used to determine the fitness function and fitness parameters (Wang et al., 2017). Based on the image analysis results and the new grading standard, we provide reliable grading results for broccoli quality (**Table 3**).

Then, we tested the accuracy of our method by comparing our grading results with manual annotation results. It was found that, using the grayscale transformation and improved Otsu algorithm, more than 80% of broccoli heads were graded correctly; specifically, for *Zheqing95*, *Tailv1*, and *Tailv2*, the prediction Accuracy was 0.879, 0.853, and 0.841, respectively compared with manual annotation results. Therefore, the performance of the proposed method is appropriate for practical use. Further research should focus on the introduction of roundness information and bud number to build a regression model (**Table 4**).

Further, an experimental comparison result was provided to show how much segmentation quality affects the grading results. With other approaches displayed relatively lower values, the Improved ResNet had higher mean value of evaluation indices which indicate that the better segmentation quality could improve the accuracy of grading results (**Table 5**).

TABLE 3 | Grading results for broccoli quality obtained by using the Particle Swarm Optimization Algorithm (PSOA) and the Otsu algorithm.

	Level 0	Level 1	Level 2	Level 3	Level 4	Accuracy
Zheqing95	31.2%	29.7%	12.5%	13.4%	13.2%	0.879
Tailv1	38.9%	37.7%	11.6%	7.8%	4.0%	0.853
Tailv2	25.1%	22.4%	25.3%	10.4%	16.8%	0.841

TABLE 4 | Comparison of the analysis accuracy of the quality of broccoli head based on Particle Swarm Optimization Algorithm (PSOA) and the Otsu algorithm in different varieties.

	Zheqing95	*Tailv1*	*Tailv2*
Accuracy	0.879	0.853	0.841
IOU	0.855	0.844	0.843
Precision	0.825	0.894	0.899
Recall	0.838	0.841	0.873
F-measure	0.851	0.865	0.895

TABLE 5 | Comparison of the analysis accuracy of the quality of broccoli head using different segmentation models.

	Improved ResNet	GoogleNet	VggNet	ResNet
Accuracy	0.867	0.814	0.827	0.835
IOU	0.848	0.801	0.789	0.811
Precision	0.877	0.798	0.767	0.825
Recall	0.858	0.799	0.772	0.813
F-measure	0.867	0.768	0.797	0.809

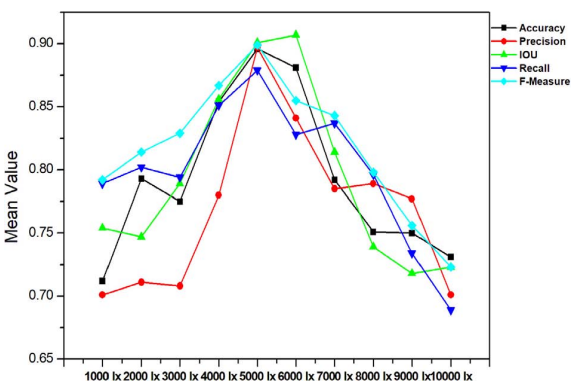

FIGURE 9 | Performance of different segmentation methods under different light condition.

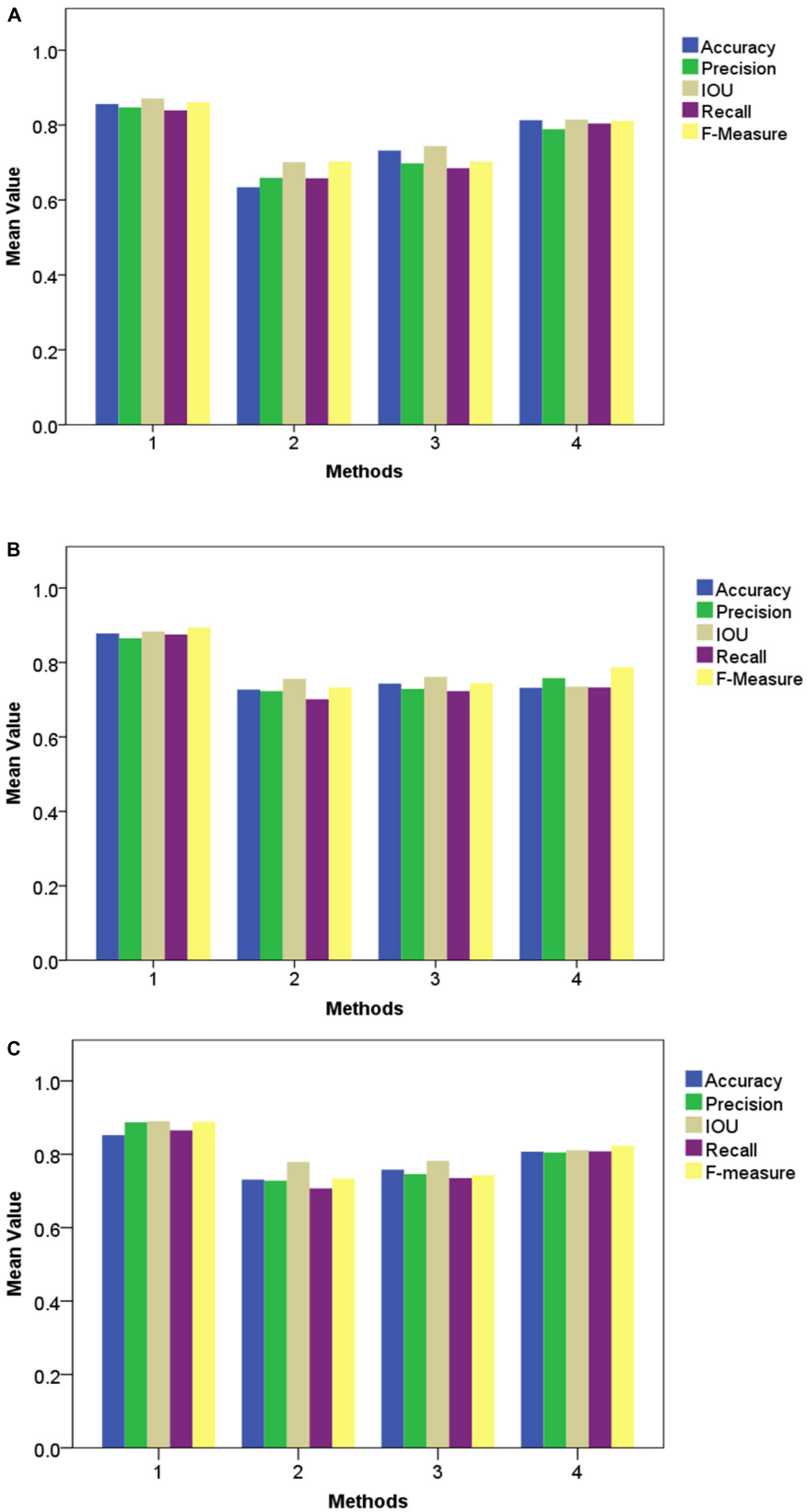

FIGURE 10 | Performance of different segmentation methods under various types of noise interference. **(A)** Gaussian noise, **(B)** Salt and Pepper noise and **(C)** Rayleigh noise.

Robustness and Efficiency Analysis

The accurate analysis of plant images acquired outdoors is a challenging task for researchers. Compared to pot experiments in the laboratory, segmenting field-grown vegetation is more complex due to varying light intensity, high specular reflectance, and ambient noise; each of these lead to the reduction of segmentation accuracy. Therefore, in order to be appropriate for the collection of plant phenotyping information, the segmentation method should be sufficiently robust to handle these unfavorable conditions. **Figures 9**, **10** present metrics indicating the quality of segmentation (Accuracy, Precision, IOU, Recall, and F-Measure) for the proposed method under different light condition (from 1000 to 10,000 lx) and under various types of noise interference (Gaussian noise, Salt and Pepper noise, Rayleigh noise). In comparison, GoogleNet, VggNet, and ResNet presented a similar analysis throughout our experiment.

As presented in these two experiments, although the three traditional architectures (GoogleNet, VggNet, and ResNet) can achieve satisfactory results under moderate brightness conditions, they performed poorly overall, particularly in circumstances with excessive or insufficient incident light. Additionally, the segmentation accuracy of GoogleNet and VggNet are greatly affected by Gaussian noise, while the traditional ResNet achieved much lower Recall and F-Measure when Salt and Pepper noise was introduced. The Improved ResNet achieved the highest mean Accuracy and the highest Precision, IOU, Recall, and F-Measure of the four compared methods. It must be noted that, we conducted all of our experiments in a confined environment (greenhouse etc.) because of the limitation of stability and power supply of our image acquisition platform. In other words, it is not a comprehensive robustness test for use under outdoor conditions. After the vehicle is further improved, we will apply it to outdoor conditions.

Moreover, the proposed algorithm achieved an average running time of 0.18 s for segmenting a single image, and thus represents a high-throughput processing method to measure the size of broccoli heads to inform decision-making in large-scale breeding. In our experiments, the 60 m^2 field contained 1000s of broccoli plants, and the proposed method was able to calculate the area of the flower ball and grade all the plants within 30 s. Considering the target number of broccoli plants and image size, the Improved ResNet can adequately estimate biomass or yield for online measurement. The running time for each image ranges from 0.13 to 0.20 s. Additionally, we calculated the average time required for each step in the Improved ResNet procedure, and found that the most time-consuming step was segmentation. In our future work, we will attempt to improve hardware (including the I/O speed of the CPU and memory) in order to reduce the segmentation time.

Limitations and Future Work

The imaging and processing procedure presented in this paper resulted in broccoli-head images with a high resolution, and allows the dynamic monitoring of the growth of individual broccoli plants at different growth stages. It must be noted that, the processing pipeline contains many steps which rely on manual settings and tuning which questions the wider applicability of the presented system. In the future, methods could be developed to improve the segmentation process in an attempt to eliminate the need for image labeling, such as semi-supervised learning (SSL) technology (Ma et al., 2015). In cases when there is only a small number of labeled samples, unlabeled samples can be labeled based on the similarity between unlabeled samples and labeled samples and the potential distribution of unlabeled samples and other strategies which will reduce the workload of annotation. Additionally, higher-resolution cameras would improve the overall process, albeit at the cost of increased processing time. Moreover, it is conceivable that an industrial robot could be used to perform both broccoli monitoring and automated harvesting based on the results of image processing, however this may become excessively cost prohibitive.

CONCLUSION

In this study, we establish a robust image processing method for segmenting and grading broccoli head in field conditions based on a deep learning architecture and color information. Compared with three other approaches—namely GoogleNet, VggNet, and ResNet—the proposed Improved ResNet algorithm has better segmentation performance and grading accuracy. Moreover, our model was tested under different light intensities and noise categories to confirm its applicability. Realizing accurate segmentation is merely the first step, being a prerequisite for extracting image-based traits. Based on our experimental results, many other traits, such as Above Ground Biomass (AGB), yield, and biologic rhythm, could be obtained using the Improved ResNet algorithm. Therefore, this algorithm represents a powerful tool for the large-scale phenotyping analysis of broccoli in a non-invasive and automatized way, and could potentially facilitate breeding research in future.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

CZ developed the detection and measurement algorithm, implemented the existing methods for comparison purposes, conducted the comparisons, obtained the results, and drafted the manuscript. GY developed the stomata detection algorithm. HY supervised the project, identified the main goals of the project, and contributed to writing the manuscript. JH and JY prepared the images for the research. ZX provided expert supervision for marking the ground-truth. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2016YFD0700303), the National Natural Science Foundation of China (31401287), the Public Projects of Zhejiang Province (2017C32024), the Beijing Natural Science Foundation (6182011), the Natural Science Foundation of China (61661136003 and 41471351).

REFERENCES

Cates, J. E., Lefohn, A. E., and Whitaker, R. T. (2004). Gist: an interactive, GPU-based level set segmentation tool for 3d medical images. *Med. Image Anal.* 8, 217–231. doi: 10.1016/j.media.2004.06.022

Chantal, L. M., Kirchgessner, N., Marschall, D., and Walter, A. (2014). Rhizoslides: paper-based growth system for non-destructive, high throughput phenotyping of root development by means of image analysis. *Plant Methods* 10:13. doi: 10.1186/1746-4811-10-13

Chaudhuri, B. B., and Chanda, B. (1984). The equivalence of best plane fit gradient with robert's, prewitt's and sobel's gradient for edge detection and a 4-neighbour gradient with useful properties. *Signal Process.* 6, 143–151. doi: 10.1016/0165-1684(84)90015-X

Cho, S. W., Cho, K., Bang, G., and Park, C. S. (2018). Molecular profiling of a y-type high molecular weight glutenin subunit at glu-d1 locus from a north korean landrace wheat (*Triticum aestivum* l.). *Plant Biotechnol. Rep.* 12, 139–148. doi: 10.1007/s11816-018-0479-6

Dell' Aquila, A. (2009). Digital imaging information technology applied to seed germination testing. a review. *Agron. Sustain. Dev.* 29, 213–221. doi: 10.1051/agro:2008039

Guo, L., Wang, P., Gu, Z., Jin, X., and Yang, R. (2017). Proteomic analysis of broccoli sprouts by itraq in response to jasmonic acid. *J. Plant Physiol.* 218, 16–25. doi: 10.1016/j.jplph.2017.07.003

Hamuda, E., Glavin, M., and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* 125, 184–199. doi: 10.1016/j.compag.2016.04.024

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Identity mappings in deep residual networks," in *Computer Vision – ECCV 2016. ECCV 2016: Lecture Notes in Computer Science*, eds B. Leibe, J. Matas, N. Sebe, and M. Welling (Cham: Springer), doi: 10.1007/978-3-319-46493-0_38

Islam, M., Dinh, A., Wahid, K., and Bhowmik, P. (2017). "Detection of potato diseases using image segmentation and multiclass support vector machine," in *Proceedings of the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, (Windsor, ON: IEEE), doi: 10.1109/CCECE.2017.7946594

Jeon, H. Y., Tian, L., Grift, T., Bode, L., and Hager, A. (2008). Plant specific direct chemical application field robot. *Diss. Theses Gradworks* 095697. doi: 10.13031/2013.26937

Ji, R., Fu, Z., and Qi, L. (2007). Real-time plant image segmentation algorithm under natural outdoor light conditions. *N. Z. J. Agric. Res.* 50, 847–854. doi: 10.1080/00288230709510359

Kalisz, A., Sêkara, A., Grabowska, A., Cebula, S., and Kunicki, E. (2015). The effect of chilling stress at transplant stage on broccoli development and yield with elements of modeling. *J. Plant Growth Regul.* 34, 532–544. doi: 10.1007/s00344-015-9488-7

Kusumam, K., Krajník, T., Pearson, S., Duckett, T., and Cielniak, G. (2017). 3d-vision based detection, localization, and sizing of broccoli heads in the field. *J. Field Robot.* 34, 1505–1518. doi: 10.1002/rob.21726

Lee, J., Kim, M. D., Park, J. J., and Chong, Y. J. (2018). Field-measurement-based received power analysis for directional beamforming millimeter-wave systems: effects of beamwidth and beam misalignment. *ETRI J.* 40, 26–38. doi: 10.4218/etrij.2017-0188

Ma, J., Wang, X., and Xiao, B. (2015). "An image segmentation method based on simple linear iterative clustering and graph-based semi-supervised learning," in *Proceedings of the 2015 International Conference on Orange Technologies (ICOT)*, (Hong Kong: IEEE), doi: 10.1109/ICOT.2015.7498477

Mehdipour Ghazi, M., Yanikoglu, B., and Aptoula, E. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 235, 228–235. doi: 10.1016/j.neucom.2017.01.018

Morais, C. L. M., Martin-Hirsch, P. L., and Martin, F. L. (2019). A three-dimensional principal component analysis approach for exploratory analysis of hyperspectral data: identification of ovarian cancer samples based on raman microspectroscopy imaging of blood plasma. *Analyst* 144, 2312–2319. doi: 10.1039/c8an02031k

Ramirez, R. A. (2006). *Computer Vision Based Analysis of Broccoli for Application in a Selective Autonomous Harvester*. Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Tang, P., Wang, H., and Kwong, S. (2016). G-ms2f: googlenet based multi-stage feature fusion of deep cnn for scene recognition. *Neurocomputing* 225, 188–197. doi: 10.1016/j.neucom.2016.11.023

Tu, K., Ren, K., Pan, L., and Li, H. (2007). "A study of broccoli grading system based 644 on machine vision and neural networks," in *Proceedings of the 2007 International Conference on Mechatronics and Automation*, (Harbin: IEEE) doi: 10.1109/ICMA.2007.4303917

Wang, D., Tan, D., and Liu, L. (2017). Particle swarm optimization algorithm: an overview. *Soft Comput.* 22, 387–408. doi: 10.1007/s00500-016-2474-6

Wang, Y., Wang, D., Zhang, G., and Wang, J. (2013). Estimating nitrogen status of rice using the image segmentation of g-r thresholding method. *Field Crops Res.* 149, 33–39. doi: 10.1016/j.fcr.2013.04.007

Wang, Y., and Xu, L. (2018). Unsupervised segmentation of greenhouse plant images based on modified latent dirichlet allocation. *PeerJ* 6:e5036. doi: 10.7717/peerj.5036

Xiong, X., Duan, L., Liu, L., Tu, H., Yang, P., Wu, D., et al. (2017). Panicle-seg: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* 13:104. doi: 10.1186/s13007-017-0254-7

Yamamoto, K., Guo, W., Yoshioka, Y., and Ninomiya, S. (2014). On plant detection of intact tomato fruits using image analysis and machine learning methods. *Sensors* 14, 12191–12206. doi: 10.3390/s140712191

Yang, L., Junfei, L., Liwei, T., and Lianbo, M. (2016). Hybrid artificial root foraging optimizer based multilevel threshold for image segmentation. *Comput. Intell. Neurosci.* 2016:1476838. doi: 10.1155/2016/1476838

Zhang, P., and Xu, L. (2018). Unsupervised segmentation of greenhouse plant images based on statistical method. *Sci. Rep.* 8:4465. doi: 10.1038/s41598-018-22568-3

Zheng, D., Da, F., Kemao, Q., and Seah, H. S. (2017). Phase-shifting profilometry combined with gray-code patterns projection: unwrapping error removal by an adaptive median filter. *Opt. Express.* 25:4700. doi: 10.1364/OE.25.004700

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhou, Hu, Xu, Yue, Ye and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 08 May 2020
doi: 10.3389/frai.2020.00028

Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest

Kasper Johansen[1], Mitchell J. L. Morton[2], Yoann Malbeteau[1], Bruno Aragon[1], Samer Al-Mashharawi[1], Matteo G. Ziliani[1], Yoseline Angel[1], Gabriele Fiene[2], Sónia Negrão[2,3], Magdi A. A. Mousa[4,5], Mark A. Tester[2] and Matthew F. McCabe[1]*

[1] Water Desalination and Reuse Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [2] Center for Desert Agriculture, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [3] School of Biology and Environmental Science, University College Dublin, Dublin, Ireland, [4] Department of Arid Land Agriculture, Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia, [5] Department of Vegetables, Faculty of Agriculture, Assiut University, Assiut, Egypt

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Yeyin Shi, University of Nebraska-Lincoln, United States
Shawn Carlisle Kefauver, University of Barcelona, Spain

***Correspondence:**
Kasper Johansen
kasper.johansen@kaust.edu.sa

Specialty section:
This article was submitted to AI in Food, Agriculture and Water, a section of the journal Frontiers in Artificial Intelligence

Received: *16 December 2019*
Accepted: *31 March 2020*
Published: *08 May 2020*

Citation:
Johansen K, Morton MJL, Malbeteau Y, Aragon B, Al-Mashharawi S, Ziliani MG, Angel Y, Fiene G, Negrão S, Mousa MAA, Tester MA and McCabe MF (2020) Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest. Front. Artif. Intell. 3:28. doi: 10.3389/frai.2020.00028

Biomass and yield are key variables for assessing the production and performance of agricultural systems. Modeling and predicting the biomass and yield of individual plants at the farm scale represents a major challenge in precision agriculture, particularly when salinity and other abiotic stresses may play a role. Here, we evaluate a diversity panel of the wild tomato species (*Solanum pimpinellifolium*) through both field and unmanned aerial vehicle (UAV)-based phenotyping of 600 control and 600 salt-treated plants. The study objective was to predict fresh shoot mass, tomato fruit numbers, and yield mass at harvest based on a range of variables derived from the UAV imagery. UAV-based red–green–blue (RGB) imageries collected 1, 2, 4, 6, 7, and 8 weeks before harvest were also used to determine if prediction accuracies varied between control and salt-treated plants. Multispectral UAV-based imagery was also collected 1 and 2 weeks prior to harvest to further explore predictive insights. In order to estimate the end of season biomass and yield, a random forest machine learning approach was implemented using UAV-imagery-derived predictors as input variables. Shape features derived from the UAV, such as plant area, border length, width, and length, were found to have the highest importance in the predictions, followed by vegetation indices and the entropy texture measure. The multispectral UAV imagery collected 2 weeks prior to harvest produced the highest explained variances for fresh shoot mass (87.95%), fruit numbers (63.88%), and yield mass per plant (66.51%). The RGB UAV imagery produced very similar results to those of the multispectral UAV dataset, with the explained variance reducing as a function of increasing time to harvest. The results showed that predicting the yield of salt-stressed plants produced higher accuracies when the models excluded control plants, whereas predicting the yield of control plants was not affected by the inclusion of salt-stressed plants within the models. This research demonstrates that it is possible to predict the average biomass and yield up to 8 weeks prior to harvest within 4.23% of

field-based measurements and up to 4 weeks prior to harvest at the individual plant level. Results from this work may be useful in providing guidance for yield forecasting of healthy and salt-stressed tomato plants, which in turn may inform growing practices, logistical planning, and sales operations.

Keywords: UAV, yield, biomass, tomato plants, salinity, random forest, RGB, multi-spectral

INTRODUCTION

Along with growing populations and the challenges of climate change, salt-stress presents as a major threat to global food production. While soil salinity in irrigated agriculture is a global concern, it is particularly so in arid and semiarid climates (Pitman and Lauchli, 2002; Rao et al., 2013; Machado and Serralheiro, 2017). Breeding of crop cultivars with improved salt tolerance represents one potential pathway toward improving food and water security (Hickey et al., 2019; Johansen et al., 2019a). To do this requires the identification of salt-tolerant genotypes/accessions, whose tolerance traits can then be introgressed into commercial varieties (Munns and Tester, 2008; Messerer et al., 2018; Morton et al., 2019). In order to identify the potential salt tolerance of plant accessions, phenotyping and related approaches that can effectively map, monitor, and predict plant biophysical and biochemical properties are required (Johansen et al., 2019a). Two indicators of plant response that can offer insight into performance are biomass and yield. While the effects of salinity are to generally reduce a plant's biomass and yield, what is not well-understood is how salt stress affects the ability to predict these variables ahead of harvest time (Flowers and Flower, 2005; Verslues et al., 2006; Stavridou et al., 2017; Johansen et al., 2019a,b).

Measurements of biomass provide information on a plant's ability to capture sunlight, water, and minerals, and the rates at which it can turn these into physical growth (Johansen et al., 2019b). It is also useful in informing the amount and timing of fertilizer, pesticides, and water to be applied to optimize crop performance and improve agricultural management practices (Jaleel et al., 2009; Bendig et al., 2015): key metrics behind the concept of precision agriculture. Predicting yield prior to harvest can facilitate logistical planning and scheduling of field and harvest operations, e.g., fruit picking, storage, packaging, and transportation (Robson et al., 2017), as well as help in financial planning and management. In recent years, predicting biomass and yield (along with other biophysical and biochemical properties) through various types of sensing technologies has become a focus for both precision agriculture and smart farming. While precision agriculture attempts to observe, measure, and respond to inter- and intrafield crop variability, smart farming encompasses a focus on agricultural systems management via big data analytics, using context and situation awareness, often in real time (Zhang et al., 2002; Gebbers and Adamchuk, 2010; Wolfert et al., 2017).

Both precision agriculture and smart farming require large amounts of data to ensure informed decision-making at the specific plant, tree, or plot level. Data required to drive timely or near real-time information-based decisions may be obtained using a range of remote sensing platforms, as well as in-field robotics and sensing technologies (Wolfert et al., 2017). However, satellite-based remote sensing is often unable to provide the spatial resolution required for per-plant or per-tree assessment, with timely data acquisition often affected by cloud cover or other adverse atmospheric conditions (Nevavuori et al., 2019). Airborne remote sensing remains costly, especially if used for high-frequency assessment of growth patterns and other crop parameters relying on multitemporal data (Koh and Wich, 2012). Field-based robotics have proved useful for fruit counting, fruit ripening assessment, flower identification, yield prediction, and measurements of three-dimensional structure (Underwood et al., 2016; Bargoti and Underwood, 2017; Wang Z. et al., 2018; Wendel et al., 2018; Westling et al., 2018) but are generally restricted to smaller areas with no ground obstacles to hinder access (Kragh and Underwood, 2019). At the other end of the spectrum, it can be very time consuming, labor intensive, and subjective to consistently collect field data suitable for predicting biomass and yield at harvest (Sugiura et al., 2015; Holman et al., 2016). The divide between space- and ground-based data collection has recently been filled by the use of unmanned aerial vehicles (UAVs), which provide a means for efficient, regular, and flexible collection of imagery at very high spatial resolutions, suitable for regular assessment of crops, their properties, and stress factors (Gil-Docampo et al., 2018). Deployment of UAVs for data collection also reduces the requirement for human-based on-site observations (and potential for investigator bias), increases safety and access, and facilitates the implementation of management practices in the agricultural sector (Shi et al., 2016; Barbedo, 2019).

Despite the commercial importance of tomatoes, relatively few studies have assessed the potential for using UAV-based imagery for modeling their biomass and yield [annual global production is ∼171 million tons; (FAOSTAT, 2017)]. Senthilnath et al. (2016) used two UAV-derived images to delineate and classify tomato fruits on individual plants but found that many fruits were omitted, as they were visually occluded by leaves and stalks. Johansen et al. (2019a) used a time series of RGB and multispectral UAV imagery to accurately monitor phenotypic traits of individual tomato plants, including plant area, plant projective cover, condition, and growth rate, and used these variables to successfully identify tomato plant accessions that performed the best in terms of yield. Moeckel et al. (2018) estimated crop height and biomass of eggplant, tomato, and cabbage plants from a time series of five red, green, and blue (RGB) UAV-based data sets and found measured crop height to correlate well with biomass when using random forest and

support vector regression. Johansen et al. (2019b) provided some initial findings for using UAV imagery to predict biomass and yield at harvest. However, this initial work only assessed prediction accuracies of individual plants from RGB imagery and excluded an evaluation on variable importance for the prediction models. This paper significantly expands on the interpretation of these preliminary results and further explores the use of multispectral imagery, variable importance, model predictions of biomass, fruit numbers, and yield mass, and a comparison of model results for control and salt-treated plants.

The use of UAVs for plant phenotyping purposes has witnessed their application for plant height assessment (Hu et al., 2018; Wang X. et al., 2018), genotype performance under low nitrogen conditions (Buchaillot et al., 2019), crop growth monitoring (Holman et al., 2016), among many others (Yang et al., 2017). While UAV imagery has only been exploited to a limited extent for predicting biomass and yield of tomato plants, it has been applied to assess biomass and yield of other crops. For instance, Fathipoor et al. (2019) used RGB UAV imagery acquired at the mid-season growth stage to predict corn forage yield using a combination of plant height and vegetation indices for partial least square regression. Ballesteros et al. (2018) used height derived from UAV-based RGB imagery as well as green canopy cover and canopy volume to estimate the biomass of onion, with canopy volume found to be particularly informative. Han et al. (2019) used various predictor variables, including plant height, canopy shape, and vegetation indices, to predict aboveground biomass of maize from UAV-based imagery and achieved the best results using a random forest model. Nevavuori et al. (2019) used convolutional neural networks (CNN) to build a model for predicting yields of wheat and barley fields using multispectral UAV imagery. They found that yield prediction errors were lower for UAV data acquired early in the growth season than using data acquired later and closer to harvest. Other examples of UAV-based studies using machine learning approaches for biomass estimation include the mapping of wheat (Lu et al., 2019), grass sward (Nasi et al., 2018), rice (Jiang et al., 2019), and maize (Han et al., 2019). Related crop yield estimation studies include the mapping of oilseed rape (Peng et al., 2019), barley (Escalante et al., 2019), rice (Yang et al., 2019), and cotton (Zou et al., 2018). Apart from the work by Moeckel et al. (2018) and Johansen et al. (2019a,b), no other research was identified using UAV-based time series for the prediction of biomass and yield of tomato plants at harvest.

The UAV-based studies on yield and biomass mapping reviewed above used a variety of artificial intelligence approaches, including both machine learning and deep learning techniques. In fact, Liakos et al. (2018) identified yield prediction as one of the most common applications of machine learning in agriculture. Although big data analysis is becoming more common in the agricultural sector (Kamilaris et al., 2017), obtaining the required data volumes of suitable quality needed for large-scale application of machine learning approaches remains a challenge. Still, more traditional and established machine learning approaches are often beneficial for smaller scale studies, where interpretability may also be important. From the selection of machine learning algorithms used in agricultural UAV-based studies, the random forest approach has been regularly identified as producing the best results, which is commonly attributed to its lower sensitivity to data skewness and prevention of model overfitting (e.g., Moeckel et al., 2018; Han et al., 2019; Lu et al., 2019).

Overall, the objectives of this research were three-fold: (1) to predict biomass (fresh shoot mass) and yield (tomato fruit numbers and yield mass) at harvest from a time series of RGB UAV-based imagery; (2) to determine if prediction accuracies varied between control and salt-treated plants; and (3) to compare results of RGB and multispectral UAV imagery collected 1 and 2 weeks before harvest. To do this, a random forest machine learning approach was employed to predict biomass and yield for both control and salt-treated plants using a time series of six RGB and two multispectral UAV image data sets collected prior to harvest. An ability to forecast at-harvest biomass and yield would offer growers a capacity to identify and understand biomass and yield variability in areas affected by salinity and could be used to optimize per plant inputs during the growing season, while also informing logistical and sales related operations.

MATERIALS AND METHODS

Study Area and Experimental Design

The study area was located at the King Abdulaziz University Agricultural Research Station in Hada Al-Sham (21.7967°N, 39.7264°E), about 60 km east of Jeddah in the Makkah region of Saudi Arabia. The area receives an annual rainfall of <100 mm and has a predominantly sandy loam soil type. A tomato field-trial experiment was initiated in November 2017, with the field arrangement consisting of four 30 × 30 m plots. In each of the four plots, 15 rows of 20 tomato plants were planted, producing a combined total of 1,200 plants (**Figure 1**). These 1,200 plants included 200 different genotypes, consisting of 199 *Solanum pimpinellifolium* accessions and one *Solanum lycopersicum* accession (the commercial tomato, Heinz 1706). The 199 *S. pimpinellifolium* accessions originated from different parts of Peru and Ecuador.

Seeds for these accessions were propagated at King Abdullah University of Science and Technology (KAUST) (Johansen et al., 2019a). Using these, the tomato plants were sown at a greenhouse nursery at KAUST a month before transplanting, which occurred on November 1-2, 2017. Plants were allocated into two control and two salt-treated plots, following a randomized design (**Figure 1**). Three replicates of the 200 accessions were planted in each of the two treatments. The two control plots were irrigated with low salinity water (27 mM NaCl, 900-1,000 ppm) throughout the growing season. The two salt-treated plots were irrigated with saline water of 127 mM NaCl (4,500 ppm) from November 14, 2017, 197 mM NaCl (7,000 ppm) from December 4, 225 mM NaCl (8,000 ppm) from December 10, 254 mM NaCl (9,000 ppm) from December 18, and 183 mM NaCl (6,500 ppm) from January 12, 2018 until harvest, which occurred between January 16-22 (Johansen et al., 2019a). Drip irrigation was applied once in the morning for 10 min and again in the evening for 10 min until November 9, then for 15 min in the morning and evening until December 17, and 30 min in the morning and evening until harvest, in response to the

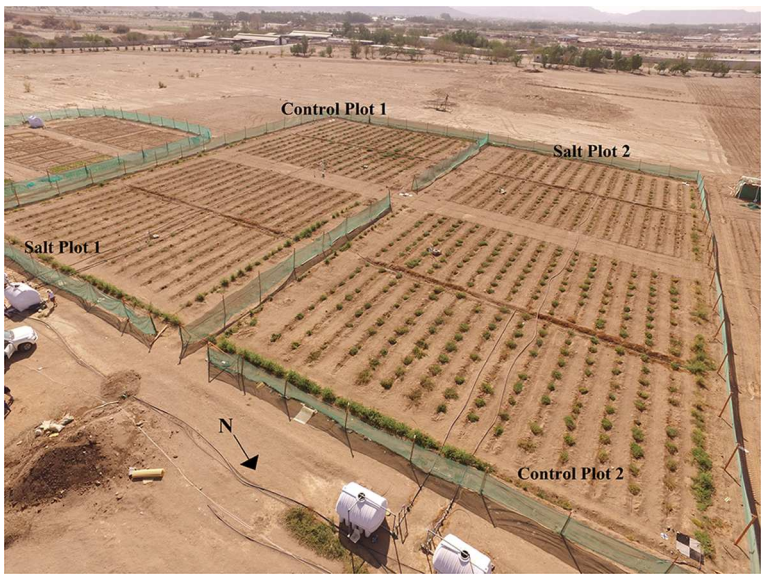

FIGURE 1 | Setup of the tomato plant trial with 300 plants in each of the two control and two salt-treated plots. The plant trial covered an area of ∼75 × 75 m, with each plot of 300 plants being 30 × 30 m.

increasing irrigation requirements of growing plants. Weeds within each plot were manually removed before each of the UAV flights. During the growing season, maximum day and minimum night temperatures ranged from 27 to 37°C and 12 to 24°C, respectively. During the growing season, no rainfall was recorded, but several sandstorms occurred, with the most severe and damaging on December 8 and 16, 2017. After each sandstorm, farm employees washed the plants with non-saline water to remove dust from the leaves (Johansen et al., 2019a).

Field Data Collection

Field and UAV data were collected on November 23 and 30, December 6 and 20, 2017, and January 7 and 14, 2018, i.e., 1, 2, 4, 6, 7, and 8 weeks before harvest. Five GCPs were deployed for geo-referencing the UAV imagery and their coordinates measured at the planting date (November 2) using a Leica GS10 base station with an AS10 antenna and a Leica GD15 smart antenna as a rover (Leica Geosystems, St. Gallen, Switzerland). GCPs were placed at the center and at each of the four corners of the study site. All collected Global Navigation Satellite System data were postprocessed using the Leica Geo Office software (Leica Geosystems, St. Gallen, Switzerland). Six near-Lambertian panels in white, four shades of gray, and in black (Johansen et al., 2018, 2019a) were placed within the field and measured with an ASD FieldSpec4 spectrometer (Malvern Panalytical, Malvern, UK) for radiometric calibration of the collected imagery.

Between January 16 and 26, fresh shoot mass, fruit numbers, and yield mass were measured for all tomato plants remaining at harvest. The fresh shoot mass, including aboveground plant material and fruit, was measured first. Fresh shoot mass ranged from 17 to 5,402 g per plant, averaging 715 g/plant (**Table 1**). Fruit numbers were manually counted for both mature (fruits with some redness) and immature (green fruits) fruits >3 mm in diameter on each plant. The fruit was subsequently weighed for each plant. For plants with >1 kg shoot mass, a representative sample of the whole plant was selected to count and weigh all fruits >3 mm in diameter. To estimate the fruit numbers and their weight for the whole plant, the sample was used for extrapolation and multiplied by the weight ratio of the whole plant and the selected sample. Based on all harvested plants, the number of tomatoes ranged from 1 to 3,349 per plant, averaging 532 fruits/plant. Yield mass ranged from 0.1 to 1,433 g/plant, averaging 227 g/plant (Johansen et al., 2019a). The total number of observations at harvest for fresh shoot mass, fruit number, and yield mass were 1,027 (514 control and 513 salt treated), 980 (496 control and 484 salt treated), and 979 (497 control and 482 salt treated) plants, respectively (**Table 1**). The present study only examines those plants that survived until harvest.

UAV Data Collection and Processing

A gimbal-stabilized Zenmuse X3 camera (Dà-Jiāng Innovations, Shenzhen, China) installed on a DJI Matrice 100 (Dà-Jiāng Innovations, Shenzhen, China) quadcopter was used to collect RGB imagery for all of the six field campaigns, i.e., 1, 2, 4, 6, 7, and 8 weeks before harvest. The Zenmuse X3 camera has a Sony EXMOR 1/2.3″ complementary metal-oxide semiconductor (CMOS) sensor with the full width at half-maximum being ∼400–510 nm for the blue band, 480–600 nm for the green band, and 580-700 nm for the red band (Sato et al., 2016). The Parrot Sequoia sensor (Parrot SA, Paris, France), also installed on the DJI Matrice 100, was used to collect coincident multispectral green (530–570 nm), red (640–680 nm), red edge (730–740 nm), and near-infrared (NIR) (770–810 nm) imagery for the last two campaigns, i.e., 1 and 2

TABLE 1 | Minimum, maximum, and mean values and the total number of observations (obs.) as well as control and salt observations of fresh shoot mass, fruit numbers, and yield mass.

	Minimum	Maximum	Mean	Total obs.	Control obs.	Salt obs.
Fresh shoot mass	17 g/plant	5,402 g/plant	715 g/plant	1,027	514	513
Fruit numbers	1 fruit/plant	3,349 fruits/plant	532 fruits/plant	980	496	484
Yield mass	0.1 g/plant	1,433 g/plant	272 g/plant	979	497	422

Note that only fruits >3 mm in diameter were included.

weeks before harvest. All UAV campaigns occurred under cloud-free conditions at around solar noon. The Universal Ground Control Station (UgCS) Client application (SPH Engineering, SIA, Riga, Latvia) was used for flight planning. For each UAV campaign, the image data were collected at an altitude of 13 m above ground level and at a speed of 2 m/s. This produced a forward overlap and sidelap of 78 and 82%, respectively, for the RGB imagery (3-s image intervals) and 83 and 68%, respectively, for the multispectral imagery (1-s image intervals) (Johansen et al., 2019a). A geo-referenced orthomosaic and digital surface model (DSM) were produced at a pixel size of 0.5 and 1.12 cm for the RGB and multispectral imagery, respectively, using Agisoft PhotoScan (Agisoft LLC, St. Petersburg, Russia). A digital terrain model (DTM), produced from RGB UAV imagery collected of the study site before planting, was subtracted from the DSMs to produce canopy height models (CHMs) for each of the six UAV campaign (Johansen et al., 2019a).

Buchaillot et al. (2019), Madec et al. (2017), and Singh et al. (2019) used RGB UAV imagery for phenotyping maize and wheat and argued that no radiometric correction of the UAV imagery was required, at least when using point cloud information for plant height measurements and field-based RGB imagery acquired coincidently with the UAV data. However, it is considered good practice to radiometrically calibrate UAV imagery to remove or normalize effects of sun-object-sensor geometry and illumination conditions (Tmusic et al., 2020). The empirical line calibration method has become a standard approach for UAV-based studies to correct the imagery to at-surface reflectance using either natural features (Lelong et al., 2008) or other types of commercial (e.g., spectralon) or more cost-effective Lambertian panels, e.g., made from masonite, plywood, or ethylene-vinyl acetate (Wang and Myint, 2015; Jeong et al., 2018; Barreto et al., 2019; Tmusic et al., 2020). An empirical line correction, produced between field-derived spectrometer measurements and the digital numbers of the radiometric calibration panels within the multispectral and RGB orthomosaics, was used to convert the orthomosaics to at-surface reflectance (Ahmed et al., 2017; Johansen et al., 2018). For the empirical line corrections, the coefficients of determination (R^2) were >0.98 for all band combinations and data sets. While the multispectral imagery produced a linear empirical line, an exponential relationship was required for the RGB imagery. An exponential relationship of the empirical line correction for RGB UAV imagery was also identified by Jeong et al. (2018).

Delineation of Tomato Plants

The eCognition Developer 9.3 software (Trimble, Munich, Germany) was used to develop an object-based rule set to delineate all individual tomato plants within the six RGB and two multispectral orthomosaics. After a fine-scale multiresolution segmentation, vegetation indices, and spectral band combinations were used to classify objects, representing the green parts of the tomato plants. A region-growing algorithm was subsequently applied to expand tomato plant objects into neighboring unclassified objects using more relaxed vegetation index thresholds. Unclassified objects enclosed by tomato plant objects were merged with the enclosing tomato plant objects. To remove small incorrectly classified tomato plant objects, an area threshold of <150 cm^2 was applied. Remaining tomato plants were then resized using a number of growing and shrinking processes and the use of the produced CHM to adjust the edges of the tomato plant objects. The delineation results of the six RGB and two multispectral orthomosaics were visually assessed and manually edited if necessary. A more detailed description of the object-based approach can be found in Johansen et al. (2019a).

Extraction of Image Variables

Based on the delineated plants, shape, spectral, and texture variables were extracted from both the RGB and multispectral UAV imagery. For the RGB imagery, the extracted variables included the three RGB bands, the Green–Red Index (Motohka et al., 2010), nine shape features (border length, width, length, length/width ratio, elliptic fit, shape index, compactness, roundness, and border index) exported directly from the eCognition Developer software, four gray-level co-occurrence textural measures [homogeneity, contrast, entropy, and dissimilarity; (Haralick et al., 1973)] based on the three spectral bands and the Green–Red Index, maximum plant height, and the standard deviation of maximum height (see **Table 2**). For the multispectral UAV imagery, the extracted variables included the four spectral bands (green, red, red edge, and NIR), six vegetation indices (Robson et al., 2017) (**Table 3**), nine shape features, four gray-level co-occurrence textural measures [homogeneity, contrast, entropy, and dissimilarity; (Haralick et al., 1973; Lofstedt et al., 2019)] based on the four spectral bands and the normalized difference vegetation index (NDVI), as well as maximum height and the standard deviation of maximum height (see **Table 2**).

Image-based texture measures have been found useful for UAV-based biomass estimation in other studies (Zheng et al., 2019) and may provide additional spatial information useful for

TABLE 2 | Variables extracted (gray fields) from the red–green–blue (RGB) (35 in total) and multispectral (46 in total) unmanned aerial vehicle (UAV) imagery for each individual tomato plant.

Image-derived variables	MS	RGB	Image-derived variables	MS	RGB
Border length	■	■	Homogeneity blue		■
Width	■	■	Homogeneity green	■	■
Length	■	■	Homogeneity Green–Red Index		■
Length/width ratio	■	■	Homogeneity red	■	■
Elliptic fit	■	■	Homogeneity red edge	■	
Shape index	■	■	Homogeneity NIR	■	
Compactness	■	■	Homogeneity NDVI	■	
Roundness	■	■	Contrast blue		■
Border index	■	■	Contrast green	■	■
Area	■	■	Contrast Green–Red Index		■
Maximum height	■	■	Contrast red	■	■
Standard deviation height	■	■	Contrast red edge	■	
Blue		■	Contrast NIR	■	
Green	■	■	Contrast NDVI	■	
Red	■	■		■	■
Red edge	■		Entropy green	■	■
NIR	■		Entropy Green–Red Index		■
Green–Red Index	■	■	Entropy red	■	■
NDVI	■		Entropy red edge	■	
RENDVI	■		Entropy NIR	■	
NDRE	■		Entropy NDVI	■	
NIRRENDVI	■		Dissimilarity blue		■
Green NDVI	■		Dissimilarity green	■	■
Standard deviation blue		■	Dissimilarity Green–Red Index		■
Standard deviation green	■	■	Dissimilarity red	■	■
Standard deviation red	■	■	Dissimilarity red edge	■	
Standard deviation red edge	■		Dissimilarity NIR	■	
Standard deviation NIR	■		Dissimilarity NDVI	■	

TABLE 3 | Vegetation indices calculated from the multispectral unmanned aerial vehicle (UAV) imagery and extracted for each individual tomato plant.

Vegetation Indices	Equation	References
Green–Red index	(Green – red)/(Green + red)	Motohka et al., 2010
NDVI	(NIR – red)/(NIR + red)	Rouse et al., 1973
RENDVI	(Red edge – red)/(red edge + red)	Sims and Gamon, 2002b
NDRE	(NIR – red edge)/(NIR + red edge)	Jorge et al., 2019
NIRRENDVI	{[(NIR + red edge)/2] – red}/{[(NIR + red edge)/2] + red}	Xie et al., 2018
Green NDVI	(NIR – green)/(NIR + green)	Gitelson and Merzlyak, 1998

improving classification results (Johansen et al., 2007). To achieve directional invariance, the sum of all four directions (0, 45, 90, and 135°) were calculated before the texture calculation. The calculation of texture [following (Gil-Docampo et al., 2018)] was independent of the image data bit depth, as the dynamic range was interpolated to 8 bit before evaluating the co-occurrence (Trimble eCognition Developer, 2017; Lofstedt et al., 2019). All extracted per-plant image variables from all eight image data sets (six RGB and two multispectral) were then used to assess their linear, exponential, logarithmic, and second-order polynomial correlation with the corresponding field-derived measurements of biomass, fruit numbers, and yield mass. Those variables with an $R^2 < 0.10$ in all four correlations were omitted from further analysis, as they were considered insignificant. The remaining variables were used for predicting biomass and yield using a random forest machine learning approach.

Random Forest Modeling and Analysis

The random forest machine learning approach has been applied widely in ecological studies (e.g., Prasad et al., 2006; Cutler et al., 2007) and in various remote-sensing-based analyses (e.g., Belgiu and Dragut, 2016; Shi and Yang, 2016; Ma et al., 2017; Sarron et al., 2018; Tu et al., 2019). Some of the identified benefits of random forest is its capability to model complex variable interactions and prevent overfitting (Maxwell et al., 2018). Breiman (2001) found the random forest approach to perform better than other classifiers, including discriminant analysis, support vector machines, and neural networks. Similar to the statistical approach of bagging, the random forest approach is used to determine the optimal set of decision trees. Successive classification trees are independently constructed using a random sample of the data that does not depend on earlier decision trees (Johansen et al., 2015). The best split at each node is based on a subset of randomly selected predictor variables: in this case, set to 10 (mtry), or if the number of variables were fewer than 10, then all variables were considered. For those data sets with >10 variables, it was found through an iterative process that reducing the number of variables considered at each node to 10 did not affect the results. A value of 10 was selected because below that value, more variation in the accuracy between multiple runs of the models was noted. The final aggregation approach for all decision trees produced by the random forest algorithm prevents overfitting. A total of 1,000 decision trees (ntree) were used in this process to ensure stable predictions that were not too computationally intensive (Oliveira et al., 2012). Every training set is randomly sampled from the whole data set with replacement, i.e., the same observation can be used multiple times. Hence, in each decision tree, a bootstrap sample was selected, containing 63.2% of the data, with the remaining data used as evaluation data and to calculate the out-of-bag (OOB) error rate. In this study, the OOB error rate was used as an unbiased estimate of prediction error (Braga-Neto and Dougherty, 2004; Ghosh et al., 2014). Using

a bootstrap sample for each decision tree further prevents overfitting. The experiments were implemented with the R package "randomForest" (https://CRAN.R-project.org/package=randomForest) (Liaw and Wiener, 2002).

Each of the six RGB and two multispectral image data sets were used independently to produce random forest models for predicting fresh shoot mass, fruit numbers, and yield mass per plant, using all plant observations, only salt-treated observations, and only control observations, yielding a total of 72 models (Johansen et al., 2019b). The models based on all plant observations were also applied to a subset consisting only of salt-treated plants and another subset including only control plants. This assessment was undertaken to evaluate how separate models adapted specifically to either salt-treated or control plants performed compared to those incorporating all plant observations. To determine the accuracy of each model, the importance of each predictor variable, the percentage of explained variance, and the root mean square error (RMSE) between the OOB observations of the field and UAV-derived measurements of fresh shoot mass, fruit numbers, and yield mass per plant were assessed. The relative root mean square error (rRMSE), defined as the RMSE divided by the mean values of the field observations, was also calculated. The decision trees were fully grown, and each was used to predict the OOB observations for that bootstrap sample. The final predicted value for an observation was the average of the OOB predictions for that observation based on the 1,000 decision trees. The permutation importance measure was used in this study. The importance of each variable is estimated by determining how much prediction error increases for each decision tree when OOB data for a selected variable is permuted and while all other variables are left unchanged. The increase in prediction error is then averaged over all trees and normalized by the standard deviation and measured as the percentage increase in mean squared error (%IncMSE) (Liaw and Wiener, 2002; Gregorutti et al., 2017).

RESULTS

Object-Based Variable Importance for Predicting Biomass and Yield

For the eight UAV image data sets (six RGB and two multispectral), 94.6–99.1% of all plants were automatically detected, with 5–16% of the plants subjected to additional manual adjustment to improve the delineation results. Plant length (longest axis) measured in the field with a tape measure produced an R^2 value and RMSE of 0.85 and 0.052 m (n = 132), respectively, when compared with measurements from the automatically delineated plants. There was a tendency of smaller plants being slightly overestimated, whereas larger plants were slightly underestimated in length (Johansen et al., 2019a,b). Manual adjustment of the delineation results increased the R^2 value to 0.97 and reduced the RMSE to 0.018 m. **Figure 2** depicts the delineation results for January 7, and further details on the delineation results can be found in Johansen et al. (2019a).

Based on the delineated tomato plant objects, extracted parameters including shape features, spectral information, vegetation indices, texture features, and height information were used for predicting biomass and yield for each individual tomato plant. **Table 4** shows the number of variables used for the random forest models, i.e., those variables with $R^2 > 0.10$ when assessed against field-derived measurements of biomass and yield. When performing the random forest machine learning predictions of biomass and yield, the variable importance was calculated. As expected, some bands proved to be of high importance in most of the predictions, whereas other bands consistently appeared with low importance. When all models were rerun omitting the bands with low importance, it was found that reducing the number of bands with low importance did not increase the variance explained or reduce the RMSE. Hence, in this study, there was little benefit in further reducing the number of predictor variables included in the predictions.

Prediction models of fresh shoot mass, fruit numbers, and yield mass were significantly improved by inclusion of shape features, including plant area, border length, width, and length, with plant area consistently achieving the highest importance values (**Figure 3**). However, there was a tendency of the importance scores for the plant area to reduce as a function of increasing time to harvest. Interestingly, the importance values for plant area were higher for the December 20 and January 7 data sets than the January 14 data set (closest to harvest), at least for fruit numbers and yield mass. After December 20, deterioration of many of the tomato plants occurred due to a number of destructive sandstorms on December 8 and 16, which damaged many plants (≈9%). That, combined with the increasing salt stress toward the end of the growing season, caused many of the remaining plants to exhibit signs of poor condition and senescence. Hence, the physical appearance and plant area of some plants may not have corresponded as well on January 14 to their measured yield as they did on December 20 and January 7. Plant area on December 20 and January 7 may reflect how well-plants coped with the sandstorms, with those coping well also yielding well. However, it is possible that the correlation between area and yield uncoupled in the final days of the experiment, i.e., January 14 (**Figure 3**), because of a lag between how well a plant performed at a particular time (e.g., measured by area) and how that translated to effects on fruiting and yield, which might not have manifested before January 14. The four shape features, i.e., plant area, border length, width, and length of individual plant objects, were important model input parameters for both the six RGB UAV collections (**Figure 3**) and the two multispectral UAV data sets (**Table 5**).

In terms of spectral metrics, the Green–Red Index had high importance for predicting fresh shoot mass for all six RGB UAV captures (**Figure 3**) as well as the multispectral UAV imagery (**Table 5**). Several of the other multispectral vegetation indices were also highly ranked in their importance for predicting fresh shoot mass, fruit numbers, and yield mass. In fact, for the prediction of fresh shoot mass on January 14 using the multispectral UAV imagery, the 10 highest ranked predictor variables in terms of importance consisted entirely of shape features and vegetation indices (**Table 5**). Many UAV-based crop studies have reported positive correlation between vegetation indices and both biomass and yield (e.g.,

FIGURE 2 | Delineation results (yellow outlines) based on unmanned aerial vehicle (UAV) imagery collected on January 7, 2018 for **(A)** the whole field site and **(B)** a subset of 20 tomato plants outlined in the white rectangle.

Hassan et al., 2019; Niu et al., 2019; Zheng et al., 2019). In our study, the Green–Red Index showed relatively lower importance for the prediction of fruit numbers and yield mass based on the RGB UAV imagery. Again, this may have been attributed to the severe sandstorms prior to the December 20 data collection, which could have limited the Green–Red Index from providing greater predictive value in terms of fruit numbers and yield mass. In addition, some senescent tomato plants still produced high fruit numbers at the end of the growing season, which again also would have reduced the capacity of vegetation indices successfully predicting fruit numbers and yield mass. It should be noted that the Green–Red index achieved the highest importance in predicting fresh shoot mass on December 20. This date corresponds to the highest recorded green biomass during the season, as the damage from the preceding sandstorm events and increasing salt-stress resulted in the senescence of many plants after December 20 (Johansen et al., 2019a).

Several of the texture features extracted for the individual tomato plants were of importance for predicting fresh shoot mass, fruit number, and yield mass from the RGB and multispectral UAV imagery. The entropy texture measure of the spectral bands and vegetation indices proved especially useful for these predictions. The entropy texture measure tended to show an exponential relationship with biomass and yield. Entropy texture extracted from a gray level co-occurrence matrix measures the spatial disorder of pixels, with texturally uniform image objects having small values (Kekre et al., 2010). Tomato plants with increasing biomass and yield appeared with

TABLE 4 | Number of variables used for predicting fresh shoot mass, fruit numbers, and yield mass from the six red–green–blue (RGB) and two multispectral (MS) unmanned aerial vehicle (UAV) data sets.

	Jan 14 RGB/MS	Jan 7 RGB/MS	Dec 20 RGB	Dec 6 RGB	Nov 30 RGB	Nov 23 RGB
Fresh shoot mass	20/31	20/35	19	14	14	8
Fruit numbers	20/31	18/26	15	14	8	4
Yield mass	20/31	19/27	15	14	8	8

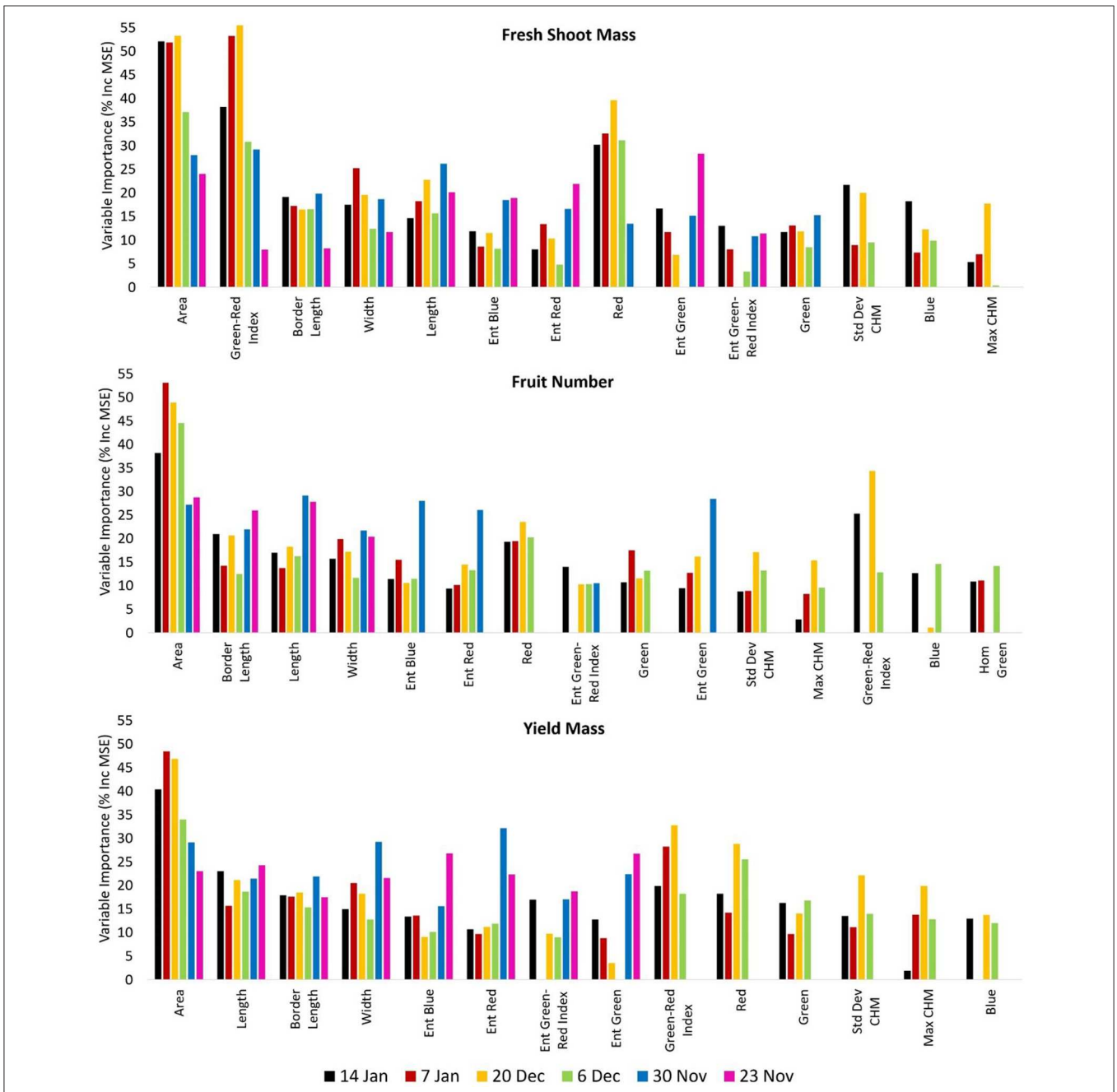

FIGURE 3 | Variable importance in the random forest models based on the six unmanned aerial vehicle (UAV) red–green–blue (RGB) image data sets for the prediction of fresh shoot mass, fruit number, and yield mass. Only variables that were included in ≥3 out of the six RGB unmanned aerial vehicle (UAV) data sets are presented. Variables were sorted based on the number of data sets within which they were included and their importance value.

TABLE 5 | Ranking of the 10 most important variables in the random forest models based on the two multispectral unmanned aerial vehicle (UAV) image data sets for the prediction of fresh shoot mass, fruit number, and yield mass.

Importance	Fresh shoot mass		Fruit number		Yield mass	
Ranking	14-Jan	7-Jan	14-Jan	7-Jan	14-Jan	7-Jan
1	Area	Area	Area	Area	Area	Area
2	Width	Width	Length	Width	Length	Width
3	Length	NIRRENDVI	Ent RedEdge	Ent RedEdge	Ent RedEdge	Ent RedEdge
4	NDVI	Length	Width	RENDVI	Width	NIRRENDVI
5	NIRRENDVI	RENDVI	Green-Red index	Green-Red index	Green NDVI	Green-Red index
6	Green-Red index	Green-Red index	Ent NIR	NIRRENDVI	Dis NDVI	RENDVI
7	Red	Red	Green NDVI	Length	Border length	Length
8	RENDVI	Hom NDVI	Ent green	Hom NDVI	Con red	Hom NDVI
9	Border length	Ent Green	Border length	Border length	Ent NIR	Green NDVI
10	Green NDVI	Ent RedEdge	Dis NDVI	Dis NDVI	Ent green	Border length

increasing entropy texture values, indicating a more complex and "disorderly" plant canopy architecture. However, once the biomass and yield reached a set threshold in size, the entropy texture values became unsuited for accurately predicting biomass and yield because of the exponential relationship. Hence, large plants with large amounts of biomass and high fruit numbers and yield mass represented the maximum spatial disorder of pixels accounted for by the entropy measure. While the entropy texture measure might be a useful predictor variable for smaller tomato plants with limited biomass and yield, the measure became increasingly unsuited as a predictive variable with increasing biomass and yield. This is likely the reason why the entropy texture measure was found informative during the earliest image campaigns (i.e., when the plants were still small) but had a lower importance for those data sets collected closer to harvest (i.e., when the plants were larger) (see **Figure 3**).

The measurements of plant height and their standard deviation was in most cases identified with low importance as a predictor variable. A cause of this might have been the fact that plant height did not vary much between large and small plants, with plant growth mainly occurring horizontally. In addition, the sandstorm events that occurred during the growing season caused many of the branches to break, which also affected plant height in some cases. The importance of predictor variables was also assessed separately for the control and salt-treated tomato plants. While variations in the ranking of predictor variables occurred, the shape features were still found to be most important, followed by vegetation indices and then texture measures (the entropy texture measure in particular). However, one noticeable variation was the much higher importance of the standard deviation of maximum plant height for the salt-treated tomato plants when using the RGB UAV imagery. This may have been due to the much lower yield and hence limited effect of the tomato fruit weight on the plant height. It is likely that the standard deviation of plant height played an important role in the salt-treated plants because of their smaller plant size and lower degree of sprawling. Hence, plant height was a stronger determinant of plant size overall, compared with the control plants.

Prediction of Biomass and Yield of All Plants

Distribution of Predictions for All Plants

The average field-derived fresh shoot mass, fruit numbers, and yield mass for all plants remaining at harvest was 715.29, 532.06, and 226.96 g, respectively, whereas those for the control plants were 1070.67, 810.00, and 362.55 g, respectively, while those of the salt-treated plants were 355.56, 241.89, and 87.47 g, respectively. When assessing the field-based average values against those predicted from all eight UAV image data sets (November 23-January 14) using the random forest models developed on (and applied to) all plants, it was found that all the average predicted values were within 16.97 g (2.37%), 20.12 (3.78%), and 9.59 g (4.23%) for fresh shoot mass, fruit numbers, and yield mass, respectively (**Figure 4**). This indicates that the average biomass and yield can be predicted with high accuracy as much as 8 weeks in advance, even with the disruptive sandstorms and increasing salt-stress affecting those plants remaining at harvest. However, the range of values for the field observations was much larger than from the predictions of the earliest UAV data collections (e.g., see interquartile range between field observations and predictions for November 23 in **Figure 4**). The whisker and interquartile range of fresh shoot mass, fruit numbers, and yield mass predictions approached those of the field observations the closer to harvest the UAV data were collected (**Figure 4**). Therefore, while the average fresh shoot mass, fruit numbers, and yield mass could be accurately mapped well in advance of harvest, the prediction accuracy for an individual plants' fresh shoot mass, fruit numbers, and yield mass increased as a function of decreasing time until harvest. For all eight UAV data sets, the median of the predicted fruit numbers and yield mass was overestimated (**Figure 4**): although this appeared to be less of an issue for the predicted fresh shoot mass. It can also be seen from **Figure 4** that those plants with very high fresh shoot mass, fruit numbers, and yield mass were underestimated based on the UAV imagery.

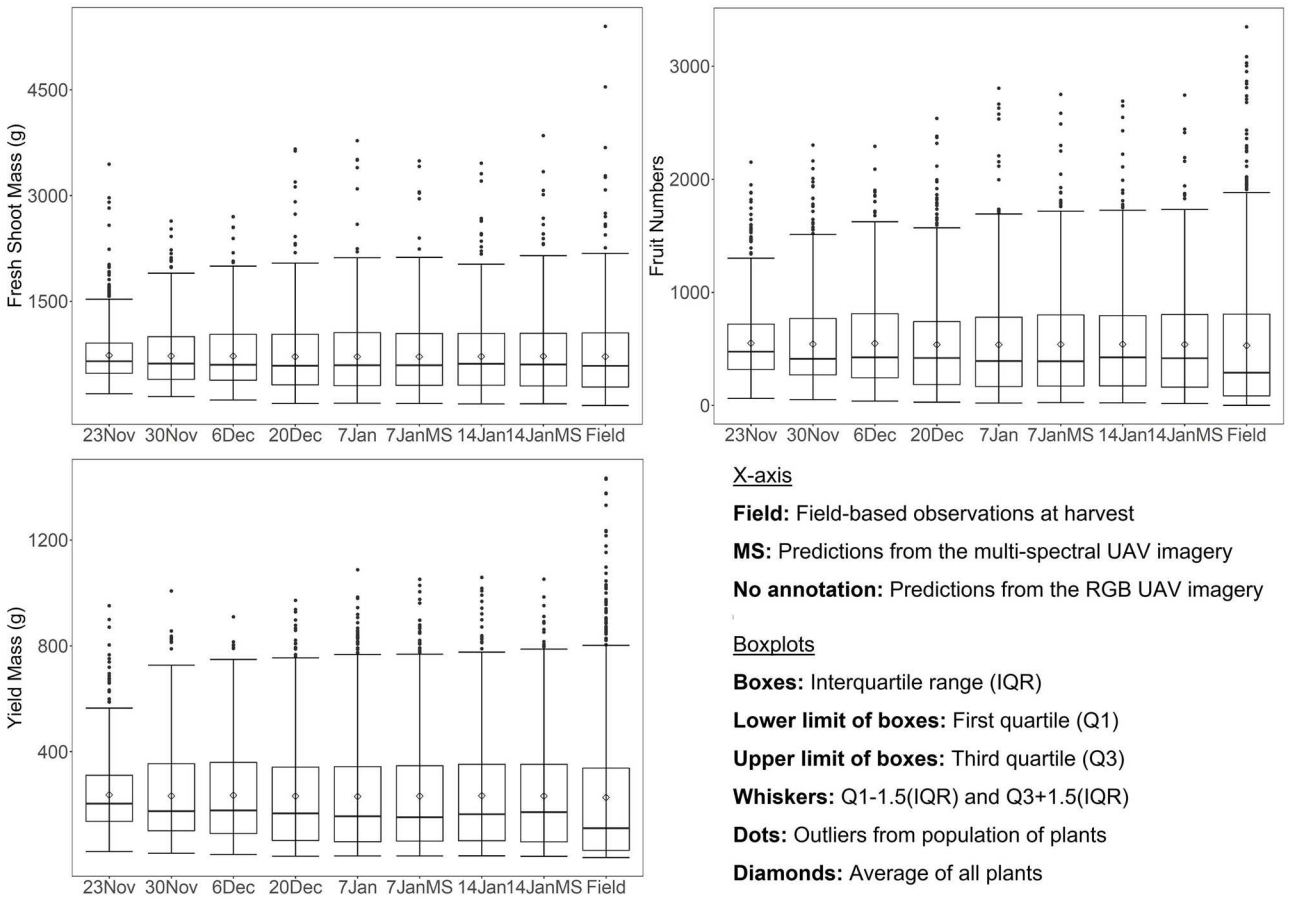

FIGURE 4 | Box-and-whisker plots, showing the variation in red–green–blue (RGB) and multispectral unmanned aerial vehicle (UAV) predicted fresh shoot mass, fruit numbers, and yield mass throughout the growing season in relation to field-based observations at the time of harvest.

Distribution of Predictions for Control and Salt-Treated Plants

When assessing the average values of predicted fresh shoot mass, fruit numbers, and yield mass for just the control plants using the random forest models developed on all plants, the first three UAV data sets (November 23 and 30, December 6) significantly underestimated average values of 108.76–238.14 g (10.16–22.24%) for fresh shoot mass, 52.14–175.87 (6.44–21.71%) for fruit numbers, and 30.00–87.85 g (8.27–24.23%) for yield mass (**Table 6**). Assessing just the salt-treated plants showed the opposite pattern, with the first three UAV data sets (November 23 and 30, December 6) producing significantly overestimated average values of 121.14–272.58 g (34.07–76.66%) for fresh shoot mass, 92.27–221.76 (38.15–91.68%) for fruit numbers, and 47.67–110.05 g (54.50–125.81%) for yield mass. Hence, the underestimation of average values for the control plants and the overestimation of the salt-treated plants balanced the average values for predicting fresh shoot mass, fruit numbers, and yield mass for all tomato plants for the first three UAV data captures. It also shows that while the overall average biomass and yield for all plants were accurately predicted, the predictions were not accurate for individual plants. This was likely because the response to the salt treatment had not yet come into full effect on November 23 and 30 and December 6, and hence, the predictions did not account for the further increase in salt treatment and consequent impact on plant growth. On the other hand, biomass and yield of control plants were underestimated as the initial impact of the salt treatment on half the plants reduced the overall average. The data sets collected on December 20 and January 7 and 14 produced much better average estimates for the control-only and salt-treated-only data sets, with all being within 14% of the average value estimated from the field data. This was because the variables extracted from the UAV imagery on those dates and used for predicting biomass and yield at harvest better represented each individual plant's condition at harvest.

Assessment of All Individual Plants From Single Date UAV Data

The initial assessment of prediction accuracies of fresh shoot mass, fruit numbers, and yield mass was based on the OOB observations of tomato plants, including both the control and the salt-treated plants. The RGB image-based results showed that in the week before harvest, the explained variance of fresh shoot mass, number of tomatoes, and yield mass were 86.60%

TABLE 6 | Percentage average over- and underestimation of predicted fresh shoot mass, fruit numbers, and yield mass in relation to the field measurements when assessing control and salt-treated plants separately.

	14-Jan MS	14-Jan	7-Jan MS	7-Jan	20-Dec	6-Dec	30-Nov	23-Nov
Fresh shoot mass—control (%)	−1.53	−2.04	−2.76	−2.45	−4.36	−10.16	−12.72	−22.24
Fresh shoot mass—salt (%)	7.03	6.74	7.19	6.04	12.49	34.07	42.84	76.66
Fruit numbers—control (%)	2.84	3.12	−0.44	−0.61	0.24	−6.44	−12.29	−21.71
Fruit numbers—salt (%)	−1.50	−1.00	10.22	8.62	6.15	38.15	52.58	91.68
Yield mass—control (%)	2.01	2.34	2.21	1.96	−0.66	−8.27	−14.24	−24.23
Yield mass—salt (%)	3.54	4.82	1.37	−0.22	13.79	54.50	73.81	125.81

Positive values indicate overestimation and negative values indicate underestimation.

(RMSE = 208.4 g, rRMSE = 29.14%), 59.46% (RMSE = 379.7, rRMSE = 71.36%), and 61.09% (RMSE = 168.9 g, rRMSE = 74.40%), respectively. Two weeks before harvest, the explained variance was slightly higher and the RMSE slightly lower than the week prior to harvest for the predicted fresh shoot mass, fruit numbers, and yield mass, using the RGB imagery (**Table 7**), possibly attributed to deterioration of some plants toward the end of the growing season. The results derived from the multispectral and RGB UAV data sets collected on January 7 and 14 were very similar and will be further compared in *Comparison of Model Results for Control and Salt-Treated Plants* below. On December 20, the explained variance of fresh shoot mass, fruit numbers, and yield mass was 79.20% (RMSE = 259.8 g, rRMSE = 36.31%), 55.90% (RMSE = 395.5, rRMSE = 74.33%), and 57.73% (RMSE = 175.7 g, rRMSE = 77.42%). On December 6, a reduction of >20% in explained variance and associated increases in RMSE was observed when comparing with December 20 for all three variables. As destructive sandstorms damaged many plants (≈9%) on December 8 and 16, the imagery collected after the sandstorm events provided more representative information on biomass and yield at harvest for the individual plants than the data collected prior to the sandstorms (Johansen et al., 2019b). The data collected on November 23 and 30 were found to be unsuitable for predicting biomass and yield at harvest for individual plants. Although the explained variance of fresh shoot mass was 46.57% on November 30, the RMSE was still more than double those for January 7 and 14.

Figure 5 provides an example, based on the multispectral UAV imagery from January 7, of the distribution of predicted fresh shoot mass, fruit numbers, and yield mass in relation to the field-based measurements. The best-fit regression lines show that there was a tendency of small plants being overestimated in terms of predicted fresh shoot mass, fruit numbers, and yield mass. However, those plants with high fresh shoot mass, fruit numbers, and yield mass had their values underestimated. The higher explained variance for fresh shoot mass was related to the ability to integrate shape features such as plant area and other plant dimensions. Those shape features and the greenness of the plants, expressed through the use of vegetation indices, are closely related to biomass (Bendig et al., 2015). Fruit numbers and yield mass have a more indirect relationship to the shape and greenness of the plants. For example, some smaller and senescent field-assessed tomato plants were identified with large numbers of fruit, which would likely have been influenced by the different types of accessions and the two treatments.

Figure 6 clearly shows the difference between the control and salt-treated tomato plants in terms of fresh shoot mass, fruit number, and yield mass based on the multispectral UAV imagery collected on January 7. Most of the plants (96.7%) in the salt-treated plots had a fresh shoot mass between 0 and 1,000 g, with only 3.3% plants appearing in the category between 1,000 and 1,500 g. Although ∼48% of the control plants appeared in the category between 0 and 1,000 g of fresh shoot mass, ∼48 and 4% occurred with fresh shoot mass weights between 1,000 and 2,000 g and 2,000 and 3,500 g, respectively. The predicted distribution in terms of fruit number and yield mass between control and salt-treated plants was similar to that of fresh shoot mass, with all but six and five plants appearing with >800 fruits and >450 g of yield mass, respectively, for the salt-treated plots. The control plots on the other hand produced fruit numbers and yield mass for some plants up to 2,751 fruits and 1,050 g of fruit, respectively.

Assessment of All Individual Plants From Multitemporal UAV Data

Combining data from different dates was also investigated to assess if the use of multitemporal information improved the prediction accuracies of fresh shoot mass, fruit numbers, and yield mass. Using the predictor variables from the multispectral imagery from both January 7 and 14, the explained variance increased by 1.66% (from 87.95 to 89.61%) and 0.10% (from 63.88 to 63.98%), and the rRMSE decreased by 1.94 and 0.04% for fresh shoot mass and fruit numbers, respectively, in comparison to the results using only data from January 7. For the yield mass predictions, the use of multispectral data from both January 7 and 14 caused a decrease in the explained variance of 0.64% and an increase in rRMSE of 0.76%. For the RGB imagery, the combination of data from January 7 and 14 and December 20 produced increases in explained variance of 1.44 (from 87.61 to 89.05%), 2.31 (from 60.61 to 62.92%), and 1.45% (from 64.37 to 65.82%) and decreases in rRMSE of 1.67, 2.01, and 1.36% for fresh shoot mass, fruit numbers, and yield mass, respectively. The addition of RGB data from November 23 and 30 reduced the prediction accuracies. As such, negligible differences in prediction accuracies were identified from using

TABLE 7 | Percentage explained variance (EV) and relative root mean square error (rRMSE) based on predictions of fresh shoot mass, fruit numbers, and yield mass for all eight unmanned aerial vehicle (UAV) data sets based on the joint analysis of control and salt-treated tomato plants.

	14-Jan MS	14-Jan	7-Jan MS	7-Jan	20-Dec	6-Dec	30-Nov	23-Nov
EV fresh shoot mass (%)	87.69	86.6	87.95	87.61	79.2	59.14	46.57	17.77
rRMSE fresh shoot mass (%)	27.90	29.14	27.64	28.02	36.31	50.89	58.23	72.21
EV fruit numbers (%)	57.86	59.46	63.88	60.61	55.9	33.77	25.3	2.15
rRMSE fruit numbers (%)	72.89	71.36	67.37	70.25	74.33	91.20	96.77	110.79
EV yield mass (%)	59.47	61.09	66.51	64.37	57.73	37.51	26.36	6.31
rRMSE yield mass (%)	76.08	74.40	69.09	71.09	77.42	94.27	102.25	115.36

The two multispectral UAV data sets are denoted with "MS," whereas the other data sets represent the RGB UAV data.

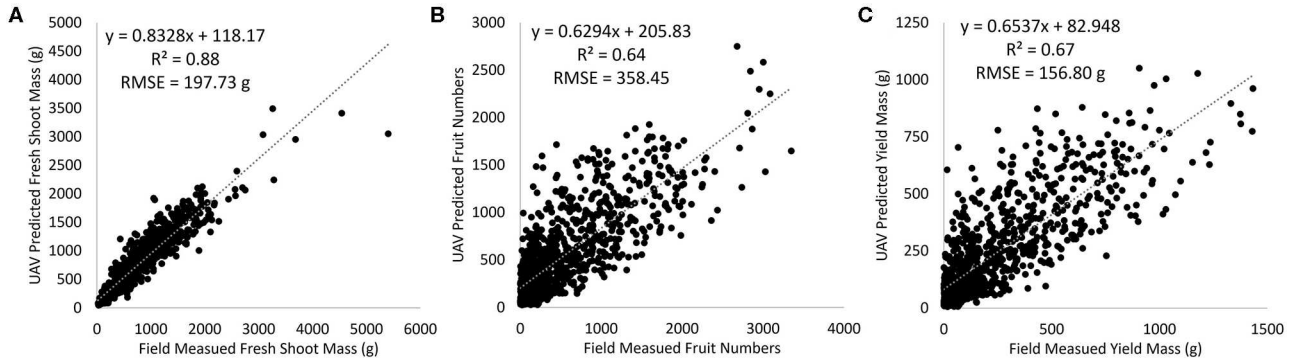

FIGURE 5 | Scatterplots showing the linear relationships between field-measured and predicted **(A)** fresh shoot mass, **(B)** fruit numbers, and **(C)** yield mass based on the multispectral unmanned aerial vehicle (UAV) imagery collected on January 7, 2018.

multitemporal data as opposed to a single date (January 7 or 14) prior to harvest.

Comparison of Model Results for Control and Salt-Treated Plants

Different random forest models were developed for (1) all tomato plants, (2) the control plants only, and (3) the salt-treated plants only. The models developed for all tomato plants were applied to all plants, the control plants only, and the salt-treated plants only, whereas the models developed for the control plants only and the salt-treated plants only were only applied to those specific subexperiments. As a general trend, the variance explained decreased and the rRMSE increased as a function of increasing time until harvest, indicating that approaching harvest time, the predictions of fresh shoot mass, fruit numbers, and yield mass improved (**Figure 7**). The random forest model developed on and applied to all (i.e., both control and salt-treated) tomato plants produced the highest proportion of explained variance and the lowest rRMSE for predicting fresh shoot mass, fruit numbers, and yield mass on December 20, January 7 and 14 (**Figure 7**). It is likely that the larger range of measurements included when using both control and salt-treated plants in the model development improved the prediction, as the salt-treated and control plants included some of the plants with the lowest and highest measures, respectively, of fresh shoot mass, fruit numbers, and yield mass.

Assessing Models Developed for Control Plants Only and Salt-Treated Plants Only

The random forest models developed on and applied to just the control plants had very similar amounts of explained variance (within 0.011) to the results produced when using the random forest model developed for all plants and applied only to the control plants for the UAV data sets collected on December 20 and January 7 and 14. On December 6, a reduction in explained variance of 3.49% was observed when using the model developed on and applied to just the control plants. However, for fresh shoot mass, fruit numbers, and yield mass, the explained variance was higher (by up to 6.29 on December 6, 5.28 on December 20, and 5.05% on December 20, respectively) between December 6 and January 14, when using the random forest models specifically developed on and applied to only the salt-treated plants compared to using the model developed on all plants (**Figure 7**). In addition, within the same period, a lowering of the RMSE was observed for predicted fresh shoot mass, fruit numbers, and yield mass of 6.06–70.15 g (1.70–19.73%), 5.27–31.05 (2.18–12.84%, excluding January 7), and 1.95–20.33 g (2.23–23.24%), respectively, using the model for the salt-treated plants only (**Figure 7**). These results indicate that models to predict biomass and yield of salt-stressed plants may need to be separately developed, whereas predicting the yield of control plants was not affected by the inclusion of salt-stressed plants within the models. The results also demonstrate the ability to predict

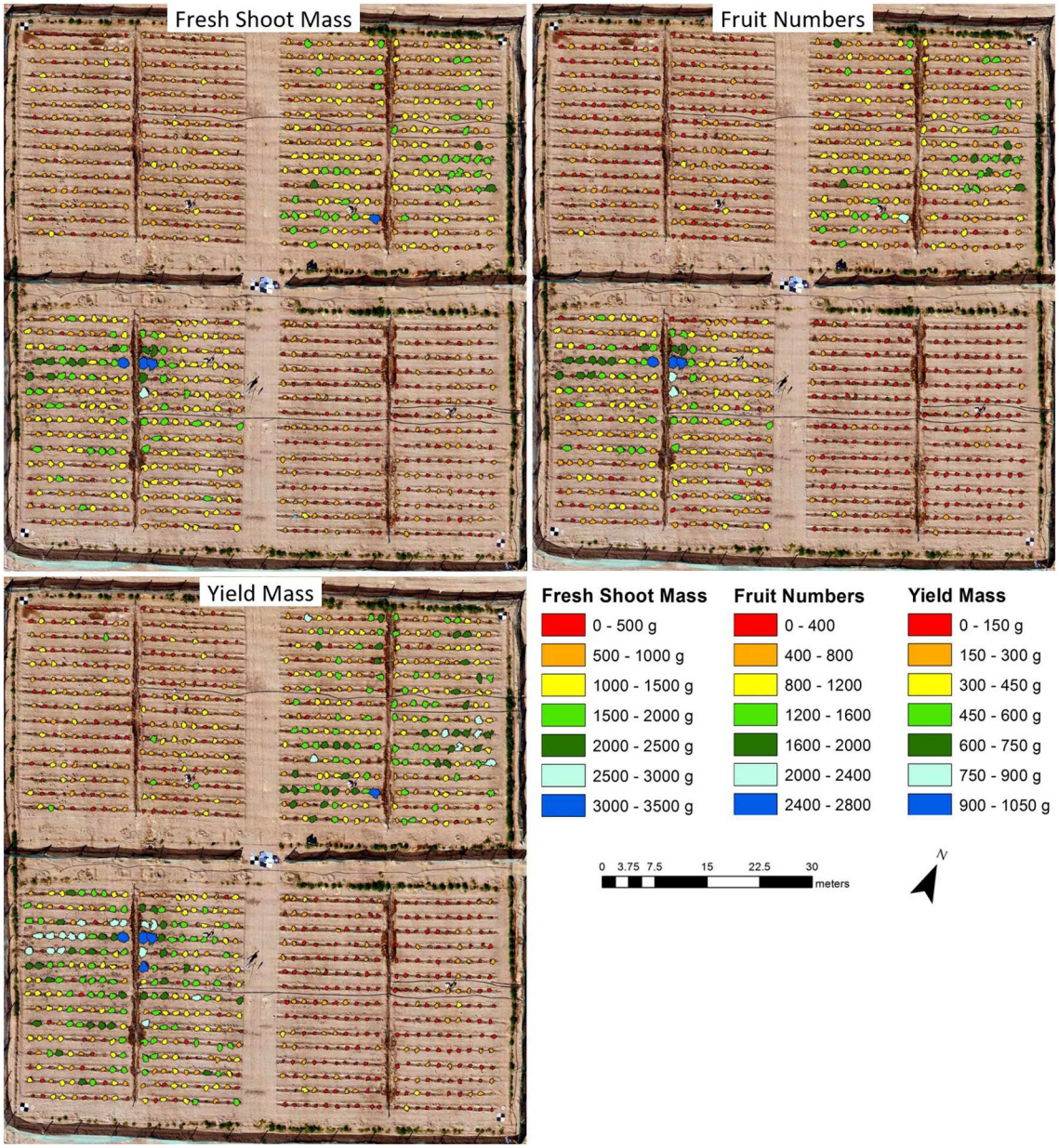

FIGURE 6 | Maps of predicted fresh shoot mass, fruit numbers, and yield mass per tomato plant based on the multispectral unmanned aerial vehicle (UAV) imagery collected on January 7.

biomass and yield of individual tomato plants up to 4 weeks before harvest.

Comparing Model Results of the Multispectral and RGB UAV Imagery

The difference in explained variance between the multispectral and RGB UAV-derived predictions of fresh shoot mass, fruit numbers, and yield mass was within 3.27% for both January 7 and 14 when assessing all plants. While the multispectral UAV data produced predictions with slightly higher explained variance for all variables than the RGB imagery collected on January 7, the RGB UAV imagery produced slightly higher explained variance for fruit numbers (1.60%) and yield mass (1.62%) for all plants on January 14 (**Figure 8**). Corresponding lowering of the RMSE on January 7 for all three variables predicted from the multispectral UAV imagery was observed, while the multispectral UAV-based RMSEs for fruit numbers and yield mass on January 14 increased slightly (8.14 and 3.81 g, respectively). An objective comparison of the results from the multispectral and RGB imagery was difficult to achieve because of the differences in suitable predictor variables (**Table 2**), number and wavelength locations of spectral bands, pixel sizes (0.50 cm for RGB and

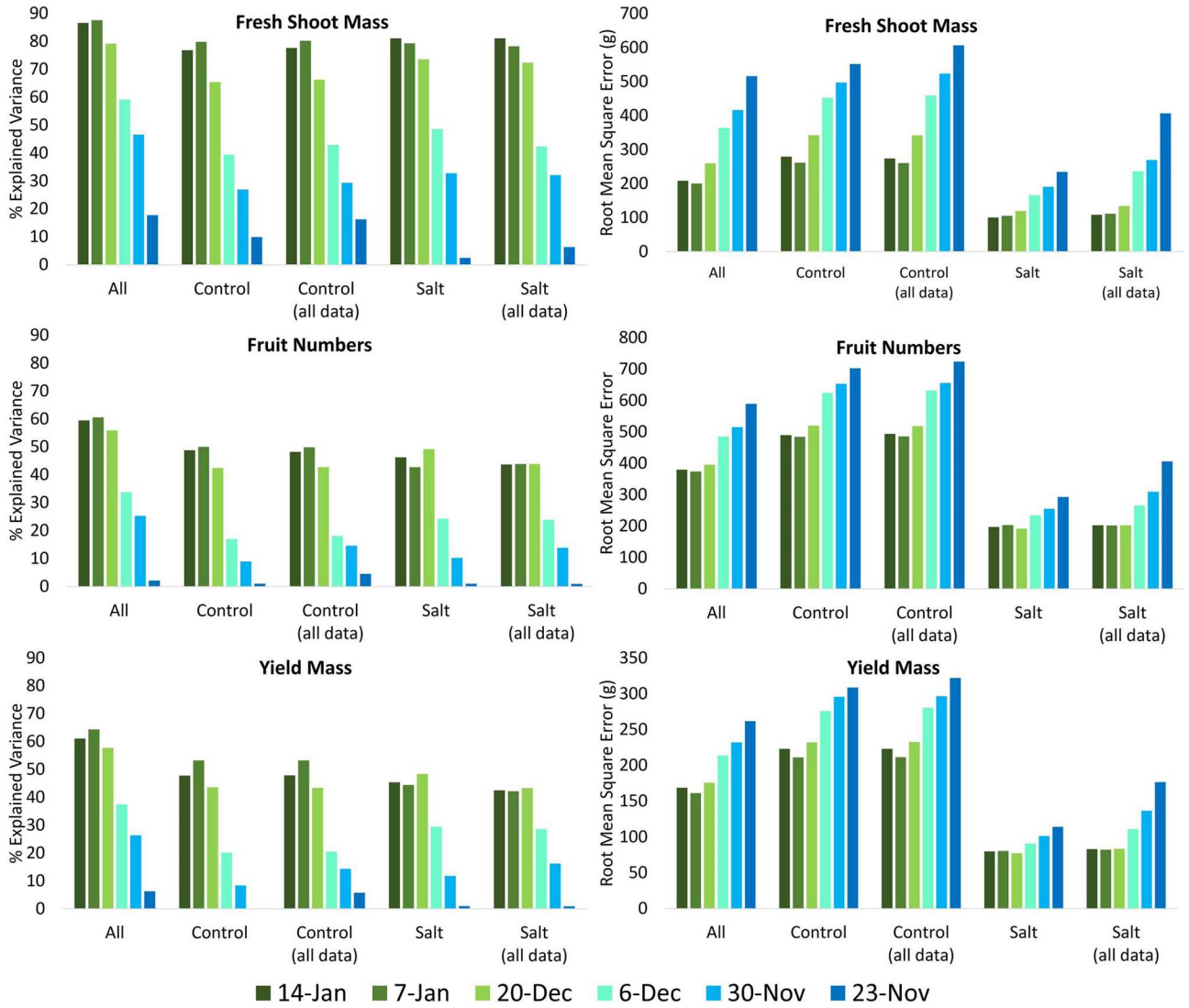

FIGURE 7 | Percentage explained variance and root mean square error of predicted measurements of fresh shoot mass, fruit numbers, and yield mass based on the RGB unmanned aerial vehicle (UAV) imagery collected on November 23 and 30, December 6 and 20, and January 7 and 14 using random forest models based on (1) all tomato plants, (2) only control plants, (3) all plants but applied only to the control plants, (4) only salt-treated plants, and (5) all plants but applied only to the salt-treated plants.

1.12 cm for multispectral), camera specifications (12 MP for RGB vs. 1.2 MP for the multispectral), as well as field of view and focal length differences, among others. Despite this, **Figure 8** shows that neither multispectral nor RGB UAV imagery were clearly advantageous in any combination of variables, models, and treatments, as long as accurate delineation of tomato plants can be achieved for derivation of plant shape features, as these were the most important variables for the prediction of biomass and yield.

For instance, when comparing the multispectral and RGB UAV imagery for the control plants only, the RGB imagery explained the highest amount of variance for fruit numbers and yield mass using both the random forest model based on the control plants (3.72–4.81% higher) and on all plants (2.89–3.65% higher) on January 14 (**Figure 8**). In contrast, the multispectral imagery could be used to predict fresh shoot mass with an explained variance of 1.43 and 1.46% higher than the RGB imagery for the control plants using control plants only and all plants for modeling, respectively. On January 7, the multispectral UAV imagery produced the highest proportion of explained variance in all cases for the control plants (0.32–5.83% higher than the RGB imagery), with the highest difference of 5.83% for fruit numbers when using a model developed on all plants and applied only to the control plants (**Figure 8**).

When predicting the salt-treated plants only based on the random forest model specifically developed on only the salt-treated plants, the multispectral UAV imagery produced higher proportions of explained variance than the RGB UAV imagery

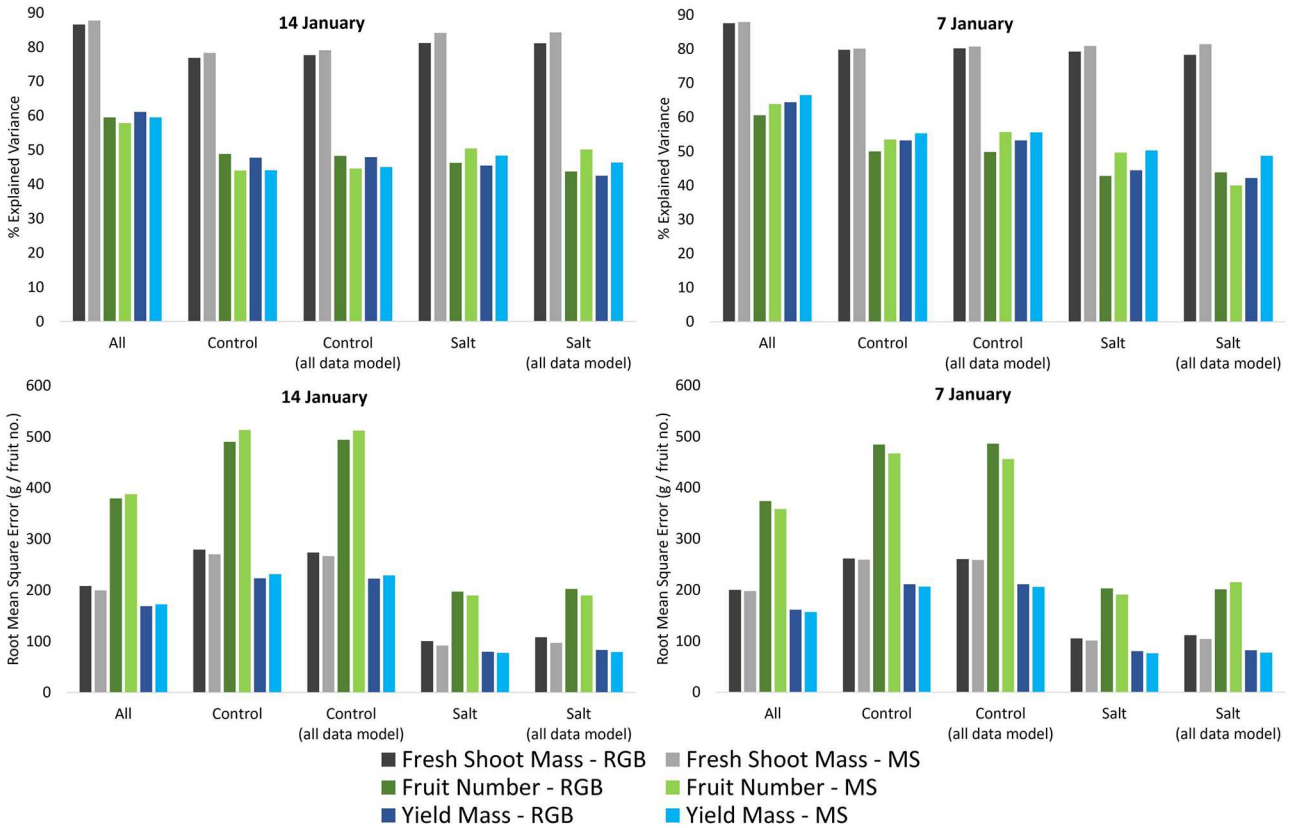

FIGURE 8 | Comparison between the percentage explained variance and root mean square error of predicted measurements of fresh shoot mass, fruit numbers, and yield derived from the multispectral and red–green–blue (RGB) unmanned aerial vehicle (UAV) imagery acquired on January 7 and 14, using random forest models based on (1) all tomato plants, (2) only control plants, (3) all plants but applied only to the control plants, (4) only salt-treated plants, and (5) all plants but applied only to the salt-treated plants.

for both January 7 (higher by 1.66% for fresh shoot mass, 5.82% for yield mass, and 6.89% for fruit numbers) and 14 (higher by 2.90% for yield mass, 3.01% for fresh shoot mass, and 4.13% for fruit numbers). Similar observations were identified when using the random forest model based on all plants for salt-treated plants to predict their fresh shoot mass, fruit numbers, and yield mass, where the multispectral UAV imagery explained the highest amount of variance (3.13–6.51% higher) for both January 7 and 14, with the exception of fruit numbers on January 7, when the RGB UAV imagery (43.86%) explained 3.88% more variance than the multispectral UAV imagery (39.98%) (**Figure 8**). Overall, the increase in explained variance were generally accompanied by a decrease in RMSE.

DISCUSSION

Model Transferability of UAV Data

Using the random forest machine learning approach, our results showed that UAV imagery collected within 4 weeks of harvest provided the best results for predicting biomass and yield at harvest for individual tomato plants. Our results also indicated that separate random forest models for predicting yield of salt-stressed plants might be required. In contrast, the yield of control plants could be predicted with similar accuracies when using models developed on both all plants and control plants only. This may be attributed to the fact that the control plants covered a similar range of values to that for all plants, whereas the salt-treated plants mainly covered the lower range of recorded values. For instance, 157 (out of 497) control plants had higher numbers of fruit yield than any of the salt-treated plants. Hence, those models developed specifically on and applied to the salt-treated plants performed better than using those developed on all plants when applied to the salt-treated plants. This emphasizes the importance of carefully selecting data sets in terms of size, variability, and representativeness for model development to ensure transferability (Liu et al., 2018; Maxwell et al., 2018; Ma et al., 2019). Model transferability will also require data that are representative in other contexts, e.g., including multiple growing seasons, different areas, different climate and weather conditions, different soils, and different management practices (Maxwell et al., 2018). Hence, future work should focus on assessing the model transferability of machine learning approaches for UAV-based mapping applications. It will be valuable for growers to

know if the same machine learning model can be used under different contexts or if the same model can be used universally for the same plant species.

The analysis included 199 *S. pimpinellifolium* accessions and one *S. lycopersicum* accession (the commercial tomato, Heinz 1706). Considering the high mapping accuracies of the UAV data collected in the weeks preceding harvest, the developed models did not seem to be affected by the large variety of accessions included in the training data. The three salt-treated *S. lycopersicum* plants all died prior to harvest, and while the biomass of the three control plants was predicted (based on the multispectral UAV imagery collected on January 7) to be within 6.84–14.48% of the field-based observations (better than the RMSE and rRMSE of 197.7 g and 27.64%, respectively, **Figure 5A**), the predicted yield was significantly overestimated, with two plants not producing fruit and one plant producing 630 fewer fruits and 283 g of yield mass less than predicted. Despite the small sample size of the *S. lycopersicum* species, these results raise questions on whether the developed models can be employed for different tomato plant species. Hence, future studies testing model transferability of machine learning approaches should also test different species.

The Issue of Data Dimensionality and Errors

The availability of UAV data for supporting smart farming is expected to grow significantly in the future (Wolfert et al., 2017; Liakos et al., 2018). The dimensionality of available data is also expected to increase, with the availability of hyperspectral imagery providing hundreds of spectral bands for analysis (Torresan et al., 2016; Adao et al., 2017). Likewise, additional sensors provide new observation capacity, e.g., LiDAR data providing three-dimensional information on crops, or thermal data providing temperature measurements at high spatial resolution, both of which are increasingly being integrated with other information and used for crop assessment (Calderon et al., 2013; Ivushkin et al., 2019). The growing availability of such UAV-based data sets will likely make predictions of biomass, yield, and other biophysical and biochemical properties not only more accurate but also more complex. In many instances, this results in higher computational costs and longer processing times, which limits the proficiency of real-time delivery. With the increasing UAV data dimensionality, machine learning approaches become the only feasible option for big data analytics. In most cases, preprocessing approaches (standard UAV processing workflow to produce orthomosaics, geo-referencing, radiometric corrections, etc.) are time consuming for large data sets. To increase the ability to achieve well-calibrated and analyzed near real-time UAV map outputs from big data, machine learning approaches suitable for converting raw data into final map outputs should be explored (Yang et al., 2020).

While deep learning models are designed for high-quality data feature learning (Zhang et al., 2018), some research has experimented with deep learning models for low-quality data. Approaches such as those by Vincent et al. (2010) and Wang and Tao (2016) have focused on data denoising routines and identifying reliable features within corrupted data, and these might be explored in future research for deep learning models applied to multiple UAV data sets collected under various conditions and with different acquisition parameters. Zhang et al. (2018) also discussed a multimodel deep learning approach specifically suited to heterogeneous data, which might be explored in future research as well to regionalize UAV data sets for optimization of analysis and results. Despite the availability of some approaches suited for reducing the effects of uncertainties and noise inherent in UAV data, further exploration is still required to effectively reduce processing time and alleviate the need for complex intermediate processing steps of UAV data, preventing near real-time delivery of information on crop variables. For instance, uncertainty in UAV-based thermal data may be introduced by wind speed, wind direction, and flight direction. UAV optical data collected by RGB and multispectral and hyperspectral sensors are all sensitive to the time, season, and latitude of data acquisition, as that will affect the solar angle and shadowing effects. In addition, the quality of sensor calibration may impact data quality (Barreto et al., 2019). The current basic preprocessing chain of optical UAV data includes many steps, where multiple filtering modes, blending modes, color correction, and interpolation approaches will affect the orthomosaic and hence potentially introduce data noise. In addition, the spectral and radiometric properties of different camera systems tend to differ, making direct comparison of data unfeasible (Tmusic et al., 2020). In addition to all of these potential uncertainties, flight planning parameters, such as forward overlap, sidelap, speed, flight direction, and flying height (Tu et al., 2020), as well as weather conditions during data acquisition, e.g., wind speed and direction, cloud shadows, dust near the ground, and variations in the atmosphere's composition (Zhang et al., 2014; Ziliana et al., 2018), will all influence and affect the quality of the acquired data. With all of these potential issues introduced inherent in UAV data, it is important that future work explore the sensitivity of machine learning approaches to these uncertainties.

Variable Standardization of Model Inputs

Ultimately, our research showed that plant-based shape parameters had the highest importance for predicting biomass and yield of the tomato plants using random forest models. As long as the UAV-derived imagery is geometrically registered and a suitable delineation approach can be developed, the mapping of shape parameters is likely to be influenced less than the spectral characteristics when using cameras with different spectral and radiometric characteristics or if flight planning parameters vary between data acquisitions. Hence, if integrating multiple data sets from different UAV-based sources for mapping tomato plants, shape parameters should be a main focus to ensure consistency between data sets. Vegetation indices proved useful in this research as well, and they have been utilized previously to reduce atmospheric effects and limit the need for image calibration (Lillesand et al., 2015; Xue and Su, 2017; Fernandez-Gallego et al., 2019). However, ratio-based indices are still affected by the spectral resolution and band width of the sensor used. Vegetation indices often fail to account for canopy-background interactions and canopy bidirectional reflectance anisotropies, especially those associated with shadowing effects, and become insensitive to vegetation with high leaf area index

values, which leads to insensitivities to vegetation variation of dense plants and trees (Gitelson et al., 1996; Sims and Gamon, 2002a; Asner and Warner, 2003). As such, the application of vegetation indices for mapping biophysical and biochemical parameters lacks generality, making them time, space, and crop type specific (Houborg et al., 2007). Mapping actual biophysical and biochemical parameters from optical data and using these as model inputs for predicting crop parameters such as biomass and yield might ensure further standardization when integrating data collected for different areas and from different sensors with machine learning approaches (Houborg et al., 2007; Gholizadeh et al., 2015). That will generally require field-based measurements to be collected for model calibration purposes. In fact, the use of high-quality field data is imperative for both calibration and validation to standardize UAV-derived outputs suited as model input for prediction of crop parameters (Von Bueren et al., 2014). Radiometric correction of imagery is equally important to ensure consistency of optical image data over time and between sites (Jeong et al., 2018; Tmusic et al., 2020). Hence, despite the large array of UAV image data sets and acquisition and processing routines for crop assessment, potential avenues exist to improve the consistency of diverse data used as input for machine learning approaches.

CONCLUSIONS

A novel approach for using UAV-based imagery collected at various intervals prior to harvest was employed to predict fresh shoot mass, fruit numbers, and yield mass of tomato plants at harvest using a random forest machine learning approach. Shape features derived from individual tomato plants were determined to be the most important predictor variables, followed by vegetation indices and image texture. While the average biomass and yield at the field level could accurately be predicted up to 8 weeks prior to harvest, a significant reduction in prediction accuracy of biomass and yield of individual plants was identified when using UAV imagery collected more than 4 weeks before harvest. This was attributed to sandstorm events, where the imagery collected after the sandstorms provided more representative predictions of biomass, fruit numbers, and yield mass at harvest for individual plants than the data collected prior to the sandstorms. Models specifically developed for predicting yield from the salt-stressed plants increased the explained variance by up to 6.29% (relative to those models for all plants), whereas little (<1.1% explained variance) variation occurred in the results for predicting yield of the control plants irrespective of which models were used.

The research demonstrates the suitability of using UAV imagery and a random forest machine learning approach for biomass and yield prediction of tomato plants but highlights the need for careful consideration in terms of data inputs (e.g., parameters of control vs. salt-stressed plants) for model development. It is important to be mindful of data quality both with regard to UAV-based imagery and field-based observations, as machine learning approaches are inevitably influenced by errors. It is therefore imperative to follow standardized procedures when extracting data used as input into machine learning algorithms. Future work should compare the results of different machine and deep learning approaches for predicting biomass, yield, and other biophysical and biochemical properties of agricultural crops and explore the sensitivities of these approaches to typical UAV data inconsistencies (preprocessing steps, cloud shadow contamination of imagery, sensor noise, etc.). Assessing the transferability of developed machine and deep learning models to test their applicability to a wider context also needs to be explored (e.g., for UAV imagery collected of the same crops but for different areas or during different growing seasons for the same area). Finally, heritability calculations to evaluate the variation between phenotypic traits of the tomato plants in response to genetic variation among the different tomato plant accessions will be an important extension of this research. This research and identified future directions may provide growers with valuable UAV-derived information on how to manage plant growth, increase yield, and obtain advance knowledge on harvesting, sales, and distribution requirements of tomatoes and other fruit.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

MT conceived the whole plant experiment and was involved in all aspects of the project. MFM designed the UAV-focused aspects of the experiment, as well as associated ground-based data collections, and was involved in all aspects of the project. KJ undertook all UAV image processing and analysis and led the writing of the manuscript, with MFM, MT, and MJLM also contributing. MJLM and MT designed the ground-based component of the plant experiment and, with GF, SN, and MAAM, coordinated the experiment, including field data collection and the final harvest. KJ, YM, BA, SA-M, MZ, and YA were responsible for field equipment and collection of the UAV imagery and field data. MAAM led a team of workers to undertake planting, irrigation, fertilization, observation, and washing of plants after sandstorms, and harvesting.

FUNDING

The research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST) Office and Sponsored Research (OSR) under Award No. 2302-01-01.

ACKNOWLEDGMENTS

We would like to thank all the workers at the King Abdulaziz University Agricultural Research Station in Hada Al-Sham for their extensive help with removal of weeds, plant maintenance, and harvesting. Khadija Zemmouri and Dinara Utarbayeva prepared plots and undertook sowing of all plants.

REFERENCES

Adao, T., Hruska, J., Padua, L., Bessa, J., Peres, E., Morais, R., et al. (2017). Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens.* 9:1110. doi: 10.3390/rs9111110

Ahmed, O. S., Shemrock, A., Chabot, D., Dillon, C., Williams, G., Wasson, R., et al. (2017). A remote sensing technique for detecting Laurel Wilt disease in avocado in presence of other biotic and abiotic stresses. *Comput. Electron. Agric.* 156, 549–557. doi: 10.1016/j.compag.2018.12.018

Asner, G. P., and Warner, A. S. (2003). Canopy shadow in IKONOS satellite observations of tropical forests and savannas. *Remote Sens. Environ.* 87, 521–533. doi: 10.1016/j.rse.2003.08.006

Ballesteros, R., Ortega, J. F., Hernandez, D., and Moreno, M. A. (2018). Onion biomass monitoring using UAV-based RGB imaging. *Precis. Agric.* 19, 840–857. doi: 10.1007/s11119-018-9560-y

Barbedo, J. G. A. (2019). A review on the use of Unmanned Aerial Vehicles and imaging sensors for monitoring and assessing plant stresses. *Drones* 3, 1–27. doi: 10.3390/drones3020040

Bargoti, S., and Underwood, J. P. (2017). Image segmentation for fruit detection and yield estimation in apple orchards. *J. Field Robot.* 34, 1039–1060. doi: 10.1002/rob.21699

Barreto, M. A. P., Johansen, K., Angel, Y., and McCabe, M. F. (2019). Radiometric assessment of a UAV-based push-broom hyperspectral camera. *Sensors* 19:4699. doi: 10.3390/s19214699

Belgiu, M., and Dragut, L. (2016). Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. doi: 10.1016/j.isprsjprs.2016.01.011

Bendig, J. Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl.Earth Observ. Geoinfo.* 39, 79–87. doi: 10.1016/j.jag.2015.02.012

Braga-Neto, U. M., and Dougherty, E. R. (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics* 20, 374–380. doi: 10.1093/bioinformatics/btg419

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Buchaillot, M. L., Garcia-Romero, A., Vergara-Diaz, O., Zaman-Allah, M. A., Tarekegne, A., Cairns, J. E., et al. (2019). Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques. *Sensors* 19:1815. doi: 10.3390/s19081815

Calderon, R., Navas-Cortes, J. A., Lucena, C., and Zarco-Tejada, P. J. (2013). High-resolution airborne hyperspectral and thermal imagery for early detection of Verticilliumwilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sens. Environ.* 139, 231–245. doi: 10.1016/j.rse.2013.07.031

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology* 88, 2783–2792. doi: 10.1890/07-0539.1

Escalante, H. J., Rodriguez-Sanchez, S., Jimenez-Lizarraga, M., Morales-Reyes, A., De La Calleja, J., and Vazquez, R. (2019). Barley yield and fertilization analysis from UAV imagery: a deep learning approach. *Int. J. Remote Sens.* 40, 2493–2516. doi: 10.1080/01431161.2019.1577571

FAOSTAT (2017). *Production – Crops – Area harvested / Production quantity – Tomatoes – 2014*. Food and Agriculture Organization. Available online at: www.fao.org/faostat/en (accessed 19 September, 2019)

Fathipoor, H., Arefi, H., Shah-Hosseini, R., and Moghadam, H. (2019). Corn forage yield prediction using unmanned aerial vehicle images at mid-season growth stage. *J. Appl. Remote Sens.* 13:034503. doi: 10.1117/1.JRS.13.034503

Fernandez-Gallego, J. A., Kefauver, S. C., Vatter, T., Gutierrez, N. A., Nieto-Taladriz, M. T., and Araus, J. L. (2019). Low-cost assessment of grain yield in durum wheat using RGB images. *Euro. J. Agronomy* 105, 146–156. doi: 10.1016/j.eja.2019.02.007

Flowers, T. J., and Flower, S. A. (2005). Why does salinity pose such a difficult problem for plant breeders? *Agric. Water Manage.* 78, 15–24. doi: 10.1016/j.agwat.2005.04.015

Gebbers, R., and Adamchuk, V. I. (2010). Precision agriculture and food security. *Science* 327, 828–831. doi: 10.1126/science.1182768

Gholizadeh, H., Robeson, S. M., and Rahman, A. F. (2015). Comparing the performance of multispectral vegetation indices and machine-learning algorithms for remote estimation of chlorophyll content: a case study in the Sundarbans mangrove forest. *Int. J. Remote Sens.* 36, 3114–3133. doi: 10.1080/01431161.2015.1054959

Ghosh, A., Fassnacht, F. E., Joshi, P. K., and Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensors across three spatial scales. *Int. J. Appl. Earath Observ. Geoinfo.* 26, 49–63. doi: 10.1016/j.jag.2013.05.017

Gil-Docampo, M. L., Arza-Garcia, M., Ortiz-Sanz, J., Martinez-Rodriguez, S., Marcos-Robles, J. L., and Sanchez-Sastre, L. F. (2018). Above-ground biomass estimation of arable crops using UAV-based SfM photogrammetry. *Geocarto Int.* 35, 687–699. doi: 10.1080/10106049.2018.1552322

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58, 289–298.

Gitelson, A. A., and Merzlyak, M. N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space Res.* 22, 689–692. doi: 10.1016/S0273-1177(97)01133-2

Gregorutti, B., Michel, B., and Saint-Pierre, P. (2017). Correlatioon and variable importance in random forests. *Stat. Comput.* 27, 659–678. doi: 10.1007/s11222-016-9646-1

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., et al. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods* 15, 1–19. doi: 10.1186/s13007-019-0394-z

Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* 3, 610–621.

Hassan, M. A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., et al. (2019). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Sci.* 282, 95–103. doi: 10.1016/j.plantsci.2018.10.022

Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi: 10.1038/s41587-019-0152-9

Holman, F. H., Riche, A. B., Michalski, A., Castle, M., Wooster, M. J., and Hawkesford, M. J. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sens.* 8:1031. doi: 10.3390/rs8121031

Houborg, R., Soegaard, H., and Boegh, E. (2007). Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data. *Remote Sens. Environ.* 106, 39–58. doi: 10.1016/j.rse.2006.07.016

Hu, P., Chapman, S. C., Wang, X., Potgieter, A., Duan, T., Jordan, D., et al. (2018). Estimation of plant height using a high throughput phenotyping platform based on unmanned aerial vehicle and self-calibration: example of sorghum breeding. *Eur. J. Agron.* 95, 24–32. doi: 10.1016/j.eja.2018.02.004

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Franceschini, M. H. D., Kramer, H., et al. (2019). UAV based soil salinity assessment of cropland. *Geoderma* 338, 502–512. doi: 10.1016/j.geoderma.2018.09.046

Jaleel, C. A., Manivannan, P., Wahid, A., Farooq, M., Al-Juburi, H. J., Somasundaram, R., et al. (2009). Drought stress in plants: a review on morphological characteristics and pigments composition. *Int. J. Agric. Biol.* 11, 100–105.

Jeong, Y., Yu, J., Wang, L., Shin, H., Koh, S.-M., and Park, G. (2018). Cost-effective reflectance calibration method for small UAV images. *Int. J. Remote Sens.* 39, 7225–7250. doi: 10.1080/01431161.2018.1516307

Jiang, Q., Fang, S., Peng, Y., Gong, Y., Zhu, R., Wu, X., et al. (2019). UAV-based biomass estimation for rice-combining spectral, TIN-based structural and meteorological features. *Remote Sens.* 11:890. doi: 10.3390/rs11070890

Johansen, K., Coops, N., Gergel, S., and Stange, Y. (2007). Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sens. Environ.* 110, 29–44. doi: 10.1016/j.rse.2007.02.014

Johansen, K., Morton, M., Malbeteau, Y., Aragon, B., Al-Mashharawi, S., Ziliani, M., et al. (2019b). "Predicting biomass and yield at harvest of salt-stressed tomato plants using UAV imagery. the international archives of

the photogrammetry, remote sensing and spatial information sciences, XLII-2/W13, 407-411," *UAV-g Conference*, (Enschede), 10–12.

Johansen, K., Morton, M. J. L., Malbeteau, Y. M., Aragon, B., Al-Mashharawi, S. K., Ziliani, M. G., et al. (2019a). Unmanned Aerial Vehicle-based phenotyping using morphometric and spectral analysis can quantify responses of wild tomato plants to salinity stress. *Front. Plant Sci.* 10:370. doi: 10.3389/fpls.2019.00370

Johansen, K., Phinn, S., and Taylor, M. (2015). Mapping woody vegetation clearing in Queensland, Australia from Landsat imagery using the Google Earth Engine. *Remote Sens. Appl.* 1, 36–49. doi: 10.1016/j.rsase.2015.06.002

Johansen, K., Raharjo, T., and McCabe, M. F. (2018). Using multi-spectral UAV imagery to extract tree crop structural properties and assess pruning effects. *Remote Sens.* 10:854. doi: 10.3390/rs10060854

Jorge, J., Vallbe, M., and Soler, J. A. (2019). Detection of irrigation inhomogeneities in an olive grove using the NDRE vegetation index obtained form UAV images. *Eur. J. Remote Sens.* 52, 169–177. doi: 10.1080/22797254.2019.1572459

Kamilaris, A., Kartakoullis, A., and Prenafeta-Boldu, F. X. (2017). A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* 143, 23–37. doi: 10.1016/j.compag.2017.09.037

Kekre, H. B., Thepade, S. D., Sarode, T. K., and Suryawanshi, V. (2010). Image retrieval using texture features extracted from GLCM, LBG and KPE. *Int. J. Comput. Theory Eng.* 2, 695–700. doi: 10.7763/IJCTE.2010.V2.227

Koh, L. P., and Wich, S. A. (2012). Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation. *Trop. Conserv. Sci.* 5, 121–132. doi: 10.5167/uzh-72781

Kragh, M., and Underwood, J. (2019). Multimodal obstacle detection in unstructured environments with conditional random fields. *J. Field Robot.* 37, 53–72. doi: 10.1002/rob.21866

Lelong, C. C. D., Berger, P., Jubelin, G., Roux, B., Labbe, S., and Baret, F. (2008). Assessment of Unmanned Aerial Vehicle imagery for quantitative monitoring of wheat crop in small plots. *Sensors* 8, 3357–3585. doi: 10.3390/s8053557

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: a review. *Sensors* 18:2674. doi: 10.3390/s18082674

Liaw, A., and Wiener, M. (2002). Classification and regression by randForest. *R News* 21, 17–22.

Lillesand, T., Kiefer, R. W., and Chipman, J. (2015). *Remote Sensing and Image Interpretation. 7th Edn*. Hoboken, NJ: Wiley.

Liu, Y., Chen, X., Wang, Z., Wang, Z. J., Ward, R. K., and Wang, X. (2018). Deep learning for pixel-level image fusion: recent advances and future prospects. *Inform. Fusion* 42, 158–173. doi: 10.1016/j.inffus.2017.10.007

Lofstedt, T., Brynolfsson, P., Asklund, T., Nyholm, T., and Garpebring, A. (2019). Gray-level invariant Haralick texture features. *PLoS ONE*. 14:e0212110. doi: 10.1371/journal.pone.0212110

Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., et al. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods* 15, 1–16. doi: 10.1186/s13007-019-0402-3

Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., et al. (2017). Evaluation of feature selection methods for object-based land cover mapping of Unmanned Aerial Vehicle imagery using random forest and support vector machine classifiers. *ISPRS Int. J. Geo-Informa.* 6:51. doi: 10.3390/ijgi6020051

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019). Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* 152, 166–177. doi: 10.1016/j.isprsjprs.2019.04.015

Machado, R. M. A., and Serralheiro, R. P. (2017). Soil salinity: Effects on vegetable crop growth. management practices to prevent and mitigate soil salinization. *Hurticulturae* 3:30. doi: 10.3390/horticulturae3020030

Madec, S., Baret, F., de Solan, B., Thomas, S., Sutartre, D., Jezequel, S., et al. (2017). High-Throughput phenotyping of plant height: comparing Unmanned Aerial Vehicles and ground LiDAR estimates. *Front. Plant Sci.* 8:2002. doi: 10.3389/fpls.2017.02002

Maxwell, A. E., Warner, T. A., and Fang, F. (2018). Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* 39, 2784–2817. doi: 10.1080/01431161.2018.1433343

Messerer, M., Lang, D., and Mayer, K. F. X. (2018). Analysis of stress resistance using next generation techniques. *Agronomy* 8:130. doi: 10.3390/agronomy8080130

Moeckel, T., Dayananda, S., Nidamanuri, R. R., Nautiyal, S., Hanumaiah, N., Buerkert, A., et al. (2018). Estimation of vegetation crop parameter by multi-temporal UAV-borne images. *Remote Sens.* 10:805. doi: 10.3390/rs10050805

Morton, M. J. L., Awlia, M., Al-Tamimi, N., Saade, S., Pailles, Y., Negrao, S., et al. (2019). Salt stress under the scalpel – dissecting the genetics of salt tolerance. *Plant J.* 97, 148–163. doi: 10.1111/tpj.14189

Motohka, T., Nasahara, K. N., Oguma, H., and Tsuchida, S. (2010). Applicability of green-red vegetation index for remote sensing of vegetation phenology. *Remote Sens.* 2, 2369–2387. doi: 10.3390/rs2102369

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Nasi, R., Viljanen, N., Oliveira, R., Kaivosoja, J., Niemelainen, O., Hakala, T., et al. (2018). "Optimizing radiometric processing and feature extraction of drone based hyperspectral frame format imagery for estimation of yield quantity and quality of a grass sward," in: ISPRS TC III Mid-term Symposium "Developments, Technologies and Applications in Remote Sensing", Beijing, China, 7-10 May 2018. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Nevavuori, P., Narra, N., and Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.* 163:104859. doi: 10.1016/j.compag.2019.104859

Niu, Y., Zhang, L., Zhang, H., Han, W., and Peng, X. (2019). Estimating above-ground biomass of maize using features derived from UAV-based RGB imagery. *Remote Sens.* 11:1261. doi: 10.3390/rs11111261

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., and Pereira, J. M. C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. *Forest Ecol. Management*, 275, 117–129. doi: 10.1016/j.foreco.2012.03.003

Peng, Y., Zhu, T., Li, Y., Dai, C., Fang, S., Gong, Y., et al. (2019). Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications. *Agric. Forest Meteorol.* 271, 116–125. doi: 10.1016/j.agrformet.2019.02.032

Pitman, M. G., and Lauchli, A. (2002). "Global impact of salinity and agricultural ecosystems," in *Salinity: Environment – Plants – Molecules*, eds Lauchli, A., and Luttge, U (Dordrecht: Kluwer Academic Publishers), 3–20. doi: 10.1007/0-306-48155-3_1

Prasad, A. M., Iverson, L. R., and Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199. doi: 10.1007/s10021-005-0054-1

Rao, E. S., Kadirvel, P., Symonds, R. C., and Ebert, A. W. (2013). Relationship between survival and yield related traits in Solanum pimpinellifolium under salt stress. *Euphytica* 190, 215–228. doi: 10.1007/s10681-012-0801-2

Robson, A., Rahman, M., and Muir, J. (2017). Using Worldview satellite imagery to map yield in avocado (Persea americana): a case study in Bundaberg, Australia. *Remote Sens.* 9:1223. doi: 10.3390/rs9121223

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1973). "Monitoring vegetation systems in the Great Plains with ERTS," in *Third Earth Resources Technology Satellite-1 Symposium,* NASA SP-351, vol. 1 (Washington, DC; NASA), 309–317.

Sarron, J. M., E., and Sane, C. A. B., Faye E (2018). Mango yield mapping at the orchard scale based on tree structure and land cover assessed by UAV. *Remote Sens.* 10:1900. doi: 10.3390/rs10020226

Sato, K., Shimada, S., Sekiyama, A., Toyoda, H., Yanai, T., and Nori, N. (2016). Extraction of near-infrared information from aerial photography using UAV with supergel color filter aiming for *in-situ* tea growing monitoring. *J. Remote Sens. Soc. Japan* 36, 131–135. doi: 10.11440/rssj.36.131

Senthilnath, J., Dokania, A., Kandukuri, M., Ramesh, K. N., Anand, G., and Omkar, S. N. (2016). Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosyst. Eng.* 146, 16–32. doi: 10.1016/j.biosystemseng.2015.12.003

Shi, D., and Yang, X. (2016). An assessment of algorithmic parameters affecting image classification accuracy by random forests. *Photogramm. Eng. Remote Sensing* 82, 407–417. doi: 10.14358/PERS.82.6.407

Shi, Y., Thomasson, J. A., Murray, S. C., Pugh, N. A., Rooney, W. L., Shafian, S., et al. (2016). Unmanned Aerial Vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE* 11:e0159781. doi: 10.1371/journal.pone.0159781

Sims, D. A., and Gamon, J. A. (2002a). Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sens. Environ.* 84, 526–537. doi: 10.1016/S0034-4257(02)00151-7

Sims, D. A., and Gamon, J. A. (2002b). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81, 337–354. doi: 10.1016/S0034-4257(02)00010-X

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., et al. (2019). High-Throughput phenotyping enabled genetic dissection of crop lodging in wheat. *Front. Plant Sci.* 10:394. doi: 10.3389/fpls.2019.00394

Stavridou, E., Hastings, A., Webster, R. J., and Robson, P. R. H. (2017). The impact of soil salinity on the yield, composition and physiology of the bioenergy grass Miscanthus 3 giganteus. *Glob. Change Biol. Bioenergy* 9, 92–104. doi: 10.1111/gcbb.12351

Sugiura, R., Itoh, A., Nishiwaki, K., Murakami, N., Shibuya, Y., Hirafuji, M., et al. (2015). "Development of high-throughput field phenotyping system using imagery from unmanned aerial vehicle," in: *American Society of Agricultural and Biological Engineers Annual International Meeting 2015* (New Orleans, LA), 126–133.

Tmusic, G., Manfreda, S., Aasen, H., James, M. R., Goncalves, G., Ben-Dor, E., et al. (2020). Current practices in UAS-based environmental monitoring. *Remote Sens.* 12:1001. doi: 10.3390/rs12061001

Torresan, C., Berton, A., Carotenuto, F., Di Gennaro, S. F., Gioli, B., Matesa, A., et al. (2016). Forestry applications of UAVs in Europe: a review. *Int. J. Remote Sens.* 38, 2427–2447. doi: 10.1080/01431161.2016.1252477

Trimble eCognition Developer (2017). *Reference book. Trimble Documentation eCognition Developer 9.3, document version 9.3.0, revision 1.0, November 2017.* Munich: Trimble Germany GmbH.

Tu, Y., Johansen, K., Phinn, S., and Robson, A. (2019). Measuring canopy structure and condition using multi-spectral UAS imagery in a horticultural environment. *Remote Sens.* 11:269. doi: 10.3390/rs11030269

Tu, Y., Phinn, S., Johansen, K., Robson, A., and Wu, D. (2020). Optimising drone flight planning for measuring horticultural tree crop structure. *ISPRS Photogramm. Remote Sens. J.* 160, 83–96. doi: 10.1016/j.isprsjprs.2019.12.006

Underwood, J. P., Hung, C., Whelan, B., and Sukkarieh, S. (2016). Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. *Comput. Electron. Agric.* 130, 83–96. doi: 10.1016/j.compag.2016.09.014

Verslues, P. E., Agarwal, M., Katiyar-Agarwal, S., Zhu, J., and Zhu, J.-K. (2006). Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J.* 45, 523–539. doi: 10.1111/j.1365-313X.2005.02593.x

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* 11, 3371–3408.

Von Bueren, S., Burkart, A., Hueni, A., Rascher, U., Tuohy, M., and Yule, I. (2014). Comparative validation of UAV based sensors for the use in vegetation monitoring. *Biogeosci. Discussions* 11, 3837–3864. doi: 10.5194/bgd-11-3837-2014

Wang, C., and Myint, S. W. (2015). A simplified empirical line method of radiometric calibration for small Unmanned Aircraft systems-based remote sensing. *IEEE J. Selected Top. Appl. Earth Observ. Remote Sens.* 8, 1876–1885. doi: 10.1109/JSTARS.2015.2422716

Wang, R., and Tao, D. (2016). Non-local auto-encoder with collaborative stabilization for image restoration. *IEEE Transact. Image Process.* 25, 2117–2129. doi: 10.1109/TIP.2016.2541318

Wang, X., Singh, D., Marla, S., Morris, G., and Poland, J. (2018). Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. *Plant Methods* 14:53. doi: 10.1186/s13007-018-0324-5

Wang, Z., Underwood, J., and Walsh, K. B. (2018). Machine vision assessment of mango orchard flowering. *Comput. Electron. Agric.* 151, 501–511. doi: 10.1016/j.compag.2018.06.040

Wendel, A., Underwood, J., and Walsh, K. (2018). Maturity estimation of mangoes using hyperspectral imaging from a ground based mobile platform. *Comput. Electron. Agric.* 155, 298–313. doi: 10.1016/j.compag.2018.10.021

Westling, F., Underwood, J., and Orn, S. (2018). Light interception modelling using unstructured LiDAR data in avocado orchards. *Comput. Electron. Agric.* 153, 177–187. doi: 10.1016/j.compag.2018.08.020

Wolfert, S., Ge, L., Verdouw, C., and Bogaardt, M.-J. (2017). Big data in smart farming - a review. *Agric. Syst.* 153, 69–80. doi: 10.1016/j.agsy.2017.01.023

Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., et al. (2018). Vegetation indices combining the red and red-edge spectral information for leaf area index retrieval. *IEEE J Selected Topics Appl. Earth Observ. Remote Sens.* 11, 1482–1493. doi: 10.1109/JSTARS.2018.2813281

Xue, J., and Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* 17:1353691. doi: 10.1155/2017/1353691

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned Aerial Vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front. Plant Sci.* 8:1111. doi: 10.3389/fpls.2017.01111

Yang, Q., Shi, L., Han, J., Yu, J., and Huang, K. (2020). A near real-time deep learning approach for detecting rice phenology based on UAV imagers. *Agric. Forest Meteorol.* 287:107938. doi: 10.1016/j.agrformet.2020.107938

Yang, Q., Shi, L., Han, J., Zha, Y., and Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Res.* 235, 142–153. doi: 10.1016/j.fcr.2019.02.022

Zhang, B., Tang, L., and Roemer, M. (2014). Probabilistic weather forecasting analysis for unmanned aerial vehicle path planning. *J. Guid. Control Dynam.* 37, 309–312. doi: 10.2514/1.61651

Zhang, N., Wang, M., and Wang, N. (2002). Precision agriculture - a worldwide overview. *Comput. Electron. Agric.* 36, 113–132. doi: 10.1016/S0168-1699(02)00096-0

Zhang, Q., Yang, L. T., Chen, Z., and Li, P. (2018). A survey on deep learning for big data. *Inform. Fusion* 42, 146–157. doi: 10.1016/j.inffus.2017.10.006

Zheng, H., Cheng, T., Zhou, M., Li, D., Yao, X., Tian, Y., et al. (2019). Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery. *Precis. Agric.* 20, 611–629. doi: 10.1007/s11119-018-9600-7

Ziliana, M., Parkes, S., Hoteit, I., and McCabe, M. (2018). Intra-season crop height variability at commercial farm scales using a fixed-wing UAV. *Remote Sens.* 10:2007. doi: 10.3390/rs10122007

Zou, K., Zhang, R., and Jiang, Y. (2018). "Yield estimation using unmanned aerial vehicle low-altitude imaging for dense planting cotton field," in *ASABE 2018 Annual International Meeting.* (Detroit, MI Cobo Center).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Johansen, Morton, Malbeteau, Aragon, Al-Mashharawi, Ziliani, Angel, Fiene, Negrão, Mousa, Tester and McCabe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 June 2020
doi: 10.3389/fpls.2020.00898

Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network

Jun Liu and Xuewei Wang *

Facility Horticulture Laboratory of Universities in Shandong, Weifang University of Science and Technology, Weifang, China

Tomato is affected by various diseases and pests during its growth process. If the control is not timely, it will lead to yield reduction or even crop failure. How to control the diseases and pests effectively and help the vegetable farmers to improve the yield of tomato is very important, and the most important thing is to accurately identify the diseases and insect pests. Compared with the traditional pattern recognition method, the diseases and pests recognition method based on deep learning can directly input the original image. Instead of the tedious steps such as image preprocessing, feature extraction and feature classification in the traditional method, the end-to-end structure is adopted to simplify the recognition process and solve the problem that the feature extractor designed manually is difficult to obtain the feature expression closest to the natural attribute of the object. Based on the application of deep learning object detection, not only can save time and effort, but also can achieve real-time judgment, greatly reduce the huge loss caused by diseases and pests, which has important research value and significance. Based on the latest research results of detection theory based on deep learning object detection and the characteristics of tomato diseases and pests images, this study will build the dataset of tomato diseases and pests under the real natural environment, optimize the feature layer of Yolo V3 model by using image pyramid to achieve multi-scale feature detection, improve the detection accuracy and speed of Yolo V3 model, and detect the location and category of diseases and pests of tomato accurately and quickly. Through the above research, the key technology of tomato pest image recognition in natural environment is broken through, which provides reference for intelligent recognition and engineering application of plant diseases and pests detection.

Keywords: deep learning, K-means, multiscale training, small object, object detection

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Stefano Speranza, University of Tuscia, Italy
Marina Regina Frizzas, University of Brasilia, Brazil

***Correspondence:**
Xuewei Wang
wangxuewei@wfust.edu.cn

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *19 December 2019*
Accepted: *02 June 2020*
Published: *16 June 2020*

Citation:
Liu J and Wang X (2020) Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. Front. Plant Sci. 11:898. doi: 10.3389/fpls.2020.00898

INTRODUCTION

The continuous development of economy and society has brought about global climate and environmental problems. The occurrence of diseases and insect pests seriously affects people's life. The incidence and occurrence of plant diseases and insect pests is higher and higher and more complex (Food and Agriculture Organization of the United Nations,). Therefore, it is very

important to study the prevention of plant diseases and insect pests, as well as the diagnosis and remedial measures of plant diseases and insect pests.

The origin of tomato is South America. Tomato is one of the important economic crops, which not only contains rich vitamins, but also can be used as fruit. In recent years, with the popularity of Western food, tomato sauce is more and more popular. The demand for tomato is increasing, and it has gradually become an important food in people's daily life. Therefore, tomato plays an extremely important role in agricultural vegetable production and vegetable trade. As one of the most widely cultivated vegetables in the world, tomato has not only high yield, wide adaptability, but also high nutritional value. But, like other crops, tomato is affected by various diseases and pests in its growth process. Diseases include tomato virus disease, tomato nematode disease, tomato deficiency disease, tomato physiological disease, tomato bacterial disease, and tomato fungal disease. Pests mainly include leaf miner, greenhouse whitefly, Alfalfa noctuid moth, tobacco green worm, cotton bollworm, and *Polyphagotarsonemus latus*, etc. Usually, the occurrence of tomato diseases and pests is greatly affected by the local environment, varieties, cultivation, and management factors, so the types of tomato diseases and pests in different regions are different. China is one of the largest country in world vegetable production and export. The development of Shandong Shouguang tomato industry has a strong comparative advantage in the world. The cultivation area of facilities and open-field tomatoes has gradually increased. If improper field management can increase the probability of infection by diseases and pests, leading to outbreaks of diseases and pests, which has a significant impact on tomato yield and quality. Field investigation revealed that common tomato diseases in Shouguang, Shandong Province mainly include early blight, late blight, yellow leaf curl virus, brown spot, coal pollution, gray mold, leaf mold, navel rot, leaf curl disease, mosaic; common tomato pests mainly include leaf miner and greenhouse whitefly. According to the statistical data analysis of field investigation, it can be seen that the incidence of tomato diseases and pests in different farm households varies greatly in the field, with the yield loss of 1%–5% in the area with mild onset and more than 30% in the area with severe onset. At present, although many studies have reported the infection and detection of tomato pests and diseases in different regions, there is a lack of systematic reports on the detection of the main pests and diseases of tomatoes commonly found in Shouguang, Shandong Province. Therefore, it is very necessary to take 10 common diseases and 2 common insect pests in Shouguang area of Shandong Province as the research objects, and collect and collate the relevant data to provide theoretical basis for targeted early warning and prevention and control in the field.

The occurrence of tomato diseases and pests in different regions seriously affects tomato production. If the control is not timely, it will lead to yield reduction or even crop failure. Disease and pest prevention is the best way to reduce yield loss and reduce pesticide application to produce pollution-free vegetables. When plants grow to the point where symptoms of pests and diseases already occur, even if people can make accurate diagnosis and appropriate treatment, it is also a passive remedy. Although this is also very necessary, at this time, the application of agricultural chemicals and pesticides has poor control effect and is easy to cause environmental pollution, which leads to excessive pesticide residues in vegetables, and at the same time leads to more and more resistance of pests and diseases, making the work of crop disease resistance more and more difficult, which is an undesirable result. Therefore, early prediction and prevention of diseases and pests are very important. Research on tomatoes (Diaz-Pendon et al., 2010; Gilbertson and Batuman, 2013) shows how susceptible a plant is to be influenced by diseases and pests. With regard to how to effectively control diseases and insect pests and help vegetable farmers to improve the yield of tomato, the most important thing is to make accurate identification of diseases and insect pests. Therefore, the identification of tomato diseases and insect pests is the most serious challenge for scientific and technical personnel.

The traditional method of artificial detection of diseases and insect pests completely depends on the observation experience of the grower, or ask experts for guidance. Such a method is not only slow, but also is of low efficiency, high cost, strong subjectivity, low accuracy, and timeliness. With the continuous development of the Internet, the application of information technology provides new methods and ideas for crop diseases and insect pests' identification. Using efficient image recognition technology can improve the efficiency of image recognition, reduce the cost, and improve the recognition accuracy. Therefore, experts and scholars at home and abroad have done a lot of research, in which deep learning has become the research focus. The application of deep learning in crop diseases and insect pests' identification can greatly reduce the workload and shorten the identification time. Complex network structure and huge data samples are the biggest characteristics of deep learning. The emergence of deep learning technology provides strong technical support for image recognition.

Among them, the convolutional neural network (CNN) is a typical model of deep learning. The diseases and pests detection method based on CNN can automatically extract the features in the original image, which overcomes the subjectivity and limitation of artificial feature extraction in the traditional methods. The end-to-end structure simplifies the recognition process and solves the problems that manually designed feature extractor cannot get the feature expression closest to the natural attribute of the object. Based on the application of CNN object detection, not only can save time and effort, but also can achieve real-time judgment, greatly reduce the huge loss caused by diseases and pests, which has important research value and significance. Based on the latest research results of detection theory CNN object detection and the characteristics of tomato diseases and pests images, this study will build the dataset of tomato diseases and pests under the real natural environment, optimize the feature layer of Yolo V3 model by using image

pyramid to achieve multi-scale feature detection, improve the detection accuracy and speed of Yolo V3 model, and detect the location and category of diseases and pests of tomato accurately and quickly. Through the above research, the key technology of tomato pest image recognition in natural environment is broken through, which provides reference for intelligent recognition and engineering application of plant diseases and pests detection.

In this paper, image data of tomato diseases and insect pests under real natural environment were collected, and corresponding data processing was carried out to build tomato diseases and pests detection dataset. Based on the YOLO v3 model as the main body, the image pyramid structure is adopted to fuse features of different levels to obtain feature maps of different scales for location and category prediction. Then, the dimension of the object box is clustered, and the number of anchor box is increased, so that the model can obtain more edge information of the object. Finally, in the training process, multi-size images are used for training, so that the model can adapt to images of different resolutions. Experiments show that the improved YOLO v3 algorithm can improve the detection speed while ensuring the detection accuracy.

RELATED WORK

Comparison Between Traditional Machine Learning Technology and Deep Learning Technology

Before the development of deep learning technology, image recognition of plant diseases and insect pests was realized by traditional technology. With the development of deep learning technology, researchers begin to apply deep learning to the image recognition of diseases and pests, and have made a lot of achievements in recent years (Mohanty et al., 2016; Sladojevic et al., 2016; Wang et al., 2017; Liu et al., 2018; Ferentinos, 2018; Brahimi et al., 2018; Kaur et al., 2019), especially in the disease and insect image recognition of apple, tomato, cucumber, and other common crops. The efficiency and effect of image recognition are much better than that of traditional recognition methods. The comparison between traditional recognition method and deep learning recognition method is shown in **Table 1**.

TABLE 1 | Contrast between traditional machine learning and deep learning.

Technology	Advantages	Disadvantages
Machine learning	It has advantages in training small data samples. No need of expensive hardware. The algorithm structure is simple and parameter adjustment is relatively simple.	Needs complex feature engineering and data dimensionality reduction. The classification accuracy is low and difficult for image recognition in complex background.
Deep learning	The deep feature extraction function makes it perform well in image, audio, and text data. Easy to update data through back propagation. Different architectures are suitable for different problems. The hidden layer reduces the dependence of algorithm on feature engineering.	The requirement of machine configuration is high. Needs massive data and relies on large-scale datasets.

Object Detection Method of Plant Diseases and Insect Pests Based on CNN

Traditional machine vision methods have poor robustness in complex scenes, so it is difficult to meet the work requirements in complex scenes. The performance of CNN in image recognition has made great progress in the past few years, and in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), a lot of deep learning architectures have emerged, such as AlexNet (Krizhenvshky et al., 2012), GoogleLeNet (Szegedy et al., 2015), VGGNet (Simonyan and Zisserman, 2014), ResNet (Xie et al., 2017), and the accuracy of general image recognition is constantly being refreshed. CNN's hard work and great breakthrough in the large-scale image classification competition prompted people to consider applying it to the problem of object detection. Different from image classification, object detection needs to detect and locate specific multiple objects from the image, which is mainly divided into two categories. One is to generate a series of candidate frames as samples by the algorithm, and then classify the samples by CNN, such as RCNN (Girshick et al., 2014), faster RCNN (Ren et al., 2016), and R-FCN (Dai et al., 2016). The other one directly transforms the problem of object bounding box location into regression problem, which does not need to generate candidate boxes. The landmark algorithms include SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016).

Arsenovic et al. (2019) established a plant disease dataset with 79,265 pieces collected under different meteorological conditions. A two-level structure, PlantDiseasenet, which is trained by PDNet-1 and PDNet-2 at the same time, is proposed. PDNet-1 uses the detection method proposed in the Yolo algorithm as the detection tool of plant leaves, and PDNet-2 is responsible for the classification of plant leaves. After training, the accuracy of the model is 93.67%. Jiang et al. (2019) proposed an improved CNN-based deep learning method for real-time detection of apple leaf diseases and insect pests. Firstly, through data expansion and image annotation technology, an apple leaf disease dataset (ALDD) composed of laboratory images and complex images under real field conditions is constructed. On this basis, a new method is proposed by introducing GoogleNet inception structure and rainbow concatenation. Finally, the proposed INAR-SSD (SSD with perception module and rainbow condition) model was trained to detect the five common apple leaf diseases and insect pests. The experimental results show that the model achieved 78.80% mAP and the detection speed is as high as 23.13 FPS. Tian et al. (2019) proposed a method of Apple Anthracnose damage detection based on deep learning and optimized the feature layer of Yolo V3 model with low resolution by using Densenet, which has greatly improved the utilization of neural network features and improved the detection results of Yolo V3 model. The experimental results show that the model achieved 95.57% mAP. Zheng et al. (2019) established a CropDeep species

classification and detection dataset, including 31,147 images and more than 49,000 annotation examples from 31 different categories. These images are collected in real natural environment with different cameras, and the most advanced deep learning classification and detection model is used to provide a wide range of baseline experiments. The results show that the existing classification method based on deep learning can achieve more than 99% in classification accuracy, and only 92% in object detection accuracy. Meanwhile, Yolo V3 model has good application potential in agricultural detection tasks.

Object Detection of Tomato Diseases and Insect Pests Based on CNN

Fuentes et al. (2017) proposed a real-time system for the identification and location of tomato plant diseases and insect pests. The system is consisted of three network structures, Fast R-CNN, SSD, and R-FCN are compared. Each network structure is combined with different feature extraction networks (such as VGG, RESNET). The system has the following advantages: (1) the images used in the system are collected in the field; (2) considering that a crop may be affected by multiple diseases and insect pests at the same time; (3) the system has strong robustness and can be used in the actual field environment to achieve detection; (4) the system can also solve complex tasks, such as infection status (early, late), infection location (leaves, culm, fruit) and both sides of leaves (obverse and reverse). Fuentes et al. (2018) proposed an improved framework of tomato plant diseases and insect pests detection algorithm aiming at the problem of false alarm and unbalanced classification, which is composed of three main units: (1) the main diagnosis unit (bounding box generator), which generates the bounding box containing the location of infected area and class; (2) the auxiliary diagnosis unit (CNN filter bank), which trains each independent CNN to filter the samples of misclassification; (3) the integration unit, which combines the information of the autonomous diagnosis unit and the auxiliary diagnosis unit, and keeps the true positive samples at the same time, so as to eliminate the false positive of the classification errors in the first unit. The experiment shows that the recognition rate of this method is about 96%. Fuentes et al. (2019) proposed a method which can not only effectively detect and locate plant anomalies, but also generate diagnostic results, display the location of anomalies and describe the symptoms of sentences as output, and has achieved 92.5% average precision (mAP) in the newly created tomato plant anomaly description dataset. However, it uses Faster R-CNN to detect the object, which needs to be carried out in two steps. First, extract the region proposal, and then detect it. However, Yolo directly generates coordinates and probability of each category through expression once. Therefore, the real-time performance of existing research needs to be improved.

Problems of Existing Research and Development Trends

The development of the existing research provides a reference and feasibility basis for the application of CNN in the detection of plant diseases and pests, and can avoid the shortcomings of the traditional machine vision method in the feature extraction process. Through CNN object detection technology, it can greatly facilitate plant breeding and help farmers to supervise their fields better. However, the automatic identification and detection of plant diseases and insect pests are mainly faced with the following difficulties (Barbedo, 2016a; Barbedo, 2016b; Barbedo, 2017; Barbedo, 2018a; Barbedo, 2018b).

i. The background of the image is complex. In addition to the infected leaves, the image may contain other elements, such as crop stalks, soil, etc. Especially for the photos taken in the field environment, not only the background is messy, there are also factors such as light, angle difference, etc.
ii. There may be no clear boundary between the infected area and the healthy area.
iii. The same disease has different characteristics in different stages of development, even in different locations.
iv. The characteristics of different kinds of diseases and insect pests may be the same or slightly different, and there may be multiple diseases and insect pests at the same location at the same time.
v. It is difficult to distinguish the infected area from other dead plant tissues.
vi. Throughout the different stages of development of the insect (nymph or larva and adult) and the different instars, the morphological characteristics of insect vary greatly. Especially for small target pests, it is difficult to identify.

At present, the research of plant diseases and insect pests based on deep learning involves a wide range of crops, including all kinds of vegetables, fruits, and food crops. The completed tasks not only include the basic tasks of classification and detection of diseases and insect pests, but also more complex tasks such as the judgment of infection degree. Because deep learning relies on large-scale datasets, there are few open datasets of plant diseases and insect pests at present, the researchers usually find the best solution by comparing different training set and test set proportion and using different network models. However, there is still a certain gap between the complexity of these susceptible images and the real field scene, and there is a certain gap between the real-time detection of diseases and pests in fields based on mobile devices.

THE PRINCIPLE OF YOLO V3 MODEL

Yolo V3 Design Idea

Yolo algorithm proposed by Redmon et al. (2016) in 2016, the object detection task is transformed into regression problem, which greatly speeds up the detection speed. Yolo V3 (Redmon and Farhadi, 2018) was proposed on the basis of Yolo V2 (Redmon and Farhadi, 2016), the detection speed of Yolo V2 is maintained, and the detection accuracy is greatly improved. Yolo V3 uses the idea of residual neural network (He K et al., 2016). The introduction of multiple residual network modules and the use of multi-scale prediction improve the shortcomings of Yolo V2 network in small object recognition. Because of the high accuracy and timeliness of detection, this algorithm is one of the

best algorithms in the field of object detection. This model uses a lot of 3 × 3 and 1 × 1 convolution layers with good performance, and some residual network structures are also used in the subsequent multi-scale prediction. Finally, it has 53 convolution layers, so it can also be called Darknet-53.

Yolo V3 network introduces the idea of using anchor boxes in Faster R-CNN. For coco dataset and VOC dataset, three scales are used for prediction. Each scale has three anchor boxes, and the feature map with large scale uses small priori box, so you can select the appropriate priori box according to the target you want to identify, and the network structure can be changed according to the scale that needs to predict.

Feature Extraction Network Darknet-53

Yolo V3 adopts the network structure of Darknet-53, which consists of 53 (2 + 1*2 + 1 + 2*2 + 1 + 8*2 + 1 + 8*2 + 1 + 4*2 + 1 = 53) convolution layers. The network structure contains 53 convolutional layers and 5 maximum pooling layers. Batch normalization and dropout operations are added after each convolutional layer to prevent overfitting. Darknet-53 is composed of five residual blocks, which uses the idea of residual neural network for reference. **Figure 1** shows the Darknet-53 network structure. Yolo V3 has increased the network depth by introducing residual unit to avoid gradient disappearance.

THE METHOD OF IMPROVING YOLO V3 MODEL

Compared with the whole tomato plant image, the disease spots and pests are small objects. Therefore, Yolo V3 model is a little inadequate in the scale when recognizing tomato disease spots and pests. In view of this situation, Yolo V3 model is improved to adapt to specific tomato diseases and pests detection tasks. The improved Yolo V3 network structure is shown in **Figure 2**.

The improved Yolo V3 network structure uses feature fusion to increase the number of feature maps, which saves a lot of computation and is very important for improving the running speed. Two residual units were added to the detection network to further enhance the expression of image features. Also, the image pyramid structure and improved conv-residual unit are used to reduce the impact of depth reduction on the precision of the backbone network. Therefore, the improved Yolo V3 network structure inherits Darknet-53's structure in ideology but reduces the dependence of the whole network on ResNet, and optimize the network forward propagation speed and model size, making it more suitable for mobile platforms with weak computing power.

Based on the image pyramid structure, Yolo V3 model is improved. Upper sampling is used to fuse the high-level features with the low-level features, and three sets of prediction feature maps of different scales are obtained. Then, the location and category are predicted on these three sets of prediction feature maps.

K-means algorithm is used to calculate the prior box dimension in the self-built tomato diseases and pests dataset. The priori box dimension obtained in Yolo V3 model is trained on the COCO dataset, and its parameters are divided into three different scales. For the specific task of tomato diseases and pests detection, cluster operation should be carried out for the specific dataset, and smaller priori box should be used for the larger scale feature map to get the corresponding cluster center.

The multi-scale training strategy was used to train the self-built dataset of tomato diseases and insect pests. In Yolo V3 model, the full connection layer is removed and uses the full convolution operation. In the training process, the input size can be changed at any time, so that the trained model can adapt to different scales of tomato diseases and pests images.

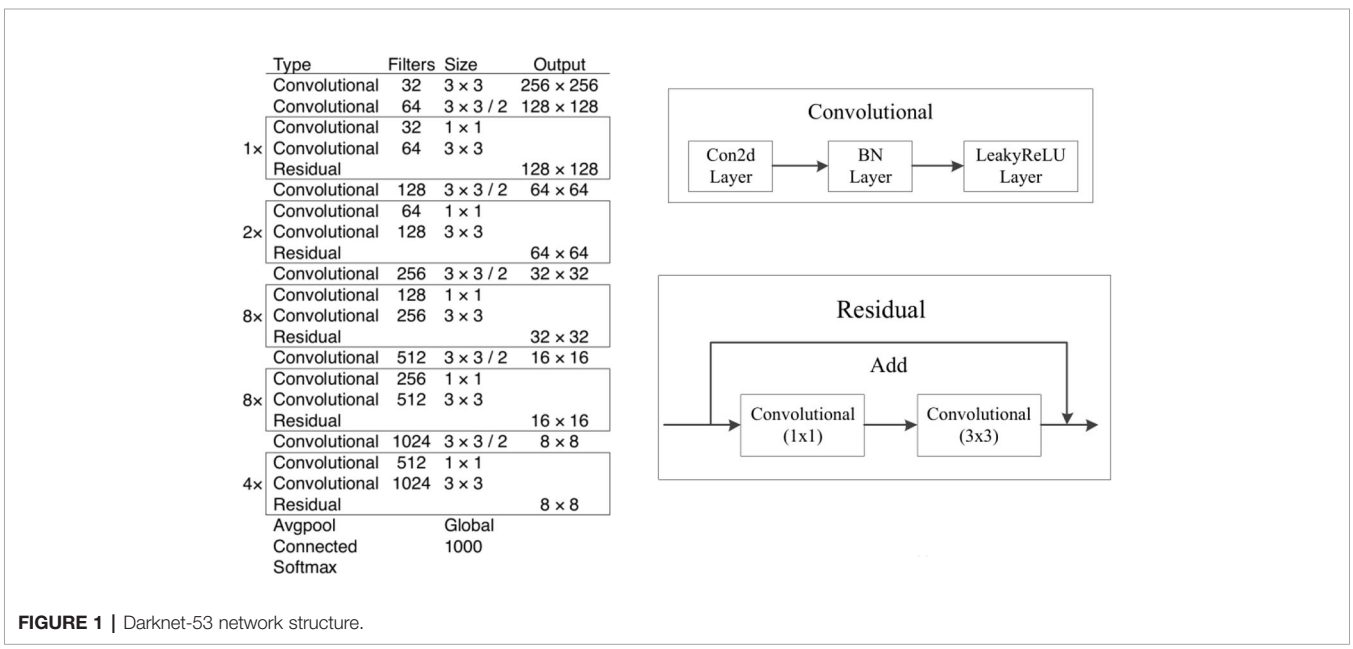

FIGURE 1 | Darknet-53 network structure.

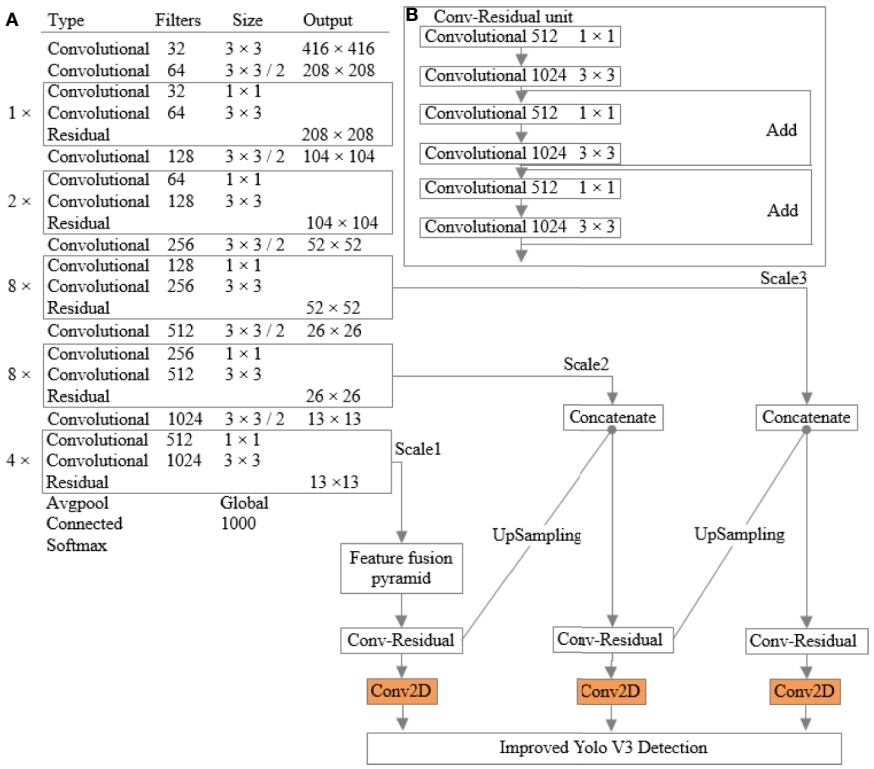

FIGURE 2 | The improved Yolo V3 network structure **(A)** network structure; **(B)** conv-residual unit.

Multiscale Feature Detection Based on Image Pyramid

In Yolo V3, Darknet-53 network is used to extract features, and feature visualization technology is used to clearly show the effect of each level of features. Low-level features have rich details and location information, while high-level features have rich semantic features. From low-level to high-level, the details are decreasing, while the semantic information is increasing. For location prediction, more low-level feature information is needed, for category prediction, more high-level local information is needed. Therefore, based on the image pyramid model, the up sampling method is used to fuse the high-level features with the low-level features to obtain the feature maps of different scales for location and category prediction.

The feature pyramid on the right side of **Figure 3** is generated by the feature pyramid on the left side. The whole process is as follows: first, the input image is deeply convoluted, then the features on layer2 are convoluted, and the features on layer4 are sampled to make them have the same size. Then, the processed layers layer2 and layer4 are convoluted and operated to input the obtained results to layer5. In the same way, feature fusion is carried out among multiple layers to get multiple sets of feature maps for prediction. Based on this scheme, the processed low-level features and high-level features are accumulated. The purpose of doing this is that because the low-level features can provide more accurate location information, and multiple down

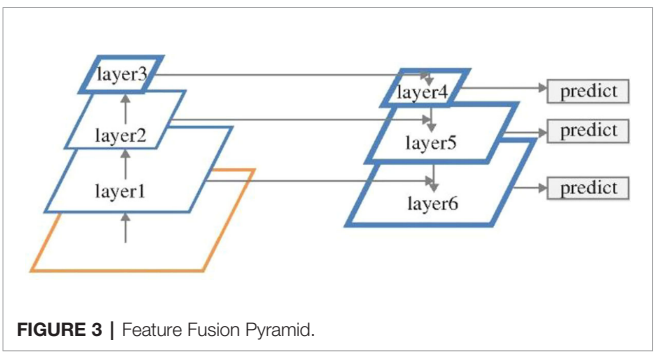

FIGURE 3 | Feature Fusion Pyramid.

sampling and up sampling operations make the location information of the deep-seated network have errors. Therefore, they are combined to build the deep feature pyramid, and integrate the multi-layer feature information on different feature maps for category and location prediction.

Based on the above idea of feature fusion, Yolo V3 algorithm is improved. Up sampling is used to fuse high-level features with low-level features. Finally, three sets of feature maps are obtained, and these three sets of feature maps of different scales are used for prediction. The improved network structure is shown in **Figure 4**.

Specific network structure improvement details: first, get the characteristic pyramid through the Darknet-53, and carry out continuous 1 × 1 and 3 × 3 convolution operations on the conv53

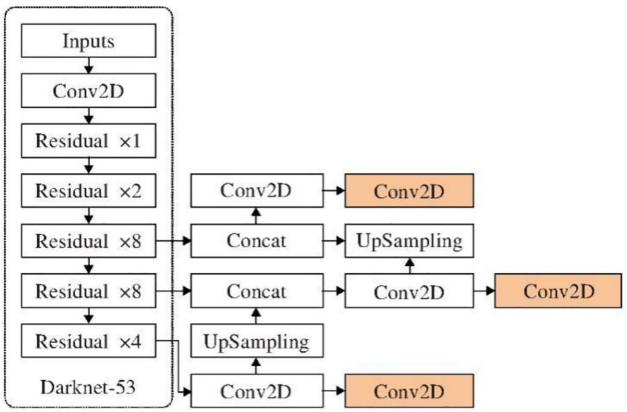

FIGURE 4 | Improved network structure.

layer to get a group of Yolo layers to be processed, and then carry out a group of 1 × 1 and 3 × 3 convolution operations on the layer to get a small scale Yolo layer. At the same time, perform upper sampling operation on this layer, and carry out convolution operation with the conv45 layer in the Darknet-53, as well using continuous 1 × 1 and 3 × 3 convolution operations, the second group of Yolo layers to be processed are obtained, and a group of 1 × 1 and 3 × 3 convolution operations are performed on this layer to obtain the mesoscale Yolo layer. At the same time, perform upper sampling operation on this layer, convoluted and operated with conv29 layer in Darknet-53, and the third group of Yolo layers to be processed are also obtained by using continuous 1 × 1 and 3 × 3 convolution operations, and a group of 1 × 1 and 3 × 1 layers are performed on this layer 3 convolution operation to get large scale Yolo layer. After the above operations, three sets of Yolo feature layers with different scales are obtained, and these three sets of feature layers are used for location and category prediction.

K-Means Dimension Clustering Algorithm

Yolo V3 uses the prior box to predict the coordinates of the bounding box in Yolo V2. The difference is that Yolo V3 uses k-means algorithm to predict and gets nine prior boxes, and divides them into three scale feature maps, among which the larger scale feature map uses the smaller prior box to obtain more edge information of the object.

The dimensions of nine sets of prior boxes calculated in Yolo V3 are respectively as follows: (10, 13), (16, 30), (33, 23), (30, 61), (62, 45), (59, 119), (116, 90), (156, 198), and (373, 326). However, in the actual detection task of tomato diseases and insect pests, the prior box dimension calculated by Yolo V3 algorithm is not suitable for the detection scene of tomato diseases and insect pests, so it is difficult to get accurate object bounding box information by using the original prior box dimension in Yolo V3 algorithm.

Therefore, in the tomato diseases and pests detection scenario, K-means algorithm is used to cluster the self-built tomato diseases and pests dataset, and nine sets of prior box dimension centers are obtained respectively as follows: (44, 11), (53, 18), (60, 22), (73, 25), (78, 27), (81, 31), (89, 41), (97, 43), and (109,51). The region is arranged from small to large, and divided into feature maps of three different scales, among which the larger scale feature map uses the smaller prior frame. Finally, the cluster center is used to detect tomato diseases and insect pests.

Multiscale Training

In the Yolo detection algorithm, convolutional network is used to extract the features, and then the full connection layer is used to get the prediction value. However, due to the existence of full connection layer, the input image size of the network must be fixed in the training process, so the final training network does not have robustness to different sizes of test images.

According to the self-built tomato diseases and pests dataset, the size of the input image is different. Therefore, in order to enhance the robustness of the model to different image sizes, multi-scale training strategy is adopted. Specifically, remove the full connection layer in Yolo V3 network and change it to full convolution operation. **Figure 5** shows the process of converting a full connection layer to a convolution layer.

Figure 5A uses the full connection layer for prediction, and **Figure 5B** converts the full connection layer into a convolution layer for prediction. When the input image size is 416 × 416, after passing through the Darknet-53 network, 13 × 13 × 1024 feature maps are output. Passing through a full connection layer containing 4,096 neurons, a set of 4,096 × 1 feature maps are obtained in **Figure 5A**. **Figure 5B** uses 4,096 convolution kernels of 13 × 13, and finally obtains 1 × 1 × 4,096 feature maps, which is essentially equivalent to 4,096 nerves obtained by full connection. For the above two network structures, when the image input size is 416 × 416, the network can operate normally, but there are other sizes of pictures in the self-built tomato diseases and pests dataset. For example, when inputting 608 × 608 size images, after passing through the Darknet-53 network, the 19 × 19 × 1024 feature map is output. For the structure of **Figure 5A**, the next step is to connect the feature map with 4,096 neurons. The size of the original architecture is 13 × 13, and now it is 19 × 19. Therefore, during the propagation of the network, the previous parameter matrix cannot be used, and the network cannot operate normally. For the structure of **Figure 5B**, after the full connection is changed to convolution operation, the network can continue to run, and finally get the correct output of 7 × 7 × 4,096. Therefore, after the full connection layer is changed to full convolution operation, input images of different sizes are used for training, and the improved algorithm can adapt to different sizes of test images.

In addition, after changing the full connection layer to full convolution operation, the performance of the network will be improved. When using the full connection layer in **Figure 5A** in the forward propagation operation of the network, it can be calculated that 708,837,377 parameters are needed. While using the full convolution operation in **Figure 5B**, only 696,320 parameters are needed. Therefore, after converting the full connection layer into convolution operation, the number of parameters can be reduced, the network operation amount can be reduced, and the network performance can be improved.

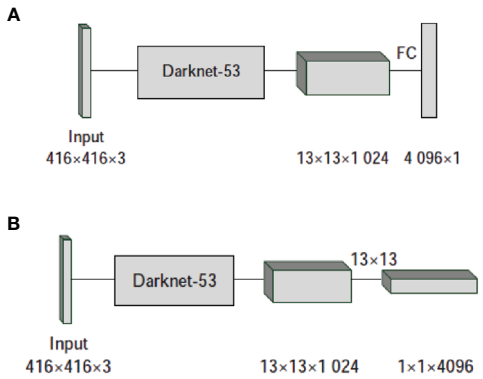

FIGURE 5 | Conversion of full connection layer to convolution layer. **(A)** Full connection operation. **(B)** Convulation operation.

Based on the above analysis, this paper will use multi-scale training strategy to train the self-built tomato diseases and pests dataset. Because the whole network has five maximum pool layers, the down sampling rate of the network is 32. In the training process, the input size of the training picture of the dataset of tomato diseases and pests is divided into a series of values of multiple of 32, and the calculation formula of the size is as follows:

$$S_{n+1} = 32 + S_n, n \leq 9 (1)$$

Among them, S_n is the size of the group n input image. During network initialization, S_1 is 320 × 320. Through formula (1), it can be concluded that the size of the input image is: 320, 352, 384, 416, 448, 480, 512, 544, 576, and 608. In the process of training, one kind of input image size is selected randomly every 10 rounds to achieve the effect that the model can adapt to images of different sizes.

EXPERIMENTAL RESULTS AND COMPARATIVE ANALYSIS

Experimental Operation Environment

The experimental hardware environment of this paper is shown in **Table 2**. On this basis, the software environment is built as follows: Ubuntu 16.04, python, OPPENCV, CUDA, etc. The framework uses the Caffe framework and Darknet-53 framework.

Experimental Dataset Building

In the process of building tomato diseases and pests detection dataset, different equipment including monitoring, digital camera, and smart phone are used to collect tomato plant photos in the local tomato planting greenhouses. In different time periods, different weather conditions and different scenes, single frame images and photos in the monitoring video are randomly selected as the dataset. The number of samples of each category is shown in **Table 3**. The labeling tool is used to mark the images. With this tool, when the manual operation is carried out, we only need to mark the user-defined objects in the image, and the tool can automatically generate the corresponding configuration file.

TABLE 2 | Configuration of experimental hardware environment.

Hardware name	Model	Number
Main board	Asus WS X299 SAGE	1
CPU	INTEL I7-9800X	1
Memory	The Kingston 16G DDR4	2
Graphics card	GEFORCE GTX1080Ti	2
Solid state drives	Kingston 256G	1
Hard disk	Western digital 1T	1

Model Training

Use the weight parameters provided on Yolo V3 official website as the initialization parameters of network training, randomly use the images in the self-built tomato diseases and pests detection training dataset to finetune the network parameters, so that the detection effect of the whole model is optimal, and the model parameter settings are shown in **Table 4**.

Comparative Analysis of Experiments

In this paper, Yolo V3 algorithm is mainly used for the experiment. The improvements include multi-scale feature detection based on image pyramid, object frame dimension clustering, and multi-scale training. At the same time, in order to verify the validity, accuracy and stability of the model, SSD, Faster R-CNN, and Yolo V3 algorithms are used for experimental comparative analysis, and the detection accuracy mAP (Mean Average Precision) and detection time are used as the evaluation indexes of detection effect. The results are shown in **Table 5**.

It can be seen from the experimental results that in terms of object detection accuracy, Faster R-CNN and the improved Yolo V3 are superior to the other two algorithms. However, due to the need to establish RPN network in the process of object detection, which involves a lot of calculation, Faster R-CNN is inferior to the

TABLE 3 | Number of samples of each disease type.

Serial Number	Disease/ Pest	Sample size	Number of labeled samples (bounding box)	Percent of bounding box samples
1	Early blight	1,209	12,187	8.30%
2	Late blight	1,303	12,362	8.41%
3	Yellow leaf curl virus	1,286	12,138	8.26%
4	Brown spot	1,348	11,726	7.98%
5	Coal pollution	1,287	13,025	8.87%
6	Gray mold	1,263	12,184	8.29%
7	Leaf mold	1,377	12,399	8.44%
8	Navel rot	1,106	12,026	8.19%
9	Leaf curl disease	1,198	11,734	7.99%
10	Mosaic	1,242	13,092	8.91%
11	Leaf miner	1,228	11,580	7.88%
12	Greenhouse whitefly	1,153	12,459	8.48%
	Total	15,000	146,912	100.00%

TABLE 4 | Model parameter settings.

Name	Value
Batch Size	64
Learning Rate	0.01
Epoch	13,000
Momentum	0.9
Match Threshold	0.5
NMS	0.3

TABLE 5 | Comparison of experimental results.

Algorithm name	Accuracy (%)	Time/(ms)
SSD	84.32	25.69
Faster R-CNN	90.67	2868.94
Yolo V3	88.31	21.18
Improved Yolo V3	92.39	20.39

improved Yolo V3 in detection speed. In general, the improved Yolo V3 algorithm is superior to the other three algorithms in detection accuracy and speed, especially the detection time is the shortest. Therefore, the improved Yolo V3 performs best for real-time detection tasks. The improved Yolo V3 algorithm can complete the detection task of tomato diseases and insect pests well on the premise of considering both the detection accuracy and detection speed.

The loss curves of the four algorithms are shown in **Figure 6**. It can be seen that the loss of the algorithm in this paper is the smallest, while Faster R-CNN is the largest. Therefore, the convergence speed of this algorithm is the fastest.

Performance Analysis in Small Object Scenario

Performance analysis in small object scenario is important for different sizes of diseases and pests objects throughout their different stages. The multi-scale detection method based on image pyramid proposed in this paper improves the detection effect of small objects. In order to verify the detection effect of the small object detection performance, the test dataset is sorted according to the size of the object. The 0-10%, 10% - 30%, 30% - 70%, 70% - 90% of the object size are divided into five sub categories: XS, S, M, L, XL, which represent the size of different objects. **Figure 7** shows the detection performance of the original Yolo V3 algorithm and the improved algorithm for objects of different sizes.

As can be seen from **Figure 7**, the detection accuracy of the improved Yolo V3 algorithm for objects of different sizes is higher than that of the other three algorithms. Therefore, when detecting tomato diseases and insect pests, using the strategy of multiscale feature fusion, combining the high-level features with the low-level features, the improved algorithm has achieved the best results in the detection of small objects. Throughout the different stages of development of the insect (nymph or larva and adult) and the different instars, the morphological characteristics of insect vary greatly. For small object pests, the improved algorithm can achieve good results.

Performance Analysis of Different Resolution Images

The multi-scale training method proposed in this paper can enhance the robustness of the model for the detection of different resolution images. In order to verify the detection effect of the model on the input images of different resolutions, this paper divides the test dataset images into three different resolution

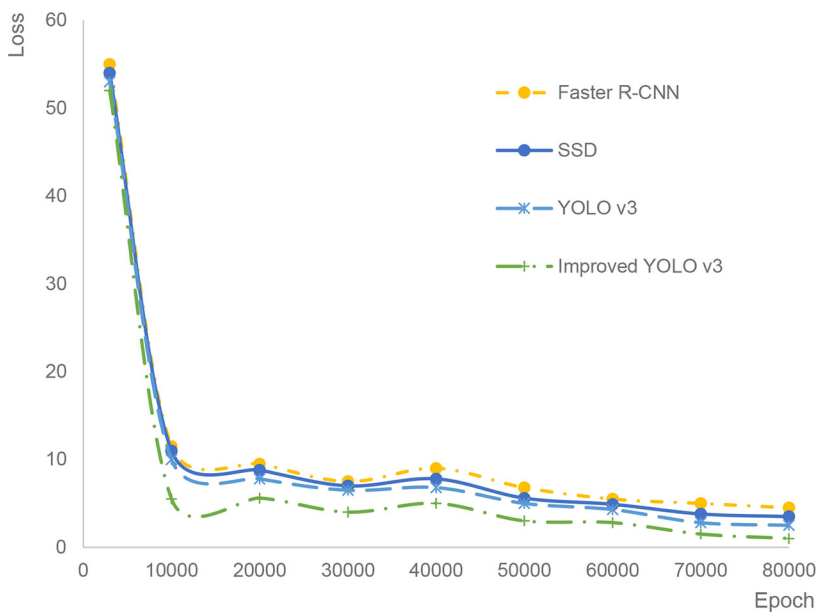

FIGURE 6 | Loss function contrast.

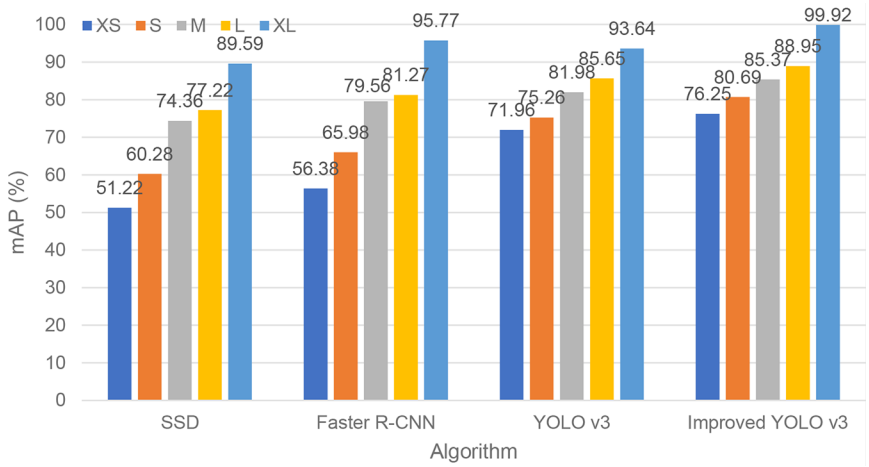

FIGURE 7 | Object size sensitivity analysis of four algorithms.

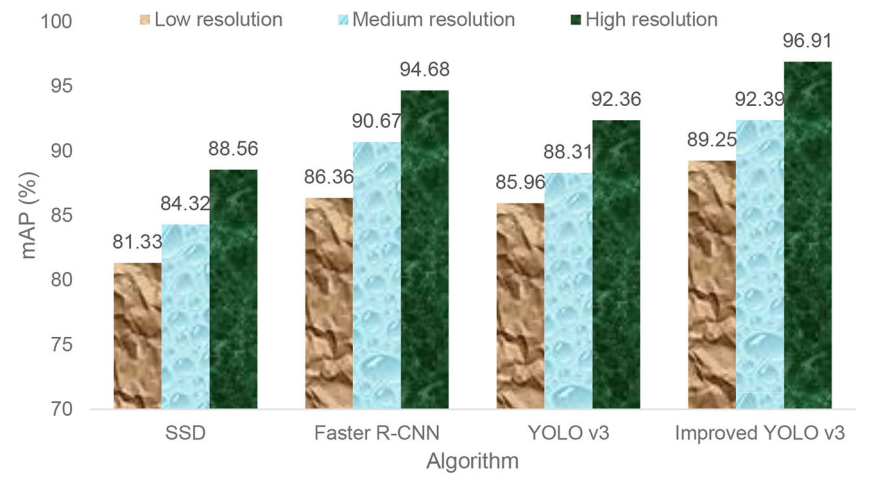

FIGURE 8 | Object image resolution sensitivity analysis of four algorithms.

sizes, namely {320, 608, 1024}, representing three types of images: low resolution, medium resolution and high resolution. **Figure 8** represents the detection accuracy performance of four algorithms for different resolution images.

As can be seen from **Figure 8**, the detection accuracy of the improved Yolo V3 algorithm for different resolution images is higher than that of the other three algorithms. It can be seen that the multi-scale training detection strategy adopted in this paper not only can enhance the adaptability of the model to different resolution images, but also can detect the location and category of tomato diseases and insect pests accurately and quickly.

In addition, in order to show the detection effect of the algorithm in this paper more intuitively, some detection images are selected, as shown in **Figure 9**. It can be seen that the algorithm in this paper can correctly detect the location and category of the object when there are multiple objects and small objects in the image, and can effectively avoid the problem of false detection and missing detection.

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

1. In this paper, an improved Yolo V3 algorithm is proposed to detect tomato diseases and insect pests. Yolo V3 network was improved by using multi-scale feature detection based on image pyramid, object bounding box dimension clustering and multi-scale training. The

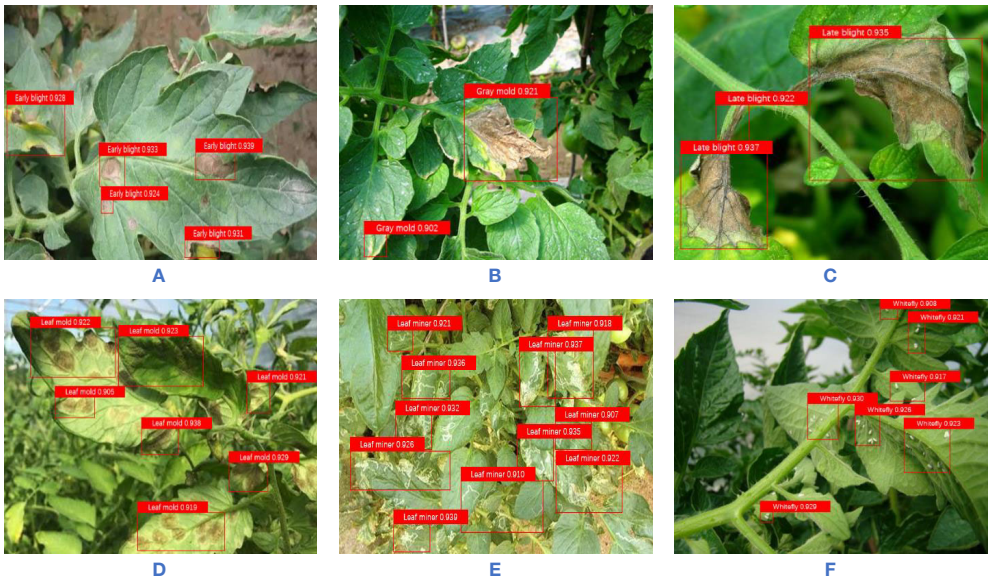

FIGURE 9 | The detection effect diagram of the improved YOLO v3 algorithm **(A)** Early bight; **(B)** Gray mold; **(C)** Late blight; **(D)** Leaf mold; **(E)** Leaf miner; **(F)** Whitefly.

experimental results show that the detection accuracy of the algorithm is 92.39% and the detection time is only 20.39 ms. Therefore, for the task of tomato diseases and pests detection, the improved Yolo V3 algorithm proposed in this paper can not only maintain a high detection rate, but also meet the real-time detection requirements, and can detect the location and category of tomato diseases and insect pests accurately and quickly.

2. Compared with SSD, Faster R-CNN and the original Yolo V3, the improved Yolo V3 CNN can achieve higher detection accuracy and shorter detection time, and meet the requirements of real-time detection accuracy and speed of tomato diseases and pests.
3. Performance analysis in small object scenario and different resolution images further verifies that the improved Yolo V3 network has strong robustness for detection of different object sizes and different resolution images in complex environment, and has high detection and positioning accuracy, which can meet the needs of tomato diseases and pests detection in complex environment.

Future Directions

1. In different growth cycles, the appearances of diseases and pests are different. Therefore, the images of diseases and pests should be divided more carefully, and the same class of diseases and pests should be divided according to the growth period as the standard. In the future, the division of dataset will be improved.
2. In order to make the model more widely applied, the next work will collect a large number of high-quality images of different types of diseases and pests, and proceed to insert other insect pests, optimize and adjust the model, and extend this to other crops, so as to improve the practicability and accuracy of crop diseases and pests image recognition.
3. Construct an intelligent patrol robot for tomato diseases. In a real greenhouse tomato planting base, the intelligent patrol inspection robot for Tomato Diseases can work around the clock. The utility of the improved Yolo V3 model will be further enhanced by carrying a mobile robot arm with a detection sensor for disease detection. The purpose of early detection and early diagnosis of tomato diseases will be achieved by capturing the lesions in real time and non-destructive detection of tomato diseases.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this article are not publicly available. Requests to access the datasets should be directed to liu_jun860116@wfust.edu.cn.

AUTHOR CONTRIBUTIONS

JL designed research. JL and XW conducted experiments and data analysis and wrote the manuscript. XW revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study is supported by the Facility Horticulture Laboratory of Universities in Shandong with project numbers 2019YY003, 2018RC002, 2018YY016, 2018YY043, and 2018YY044; Research

and Development Plan of Applied Technology in Shouguang with project number 2018JH12; 2018 innovation fund of science and technology development center of the China ministry of education with project number 2018A02013; 2019 basic capacity construction project of private colleges and universities in Shandong province; Weifang Science and Technology Development Program with project numbers 2019GX081 and 2019GX082. Key research and development plan of Shandong Province with project number 2019RKA07012 and 2019GNC106034.

REFERENCES

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., and Stefanović, D. (2019). Solving Current Limitations of Deep Learning Based Approaches for Plant Disease Detection. *Symmetry* 11, 21. doi: 10.3390/sym11070939

Barbedo, J. G. A. (2016a). A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. *Trop. Plant Pathol.* 41 (4), 210–224. doi: 10.1007/s40858-016-0090-8

Barbedo, J. G. A. (2016b). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60. doi: 10.1016/j.biosystemseng.2016.01.017

Barbedo, J. G. A. (2017). A new automatic method for disease symptom segmentation in digital photographs of plant leaves. *Eur. J. Plant Pathol.* 147 (2), 349–364. doi: 10.1007/s10658-016-1007-6

Barbedo, J. G. A. (2018a). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. doi: 10.1016/j.biosystemseng.2018.05.013

Barbedo, J. G. A. (2018b). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53. doi: 10.1016/j.compag.2018.08.013

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., and Moussaoui, A. (2018). "Deep learning for plant diseases: detection and saliency map visualisation," in *Human and Machine Learning*. Eds. J. Zhou and F. Chen (Cham, Switzerland: Springer International Publishing), 93–117.

Dai, J., Li, Y., He, K., and Sun, J. (2016). R-FCN: Object Detection via Region-based Fully Convolutional Networks. *arXiv*. arXiv:1605.06409v2.

Diaz-Pendon, J. A., Canizares, M. C., Moriones, E., Bejarano, E. R., Czosnek, H., and Navas-Castillo, J. (2010). Tomato yellow leaf curl viruses: Menage a trois between the virus complex, the plant and whitefly vector. *Mol. Plant Pathol.* 11, 414–450. doi: 10.1111/j.1364-3703.2010.00618.x

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009

Food and Agriculture Organization of the United Nations. Plant Diseases and pests. Available online: http://www.fao.org/emergencies/emergency-types/plant-pests-and-diseases/en/ (accessed on 5 February 2020).

Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S. A. (2017). Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. *Sensors* 17, 2022. doi: 10.3390/s17092022

Fuentes, A. F., Yoon, S., Lee, J., and Park, D. S. (2018). High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9, 1162. doi: 10.3389/fpls.2018.01162

Fuentes, A. F., Yoon, S., and Park, D. S. (2019). Deep Learning-Based Phenotyping System With Glocal Description of Plant Anomalies and Symptoms. *Front. Plant Sci.* 10, 1321. doi: 10.3389/fpls.2019.01321

Gilbertson, R. L., and Batuman, O. (2013). Emerging Viral and Other Diseases of Processing Tomatoes: Biology Diagnosis and Management. *Acta Hortic.* 1, 35–48. doi: 10.17660/ActaHortic.2013.971.2

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv*, 580–587.

He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J. (2016). *Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. New York: IEEE, June 27–30, 2016. pp. 770–778.

Jiang, P., Chen, Y., Liu, B., He, D., and Liang, C. (2019). Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE*, 7, 59069–59080. doi: 10.1109/ACCESS.2019.2914929

Kaur, S., Pandey, S., and Goel, S. (2019). Plants disease identification and classification through leaf images: A survey. *Arch. Comput. Methods Eng.* 26, 507–530. doi: 10.1007/s11831-018-9255-6

Krizhenvshky, A., Sutskever, I., and Hinton, G. (2012).Imagenet classification with deep convolutional networks, in: *Proceedings of the Conference Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 3–8 December. pp. 1097–1105.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016).SSD: Single Shot MultiBox Detector, in: *Proceedings of the European Conference on Computer Vision—ECCV*, Amsterdam, The Netherlands, 8–16 October. pp. 21–37.

Liu, B., Zhang, Y., He, D., and Li, Y. (2018). Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* 10, 11.

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 1419. doi: 10.3389/fpls.2016.01419

Redmon, J., and Farhadi, A. (2016). YOLO9000: Better, Faster, Stronger. *IEEE Conference on Computer Vision and PatternRecognition*. 2017, 6517–6525. doi: 10.1109/CVPR.2017.690

Redmon, J., and Farhadi, A. (2018). Yolo V3: An incremental improvement. *arXiv*. arXiv:1804.02767 [cs], 1–6.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016).You only look once: Unified, Real-Time Object Detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 26 June–1 July. pp. 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv*. arXiv:1409.1556.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* 2016, 1–11. doi: 10.1155/2016/3289801

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015).Going deeper with convolutions, in: *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June. pp. 1–9.

Tian, Y., Yang, G., Wang, Z., Li, E., and Liang, Z. (2019). Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense. *J. Sens.* 2019, 1–13. doi: 10.1155/2019/7630926

Wang, G., Sun, Y., and Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Comput. Intell. Neurosci.* 2017, 1–8. doi: 10.1155/2017/2917536

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. *arXiv*. arXiv:1611.05431.

Zheng, Y., Kong, J., Jin, X., Wang, X., and Zuo, M. (2019). CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture. *Sensors* 19, 1058. doi: 10.3390/s19051058

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 July 2020
doi: 10.3389/fpls.2020.01086

A Cloud-Based Environment for Generating Yield Estimation Maps From Apple Orchards Using UAV Imagery and a Deep Learning Technique

Orly Enrique Apolo-Apolo[1], *Manuel Pérez-Ruiz*[1]*, *Jorge Martínez-Guanter*[1] *and João Valente*[2]

[1] *Área de Ingeniería Agroforestal, Dpto. de Ingeniería Aeroespacial y Mecánica de Fluidos, Universidad de Sevilla, Sevilla, Spain,* [2] *Information Technology Group, Wageningen University & Research, Wageningen, Netherlands*

OPEN ACCESS

Edited by:
Spyros Fountas, Agricultural University of Athens, Greece

Reviewed by:
Gambella Filippo, University of Sassari, Italy
Jordi Llorens Calveras, Universitat de Lleida, Spain

***Correspondence:**
Manuel Pérez-Ruiz
manuelperez@us.es

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *03 March 2020*
Accepted: *01 July 2020*
Published: *15 July 2020*

Citation:
Apolo-Apolo OE, Pérez-Ruiz M, Martínez-Guanter J and Valente J (2020) A Cloud-Based Environment for Generating Yield Estimation Maps From Apple Orchards Using UAV Imagery and a Deep Learning Technique. Front. Plant Sci. 11:1086. doi: 10.3389/fpls.2020.01086

Farmers require accurate yield estimates, since they are key to predicting the volume of stock needed at supermarkets and to organizing harvesting operations. In many cases, the yield is visually estimated by the crop producer, but this approach is not accurate or time efficient. This study presents a rapid sensing and yield estimation scheme using off-the-shelf aerial imagery and deep learning. A Region-Convolutional Neural Network was trained to detect and count the number of apple fruit on individual trees located on the orthomosaic built from images taken by the unmanned aerial vehicle (UAV). The results obtained with the proposed approach were compared with apple counts made *in situ* by an agrotechnician, and an R^2 value of 0.86 was acquired (MAE: 10.35 and RMSE: 13.56). As only parts of the tree fruits were visible in the top-view images, linear regression was used to estimate the number of total apples on each tree. An R^2 value of 0.80 (MAE: 128.56 and RMSE: 130.56) was obtained. With the number of fruits detected and tree coordinates two shapefile using Python script in Google Colab were generated. With the previous information two yield maps were displayed: one with information per tree and another with information per tree row. We are confident that these results will help to maximize the crop producers' outputs *via* optimized orchard management.

Keywords: deep learning, apple, yield map, Google Colab, photogrammetry, fruit

INTRODUCTION

The successful management of modern, high-density apple orchards depends on the ability to improve processes such as planting, cultivation, harvesting, and the optimization of fruit commercialization (González-Araya et al., 2015). The efficient management of these tasks, where harvesting and fruit processing are considered high-cost, high value-added operations, is key for producers (Silwal et al., 2016). Consequently, an accurate yield estimation is crucial for the stakeholders (apple growers and sellers), since this information can significantly contribute to their decisions-making process (Gongal et al., 2015; Tian et al., 2019b).

The traditional management of agricultural crops has been inherently subjective and based on past experience, manually counting, and historical data collected by farmers (Rahnemoonfar and Sheppard, 2017). These methods can be inaccurate, subjected to bias, and inefficient, since they do not reflect the yield distribution across the orchard, especially in orchards with a high spatial variability (Aggelopooulou et al., 2013; Bargoti and Underwood, 2017). Currently, with the breakthrough of new agricultural technologies, many farm tasks are becoming automated, and researchers and companies have carried out studies based on artificial intelligence algorithms which automatically learns decision rules from data (Abiodun et al., 2018). A particular success has been the use of deep learning (DL) and, in particular, the development and application of a branch of these techniques known as Convolutional Neural Networks (CNN). These complex algorithms use images tagged by technicians or crop experts as inputs. These are laid out through various convolutional filters that activate image features to generate a trained model. As reviewed by other authors, the use of this models makes it possible to simplify and automate some of the analytical tasks in the agricultural domain (Kamilaris et al., 2017; Jha et al., 2019). Therefore, for example, a model for detecting and mapping every piece of fruit in a commercial mango orchard was proposed by Stein et al. (2016). The fruits were detected using a model based on Faster R-CNN. Koirala et al. (2019) tested several deep learning architectures to detect mango fruits on RGB images taken from a terrestrial vehicle during the night. Additionally, a method where synthetic images were used to train the model and tested on actual images was suggested by Rahnemoonfar and Sheppard (2017). Moreover, Fu et al. (2018) presented a system to detect kiwifruit in field images under different lighting conditions.

In the specific case of apple orchards, works employing different approaches have been explored by many researchers. Tian et al. (2019b) developed an improved model for apple detection during different growth stages. An object detection architecture named Yolo-V3 was used, and images with different light conditions at ground level were obtained. The pre-harvest yield mapping of apple orchards using segmentation techniques was suggested by Roy et al. (2019). Their contribution was the use of two clustering methods: semi-supervised (to separate the apple pixels from others in the input images) and unsupervised (to automatically identify the apples). Fruit size was estimated by Gongal et al. (2018) using the 3D coordinates of pixels from images taken by a 3D-camera as a tool for harvesting robots. A fine-tuned model for apple flower detection was deployed by Dias et al. (2018). The high accuracy of these approaches opened the door for the possible integration of these models into complex automated decision-making systems in the future. Nevertheless, existing methods can be improved, since many of the images used were taken by terrestrial vehicles and at ground level. This means that labor remains an inefficient aspect, since specific platforms are required for the taking of images, which constitutes a time-consuming task and can accentuate soil compaction problems.

Unmanned aerial vehicles (UAVs) are currently modernizing the farming industry by helping farmers to monitor their crops in a more timely manner (Mogili and Deepak, 2018). These aerial platforms usually mount high-resolution cameras that are capable of acquiring quality images (thermal, spectral, multispectral, or RGB-visible images), which can be used for various kinds of analysis (Maes and Steppe, 2019). Moreover, these vehicles can integrate an RTK-GNSS system for precise real-time positioning allowing the generation of crop maps with a centimeter-level accuracy at the field level (Chlingaryan et al., 2018). A general method used for creating crop maps is based on the structure from motion (SfM) algorithm (Turner et al., 2012). This algorithm selects important features known as keypoints from individual images to build a georeferenced orthomosaic (Anders et al., 2019). However, despite its suitability, producing these kinds of maps requires costly commercial software, a powerful computer, and multiple supervised steps to generate the new composite images (Torres-Sánchez et al., 2018). According to the literature reviewed, the most common types of photogrammetry software under private licenses used for this purpose are: Pix4D® (www.pix4d.com), AgisSoft PhotoScan® (www.agisoft.com), and Photomodeler® (www.photomodeler.com). However, in the recent years, the emergence of platforms, such as Docker (www.docker.com) or Django (www.djangoproject.com), have opened up the possibility of implementing the SfM algorithm in the cloud and developing open-source tools that are affordable for everyone at both professional and educational levels.

On the other hand, many of the remote sensing applications in agriculture are based on using Geographical Information Systems (GIS) to bring value to the farmers (Machwitz et al., 2019; Maes and Steppe, 2019). These tools allow us to prepare and manage agricultural georeferenced data and build geospatial snapshots of cropland from remote sensors mounted on both aerial and terrestrial platforms (Sharma et al., 2018). The information generated enables the automation of field operations, the reduction of costs, and maximization, acting as a steward of the land (Kaloxylos et al., 2012). Until a few years ago, the most popular types of software for GIS applications were Quantum GIS (www.qgis.org) and Esri's ArcGIS (www.arcgis.com). The first is open-source but the other needs a commercial license (Duarte et al., 2017). The use of this software requires the user to have a basic knowledge of how to work and interpret the data contained in raster and shape files (the most common files used in GIS), although it is not always an easy task, especially for farmers (Abdelrahman et al., 2016). In recent years, a collection of open-source GIS libraries that work with Python language have been developed and made available to the general public (Gillies, 2007; Jordahl, 2014; Rapiński and Zinkiewicz, 2019). Examples of this type of library are GeoPandas (www.geopandas.org), GeoServer (www.geoserver.org), and Qhull (www.qhull.org), among others. At the same time, the development of platforms such as Google Colaboratory (www.colab.research.google.com), a cloud service based on Jupyter Notebooks, which allows the integration of deep learning models and GIS tools in a simple python script, has occurred (Carneiro et al., 2018; Bisong, 2019). This provides the opportunity to develop geospatial analysis tools that can be readily integrated into web platforms, allowing their adoption by farmers.

Based on the above, it can be asserted that the high cost of data collection and the difficulty of interpretation currently prevents farmers from implementing data-driven agriculture (Thompson et al., 2019). With specific regard to yield mapping in apple orchards, based on the detection of the number of fruits, although the proposed methods have shown promising results and a high accuracy, they do not provide a final product with a high potential to be exploited by the farmers. Additionally, most of them use ground-level platforms that may increase the data collection time and hinder their application in large agricultural areas.

Therefore, the objectives of this project were the following: (1) exploring the feasibility of yield estimation by detecting apple fruits on images taken by a UAV; (2) training and testing a model based on CNNs to automatically detect apple fruits, with the aim of making the weights and models used for apple detection available for the general public; and (3) building an apple tree variable yield map for each tree and one with information per each tree row.

MATERIALS AND METHODS

Location and Imagery Acquisition

This study was undertaken during the 2018 and 2019 seasons in an orchard fields of apple (*Malus x Dornestica* Borkh. cv 'Elstar') in Randwijk (latitude: 51°56'18.5"N; longitude: 5°42'24.8"E) near Wageningen (The Netherlands). The crop field had 0.47 ha with 592 trees allocated in 14 rows with approximately 41 trees in each row and a pollinator tree every 10 m. The average tree height was 3 m with a tree spacing was 3 × 1 m (inter-row and intra-row), rows were NW-SE oriented, and the crop management tasks (fertilization, thinning, pruning, etc.) were performed following the conventional farm practices.

The UAV platform employed to take the pictures was a DJI Phantom 4 Pro (DJI Technology Co., Ltd., Shenzhen, China) at a set flying altitude of 10 m (**Figure 1A**). The onboard camera had a 1/2.3" CMOS sensor (with an effective pixel count of 20M), a lens FOV of 84°, a focal length of 8.8 mm, a focal ratio of f/4.5, and a focus to infinity. This UAV was equipped with dual-band satellite positioning (GPS and GLONASS), which provided a sub-meter precision location.

A grid-shaped flight plan was designed using the DJI Ground Station Pro (DJI Technology Co., Ltd., Shenzhen, China) iPad application, which allowed us to control or plan automatic flights for DJI aircrafts (Chen et al., 2019). The flights, in the two seasons (2018 and 2019), to take the pictures were made 2 days before the first harvest (40%). It was a sunny day with low wind speed. A total of 806 pictures at 15 m above the ground were taken in a nadiral view (vertically downward at 90°) (**Figure 1B**). The image resolution was set to 5,472 × 3,648 pixels (JPG format). A total of 354 images taken in 2019 were used to build the dataset for training the CNN, while the rest (taken in 2018) were used for creating the visible orthomosaic. These latter were obtained with a forward overlap of 85% and a sideway overlap of 75%. The UAV flight made in 2018 had to be made over a portion of the trees because the rest of the field had already been harvested by the farmer.

Five ground control points (GCPs) were established during each flight as an indirect georeferencing of UAV images and for an accuracy assessment of the orthomosaic obtained (**Figure 1C**). The precise locations of the GCP (black and white targets) were obtained using a Topcon RTK GNSS equipment with an accuracy below 2.5 cm. A total of 452 pictures were used for orthomosaic creation.

Ground Truth Acquisition for Yield Estimation

According to Moltó et al. (1992) and Jiménez et al. (2000), only approximately 60–70% of crop production is visible from the

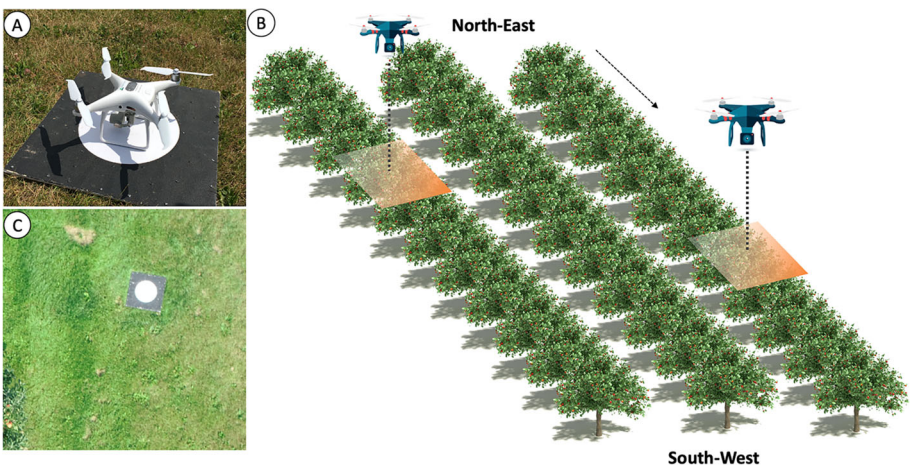

FIGURE 1 | Field test flight design and equipment used **(A)** unmanned aerial vehicle (UAV; DJI Phantom 4) used during the flights, over one of the ground control points to associate projection coordinates with locations on the images; **(B)** workflow for UAV image acquisition; and **(C)** an image of an apple tree in the field.

outside of a tree; here lies the complexity of yield estimation, as not all existing fruits can be detected with only external images of the tree. Moreover, previous studies have been based on ground-level observations on both sides of the tree canopy. However, zenithal pictures shown only a fraction of the total fruits, making it a challenge to generate complex models for yield estimation in this type of study (Chen et al., 2019). On this basis, a previous step in this research was to check the percentage of fruit visible from the aerial pictures. At the same time as the pictures to build the orthomosaic were taken by the UAV before harvesting, a representative sample of 19 trees was randomly selected from row 5 of the crop field. We assumed the number of fruits by row remain consistent based on historical data provide by the farmer. The tree architecture was divided according to **Figure 2**. Then, visual counting of the fruit was conducted on each side (right and left) of the tree, and the data were collected in a Microsoft Excel (Version 16.37) file. For avoiding duplicated counting of the fruits, a plastic tape was used to delimit the areas of interest.

Later, the apple fruits in all trees were hand-harvested and weighed to give an average weight in kilograms per meter (kg/m) of fruit per row. The collection of fruit was conducted in three stages, since the market demand for fresh fruit is variable over the time during a harvest season (Lötze and Bergh, 2004). Moreover, farmers tend to choose the best moment to be able to find a good price for their product.

Orthomosaic Construction and Data Pre-Processing for Yield Map Estimation

A total of 452 images were used to build the orthomosaic [an aerial image of an area, composed of multiple images stitched together using photogrammetry which has been geometrically corrected (Chen et al., 2019)]. The imagery was automatically processed using Agisoft PhotoScan Professional 1.2.3 software (Agisoft LCC, St. Petersburg, Russia). Following the software recommendations, the first step was to "Align Photos" with the "High" accuracy set up. This option uses the original resolution of images to generate a sparse 3D point cloud with a low resolution as a necessary first step towards building the orthomosaic. After that, GCPs were manually located in each image. This process is necessary, as, despite of the images taken by the UAV being geotagged using the onboard GNSS receiver, the accuracy of this sensor is low. Then, a 3D dense point cloud (110449395 points) with a "High" accuracy was generated in a previous step to build the final raster file (**Figure 3**). Finally, the orthomosaic in the coordinate system WGS 84 (EPSG: 4326) was exported as a GeoTIFF file with 4.18 mm/pixel to be further used in fruit detection and to build the yield map based on the number of fruits detected.

Currently, apple orchards are being planted using advanced machinery that records GNSS coordinates of each tree in a standard vectorial format (shapefile) of the GIS. These files allow storage spatial information and vector operations with other files, such as raster files (Oliver, 2013; Maes and Steppe, 2019). On this basis, a python script was developed to create a circular mask (1-meter diameter) with the coordinates of each tree. The output's script was an individual shapefile for each tree avoiding the edges of the canopy. Then, the orthomosaic was cropped using these shapefiles, and a TIF file was obtained for each tree as an output. Finally, each TIF file was tested using the Faster R-CNN model to count apple fruits. Considering the number of fruits detected and taking into account and the distribution of fruits on the structure of the apple canopy, a yield map estimation was created using Qgid (3.12).

Building and Labeling Image Datasets for Apple Fruit Detection

Dataset size plays a critical role in making DL models successful. A model without sufficient and representative training data is not able to learn the fundamental discriminative patterns required to carry out robust detection of fruits (Sonka et al., 1993). The features of apples on the trees may dramatically diverge (e.g., green fruits, fruits of different sizes, fruits occluded by branches and leaves, and overlapping fruits). Moreover, the images might suffer from distortions, especially those generated by outdoor light conditions and the rolling shutter effect (Chen et al., 2019).

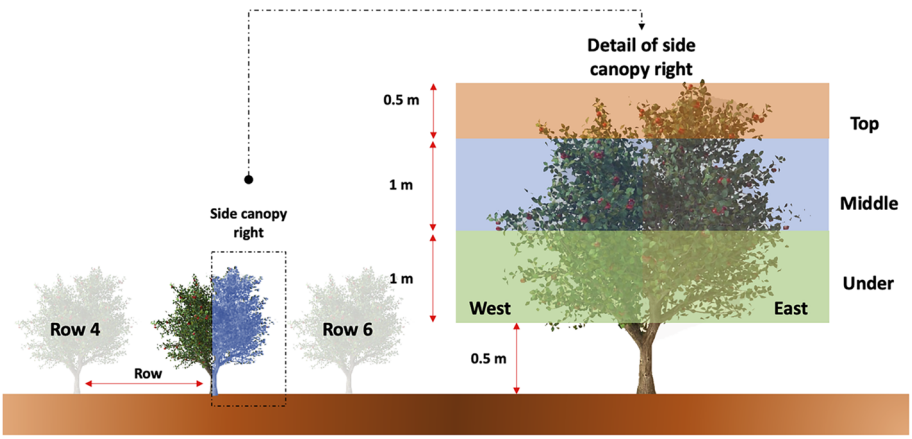

FIGURE 2 | Detailed illustration of apple tree canopy architecture division for visual counting of fruit from both the right and left sides of the row.

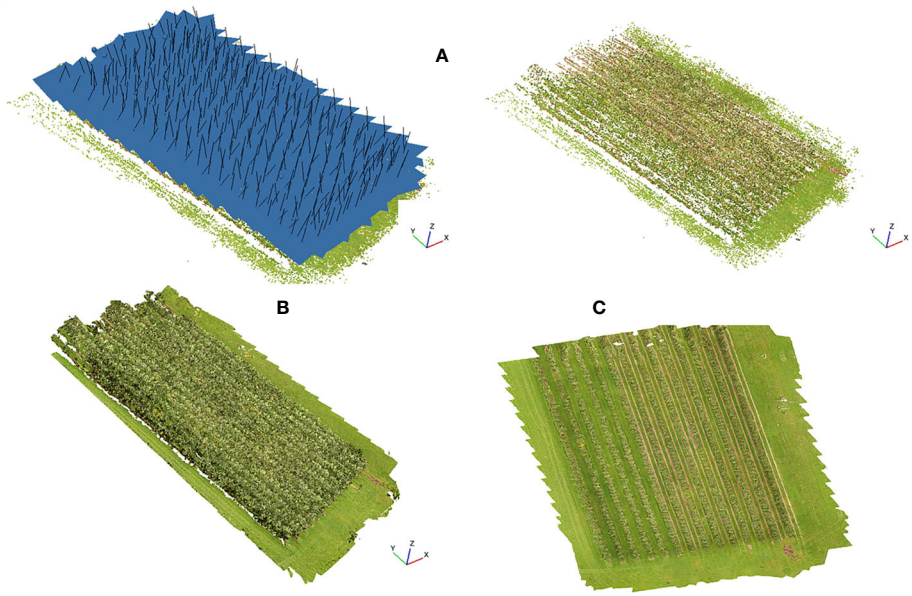

FIGURE 3 | Workflow to build an orthomosaic using Agisoft PhotoScan: **(A)** photo alignment and 3D sparse point cloud (each blue square represents the estimated viewpoint of the input images). **(B)** dense point cloud, and **(C)** orthomosaic.

The set of processes carried out by a CNN requires images with an appropriate resolution, since high-quality images increase the computational resources needed (Lecun et al., 2015; Chollet, 2017). Therefore, the images taken by the UAV in this study were cropped to produce smaller images with a resolution of 416 × 416 px without applying any resizing process. As a result, a preliminary sample of 1,000 images was selected to train the model. Additionally, in order to achieve a high accuracy and avoid overfitting problems, data augmentation techniques were applied (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014). Data augmentation is a common technique used to transform pictures based on rotation, changing color channels, and the addition of filters among others. In this paper, images were rotated by 90, 180, and 270 degrees. The contrast and brightness were changed by varying α and β values responsible for the color difference settings using a Python script developed by the authors (**Figure 4**). Consequently, a dataset containing a total of 3,000 pictures was used to train the CNN.

As suggested by Rahnemoonfar and Sheppard (2017), CNN requires a huge amount of annotated pictures with the coordinates of each fruit on the images from the training dataset. In this project, a

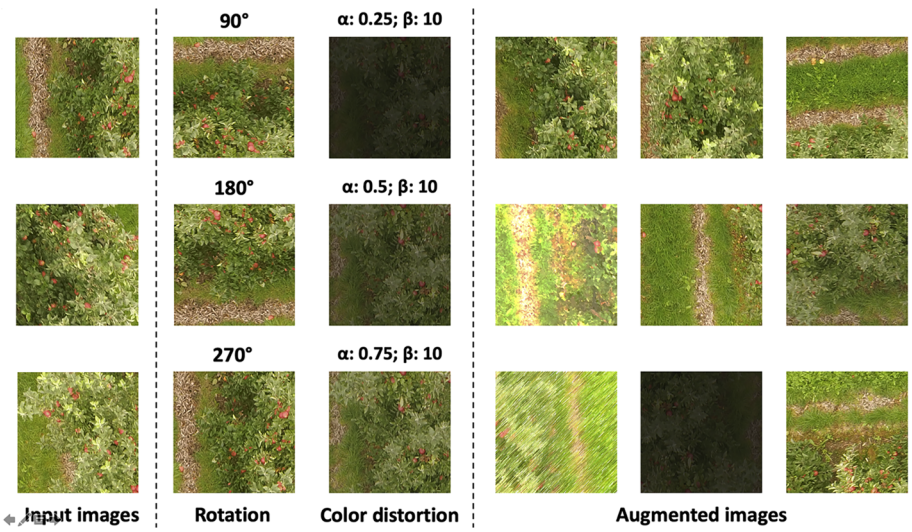

FIGURE 4 | Dataset augmentation process. Original images were rotated 90, 180, and 270 degrees and color distortion ware setting changing both α and β values.

free and open-source labeling tool called LabelImg (v1.8.3) was used (Tzutalin, 2015). The process was done manually and very carefully to prevent mislabeling or occlusion, since, due to the nature of fruit trees, many of them were completely occluded by others or even attached to each other (**Figure 5**). Once all fruits had been labeled with a bounding box, an Extensible Markup Language (XML) file in PASCAL Visual Object Classes (VOC) format was generated.

Once the labeling process was complete, the configuration details for the model and labels were implemented in the TensorFlow API (www.tensorflow.org). Due to CNN's high demand for hardware and GPU resources, Google Colaboratory (also known as Colab) offered by Google was used to implement and train the model. Colab, a cloud service based on Jupyter Notebooks, provides a free single 12GB NVIDIA Tesla K80 GPU that can be continuously used for up to 12 h. The advantage of this particular tool lies in the fact that its access is completely free and open-source. It also allowed us to work in the same work space with geospatial data and DL algorithms. We consider this platform to be a powerful tool that may in the future play a determining role in research and education with aggregated data and expert decision-making systems based on georeferenced data and ML (Machine Learning) algorithms.

For the local computing processes, a MacBook Pro laptop (MacOs High Sierra 10.13.4) with a 2.5 GHz Intel Core i7 processor, 16 GB of RAM, and Graphics AMD Radeon R9 M370X 2048 MB Intel Iris Pro 1536 MB was used. The Open-Source Computer Vision (OpenCV) library (http://opencv.org/), which includes several hundred computer vision algorithms, was used to process images (Rosebrock, 2016). The Keras (Chollet, 2017) open-source library was used in combination with TensorFlow backend tools to build and deploy the DL architecture.

Fine-Tuning and Training of the Faster-RCNN

Convolutional neural networks have been proven to be powerful visual models that use complex data as inputs that are capable of conducting automated fruit counting in the images. These algorithms consider an image as a matrix of pixels whose size (kernel) is (height × width × depth), where the depth is the number of image channels (3 for our RGB crop images). Hidden layers with a hierarchical structure (Lecun et al., 2015) are the main components of a CNN; the first layers can detect lines, corners, and simple shapes, whereas deeper layers can recognize complex shapes (Rosebrock, 2018). A common CNN architecture consists of several convolutional blocks (composed of convolutional layer + pooling layers + non-linearity) and one or more fully connected layers (**Figure 6**). Feature extraction, non-linearity operations, and dimension reduction were performed with this common architecture. Additionally, a fully connected layer was used to classify data from images (Guo et al., 2016), while a softmax function assigned the probability of belonging to the class (apple).

Despite the advances in computational processes and the available power offered by the graphics processing unit (GPU), training a neural network from scratch is still highly computationally expensive and requires large datasets for learning (Patrício and Rieder, 2018). To overcome these obstacles, a method named transfer learning (Gu et al., 2018) was used. The main objective of this procedure is to transfer the knowledge from one model trained on large datasets, such as ImageNet (Gopalakrishnan et al., 2017), to another model to solve a specific task (Talukdar et al., 2018). Several popular pretrained networks using transfer learning, such as VGG-16, ResNet 50,

FIGURE 5 | Labeling process used to annotate pictures: **(A)** original picture taken by UAV; **(B)** picture with 416 × 416 px resolution; and **(C)** picture with each one of the bounding boxes.

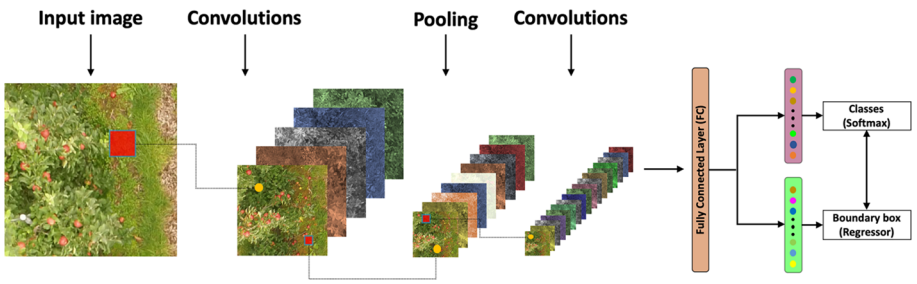

FIGURE 6 | General architecture of a convolutional neural network.

DeepNet, and AlexNet Inception V2, are described in the literature (Rosebrock, 2018).

The Faster R-CNN model was selected, since this network can use several architectures, such as ResNet, Inception, and Atrous, and thus increase the efficiency and precision of fruit detection (Dias et al., 2018). In this study, the Faster R-CNN Inception Resnet V2 Atrous Coco (Ren et al., 2017) model with a TensorFlow object detection application programming interface (API) was used. TensorFlow is an open-source software library for numerical computations (Kamilaris and Prenafeta-Boldú, 2018) and was used because of its flexibility and the ability to deploy network computations in multiple central processing units (CPUs), GPUs, and servers. The model comprises three steps, with an apple tree image as the input. Faster R-CNN extracts feature maps from the image using a CNN and then passes these maps through a region proposal network (RPN), which returns object proposals (Rosebrock, 2018). Finally, these maps are classified, and the bounding boxes enclosing the apple fruits are predicted (**Figure 7**).

The model was trained for 6 h, until the loss function reached the value of 0.06. This function allowed an accurate quantification of the model to ensure correct classification of the apples in our dataset (Kamilaris and Prenafeta-Boldú, 2018). The batch size (the parameter that defines the number of samples, which are images in this case, that will be propagated through the CNN) was two images in each step. The learning rate (a hyperparameter which determines the learning speed of the new information over the old) was 0.001.

Statistical Analyses

To evaluate the accuracy of the trained model, 20 randomly selected pictures cropped from the orthomosaic randomly selected were used. The total number of fruits per picture (Nfp) was manually counted using the Photoshop count tool (Adobe Systems Inc., San Jose, United States), as suggested by Payne et al. (2014). Consequently, with this data, the precision (P, Eqn. 1), recall (R, Eqn. 2), F1score (Eqn. 3), and Accuracy (A, Eqn. 4) were used as the evaluation metrics for fruit detection (Rosebrock, 2018). These model evaluation metrics are defined as follows:

$$Precision\ (P) = \frac{TP}{TP + FP} \quad \text{Eqn. 1}$$

$$Recall\ (R) = \frac{TP}{TP + FN} \quad \text{Eqn. 2}$$

$$F1score = 2x\frac{PxR}{P + R} \quad \text{Eqn. 3}$$

$$Accuracy\ (A) = \frac{TP}{Nfp} \quad \text{Eqn. 4}$$

where TP corresponds to true positives, i.e., when the algorithm correctly detects a fruit with a bounding box; FP indicates false positives, i.e., when a box is computed in a location where a fruit is not located; and FN denotes false negatives, i.e., when a target fruit is not detected.

Linear regressions were used for comparisons of the number of fruits counted visually (in the field and on the pictures) and the number of fruits harvested. The analysis was performed with RStudio® (http://www.rstudio.com). A comparison of visually counted fruits and harvested fruits was performed using the Mean Absolute Error (MAE, Eqn. 5) and the Root Mean Square Error (RMSE, Eqn. 6):

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \quad \text{Eqn. 5}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}} \quad \text{Eqn. 6}$$

where n refers to the number of compared values, A_t is the actual observed value, and F_f is the forecast value.

RESULTS

Distribution of Fruits in an Apple Orchard Canopy

The distribution of the fruits inside an apple canopy tree can be strongly variable. It depends on several factors, such as the tree height, the effect of row orientation on daily light absorption, and the apple cultivar planted in the field, among others (Gongal et al., 2018). **Table 1** shows that the largest amount of fruits was found between the middle and underside of the tree. This could be explained by the canopy architecture, since, on the top of the tree, generally, there is a smaller number of branches (Willaume et al., 2004). Furthermore, farmers tend to prune apple trees to concentrate the majority of the fruits in the middle and

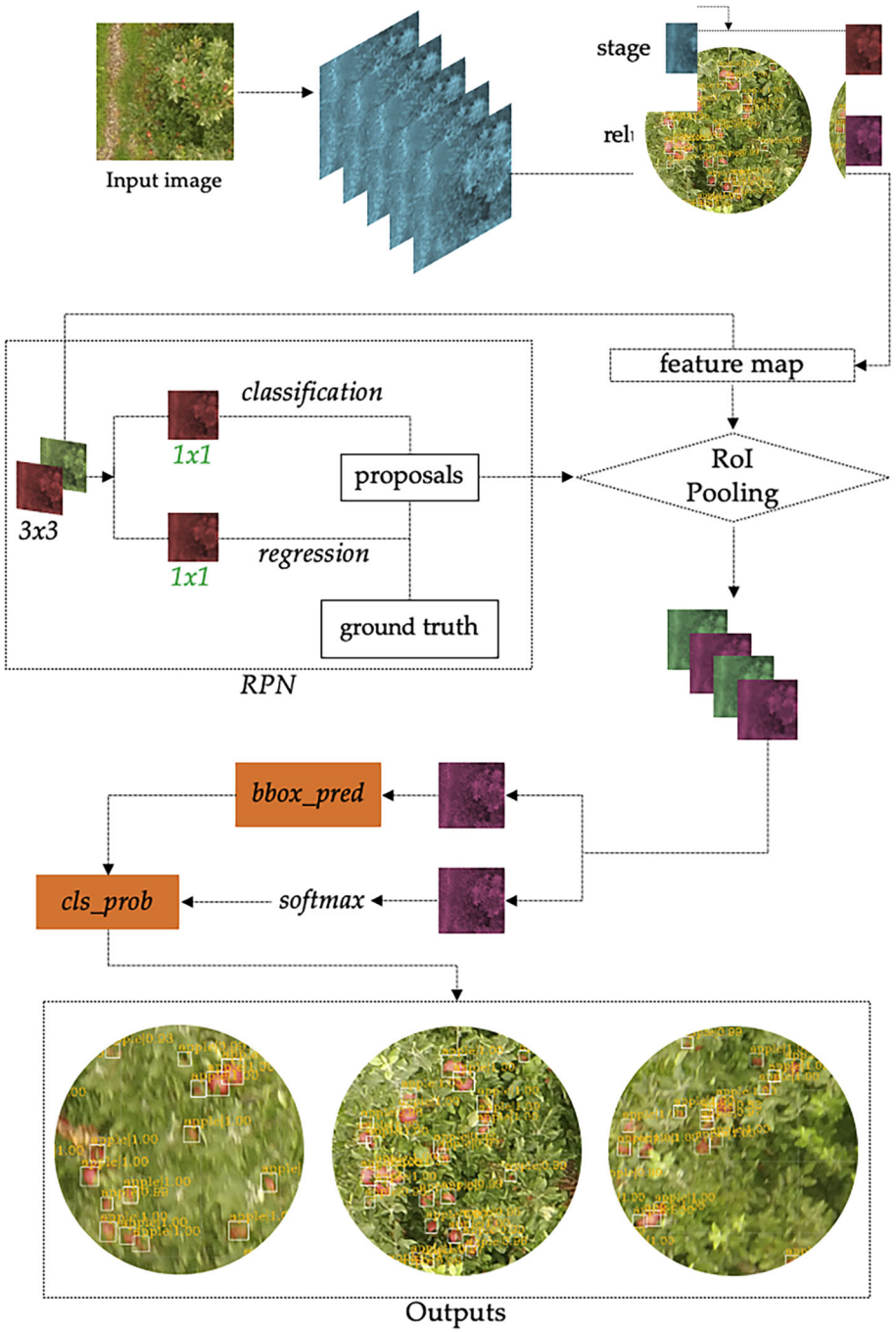

FIGURE 7 | The architecture of Faster R-CNN. "conv" represents the convolutional layer, the "relu" represents the activation function, and the "fc layer" represents the fully connected layer. The network outputs intermediate layers of the same size in the same "stage." "bbox_pred" represents the position offset of the object, "cls_prob" represents the probability of the category, and the outputs show the fruits detected.

underside of the tree. This fruit distribution makes it much easier for the fruit picking operator during the harvesting process (Brendon et al., 2019).

It can also be observed that each apple tree contained between 175 and 308 fruits, with an average of 255. On the other hand, the percentage of fruits on the top of the tree had an average value of 27.31%. Hence, it must be realized that only a part of this percentage of fruits was detected on the images obtained with the UAV.

When visual counts of fruits are made before harvesting, the total number of them can be affected by many factors. The main reasons for this are natural fruit drop and biotic and abiotic

TABLE 1 | Data on the trees randomly selected in the crop field.

Id	Nfcvt	Nfcvm	Nfcvun	Total of fruits	Top (%)	Middle (%)	Underside (%)
1	88	95	92	275	32.00	34.55	33.45
5	85	101	99	285	29.82	35.44	34.74
9	66	77	95	238	27.73	32.35	39.92
14	83	68	77	228	36.40	29.82	33.77
17	56	75	82	213	26.29	35.21	38.50
24	51	93	82	226	22.57	41.15	36.28
25	86	87	70	243	35.39	35.80	28.81
27	77	107	124	308	25.00	34.74	40.26
32	53	78	125	256	20.70	30.47	48.83
33	88	117	64	269	32.71	43.49	23.79
34	56	89	107	252	22.22	35.32	42.46
43	35	202	82	319	10.97	63.32	25.71
48	80	117	70	267	29.96	43.82	26.22
50	90	95	72	257	35.02	36.96	28.02
51	63	44	68	175	36.00	25.14	38.86
56	77	84	78	239	32.22	35.15	32.64
57	48	128	97	273	17.58	46.89	35.53
59	69	88	98	255	27.06	34.51	38.43
62	52	110	108	270	19.26	40.74	40.00
Average	68.58	97.63	88.95	255.16	27.31	37.63	35.06

Id, number of the tree in the field; Nfcvt, Number of fruits visually counted on the top of the tree; Nfcvn, Number of fruits visually counted in the middle of the tree; Nfcvun, Number of fruits visually counted in the lowest part of the tree; (%), percentage values for each part of the tree regarding the total.

factors. Another reason it may be due to visual errors committed by the staff devoted to counting the fruits (i.e., they may count the same fruit twice). In **Figure 8**, a linear regression between the number of fruits counted visually in the field and the number of fruits harvested is shown. An R^2 value of 0.86 was obtained, which indicates a good correlation for both numerical variables. However, the MAE and RMSE values obtained were high, which indicates a bad model adjustment. The low consistency between the number of fruits counted visually and the fruits harvested is probably due to losses during the counting process when using a fruit grading machine. This kind of machine does not detect small-sized fruits; hence, the use of a manual process to count the fruit can improve model adjustment.

The starting point was the premise that the human eye is the most accurate method for detecting fruit on the images (Rosebrock, 2018). In this sense, in **Figure 9**, a linear regression between the number of fruits counted on the image and the number of fruits harvested is displayed. An R^2 value of 0.80 can be observed, which indicates a low correlation. As expected, the number of fruits detected in the images taken by the UAV is insufficient for estimating the rest of the fruits present in the canopy tree with traditional mathematical models. The results show that it is possible, although with a low accuracy, to make predictions of the total number of fruits in each tree using these kinds of images. The high values for MAE and RMSE suggest that, despite of all the fruits being detected using DL algorithms, the variability in the number of fruits harvested with respect to the number of detected fruits cannot be modeled well using standard linear regression.

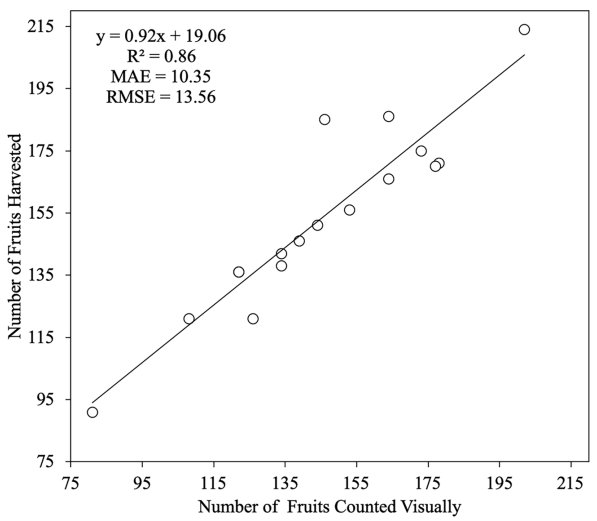

FIGURE 8 | Linear regression between both the Number of Fruits Counted Visually in the field and the Number of Fruits Harvested. The number of trees analyzed was n = 19. The straight line represents the best-fit linear regression ($p < 0.001$).

Distribution of Fruits in an Apple Orchard Canopy

In **Figure 10**, the workflow from an input image until the fruits are detected is shown. Over each detected apple fruit, a blue bounding box with the probability of containing the fruits and a legend with TP, FP, and FN are shown in (**Figure 10B**). Based on the above, it was concluded that the 3,000 images tagged for apple detection were sufficient for explaining the wide variability in the data set, as the number of fruits detected was high. Apparently, the application of data augmentation helped to overcome the problems in object detection caused by illumination conditions, the distance to the fruit, and fruit clustering, among others, as suggested by Voulodimos et al. (2018).

In outdoor conditions, the model could not detect all fruits, but it was able to detect most of the visible fruits. It was observed

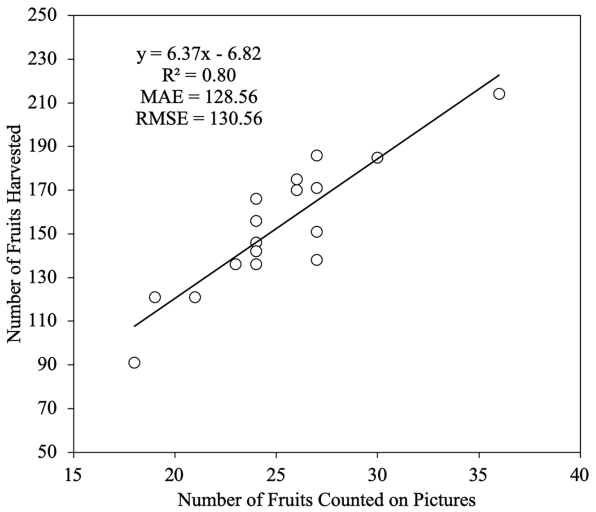

FIGURE 9 | Linear regression between the Number of Fruits Counted on Pictures and the Number of Fruits Harvested. The number of trees analyzed was n = 19. The straight line represents the best-fit linear regression ($p < 0.001$).

that the pictures taken by an UAV suffered from notable changes, mainly due to unstructured light conditions and the camera's rolling shutter effect. Moreover, invisible fruits that are occluded by foliage or other fruits are the main challenge for DL models based on object detection (Kamilaris and Prenafeta-Boldú, 2018). Therefore, in **Table 2**, an analysis of the precision of fruit detection is presented. The values for each of the metrics used to assess the obtained results were greater than 90% in terms of precision (P). Similar results were obtained by Chen et al. (2019), although their results were slightly lower, probably due to the size of strawberry fruits, which are smaller than apple fruits. False positives were observed in the pictures that corresponded to immature fruits (fruits green), and where the brightness of sunlight was slightly greater, and in those pictures that suffered from rolling shutter. These results can be significantly improved by taking pictures several times throughout the day, as suggested by Fu et al. (2018), or by flying the UAV at a low speed. Finally, the F1-score exhibited values greater than 87%, indicating the high robustness of the trained model. On the other hand, with visual counting (Nfp), considered to be the most reliable method, an accuracy of 88.96% was obtained. The errors between visual counts and object detection were similar to those obtained by Neupane et al. (2019) when counting banana fruits. These results demonstrate that the use of simple data augmentation techniques such as picture rotation, filters, and transfer learning can facilitate the building of tools with a high potential for apple fruit detection.

Yield Map Creation

As seen in the previous sections, a highly accurate estimation of the number of fruits per tree is not easy or straightforward. Nevertheless, it is possible to build an apple yield map as a tool to at least approximately determine the number of fruits in each tree of the crop field. This foreseen information could be useful for both farmers (to know the number of staff needed to be contracted) and contractors (to know the volume of production to be transported). In **Figure 11**, there is an apple yield map in which the number of fruits per tree detected is shown. It allows a visualization of the high spatial variability in the field, as well as the expected number of fruits per tree. It also may be affirmed that there is a low percentage (9.12%) of trees with a number of fruits between 30 and 40.

Tree-level information can be useful, but it could be more interesting to have the same information for each row of the crop field. In **Figure 12**, a more actionable apple yield map with the total number of fruits for each row is shown. The results show that row 5 and row 10 contain less fruits in their trees. Meanwhile, row 1 and 14 are the rows with the greatest volumes of fruits. The rest of the rows have a similar number of fruits.

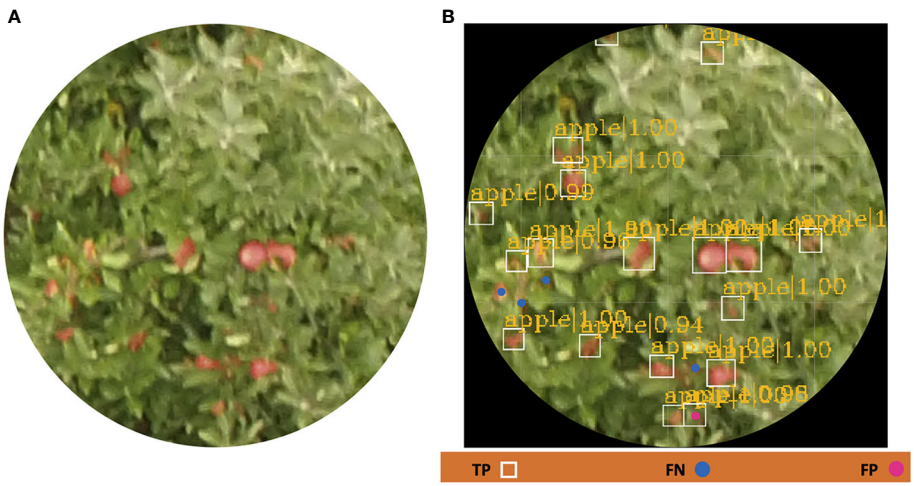

FIGURE 10 | (A) Original RGB picture taken from an UAV and **(B)** apple fruits automatically detected with bounding boxes and their probabilities obtained by the Faster R-CNN model. The errors (TP, FP, and FN) are shown as a legend below the picture.

TABLE 2 | Fruit detection analyses for each of the pictures selected.

Picture	TP	FP	FN	P	R	F1	N_{fp}	A
1	70	2	8	0.97	0.90	0.93	78	0.90
2	44	3	9	0.94	0.83	0.88	53	0.83
3	62	3	6	0.95	0.91	0.93	68	0.91
4	53	4	5	0.93	0.91	0.92	58	0.91
5	20	2	4	0.91	0.83	0.87	24	0.83
6	41	3	9	0.93	0.82	0.87	50	0.82
7	67	3	5	0.96	0.93	0.94	72	0.93
8	83	5	8	0.94	0.91	0.93	91	0.91
9	86	7	10	0.92	0.90	0.91	96	0.90
10	77	6	4	0.93	0.95	0.94	81	0.95
11	80	7	6	0.92	0.93	0.92	86	0.93
12	61	3	6	0.95	0.91	0.93	67	0.91
13	75	4	1	0.95	0.99	0.97	76	0.99
14	73	5	8	0.94	0.90	0.92	81	0.90
15	30	6	4	0.83	0.88	0.86	34	0.88
16	38	5	7	0.88	0.84	0.86	45	0.84
17	56	3	8	0.95	0.88	0.91	64	0.88
18	53	7	3	0.88	0.95	0.91	56	0.95
19	67	8	9	0.89	0.88	0.89	76	0.88
20	91	2	7	0.98	0.93	0.95	98	0.93
Avg	61.35	4.40	6.35	0.93	0.90	0.91	67.70	0.90

TP, True Positive; FP, False Positive; P, Precision; R, Recall; F1, F1score value; N_{fp}, Number of fruits counted on pictures visually; A, Accuracy.

DISCUSSION

Computational Time Required

According to Torres-Sánchez et al. (2015), the computational and processing time is a crucial aspect in this kind of work. In this line, the time needed for each step is discussed in the following lines. The alignment process took 68 min and the dense point cloud analysis took 159 min, without taking into account the quiet time needed to upload images and carry out the image georeferencing process. On the other hand, the time required for these steps mainly depends on the covered area, the number of images and their resolution, and the computer used, as suggested by Ai et al. (2015).

Most of the processing time was spent training the Faster R-CNN model, which took approximately 5 h using Google Colab. This depends on the number of images used, the batch size, the learning rate, and the hardware used, among other factors. This step did not take the time required for image labeling, which is highly time-consuming, into account. It usually takes several working days, since it is a process that is completely manual. In a study by Tian et al. (2019a), a similar number of images was obtained using an analogous training time to this study. However, the studies cannot be completely compared, since the hardware and images used were not the same. The research carried out by Id et al. (2019) to detect banana trees on images taken by an UAV took 180 min to train a similar model. They used over 2,000 images with a resolution of 600 × 600 px. However, the CNN architecture used was Yolo-V3, which is slower than Faster R-CNN, according to Rosebrock (2018).

Assessment of Apple Fruit Detection on UAV Images and Orthomosaics Construction

The main challenge in fruit detection when applying images taken from a UAV is fruit size. In addition, the size of TIF files increases the amount of computational resources needed to train the models. Senthilnath et al. (2016) demonstrated a novel method for detecting tomatoes using UAV images taken with a multispectral camera. They used spectral clustering based on the K-means algorithm to detect tomato fruits. The main problem

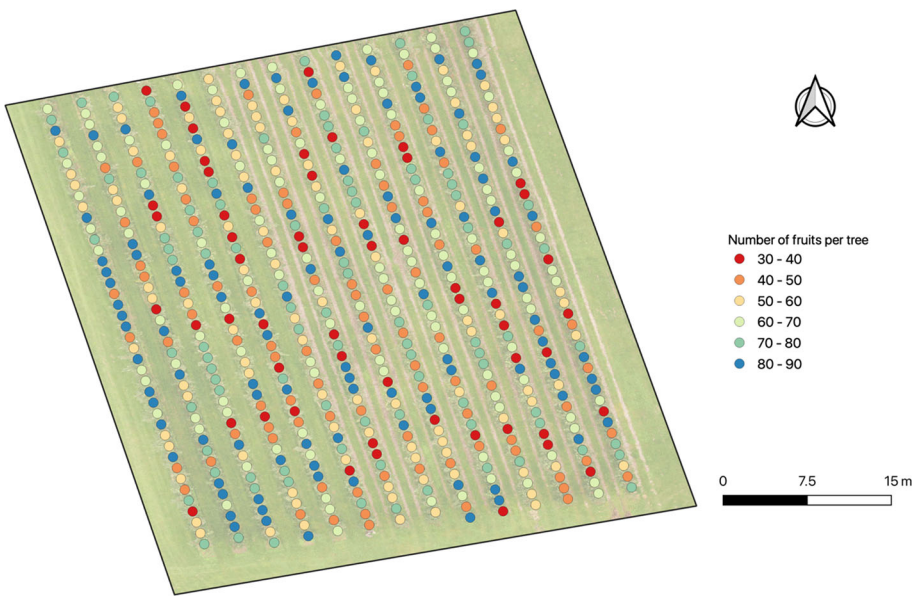

FIGURE 11 | Apple yield map with the number of fruits per tree detected with the Faster R-CNN model trained.

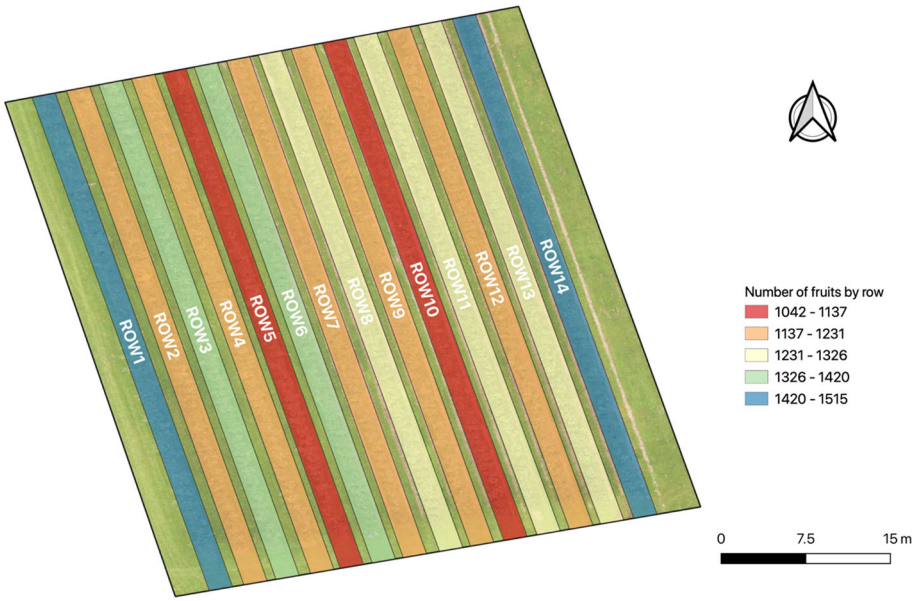

FIGURE 12 | Apple yield map with the total number of fruits by row detected with the Faster R-CNN trained model.

that they found was the inability to detect fruits covered by the leaves. Other studies, such as that proposed by Id et al. (2019) to detect banana trees, obtained similar results in terms of accuracy to what this study concluded. Hence, we can conclude that the method proposed in this study was highly accurate for fruit detection tasks. In addition, the maps generated from the detections in images taken from a UAV represent an innovative proposal that, until now, has not been implemented in an apple crops field.

Regarding the creation of orthomosaics, many of the tools that make use of them apply segmentation techniques to detect objects (fruits, trees, rows, etc.). Csillik et al. (2018) developed an algorithm for citrus tree identification. They applied the CNN workflow using Trimble's eCognition Developer 9.3 (www.ecognition.com). Johansen et al. (2018) also proposed a methodology using multispectral images to detect tree canopies with the intention of determining the number of trees. Although these methods have a high level of accuracy, the process is not completely automated; hence, it can be improved. On the other hand, much of the research that currently applies DL algorithms operates with individual images without georeferencing (Kamilaris et al., 2017). Knowing the accurate position of each element (plants, machinery, sensors, etc.) available on any farm is crucial (Ramin Shamshiri et al., 2018). To our knowledge, our methodology is the first that allows the orchard yield to be estimated based on the number of fruits detected a tree-scale precision on images taken by an UAV.

Figure 13 compares two schematic workflows for the purposes of applying the DL algorithm and other common indexes used in agriculture. On the left (**Figure 13A**), the traditional workflow as used in reference A is presented. This is characterized by the performance of the detection processes on different platforms and in separate steps. For example, the preparation of the datasets is usually done on a conventional computer, while the training of the algorithms is done on a more powerful computer (mainly with advanced GPU hardware). The main advantage of the proposed method (**Figure 13B**) is that Colab allows the data to be prepared and applies fruit detection in georeferenced images on the same platform, which reduces the processing time and leverages the interoperability.

Integration of Automated Yield Estimation Systems Into the Agricultural Domain

Finally, we would like to focus on the translation of this type of fruit detection and counting systems to the agri-economic terrain. Being aware of the advance that this type of technique implies for an early forecast of yield, we think that it may have an impact on the way in which the management of farms is carried out in the coming years. The organization of harvesting tasks, the pruning of trees, or the fruit purchase process itself can be optimized with this type of system. However, we would like to point out that this type of development, although employing collaborative platforms such as the one shown here, *a priori* does not have the average producer as an end user. We envisage that the development of an automated fruit detection system and the possibility of generating variable crop maps, can be marketed as a service within agricultural cooperatives. When demonstrated in a real environment, with a model with several learning campaigns, it can represent an important advance in the adoption of new agricultural management systems. Although the development in this work implements an open data model, with open-source algorithms, the algorithm-as-a-service model is still far from a firm implementation in the agricultural field. Cloud computing and development platforms such as Google Colab have great

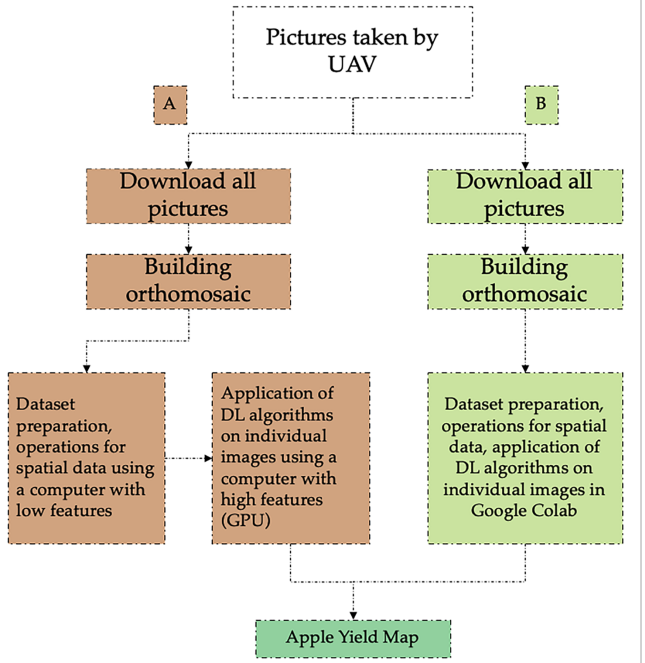

FIGURE 13 | Schematic workflow of the main steps used for orchard yield mapping: **(A)** traditional methodology and **(B)** proposed methodology.

potential in the near future to serve as tools for the creation of advanced services in precision agriculture. Moreover, thanks to advances in GPS position enabling farmers to accurately navigate to specific location in the field, a door opens to an automatic harvested in combination with yield maps and autonomous farm equipment's. The object of these developments can be anticipated as they will be integrated into software solutions and much more automated platforms (Farm Management Information Systems), in which the user will hardly have to interact with the data to obtain reliable forecasts.

CONCLUSIONS

This paper introduces a novel methodology for sampling an apple orchard at the tree level to infer the final yield. It was found that it is possible to detect the number of fruits in apple trees from images taken from a UAV. The assessment of the DL model showed very promising values and, therefore, a great potential of the method is foreseen for the estimation of apple yields and probably the yield of other fruits.

Google Colab's usefulness as a tool for training DL algorithms to build useful tools for farmers was assessed. This cloud environment will make the tool more available for further research and improve orchard management. Moreover, the use of python opens the door to developing web tools with the aim of automating the process. In this case we provide the code used in the **Supplementary Material**.

Future works will involve the automation of all of the processes: the creation of the orthomosaic, individual tree identification, the detection of all the fruits in each tree, and the generation of the yield map on a single platform integrated in a graphical user interface (GUI). This will provide stakeholders a useful and easy-to-use tool. Moreover, the combination of historical data from several seasons will be tested to build models where data and images converge to obtain accurate results.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version. OEA-A wrote the first draft of the paper and analyzed data. JV conceived the experiments, flew the UAV to take the pictures, and conducted the field measurements. MP-R provided guidance for the analysis of data and writing of the manuscript. JM-G provided suggestions on the structure of the manuscript and participated in discussions of the results.

FUNDING

This work was partially supported by the project MARS4Earth: Modular Aerial Robotic Systems for Sustainable Living on Earth (RAAK.PRO03.112), which is funded by the Netherlands Organisation for Scientific Research and the project AGL2016-78964-R funded by the Spanish Ministry of Economic and Competence.

ACKNOWLEDGMENTS

The authors would like to thank Pieter van Dalfsen and Peter Frans de Jong from Wageningen Plant Research for providing assistance during the field work at the orchard and for sharing their knowledge, which helped us understand more about the apple orchard lifecycle. Additionally, the authors want to thank the Predoctoral Research Fellowship for the development of the University of Seville R&D&I program (IV.3 2017) granted to OEA-A.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01086/full#supplementary-material

REFERENCES

Abdelrahman, M. A. E., Natarajan, A., and Hegde, R. (2016). Assessment of land suitability and capability by integrating remote sensing and GIS for agriculture in Chamarajanagar district, Karnataka, India. *Egypt. J. Remote Sens. Sp. Sci.* 19, 125–141. doi: 10.1016/j.ejrs.2016.02.001

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A. E., and Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon* 4, e00938. doi: 10.1016/j.heliyon.2018.e00938

Aggelopooulou, K., Castrignanò, A., Gemtos, T., and De Benedetto, D. (2013). Delineation of management zones in an apple orchard in Greece using a multivariate approach. *Comput. Electron. Agric.* 90, 119–130. doi: 10.1016/j.compag.2012.09.009

Ai, M., Hu, Q., Li, J., Wang, M., Yuan, H., and Wang, S. (2015). A robust photogrammetric processing method of low-altitude UAV images. *Remote Sens.* 7, 2302–2333. doi: 10.3390/rs70302302

Anders, N., Valente, J., Masselink, R., and Keesstra, S. (2019). Comparing Filtering Techniques for Removing Vegetation from UAV-Based Photogrammetric Point Clouds. *Drones* 3, 61. doi: 10.3390/drones3030061

Bargoti, S., and Underwood, J. P. (2017). Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *J. F. Robot.* 34, 1039–1060. doi: 10.1002/rob.21699

Bisong, E. (2019). Building Machine Learning and Deep Learning Models on Google Cloud Platform. *Apress* doi: 10.1007/978-1-4842-4470-8

Brendon, A., Serra, S., and Musacchi, S. (2019). Optimizing crop load for new apple cultivar: "WA38". *Agronomy* 9. doi: 10.3390/agronomy9020107

Carneiro, T., Da Nobrega, R. V. M., Nepomuceno, T., Bian, G., De Albuquerque, V. H. C., and Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access* 6, 61677–61685. doi: 10.1109/ACCESS.2018.2874767

Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., et al. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sens.* 11, 1–21. doi: 10.3390/rs11131584

Chlingaryan, A., Sukkarieh, S., and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* 151, 61–69. doi: 10.1016/j.compag.2018.05.012

Chollet, F. (2017). *Deep Learning with Python*, Available online: https://github.com/keras-team/keras (accessed on 10 October 2019).

Csillik, O., Cherbini, J., Johnson, R., Lyons, A., and Kelly, M. (2018). Identification of Citrus Trees from Unmanned Aerial Vehicle Imagery Using Convolutional Neural Networks. *Drones* 2, 39. doi: 10.3390/drones2040039

Dias, P. A., Tabb, A., and Medeiros, H. (2018). Apple flower detection using deep convolutional networks. *Comput. Ind.* 99, 17–28. doi: 10.1016/j.compind.2018.03.010

Duarte, L., Teodoro, A. C., Moutinho, O., and Gonçalves, J. A. (2017). Open-source GIS application for UAV photogrammetry based on MicMac. *Int. J. Remote Sens.* 38, 3181–3202. doi: 10.1080/01431161.2016.1259685

Fu, L., Feng, Y., Majeed, Y., Zhang, X., Zhang, J., Karkee, M., et al. (2018). Kiwifruit detection in field images using Faster R-CNN with ZFNet. *IFAC-PapersOnLine* 51, 45–50. doi: 10.1016/j.ifacol.2018.08.059

Gillies, S. (2007). *Shapely: manipulation and analysis of geometric objects*, Available online: https://github.com/Toblerity/Shapely (accessed on 15 November 2019).

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., and Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. *Comput. Electron. Agric.* 116, 8–19. doi: 10.1016/j.compag.2015.05.021

Gongal, A., Karkee, M., and Amatya, S. (2018). Apple fruit size estimation using a 3D machine vision system. *Inf. Process Agric.* 5, 498–503. doi: 10.1016/j.inpa.2018.06.002

González-Araya, M. C., Soto-Silva, W. E., and Espejo, L. G. A. (2015). Harvest planning in apple orchards using an optimization model. *Int. Ser. Oper. Res. Manage. Sci.* 224, 79–105. doi: 10.1007/978-1-4939-2483-7_4

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., and Agrawal, A. (2017). Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr. Build. Mater.* 157, 322–330. doi: 10.1016/j.conbuildmat.2017.09.110

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognit.* 77, 354–377. doi: 10.1016/j.patcog.2017.10.013

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing* 187, 27–48. doi: 10.1016/j.neucom.2015.09.116

Neupane, B., Horanont, T., and Hung, N. D. (2019). Deep learning based banana plant detection and counting using high-resolution red-green- blue (RGB) images collected from unmanned aerial vehicle (UAV) 1–22. doi: 10.1371/journal.pone.0223906

Jha, K., Doshi, A., Patel, P., and Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artif. Intell. Agric.* 2, 1–12. doi: 10.1016/j.aiia.2019.05.004

Jiménez, A. R., Ceres, R., and Pons, J. L. (2000). a Survey of Computer Vision Methods for Locating Fruit on Trees. *Trans. ASAE* 43, 1911–1920. doi: 10.13031/2013.3096

Johansen, K., Raharjo, T., and McCabe, M. F. (2018). Using multi-spectral UAV imagery to extract tree crop structural properties and assess pruning effects. *Remote Sens.* 10, 1–21. doi: 10.3390/rs10060854

Jordahl, K. (2014). *GeoPandas Jordahl. Python tools for geographic data*, Available online: https://github.com/geopandas/geopandas (accessed on 30 November 2019).

Kaloxylos, A., Eigenmann, R., Teye, F., Politopoulou, Z., Wolfert, S., Shrank, C., et al. (2012). Farm management systems and the Future Internet era. *Comput. Electron. Agric.* 89, 130–144. doi: 10.1016/j.compag.2012.09.002

Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016

Kamilaris, A., Kartakoullis, A., and Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* 143, 23–37. doi: 10.1016/j.compag.2017.09.037

Koirala, A., Walsh, K. B., Wang, Z., and McCarthy, C. (2019). Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. *Precis. Agric.* 20 (6), 1107–1135. doi: 10.1007/s11119-019-09642-0

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks Alex. *Adv. Neural Inf. Process Syst.* 25, 1097–1105. doi: 10.1061/(ASCE)GT.1943-5606.0001284

Lötze, E., and Bergh, O. (2004). Early prediction of harvest fruit size distribution of an apple and pear cultivar. *Sci. Hortic. (Amsterdam).* 101, 281–290. doi: 10.1016/j.scienta.2003.11.006

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Machwitz, M., Hass, E., Junk, J., Udelhoven, T., and Schlerf, M. (2019). CropGIS – A web application for the spatial and temporal visualization of past, present and future crop biomass development. *Comput. Electron. Agric.* 161, 185–193. doi: 10.1016/j.compag.2018.04.026

Maes, W. H., and Steppe, K. (2019). Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. *Trends Plant Sci.* 24, 152–164. doi: 10.1016/j.tplants.2018.11.007

Mogili, U. R., and Deepak, B. B. V. L. (2018). Review on Application of Drone Systems in Precision Agriculture. *Proc. Comput. Sci.* 133, 502–509. doi: 10.1016/j.procs.2018.07.063

Moltó, E., Plá, F., and Juste, F. (1992). Vision systems for the location of citrus fruit in a tree canopy. *J. Agric. Eng. Res.* 52, 101–110. doi: 10.1016/0021-8634(92)80053-U

Oliver, M. A. (2013). Precision agriculture and geostatistics: How to manage agriculture more exactly. *Significance* 10, 17–22. doi: 10.1111/j.1740-9713.2013.00646.x

Patrício, D. I., and Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* 153, 69–81. doi: 10.1016/j.compag.2018.08.001

Payne, A., Walsh, K., Subedi, P., and Jarvis, D. (2014). Estimating mango crop yield using image analysis using fruit at "stone hardening" stage and night time imaging. *Comput. Electron. Agric.* 100, 160–167. doi: 10.1016/j.compag.2013.11.011

Rahnemoonfar, M., and Sheppard, C. (2017). Deep count: Fruit counting based on deep simulated learning. *Sensors (Switzerland)* 17, 1–12. doi: 10.3390/s17040905

Ramin Shamshiri, R., Weltzien, C., A. Hameed, I., J. Yule, I., E. Grift, T., K. Balasundram, S., et al. (2018). Research and development in agricultural robotics: A perspective of digital farming. *Int. J. Agric. Biol. Eng.* 11, 1–11. doi: 10.25165/j.ijabe.20181104.4278

Rapiński, B., and Zinkiewicz, (2019). JupyTEP IDE as an Online Tool for Earth Observation Data Processing. *Remote Sens.* 11, 1973. doi: 10.3390/rs11171973

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Rosebrock, A. (2016). *Practical Python and OpenCV + Case Studies* (Baltimore: PyImageSearch.com).

Rosebrock, A. (2018). *Deep Learning for Computer Vision with Python. ImageNet Bundle* (Baltimore: PyImageSearch.com).

Roy, P., Kislay, A., Plonski, P. A., Luby, J., and Isler, V. (2019). Vision-based preharvest yield mapping for apple orchards. *Comput. Electron. Agric.* 164, 104897. doi: 10.1016/j.compag.2019.104897

Senthilnath, J., Dokania, A., Kandukuri, M., K.N, R., Anand, G., and Omkar, S. N. (2016). Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosyst. Eng.* 146, 16–32. doi: 10.1016/j.biosystemseng.2015.12.003

Sharma, R., Kamble, S. S., and Gunasekaran, A. (2018). Big GIS analytics framework for agriculture supply chains: A literature review identifying the current trends and future perspectives. *Comput. Electron. Agric.* 155, 103–120. doi: 10.1016/j.compag.2018.10.001

Silwal, A., Karkee, M., and Zhang, Q. (2016). A hierarchical approach to apple identification for robotic harvesting. *Trans. ASABE* 59, 1079–1086. doi: 10.13031/trans.59.11619

Simonyan, K., and Zisserman, A. (2014).Very Deep Convolutional Networks For Large-Scale Image Recognition, in: *Published as a conference paper at ICLR 2015*, . pp. 1–14.

Sonka, M., Hlavac, V., and Boyle, R. (1993). *Image Processing, Analysis and Machine Vision*. SPRINGER-SCIENCE+BUSINESS MEDIA, BV. doi: 10.1007/978-1-4899-3216-7

Stein, M., Bargoti, S., and Underwood, J. (2016). Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. *Sensors* 16, 1915. doi: 10.3390/s16111915

Talukdar, J., Gupta, S., Rajpura, P. S., and Hegde, R. S. (2018).Transfer Learning for Object Detection using State-of-the-Art Deep Neural Networks. *Proceedings of the 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*. Noida, India pp. 78–83. doi: 10.1109/SPIN.2018.8474198

Thompson, N. M., Bir, C., Widmar, D. A., and Mintert, J. R. (2019). Farmer perceptions of precision agriculture technology benefits. *J. Agric. Appl. Econ.* 51, 142–163. doi: 10.1017/aae.2018.27

Tian, Y., Yang, G., Wang, Z., Li, E., and Liang, Z. (2019a). Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense. *J. Sensors* 2019, 1–13. doi: 10.1155/2019/7630926

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., and Liang, Z. (2019b). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* 157, 417–426. doi: 10.1016/j.compag.2019.01.012

Torres-Sánchez, J., López-Granados, F., Serrano, N., Arquero, O., and Peña, J. M. (2015). High-throughput 3-D monitoring of agricultural-tree plantations with Unmanned Aerial Vehicle (UAV) technology. *PloS One* 10, 1–20. doi: 10.1371/journal.pone.0130479

Torres-Sánchez, J., de Castro, A. I., Peña, J. M., Jiménez-Brenes, F. M., Arquero, O., Lovera, M., et al. (2018). Mapping the 3D structure of almond trees using UAV acquired photogrammetric point clouds and object-based image analysis. *Biosyst. Eng.* 176, 172–184. doi: 10.1016/j.biosystemseng.2018.10.018

Turner, D., Lucieer, A., and Watson, C. (2012). An automated technique for generating georectified mosaics from ultra-high resolution Unmanned Aerial Vehicle (UAV) imagery, based on Structure from Motion (SFM) point clouds. *Remote Sens.* 4, 1392–1410. doi: 10.3390/rs4051392

Tzutalin (2015). *LabelImg. Git code*, Available at: https://github.com/tzutalin/labelImg.

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* 2018, 1–13. doi: 10.1155/2018/7068349

Willaume, M., Lauri, P. É., and Sinoquet, H. (2004). Light interception in apple trees influenced by canopy architecture manipulation. *Trees - Struct. Funct.* 18, 705–713. doi: 10.1007/s00468-004-0357-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Apolo-Apolo, Pérez-Ruiz, Martínez-Guanter and Valente. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 02 September 2020
doi: 10.3389/fpls.2020.534853

Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning

Zhe Lin[1] *and Wenxuan Guo*[1,2]*

[1] Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, United States, [2] Department of Soil and Crop Sciences, Texas A&M AgriLife Research, Lubbock, TX, United States

Machine learning and computer vision technologies based on high-resolution imagery acquired using unmanned aerial systems (UAS) provide a potential for accurate and efficient high-throughput plant phenotyping. In this study, we developed a sorghum panicle detection and counting pipeline using UAS images based on an integration of image segmentation and a convolutional neural networks (CNN) model. A UAS with an RGB camera was used to acquire images (2.7 mm resolution) at 10-m height in a research field with 120 small plots. A set of 1,000 images were randomly selected, and a mask was developed for each by manually delineating sorghum panicles. These images and their corresponding masks were randomly divided into 10 training datasets, each with a different number of images and masks, ranging from 100 to 1,000 with an interval of 100. A U-Net CNN model was built using these training datasets. The sorghum panicles were detected and counted by a predicted mask through the algorithm. The algorithm was implemented using Python with the Tensorflow library for the deep learning procedure and the OpenCV library for the process of sorghum panicle counting. Results showed the accuracy had a general increasing trend with the number of training images. The algorithm performed the best with 1,000 training images, with an accuracy of 95.5% and a root mean square error (RMSE) of 2.5. The results indicate that the integration of image segmentation and the U-Net CNN model is an accurate and robust method for sorghum panicle counting and offers an opportunity for enhanced sorghum breeding efficiency and accurate yield estimation.

Keywords: deep learning, computer vision, sorghum panicle, unmanned aerial systems, convolutional neural networks, python, TensorFlow, image segmentation

OPEN ACCESS

Edited by:
Spyros Fountas, Agricultural University of Athens, Greece

Reviewed by:
Zhanguo Xin, United States Department of Agriculture, United States
Lingfeng Duan, Huazhong Agricultural University, China
Wanneng Yang, Huazhong Agricultural University, China

***Correspondence:**
Wenxuan Guo
wenxuan.guo@ttu.edu

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *14 February 2020*
Accepted: *17 August 2020*
Published: *02 September 2020*

Citation:
Lin Z and Guo W (2020) Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning. Front. Plant Sci. 11:534853. doi: 10.3389/fpls.2020.534853

INTRODUCTION

Sorghum (*Sorghum bicolor* L. Moench) is the fifth top cereal crop in the world, which provides nutrition to humans and livestock, particularly in warm and arid climates (FAO, 1999). Sorghum is one of the most efficient crops in the conversion of solar energy and the use of water. It has numerous varieties, including grain sorghums used for human food, and forage sorghum for livestock hay and fodder (Dahlberg et al., 2015). By measuring the plant population and the weight per panicle, growers can estimate the potential final grain yield (Norman et al., 1995). However, it is challenging to determine

plant population by manually counting sorghum panicles, especially for large fields. Traditional counting methods for yield estimation are mainly focused on hand-sampling in the field, which is tedious, time-consuming, labor-intensive, and prone to human errors. Therefore, it is critical to develop alternative methods to efficiently and accurately count sorghum panicles for determining population and estimating yield.

Technological innovations in platforms and advanced sensors such as unmanned aerial systems (UAS) and efficient image processing capabilities provide opportunities to automate high-throughput plant phenotyping through computer vision. UAS imaging has been widely used in plant phenotyping and precision agriculture-related research. Many low-cost sensors onboard UAS can provide aerial images with centimeter-level spatial resolutions. Further, UAS allows for more flexibility in image acquisition in terms of flight height, flight area, and weather conditions. Different sensors onboard the UAS offer various ways for researchers and growers to characterize plant attributes at different scales. As a result, UAS has become a useful platform for crop growers and researchers to acquire aerial images with high spatial and temporal resolutions for quantifying within-field variations (Gómez-Candón et al., 2014). For example, RGB (red, green, and blue bands) cameras, multispectral and thermal sensors were applied to estimate LAI (Hunt et al., 2010; Verger et al., 2014), biomass (Hunt et al., 2005; Bendig et al., 2015), water stress (Gago et al., 2015; Ballester et al., 2018), plant height (Bendig et al., 2015; Díaz-Varela et al., 2015), plant density (Jin et al., 2017; Liu et al., 2017), plant counts (Chen et al., 2017b; Gnädinger and Schmidhalter, 2017; Guo et al., 2018; Olsen et al., 2018; Oh et al., 2019), plant and soil temperature (Gómez-Candón et al., 2016; Zhang et al., 2018), and plant nitrogen status (Hunt et al., 2005; Tokekar et al., 2016). Yang et al. (2017) provided a review on how UAS remote sensing and multiple sensors were applied in field-based plant phenotyping.

Image segmentation is commonly the first step to extract information of targets from an image by separating a set of pixels containing the objects of interest (Mochida et al., 2018). The application of image segmentation for plant phenotyping is typically implemented at small scales because the input requires detailed information with accurate labels, which is time-consuming and labor-intensive. Machine learning, together with computer vision, offer opportunities for high-throughput plant phenotyping in recent years. Machine learning can be broadly defined as computational methods to make accurate predictions or improve performance using experience (Mohri et al., 2018). Deep learning refers to a class of machine learning techniques that leverage multiple layers of non-linear information processing for unsupervised or supervised feature extraction and transformation, and for classification and pattern analysis (Deng et al., 2014). Deep learning algorithms learn high-level features in an incremental way, which eliminates the need for feature identification and extraction (LeCun et al., 2015). The deep networks have the capacity to learn complex models that involve crop phenotypic attributes. A variety of vision-based algorithms have been proven effective with high accuracy in plant phenotyping, such as crop or leaf recognition (Sankaran et al., 2015; Gómez-Candón et al., 2016; Sladojevic et al., 2016), disease detection (Barbedo, 2014; Pérez-Ortiz et al., 2015; Too et al., 2019), crop classification (Makantasis et al., 2015; Dyrmann et al., 2016; Kussul et al., 2017), and crop or fruit counting (Pape and Klukas, 2015; Chen et al., 2017b; Qureshi et al., 2017; Guo et al., 2018; Hasan et al., 2018; Olsen et al., 2018; Ubbens et al., 2018; Madec et al., 2019; Oh et al., 2019; Xiong et al., 2019). In recent years, traditional machine learning and deep learning algorithms were used on image segmentation, especially in the areas of biomedical and object detection. For example, Chen et al. (2017a) developed the Deeplab system and Fully Convolutional Network for semantic image segmentation. Ronneberger et al. (2015) used a U-Net convolutional neural networks (CNN) algorithm with limited training images for the segmentation of neuronal structures in electron microscopic images. Few studies integrated image segmentation in traditional machine learning or deep learning models for plant phenotyping applications. Islam et al. (2017) detected potato diseases on individual leaves using image segmentation and the multiclass support vector machine. Wu et al. (2019) combined image segmentation with VGG-16 CNN on automatic counting of rice seedlings from UAS images. Traditional machine learning and deep learning architectures have been applied to sorghum panicle detection and counting. Guo et al. (2018) used a two-step, decision-tree-based pixel segmentation model (DTSM), and Support Vector Machine (SVM) method with the Classification Learner in sorghum panicle detection. Olsen et al. (2018) developed a machine learning algorithm using image annotation to detect and count sorghum panicles with a mean absolute error of 2.66. Ghosal et al. (2019) proposed a weakly supervised semi-trained CNN model using UAS images for sorghum panicle detection and rough localization. Therefore, image segmentation, together with machine learning, has the potential to detect sorghum panicles and estimate the panicle shape, which can further improve the accuracy of yield prediction.

For effective deep learning algorithms in agricultural applications, model selection and feature definition are critical, which heavily rely on specialized knowledge in both plant phenotyping and computer science (Singh et al., 2016). Environmental factors such as cloud and windy weather can significantly impact the quality of agricultural images (Ghosal et al., 2019). In addition, plant phenotyping based on UAS images is also sensitive to plant genotypes, sensor-target angles, overlap among leaves and panicles, panicle damages, and field conditions. As a result, a large number of training images are required to accommodate various environmental conditions to obtain robust and accurate machine learning algorithms for plant phenotyping tasks. However, building a large number of training samples requires a long time and heavy labor. As a result, datasets of crop images are not yet available on a large scale due to the expenses involved in collecting and preparing the corresponding training data. Therefore, it is critical to develop algorithms that determine the appropriate number of images to meet the requirement of accurate plant phenotyping, such as sorghum panicle counting. The objectives of this study were to

1) develop a deep learning CNN image segmentation algorithm to detect and quantify sorghum panicles; 2) evaluate the performance of this algorithm with respect to the number of training images.

MATERIALS AND METHODS

Experimental Sites

This study was conducted in a research field (33° 35' 50.53" N, 101° 54' 27.30" W) in Lubbock, Texas, in 2018. The climate in this region is semiarid, with an average annual rainfall of 487 mm, mostly falling between May and September, frequently as the result of convective thunderstorms (US Climate Data, 2019). Three sorghum varieties, including NK180, XM217, and KS585 (S&W Seed Company, Sacramento, CA) with two seed populations of 120 and 180, were planted on May 26, 2018. In total, there were 120 plots, each of 6 m long and eight rows wide in an east-west direction. A 1.5-m alley was arranged between plots. The NK180 is a drought-tolerant, bird resistant, and early-maturity variety. The NK180 has a whitish color and a relatively large sorghum panicle. The average measured panicle length for this variety in this study was 22 cm. The XM217 has a red color and a relatively small sorghum panicle. The average panicle length was 13 cm. The KS585 is a drought-tolerant, medium height, and medium-maturity variety. The KS585 has a light brown color, which is close to the soil color, and a relatively small sorghum panicle. The average panicle length was 14 cm. A subsurface drip irrigation system was used for irrigation in this field during the growing season.

UAS Image Collection

A DJI Phantom 4 Pro UAS (DJI, Shenzhen, China) with a 4K RGB camera was applied in image acquisition. The UAS has a 2-axis gimbal that can maintain the orientation of the camera independently from the movement. The UAS is controlled with a 2.4 GHz frequency bidirectional transmission that receives data of the battery voltage, Global Positioning System (GPS) reception, the distance, and the height differences from the home point. The maximum flight duration of the UAS is about 30 min. The flight plan was created using the Pix4Dcapture software (Pix4D S.A., Switzerland). The flight plan included 80% front overlap and 80% side overlap. The angle of the camera was set at 90 degrees to the land surface during flight. The UAS was flying at an altitude of 10 m at 2.7 m s^{-1} speed. The spatial resolution was 2.7 mm for 10 m altitude. Two image datasets were acquired on August 24 and September 10, 2018. All image acquisitions were completed under sunny conditions with light to moderate wind around local solar noon. Raw images were stitched into a whole image using the Pix4DMapper software (Pix4D S.A., Switzerland).

This study applied an integrated method of image segmentation and deep learning for sorghum panicle detection and counting. **Figure 1** shows the steps of the algorithm for sorghum panicle detection and counting. The U-Net CNN (Ronneberger et al., 2015) was adopted as the deep learning framework to train and test the image data.

Preparing Training Images and Masks

The training images were prepared by randomly cropping the raw UAS images using the Microsoft Paint 3D software (Microsoft Corporation, Redmond, WA). To accurately separate sorghum panicles from other objects in the image, a segmentation mask for each training image was created by encircling the sorghum panicle pixels using the Adobe Photoshop CC software (Adobe Systems Inc., San Jose, CA). Specifically, for each training image of 1024 x 1024 pixels, pixels were divided into two classes, the sorghum panicle class and the non-panicle class. In the mask, the pixels containing sorghum panicles were digitized as white and assigned a value of 1, while the other pixels were set black and assigned the value of 0 (**Figure 2**). These mask images were saved separately to ensure that each mask matched its corresponding training image when running the U-Net CNN model. The full training dataset contained 1,000 images. To test the model performance as a function of the number of training images, a series of 10 randomly selected sub-datasets, ranging from 100 to 1,000 with an interval of 100 images (i.e., 100, 200, …, 1000 images), were generated from the full training dataset. Each sub-dataset was used to train a model and tested for the accuracy of the panicle count for the corresponding number of training images.

U-Net Convolutional Neural Networks

The general procedure of the U-Net CNN in this study is described as follows. The U-Net architecture consists of three sections: the contraction, the bottleneck and the expansion. In this study, there were six blocks in the contraction and the expansion sections. The kernel size was 3 x 3 and the strides were 1 x 1 in the contraction section. In the expansion section, the kernel size was 2 x 2 and the strides were 2 x 2. No padding was applied in either section. In the contraction section, each block contained two convolution layers, followed by a down-sampling layer. Once every pixel was processed after the convolution layers, the result was saved into a new feature map in the same arrangement as the input image. The down-sampling layer was used to reduce the feature map dimension, so only the most essential bits of the feature map were kept. The reduced feature map was then utilized as an input to the next contraction block. The spatial dimensions of the feature maps were halved and the number of feature maps was doubled repeatedly through the down-sampling layer (Guan et al., 2019; Weng et al., 2019). The bottleneck layer, which contained two convolution layers but without max pooling, mediated the contraction section and the expansion layer. The data at the bottleneck had the spatial dimension of 32 x 32 with 2048 feature maps. In the expansion section, the block contained two convolution layers followed by an up-sampling layer. After each up-sampling layer, the number of feature maps was halved and the spatial dimensions of the feature maps were doubled to

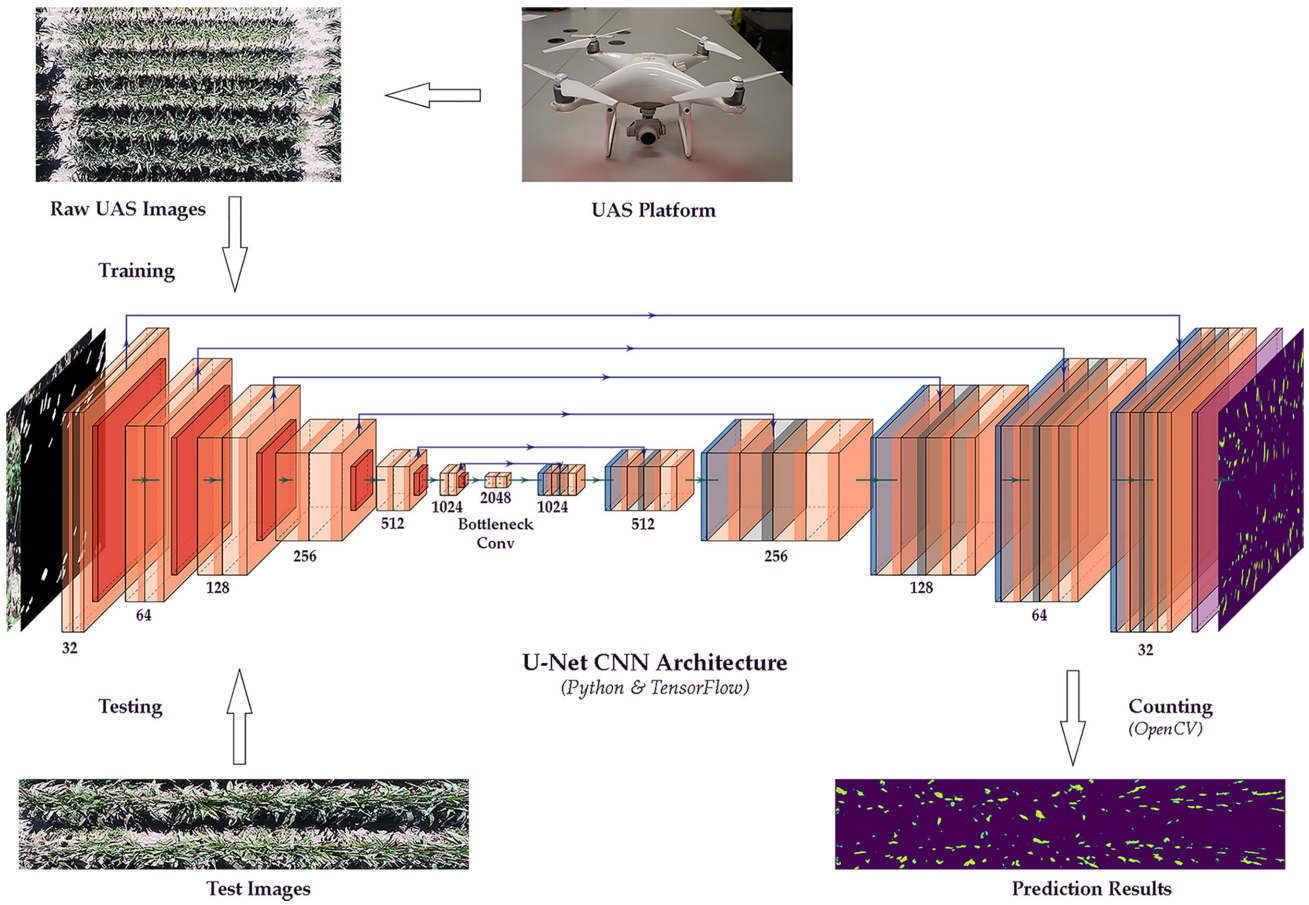

FIGURE 1 | Flow chart of a sorghum panicle detection and counting algorithm using a U-net Convolutional Neural Networks model on unmanned aerial system images.

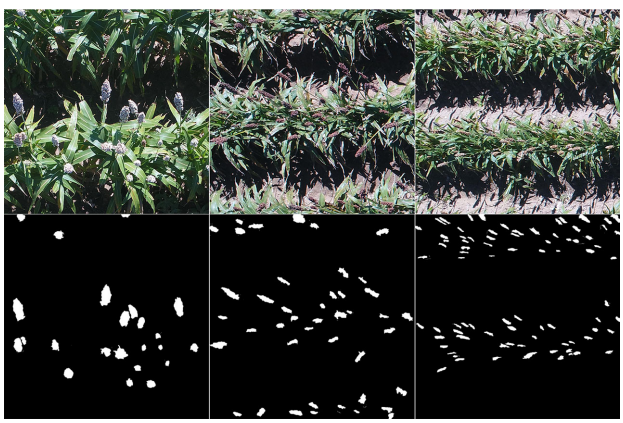

FIGURE 2 | Examples of training images (Top) and corresponding masks (Bottom) for a sorghum panicle detection and counting algorithm using a Convolutional Neural Networks model on unmanned aerial system images (Left to right: NK180, XM217, and KS585).

maintain the whole architecture symmetry. In the meantime, the input from the corresponding contraction block was appended to the feature maps. After running all the expansion blocks, the final output feature map with the same spatial dimension as the original input image included the sorghum panicle class and the non-panicle class.

Segmentation Model Training and Validation

For each training dataset, 90% of the images were set as training, and the rest 10% was used as validation for the training models. For example, in a 500-image dataset, 450 images were trained through the model, and the rest 50 images were used as validation. Before training the segmentation model, all training images and masks went through the image augmentation processes. The hue of each RGB training image was adjusted by a factor of 0.1. Both the training images and corresponding masks were flipped horizontally along the central axis with a 0.5 probability. The training images and corresponding masks were randomly shifted either horizontally or vertically. Finally, both training images and

the corresponding masks were rescaled by a factor of 1/1023. The purpose of image augmentation was to increase the amount of training data by applying some transformations to the original training images. This helps the model to generalize better to unseen data and prevent overfitting (Wang and Perez, 2017; Frid-Adar et al., 2018; Mikołajczyk and Grochowski, 2018). After the image augmentation process, a two-channel segmentation model was generated from these training images and masks using the U-Net algorithm.

The pixel-wise cross-entropy loss function was used to evaluate the training models of the U-Net CNN algorithm using the 10% validation images in the training datasets. The cross-entropy loss is commonly used as a loss function for training in deep learning networks, especially in image segmentation tasks (Ronneberger et al., 2015; Sudre et al., 2017; Martins and Zaglia, 2019). Cross-entropy loss measures the probability difference between the observed and the predicted values in a classification model (Buja et al., 2005). The cross-entropy loss (CE) for the binary classification in this study is defined as,

$$CE = -(y_i log(p_i) + (1 - y_i) log(1 - p_i)) \quad (1)$$

where y_i represents the labeled value for that sample in the mask, and $\boldsymbol{p_i}$ represents the predicted probability being the sorghum panicle in the output feature maps.

Counting and Evaluations

A test dataset containing 120 images was selected from the fully stitched image for accuracy assessment. The images in the test dataset were different from the images in the training dataset. Each test image was corresponding to two rows of sorghum plants randomly selected from a plot. Sorghum panicles in each test image were manually counted, and the number of sorghum panicles in these test images varied from 95 to 188. The size of each test image was 3800 x 1280. We found out that it was difficult for the U-Net CNN model to process the high-resolution test images directly. In this case, we horizontally split each test image into four non-overlapped subtest images. Before running the model on the test images, the subtest images for each test image were resized to dimensions of 1024 x 1024 pixels. Then the test images were run through the trained segmentation model to perform the panicle detection. Each sorghum panicle detected was treated as a contour using the *findContours* function of the OpenCV library in the prediction output feature map. Our initial assessment indicated that contours with less than six pixels were mainly noise related and not classified as panicles. A bounding polygon was applied around each panicle contour using the *drawContours* function for each subtest image. Therefore, the number of bounding polygons represented the number of predicted sorghum panicles in each subtest image. The summation of sorghum panicles of the four subtest images equaled the total number of sorghum panicles in each test image.

The mean absolute error (MAE), mean absolute percentage error (MAPE), accuracy (ACC), coefficient of the determination (R^2), and the root mean squared error (RMSE) were used as evaluation metrics to assess the performance of the sorghum panicle counting algorithm.

$$MAE = \frac{1}{n}\sum_{1}^{n}|m_i - c_i| \quad (2)$$

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left|\frac{m_i - c_i}{m_i}\right| \quad (3)$$

$$ACC = (1 - \frac{1}{n}\sum_{1}^{n}\frac{|m_i + c_i|}{m_i}) \times 100\% \quad (4)$$

$$R^2 = 1 - \frac{\sum_1^n (m_i - c_i)^2}{\sum_1^n (m_i - \overline{m_i})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_1^n (m_i - c_i)^2}{n}} \quad (6)$$

where m_i, $\bar{m}_i$, and c_i represent the manually counted sorghum panicles for the i^{th} image, the mean manual counts, and the predicted count for the i^{th} image, respectively. n is the number of test images.

Hardware and Libraries Used

The algorithm was implemented using the Python programming language (Python Software Foundation, 1995). The model was trained on a computer with 192 GB of memory at the Texas Tech University High Performance Computing Center (HPCC). Training, evaluation, and testing were performed using the Keras (Chollet, 2015) high-level neural networks application programming interface (API), running on top of the TensorFlow package (Abadi et al., 2016). The model in this study was trained using the Adam (Kingma and Ba, 2015) optimizer with a learning rate of 0.001. Fifteen epochs were performed in the training process. The number of epochs was determined based on the training image size, training required time, and the overall performance of the model. In this study, the cross-entropy loss value did not decrease significantly beyond 15 epochs. The OpenCV-Python library (Bradski, 2000) was used in model testing.

RESULTS

Training Model and the Number of Training Images

Table 1 shows the overall decreasing trend of the cross-entropy loss with the number of training images, indicating an increasing accuracy in model performance with the number of training images. The value of cross-entropy loss did not decrease rapidly from 100 to 500 training images. On the other hand, from 600 training images, every 100 more training images resulted in a decrease of more than 0.10 in cross-entropy loss. The value of cross-entropy loss and the trend indicated that there could be

potential to improve the performance of the segmentation model by increasing the number of training images. However, due to the restriction of the training period in our HPCC and the CPU memory, the value of the cross-entropy loss used here was based on 15 epochs of training for all different numbers of training images. Many studies have shown that the cross-entropy loss value could be close to 0 with a large number of training epochs (Alom et al., 2018; Zhang and Sabuncu, 2018). However, with the number of training epochs, the results of cross-entropy values showed a clear negative trend as the number of training images increased.

Sorghum Panicle Counting Performance and the Number of Training Images

Figure 3 presents the accuracy and coefficient of determination (R^2) of the model performance in relation to the number of training images. In general, the sorghum panicle count accuracy and R^2 values increased with the number of training images. This trend, however, was not consistent for the cases with training images below 500. The accuracy was low at 59% with 100 training images, increased to 75% and 80% with 200 and 300 training images, respectively, but dropped slightly to 78% with 400 training images. Similarly, the R^2 value was lowest at 0.01 with 100 training images, increased to 0.09 and 0.17 with 200 and 300 training images, but dropped to 0.08 with 400 images. For cases with more than 500 training images, the accuracies and R^2 values consistently increased with the number of training images. With 1000 training images, the highest accuracy of 95.5%, and the highest R^2 value of 0.90 were achieved. In addition, the rate of change in relation to the number of images for R^2 was greater than that for the accuracy. From 500 to 1,000 images, the accuracy increased by 16.5% from 82% to 95.5%, while the R^2 value increased by 900% from 0.09 to 0.90. This indicates the accuracy is a better parameter for evaluating this type of algorithm performance. In summary, the algorithm performance was not stable with less than 500 training images. With more than 500 training images, the algorithm performance steadily improved with respect to accuracy.

MAE, MAPE, and RMSE consistently decreased with the increasing number of training images (**Table 2**). These trends were not consistent with the trends for the accuracy and R^2 values, which had fluctuations in the relation between the magnitude and number of training images. For a low number of training images, the MAE was relatively large; it was 53.1 for 100

TABLE 1 | Cross-entropy loss values for 10 sets of training images for a sorghum panicle detection and counting algorithm using a U-Net Convolutional Neural Networks model on unmanned aerial system images.

No. of Images	100	200	300	400	500	600	700	800	900	1000
Cross-entropy loss	0.71	0.66	0.63	0.59	0.53	0.50	0.39	0.34	0.20	0.11

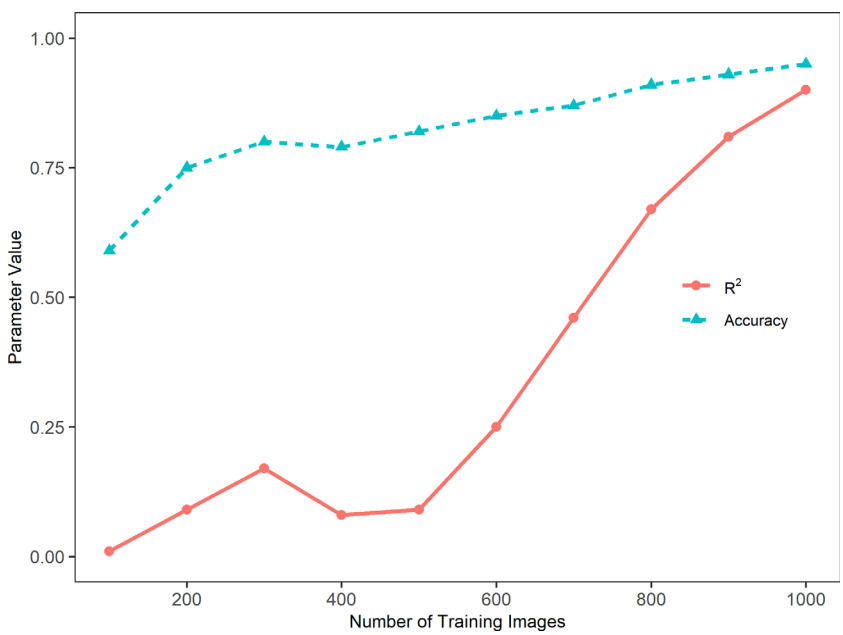

FIGURE 3 | Trends of accuracy, Cross-Entropy Loss, and coefficient of determination (R^2) with the number of training images in a sorghum panicle detection and counting algorithm using a Convolutional Neural Networks model on unmanned aerial system images.

TABLE 2 | Mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE) for 10 sets of training images for a sorghum panicle detection and counting algorithm using a U-Net Convolutional Neural Networks model on unmanned aerial system images.

No. of Images	100	200	300	400	500	600	700	800	900	1000
MAE	53.1	35.2	28.1	27.2	23.7	19.6	16.9	11.4	9.9	6.3
MAPE	0.41	0.25	0.20	0.21	0.18	0.15	0.13	0.09	0.07	0.05
RMSE	7.3	5.9	5.3	5.2	4.9	4.4	4.1	3.4	3.2	2.5

training images and 35.2 for 200 training images. This value dropped to 6.3 for 1,000 images. MAPE was 0.41 for 100 training images and 0.25 for 200 training images, then it dropped to 0.05 for 1,000 training images. A similar trend was observed for the RMSE values. Its change, however, was not as extreme as those MAE values. RMSE was 7.3 for 100 training images, and it gradually dropped to 2.5 for 1,000 training images. Considering the range of sorghum panicles (95 to 188) in the test dataset, the MAE and RMSE values for 1,000 training images are within an acceptable range.

To better evaluate the algorithm performance with respect to the patterns of over- and under-estimations, **Figure 4** shows the error, the difference in sorghum panicles between the model prediction and the manual count result, in relation to the number of training images. If the error is positive, then the algorithm overestimates sorghum panicles; otherwise, the algorithm underestimates. For the cases of 100, 400, 600, 700, and 900 training images, the results represented a mean overestimation of 50.1, 23.7, 20.1, 16.9, and 10.1 panicles, respectively. For 200, 300, and 800 training images, the results represented a mean underestimation of 35.2, 28.0, and 11.4 panicles, respectively. The mean errors were 3.9 and 2.6 for the cases of 500 and 100 images, respectively. However, the variance of prediction results for the 500 training images was larger than that for the 1000 training images. A key to the success of deep learning in object detection tasks is abundant training images. A larger number of training images results in better accuracy and performance (Kamnitsas et al., 2017; Aggarwal et al., 2019). Therefore, the accuracy and robustness of this algorithm increased with the number of training images, with 1,000 images providing the best performance.

It appeared that the counting accuracy was related to the soil background. **Figures 5–7** show the examples of sorghum panicle detection results for the three varieties with 100, 500, and 1,000 training images. For the case of 100 training images, the prediction was overestimated by 53 on average. This substantial overestimation was due to some soil pixels between plots being counted as sorghum panicles, especially for XM217 and KS585. The sorghum panicle colors of these two varieties were similar to the soil background. Therefore, with only 100 training images, the U-Net CNN algorithm was not able to distinguish the soil and sorghum panicles with similar colors. For the case of 500 training images, both overestimation and underestimation were observed. For

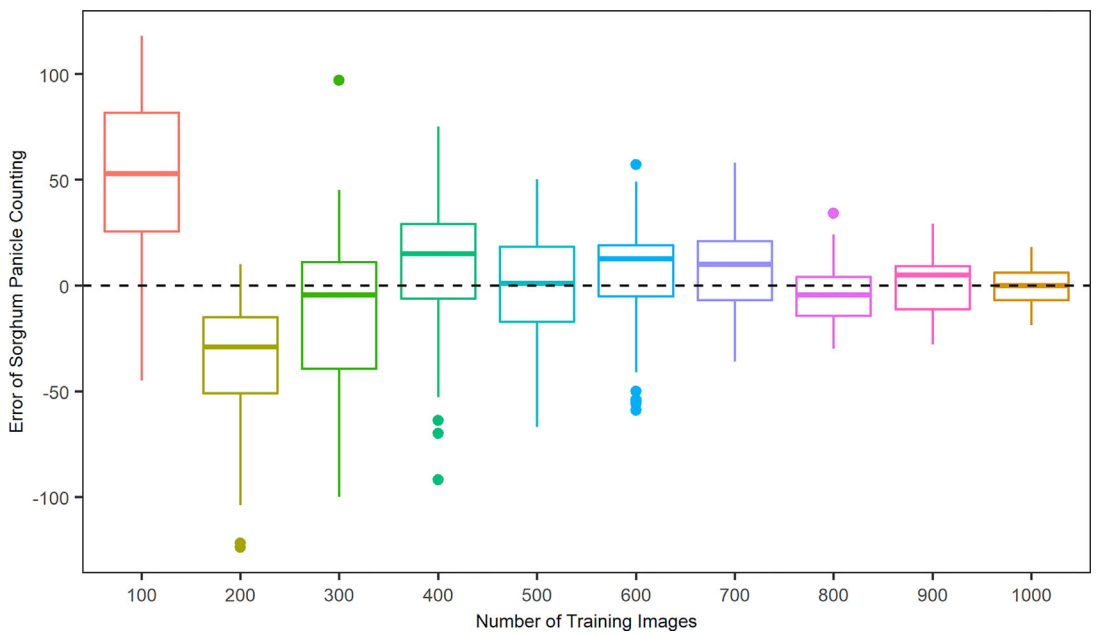

FIGURE 4 | Distributions of counting errors between predicted and observed sorghum panicles for ten sets of training images using a Convolutional Neural Networks model on unmanned aerial system images.

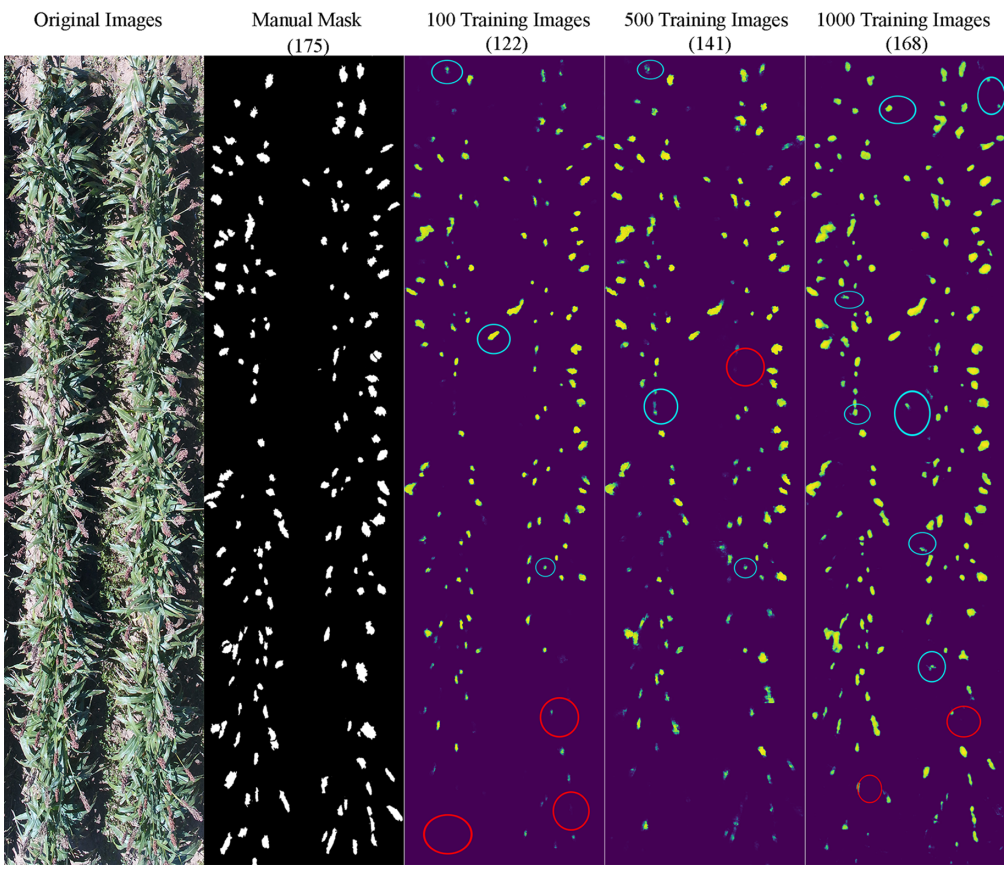

FIGURE 5 | Sample results of sorghum panicle detection for Variety XM217 with 100, 500, and 1000 training images using a Convolutional Neural Networks model and UAS images. Red circles represent underestimation; blue circles represent overestimation compared to the manual masks.

XM217, 141 sorghum panicles were predicted compared to 175 values observed. For KS585, the predicted number of sorghum panicles was 163 compared with the observed number of 215. For NK180, 99 sorghum panicles were detected, while the observed was 114. Overlapping sorghum panicles and the misclassification between white soil background and sorghum panicles caused the overestimation and underestimation with 500 training images.

For the case of 1,000 training images, sorghum panicles were overestimated for variety XM217 (168 predicted vs. 175 observed). For KS585, 199 sorghum panicles were predicted compared to 215 observed. For NK180, 106 panicles were predicted compared to 114 observed. The errors in these cases were mainly caused by overlapping sorghum panicles. For some images with bright soil background and leaves, the algorithm could not perfectly separate sorghum panicles from surroundings, which led to the underestimation errors. This situation was more widespread with a small number of training images, especially for sorghum varieties KS585 and NK180, which had bright panicles similar to the soil and shiny leaves.

DISCUSSION

Previous studies on sorghum panicle detection and counting used points or rectangular bounding boxes to label sorghum panicles for preparing the training datasets and outputting the predicted results (Guo et al., 2018; Ghosal et al., 2019; Oh et al., 2019). For example, Wu et al. (2019) combined the image segmentation technique and basic CNN algorithm to create a density map of sorghum panicles. The application of image segmentation can exclude the areas that are not directly involved in training dataset preparation and final output. Malambo et al. (2019) applied a semantic segmentation-based CNN algorithm to separate sorghum panicles from the soil and other parts in images. These machine learning algorithms for sorghum panicle detection were mainly based on image classification. The use of points or bounding boxes does not provide direct information about the sorghum panicle shape and size. Compared to previous similar studies, the U-Net CNN segmentation adopted in this study not only detect but also localize and delineate individual sorghum panicles. Therefore, the use of sorghum panicle masks and deep learning from this

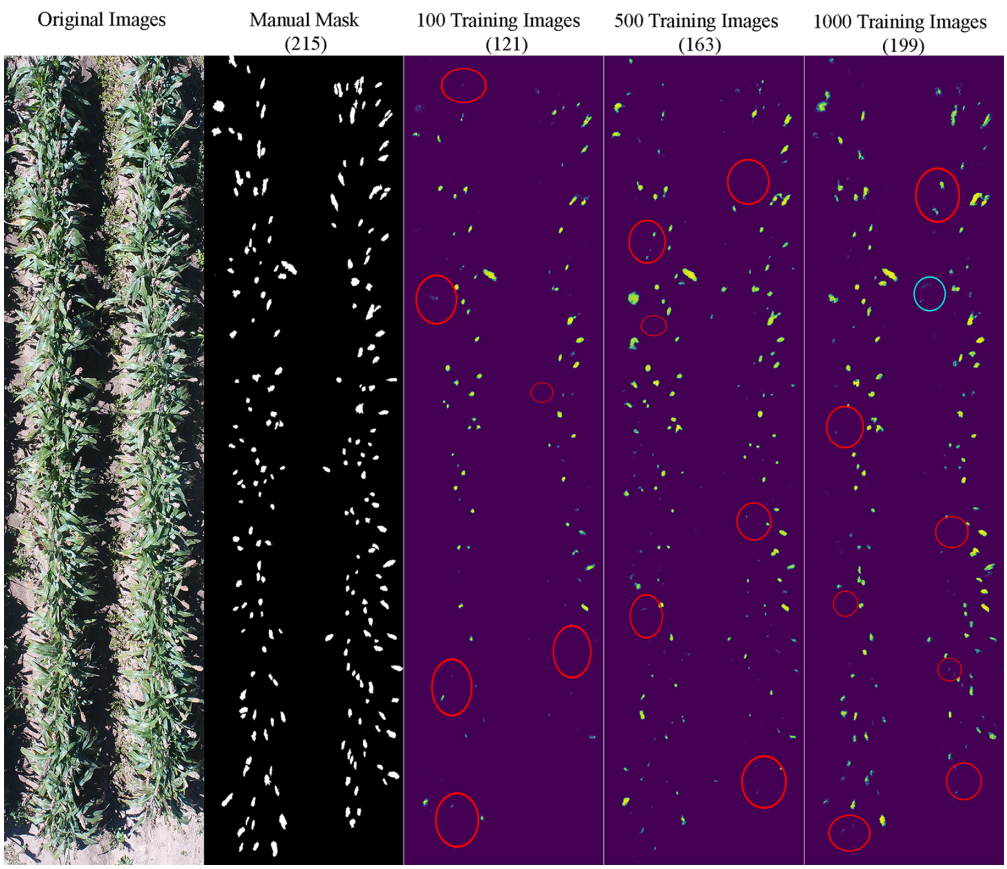

FIGURE 6 | Sample results of sorghum panicle detection for Variety KS585 with 100, 500, and 1000 training images using a Convolutional Neural Networks model on UAS images. Red circles represent underestimation; blue circles represent overestimation compared to the manual masks.

study enables the characterization of individual sorghum panicles, leading to more accurate yield estimation. This, however, does not mean sorghum yield can be directly calculated from the images because the sorghum panicles are typically not orthogonal to the UAS sensor during image acquisition. Further research is required to more accurately determine the size and shape of the sorghum panicle if yield prediction is needed.

By using masks, our algorithm also minimized the errors in sorghum panicle detection due to panicle overlaps and mixing with other elements in the image. Agricultural images acquired using UAS typically have a mixture of target items and background elements due to the deformation caused by camera angel and other factors (Kamilaris and Prenafeta-Boldú, 2018; Pradeep et al., 2018). This makes object detection in computer vision tasks challenging, especially for the multiple overlapping panicles and panicles that are obscured partially by plant leaves (Guo et al., 2018). Chen et al. (2017a) used the DeepLabv3+ to detect object boundaries, with a high accuracy using 11,530 high quality pixel-level annotated images. However, this proposed algorithm was only able to separate the object boundaries between two different classes, but could not detect overlapping sorghum panicles described in this study. Similar methods all required a large number of well labeled training images. Compared with these methods, our algorithm was able to separate and count sorghum panicles individually. For example, **Figure 8** shows the overlapping panicles situation and the prediction results using training images and corresponding masks. As shown, the algorithm was able to detect overlapping sorghums by providing masks that mark overlapping panicles. This algorithm, however, could not detect all overlapping panicles due to the lack of training masks in such cases. We believe the performance in detecting overlapping panicles can be improved by increasing the number of overlapping training images.

One of the limitations encountered in this study was the split of a full image into pieces for counting sorghum panicles due to computation restrictions. Previous studies have also shown such challenges in machine learning and deep learning algorithms to directly process high-resolution images. It is common to crop or split the original large dimension images to smaller images for detecting and counting objects (Aich et al., 2018; Wu et al., 2019;

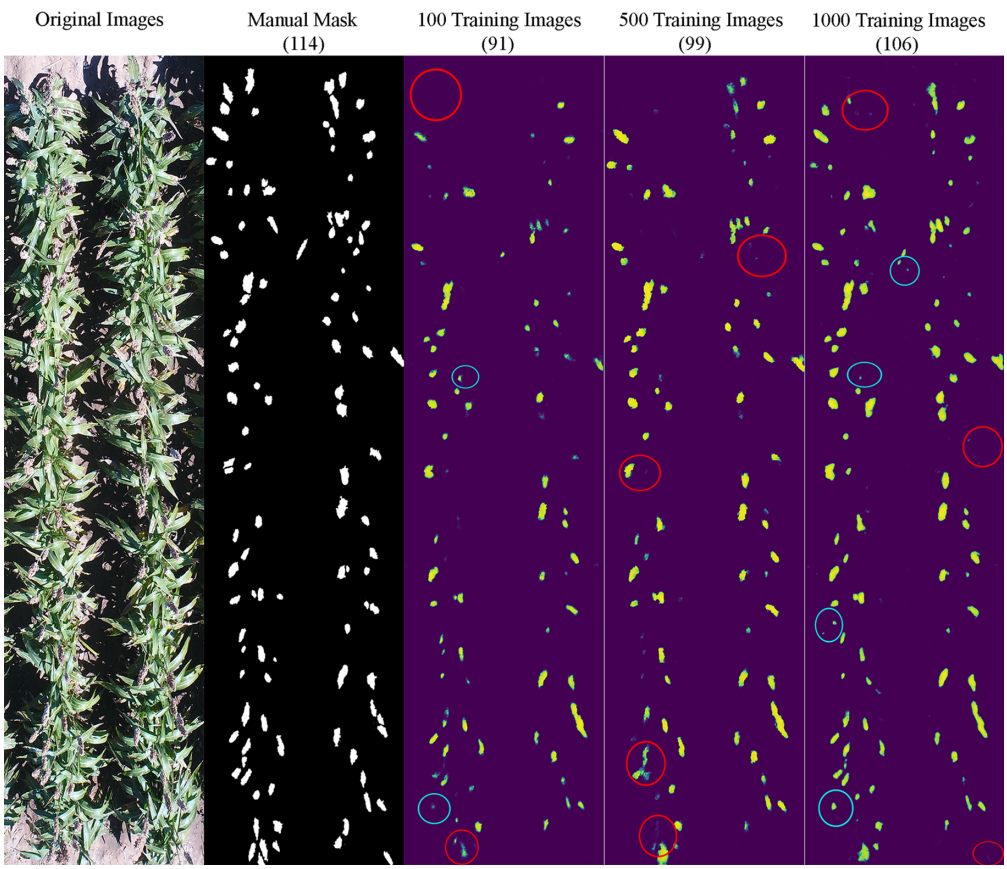

FIGURE 7 | Sample results of sorghum panicle detection for Variety NK180 with 100, 500, and 1000 training images using a Convolutional Neural Networks model on UAS images. Red circles represent underestimation and blue circles represent overestimation compared to the manual masks.

Chandra et al., 2020). This potentially leads to overestimation. In our study, some sorghum panicles were cut into two parts and counted twice because we horizontally split the test image into four sub-images. However, our visual check indicated that most of panicles were not split evenly, resulting in the smaller pieces with less than six pixels not being counted. Therefore, the double counting issue had no significant effect on the accuracy of the algorithm. Future studies are required to address this limitation by adopting more efficient image processing algorithms to avoid potential double counting.

In this study, sorghum panicles with greater contrast in color and brightness with surrounding elements were easily detected and counted, while some other panicles, especially for the variety KS585, were challenging to detect due to their similarity to the surrounding features, including soil and dry leaves. Environmental factors, such as wind and clouds, have a significant impact on UAS image quality, which can affect the performance of deep learning algorithms. Field condition and plant genotypes also affect the accuracy of machine learning tasks (Torres-Sánchez et al., 2013; Rasmussen et al., 2019). The similar colors between soil background and crops could also cause errors in computer vision tasks (El-Faki et al., 2000; Lee et al., 2018). In this study, it appears some soil clusters and leaves were mislabeled as sorghum panicles, probably due to strong sunlight conditions. We acquired most of the images around local noon time. As a result, both the soil surface and some sorghum panicles were relatively bright in full sunlight. In future studies, users may consider acquiring UAS images under relatively soft light environments, such as late afternoon or early morning. For the improvement of the algorithm performance, adding a separate mask for soil pixels can be an effective alternative to separate the sorghum panicles from soil background.

Abundant training datasets are critical for effective deep learning tasks (Deng et al., 2014; LeCun et al., 2015), especially for complex computer vision tasks such as sorghum panicle detection and counting. This study provides useful information regarding the number of training images required for such deep learning tasks. The algorithm produces inconsistent predictions and low accuracy with below 500 training images. It is reasonably accurate with 1,000 training images. It is expected that with more training images, the accuracy and robustness can be further improved. Aggarwal et al. (2019) demonstrated that a large number of training images could improve the

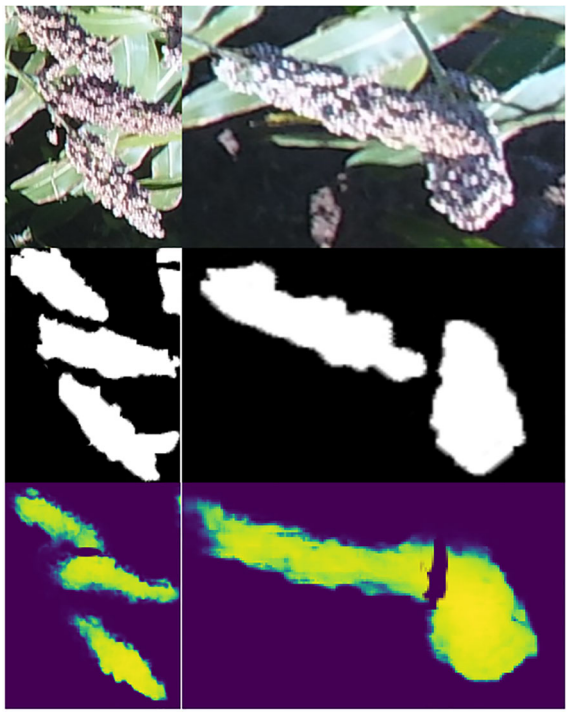

FIGURE 8 | Sample images showing the minimization of errors in sorghum panicle detection due to overlaps using masks and deep learning. Upper images are raw images. Middle images are manually training masks. Bottom ones are predicted masks.

performance of the U-Net CNN model, especially in complex models. However, there are not enough public ready-to-use data as training datasets for specific crops and their phenotypic traits. The development of large training datasets for plant phenotyping is time-consuming and labor-intensive. The drawback of the pre-label based algorithm lies in the fact that it is time consuming to prepare these training masks of sorghum panicles. In this study, 1,000 training images and masks were manually prepared and applied to develop the algorithm. It took a considerably longer time to prepare the training datasets compared with previous studies that used dot-labeled training images. The automatic annotation technique has shown its potential in similar algorithms (Zahavy et al., 2016; Komura and Ishikawa, 2018; Ghosal et al., 2019). Predicted outputs from automatic annotation can be used as new training input, which reduces the workload of manual preparation and can improve the efficiency and the robustness of the algorithm.

CONCLUSIONS

In this study, we developed an algorithm to integrate deep learning and segmentation to detect and count sorghum panicles using high-resolution UAS images. A dataset of 1,000 randomly selected images and their corresponding manually labeled masks were constructed for training this algorithm. The performance and efficacy of the algorithm were assessed with a different number of subset training images. The performance of the algorithm improved with the number of training images. The performance of the algorithm was not stable with less than 500 training images. With 1,000 training images, the algorithm had the best performance, with an accuracy of 95.5% and an RMSE of 2.5. The algorithm is sufficiently accurate for varying orientations and sizes of three sorghum varieties. Therefore, future studies are required to test the robustness of our algorithm with other varieties. In addition, compared to previous similar studies, our algorithm integrated image segmentation and CNN deep learning, which not only detect but also localize and delineate individual sorghum panicles. The algorithm is also capable of detecting overlapping sorghum panicles. This offers an opportunity for enhanced sorghum breeding efficiency and accurate yield estimation. To achieve this, however, further research is needed to improve the algorithm to quantify panicle dimension in relation to yield.

The development of large training datasets for plant phenotyping is time consuming and labor intensive. Therefore, this study provides a benchmark for the requirement for the number of training images for such phenotyping tasks. On the other hand, a more effective method, such as automatic annotation, is needed to prepare reliable training images. The performance of this algorithm was evaluated at the small-plot scale. Further studies are required to expand this algorithm to detect and count sorghum panicles at the commercial field scale. In addition, sorghum panicle detection accuracy as influenced by environmental factors, including image resolution, soil background, and illumination levels, requires further evaluation.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization: WG and ZL. Methodology: ZL and WG. Software: ZL. Validation: ZL and WG. Formal analysis: ZL and WG. Investigation: ZL and WG. Resources: WG. Data curation: ZL. Writing—original draft preparation: ZL. Writing—review and editing: ZL and WG. Visualization: ZL and WG. Supervision: WG. Project administration: WG. Funding acquisition: WG.

ACKNOWLEDGMENTS

We would like to acknowledge the Texas Tech University High Performance Computing Center for support of the study.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. Available at: www.tensorflow.org (Accessed November 20, 2019).

Aggarwal, H. K., Mani, M. P., and Jacob, M. (2019). MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Trans. Med. Imaging* 38, 394–405. doi: 10.1109/TMI.2018.2865356

Aich, S., Josuttes, A., Ovsyannikov, I., Strueby, K., Ahmed, I., Duddu, H. S., et al. (2018). "DeepWheat: Estimating Phenotypic Traits from Crop Images with Deep Learning," *2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, 323–332. doi: 10.1109/WACV.2018.00042

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). *Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation*. Available at: http://arxiv.org/abs/1802.06955 (Accessed April 27, 2020). arXiv Prepr. arXiv1802.06955.

Ballester, C., Zarco-Tejada, P. J., Nicolás, E., Alarcón, J. J., Fereres, E., Intrigliolo, D. S., et al. (2018). Evaluating the performance of xanthophyll, chlorophyll and structure-sensitive spectral indices to detect water stress in five fruit tree species. *Precis. Agric.* 19, 178–193. doi: 10.1007/s11119-017-9512-y

Barbedo, J. G. A. (2014). Using digital image processing for counting whiteflies on soybean leaves. *J. Asia Pac. Entomol.* 17, 685–694. doi: 10.1016/J.ASPEN.2014.06.014

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* 39, 79–87. doi: 10.1016/J.JAG.2015.02.012

Bradski, G. (2000). The OpenCV Library. *Dr Dobbs J. Software Tools* 25, 120–125. doi: 10.1111/0023-8333.50.s1.10

Buja, A., Stuetzle, W., and Shen, Y. (2005).Loss functions for binary class probability estimation and classification: structure and application (Accessed April 27, 2020).

Chandra, A. L., Desai, S. V., Balasubramanian, V. N., Ninomiya, S., and Guo, W. (2020). Active learning with point supervision for cost-effective panicle detection in cereal crops. *Plant Methods* 16, 1–16. doi: 10.1186/s13007-020-00575-8

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017a). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 834–848. doi: 10.1109/TPAMI.2017.2699184

Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., et al. (2017b). Counting Apples and Oranges with Deep Learning: A Data-Driven Approach. *IEEE Robot. Autom. Lett.* 2, 781–788. doi: 10.1109/LRA.2017.2651944

Chollet, F. (2015). *Keras: Deep learning library for theano and tensorflow*. (GitHub).

Dahlberg, J., Hutmacher, B., and Wright, S. (2015).Sorghum: an alternative feed, hay and forage. In: *Proceedings, 2015 Weastern Alfalfa & Forage Symposium, Reno, NV*. Available at: https://alfalfa.ucdavis.edu/+symposium/2015/PDFfiles/DahlbergJeff.pdf (Accessed October 10, 2019).

Deng, L., Yu, D., Deng, L., and Yu, D. (2014). the essence of knowledge Deep Learning Methods and Applications Foundations and Trends ® in Signal Processing Deep Learning Methods and Applications Deep Learning: Methods and Applications. *Found. Trends R Signal Process* 7, 197–387. doi: 10.1561/2000000039

Díaz-Varela, R. A., de la Rosa, R., León, L., and Zarco-Tejada, P. J. (2015). High-resolution airborne UAV imagery to assess olive tree crown parameters using 3D photo reconstruction: Application in breeding trials. *Remote Sens.* 7, 4213–4232. doi: 10.3390/rs70404213

Dyrmann, M., Karstoft, H., and Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. doi: 10.1016/j.biosystemseng.2016.08.024

El-Faki, M. S., Zhang, N., and Peterson, D. E. (2000). Factors affecting color-based weed detection. *Trans. Am. Soc Agric. Eng.* 43, 1001–1009. doi: 10.13031/2013.2968

FAO (1999). *SORGHUM: Post-harvest Operations*. Available at: http://www.fao.org/in-action/inpho/crop-compendium/cereals-grains/en/ (Accessed October 10, 2019).

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 321, 321–331. doi: 10.1016/j.neucom.2018.09.013

Gago, J., Douthe, C., Coopman, R. E., Gallego, P. P., Ribas-Carbo, M., Flexas, J., et al. (2015). UAVs challenge to assess water stress for sustainable agriculture. *Agric. Water Manage.* 153, 9–19. doi: 10.1016/j.agwat.2015.01.020

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., et al. (2019). A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting. *Plant Phenomics* 2019, 1–14. doi: 10.34133/2019/1525874

Gnädinger, F., and Schmidhalter, U. (2017). Digital counts of maize plants by Unmanned Aerial Vehicles (UAVs). *Remote Sens.* 9, 544. doi: 10.3390/rs9060544

Gómez-Candón, D., De Castro, A.II, and López-Granados, F. (2014). Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. *Precis. Agric.* 15, 44–56. doi: 10.1007/s11119-013-9335-4

Gómez-Candón, D., Virlet, N., Labbé, S., Jolivot, A., and Regnard, J. L. (2016). Field phenotyping of water stress at tree scale by UAV-sensed imagery: new insights for thermal acquisition and calibration. *Precis. Agric.* 17, 786–800. doi: 10.1007/s11119-016-9449-6

Guan, S., Khan, A. A., Sikdar, S., and Chitnis, P. V. (2019). Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal. *IEEE J. Biomed.Health Informatics* 24, 568–576. doi: 10.1109/JBHI.2019.2912935

Guo, W., Zheng, B., Potgieter, A. B., Diot, J., Watanabe, K., Noshita, K., et al. (2018). Aerial imagery analysis – Quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Front. Plant Sci.* 9, 1544. doi: 10.3389/fpls.2018.01544

Hasan, M. M., Chopin, J. P., Laga, H., and Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods* 14, 100. doi: 10.1186/s13007-018-0366-8

Hunt, E. R., Cavigelli, M., Daughtry, C. S. T., McMurtrey, J. E., and Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precis. Agric.* 6, 359–378. doi: 10.1007/s11119-005-2324-5

Hunt, E. R., Hively, W., Fujikawa, S. J., Linden, D. S., Daughtry, C. S. T., and McCarty, G. W. (2010). Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. *Remote Sens.* 2, 290–305. doi: 10.3390/rs2010290

Islam, M., Dinh, A., Wahid, K., and Bhowmik, P. (2017). "Detection of potato diseases using image segmentation and multiclass support vector machine," *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. (Windsor, ON), 1–4. doi: 10.1109/CCECE.2017.7946594

Jin, X., Liu, S., Baret, F., Hemerlé, M., and Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sens. Environ.* 198, 105–114. doi: 10.1016/j.rse.2017.06.007

Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016

Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78. doi: 10.1016/j.media.2016.10.004

Kingma, D. P., and Ba, J. L. (2015). *Adam: A method for stochastic optimization. in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (International Conference on Learning Representations, ICLR). Available at: https://arxiv.org/abs/1412.6980v9 (Accessed July 2, 2020).

Komura, D., and Ishikawa, S. (2018). Machine Learning Methods for Histopathological Image Analysis. *Comput. Struct. Biotechnol. J.* 16, 34–42. doi: 10.1016/j.csbj.2018.01.001

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* 14, 778–782. doi: 10.1109/LGRS.2017.2681128

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lee, U., Chang, S., Putra, G. A., Kim, H., and Kim, D. H. (2018). An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image analysis. *PloS One* 13, e0196615. doi: 10.1371/journal.pone.0196615

Liu, S., Baret, F., Andrieu, B., Burger, P., and Hemmerlé, M. (2017). Estimation of wheat plant density at early stages using high resolution imagery. *Front. Plant Sci.* 8, 739. doi: 10.3389/fpls.2017.00739

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. Forest Meteorol.* 264, 225–234. doi: 10.1016/j.agrformet.2018.10.013

Makantasis, K., Karantzalos, K., Doulamis, A., and Doulamis, N. (2015). Deep supervised learning for hyperspectral data classification through convolutional neural networks. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. (Milan), 4959–4962. doi: 10.1109/IGARSS.2015.7326945

Malambo, L., Popescu, S., Ku, N. W., Rooney, W., Zhou, T., and Moore, S. (2019). A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. *Remote Sens.* 11, 2939. doi: 10.3390/rs11242939

Martins, F. F., and Zaglia, M. C. (2019). *Application of convolutional neural network to pixel-wise classification in deforestation detection using PRODES data. GEOINFO, 20 Years After!, 57*. Available at: http://www.obt.inpe.br/OBT/assuntos/programa/amazonia/prodes (Accessed April 30, 2020).

Mikołajczyk, A., and Grochowski, M. (2018). "Data augmentation for improving deep learning in image classification problem," *2018 International Interdisciplinary PhD Workshop (IIPhDW)*. (Swinoujście), 117–122. doi: 10.1109/IIPHDW.2018.8388338

Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., et al. (2018). Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective. *Gigascience* 8, 1–12. doi: 10.1093/gigascience/giy153

Mohri, M., Afshin, R., and Ameet, T. (2018). *Foundations of Machine Learning*. (MIT Press). doi: 10.2139/ssrn.3399990

Norman, D. W., Worman, F. D., Siebert, J. D., and Modiakgotla, E. (1995). *The farming systems approach to development and appropriate technology generation* (Food and Agriculture Organization of the United Nations). Available at: http://www.fao.org/3/v5330e/V5330e00.htm#Contents (Accessed October 10, 2019).

Oh, M., Olsen, P., and Ramamurthy, K. N. (2019). *Counting and Segmenting Sorghum Heads*. Available at: https://arxiv.org/abs/1905.13291 (Accessed October 10, 2019). arxiv.org.

Olsen, P. A., Natesan Ramamurthy, K., Ribera, J., Chen, Y., Thompson, A. M., Luss, R., et al. (2018). "Detecting and counting panicles in sorghum images," in *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018* (Turin, Italy: IEEE), 400–409. doi: 10.1109/DSAA.2018.00052

Pape, J.-M., and Klukas, C. (2015). Utilizing machine learning approaches to improve the prediction of leaf counts and individual leaf segmentation of rosette plant images. in. *bmva.org* 3, 1–3.12. doi: 10.5244/c.29.cvppp.3

Pérez-Ortiz, M., Peña, J. M., Gutiérrez, P. A., Torres-Sánchez, J., Hervás-Martínez, C., and López-Granados, F. (2015). A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Appl. Soft Comput. J.* 37, 533–544. doi: 10.1016/j.asoc.2015.08.027

Pradeep, P., Park, S. G., and Wei, P. (2018). "Trajectory optimization of multirotor agricultural UAVs," in *IEEE Aerospace Conference Proceedings* (Big Sky, MT: IEEE Computer Society), 1–7. doi: 10.1109/AERO.2018.8396617

Python Software Foundation (1995). Python tutorial. *CWI Rep.*, 1–65. http://oai.cwi.nl/oai/asset/5007/05007D.pdf. CS-R9526.

Qureshi, W. S., Payne, A., Walsh, K. B., Linker, R., Cohen, O., and Dailey, M. N. (2017). Machine vision for counting fruit on mango tree canopies. *Precis. Agric.* 18, 224–244. doi: 10.1007/s11119-016-9458-5

Rasmussen, J., Nielsen, J., Streibig, J. C., Jensen, J. E., Pedersen, K. S., and Olsen, S.II (2019). Pre-harvest weed mapping of Cirsium arvense in wheat and barley with off-the-shelf UAVs. *Precis. Agric.* 20, 983–999. doi: 10.1007/s11119-018-09625-7

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. (Cham: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Sankaran, S., Khot, L. R., and Carter, A. H. (2015). Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Comput. Electron. Agric.* 118, 372–379. doi: 10.1016/j.compag.2015.09.001

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Comput. Intell. Neurosci.* 2016. doi: 10.1155/2016/3289801

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M. (2017). "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. (Cham: Springer), 240–248. doi: 10.1007/978-3-319-67558-9_28

Tokekar, P., Hook, J., Mulla, D., and Isler, V. (2016). Sensor Planning for a Symbiotic UAV and UGV System for Precision Agriculture. *IEEE Trans. Robot.* 32, 1498–1511. doi: 10.1109/TRO.2016.2603528

Too, E. C., Yujian, L., Njuki, S., and Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161, 272–279. doi: 10.1016/j.compag.2018.03.032

Torres-Sánchez, J., López-Granados, F., De Castro, A.II, and Peña-Barragán, J. M. (2013). Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management. *PloS One* 8, e58210. doi: 10.1371/journal.pone.0058210

Ubbens, J., Cieslak, M., Prusinkiewicz, P., and Stavness, I. (2018). The use of plant models in deep learning: An application to leaf counting in rosette plants. *Plant Methods* 14, 6. doi: 10.1186/s13007-018-0273-z

US Climate Data (2019). *Weather history Lubbock* (Texas: U.S. Clim. Data). Available at: https://www.usclimatedata.com/climate/lubbock/texas/united-states/ustx2745/2014/1 (Accessed February 25, 2019).

Verger, A., Vigneau, N., Chéron, C., Gilliot, J. M., Comar, A., and Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens. Environ.* 152, 654–664. doi: 10.1016/j.rse.2014.06.006

Wang, J., and Perez, L. (2017).The Effectiveness of Data Augmentation in Image Classification using Deep Learning. Available at: http://arxiv.org/abs/1712.04621 (Accessed April 29, 2020). arXiv Prepr. arXiv1712.04621

Weng, Y., Zhou, T., Li, Y., and Qiu, X. (2019). NAS-Unet: Neural architecture search for medical image segmentation. *IEEE Access*. 7, 44247–44257. doi: 10.1109/ACCESS.2019.2908991

Wu, J., Yang, G., Yang, X., Xu, B., Han, L., and Zhu, Y. (2019). Automatic Counting of in situ Rice Seedlings from UAV Images Based on a Deep Fully Convolutional Neural Network. *Remote Sens.* 11, 691. doi: 10.3390/rs11060691

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., and Shen, C. (2019). TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* 15, 150. doi: 10.1186/s13007-019-0537-2

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* 8, 1111. doi: 10.3389/fpls.2017.01111

Zahavy, T., Sivak, A., Kang, B., Feng, J., Xu, H., and Mannor, S. (2016). "Ensemble robustness and generalization of stochastic deep learning algorithms," *arXiv preprint arXiv:1602.02389*.

Zhang, Z., and Sabuncu, M. R. (2018). "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Advances in Neural Information Processing Systems*, 8778–8788.

Zhang, Y., Zhou, J., Meng, L., Li, M., Ding, L., and Ma, J. (2018). "A method for deriving plant temperature from UAV TIR Image," *2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics)*. (Hangzhou), 1–5. doi: 10.1109/Agro-Geoinformatics.2018.8475995

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lin and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 November 2020
doi: 10.3389/fpls.2020.575810

Wheat Kernel Variety Identification Based on a Large Near-Infrared Spectral Dataset and a Novel Deep Learning-Based Feature Selection Method

Lei Zhou[1,2], Chu Zhang[1,2], Mohamed Farag Taha[1,2], Xinhua Wei[3,4], Yong He[1,2], Zhengjun Qiu[1,2] and Yufei Liu[1,2]**

[1] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2] Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou, China, [3] Synergistic Innovation Center of Jiangsu Modern Agricultural Equipment and Technology, Zhenjiang, China, [4] School of Agricultural Equipment Engineering, Jiangsu University, Zhenjiang, China

OPEN ACCESS

Edited by:
Spyros Fountas,
Agricultural University of Athens,
Greece

Reviewed by:
Song Li,
Virginia Tech, United States
Robert Winkler,
National Polytechnic Institute
of Mexico (CINVESTAV), Mexico

***Correspondence:**
Zhengjun Qiu
zjqiu@zju.edu.cn
Yufei Liu
yufeiliu@zju.edu.cn

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: 24 June 2020
Accepted: 09 October 2020
Published: 10 November 2020

Citation:
Zhou L, Zhang C, Taha MF, Wei X, He Y, Qiu Z and Liu Y (2020) Wheat Kernel Variety Identification Based on a Large Near-Infrared Spectral Dataset and a Novel Deep Learning-Based Feature Selection Method. Front. Plant Sci. 11:575810. doi: 10.3389/fpls.2020.575810

Near-infrared (NIR) hyperspectroscopy becomes an emerging nondestructive sensing technology for inspection of crop seeds. A large spectral dataset of more than 140,000 wheat kernels in 30 varieties was prepared for classification. Feature selection is a critical segment in large spectral data analysis. A novel convolutional neural network-based feature selector (CNN-FS) was proposed to screen out deeply target-related spectral channels. A convolutional neural network with attention (CNN-ATT) framework was designed for one-dimension data classification. Popular machine learning models including support vector machine (SVM) and partial least square discrimination analysis were used as the benchmark classifiers. Features selected by conventional feature selection algorithms were considered for comparison. Results showed that the designed CNN-ATT produced a higher performance than the compared classifier. The proposed CNN-FS found a subset of features, which made a better representation of raw dataset than conventional selectors did. The CNN-ATT achieved an accuracy of 93.01% using the full spectra and keep its high precision (90.20%) by training on the 60-channel features obtained via the CNN-FS method. The proposed methods have great potential for handling the analyzing tasks on other large spectral datasets. The proposed feature selection structure can be extended to design other new model-based selectors. The combination of NIR hyperspectroscopic technology and the proposed models has great potential for automatic nondestructive classification of single wheat kernels.

Keywords: NIR hyperspectroscopy, wheat kernel classification, feature selection, convolutional neural network, attention mechanism

INTRODUCTION

Wheat is one of the most important agricultural products. Various varieties of wheat are cultivated to adapt to different planting environments and to improve the yield and quality. Different varieties of wheat kernels have different characters and values. The purity of wheat kernels is of great concern by breeders, planters, and consumers (Ebrahimi et al., 2014). Wheat kernels of different variety

share some similar characteristics, which makes it difficult to distinguish with the naked eye. Mass spectrometry-based methods have been widely accepted for inspection of wheat quality owing to their high sensitivity (Koistinen and Hanhineva, 2017). However, they are destructive methods, and expensive instrument is required. Recent advances in machine learning, computer vision, and spectroscopic sensing promote a series of nondestructive testing techniques for crop seeds evaluation (Sabanci et al., 2017; Ding et al., 2019; Xia et al., 2019; Fabiyi et al., 2020). Hyperspectral imaging (HSI) is an emerging tool with the advantages of collecting spectral and spatial information simultaneously. The obtained data is in the shape of a hypercube (width * height * number of channels). Each spatial pixel is a spectral vector. It allows a user to collect data of many samples by scanning a single HSI image. Therefore, it is very suitable for analyzing large quantities of crop kernels (Feng et al., 2019).

The digital information collected by hyperspectroscopic instruments are always in large volume. It contains a lot of redundant information and causes troubles for data analysis. Feature selection becomes a critical procedure in the pretreatment of high-dimensional spectral data (Lin et al., 2012; Zhang et al., 2016). Feature selection methods are expected to improve the performance of classification/regression models by screening out a subset of informative features and to accelerate the model training procedure as well. Another way for compressing the volume of a dataset is feature extraction.

Research efforts are attracted in the field of feature selection in spectral data processing (Balabin and Smirnov, 2011; Lei and Sun, 2020). Some methods aim at selecting the most important subset by univariate statistical tests, such as univariate feature selection (UFS) (Emura et al., 2019). This kind of methods operate with a high efficiency. However, as a fact, the feature selectors based on checking the amount of information in each channel only use the features, excluding corresponding labels. It is possible to screen out the features, which are not highly related to the final target. Similar problems were discussed in an article about feature extraction (Yuan et al., 2018). Some methods search the important features by training a linear machine learning model. The coefficients of the trained models are considered as the importance scores of each feature. A typical example is support vector machine (SVM) feature selector (SVM-FS) (Khaled et al., 2018; Pes, 2019). The model-based feature selectors consider the target labels. However, the quality of the selected features depends on the capacity of the machine learning model. It is hard to find a subset of effective features by a low-performance model. The model-based feature selection methods have great potential to screen out informative and output-related features by improving the used model. There are also several tree-based feature selection approaches, such as extra-trees classifier (ETC) (Geurts et al., 2006; Krol and Polanska, 2017) and random forest (RF) (Khaled et al., 2018).

Concerning large dataset processing, there are some limitations in traditional modeling methods. Deep learning algorithms become emerging tools to solve the complex modeling tasks. Different deep architectures consisting of nonlinear processing units have been introduced for seed variety identification based on spectral datasets. Ozkan et al. (2019) designed a convolutional neural network (CNN) for variety discrimination of wheat grain. Spectral images of the grains were collected by multiple spectroscopic sensors to make a dataset. It was found that the CNN method could identify the category of the grains based on a spectral image with hundreds of grains. The recognition of each individual grain was not involved. Several researchers investigated the application of deep learning algorithms on kernel-level analysis of crop seeds. Qiu et al. (2018) proposed a CNN classifier for rice seed variety identification using NIR spectroscopy. The CNN model showed its superiority in the classification task, achieving a higher precision than the compared K nearest neighbors (KNN) and SVM. Zhu et al. (2019b) applied CNN methods to discriminate three varieties of soybean seeds by processing NIR spectral data. A satisfactory result was achieved.

According to the surveyed articles above, deep learning classifiers performed better than conventional ones in spectra pattern recognition. In general, a deep learning model is a combination of linear/nonlinear data processing layers with different operation rules. Different structures of the layers or the whole networks can be custom defined to implement different applications. The existing deep learning applications involve classification, regression, feature extraction, objective detection, and so on (Zhang et al., 2018). The application of CNN architecture was also extended to variable selection (Liu et al., 2019), which calculated the importance score according to weights of the first convolutional layer. Very few of the abovementioned articles about seed classification used a very large dataset to evaluate the models. The deep learning methods can show its advantages more prominently in big datasets. In this study, a novel CNN-based feature selection algorithm was proposed for searching the informative spectral channels beneficial to the final classification problem. A large NIR spectral dataset of 30 varieties of wheat kernel (147,096 kernels in total) was prepared for experiment. A CNN model with attention mechanism was designed to process the selected features. Several well-known feature selection methods including UFS, SVM-FS, ETC, and classification methods including RBF-SVC, partial least squares discrimination analysis (PLSDA) were employed as benchmark methods. The selected features were visualized, and the classification accuracies were compared.

MATERIALS AND METHODS

Wheat Kernel Preparation

Thirty varieties of wheat kernels harvested in 2019 were collected from a local seed company in Shuyang, Jiangsu Province, China. The wheat kernels were stored under the same condition after harvest (dried, packaged by woven plastic bags, and delivered to the laboratory). The wheat plants were grown in the same field, and the kernels were harvested in the same year. Based on the time-sequence, the kernels were taken out of the package and sent for analysis, without any physical or chemical operation used on the kernels as preprocessing. Indeed, there were some differences among all the kernel in shape, weight, water content, and so on. A large number of samples of each variety were scanned

to build the dataset, which was expected to provide adequate knowledge for the deep learning models. The category number, wheat variety, and number of samples are listed in **Table 1**.

Near-Infrared Hyperspectral Image Scanning

Spectral images of the wheat kernels were collected by an NIR hyperspectral system with the spectral range from 874 to 1,734 nm. The main components included a spectrograph, a camera with lens, tungsten halogen light source, and a conveyer belt driven by a stepper motor for line scan. The detail information and the operation process of the sensing system were the same as the descriptions in Qiu et al. (2018). In this research, the moving speed of the conveyer belt, the exposure time of the camera, and the distance from the lens to the plate were set as 8.7 mm/s, 3 ms, and 20 cm, respectively.

The kernels were placed in a plate with grid for spectral image acquisition (see **Figure 1**). The plate was made of special materials that generate a very low reflectance in near-infrared range. A total of 147,096 wheat kernels were sampled. The samples of each variety were randomly separated into training, validation, and prediction set. In each variety, 2,400 kernels were used for calibration, 1,200 kernels were for validation, and the rest was for prediction. Details are listed in **Table 1**.

TABLE 1 | Overview of the dataset properties.

No.	Variety name	Abbreviation	Number of samples (Cal/Val/Pre*)
1	aikang58	AK58	2,400/1,200/1,054
2	bainong207	BN207	2,400/1,200/1,431
3	bainong4199	BN4199	2,400/1,200/1,332
4	bainong889	BN889	2,400/1,200/1,335
5	baomai218	BM218	2,400/1,200/1,259
6	baomai330	BM330	2,400/1,200/1,277
7	baomai5	BM5	2,400/12,00/1,355
8	fengdecunmai12	FDCM12	2,400/1,200/1,353
9	fengdecunmai20	FDCM20	2,400/1,200/1,360
10	guanmai1	GM1	2,400/1,200/1,361
11	huaimai20	HM20	2,400/1,200/1,283
12	huaimai40	HM40	2,400/1,200/1,273
13	huaimai41	HM41	2,400/1,200/1,280
14	huaimai920	HM920	2,400/1,200/1,134
15	jiangmai816	JM816	2,400/1,200/1,356
16	jiangmai919	JM919	2,400/1,200/1,356
17	jimai211	JM211	2,400/1,200/1,353
18	jimai22	JM22	2,400/1,200/1,241
19	lunxuan99	LX99	2,400/1,200/1,285
20	luomai9	LM9	2,400/1,200/1,357
21	ruihuamai518	RHM518	2,400/1,200/1,346
22	saidemai1	SDM1	2,400/1,200/1,354
23	tiechuaifu99	TCAF99	2,400/1,200/1,353
24	weilong169	WL169	2,400/1,200/1,204
25	xinong20	XN20	2,400/1,200/1,342
26	xinong979	XN979	2,400/1,200/1,350
27	xumai36	XM36	2,400/1,200/1,356
28	xumai818	XM818	2,400/1,200/1,354
29	yannong19	YN19	2,400/1,200/1,279
30	yunong035	YN035	2,400/1,200/1,123
Total			72,000/24,000/39,096

*Cal/Val/Pre means the calibration (Cal), validation (Val), and prediction (Pre) sets.

Spectra Extraction and Preprocessing

The collected spectral image should be corrected by black and white calibration first:

$$S_c = \frac{S_{raw} - S_{dark}}{S_{white} - S_{dark}} \quad (1)$$

The dark reference S_{dark} was obtained by covering the camera lens with opaque cap, and the white reference S_{white} was obtained by scanning a white Teflon tile. S_{raw} is the raw spectral data, and S_c is the dark-white calibrated data.

Due to the high-level noises at the start and end of the band range, only the range of 975–1,645 nm (200 bands in total) was selected in this research. The pseudo-image of the 20th band is shown in **Figure 2A**. An adaptive threshold segmentation method was used to remove the background and get the region of interest (ROI). The adaptive threshold was calculated by Otsu method (Otsu, 1979). Morphology erosion and dilation operations (a filter in disk shape with diameter of 5 pixels) were adopted for removing the small noisy pixels. The connected component labeling method was used to detect the region of each kernel in the image. Due to the use of the plate shown in **Figure 1**, the process of background removing and kernel region segmentation becomes very easy. Then, a square region with side length of 30 pixels for each kernel was extracted. The pseudo-image of ROI of the 20th band is shown in **Figure 2B**. Thus, the information of each kernel was recorded in a hypercube (width: 30 pixels, height: 30 pixels, channels: 200). Take one kernel for example; the pseudo-image of the 20th band is shown in **Figure 2C**. The average spectra of each individual hypercube was calculated and then processed by mean filter (window size of 5) and average normalization. The normalized spectra were used for representing the sample for further analysis. As an example, the normalized average spectra of the wheat kernel in **Figure 2C** is shown in **Figure 2D**. The pre-processed average spectra of all 30 different varieties of wheat kernels are shown in **Figure 3**.

Conventional Feature Selection Methods and Classifiers

Feature Selection Methods

Univariate feature selection (UFS) is a simple approach, which evaluates the importance of features by univariate statistical tests. For the data of classification purpose, Chi-square test was commonly used (Pes, 2019). Tree-based methods are also a choice for feature selection. These models calculate impurity-based importance. The importance value in the trained tree could be regarded as the importance score of each individual feature (Khaled et al., 2018). Linear estimators with sparse penalty term can be used as feature selectors. Due to the L1 norm term added in the loss function, some of the coefficients are fixed as 0. The features with nonzero coefficients are

FIGURE 1 | A batch of wheat kernels.

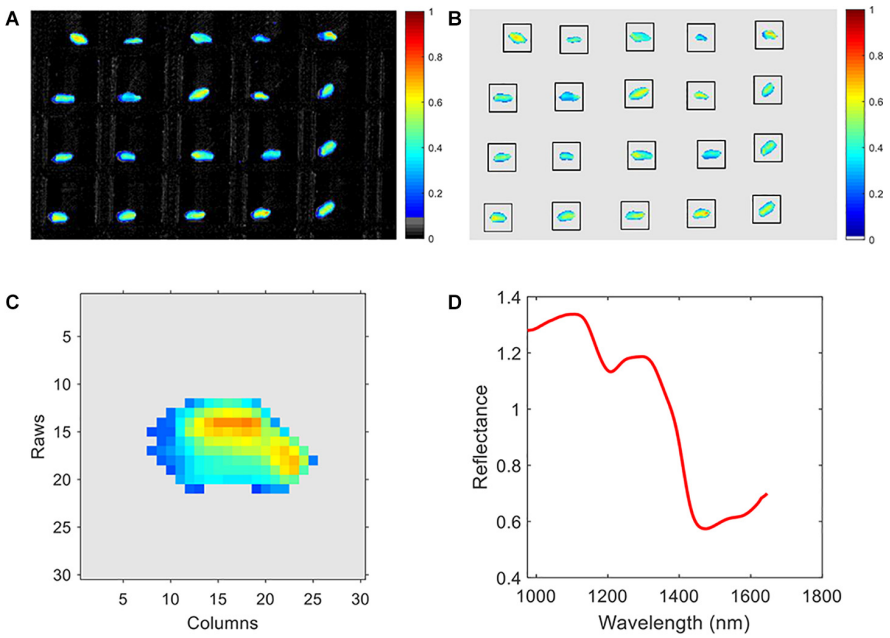

FIGURE 2 | Spectral data pre-processing. **(A)** Pseudo-image of the raw spectral image with background, **(B)** pseudo-image of the spectral image without background, **(C)** pseudo-image of an individual wheat kernel, and **(D)** the mean spectra of the wheat kernel in **(C)**.

considered as key features. As typical examples, the least absolute shrinkage and selection operator (LASSO) method (Kok et al., 2019; Chakraborty et al., 2020) is commonly used in regression problems. Logistic regression and linear SVM are adopted in classification problems (Pes, 2019). The selected features are further processed by another model. UFS with Chi-square test, ETC, and linear SVM were used in this study. These three methods were implemented by using scikit-learn[1], a very popular machine learning tool kit for Python.

Classifiers

Radial basis function support vector machine classifier (RBF-SVC) is a popular pattern recognition method (Burges, 1998). Due to its high capacity of handling nonlinearity, RBF-SVC has been widely used in spectral data classification (Jimenezcarvelo et al., 2019). In the RBF-SVC model, penalty coefficient C and the kernel parameter gamma should be optimized to realize high performances. The optimized value of C and gamma in RBF-SVC models were determined by grid search strategy and checking the validation accuracies. PLSDA is another widely accepted linear classification algorithm. It can be realized by modifying partial least squares regression (PLSR) method (Lee et al., 2018). In PLSDA model, the number of the principle component (comp) was optimized by checking the precision on the validation set. RBF-SVC and PLSDA, a nonlinear and a linear classifier, were selected as the benchmark classifiers in this research.

The Proposed Convolutional Neural Network Architectures for Feature Selection and Classification

CNN-based models have become very important machine intelligence algorithms for big data analysis. These models were popular in computer vision for RGB image processing

[1]https://scikit-learn.org/

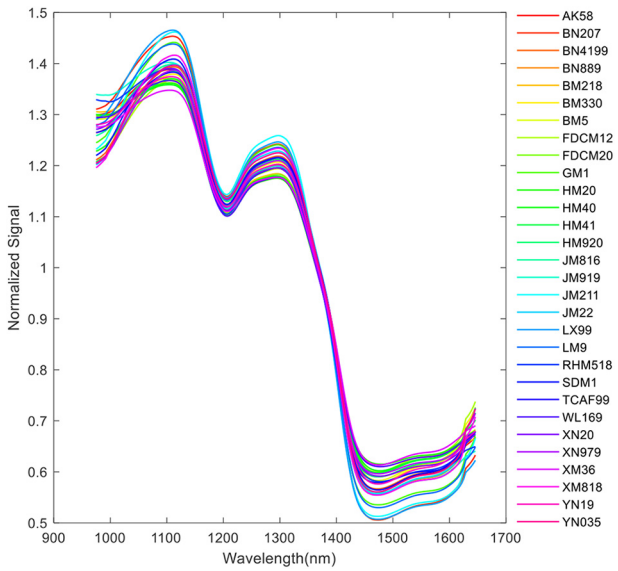

FIGURE 3 | Spectral profiles of the 30 varieties of wheat kernels.

(Ou et al., 2020). The applications of CNN were also extended to one-dimension data (such as pixel-level spectra) (Zhou et al., 2019) and three-dimension data (Zhang et al., 2018; Mehrkanoon, 2019). In this research, the features of the wheat kernel were in a shape of 200*1. A one-dimension CNN architecture with attention mechanism was presented for classification. Another CNN model with feature selection block was designed for optimal band selection. We define them as CNN-ATT and CNN-FS, respectively. The architectures of the designed networks are shown in **Figure 4**. The custom blocks were annotated by a dashed box. The CNN-ATT and CNN-FS shared a similar structure. First, the input data with a shape of N*Ch is processed by a custom block. N denotes batch size, and Ch is the number of feature channels. The output of the custom block was the weighted spectra vector. The weighted spectra vector should be reshaped to N*1*Ch for convolution operation. Then, the reshaped data was processed by three one-dimension convolution (1D Conv.) blocks (the number of the kernels, the kernel size, the strides were set as 16, 3, 1, respectively) and three dense layers (the number of neurons in three dense layers were 512, 128, 30, respectively). The last part was the SoftMax layer. The ELU activation (Qiu et al., 2018) was used in the networks (excluding the activation function in the two custom blocks). The MaxPooling layer (pool size and stride are 2, 2, respectively) batch normalization operation was used after each convolution layer.

CNN-FS

As for the CNN-FS, the custom block was defined as the selection block (SL block). Referring to the idea of LASSO algorithm (Lee and Cai, 2020), our aim was to calculate a vector of scores to indicate the importance of each channel on the target classification task. The weights of this block (WSL) were generated as the same shape as the input. A nonlinear activation function was used to process the WSL. The output of the SL block was defined as:

$$\begin{aligned} Y_{SL} &= f_A\left(W_{SL}\right) \cdot X \\ W_{SL} &= [W_1, W_2, ..., W_{Ch}], X = [X_1, X_2, ..., X_{Ch}] \end{aligned} \tag{2}$$

where · denotes multiplication of corresponding elements in the two matrixes. W_{SL} is the weights of the SL block, Y_{SL} is the weighted spectra, and f_A is the activation function. Define $\text{score}_{SL} = f_A\left(W_{SL}\right)$ as the importance scores. Here, some alternative activation functions could be SoftMax activation, Sigmoid, and ReLU. However, SoftMax makes the scores too small, which is unfavorable for model training. Although Sigmoid function can map the values to (0,1), it is also not suitable here. The target was to remove some non-informative channels and tune the scores of these channels to 0. If the Sigmoid was used, the Wi for a non-informative channel should be tuned close to negative infinity to get a score of 0 (or a value smaller than −4 to get a score close to 0). ReLU function could add some nonlinearity to the SL block and ensure the scores are all positive. Thus, ReLU was finally employed:

$$\text{ReLU}\left(a\right) = \begin{cases} a & a > 0 \\ 0 & a <= 0 \end{cases} \tag{3}$$

In this block, random initialization not suitable for the SL block. The WSL was initialized as [1,1,. . .,1], and the activated W_{SL} (or the initial scores) was also [1,1,. . .,1], which controls the initial importance scores all equal for each channel.

Then, the output of the SL block was reshaped and calculated by a one-dimension CNN classifier. The feature selection model was trained to minimize the classification error and keep the value of $f_A\left(W_{SL}\right)$ small. The loss function is defined as follows:

$$Loss\left(y, \tilde{y}\right) = \frac{1}{N}\sum_{i=1}^{S}\left(-\sum_{j=1}^{q} y_j^i \log\left(\tilde{y}_j^i\right)\right) + \lambda\sum_{j=1}^{q}\left(\text{ReLU}\left(\left|W_j\right|\right)\right) \tag{4}$$

The loss function consisted of two terms. The first part was cross entropy loss, which controls the classification precision. The second part was the sum of the activated weights in the SL block, which led the scores of unimportant features close to zero.

CNN-ATT

As for the CNN-ATT, the custom block was defined as attention block (ATT block). Attention mechanism enabled the network to pay more attention to some specific regions of the input data and weaken the focus on other regions. The attention scores were calculated by encoding the input via two dense layers:

$$Y_{ATT} = f_2\left(W_{ATT}^{(2)} \cdot f_1\left(W_{ATT}^{(1)} \cdot X + b_1\right) + b_2\right) \cdot X \tag{5}$$

where the symbol "·" denotes matrix multiplication, W_{ATT} is the weights of the ATT block, and Y_{ATT} is the weighted spectra by attention. Define $\text{score}_{ATT} = f_2\left(W_{ATT}^{(2)} \cdot f_1\left(W_{ATT}^{(1)} \cdot X + b_1\right) + b_2\right)$ as the attention scores.

The number of the neurons in the second dense layer should be equal to the dimension of the input, while that in the first dense

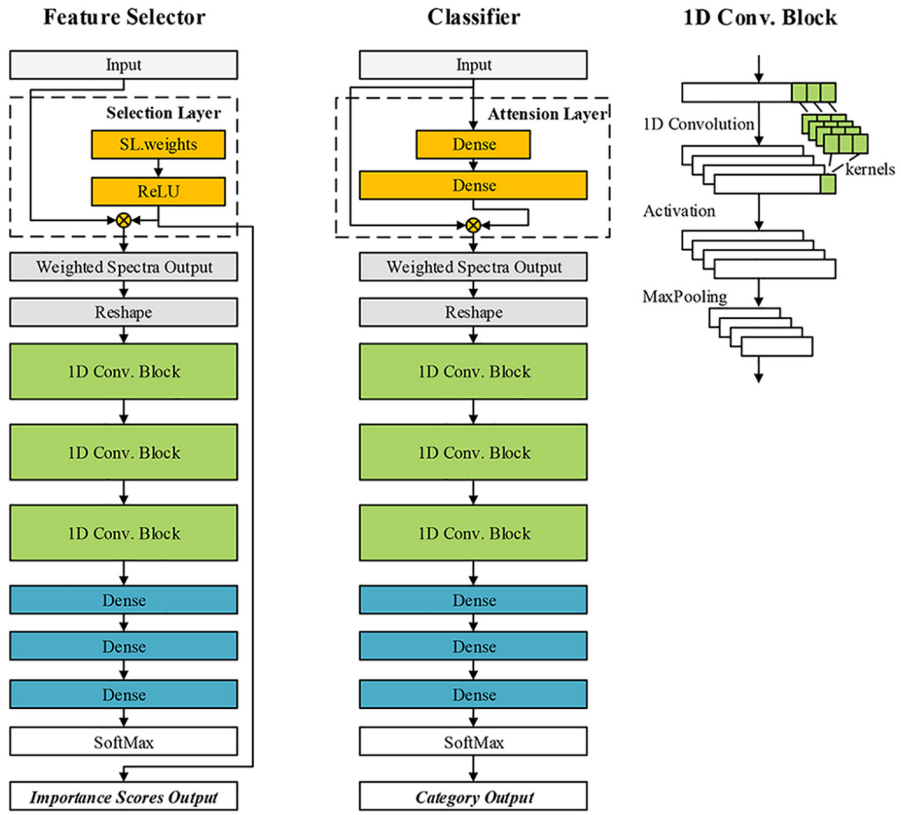

FIGURE 4 | The architectures of the proposed convolutional neural network (CNN)-based models for classification and feature selection.

layer could be set by developers. Although both the ATT block and SL block could generate a vector of scores, the functions of them differed. In the trained model, $score_{SL}$ is constant, while the value of $score_{ATT}$ varies with different X. The $score_{ATT}$ cannot adopted for feature selection, but it can be effective for building the classifier. The CNN-based models were programmed base on the MXNET framework[2].

RESULTS AND DISCUSSION

Training Procedures for the Convolutional Neural Network-Based Feature Selector and Convolutional Neural Network With Attention

In this study, the calibration dataset was used for training the models, the validation set was only for model adjustment, and the independent prediction set was for performance evaluation of the models. The methods for constructing conventional feature selectors and classifiers are described in the Conventional Feature Selection Methods and Classifiers section. The searching ranges were set as 10^1–10^{12} and 10^{-9}–10^2. The searching range for comp was set from 1 to the number of the input features. The optimized parameters of PLSDA models and RBF-SVC models are listed in **Table 2**. The result of grid-search (take the modeling based on full spectra as an example) is shown in **Figure 5**; the best combination of the parameters is marked by "*."

The CNN-ATT model was trained by minimizing the output of the cross-entropy loss using the Adam algorithm with a dynamic learning rate. The loss function used for CNN-ATT was the cross-entropy loss, while that for CNN-FS is defined in Eq. 4. The hyper parameters in CNN-based models were adjusted by repeated training for many times and observing the accuracy of the validation set. A group of parameters was found for obtaining a stable model. The number of epochs was set as 500. The value of λ in Eq. 4 was optimized as 0.1. We used a scheduled learning rate, the initial value was 0.0005, the period was defined as 100, and learning rate decay was set as 0.1. In other words, the learning rate was 0.0005 in the first 100 epochs, and it was decreased to 0.00005 in the next 100 epochs, and so on. At the start of the training, a relatively large learning rate was used to speed up training, and smaller values were used with the gradual convergence of the model. With the mentioned training procedure, the CNN models produced a low and stable loss value and a high classification accuracy on validation set.

For training the CNN-FS, the custom loss function defined in Eq. 4 was utilized. Other properties of the trainer were configured as the same as that for the mentioned CNN classifier. Each of the feature selectors provides a vector of importance scores.

[2]http://mxnet.incubator.apache.org/

TABLE 2 | The classification accuracies achieved by different feature selectors and classifiers.

Methods	Features selectors	Selected channels	Parameters*	Calibration	Validation	Prediction
Partial least squares discrimination analysis (PLSDA)	/	200	comp = 200	0.7008	0.6937	0.6880
	Convolutional neural network-based feature selector (CNN-FS)	60	comp = 60	0.5980	0.5895	0.5942
	Support vector machine-feature selector (SVM-FS)	60	comp = 60	0.5133	0.5040	0.5075
	Univariate feature selection (UFS)	60	comp = 60	0.4213	0.4183	0.4132
	Extra-trees classifier (ETC)	60	comp = 60	0.5531	0.5474	0.5488
Radial basis function support vector machine classifier (RBF-SVC)	/	200	$C = 10^7$, g = 0.001	0.8706	0.8594	0.8599
	CNN-FS	60	$C = 10^6$, g = 0.01	0.8668	0.8555	0.8572
	SVM-FS	60	$C = 10^7$, g = 0.01	0.7603	0.7433	0.7516
	UFS	60	$C = 10^7$, g = 0.01	0.6590	0.6488	0.6487
	ETC	60	$C = 10^6$, g = 0.01	0.8267	0.8123	0.8167
CNN-ATT	/	200	nn = 64	***0.9686***	***0.9303***	***0.9301***
	CNN-FS	60	nn = 32	***0.9428***	***0.9012***	***0.9020***
	SVM-FS	60	nn = 32	0.8924	0.8438	0.8394
	UFS	60	nn = 32	0.7778	0.7399	0.7420
	ETC	60	nn = 32	0.9233	0.8742	0.8773

**The used hyperparameters of different models. "comp" is the number of principle components for PLSDA. "C" and "g" are the C and gamma value for SVM. "nn" is the number of neurons in the first dense layer of the attention block. The bold and italic values are the best classification results achieved by modeling on full spectra and 60 selected features, respectively.*

Theoretically, the more features, the better the modeling effect. In order to better compare the effect of feature selector and ensure the calculation efficiency, we chose the features of the highest 60 scores for study. This value could be changed according to the requirement of practical applications.

Feature Selection Results

The results of the feature selection procedure are visualized in **Figure 6**. In **Figure 6A**, the scores calculated by the proposed CNN-FS are shown. The channel was given a nonzero score, which was found important for the target classification problem. The scores of non-informative channels were adjusted close to zero. In **Figure 6B**, the locations of the selected bands by different methods are marked. The UFS method found that the important bands lie in the range of 975–985 and 1,429–1,612 nm. The SVC method selected a range of 975–982, 1,389–1,500 and some bands in 1,534–1,646 nm. The ETC methods produced high importance score at the range of 975–1,022, 1,200–1,224, 1,294–1,338, 1,359–1,419, and 1,625–1,648 nm. The channels found by the abovementioned methods gathered in several continuous ranges. The channels selected by CNN-FS scattered in the whole range of 975–1,645 nm. To further analyze the common parts of the selected features, an indicator of each channel was calculated to indicate how many selectors decide that the channel was important (see **Figure 6C**). In **Figure 6C**, "L0" means that no selector found that the corresponding channel is important, "L1" means that one selector found the importance in the corresponding channel, and so on. The common ranges were centered at about 978, 1,405, and 1,601 nm.

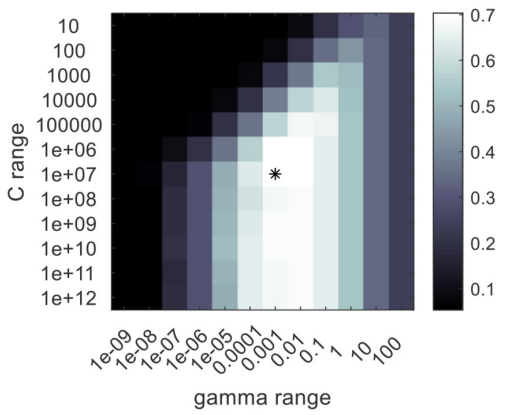

FIGURE 5 | The grid search result for the optimization of the radial basis function support vector machine classifier (RBF-SVC) model. The best combination of the RBF-SVC parameters is marked by "*".

Classification Result Analysis

The effects of feature selectors based on UFS, ETC, linear SVC, and CNN were compared. The classification accuracies reached by different classifiers with different feature extraction algorithms are summarized in **Table 2**. PLSDA is a recognized method for spectral analysis; however, the structure of the linear PLS model is simple. In this study, it is hard to predict the probability distribution of each category based on such a large and complex

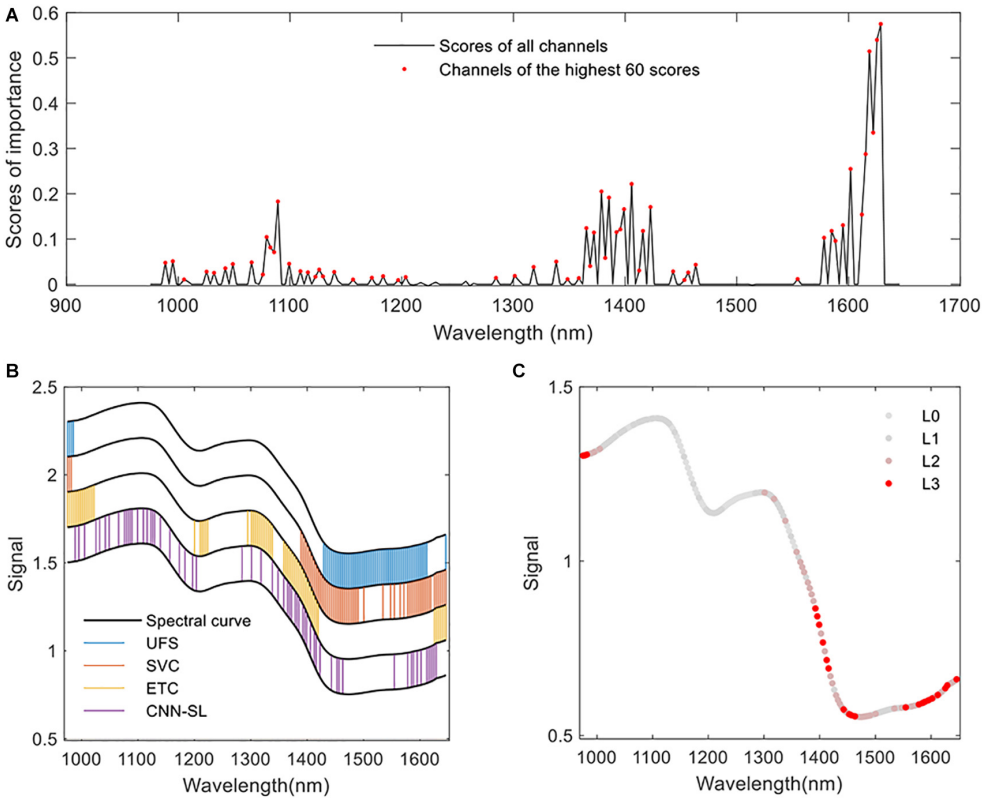

FIGURE 6 | The feature selection results. **(A)** The scores of each feature channels calculated by CNN-based feature selector (CNN-FS). **(B)** Feature selected by different algorithms. The black curves are produced by adding different offsets (0.2, 0.4, 0.6, 0.8, 1), which is only used for the convenience of visualization. The values of the y axis have no special meaning. **(C)** The overall result of feature selection. "L3" means that three of the four selectors find that the specific channel is important, "L2" means that two of them find the importance, and so on. The closer the color of the channel is to red, the more selectors select the channel.

dataset. The PLSDA-based methods achieved lowest accuracy rates (0.4132∼0.6880). The RBF-SVC method did better than the PLSDA. The CNN-ATT classifier realized higher accuracies on prediction set than those achieved by any other classifiers considered in this study. The highest classification accuracy of 0.9301 was realized by CNN-ATT with full spectra. The structures of the RBF-SVC and CNN were more complex than PLS. They could handle the dataset with high nonlinearity. Thus, the performances of these models were better than PLSDA.

Take the PLSDA classifiers as examples; the accuracies vary from 0.4132 to 0.6880. Modeling based on the UFS selected features performed the worst. The PLSDA models calibrated by SVM-FS and ETC produced better prediction results, and the model built based on features selected by CNN-FS showed the highest precision. The same phenomenon could be found in other classifiers. The feature channels selected by the proposed CNN-based feature extractor significantly improved the classification rate for all the tested classifiers. Furthermore, with 60 optimal bands screened by CNN-FS, the CNN-ATT classifier still achieved an accuracy of 0.902. The performance was not significantly decreased after feature selection and still kept high than 0.9, which proves the effectiveness/rationality of the features selected by the proposed CNN-FS. It can be inferred that CNN-based feature selectors could find the small differences lying in those selected channels, which was crucial to the classification task. The precision reached by CNN-ATT with other compared selectors was lower than 0.9 (0.7420∼0.8773). Experiments showed the superiority of the proposed CNN-ATT classifier and the CNN-FS feature selector.

The confusion maps of the best results (on the prediction set) achieved by the full spectra and the selected features are shown in **Figure 7**. Relatively high error rate was in some categories. For example, the samples of BM218 (category "5") were misjudged as HM40 (category "12"). The highest error rate in a single variety was about 20%. The low error was about 1%. The distribution of error rate was approximately the same in the two experiments (the full spectra-based modeling and 60 feature-based modeling).

DISCUSSION

NIR spectrometry has some advantages toward other technologies for processing a large number of kernels. Mass spectrometry could analyze the biochemical information of crop kernels with high sensitivity (Sorensen et al., 2002; Wadood et al., 2020). However, it is destructive, requiring high time consumption and relatively high cost. Crop seed classification

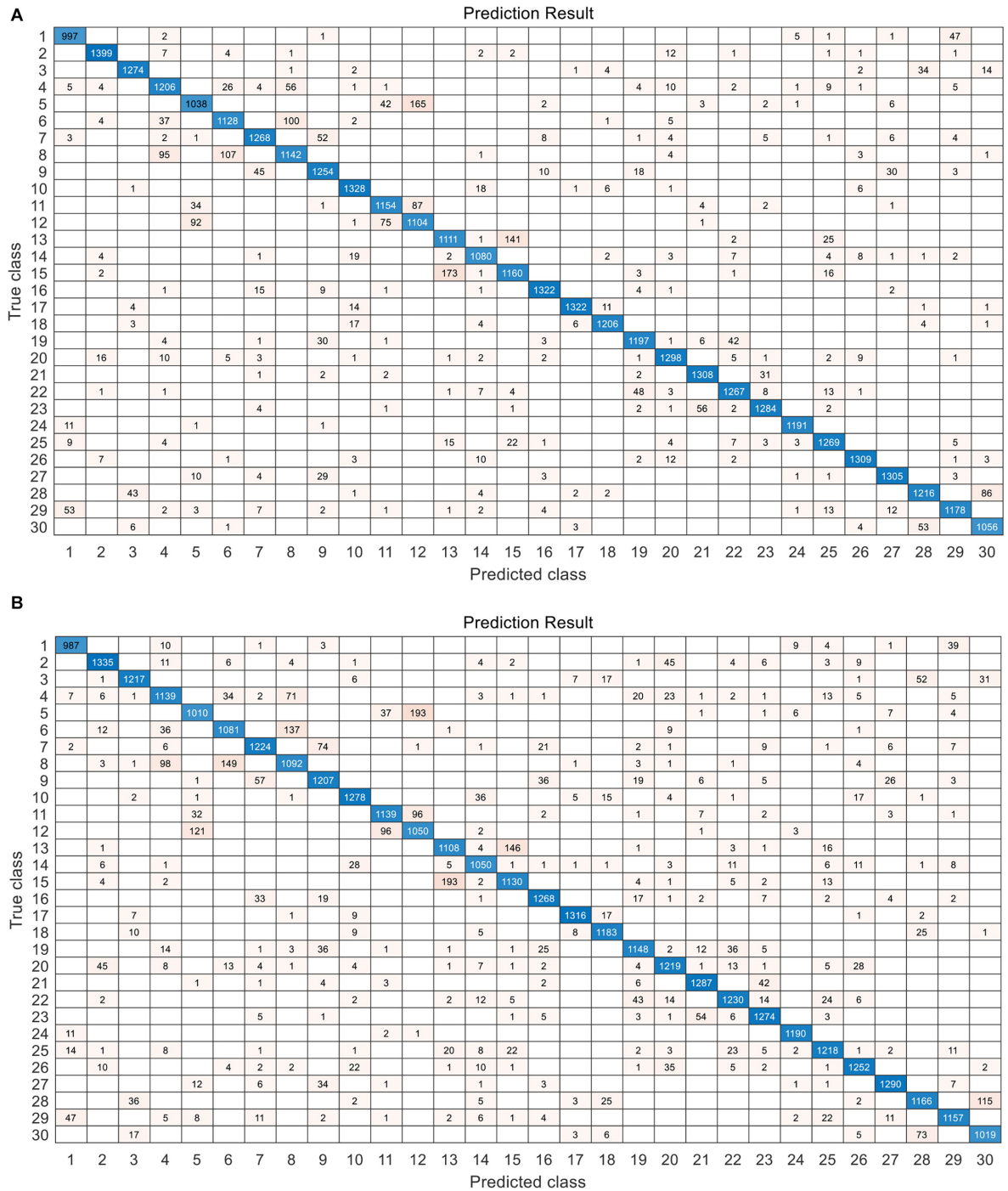

FIGURE 7 | The confusion map of the best classification result on the prediction set. **(A)** The result achieved by CNN with attention (CNN-ATT) model with 200 features. **(B)** The result achieved by CNN-ATT model with 60 features selected by CNN-FS.

applications based on electronic nose and thermal imaging were limited by the requirement of environment condition, and those based on X-ray imaging was also limited due to the high cost (Rahman and Cho, 2016). In addition, varieties of portable NIR spectral sensors (Crocombe, 2018; Deidda et al., 2019) with tiny size and low cost are available on the market, which makes the NIR spectroscopy-based technologies become promising tools for crop kernel inspection in practical application the future industry. Potential applications include breeding, seed adulteration inspection, and so on.

Feature selection methods choose the most informative features in the raw dataset. It is critical to big data analysis

and data mining. The existing feature selector can be divided into two types, unsupervised feature selection and supervised. As for the former, univariate statistical tests is used on each feature channel to calculate the importance. As for the latter, a classification or regression model is trained. The coefficients of the model are considered as the importance of the features, such as SVM-FS. Also, some supervised methods remove the non-informative feature during the training, such as uninformative variable elimination partial least square (UVE-PLS) (Wang et al., 2012). The supervised feature selector can also be considered as a model-based selector. According to the results in **Table 2**, the PLSDA methods could not discriminate the samples with high accuracy. The reliability of the features selected by low-performance models could not be guaranteed. Similar linear models could realize satisfactory performance in relatively small datasets. For example, in the article (Zhu et al., 2019a), the LR classifier calibrated by original features and CNN-extracted features reached similar accuracies of those achieved by CNN classifiers. However, PLSDA was insufficient to build a high-performance estimator on such a large and complex dataset in this study. The methods that are popular in small dataset processing may not be suitable for applications on large and complex datasets. Thus, some common methods such as SPA-PLS, UVE-PLS, and the methods based on loading weights of latent variables of PLS (Li et al., 2007, 2019; Zhang et al., 2017, 2020; Li and Hui, 2019) were not evaluated in this study. The UFS is a typical unsupervised feature selector. The performances of the classifiers did even worse based on such non-output-related selection method. The UFS tested weak for valid information identification. The CNN-based classifiers performed better than the SVM-based methods, and the classification results achieved by features from the CNN-FS was better than those from the SVM-FS. Thus, the capability of the core model is essential to the model-based feature selectors. As for the deep learning-based selector in Liu et al. (2019), the weights of the kernels in the first convolution layer was used to indicate the importance of the feature. As a fact, the 1D convolution kernel operates on the adjacent region centered on the target feature channel (rather than operates on the single target channel). Thus, the trained weights were affected by adjacent feature channels. The importance scores, which were determined by calculating the sum of weights in each convolutional kernel in the first convolutional layer, might be disturbed. Also, the scores of the unimportant channels were nonzero values in the trained model (Liu et al., 2019). These nonzero weighted features could still make a certain impact on classification.

The proposed CNN-FS used a powerful deep learning model to find the deep relation between the input and the target output. A custom linear selection layer was added for feature ranking. Each weight of the selection layer controls an individual feature, and the custom-defined loss function controls the adjustment process of the coefficients in the selection layer. The scores of non-informative channels could be tuned close to zero, as shown in **Figure 6A**. All of the three evaluated classifiers learned better knowledge from the selected features of CNN-FS than those from other feature selectors. The efficiency of the proposed CNN-FS was confirmed by the experimental results. This research provided an approach that adds a linear part at the start of the model to make the nonlinear model applicable for feature selection. In the future work, this idea can also be extended to develop other nonlinear feature selectors.

CONCLUSION

A large NIR spectral dataset of wheat kernels was collected for a 30-category classification task. The CNN-FS method was proposed for screening out the key bands of spectral data. Moreover, a CNN classifier with attention mechanism was designed for wheat kernel identification. The CNN-ATT method produced a higher precision than that realized by PLSDA and RBF-SVC. The selected key spectral channels from the CNN-FS acted more informative than those found by the compared existing feature selection methods. The CNN-ATT model achieved an accuracy of 0.9301 on the prediction set using full spectra, and it kept a high performance (accuracy of 0.9020) for classification when using 60 key channels selected by CNN-FS. The CNN-FS method was proved to be suitable for feature selection on the large dataset.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

LZ was responsible for the conceptualization, methodology, software, data curation, and writing – original draft preparation. CZ was also responsible for the conceptualization, writing – reviewing and editing, and funding acquisition. MT was responsible for data curation. XW was in charge of the resources and funding acquisition. YH was responsible for the project administration and writing – reviewing and editing. ZQ supervised and wrote – reviewing and editing, and acquired the funds. YL was responsible for the investigation, resources, writing – reviewing and editing, visualization, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been supported by the China National Key Research and Development Program (2016YFD0700304), the National Natural Science Foundation of China (61705195), the Postdoctoral Science Foundation of Zhejiang Province, China (zj2019091), and Synergistic Innovation Center of Jiangsu Modern Agricultural Equipment and Technology (Grant No. 4091600019).

ACKNOWLEDGMENTS

We would like to thank Ning Liang, Nan Zhao, and Shutao Zhao (College of Biosystems Engineering and Food Science, Zhejiang University) for the assistance of conducting the experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.575810/full#supplementary-material

REFERENCES

Balabin, R. M., and Smirnov, S. V. (2011). Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data. *Anal. Chim. Acta* 692, 63–72. doi: 10.1016/j.aca.2011.03.006

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowledge Discov.* 2, 121–167.

Chakraborty, A., Bhattacharya, A., and Mallick, B. K. (2020). Bayesian sparse multiple regression for simultaneous rank reduction and variable selection. *Biometrika* 107, 205–221. doi: 10.1093/biomet/asz056

Crocombe, R. A. (2018). Portable spectroscopy. *Appl. Spectrosc.* 72, 1701–1751. doi: 10.1177/0003702818809719

Deidda, R., Sacre, P. Y., Clavaud, M., Coic, L., Avohou, H., Hubert, P., et al. (2019). Vibrational spectroscopy in analysis of pharmaceuticals: critical review of innovative portable and handheld NIR and Raman spectrophotometers. *Trac Trends Anal. Chem.* 114, 251–259. doi: 10.1016/j.trac.2019.02.035

Ding, X.-L., Zhao, L.-X., Zhou, T.-T., Li, Y.-B., Huang, X.-M., and Zhao, Y.-L. (2019). Research on wheat leaf water content based on machine vision. *Cluster Comput.* 22, 9199–9208. doi: 10.1007/s10586-018-2112-4

Ebrahimi, E., Mollazade, K., and Babaei, S. (2014). Toward an automatic wheat purity measuring device: a machine vision-based neural networks-assisted imperialist competitive algorithm approach. *Measurement* 55, 196–205. doi: 10.1016/j.measurement.2014.05.003

Emura, T., Matsui, S., and Chen, H. (2019). compound.Cox: univariate feature selection and compound covariate for predicting survival. *Comput. Methods Prog. Biomed.* 168, 21–37. doi: 10.1016/j.cmpb.2018.10.020

Fabiyi, S. D., Vu, H., Tachtatzis, C., Murray, P., Harle, D., Dao, T., et al. (2020). Varietal classification of rice seeds using RGB and hyperspectral images. *IEEE Access* 8, 22493–22505. doi: 10.1109/access.2020.2969847

Feng, L., Zhu, S. S., Liu, F., He, Y., Bao, Y. D., and Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods* 15:91. doi: 10.1186/s13007-019-0476-y

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1

Jimenezcarvelo, A. M., Gonzalezcasado, A., Bagurgonzalez, M. G., and Cuadrosrodriguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – a review. *Food Res. Int.* 122, 25–39. doi: 10.1016/j.foodres.2019.03.063

Khaled, A. Y., Aziz, S. A., Bejo, S. K., Nawi, N. M., and Seman, I. A. (2018). Spectral features selection and classification of oil palm leaves infected by Basal stem rot (BSR) disease using dielectric spectroscopy. *Comput. Electron. Agric.* 144, 297–309. doi: 10.1016/j.compag.2017.11.012

Koistinen, V. M., and Hanhineva, K. (2017). Mass spectrometry-based analysis of whole-grain phytochemicals. *Crit. Rev. Food Sci. Nutr.* 57, 1688–1709. doi: 10.1080/10408398.2015.1016477

Kok, B. C., Choi, J. S., Oh, H., and Choi, J. Y. (2019). Sparse extended redundancy analysis: variable selection via the exclusive LASSO. *Multiv. Behav. Res.* doi: 10.1080/00273171.2019.1694477 [Epub ahead of print].

Krol, L., and Polanska, J. (2017). "Multidimensional feature selection and interaction mining with decision tree based ensemble methods," in *11th International Conference on Practical Applications of Computational Biology & Bioinformatics*, eds F. FdezRiverola, M. S. Mohamad, M. Rocha, J. F. DePaz, and T. Pinto (Cham: Springer International Publishing), 118–125. doi: 10.1007/978-3-319-60816-7_15

Lee, C., and Cai, J. Y. (2020). LASSO variable selection in data envelopment analysis with small datasets. *Omega Int. J. Manag. Sci.* 91:102019. doi: 10.1016/j.omega.2018.12.008

Lee, L. C., Liong, C. Y., and Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst* 143, 3526–3539. doi: 10.1039/c8an00599k

Lei, T., and Sun, D. W. (2020). A novel NIR spectral calibration method: sparse coefficients wavelength selection and regression (SCWR). *Anal. Chim. Acta* 1110, 169–180. doi: 10.1016/j.aca.2020.03.007

Li, Q. Y., and Hui, Y. H. (2019). Kinetic models of peroxidase activity in potato leaves infected with late blight based on hyperspectral data. *Int. J. Agric. Biol. Eng.* 12, 160–165. doi: 10.25165/j.ijabe.20191202.4574

Li, X. L., He, Y., Wu, C. Q., and Sun, D. W. (2007). Nondestructive measurement and fingerprint analysis of soluble solid content of tea soft drink based on Vis/NIR spectroscopy. *J. Food Eng.* 82, 316–323. doi: 10.1016/j.jfoodeng.2007.02.042

Li, Z., Tang, X., Shen, Z., Yang, K., Zhao, L., and Li, Y. (2019). Comprehensive comparison of multiple quantitative near-infrared spectroscopy models for *Aspergillus flavus* contamination detection in peanut. *J. Sci. Food Agric.* 99, 5671–5679. doi: 10.1002/jsfa.9828

Lin, P., Chen, Y. M., and He, Y. (2012). Identification of geographical origin of olive oil using visible and near-infrared spectroscopy technique combined with chemometrics. *Food Bioprocess Technol.* 5, 235–242. doi: 10.1007/s11947-009-0302-z

Liu, Y. S., Zhou, S. B., Han, W., Liu, W. X., Qiu, Z. F., and Li, C. (2019). Convolutional neural network for hyperspectral data analysis and effective wavelengths selection. *Anal. Chim. Acta* 1086, 46–54. doi: 10.1016/j.aca.2019.08.026

Mehrkanoon, S. (2019). Deep shared representation learning for weather elements forecasting. *Knowl. Based Syst.* 179, 120–128. doi: 10.1016/j.knosys.2019.05.009

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernet.* 9, 62–66. doi: 10.1109/tsmc.1979.4310076

Ou, J., Li, Y., and Liu, W. (2020). TDP: two-dimensional perceptron for image recognition. *Knowl. Based Syst.* 195:105615. doi: 10.1016/j.knosys.2020.105615

Ozkan, K., Isik, S., and Yavuz, B. T. (2019). Identification of wheat kernels by fusion of RGB, SWIR, and VNIR samples. *J. Sci. Food Agric.* 99, 4977–4984. doi: 10.1002/jsfa.9732

Pes, B. (2019). Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains. *Neural Comput. Appl.* 32, 5951–5973. doi: 10.1007/s00521-019-04082-3

Qiu, Z., Chen, J., Zhao, Y., Zhu, S., He, Y., and Zhang, C. (2018). Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Appl. Sci.* 8:212. doi: 10.3390/app8020212

Rahman, A., and Cho, B. K. (2016). Assessment of seed quality using non-destructive measurement techniques: a review. *Seed Sci. Res.* 26, 285–305. doi: 10.1017/s0960258516000234

Sabanci, K., Kayabasi, A., and Toktas, A. (2017). Computer vision-based method for classification of wheat grains using artificial neural network. *J. Sci. Food Agric.* 97, 2588–2593. doi: 10.1002/jsfa.8080

Sorensen, H. A., Sperotto, M. M., Petersen, M., Kesmir, C., Radzikowski, L., Jacobsen, S., et al. (2002). Variety identification of wheat using mass spectrometry with neural networks and the influence of mass spectra processing prior to neural network analysis. *Rapid Commun. Mass Spectr.* 16, 1232–1237. doi: 10.1002/rcm.709

Wadood, S. A., Guo, B., Zhang, X., Raza, A., and Wei, Y. (2020). Geographical discrimination of Chinese winter wheat using volatile compound analysis by HS-SPME/GC-MS coupled with multivariate statistical analysis. *J. Mass Spectr.* 55:e4453. doi: 10.1002/jms.4453

Wang, S., Huang, M., and Zhu, Q. (2012). Model fusion for prediction of apple firmness using hyperspectral scattering image. *Comput. Electr. Agric.* 80, 1–7. doi: 10.1016/j.compag.2011.10.008

Xia, C., Yang, S., Huang, M., Zhu, Q., Guo, Y., and Qin, J. (2019). Maize seed classification using hyperspectral image coupled with multi-linear discriminant

analysis. *Infrared Phys. Technol.* 103:103077. doi: 10.1016/j.infrared.2019.103077

Yuan, X., Huang, B., Wang, Y., Yang, C., and Gui, W. (2018). Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE. *IEEE Trans. Ind. Inform.* 14, 3235–3243. doi: 10.1109/tii.2018.2809730

Zhang, C., Guo, C. T., Liu, F., Kong, W. W., He, Y., and Lou, B. G. (2016). Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. *J. Food Eng.* 179, 11–18. doi: 10.1016/j.jfoodeng.2016.01.002

Zhang, C., Jiang, H., Liu, F., and He, Y. (2017). Application of near-infrared hyperspectral imaging with variable selection methods to determine and visualize caffeine content of coffee beans. *Food Bioprocess Technol.* 10, 213–221. doi: 10.1007/s11947-016-1809-8

Zhang, L., Li, Y., Huang, W., Ni, L., and Ge, J. (2020). The method of calibration model transfer by optimizing wavelength combinations based on consistent and stable spectral signals. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 227:117647. doi: 10.1016/j.saa.2019.117647

Zhang, Q., Yang, L. T., Chen, Z., and Li, P. (2018). A survey on deep learning for big data. *Inform. Fusion* 42, 146–157.

Zhang, X., Sun, Y., Zhang, J., Wu, P., and Jiao, L. (2018). Hyperspectral unmixing via deep convolutional neural networks. *IEEE Geosci. Rem. Sens. Lett.* 15, 1755–1759. doi: 10.1109/lgrs.2018.2857804

Zhou, L., Zhang, C., Liu, F., Qiu, Z., and He, Y. (2019). Application of deep learning in food: a review. *Compr. Rev. Food Sci. Food Saf.* 18, 1793–1811. doi: 10.1111/1541-4337.12492

Zhu, S., Zhou, L., Gao, P., Bao, Y., He, Y., and Feng, L. (2019a). Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties. *Molecules* 24:3268. doi: 10.3390/molecules24183268

Zhu, S., Zhou, L., Zhang, C., Bao, Y., Wu, B., Chu, H., et al. (2019b). Identification of soybean varieties using hyperspectral imaging coupled with convolutional neural network. *Sensors* 19:4065. doi: 10.3390/s19194065

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhou, Zhang, Taha, Wei, He, Qiu and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 November 2020
doi: 10.3389/fpls.2020.577063

Investigation on Data Fusion of Multisource Spectral Data for Rice Leaf Diseases Identification Using Machine Learning Methods

*Lei Feng[1,2], Baohua Wu[1,2], Susu Zhu[1,2], Junmin Wang[3], Zhenzhu Su[4], Fei Liu[1,2], Yong He[1,2] and Chu Zhang[1,2]**

[1] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2] Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou, China, [3] Institute of Crop Science and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [4] State Key Laboratory for Rice Biology, Institute of Biotechnology, Zhejiang University, Hangzhou, China

Rice diseases are major threats to rice yield and quality. Rapid and accurate detection of rice diseases is of great importance for precise disease prevention and treatment. Various spectroscopic techniques have been used to detect plant diseases. To rapidly and accurately detect three different rice diseases [leaf blight (*Xanthomonas oryzae* pv. *Oryzae*), rice blast (*Pyricularia oryzae*), and rice sheath blight (*Rhizoctonia solani*)], three spectroscopic techniques were applied, including visible/near-infrared hyperspectral imaging (HSI) spectra, mid-infrared spectroscopy (MIR), and laser-induced breakdown spectroscopy (LIBS). Three different levels of data fusion (raw data fusion, feature fusion, and decision fusion) fusing three different types of spectral features were adopted to categorize the diseases of rice. Principal component analysis (PCA) and autoencoder (AE) were used to extract features. Identification models based on each technique and different fusion levels were built using support vector machine (SVM), logistic regression (LR), and convolution neural network (CNN) models. Models based on HSI performed better than those based on MIR and LIBS, with the accuracy over 93% for the test set based on PCA features of HSI spectra. The performance of rice disease identification varied with different levels of fusion. The results showed that feature fusion and decision fusion could enhance identification performance. The overall results illustrated that the three techniques could be used to identify rice diseases, and data fusion strategies have great potential to be used for rice disease detection.

Keywords: hyperspectral imaging, mid-infrared spectroscopy, laser-induced breakdown spectroscopy, data fusion, rice disease

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Jakub Nalepa, Silesian University of Technology, Poland
Sailendra Nath Sarkar, University of Calcutta, India

****Correspondence:***
Chu Zhang
chuzh@zju.edu.cn

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *28 June 2020*
Accepted: *06 October 2020*
Published: *10 November 2020*

Citation:
Feng L, Wu B, Zhu S, Wang J, Su Z, Liu F, He Y and Zhang C (2020) Investigation on Data Fusion of Multisource Spectral Data for Rice Leaf Diseases Identification Using Machine Learning Methods. Front. Plant Sci. 11:577063. doi: 10.3389/fpls.2020.577063

INTRODUCTION

With the increase of population, the demand for food supply will surge. To meet such a great need of food, it is critical to improve crop efficiency to increase the food supply. Cereals are stable food supply for human beings. Due to the changes in climate and environment, biological and abiotic stresses which hinder the normal growth of crops become increasingly frequent. The disease is one

of the major stresses of crops, causing severe losses in quality and yield (Skolik et al., 2019; Yang et al., 2019).

Rice is one of the most popular staple food sources in the world, and rice is widely planted all over the world, especially in Asia and Africa. However, there are various diseases influencing rice growth. Bacterial leaf blight (*Xanthomonas oryzae* pv. *Oryzae*) (Alberto, 2018), blast (*Pyricularia oryzae*) (Gaoqiang et al., 2020), and sheath blight (*Rhizoctonia solani*) (Yuan et al., 2019) are the three major diseases of rice (Kumar et al., 2020; Molla et al., 2020). Prevention and treatment of disease is an indispensable task for rice growth management at the current time. Traditionally, on the one hand, detection of rice diseases is mainly based on the experts or experienced farmers, with their visual and manual work. On the other hand, with the development of molecular biology and the related techniques, rice diseases can be accurately detected, and these techniques have been widely used as "standard" or "reference" techniques in the related fields. The shortcomings of these techniques are also obvious. They are time consuming, expensive, and complex to be operated.

Rapid and accurate techniques for rice disease detection are of great importance for rice growth management. For the past decades, optical characteristics of plants have widely been studied (Carter and Knapp, 2001; Altangerel et al., 2017; Ribeiro et al., 2018; Liu et al., 2019; Zahid et al., 2019). Under the stress of diseases, external features such as morphology, color, and texture are changed. The plants' self-defense systems also work to alleviate the damage, resulting in the changes of physiological and biochemical parameters. These changes can be captured by various spectroscopic techniques based on different principals. Researchers have used various spectroscopic techniques for plant disease detection (Zhang et al., 2017; Thomas et al., 2018; Farber et al., 2019; Liu et al., 2019). In this study, to detect rice diseases, spectral information of visible/near-infrared hyperspectral imaging (HSI), mid-infrared spectroscopy (MIR), and laser-induced breakdown spectroscopy (LIBS) were used.

HSI integrates both visible/near-infrared spectroscopy and imaging techniques. Visible/near-infrared spectroscopy has a strong relationship with biological and physiological parameters and internal structures of plants, and it is the most widely used spectroscopic technique to monitor plant growth and plant stresses (Knauer et al., 2017; Asaari et al., 2018; Ribeiro et al., 2018). MIR is a spectroscopic technique to study the fundamental vibrations and associated rotational-vibrational structure of chemical bonds (Machado et al., 2018; Skolik et al., 2019). MIR is used to identify the chemical components of plants and monitoring the change in those components can help to identify plant growth status. LIBS is a spectroscopic technique to detect elements and their concentrations by analyzing spectral signal constituted by the light emission from laser plasma (Peng et al., 2016). LIBS is used for quantitative and qualitative analysis of elements in plants.

Among these three spectroscopic techniques, HSI (Knauer et al., 2017; Thomas et al., 2018) is the most widely used technique for plant disease detection, while fewer studies have used MIR (Hawkins et al., 2010; Zhang et al., 2017) and LIBS (Ponce et al., 2018; Liu et al., 2019) for disease detection. Thomas et al. adopted HSI to detect barley cultivars inoculation with powdery mildew. An accurate assessment of the disease severity for all six cultivars at measurements over 30 days was achieved (Thomas et al., 2018). Luo et al. (2019) applied HSI to grade the severity of rice blast, and the probabilistic neural network obtained the best performance with the highest classification accuracy of 97.8%. For MIR, Zhang et al. (2017) explored and validated the feasibility of using MIR to detect oilseed rape leaves infected with Sclerotinia stem rot. Healthy and infected leaves had a difference on the average MIR spectra, and the accuracy over 80% was achieved with three chemometric methods. In terms of disease detection with LIBS, Ponce et al. (2018) applied LIBS for discrimination between healthy and Huanglongbing-affected citrus. The wavebands that had the most obvious difference between healthy and HLB-affected trees were the same for all species. With chemometric analysis, the healthy status of plants was differentiated with a high degree of precision.

Several groups of scientists are involved in disease detection using spectral features and modeling (Alberto, 2018; Luo et al., 2019; Gaoqiang et al., 2020). However, these studies did not detect different diseases simultaneously. In natural conditions, there are various diseases affecting rice growth, and they can happen within one field. Since many external and internal symptoms of rice leaves are similar under different disease pressures, it is difficult to detect multiple diseases simultaneously. For this reason, it is of significant importance and applicability to detect various rice diseases at the same time with one model. Besides, only one detection technique was used to detect plant diseases in each study. Although the single dataset can be used to solve the same problem, a combination of information from different modalities have the potential to provide a better understanding of the problem since each technique has unique advantages as well as limitations.

Information fusion of multiple modalities is the key to combine these three techniques. However, the use of datasets with various modalities is a challenging issue. In general, information fusion can be categorized as low-level fusion (raw data are directly combined), mid-level fusion (features extracted from the raw datasets are combined), and high-level fusion (known as decision fusion, the decision results are combined) (Castanedo, 2013; Borras et al., 2015; Zhou et al., 2020). Information fusion aims to reveal the benefits of multisensor measurement, and they are expected to perform better than individual sensors, providing more robust and accurate decisions. In this study, the three fusion levels of HSI, MIR, and LIBS to detect rice diseases were explored.

The objective of this study was to use HSI, MIR, and LIBS to detect three different diseases of rice, including rice leaf blight, rice blast, and rice sheath blight. The specific objectives were to (1) explore the spectral differences among rice leaves inoculated by different diseases; (2) conduct low-level, medium-level, and high-level data fusion for disease identification; (3) develop detection models based on fused data and nonfused data.

MATERIALS AND METHODS

Sample Preparation

To verify the proposed methods in this article being effective despite rice varieties, two different rice varieties were used in this study, including a commercial variety (Zhefujing83) and a newly developed variety (AD516, which is provided and cultivated by the Institute of Crop Science and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China). After 1 month of sowing seed into the seed plots, the seedlings were transplanted into the laboratory greenhouse with regular fertilization and watering.

To obtain inoculated samples, the *in vitro* inoculation method was applied. Rice blast and rice sheath blight are fungal diseases, while rice leaf blight is a bacterial disease. Thus, funguses of rice blast and rice sheath blight were cultured on potato dextrose agar medium, and bacteria of rice leaf blight were cultured in conical flasks.

The leaves cut from healthy plants were used to inoculate fungus and bacteria. Then the leaves were put into sterilized plastic flat plates. To prevent the rice leaves from drying out, the leaves were placed on distilled water–sterilized wipes. For rice blast and rice sheath blight inoculation, the mycelial pellets were placed on the leaves, with two or three pellets per leaf. For rice leaf blight inoculation, the solutions of bacteria were sprayed on to the leaf surface. After inoculation, the plates were sealed and then placed in a room with a temperature of about 26℃ and relative humidity about 60%, and healthy leaves were used as control. Four days later, leaves with visible symptoms were collected.

Representative images of diseased leaves are shown in **Supplementary Figure 1**. The infected leaves were collected for hyperspectral image acquisition. The number of leaves used in this study is presented in **Table 1**. Six or twelve leaves were acquired in one image. If an image contained 12 leaves, this image would be divided into two subimages with six leaves in each image. After hyperspectral image acquisition, the six leaves were dried as one sample for MIR and LIBS analysis in an oven at the temperature 75°C for 12 h a day for 3 days in a row. Then the dried leaves were placed into centrifuge tubes and ground into powder using an electrical grinder for 5 min with an oscillation frequency of 60 Hz. Unground leaf veins were removed from the centrifuge tubes.

Regarding data splitting, 30 and 5 samples of each category were randomly selected into the training set and validation set, and the remaining samples of each category were all selected into an external test set. Besides, the order of samples in different sets of three spectra was the same. In this study, the category value of the healthy samples (CK) was assigned as 0, and the category values of the samples inoculated by rice leaf blight (BYK), rice blast (DWB), and rice sheath blight (WKB) were assigned as 1, 2, and 3, respectively.

TABLE 1 | The number of leaves under different disease inoculations (six leaves per sample).

Cultivar	BYK[a]	DWB[b]	WKB[c]	CK[d]
Zhefujing83	276 (46)[e]	324 (54)	312 (52)	240 (40)
AD516	294 (49)	330 (55)	336 (56)	264 (44)

The numbers in the brackets were the numbers of samples to be studied. The numbers outside the brackets were the numbers of leaves. [a]The samples inoculated by rice leaf blight. [b]The samples inoculated by rice blast. [c]The samples inoculated by rice leaf blight. [d]The samples inoculated by rice sheath blight. [e]Six leaves were used as one sample.

Hyperspectral Image Acquisition and Spectra Extraction

A visible/near-infrared hyperspectral imaging system covering the spectral range of 400–1,000 nm was used to acquire hyperspectral images of healthy and infected leaves. The hyperspectral imaging system is formed by an imaging spectrograph (ImSpector V10E; Spectral Imaging Ltd., Oulu, Finland), a highly sensitive EMCCD camera (Raptor EM285CL, Raptor Photonics limited, Larne, United Kingdom), and a long camera lens (OLES23; Specim, Spectral Imaging Ltd., Oulu, Finland). The illumination of the system is provided by 150 W tungsten halogen lamps (3900 Lightsource, Illumination Technologies Inc., United States). This hyperspectral imaging system conducts line scanning, and a moving plate driven by a stepper motor (GYB751D5-RC2, Fuji Electric (Dalian) Co., Ltd., Dalian, China) is used to move the samples.

To acquire clear and nondeformable images, the distance between the camera lens and the moving plate, the exposure time of the camera, and the moving speed of the moving plate was adjusted to 26 cm, 55 ms, and 2.6 mm/s. The acquired hyperspectral images were then corrected by using the white reference image (acquired using a piece of pure white Teflon board with nearly 100% reflectance) and the dark reference image (acquired covering the lens using a black lens cap with nearly 0% reflectance) according to the following equation:

$$I_C = \frac{I_R - I_D}{I_W - I_D} \tag{1}$$

where I_C is the corrected image, I_R is the raw image, I_W is the white reference image, and I_D is the dark reference image.

After image correction, all six leaves were defined as a region of interest (ROI); wavelet transform (Wavelet function: Daubechies 8; Decomposition level: 3) was used to denoise the pixel-wise spectra. The average spectrum of the six leaves was calculated as one sample spectrum. The head and the tail of the spectra contained obvious noises produced by the hyperspectral imaging system. Only the spectra in the range of 448–945 nm were used for analysis.

Mid-Infrared Spectra Collection

To conduct MIR spectra acquisition, rice leaves were dried, ground, and pressed into pellets. Potassium bromide (KBr) was used to be mixed with the leaf powders for mid-infrared spectra acquisition. The KBr powders were first dried at 105°C in an oven for 4 h and then mixed. To obtain the pellet, 0.02 g leaf powders and 0.98 g KBr powders were weighed and mixed thoroughly. Then 0.2 g mixtures were used for tableting. To obtain mid-infrared spectra, a Nicolet iS10 FT-IR spectrometer (Thermo

Fisher Scientific™, Madison, WI, United States) was used. The spectral range was 400–4,000 cm^{-1}, and the spectral resolution was set as 4 cm^{-1}. For each sample, 32 measurements were conducted, and the average value was used as the transmittance spectrum of the sample to reduce variations and random noises. Before spectra acquisition, background correction was performed, and the background correction was conducted every 30 min during spectra acquisition. The spectral data was saved as .csv format for further analysis.

LIBS Spectra Acquisition

To conduct LIBS spectra acquisition, 0.1 g rice leaf powders were used for tableting. The tablets were dried for 4 h at 75°C before LIBS spectra acquisition. An assembled LIBS system was used to acquire LIBS spectra. The LIBS system used in this research consists of a Q-switch Nd:YAG nanosecond pulsed laser (Vlite-200, Beamtech, Beijing, China). The second harmonic laser (532 nm, pulse duration of 8 ns, beam diameter of 7 mm) was used to ignite the sample with the help of a plano-convex lens (f = 50 mm). The detection system consists of a Mecchelle spectrograph (ME5000, Andor, Belfast, United Kingdom) and an ICCD camera (DH334, Andor, Belfast, United Kingdom), which was used to collect plasma emission spectra with the range from 230 to 880 nm. Samples were placed at an X-Y-Z translation stage.

To improve the signal to noise rate, the delay time and the integration time were optimized to 1.5–10 μs, respectively. The laser was fired with a pulse energy of 60 mJ at 1 Hz. For each sample, 10 successive spectra were acquired at each location, and 16 different locations were measured with the help of translation stage in ambient air.

Data Feature Extraction

Since full spectra of HSI, MIR, and LIBS have high dimensionalities of features, it will increase computing time and the difficulty to build models. Reducing the feature dimensions while keeping the most useful information is a good approach to make full use of the features. Feature extraction methods are effective tools to extract most informative features for dimension reduction. In this study, two feature extraction methods used in spectral data analysis were applied for feature extraction, principal component analysis (PCA), and autoencoder (AE).

PCA is a widely used feature extraction method for data compression in spectral data analysis. It can reduce the dimensionality of spectral information through calculating the linear combination of the original data (Diniz et al., 2014). The variables after transformation are called principal components (PCs). After the orthogonal transformation of PCA, the first few PCs contain a majority of the information pertaining to the original variables (Zhu S.et al., 2019). The accumulative explained variance determines the number of PCs. Considering making use of original information as much as possible, the first few PCs with accumulative explained variance over 99.99% were adopted as extracted features in this study. Therefore, the number of PCs of three different sources was not identical.

The deep learning (DL) framework has been introduced into feature extraction and data reduction in spectra analysis recently due to its powerful representation ability. As part of the DL framework, the AE network can learn abstract features through the hidden layer in an unsupervised manner (Zabalza et al., 2016), which makes AE quite popular for feature extraction. With the assistant of hidden layers in neural networks, data reduction is achieved while maintaining the effective information of the data (Zabalza et al., 2016).

AE is a neural network that reconstructs the value of output to be as the same as possible to the value of the input, which indicates the output layer has the same number of nodes as the input layer (Xing et al., 2016). A basic architecture of AE can be seen in **Figure 1**. In the encoder part, a basic AE has an input layer of i neurons, which is equal to the dimensionalities of features of the input. A hidden layer with h neurons ($h < i$) following with the input layer is also introduced. This hidden layer is used to extract features with an activation function. In the decoder part, the h is mapped to an output layer with o neurons ($o = i$) to reconstruct the input data (Lin et al., 2013). This network is used to reconstruct the original spectra by minimizing the loss of mean squares, which maintains the key information of original data.

AE can simply introduce several hidden layers between the input and the output for feature extraction. After several trials, the shallow AE network was found to be more effective for reconstructing the original spectra of rice leaves. The architectures of AE used in this work and the change of data dimensionalities are reported in **Supplementary Figure 2**. For HSI, the input layer was a fully connected layer, and dimensionalities of the output were 64. Next, the dimensionality of data became 32 after encoder, and the features were further used for classification. Then the decoder increased the dimensionality of data to 64, and further increased to the same dimensionality of input data. Apart from the last output layer in the AE, the other three fully connected layers were all followed by a batchnorm layer. Besides, all other three layers used the rectified linear unit (ReLU) as an activation function while the last output layer did not use any activation function. The data dimensionalities from beginning to the end could be simply recorded as 390–64–32–64–390. In terms of MIR spectra, the architecture with total four layers was also adopted, and the change of dimensionalities could be simply recorded as 7,468–64–16–64–7,468. Since LIBS spectra had over 20,000 dimensionalities of features, we increased the dimension of the

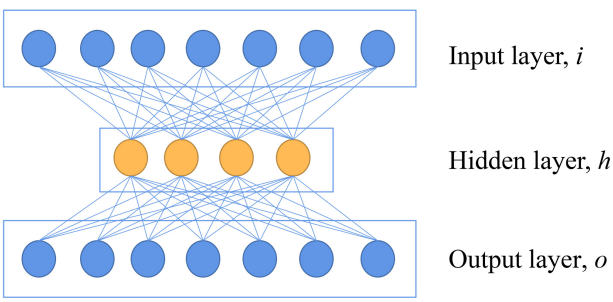

FIGURE 1 | The basic architecture of autoencoder.

output of the first layer. Thus, the change of dimensionalities could be simply recorded as 22,036–256–64–256–22,036.

Data Fusion Strategies

Data fusion is a method to fuse data collected from multiple sensors (Doeswijk et al., 2011). To fully dig effective information from different sources, different kinds of data fusion strategies were used to investigate the feasibility of combining the information with HSI, MIR, and LIBS datasets for rice disease detection.

Low-Level Data Fusion

In the case of low-level fusion strategy, the original HSI, MIR, and LIBS datasets were concatenated into a single matrix. Two or three data sources were fused. The four combinations were listed as follows: HSI-MIR (this means the fusion of HSI spectra and MIR spectra, and other abbreviations are similar to this fashion), HSI-LIBS, MIR-LIBS, and HSI-MIR-LIBS. However, the full use of the datasets of different spectra is a challenge because the informative parameters with small value will be ignored due to the existence of large value parameters during data fusion (Wang et al., 2016). Therefore, a z-score normalization was applied firstly to rescale the spectra values of different sensors before the model construction.

Mid-Level Fusion

In mid-level fusion, two strategies were adopted: (1) concatenating the PCA/AE features from three data sources, respectively. As shown in **Figure 2**, PCA features separately extracted from three kinds of spectra were concatenated. There were four combinations for PCA features: PCA-HSI-MIR (this means concatenating the PCs of HSI and the PCs of MIR, and other abbreviations are similar to this fashion), PCA-HSI-LIBS, PCA-MIR-LIBS, and PCA-HSI-MIR-LIBS. In terms of AE features, fused data could be briefly recorded as AE-HSI-MIR, AE-HSI-LIBS, AE-MIR-LIBS, and AE-HSI-MIR-LIBS; (2) concatenating the PCA features and the AE features from the same data source, and the fused data can be briefly recorded as AE-PCA-HSI, AE-PCA-MIR, and AE-PCA-LIBS. Furthermore, the aforementioned concatenated features of three different instruments were further combined. These combined features could be briefly recorded as AE-PCA-HSI-MIR, AE-PCA-HSI-LIBS, AE-PCA-MIR-LIBS, and AE-PCA-HSI-MIR-LIBS, as shown in **Figure 3**.

High-Level Fusion

The high-level fusion is also called decision-level fusion. Majority voting is commonly applied to make the final prediction according to the prediction results of single models (Ballabio et al., 2018). For the same modeling method, models were built using HSI, MIR, and LIBS datasets, and the voting weights of each dataset were set as equivalent. For the same sample, if two or three of the models predicted it as the same value, then this value was set as the final prediction value of the sample. If the prediction value of the same sample by the models using the three datasets were all different, it means that the sample was unidentified.

In summary, we conducted three fusion strategies for rice disease detection: (1) the preprocessed spectra were concatenated as a new dataset to build classification models directly; (2) to conduct mid-level fusion, PCA and AE were used as feature extraction methods. The features extracted by the same feature extraction method (PCA/AE) of the three datasets were combined. Also, the features extracted by different feature extraction methods of each dataset were combined. Furthermore,

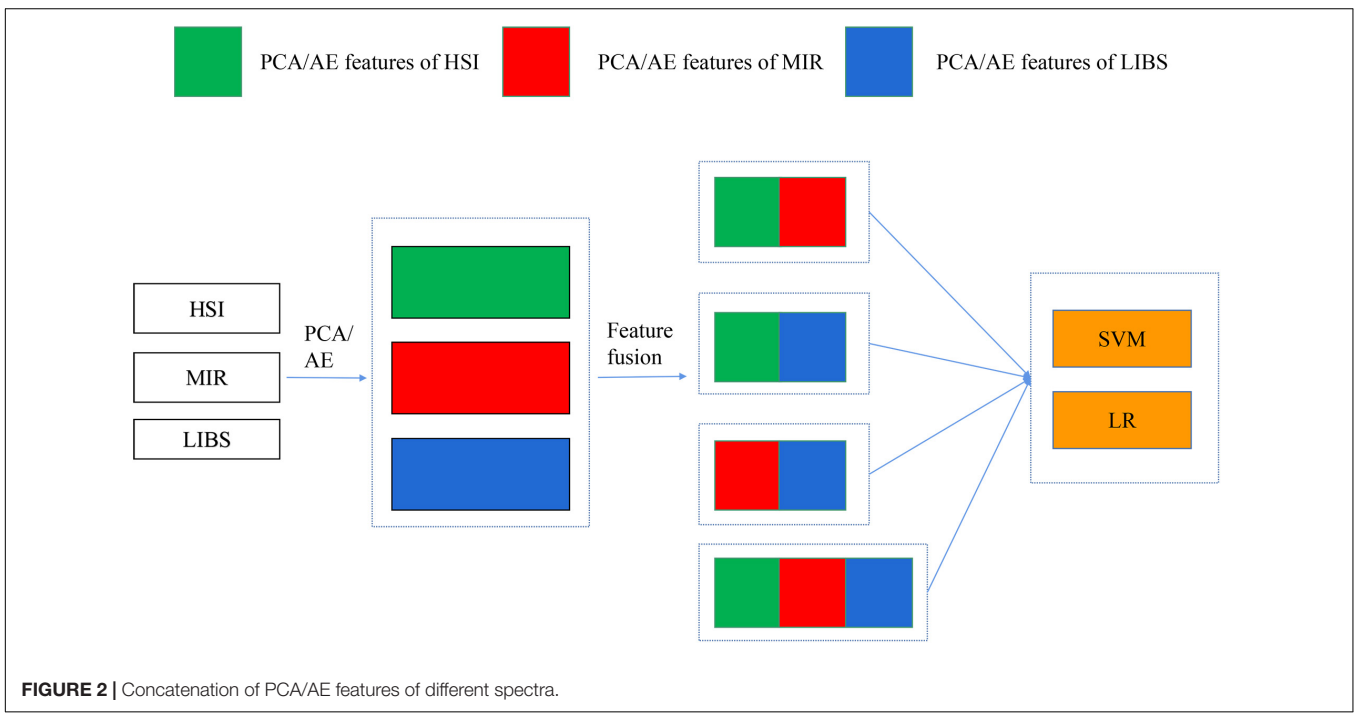

FIGURE 2 | Concatenation of PCA/AE features of different spectra.

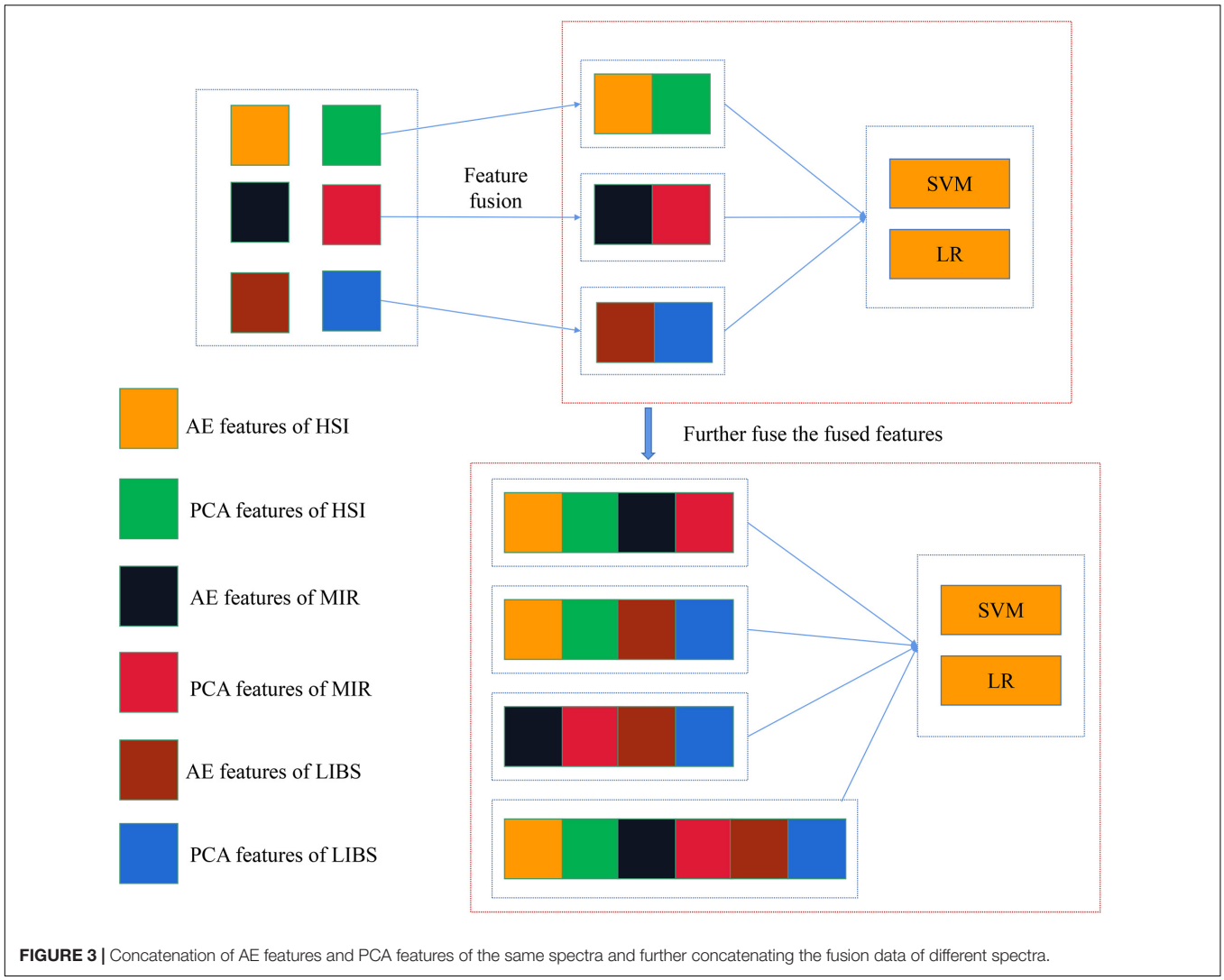

FIGURE 3 | Concatenation of AE features and PCA features of the same spectra and further concatenating the fusion data of different spectra.

aforementioned fused features of the three datasets were further combined; (3) decision fusion was conducted by combining the prediction results of different datasets using the same modeling method.

Classification Methods

Discriminant methods were used in this study to classify different rice diseases. These methods included logistic regression (LR), support vector machine (SVM), and convolutional neural network (CNN).

LR is a linear regression method used for classification. LR model is basically used to solve the binary classification problems. It outputs probabilities of two situations, and the class of the corresponding sample is determined based on the probabilities. The general idea of LR is to map the real value predicted by the linear regression model into the value in range 0–1 (probability) by a sigmoid function. LR can also be extended to multiclass classification problems by using the one vs. rest strategy (Tremblay et al., 2019; Zhu S.S.et al., 2019).

SVM is a widely used machine learning method for classification and regression. For linearly separable data, a linear equation can be obtained to construct the hyperplanes. For data which are not linearly separable, SVM maps the original data into high-dimensional spaces to transform the problem into linearly separable issues and constructs hyperplanes to maximally divide the samples from different categories in the new spaces (Feng et al., 2018; Sun et al., 2018; Zhang et al., 2018). Kernel functions are essential for the mapping, and radial basis function (RBF) is a widely used kernel function of SVM. In this study, RBF was used as the kernel function for SVM with a grid-search procedure for parameter optimization.

CNN is a promising method in various fields nowadays. CNN consists of multiple convolution layers and pooling layers, which enable this neural network to extract abstract shallow and deep features of the input automatically. Due to its powerful representative ability, CNN recently has been introduced to vibration spectral data analysis for classification (Wu et al., 2018; Feng et al., 2019; Zhu S.S.et al., 2019) and regression (Ng et al., 2020; Zhang et al., 2020).

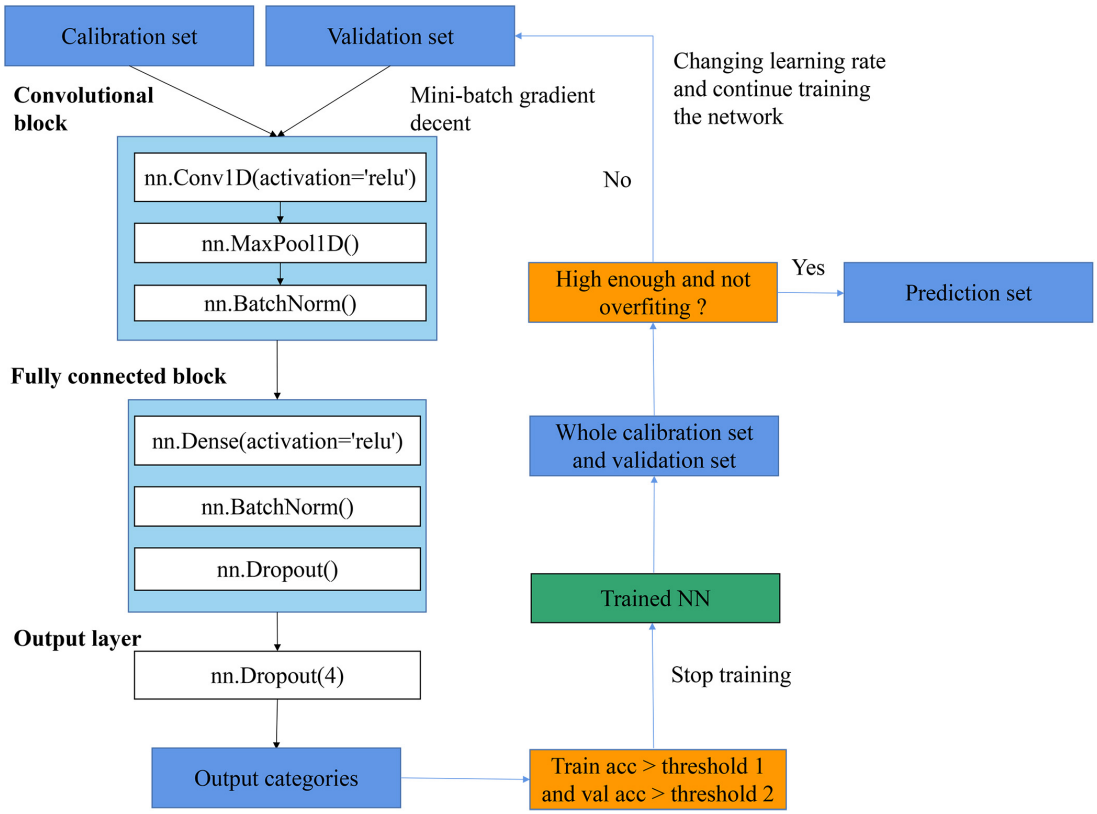

FIGURE 4 | The flowchart of procedures of training CNN.

As shown in **Figure 4**, the general flowchart of training CNN in this article included: (1) feeding training set and validation set into a network and keep training with changing learning rate until the training accuracy and validation accuracy reach up to thresholds; (2) applying the trained network to predict the whole training set and validation set and inspecting whether the training accuracy and validation accuracy were high enough and whether the overfitting problem existed; (3) if the accuracy was high enough and the overfitting did not exist, the trained network was saved and all results with this trained network were recorded; otherwise, changing the learning rate and keep training the network. The training procedure was performed by optimizing the cross-entropy loss with the SGD algorithm. Furthermore, the relationship between epochs and training performances is provided in **Supplementary Figure 3**, which illustrated the change of training accuracy and train loss as the change of epochs. This is an initial step of training process. After this step, several small learning rates were used to finetune the CNN, and a threshold (e.g., accuracy >0.98) was set to stop the training process (which is not shown in the figure).

Model Establishment, Evaluation, and Software

To develop the SVM model, the RBF kernel was set as the kernel function. The optimal combination of penalty parameter (c) and kernel coefficient parameter gamma (γ) of SVM was searched using the GridSearchCV function provided by scikit-learn (version 0.21.3), a Python machine learning package. The search range of parameters was set as 10^{-10} to 10^{10} for both c and γ, and the three-fold was used for cross-validation strategy. In addition, the scoring parameter of the function was set as "accuracy." In terms of the optimization algorithm of LR, "liblinear" was chosen for L1 penalization and "newton-cg" was chosen for L2 penalization in this article. After optimizing these parameters, a fine-tuning of the parameters was implemented on the training set and the validation set. Except for the traditional SVM and LR models, CNN was also used for the classification task. Since the dimensionalities of features of raw spectra were varied with the type of technique, the architectures of CNN were not identical. Besides, different spectra data have different data structures, and different CNN architectures were tried and a superior one was chosen for each specific input to get better results (shown in **Supplementary Figures 4–19**).

Each preprocessed spectrum of each technique was further implemented with the standardization process before being fed into SVM and LR. This standardization preprocessing standardized features of the training set by removing the mean and scaling to unit variance and performed the same standardization on the validation set and the test set by a utility class StandardScaler in scikit-learn. Concerning modeling with PCA features, the standard preprocessing was also applied.

According to self-designed AE, 32 dimensionalities features of HSI spectra extracted by AE were fed into SVM and LR models. Before being fed into SVM and LR, features of the training set, the validation set, and the test set were firstly standardized by the StandardScaler as mentioned before. This standard preprocessing was also applied to AE features of MIR and LIBS. In terms of building CNN, data without the preprocessing of StandardScaler was found to be helpful to achieve better performance. Therefore, each source of spectra, PCA features, and AE features were directly fed into CNN. In terms of features of different levels of fusion, Standardscaler transformation was applied on merged datasets to compensate for the scale differences before feeding these features into classifiers.

The spectra extraction of HSI, MIR, and LIBS were conducted on MATLAB R2015b (The Math Works, Natick, MA, United States). To evaluate the model performances, classification accuracy was used, which was the ratio of the correctly classified number of samples and the total sample number. Deep learning was conducted by python3 with MXNET framework (Amazon, Seattle, WA, United States) with GPU acceleration. A computer with Intel Core-i7 8700k CPU, NVidia GTX1060 GPU, 16 GB RAM, and 256 GB SSD was used for calculation.

RESULTS

Spectral Profiles

After standardization preprocessing for each dataset, the average spectra for both rice cultivars were plotted for visualization. **Figures 5A,B** shows the average HSI spectra of healthy and infected leaves of the two rice cultivars. **Figures 5C,D** presents the average MIR spectra of healthy and infected leaves of the two rice cultivars. **Figures 5E,F** shows the average LIBS spectra of healthy and infected leaves of the two rice cultivars. The differences between HSI spectra, MIR spectra, and LIBS spectra could be observed intuitively.

Concerning HSI spectra, there was some noise at the beginning and end of spectra due to the response of the spectroscopic instrument, which made a difference among curves of leaves in four health conditions. Therefore, this range of spectra is out of the discussion. The spectra among healthy leaves and leaves suffering from three different diseases presented an obvious difference in the range of 530–650 nm. The average spectra of healthy leaves were distinguishingly higher than leaves of DWB and WKB around the spectral region of 530–650 nm. However, some overlaps existed among the spectral curves of leaves under different conditions. Therefore, it required further study to make a better distinction.

There were obvious differences among the MIR spectra of leaves in four different health conditions as well. In the range of 500–1,500 cm^{-1}, the spectra of healthy leaves of the Zhefujing83 had higher values than the leaves infected with DWB and BYK. On the contrary, the spectra of healthy rice leaves of the Zhefujing83 had lower values than the leaves infected with DWB and BYK around 2,000–2,750 nm. In terms of the AD516, the tendency of MIR spectra was similar.

As for LIBS spectra, each peak of LIBS spectra was related to the specific element, which would vary with the change of environmental and intrinsic factors (such as the health condition). With respect to Zhefujing83, the peak intensity of leaves infected by DWB and BYK were smaller than that of healthy leaves. In terms of AD516, the peak intensity of leaves with WKB, DWB and BYK were all smaller than that of healthy leaves.

Classification Models Using Each Single Spectroscopic Datasets

The results of the classification models using each single spectroscopic datasets are shown in **Table 2**. Among the three original spectroscopic datasets, the highest prediction accuracy was obtained from HSI spectral data of two rice varieties. CNN model with HSI data of Zhefujing83 could achieve the accuracy of 100% for the test set that was better than 90.38% (SVM) and 98.08% (LR). As for AD516, both CNN and LR models using HSI data achieved the accuracy of 100% for the training set, the validation set, and the test set. Furthermore, the confusion matrix of CNN based on the full HSI spectra of AD516 is illustrated in **Supplementary Figure 20**. The confusion matrix indicated the three diseases were separable based on the spectral data.

In terms of the results of MIR, both CNN and LR using MIR data of Zhefujing83 obtained the accuracy of 82.69% for the test set, which was higher than SVM (79.69%). As for AD516, LR obtained the best performance, with the accuracy of 96.88% for the test set.

Based on Full-LIBS spectra of Zhefujing83, CNN (the accuracy of 86.54% for the test set) obtained a higher accuracy of 15.39 and 19.23% for the test set than the corresponding SVM (71.15%) and LR (67.31%), respectively. CNN with Full-LIBS spectra of AD516 were inferior to the corresponding LR but superior to the corresponding SVM. This might result from the severity of diseases varying with rice variety.

Classification Models Based on Feature Extraction

PCA-Based Classification

The results of models based on features extracted by PCA are shown in **Table 3**. Among PCA features from three spectroscopic instruments, the highest prediction accuracy was obtained from PCA features of HSI spectra for both rice varieties.

In terms of HSI, the first 14 PCs explaining 99.99% of the information in original spectra were adopted to modeling. For two rice varieties, the accuracy of the training set of SVM and LR models were all over 98%, while the accuracy of the validation set and the test set exceeded 96%. CNN with PCA features of HSI spectra obtained the accuracy over 93% for both varieties. Overall, models based on PCA features achieved better performance than models based on full spectra.

With respect to MIR, the number of PCs for two rice varieties was not identical in the condition of the same accumulative explained variance. The results of the training set and validation set all reached 100% based on both SVM and LR models, and the results of the test set were all

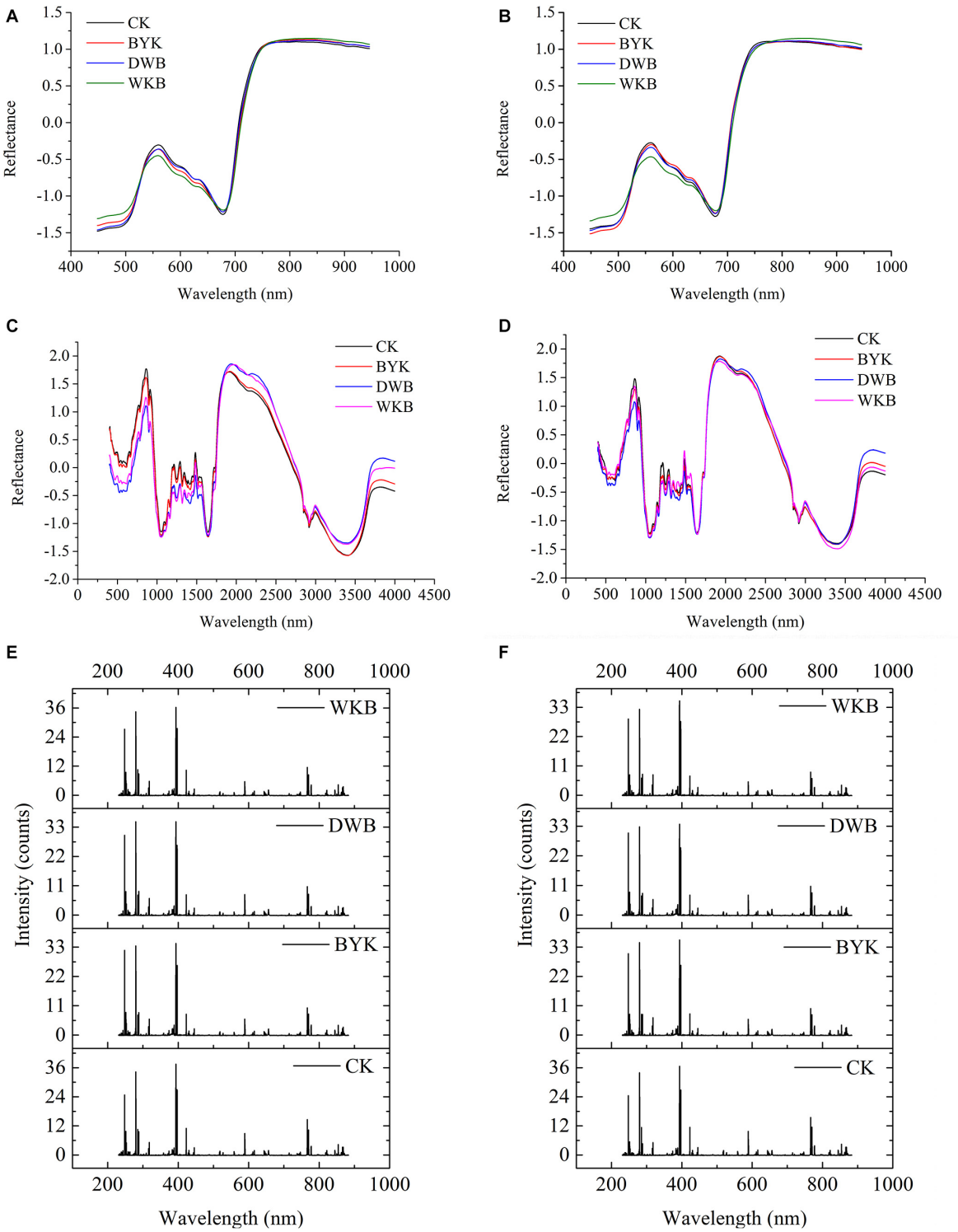

FIGURE 5 | The average spectra of two varieties of rice leaves in different health conditions. **(A)** HSI spectra of Zhefujing83; **(B)** HSI spectra of AD516; **(C)** MIR spectra of Zhefujing83; **(D)** MIR spectra of AD516; **(E)** LIBS spectra of Zhefujing83; **(F)** LIBS spectra of AD516.

above 90%. CNN obtained the accuracy of 90.38% for the test set as well. Compared with the results based on Full-MIR spectra, the performance of the three models was improved overall. This indicted PCA was not only helpful in reducing the dimension of data but also contributed to improving classification accuracy.

TABLE 2 | The classification results based on multisource spectra.

Cultivar	Features	Variables	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)
Zhefujing83	Full-HSI	390	SVM	95.00	90.00	90.38	LR	98.33	95.00	98.08	CNN	**98.33**	**95.00**	**100**
	Full-MIR	7,468		100	100	76.92		**100**	**100**	**82.69**		100	95.00	**82.69**
	Full-LIBS	22,036		100	100	71.15		100	100	67.31		**100**	**100**	**86.54**
AD516	Full-HSI	390		100	100	96.88		**100**	**100**	**100**		**100**	**100**	**100**
	Full-MIR	7,468		99.17	90.00	85.94		**100**	**100**	**96.88**		100	90.00	81.25
	Full-LIBS	22,036		100	100	75.00		**100**	**100**	**87.50**		100	100	81.25

Other abbreviations are similar to this fashion. Tr, training set; Val, validation set; Te, test set; Full-HSI, full spectra of HIS.

TABLE 3 | The classification results based on PCA/AE features of different spectral sources.

Cultivar	Features	Variables	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)
Zhefujing83	PCA-HSI	14	SVM	98.33	95.00	100	LR	**98.33**	**100**	**100**	CNN	100	95.00	94.23
	PCA-MIR	34		**100**	**100**	**90.38**		**100**	**100**	**90.38**		100	95.00	90.38
	PCA-LIBS	114		100	85.00	63.46		100	85.00	63.46		**100**	**95.00**	**78.85**
	AE-HSI	32		**98.33**	**70.00**	**76.92**		98.33	70.00	65.38		98.33	80.00	63.4
	AE-MIR	16		**94.17**	**80.00**	**84.38**		96.67	95.00	76.92		100	100	78.85
	AE-LIBS	64		99.17	95.00	63.46		99.17	90.00	65.38		**100**	**100**	**69.23**
AD516	PCA-HSI	14		100	100	96.88		**100**	**100**	**100**		100	100	93.75
	PCA-MIR	42		**100**	**100**	**96.88**		100	100	95.31		100	95.00	78.13
	PCA-LIBS	103		100	90.00	81.25		100	95.00	81.25		**95.83**	**95.00**	**85.94**
	AE-HSI	32		94.17	75.00	81.25		**97.5**	**95.00**	**84.38**		95.00	85.00	84.38
	AE-MIR	16		**96.67**	**80.00**	**87.50**		95.00	85.00	84.38		99.17	85.00	78.13
	AE-LIBS	64		100	90.00	73.44		**100**	**95.00**	**81.25**		100	95.00	73.44

Other abbreviations are similar to this fashion. Tr, training set; Val, validation set; Te, test set; PCA-HSI, the features of HSI extracted by PCA; AE-HSI, the features of HSI extracted by AE.

As for LIBS, the number of PCs for two rice varieties was also not identical for a similar reason. After parameter optimization, classification results of SVM and LR for the training set were all 100% for both varieties, and the prediction results of AD516 were 81.25%, while prediction results of the Zhefujing83 were 63.46%. These results revealed that the variety of rice might have an influence on rice leaf disease discrimination. CNN model based on PCA-LIBS of Zhefujing83 obtained the accuracy of 100 and 95% for the training set and the validation set, respectively, which were better than corresponding SVM and LR models. The accuracy for the test set increased from 63.46% (SVM and LR) to 78.85% (CNN). The change trend of results of CNN for the AD516 was similar to the Zhefujing83.

Compared with the results based on Full-HSI spectra, the accuracy of models based on PCA features was improved but with only 14 dimensionalities of features. This contributed to saving computing resources and computing time. In addition, the results based on PCA features of Full-MIR data were obviously improved than the results based on Full-MIR data, while the dimensionality of data was decreased from 7,468 to 34 for Zhefujing83 and 42 for AD516. The results based on Full-LIBS data was overall slightly better than the results based on PCA-LIBS. However, the models with PCA-LIBS only used less than 114 dimensionalities (114 for Zhefujing83 and 103 for AD516) of data, while the former used 22,036 dimensionalities of Full-LIBS data. Overall, PCA was an effective tool for feature extraction.

Autoencoder-Based Classification

As shown in **Table 3**, the accuracy of the test sets of two rice varieties all exceeded 75% by SVM models using AE features. Meanwhile, LR only achieved the accuracy of 84.38% for the test set of AD516 but obtained only 63.58% for Zhefujing83. CNN based on AE-HSI obtained the accuracy of 67.31% for the test set of Zhefujing83 and obtained 79.69% for AD516.

In terms of AE-MIR, the performance of the SVM model was superior to the LR model after parameter optimization, with the accuracy surpassing 94% for the training set, reaching 80% for the validation set, and all the accuracy over 84% for the test set. Compared with the results based on Full-MIR spectra, SVM based on features from AE performed slightly better. Meanwhile, LR based on AE-MIR obtained worse classification results than LR based on Full-MIR. For Zhefujing83, CNN based on AE-MIR obtained the accuracy of 100% for both the training set and the validation set. However, it only obtained 75% for the test set. Besides, CNN obtained 78.13% for the test set of AD516. It indicated CNN had limited power on the small dataset (only 30 samples per category of rice). Besides, the results based on AE were inferior to the results based on PCA. The features extracted from AE had no more than half of the dimensionalities of features extracted by PCA.

Considering the massive dimensionalities of the Full-LIBS spectra, 64 dimensionalities of features were extracted for classification models. The performance of SVM and LR models using AE-LIBS were similar, and both models obtained the

accuracy surpassing 99% for the training set and the accuracy surpassing 90% for the validation set for both rice varieties. CNN (71.88% for the test set) based on AE-LIBS was inferior to the corresponding SVM (73.44% for the test set) and LR (81.25% for the test set). Besides, all three classifiers obtained a higher accuracy for AD516 than Zhefujing83. It indicated that the variety of rice influenced feature extraction.

Overall, the performance of models varied with the feature extraction methods. This might attribute to that different feature extraction methods learned and represented different aspects of the features of the original data. Integrating features extracted by different methods might contribute to improving classification performance. Thus, feature fusion was conducted for classification.

Classification Models Based on Data Fusion

Low-Level Fusion

In the low-level fusion approach, spectra from different sources were directly concatenated. After parameter optimization, the results of SVM, LR, and CNN are presented in **Table 4**.

The combination of HSI and MIR helped to improve classification accuracy by about 4–6% compared with Full-MIR spectra. Moreover, models based on HSI-LIBS fusion data obtained better performance than models based on the Full-LIBS spectra. However, the accuracy based on HSI-MIR and HSI-LIBS declined compared with the results only based on the HSI data. The reason might attribute to the irrelevant information contained in full MIR and LIBS spectra. Fused data of individual spectra had a higher dimension and contained more irrelevant information, which interfered with the discriminative power of classifiers. In addition, compared with Full-MIR, the combination of Full-MIR and Full-LIBS only helped to improve the accuracy of SVM for the Zhefujing83. In contrast, the accuracy of SVM for AD516 and the accuracy of LR models for both rice varieties had all slightly declined.

In terms of Zhefujing83, CNN based on HSI-LIBS achieved an accuracy of 90.38% that was higher than 76.92% (SVM) and 69.23% (LR). For MIR-LIBS, CNN obtained an accuracy of 86.54% for the test set that was around 6% higher than the corresponding LR. CNN based on MIR-LIBS (with the accuracy of 86.54% for the test set) was better than CNN based on Full-MIR alone (with 82.69% for the test set). CNN based on HSI-MIR-LIBS performed better than the corresponding SVM and LR. Besides, CNN based on HSI-MIR-LIBS obtained the accuracy of 92.31% for the test set, which were higher than both CNN based on MIR (82.69%) and CNN based on LIBS (86.54%). In terms of AD516, though CNN achieved the accuracy of 100% for both the training set and the validation set despite the data sources, CNN obtained lower accuracy for the test set when compared with SVM and LR.

In summary, the accuracy based on low-level fusion was not improved, but the dimensionalities of the input data were multiplied. In consequence, the computing resources and computing time were increased. Therefore, the low-level fusion strategy was not efficient enough for classification improvement in this study.

Mid-Level Fusion

Fusion of AE Features and PCA Features of Each Spectral Dataset

In the case of mid-level fusion, the informative features separately extracted by AE and PCA were concatenated into a single matrix that was further used for multivariate analysis. The results are shown in **Table 5**. Satisfactory results were obtained with feature fusion analysis.

In terms of HSI, both the Zhefujing83 and the AD516 have 46 variables after concatenation. With respect to Zhefujing83, in contrast with the model performance of the AE-based model, the accuracy for the training set, the validation set, and the test set was improved after data fusion, which was not much different from the results of the PCA feature-based model. The accuracy of SVM based on AE-PCA-HSI improved to 98.33, 100 and 98.08% for the three datasets (training, validation, and test), respectively, when compared with SVM based on AE-HSI (98.33, 70, and 76.92%). The accuracy of LR based on AE-PCA-HSI improved to 98.33, 100, and 100% for the three datasets, respectively, when compared with LR based on AE-HSI (98.33, 70, and 65.38%). The accuracy of CNN based on AE-PCA-HSI improved to 98.33, 85, and 90.38% for the three datasets, respectively, when compared with CNN based on AE-HSI (99.17, 80, and 67.31%). The changing trend of results was similar for AD516. This indicted combining features extracted

TABLE 4 | The accuracy rates of classification models based on fusion data.

Cultivar	Features	Variables	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)
Zhefujing83	HSI-MIR	7,858	SVM	100	95.00	84.62	LR	**100**	**100**	**86.54**	CNN	100	95.00	84.62
	HSI-LIBS	22,426		100	100	76.92		100	100	69.23		**100**	**100**	**90.38**
	MIR-LIBS	29,504		**100**	**100**	**86.54**		100	100	80.77		**100**	**100**	**86.54**
	HSI-MIR-LIBS	29,894		100	100	90.38		100	100	82.69		**100**	**100**	**92.31**
AD516	HSI-MIR	7,858		100	100	92.19		**100**	**100**	**100**		100	95.00	85.94
	HSI-LIBS	22,426		100	100	81.25		**100**	**100**	**90.63**		100	100	84.38
	MIR-LIBS	29,504		100	100	84.38		**100**	**100**	**92.19**		100	100	84.38
	HSI-MIR-LIBS	29,894		100	100	85.94		**100**	**100**	**92.19**		100	100	81.25

Other abbreviations are similar to this fashion. Tr, training set; Val, validation set; Te, test set; HSI-MIR, the concatenation of full HSI and MIR spectra; HSI-LIBS, the concatenation of full HSI and LIBS spectra; MIR-LIBS, the concatenation of full MIR and LIBS spectra; HSI-MIR-LIBS, the concatenation of full HSI, MIR, and LIBS spectra.

TABLE 5 | The classification accuracy based on fusion data of AE and PCA features of different spectral sources.

Cultivar	Features	Variables	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)
Zhefujing83	AE-PCA-HSI	46	SVM	98.33	100	98.08	LR	**98.33**	**100**	**100**	CNN	98.33	85.00	90.38
	AE-PCA-MIR	50		100	100	86.54		**100**	**100**	**88.46**		100	100	86.54
	AE-PCA-LIBS	178		100	90.00	61.54		99.17	95.00	67.31		**100**	**95.00**	**73.08**
	PCA-HSI-MIR	48		100	100	96.15		**100**	**100**	**98.08**		100	95.00	88.46
	PCA-HSI-LIBS	128		100	95.00	96.15		**100**	**100**	**100**		100	85.00	63.46
	PCA-MIR-LIBS	148		**100**	**100**	**86.54**		**100**	**100**	**86.54**		100	95.00	48.08
	PCA-HSI-MIR-LIBS	162		**100**	**100**	**96.15**		**100**	**100**	**96.15**		100	100	80.77
	AE-PCA-HSI-MIR	96		100	100	94.23		**100**	**100**	**98.08**		100	95.00	84.62
	AE-PCA-HSI-LIBS	224		**100**	**95.00**	**75.00**		**100**	**95.00**	**75.00**		**100**	**95.00**	**75.00**
	AE-PCA-MIR-LIBS	228		100	100	88.46		**100**	**100**	**90.38**		100	100	86.54
	AE-PCA-HSI-MIR-LIBS	274		**100**	**100**	**92.31**		**100**	**100**	**92.31**		100	100	78.85
AD516	AE-PCA-HSI	46		100	100	92.19		**100**	**100**	**100**		100	95.00	92.19
	AE -PCA-MIR	58		100	100	95.31		**100**	**100**	**96.88**		100	95.00	84.38
	AE -PCA-LIBS	167		100	85.00	79.69		**100**	**100**	**84.38**		100	95.00	73.44
	PCA-HSI-MIR	56		100	100	98.44		**100**	**100**	**100**		100	95.00	90.63
	PCA-HSI-LIBS	117		**100**	**100**	**98.44**		**100**	**100**	**98.44**		100	95.00	87.5
	PCA-MIR-LIBS	145		**100**	**100**	**85.94**		**100**	**100**	**85.94**		100	95.00	85.94
	PCA-HSI-MIR-LIBS	159		100	100	95.31		**100**	**100**	**98.44**		100	100	87.5
	AE-PCA-HSI-MIR	104		100	100	98.44		**100**	**100**	**100**		100	100	85.94
	AE-PCA-HSI-LIBS	213		100	100	82.81		**100**	**100**	**90.63**		100	95.00	73.44
	AE-PCA-MIR-LIBS	225		100	100	79.69		**100**	**100**	**85.94**		100	100	78.13
	AE-PCA-HSI-MIR-LIBS	271		100	100	82.81		**100**	**100**	**100**		100	90.00	78.13

Other abbreviations are similar to this fashion. Tr, training set; Val, validation set; Te, test set; AE-PCA-HSI, the concatenation of PCA features and AE features of HSI; PCA-HSI-MIR, the concatenation of PCA features of HSI and PCA features of MIR; AE-PCA-HSI-MIR, the concatenation of AE-PCA features of HSI and AE-PCA features of MIR.

by different feature extraction methods was helpful to improve classification performance.

In the case of MIR, the performance of models based on this fusion was better than those based on the AE-based model. Concerning Zhefujing83, the results of SVM with AE-PCA-MIR were improved to 100, 100, and 86.54% for the three datasets, respectively, when compared with SVM based on AE-MIR (94.17, 80, and 84.38%). The accuracy of LR with AE-PCA-MIR was improved to 100, 100, and 88.46% for the three datasets, respectively, when compared with LR based on AE-MIR (96.67, 95, and 76.92%). The accuracy of CNN based on AE-PCA-MIR was improved to 100, 100, and 86.54% for the three datasets, respectively, when compared with CNN based on AE-MIR (100, 100, and 75%). The change trend of results was similar with respect to AD516.

In the case of LIBS, compared with the accuracy based on PCA features, the accuracy based on the fusion data were decreased by about 2% for Zhefujing83, which were increased by about 3% for AD516. Overall, integrating features extracted by different feature extraction methods was helpful for better classification performance.

Fusion of Features from Three Spectral Datasets

From the section "Classification Models Based on Feature Extraction," we found that the models based on features extracted from PCA had the best performance compared with models based on the AE features. Thus, the PCA features extracted from three spectroscopic data were further fused for classification. The results are shown in **Table 5**. SVM and LR were carried out on the fused features using the previously used parameter optimization method.

In terms of the Zhefujing83, on the one hand, in contrast with the results based on PCA-HSI, the accuracy for the training set and the validation set increased to 100% with both SVM and LR models based on the integration of PCA features of HSI and MIR. The accuracy of the test set was 96.15% for SVM and 98.08% for LR. On the other hand, compared with results based on PCA features of MIR, the accuracy for the test set respectively increased by 5.77 and 7.7% for SVM and LR based on integrated features of HSI and MIR.

As for AD516, the models based on PCA-HSI-MIR obtained the best performance among all combinations, followed by PCA-HSI-LIBS, PCA-HSI-MIR-LIBS and PCA-MIR-LIBS. In contrast with the corresponding models using single source of PCA features, the LR based on PCA-HSI-MIR achieved the classification accuracy of 100% for the training set, the validation set, and the test set, which were identical with LR based on PCA-HSI and were better than the LR based on PCA-MIR. Moreover, SVM based on PCA-HSI-MIR obtained good results for the training set, the validation set, and the test set, with the classification accuracy of 100, 100, and 98.44%, respectively. SVM based on PCA-HSI-MIR were better than SVM based on PCA-HSI. Besides, SVM based on PCA-HSI-MIR were better than SVM based on PCA-MIR as well. CNN based on PCA-HSI-MIR

achieved 100, 95, and 90.63% for the training set, the validation set, and the test set, respectively.

However, integrating PCA-MIR with PCA-LIBS deteriorated the performance of three kinds of models; the accuracy was 5–10% lower than models based on PCA-MIR. Beyond that, the results based on PCA feature fusion were close to even better than the results based on a single source of PCA features.

Except for integrating PCA features from three kinds of spectra, we further fused the AE-PCA fusion data of different data sources in the previous section. The results are shown in **Table 5**. Before being fed to SVM and LR, the fusion data were preprocessed by the StandaradScaler as mentioned before.

For Zhefujing83, in contrast with models based on AE-PCA features of MIR, the classification accuracy of the SVM and LR based on AE-PCA-MIR-LIBS was improved by about 2%. The accuracy increased more when compared with models based on the AE-PCA features of LIBS. Besides, SVM and LR based on this kind of fusion had better performance than models based on full spectra overall for both rice varieties. Among all combinations, the performances of models in descending order were as follows: AE-PCA-HSI-MIR, AE-PCA-HSI-MIR-LIBS, AE-PCA-MIR-LIBS, and AE-PCA-HSI-LIBS. CNN obtained the best results using PCA-HSI-MIR, with the accuracy of 100, 95, and 88.46% for the training set, the validation set, and the test set, respectively.

For AD516, in contrast with the best SVM and LR based on the AE-PCA feature from a single data source, the integration of AE-PCA features of HSI and MIR contributed to obtaining better classification results. Besides, the integration of AE-PCA features of three spectra was helpful to obtain the accuracy of 100% for three datasets with LR. CNN based on this kind of fusion did not exhibit good enough performance, though CNN obtained the accuracy of 100% and over 95% for the training set and the validation set, respectively. Among all combinations, CNN using PCA-HSI-MIR obtained the highest accuracy for the test set (90.38%).

In summary, this section studied the feasibility of CNN to classify rice leaves in four conditions with AE-PCA features. The results were inferior to SVM and LR overall. Considering the number of samples in the training set and the validation set being 120 and 20, respectively, CNN may not be able to exhibit its power for classification for lack of enough samples. However, the results could indicate the great potential of deep learning based approaches for rice disease detection. More samples were needed in future studies to fully reveal the advantage of deep learning.

High-Level Fusion

The high-level fusion strategy based on majority voting was applied to classification results (**Tables 2** and **3**) obtained by classification models based on a single source of spectra. High-level fusion was applied to full spectra, PCA features, and AE features, respectively. A classifier developed on one specific analytical data made its own predictions. With these predictions from different sources of data, the final decisions of high-level fusion were calculated according to a majority of vote rule. Classification results achieved by using the combination of all three sources of full spectra, the combination of all three sources of PCA features, and the combination of all three sources of AE features are listed in **Table 6**.

On the one hand, SVM based on high-level fusion using full spectra of Zhefujing83 obtained the accuracy of 100 and 100% for the training set and the validation set, which were higher than SVM based on Full-HSI spectra (95 and 90%), with the accuracy of 90.38% for the test set of the two SVM models. CNN based on high-level fusion obtained the classification accuracy of 100% for the training set, the validation set, and the test set, which was better than the corresponding CNN based on Full-HSI (98.33, 95, and 100% for the training set, the validation set, and the test set, respectively), Full-MIR (100, 95, and 82.69%) and Full-LIBS (100, 100, and 86.54%). On the other hand, in terms of AD516, the accuracy of SVM for the test set declined to 93.75% after high-level fusion, with the accuracy for the training set and the validation set being the same as 100%. Fusion results of CNN only obtained the accuracy of 87.5% for the test set since both CNN based on Full-MIR and Full-LIBS only achieved 81.25% for the test set. Besides, compared with the accuracy of SVM models based on Full-MIR, the accuracy for the test set was improved from 76.92 to 90.38% for Zhefujing83 and from 85.94 to 93.75% for AD516. When compared with the accuracy of SVM models based on Full-LIBS, the accuracy for the test set was improved by over 18% after high-level fusion.

For high-level fusion of classifiers based on PCA features, the accuracy for both the training set and the validation set was increased to 100% compared with the accuracy based on PCA features of a single type of spectra. In terms of Zhefujing83, SVM based on high-level fusion obtained the accuracy of 98.08% for the test set, which exceeded the corresponding accuracy of

TABLE 6 | The classification accuracy rate based on high-level fusion.

Feature type	Rice cultivar	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)	Model	Tr (%)	Val (%)	Te (%)
Full	Zhufujing83	SVM	100	100	90.38	LR	100	100	88.46	CNN	**100**	**100**	**100**
	AD516		100	100	93.75		**100**	**100**	**98.44**		100	100	87.50
PCA features	Zhufujing83		**100**	**100**	**98.08**		**100**	**100**	**98.08**		100	100	96.15
	AD516		100	100	96.88		**100**	**100**	**98.44**		100	100	90.63
AE features	Zhufujing83		**100**	**100**	**80.77**		99.17	90.00	75.00		100	100	73.08
	AD 516		**100**	**85.00**	**90.63**		100	95.00	89.06		100	90.00	85.94

Tr, training set; Val, validation set; Te, test set; Full, Full spectra; PCA features, featuers extracted by principal component analysis; AE features, features extracted by autoencoder.

SVM based on PCA-MIR and PCA-LIBS. Besides, LR based on high-level fusion obtained the accuracy of 98.08% for the test set, which was much higher than 90.38 and 63.46% of LR based on PCA-MIR and PCA-LIBS, respectively. CNN based on high-level fusion obtained the accuracy of 96.15% for the test set, which were much higher than 94.23, 90.38, and 78.13% of CNN based on PCA-HSI, PCA-MIR, and PCA-LIBS, respectively. With respect to AD516, SVM based on high-level fusion obtained an accuracy of 96.88%, which was 15.63% more than the SVM model based on PCA-LIBS. In addition, LR obtained the accuracy of 98.44% for the test sets, which were higher than 95.31 and 81.25% of LR based on PCA-MIR and PCA-LIBS, respectively. Besides, CNN obtained the accuracy of 90.63% for the test set, which was higher than 78.85 and 85.94% of CNN based on PCA-MIR and PCA-LIBS, respectively.

In terms of high-level fusion of classifiers based on AE features, the SVM and LR models after high-level fusion were all better than models based on individual type of spectra. In terms of Zhefujing83, the accuracy of SVM was increased to 100% for both the training set and the validation set after data fusion, and the accuracy of the test set after high-level fusion was 3.85 and 17.31% higher than that base on AE-HSI and AE-LIBS, respectively. Besides, the improvement trend of the accuracy of SVM and LR for AD516 complied with that of Zhefujing83. For Zhefujing83, CNN based on high-level fusion obtained the accuracy of 73.08% for the test set, which were 5.77 and 17.31% higher than CNN based on AE-HSI and AE-LIBS, respectively. For AD516, CNN based on high-level fusion obtained the accuracy of 85.94% for the test set, which were improved when compared with CNN based on AE-HSI (79.69% for the test set), CNN based on AE-MIR (78.13% for the test set), and CNN based on AE-LIBS (71.88% for the test set).

In all, classification performances based on the high-level fusion approach were slightly better than those based on one single analytical source.

DISCUSSION

In this study, two different rice varieties were used to verify the proposed methods in this article being effective despite rice varieties. According to **Tables 2** and **4,** classification accuracy between Zhefujing83 and AD516 was different in general. Overall, LR and SVM obtained higher prediction accuracy for AD516 than those for Zhefujing83. Besides, CNN models for Zhefujing had a better overall performance than those for AD516. There existed a variance among the performance of different models using different datasets of different varieties of rice. The deep reasons for the varietal variances would be further investigated in future studies with more samples with physiological and biochemical analyses.

In terms of data fusion strategies, low-level fusion directly integrates the original data, so it has an immense data calculation. In this study, the original HSI, MIR, and LIBS spectra had 390, 7,468, and 22,036 dimensionalities of features, respectively. After concatenating every two of them, the dimensionalities of new data would greatly increase. This would increase the computing time. Apart from the high dimension of input, the limited number of samples in the training set (only 30 samples per category) restricted the performance of CNN.

Besides, the increase in information brought by low-level fusion may not compensate for irrelevant or spurious variance brought by this fusion strategy (Biancolillo et al., 2014). The low-level fusion has some limitations which are a high data volume and the possible predominance of one data source over the others (Borras et al., 2015). In our case, models based on full HSI spectra had very satisfying classification results. However, after concatenating full HSI with full MIR or full LIBS, the accuracy of models has declined slightly when compared with results based on full HSI but increased obviously when compared with results based on full MIR or full LIBS. That indicated full HSI spectra had predominance over the other data sources. It should be addressed that the different sources of data can have a very different scale, and the appropriate preprocessing is of great importance before establishing models. Besides, there can exist some redundant information when using different instruments. In these conditions, it is critical to preprocess the raw data before data fusion, and sometimes fusion of fewer techniques might be able to obtain satisfactory results.

Mid-level fusion can partially overcome the high-data–volume problem. The data dimensionality could be significantly reduced with feature extraction methods. Besides, this fusion strategy is helpful to filter individual instrument noise and enable the interpretation of the results because of the fewer dimensionalities of inputs (Borras et al., 2015). Berdugo et al. (2014) adopted mid-level fusion with at least one feature per sensor among hyperspectral imaging, thermography, and chlorophyll fluorescence to detect cucumber disease. The most discriminant features from thermography and chlorophyll fluorescence had limited power to identify or differ plant diseases or abiotic stress. However, hyperspectral imaging was good at assessing disease-specific changes. Therefore, the features extracted from different instruments were cooperative, and the fusion of these features was helpful to filter noise existing in each instrument and obtain more complete information.

Furthermore, the fewer features within mid-level fusion were helpful to develop a real-time disease detection system. Moshou et al. (2005) assessed the real-time implementation of the Self-Organizing Map (SOM) neural network to detect wheat disease. The three selected spectral reflectance values were further fused with one fluorescence feature. The SOM classifier based on the fusion data achieved the overall classification accuracy of around 99%, which was higher than using one fluorescence feature. Besides, the fusion of features from different instruments may cause the issue of redundant information. To detect grape leaf disease, Adeel et al. (2019) implemented canonical correlation analysis for feature fusion and further performed neighborhood correlation analysis to reduce the dimensionalities and redundant information of the fused data before feeding the fused data into the classifier. This strategy helped to achieve an accuracy of 94.1% that was superior to the existing methods. In terms of disadvantages, mid-level fusion requires a preliminary feature extraction stage. Besides, taking account the many combinations of feature extraction methods

and preprocessing, testing all the combinations makes the whole process cumbersome, computationally intensive, and difficult to validate (Borras et al., 2015).

Lastly, high-level fusion is operated with the classification results of individual classifiers. These separate models are developed based on the data of different instrumental techniques. Their predictions are then integrated into a single final response. Through a majority vote, a sample is assigned to the most-frequently predicted category. In this study involving four categories and three instrumental data, if the sample was predicted as class 1 by the two of the three classifiers, it would be assigned to class 1. Moreover, there are some other more complex protocols, such as Bayesian statistics (Biancolillo et al., 2014), which can be applied for decision making of high-level fusion. Concerning this type of fusion, every individual instrument is treated independently. Therefore, the responses from inefficient techniques (like LIBS spectra in this article) do not worsen the overall performance. Moreover, it is easy to add new techniques for final decision making when a new type of data is available. This increases the versatility of the decision-making process.

To better understand the differences across different methods, the ANOVA analysis was carried out. Diagnostic analysis among models was mainly discussed here, and the variance of rice variety and spectroscopic techniques was not considered. The influence of the type of input (full spectra, PCA features, and AE features) on classification results of CNN was analyzed. Besides, the difference of CNN based on different levels of fusion was also analyzed. On the one hand, the ANOVA analysis was performed on results based on full spectra (including Full-HSI, Full-MIR, and Full-LIBS), PCA features (including PCA-HSI, PCA-MIR, and PCA-LIBS), and AE features (including AE-HSI, AE-MIR, and AE-LIBS). The analysis results are summarized in **Supplementary Table 1** (based on classification results of the training set), **Supplementary Table 2** (based on classification results of the validation set), and **Supplementary Table 3** (based on classification results of the test set), respectively. Since all CNN models achieved the accuracy of about 100% on the training set, all p values in **Supplementary Table 1** were greater than 0.2, which indicated there was no obvious difference between classification results based on full spectra and those based on features. Similar results could also be found in **Supplementary Table 2**. However, **Supplementary Table 3** showed that the significance value ($p = 0.008$) between classification results based on full spectra and classification results based on AE features was smaller than 0.05, which suggested classification results based on these two sources of data had significant differences. On the other hand, the ANOVA analysis was carried out on a different level of fusion, including nonfusion, low-level fusion, mid-level fusion, and high-level fusion. The ANOVA results are summarized in **Supplementary Table 4** (based on classification results of the training set), **Supplementary Table 5** (based on classification results of the validation set), and **Supplementary Table 6** (based on classification results of the test set), respectively. The analysis results based on the training set and the validation set revealed most p values across two different groups were greater than 0.25, which suggested different levels of fusion had little effect on classification results. Besides, a p value equaled to 0.063 in **Supplementary Table 6**, which indicated the difference between classification results based on nonfusion and results based on mid-level fusion was not significant at the $\alpha = 0.05$ level but was significant at the $\alpha = 0.1$ level. In addition, another p value equaled to 0.055 in **Supplementary Table 6**, which indicated the difference between classification results based on mid-fusion and results based on high-level fusion was significant at the $\alpha = 0.1$ level as well.

CONCLUSION

In this study, HSI, MIR, and LIBS were applied to detect rice leaves inoculated by different diseases. Models based on full HSI spectra had the best performance among three full spectra. Based on full HSI, SVM, LR, and CNN obtained the accuracy of 90.38, 98.08, and 100% for the test set, respectively. PCA was an effective tool to extract key information. All three classifiers based on PCA-HSI obtained 94% accuracy for the test set. Besides, as part of the deep learning framework, AE was proved to be effective to extract features and reduce data dimension. Three kinds of data fusion strategies were explored for classification. The low-level fusion strategy was the least effective among the three fusion strategies due to the huge dimensions of fused data. Combined with appropriate feature extraction methods, the mid-level fusion exhibited better performance when compared with nonfused data. By integrating the PCA features of HSI and the PCA features of MIR, LR achieved an accuracy of over 98% for both rice varieties. Besides, it took less time to model with features. Overall, decision level fusion was a good way to avoid the limitation of decision making based on a single kind of classifier. In terms of the high-level fusion of classifiers based on full spectra, compared with the accuracy of SVM models based on full MIR, the accuracy of the test set after fusion was improved from 76.92 to 90.38% for Zhefujing83 and from 85.94 to 93.75% for AD516. Concerning high-level fusion of classifiers based on PCA features, the accuracy of both the training set and the validation set was increased to 100% compared with the accuracy based on PCA features of a single type of spectra. In terms of the high-level fusion of AE features, the accuracy of SVM was increased to 100% for both the training set and the validation set after data fusion, and the accuracy for the test set after fusion obtained 3.85 and 17.31% higher than that based on AE-HSI and AE-LIBS, respectively. In this work, CNN did not achieve excellent performance due to the limited number of samples in the training set (only 30 samples per category), but the great potential of CNN for rice diseases detection could be observed. More samples are required to make full use of CNN. More rapid and sensitive analytical techniques are available in industrial processes and laboratories, which will keep promoting advances in data fusion in various fields. There is still room for improvement in different levels of fusion.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LF, YH, and CZ: conceptualization. CZ and BW: data curation. JW and ZS: formal analysis. YH and LF: funding acquisition. CZ and JW: investigation. SZ, FL, ZS, and CZ: methodology. LF, CZ, and YH: project administration. CZ, JW, and ZS: resources. FL and CZ: software. CZ, LF, and SZ: supervision. SZ and CZ: validation. BW and LF: visualization. LF and BW: writing – original draft. CZ and YH: writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by Key Projects of International Scientific and Technological Innovation Cooperation Among Governments Under National Key R&D Plan (2019YFE0103800), XPCC Science and Technology Projects of Key Areas (2020AB005), National Natural Science Foundation of China (61705195), and the Postdoctoral Science Foundation of Zhejiang Province, China (zj2019091).

ACKNOWLEDGMENTS

We would like to thank Zhenyu Qi, an associate research fellow in Zhejiang University for his help on rice plant growth and management.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.577063/full#supplementary-material

REFERENCES

Adeel, A., Khan, M. A., Sharif, M., Azam, F., Shah, J. H., Umer, T., et al. (2019). Diagnosis and recognition of grape leaf diseases: an automated system based on a novel saliency approach and canonical correlation analysis based multiple features fusion. *Sustain. Comput. Inform. Syst.* 24:100349. doi: 10.1016/j.suscom.2019.08.002

Alberto, R. T. (2018). Spectral characterization of bacterial leaf blight of rice through spectroscopy and remotely sensed multi-spectral imagery. *Phytopathology* 108.

Altangerel, N., Ariunbold, G. O., Gorman, C., Alkahtani, M. H., Borrego, E. J., Bohlmeyer, D., et al. (2017). In vivo diagnostics of early abiotic plant stress response via Raman spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 114, 3393–3396. doi: 10.1073/pnas.1701328114

Asaari, M. S. M., Mishra, P., Mertens, S., Dhondt, S., Inze, D., Wuyts, N., et al. (2018). Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform. *ISPRS J. Photogram. Remote Sens.* 138, 121–138. doi: 10.1016/j.isprsjprs.2018.02.003

Ballabio, D., Robotti, E., Grisoni, F., Quasso, F., Bobba, M., Vercelli, S., et al. (2018). Chemical profiling and multivariate data fusion methods for the identification of the botanical origin of honey. *Food Chem.* 266, 79–89. doi: 10.1016/j.foodchem.2018.05.084

Berdugo, C. A., Zito, R., Paulus, S., and Mahlein, A. K. (2014). Fusion of sensor data for the detection and differentiation of plant diseases in cucumber. *Plant Pathol.* 63, 1344–1356. doi: 10.1111/ppa.12219

Biancolillo, A., Bucci, R., Magri, A. L., Magri, A. D., and Marini, F. (2014). Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Anal. Chim. Acta* 820, 23–31. doi: 10.1016/j.aca.2014.02.024

Borras, E., Ferre, J., Boque, R., Mestres, M., Acena, L., and Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment - a review. *Anal. Chim. Acta* 891, 1–14. doi: 10.1016/j.aca.2015.04.042

Carter, G. A., and Knapp, A. K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am. J. Bot.* 88, 677–684. doi: 10.2307/2657068

Castanedo, F. (2013). A review of data fusion techniques. *Sci. World J.* 2013:704504. doi: 10.1155/2013/704504

Diniz, P., Gomes, A. A., Pistonesi, M. F., Band, B. S. F., and de Araujo, M. C. U. (2014). Simultaneous classification of teas according to their varieties and geographical origins by using NIR Spectroscopy and SPA-LDA. *Food Analyt. Methods* 7, 1712–1718. doi: 10.1007/s12161-014-9809-7

Doeswijk, T. G., Smilde, A. K., Hageman, J. A., Westerhuis, J. A., and van Eeuwijk, F. A. (2011). On the increase of predictive performance with high-level data fusion. *Anal. Chim. Acta* 705, 41–47. doi: 10.1016/j.aca.2011.03.025

Farber, C., Mahnke, M., Sanchez, L., and Kurouski, D. (2019). Advanced spectroscopic techniques for plant disease diagnostics. A review. *Trac Trends Analyt. Chem.* 118, 43–49. doi: 10.1016/j.trac.2019.05.022

Feng, L., Zhu, S., Zhou, L., Zhao, Y., Bao, Y., Zhang, C., et al. (2019). Detection of subtle bruises on winter jujube using hyperspectral imaging with pixel-wise deep learning method. *IEEE Access.* 7, 64494–64505. doi: 10.1109/access.2019.2917267

Feng, L., Zhu, S. S., Zhang, C., Bao, Y. D., Gao, P., and He, Y. (2018). Variety identification of raisins using near-infrared hyperspectral imaging. *Molecules* 23:2907. doi: 10.3390/molecules23112907

Gaoqiang, L., Changwen, D., Fei, M., Yazhen, S., and Jianmin, Z. (2020). Responses of leaf cuticles to rice blast: detection and identification using depth-profiling fourier transform mid-infrared photoacoustic Spectroscopy. *Plant Dis.* 104, 847–852. doi: 10.1094/PDIS-05-19-1004-RE

Hawkins, S. A., Park, B., Poole, G. H., Gottwald, T., Windham, W. R., and Lawrence, K. C. (2010). Detection of citrus huanglongbing by fourier transform infrared-attenuated total reflection Spectroscopy. *Appl. Spectrosc.* 64, 100–103. doi: 10.1366/000370210790572043

Knauer, U., Matros, A., Petrovic, T., Zanker, T., Scott, E. S., and Seiffert, U. (2017). Improved classification accuracy of powdery mildew infection levels of wine grapes by spatial-spectral analysis of hyperspectral images. *Plant Methods* 13:47. doi: 10.1186/s13007-017-0198-y

Kumar, A., Kumar, R., Sengupta, D., Das, S. N., Pandey, M. K., Bohra, A., et al. (2020). Deployment of genetic and genomic tools toward gaining a better understanding of rice-*Xanthomonasoryzae pv. oryzae* interactions for development of durable bacterial blight resistant rice. *Front. Plant Sci.* 11:1152. doi: 10.3389/fpls.2020.01152

Lin, Z. H., Chen, Y. H., Zhao, X. and Wang, G. (2013). "Spectral-spatial classification of hyperspectral image using autoencoders", *Proceedings of the 2013 9th International Conference on Information, Communications & Signal Processing*, Tainan. doi: 10.1109/ICICS.2013.6782778

Liu, F., Shen, T., Wang, J., He, Y., Zhang, C., and Zhou, W. (2019). Detection of sclerotinia stem rot on oilseed rape (*Brassica napus* L.) based on laser- induced breakdown spectroscopy. *Trans. Asabe* 62, 123–130. doi: 10.13031/trans.12206

Luo, Y. H., Jiang, P., Xie, K., and Wang, F. J. (2019). Research on optimal predicting model for the grading detection of rice blast. *Opt. Rev.* 26, 118–123. doi: 10.1007/s10043-018-0487-3

Machado, J. C., Faria, M. A., Ferreira, I. M. P. L. V. O., Páscoa, R. N. M. J., and Lopes, J. A. (2018). Varietal discrimination of hop pellets by near and

mid infrared spectroscopy. *Talanta* 180, 69–75. doi: 10.1016/j.talanta.2017.12.030

Molla, K. A., Karmakar, S., Molla, J., Bajaj, P., Varshney, R. K., Datta, S. K., et al. (2020). Understanding sheath blight resistance in rice: the road behind and the road ahead. *Plant Biotechnol. J.* 18, 895–915. doi: 10.1111/pbi.13312

Moshou, D., Bravo, C., Oberti, R., West, J., Bodria, L., McCartney, A., et al. (2005). Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using Kohonen maps. *Real Time Imag.* 11, 75–83. doi: 10.1016/j.rti.2005.03.003

Ng, W., Minasny, B., and McBratney, A. (2020). Convolutional neural network for soil microplastic contamination screening using infrared spectroscopy. *Sci. Total Environ.* 702:134723. doi: 10.1016/j.scitotenv.2019.134723

Peng, J., Liu, F., Zhou, F., Song, K., Zhang, C., Ye, L., et al. (2016). Challenging applications for multi-element analysis by laser-induced breakdown spectroscopy in agriculture: a review. *TrAC Trends Analyt. Chem.* 85, 260–272. doi: 10.1016/j.trac.2016.08.015

Ponce, L., Etxeberria, E., Gonzalez, P., Ponce, A., and Flores, T. (2018). Rapid identification of Huan ongbing-infected citrus plants using laser-induced breakdown spectroscopy of phloem samples. *Appl. Opt.* 57, 8841–8844. doi: 10.1364/ao.57.008841

Ribeiro, L. D., Klock, A. L. S., Wordell, J. A., Tramontin, M. A., Trapp, M. A., Mithofer, A., et al. (2018). Hyperspectral imaging to characterize plant-plant communication in response to insect herbivory. *Plant Methods* 14:54. doi: 10.1186/s13007-018-0322-7

Skolik, P., McAinsh, M. R., and Martin, F. L. (2019). ATR-FTIR spectroscopy non-destructively detects damage-induced sour rot infection in whole tomato fruit. *Planta* 249, 925–939. doi: 10.1007/s00425-018-3060-1

Sun, J., Tang, K., Wu, X. H., Dai, C. X., Chen, Y., and Shen, J. F. (2018). Nondestructive identification of green tea varieties based on hyperspectral imaging technology. *J. Food Process Eng.* 41:e12800. doi: 10.1111/jfpe.12800

Thomas, S., Behmann, J., Steier, A., Kraska, T., Muller, O., Rascher, U., et al. (2018). Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods* 14:45. doi: 10.1186/s13007-018-0313-8

Tremblay, M., Kammer, M., Lange, H., Plattner, S., Baumgartner, C., Stegeman, J. A., et al. (2019). Prediction model optimization using full model selection with regression trees demonstrated with FTIR data from bovine milk. *Prevent. Vet. Med.* 163, 14–23. doi: 10.1016/j.prevetmed.2018.12.012

Wang, L., Sun, D.-W., Pu, H., and Zhu, Z. (2016). Application of hyperspectral imaging to discriminate the variety of maize seeds. *Food Analyt. Methods* 9, 225–234. doi: 10.1007/s12161-015-0160-4

Wu, N., Zhang, C., Bai, X. L., Du, X. Y., and He, Y. (2018). Discrimination of chrysanthemum varieties using hyperspectral imaging combined with a deep convolutional neural network. *Molecules* 23:2831. doi: 10.3390/molecules23112831

Xing, C., Ma, L., and Yang, X. (2016). Stacked denoise autoencoder based feature extraction and classification for hyperspectral images. *J. Sensors* 2016:3632943. doi: 10.1155/2016/3632943

Yang, J., Chen, G., Li, L., Zou, L., Zhang, C., and Mao, U. (2019). Portable rice disease spores capture and detection method using diffraction fingerprints on microfluidic chip. *Micromachines* 10:289. doi: 10.3390/mi10050289

Yuan, C., Yuxiang, Z., Zhijuan, J., Yan, L., Zhihua, W., and Changdeng, Y. (2019). Identification of stable quantitative trait loci for sheath blight resistance using recombinant inbred line. *Rice Sci.* 26, 331–338. doi: 10.1016/j.rsci.2019.08.007

Zabalza, J., Ren, J., Zheng, J., Zhao, H., Qing, C., Yang, Z., et al. (2016). Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing* 185, 1–10. doi: 10.1016/j.neucom.2015.11.044

Zahid, A., Abbas, H. T., Ren, A., Zoha, A., Heidari, H., Shah, S. A., et al. (2019). Machine learning driven non-invasive approach of water content estimation in living plant leaves using terahertz waves. *Plant Methods* 15:138. doi: 10.1186/s13007-019-0522-9

Zhang, C., Feng, X., Wang, J., Liu, F., He, Y., and Zhou, W. (2017). Mid-infrared spectroscopy combined with chemometrics to detect *Sclerotinia* stem rot on oilseed rape (*Brassica napus* L.) leaves. *Plant Methods* 13:39. doi: 10.1186/s13007-017-0190-6

Zhang, C., Liu, F., and He, Y. (2018). Identification of coffee bean varieties using hyperspectral imaging: influence of preprocessing methods and pixel-wise spectra analysis. *Sci. Rep.* 8:11. doi: 10.1038/s41598-018-20270-y

Zhang, C., Wu, W. Y., Zhou, L., Cheng, H., Ye, X. Q., and He, Y. (2020). Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging. *Food Chem.* 319:126536. doi: 10.1016/j.foodchem.2020.126536

Zhou, L., Zhang, C., Qiu, Z., and He, Y. (2020). Information fusion of emerging non-destructive analytical techniques for food quality authentication: a survey. *TrAC Trends Analyt. Chem.* 127:115901. doi: 10.1016/j.trac.2020.115901

Zhu, S., Feng, L., Zhang, C., Bao, Y., and He, Y. (2019). Identifying freshness of spinach leaves stored at different temperatures using hyperspectral imaging. *Foods* 8:356. doi: 10.3390/foods8090356

Zhu, S. S., Zhou, L., Gao, P., Bao, Y. D., He, Y., and Feng, L. (2019). Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties. *Molecules* 24:3268. doi: 10.3390/molecules24183268

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Feng, Wu, Zhu, Wang, Su, Liu, He and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 30 November 2020
doi: 10.3389/frai.2020.564878

Automatic Detection of Flavescence Dorée Symptoms Across White Grapevine Varieties Using Deep Learning

Justine Boulent[1,2,3]*, *Pierre-Luc St-Charles*[4], *Samuel Foucher*[2] and *Jérome Théau*[1,3]

[1]Department of Applied Geomatics, Université de Sherbrooke, Sherbrooke, QC, Canada, [2]Computer Research Institute of Montréal, Montréal, QC, Canada, [3]Quebec Centre for Biodiversity Science (QCBS), Montréal, QC, Canada, [4]Applied Machine Learning Research Team, Mila, Montréal, QC, Canada

Flavescence dorée (FD) is a grapevine disease caused by phytoplasmas and transmitted by leafhoppers that has been spreading in European vineyards despite significant efforts to control it. In this study, we aim to develop a model for the automatic detection of FD-like symptoms (which encompass other grapevine yellows symptoms). The concept is to detect likely FD-affected grapevines so that samples can be removed for FD laboratory identification, followed by uprooting if they test positive, all to be conducted quickly and without omission, thus avoiding further contamination in the fields. Developing FD-like symptoms detection models is not simple, as it requires dealing with the complexity of field conditions and FD symptoms' expression. To address these challenges, we use deep learning, which has already been proven effective in similar contexts. More specifically, we train a Convolutional Neural Network on image patches, and convert it into a Fully Convolutional Network to perform inference. As a result, we obtain a coarse segmentation of the likely FD-affected areas while having only trained a classifier, which is less demanding in terms of annotations. We evaluate the performance of our model trained on a white grape variety, Chardonnay, across five other grape varieties with varying FD symptoms expressions. Of the two largest test datasets, the true positive rate for Chardonnay reaches 98.48% whereas for Ugni-Blanc it drops to 8.3%, underlining the need for a multi-varietal training dataset to capture the diversity of FD symptoms. To obtain more transparent results and to better understand the model's sensitivity, we investigate its behavior using two visualization techniques, Guided Gradient-weighted Class Activation Mapping and the Uniform Manifold Approximation and Projection. Such techniques lead to a more comprehensive analysis with greater reliability, which is essential for in-field applications, and more broadly, for all applications impacting humans and the environment.

Keywords: precision viticulture, smart farming, plant diseases detection, Flavescence dorée, grapevine yellows, convolutional neural networks, fully convolutional networks, explainable artificial intelligence

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Hankui Zhang, South Dakota State University, United States
Roberto Pierro, University of Pisa, Italy

***Correspondence:**
Justine Boulent
justine.boulent@usherbrooke.ca

Specialty section:
This article was submitted to AI in Food, Agriculture and Water, a section of the journal Frontiers in Artificial Intelligence

Received: *22 May 2020*
Accepted: *21 October 2020*
Published: *30 November 2020*

Citation:
Boulent J, St-Charles P-L, Foucher S and Théau J (2020) Automatic Detection of Flavescence Dorée Symptoms Across White Grapevine Varieties Using Deep Learning. Front. Artif. Intell. 3:564878. doi: 10.3389/frai.2020.564878

INTRODUCTION

Flavescence dorée (FD) is a grapevine disease raising serious concern in Europe. This disease is caused by several phytoplasmas classified according to their ribosomal DNA (16SrV subgroup C and D) (Filippin et al., 2009) and grouped under the temporary name of *Candidatus Phytoplasma vitis* (Firrao et al., 2004). The transmission of those phytoplasmas is mediated by infected leafhopper *Scaphoideus titanus* which transmit the disease when feeding on vine leaves. This leafhopper, native to North America, was first observed in France in 1958 (Bonfils and Schvester, 1960). Adult *S. titanus* have a limited dispersal distance, reaching 25–30 m, although winds can cause passive dispersal over larger areas. However, the long distance spread is mainly caused by human activities, such as the trade of plant infested material (Chuche and Thiery, 2014). Despite a quarantine status and mandatory monitoring practices, FD is still spreading in Europe (Fanjul, 2017). FD symptoms appear gradually over the summer with a peak of expression at the end of August/September. Contaminated plants show a drooping appearance caused by a lack of lignification. Infected leaves roll up and change color: yellowing for white grape varieties and reddening for red grape varieties. FD causes the death of inflorescences and of berries (Chuche and Thiery, 2014). Symptoms expressed by an FD-infected plant are the same as those expressed by *Bois noir*, another disease caused by phytoplasmas. Both diseases are gathered under the term of "grapevine yellows symptoms." To distinguish between them, laboratory analyses are necessary (Chuche and Thiery, 2014). Over the years, FD-contaminated plants will die or become less and less productive. In any case, they constitute an eventual source of contamination for the surrounding vineyards. In France, the uprooting of contaminated grapevines is mandatory. When 20% of the grapevines in a parcel are contaminated, the whole parcel must be uprooted (Ministère Français de l'Agriculture, 2013). Winegrowers are thus exposed to significant economic losses. Since there is no control method for phytoplasma and no treatment to cure an infected plant, the disease management has focused on the FD vector: the leafhopper. In some at-risk areas in France, Italy and Switzerland, insecticide treatments are mandatory to limit the number of leafhoppers (Chuche and Thiery, 2014).

Another strategy complementary to vector control has been investigated by several research teams: the automatic detection of FD symptoms in the field. This approach is based on the idea that by quickly locating all the grapevines that might be contaminated with FD, winegrowers are able to eliminate all of the at-risk plant material. Winegrowers already scout their fields, but due to a lack of time or of manpower, this tedious task is not always carried out systematically on the entire vineyard. With a camera on a drone, a tractor or a ground robot, a complete search could be carried out in a relatively short time. The winegrower or agronomist would then only have to check the suspect grapevines spotted by the detection tool and take appropriate actions. To develop such an automatic detection tool, AL-Saddik et al. (2017) opted for non-imaging and proximal hyperspectral analyses to find the optimal wavelengths to identify FD. Using several spectral measurements and considering grape variety and symptom intensity, they developed specific vegetation indices for FD detection. These indices resulted in classification accuracies of more than 90% on leaf-scale spectral signatures. Albetis et al. (2018) used multispectral drone images (combining Red-Green-Blue, or RGB, and Near-Infrared, or NIR) acquired on seven red grape varieties. From these images, 24 variables such as vegetation indices and biophysical parameters were computed and used to differentiate two study cases: 1) FD and wood diseases (Esca and Black Dead Arm) from asymptomatic grapevines, and 2) FD from wood diseases. Results were promising in the first study case, but more mixed in the second.

The processing of in-field images highlighted another challenge: the management of soil and shadow mixels, which are significant sources of misclassification. Cruz et al. (2019) used Convolutional Neural Networks (CNNs) on proximal RGB images to detect grapevine yellowing symptoms on a red grape variety, with particular emphasis on differentiating them from leafroll and *Stictocephala biosonia* symptoms. Their model achieved 98.96% sensitivity and 99.40% specificity. However, the images they used were collected under controlled conditions, which would not be suitable for an in-field automated detection tool. Nevertheless, numerous studies have shown the relevance of using CNNs to analyze in-field images on various crops (DeChant et al., 2017; Fuentes et al., 2018; Boulent et al., 2019b). Despite the difficulties in analyzing field images due to complex background, leaf entanglements and shadows, CNNs are robust enough to overcome those issues and find discriminating features to identify several diseases on different crops.

CNNs provide a modeling approach that is part of deep learning methods. CNNs consist of neural networks that can be trained to fit their internal parameters for a specific task based on a large amount of training data. More specifically, CNNs are primarily a set of convolutional and non-linear layers that extract hierarchical representations to solve a task defined through a loss function. For more detailed information on deep learning and CNNs, please refer to LeCun et al. (2015) and Goodfellow et al. (2016). The deployment of CNNs has led to great progress in several vision-related applications, particularly in image content classification.

Although FD is a complex disease whose symptoms are not easy to detect, we hypothesize that CNNs can tackle this challenging task. Several confusion factors can be expected: 1) other phytosanitary problems, such as downy mildew or deficiencies, can be expressed through similar symptoms, especially on white grape varieties (e.g., yellowing); 2) a combination of health issues can occur on the same leaves, particularly at the end of the season, which complicates the visual identification of FD; 3) the symptoms of FD can be expressed differently between varieties beyond the main two grape families (red and white) through variations in curling or coloring for example(AL-Saddik et al., 2017; Albetis et al., 2018); and 4) FD symptoms also seem to fluctuate from one

year to the next, even for the same grape variety (Albetis et al., 2018).

The objective of this study is to achieve the automatic pre-identification of FD symptoms in several white grape varieties from images taken in the field. The term "pre-identification" refers to the fact that a laboratory analysis must be performed to ensure that a grapevine is indeed contaminated with FD. Our level of detection is downstream: the idea is to visually detect FD-like symptoms. Therefore, in this study, any mention of FD detection refers to this notion of visual "pre-identification." The multi-varietal dimension is studied here because, for professional use, an automatic detection tool would be of real interest only if it could be used on several grape varieties. Acquiring images of all the targeted grape varieties would be expensive and very challenging. However, the symptoms, even within the same color family, can be quite different. So, could a model trained on one grape variety be effective on other grape varieties? To answer this question, we train a classification model on Chardonnay images and transform it into a Fully Convolutional Network (FCN, Shelhamer et al., 2017). We then test it on several independent datasets originating from different sources. To provide greater transparency to our model and its results, we use visualization techniques to understand which image features the model uses for its predictions. This also helps us understand the differences in precision obtained on the different grape varieties. By identifying the strengths and weaknesses of our model (along with ways to improve its robustness and reliability) we take one more step toward the development of a professional tool.

MATERIAL AND METHODS

Method Overview

To provide greater clarity, this preliminary section presents an overview of the processing steps performed in this study (**Figure 1**). To develop a tool to automatically detect FD symptoms, we first trained a CNN-based classifier. As an output, this model indicates whether or not there are FD-like symptoms in an input image. As a way to help the model's training, we ensure that it targets elements of interest in the grapevines by annotating portions of images (mostly leaves). In the FD class of the training dataset, there are only Chardonnay leaves. The training process follows a classical approach: a ResNet18 is pre-trained on ImageNet and then completely fine-tuned with the FD training dataset. In fact, as our training dataset is quite small, it is advantageous to use weights from a prior training on a large dataset to retrieve the learned features as a starting point for our model training (Tajbakhsh et al., 2016; Tan et al., 2018).

The image-level binary classifier output (e.g., a unique FD or Non FD predicted label) is not the most suitable for our application. Indeed, a rough localization of the affected regions in the analyzed images would be much more appropriate. It would offer a more complete analysis while targeting suspect areas, bringing more transparency to the results and thus meeting the needs of the users. To achieve this rough localization, we opted to transform our classifier into an FCN. This transformation is simple in practice and it does not require any retraining to generate coarse class segmentation maps during inference (Shelhamer et al., 2017).

To test our model, we use several FD image datasets. Two are from personal acquisitions (*Pers.* datasets). These have the advantage of being quite large (153 images for the one on Chardonnay and 155 for the one on Ugni-Blanc) and they have a certain independence from the training dataset, as the images come from separate plots. However, they allow a possible bias: the images used for training and testing were taken during the same year, by the same team, and with similar cameras. To address this issue, we introduce a dataset from an external source (*Ext.* datasets), provided by FREDON (Fédération Régionale de Défense contre les Organismes Nuisibles) Aquitaine. Although small in size (34 images), it is of significant value, as it ensures that our model can work on images from completely independent sources. Furthermore, it also provides an opportunity to evaluate our model on several white grape varieties.

Datasets

Image Collection Process

To train and test the FD pre-identification model, we use RGB images captured with standard cameras. The training images were taken in 2018, in France, in the Cognac (45°41′24‴N, 0°19′12″W) and Limoux (43°2′60″N, 2°12′36″E) regions. The data acquisition was carried out from mid-August to mid-September, allowing the recording of varied FD symptom intensities. During the scouting, we mainly found FD-infected grapevines in Chardonnay plots, hence our constraint to focus on this grape variety for the FD class. We also encountered combinations of FD symptoms with symptoms of downy mildew, pest damages and phytosanitary treatment residues. Thus, even though our training dataset for the FD class is single-varietal, it contains some diversity in the symptoms' expression. During our scouting, we also encountered other plant health problems in Chardonnay and Ugni-Blanc plots: downy mildew, Esca, Black Dead Arm, damage caused by acarians, mineral deficiencies (in magnesium, potassium, manganese and iron) as well as burns due to pesticides. Two plots were selected as acquisition sites for the test datasets (referred to as *Pers.* datasets for "Personal" acquisition): one of Ugni-Blanc and one of Chardonnay. Thus, the images used to train and test the model came from separate plots.

Image acquisition was split based on two approaches: some images were captured using a hand-held camera, while others were captured from a camera mounted on a two-meter pole. The latter approach allowed a wider view of the grapevines. Due to the changing vineyard structure (inter-row distances, row heights, trimming type), it was not possible to define fixed acquisition parameters. The distance between the camera and the foliage as well as the acquisition angle thus vary from one photo to another. Hence, the photographs have slightly

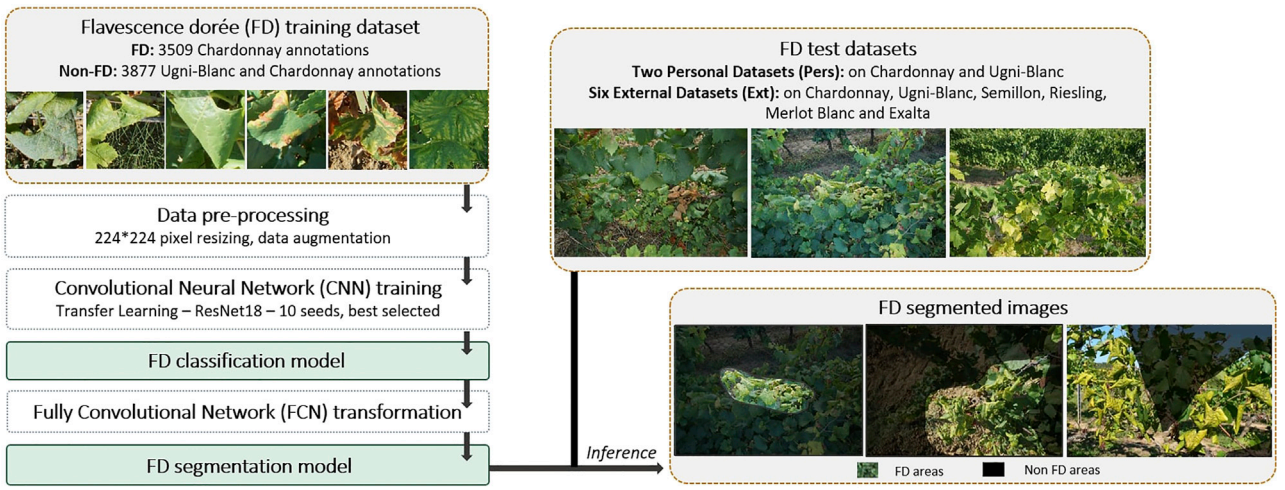

FIGURE 1 | Processing setup overview.

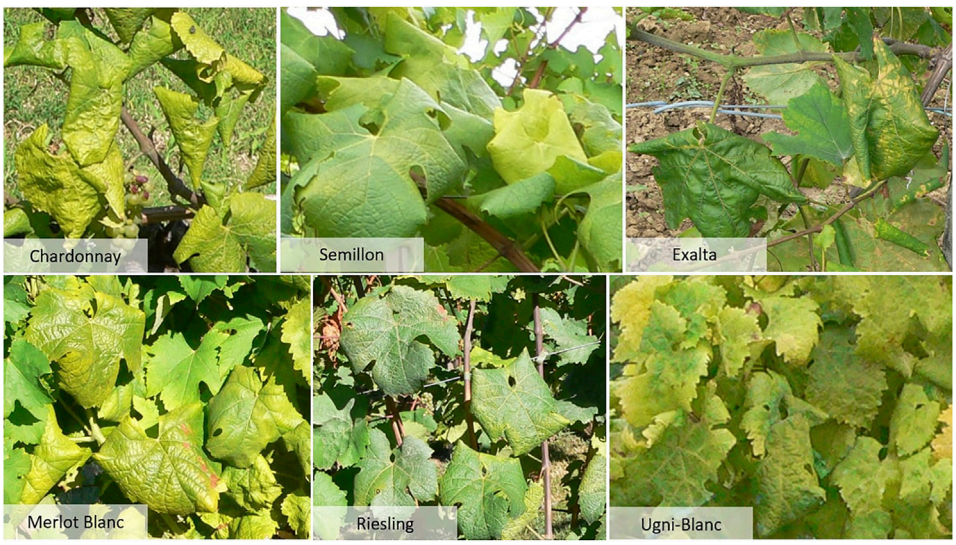

FIGURE 2 | Illustration of Flavescence dorée symptoms for the different grape varieties found in the *Ext.* test dataset. Source: D. Vergnes, FREDON Aquitaine.

different spatial focus and resolution. No restrictions in terms of weather conditions were established, and the only requirement for acquisition was that the leaves had to be dry. During the acquisitions, images were taken from both shaded and sunny rows, resulting in varying light exposures.

The set of 34 images prepared by a FREDON Aquitaine technician were reserved for testing.[1] This set covers six grape varieties, with 3 to 12 images per variety: Chardonnay, Ugni-Blanc, Merlot Blanc, Riesling, Exalta and Semillon. All the images contain symptomatic FD grapevines. **Figure 2** shows the typical expressions of the different grape varieties in this external (*Ext.*) dataset. Here, Chardonnay expresses a significant rolling with yellowing of a shade that differs from that of the training set. Exalta has rolled leaves with a withered aspect, sometimes with yellowing veins. Similar coloring is found on some Semillon leaves. Merlot Blanc and Riesling leaves show medium curling and yellowing. As for the Ugni-Blanc, it has little curling but significant coloration, sometimes with yellowing of the veins. Such variations in expression suggest generalization difficulties for our Chardonnay model on grape varieties expressing little curling. Although small in size, this test dataset is particularly relevant for the evaluation of our model because it was acquired by researchers not involved in this study. We know neither the camera model nor the year of acquisition, and had no control over the acquisition conditions. It therefore constitutes a realistic robustness test.

[1]Available online: https://flavescencedoree.jimdofree.com/

TABLE 1 | Summary of datasets used to train the CNN classifier and to evaluate the FCN model.

Dataset name	Grapevine varieties	Number of images
CNN training dataset	FD class: Chardonnay, non FD class: Chardonnay, Ugni-Blanc	Total: 894[a]
FCN evaluation datasets		
Pers. Chardonnay test	Chardonnay	FD class: 66 Non FD class: 87 Total: 153
Ext. Chardonnay test	Chardonnay	FD class: 5 Non FD class: None Total: 5
Pers. Ugni-Blanc test	Ugni-Blanc	FD class: 12 Non FD class: 143 Total: 155
Ext. Other Grapevine Varieties test	Ugni-Blanc, Exalta, Merlot Blanc, Semillon, Riesling	FD class: 29 Non FD class: None Total: 29

[a]*As we extracted samples for both classes in some images, we only indicate the total images' count for the CNN Training dataset.*

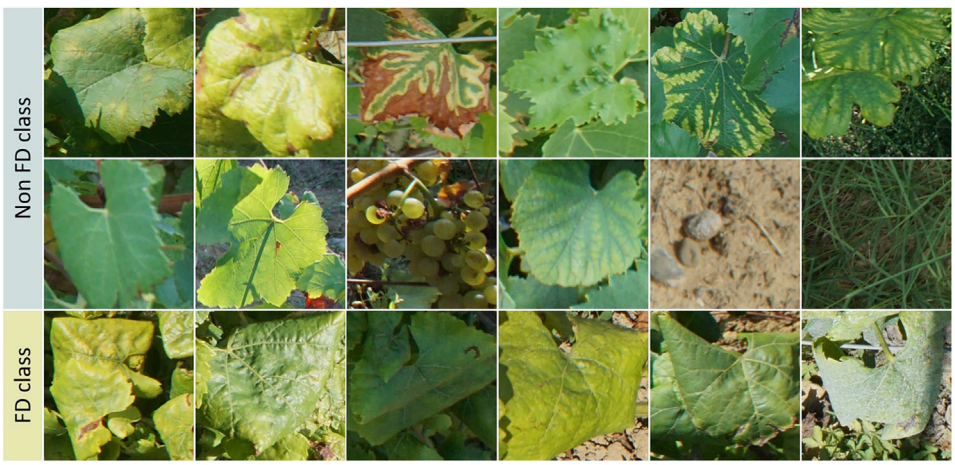

FIGURE 3 | Examples of samples from the CNN training dataset for FD and Non-FD classes.

Table 1 summarises the information about the datasets used to train the CNN classifier and to evaluate the FCN model.

Data Preparation

Taxonomy Definition and Annotation

A binary taxonomy was chosen for our classification model: FD versus Non FD. This choice was made primarily because it simplifies the annotation step. While several diseases were present in the surveyed plots, and these sometimes appeared in combinations, we do not have the expertise required to identify all the diseases we observed. The binary taxonomy therefore allows us to put everything that does not look like FD into a generic "Non FD" class. This class contains healthy grapevines, but also includes weeds, soil, sky and trellising elements. On the other hand, the FD class only contains leaves with FD-like symptoms, although no laboratory tests were conducted to confirm that FD was truly present.

To train the classifier, we manually extracted diverse samples from the captured images. For the plant elements, a similar unit was used throughout the annotation: the leaf or the bunch of grapes. We identified the minimal enclosing square for each image region of interest. Around 7,300 annotations were generated to train the classifier, with 3,509 for the FD class and 3,877 for the Non-FD class. Some samples within these classes are shown in **Figure 3**. Before each training session, the samples are randomly separated into training samples (85%, or 6 279) and validation samples (15%, or 1 107). In order to account for this random split in our evaluation results, we run a series of 10 independent experiments while shuffling these sample sets.

To test the classification model post-FCN conversion, we manually generate segmentation maps by coarsely delimiting the image regions potentially contaminated by FD. Segmentation maps were also manually generated on 15 images from the training set to help choose the best model out of our 10 experiments.

Image preprocessing

The samples used for training our binary classifier are resized to 224×224 pixels. Since we use transfer learning based on ImageNet weights, we normalize the images using the mean and standard deviation values given for the RGB channels on the PyTorch website.[2] To increase the model's robustness to geometric and dynamic effects and to add diversity to our images, the following data augmentation operations are applied on the images during training: image rotations (probability: 0.6), distortions (probability: 0.6), flips (probability: 0.8) and changes in brightness (probability: 0.6). These transformations are applied randomly to each sample using Augmentor.[3] For FCN inference, only the normalization was reapplied.

Methods

Training

As mentioned earlier, the FD pre-identification model is based on a fine-tuned ResNet-18 architecture (He et al., 2016) pre-trained on ImageNet. This architecture is well-known and performed well in our previous experiments despite its simple nature (Boulent et al., 2019a). The hyperparameter values are fixed based on several trials that involved the learning rate definition method proposed by Smith (2018) implemented in the FastAi library.[4] The values of the hyperparameters are provided in **Table 2**. The framework used to fine-tune and test the model is available on GitHub.[5]

TABLE 2 | Hyperparameters used to trained the ResNet-18 classifier.

Hyperparameter name	Hyperparameter value
Learning rate	1e–04
Learning rate scheduler	Step size: 5, Gamma: 0.7
Optimizer	Adam
Loss function	Cross entropy
Batch size	64
Epochs	15

Inference

Once trained, our image classifiers were converted to FCN models for semantic segmentation following the methodology detailed by (Shelhamer et al., 2017). By replacing the fully connected layer with a 1×1 convolutional layer, a coarse segmentation map can be produced as the result of inference. Such limited resolution is caused by the use of pooling layers as well as the stride and padding values of the convolution layers. As a result, the output segmentation map is much smaller than the original image. To overlay with the original image, we resize this result to the input image's size using bilinear interpolation.

We select the best CNN model out of our 10 experiments, using both validation accuracy and a visual quality assessment of the output segmentation maps on a handful of samples. The selected model is then evaluated on the four test sets: the two Chardonnay (*Pers.* and *Ext.*) as well as two other white grape varieties (*Pers.* and *Ext.*). Three metrics are used for performance evaluation: the True Positive Rate (TPR, **Eq. 1**) and the False Positive Rate (FPR, **Eq. 2**) to evaluate the quantity of both the correct and the incorrect detections generated by the model, and the Intersection over Union (IoU, **Eq. 3**), which provides a balanced overall look at the quality of the segmentation maps produced by the model. A Positive (P) sample is qualified as a "True" Positive (TP) if there was an intersection between the prediction and the annotation for the FD areas. A Negative (N) sample is qualified as a "True" Negative (TN) if the entire image was correctly predicted as Non-FD. The TPR was therefore more restrictive than a simple classification: it ensured that the detection was due to an FD area. Since for FD pre-identification, a false negative is more costly than a false positive, the decision threshold for the FD class has been set at 40% instead of the 50% usually used in binary classification.

$$TPR = \frac{TP}{P} \times 100, \tag{1}$$

where TP is the number of True Positive images and P is the number of Positive images;

$$FPR = \frac{FP}{N} \times 100, \tag{2}$$

where FP is the number of False Positives images and N is the total number Negatives of images; and

$$IoU_{AB} = \frac{A \cap B}{A \cup B}, \tag{3}$$

where A is the set of all FD pixels predicted by the model and B is the set of all pixels labeled as FD.

RESULTS

During the fine tuning, all models converged quickly (see **Figure 4**). By the end of the first epoch, the accuracies ranged from 96.2% to 98.7% on the validation dataset for all 10 runs. From the 11th epoch, accuracy values became very close, with a difference between the minimum and maximum accuracies of only 0.59% on average for the last 4 epochs. We transformed several of the obtained models with the highest precision values into FCNs, and used our validation dataset of 15 images to evaluate them. We decided to select the model with the highest precision value at the 9th epoch.

Inference on Chardonnay Datasets

For both the Chardonnay test datasets, the TPR values are very high: 98.5% for the *Pers. Chardonnay* dataset (i.e., 65 TP for 66 P), 100% for *Ext. Chardonnay* dataset (i.e. 5 TP for 5 P). There is a low number of false alarms for the *Pers. Chardonnay* dataset, with an FPR of 1.15% (i.e., 1 FP for 87 N). These results highlight the detection power of the trained model. Next, the IoU values provide insight on the quality of the segmentation maps. The

[2]https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html
[3]https://augmentor.readthedocs.io/en/master/
[4]https://docs.fast.ai/
[5]https://github.com/plstcharles/thelper

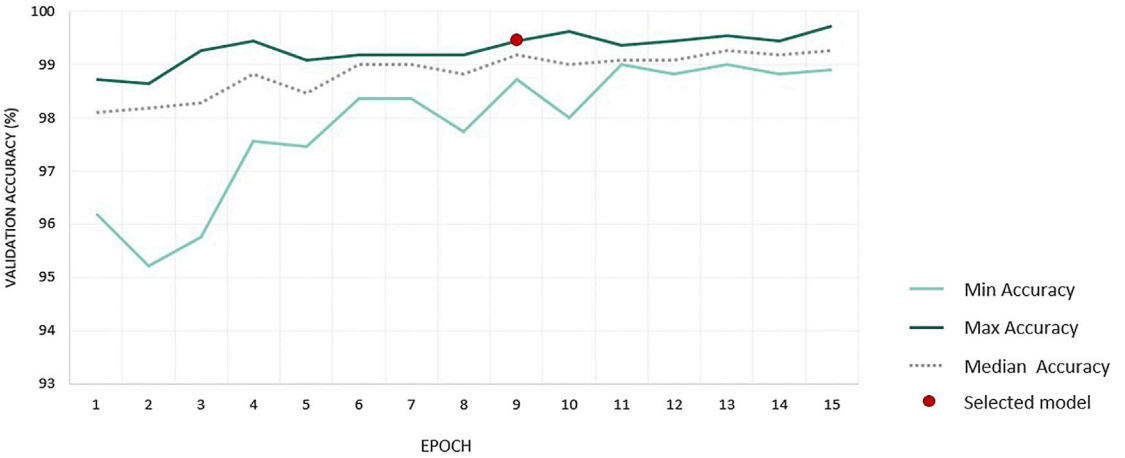

FIGURE 4 | Summary of accuracy values obtained on the validation dataset for the 10 runs.

TABLE 3 | Summary of inference results on both Chardonnay datasets, presenting the True Positive Rate (TPR), False Positive Rate (FPR), mean, min and max values of Intersection over Union (IoU).

Test Dataset Name	Number of images	TPR	FPR	IoU mean	IoU min[a]	IoU max
Pers. Chardonnay	FD class: 66 Non FD class: 87 Total: 153	98.48% (65 TP/66 P)	1.15% 1 FP/87 N	0.53	0.14	0.75
Ext. Chardonnay	FD class: 5 Non FD class: none Total: 5	100% (5 TP / 5 P)	N/A	0.39	0.26	0.6

Note: TP, True Positive; P, Positives; FP, False Positive; N, Negatives.
[a]*IoU min, considering only the True Positive predictions.*

average IoU value is 0.53 for the *Pers. Chardonnay* dataset (with values ranging from 0.14 to 0.75), and 0.39 for the *Ext. Chardonnay* dataset (with values ranging from 0.26 to 0.6) (**Table 3**). Such values indicate a rather coarse segmentation quality. This was as expected, due to the simple classifier's conversion to an FCN and its lack of advanced upsampling.

Figure 5 presents a set of typical predictions with minimum and maximum IoU values for the two Chardonnay datasets. The minimum IoU prediction on the *Pers. Chardonnay* dataset shows both over- and under-segmentation. However, two of the three leaves with FD symptoms are detected. Overall, we observe a better detection of FD when it is expressed on several leaves or on a branch. When symptoms affect only a few isolated leaves, there is a greater risk of non-detection (6, A and B). Leaves with only the underside visible are also under-detected (**Figure 6**). Furthermore, even if FD symptoms combined with downy mildew or residues of phytosanitary treatments are mostly detected, we noted a few cases of non-detection in this situation (**Figure 6**). Confusions with early downy mildew, with an embossed appearance, are present. With the minimum IoU prediction on the *Ext.* Chardonnay dataset, another type of misdetection is related to the background. Grapevines, corn or a hedge in the background can sometimes be partly identified as FD. We also observed a case where another grapevine contaminated with FD in the background was detected. We did not anticipate these two situations during our acquisitions, as we took photos that mainly focused on one grapevine. This image, as well as the maximum IoU prediction on the *Ext. Chardonnay* dataset, also highlight the coarseness of the detection – the FD is properly detected, but the detection mask also includes some of the surrounding ground. Several detections were found, including soil or areas not contaminated with FD, but close to contaminated areas (**Figure 6**). Finally, false detections that are visually difficult to explain were also found (**Figure 6**).

As a result, despite the coarseness of the detections and some cases of confusion, the classification model converted to an FCN for large-scale inference is able to detect symptoms of FD on unseen images of Chardonnay grapevines. While the *Ext. Chardonnay* dataset is small, the metrics show that the model is able to generalize on images of the same grape variety, even if they come from another source. The results are very encouraging, as the expression of FD symptoms is slightly different between the two datasets; the leaves are yellower in the *Ext. Chardonnay* images, but they remain quite green in the *Pers. Chardonnay* images. These two expressions, both present in the training set, are well-managed by the model.

Inference on Several White Grapevine Varieties

Inference using the Chardonnay model on five other white grape varieties gave mixed results (**Table 4**). For the two

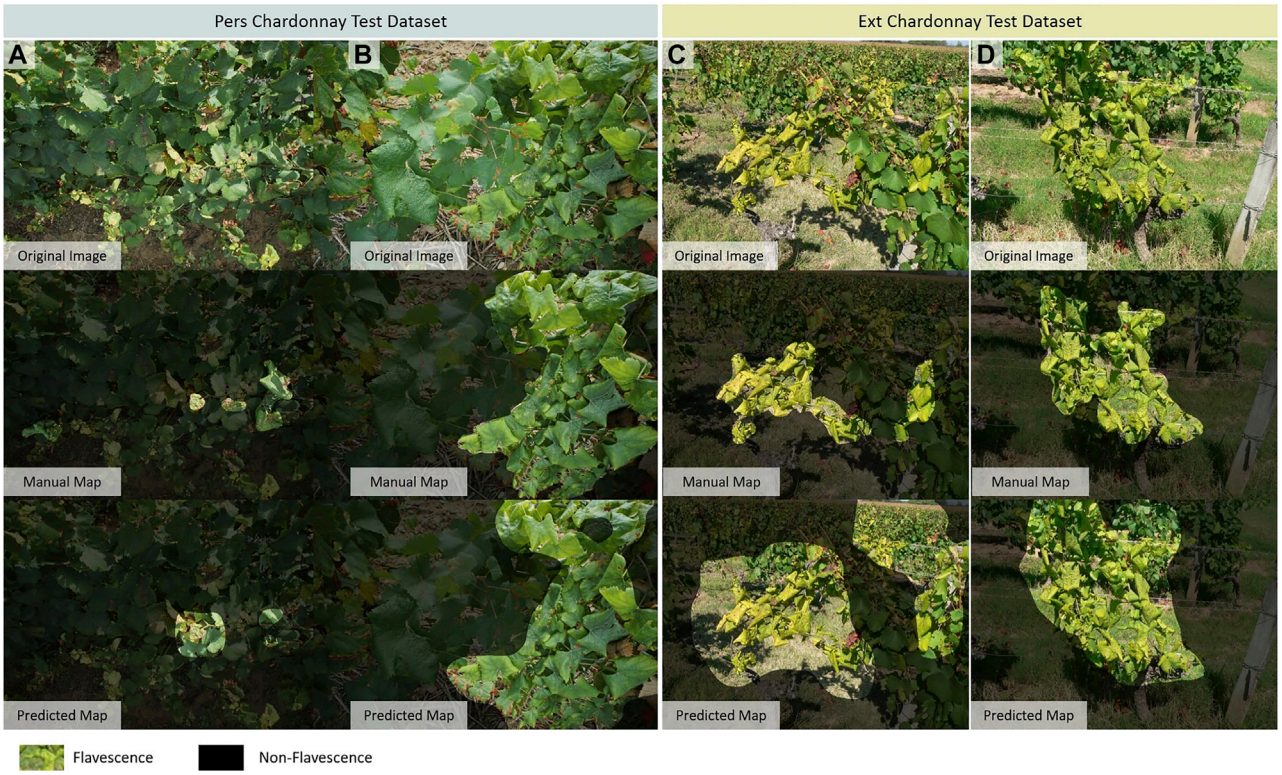

FIGURE 5 | Segmentation maps with minimum and maximum Intersection over Union (IoU) values on Chardonnay datasets. On the *Pers. Chardonnay* dataset, Images **(A)** with the lower IoU value, **(B)** with the higher IoU value. On the *Ext. Chardonnay* dataset, **(C)** with the lower IoU value, **(D)** with the higher IoU value.

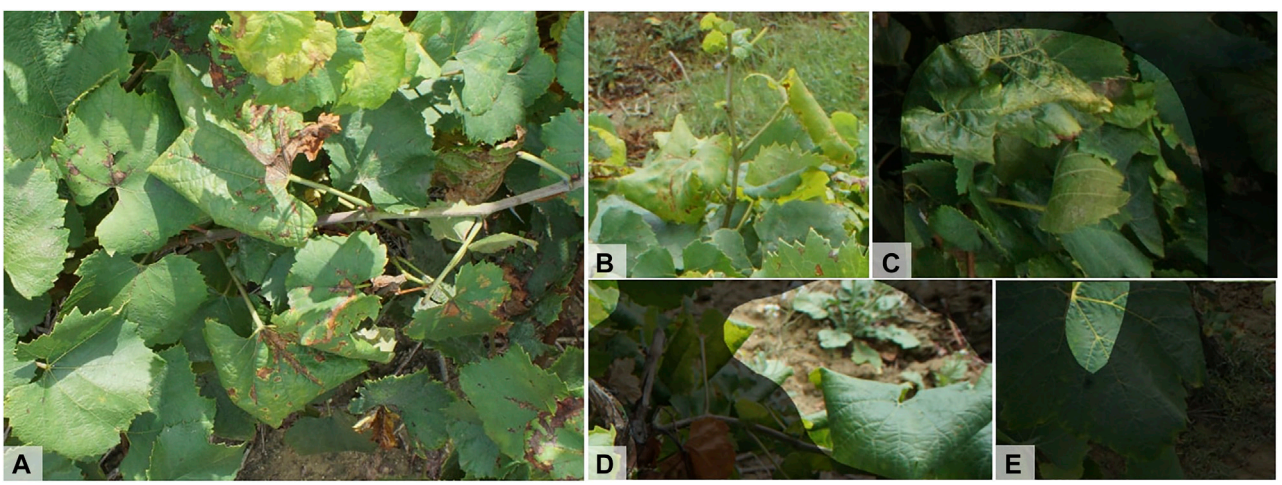

FIGURE 6 | Examples of false predictions from the *Pers. Chardonnay* Test dataset. **(A,B)** Undetected Flavescence dorée (FD) symptoms, **(C,D)** Over-detection close to FD symptomatic areas, **(E)**: Visually unexplainable false FD detection.

Ugni-Blanc datasets, the Chardonnay model is unable to detect FD symptoms: 0% of TPR for *Ext. Ugni-Blanc* dataset and 8.3% of TPR for *Pers. Ugni-Blanc* dataset, with a very poor quality segmentation map for the only TP obtained (IoU value of 0.07). Even when leaves express curling, they are sometimes identified as Non-FD (**Figure 7**). The same phenomenon was observed on other grape varieties, such as Semillon (**Figure 8**) or Exalta (**Figure 7**), which was unexpected given that the model detects Chardonnay symptoms of FD characterized by rolling and significant yellowing (**Figure 5**). However, one result can be considered positive: the FPR of 2.87% obtained on the *Pers. Ugni-Blanc* dataset is of the same order as the FPR value obtained on the *Pers. Chardonnay* dataset.

TABLE 4 | Summary of inference results on other grapevine varieties' datasets, presenting the True Positive Rate (TPR), False Positive Rate (FPR), mean, min and max values of Intersection over Union (IoU).

Test Dataset Name	Number of images	TPR	FPR	IoU mean	IoU min[a]	IoU max
Pers. Ugni-Blanc	FD class: 12 Non FD class: 143 Total: 155	8.3% 1 TP/12 P	2.87% 4 FP/139 N	N/A IoU for the only TP: 0.07		
Ext. Ugni-Blanc	FD class: 5 Non FD class: none Total: 5	0% 0 TP/5 P	N/A	N/A	N/A	N/A
Ext. Exalta	FD class: 12 Non FD class: none Total: 12	83.3% 10 TP/12 P	N/A	0.13	0.02	0.31
Ext. Merlot Blanc	FD class: 3 Non FD class: none Total: 3	100% 3 TP/3 P	N/A	0.18	0.01	0.46
Ext. Riesling	FD class: 4 Non FD class: none Total: 4	100% 4 TP/4 P	N/A	0.27	0.17	0.39
Ext. Semillon	FD class: 5 Non FD class: none Total: 5	100% 5 TP/5 P	N/A	0.14	0.01	0.24

Notes: TP, True Positive; P, Positives; FP, False Positive; N, Negatives.
[a]*IoU min, considering only the True Positive predictions.*

FIGURE 7 | Examples of false predictions from *Ext.* datasets. **(A)** Original image of Exalta leaf with Flavescence dorée (FD) symptoms; **(B)** Prediction associated to image A, only the rolled areas of the leaf are identified as FD; **(C)** Undetected FD symptoms on Ugni-Blanc grapevine variety; **(D)** Over-detection close to FD symptomatic areas; **(E)** False detection on non-grapevine elements; and **(F)** Undetected FD symptoms on Exalta grapevine variety.

The inability of the Chardonnay model to identify FD on Ugni-Blanc could be caused by confusion with downy mildew or nutrients deficiencies. This hypothesis is based on the observation of the predictions for Exalta. In this grape variety, there were more detections on green leaves with heavy rolling. Yellow vein symptoms were classified as Non-FD (**Figure 7**). This symptom's expression has not been observed on the visited Chardonnay plots and is therefore absent from the training dataset. For Exalta and Semillon varieties, even if the TPR is high, the segmentation is of poor quality, with many under-detections (**Figure 8**).

On Riesling and Merlot Blanc varieties, the segmentation is of relatively better quality, with average IoUs of 0.27 and 0.18, respectively. Again, cases of under-detection were noted (**Figure 8**). One possible explanation could be the difference in resolution between the images used for training and those used for inference. Indeed, in images found in **Figures 8C,D**, the same leaves were photographed

FIGURE 8 | Segmentation maps with minimum and maximum Intersection over Union (IoU) values on two *Ext.* datasets. On the *Ext. Semillon* dataset: Images **(A)** with the lower IoU value, and **(B)** with the higher IoU value. On the *Ext. Merlot Blanc* dataset: **(C)** with the lower IoU value, and **(D)** with the higher IoU value.

with a more or less close-up view. On the close-up view, FD is detected, while on the farther view, symptomatic leaves are labeled as Non-FD. Furthermore, as observed in the Chardonnay datasets, there is an over-detection in the predictions around areas identified as symptomatic (**Figure 7**), but there is also over-detection related to the background that is barely interpretable (**Figure 7**).

In summary, the Chardonnay model does not work to detect FD symptoms on the two Ugni-Blanc datasets. However, the low FPR obtained on the *Pers. Ugni-Blanc* dataset may translate to significant robustness in a real-world application. For the other grape varieties, even if the TPR is high, the average IoUs are very low. In particular, many under-detections were noted. Therefore, the segmentation maps do not have the required quality for an automatic detection tool.

DISCUSSION

Providing a model with high accuracy for FD symptoms recognition is a first step toward the development of an automatic detection tool, but is not sufficient in itself. Indeed, in agriculture, as in all other fields of application that may have social and environmental impacts, it is necessary to move toward artificial intelligence tools whose results and underlying processes are transparent, understandable and explainable. Relying on opaque results does not provide a sufficient degree of reliability for real-world applications. Wrong predictions in the field could have severe economic consequences for farmers. Without transparency, it would be difficult for farmers to have confidence in new AI-driven tools. Explainable AI refers to all the methods and techniques put in place to allow humans to understand the decisions and results produced by AI systems, thus avoiding a black box effect (Holzinger et al., 2019; Doran et al., 2018; Arrieta et al., 2020).

Here, we take a step toward greater transparency by increasing the interpretability of our system regarding two aspects: 1) the multi-varietal application capability of the model trained only on Chardonnay FD symptoms, and 2) the general robustness of the model. For this purpose, we use two visualization techniques. The first one, Guided Gradient-weighted Class Activation Mapping (GG-CAM, Selvaraju et al., 2019),[6] provides a finely detailed and discriminative view of class activations. This allows us to observe which characteristics of an image are used to predict a targeted class. The second visualization technique, Uniform Manifold Approximation and Projection (UMAP, McInnes et al., 2018a),[7] is a a dimensionality reduction algorithm that preserves the global structure of the projected data. This allows us to visualize the spatial arrangement of the image embeddings inferred by the model. Here, we use UMAP to observe the closeness of the embeddings according to their class and grape variety.

[6]Python library used: Captum (Kokhlikyan et al., 2019)
[7]The Python library used here: UMAP (McInnes et al., 2018b)

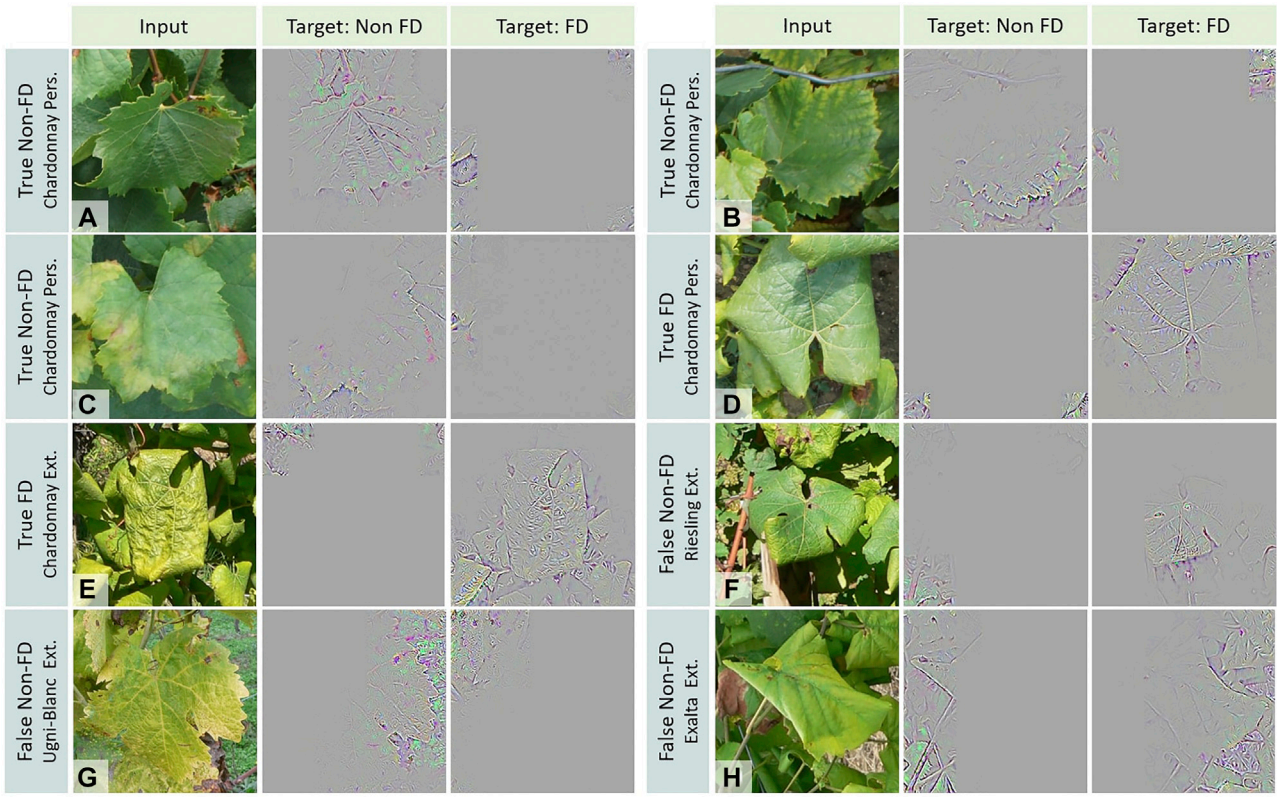

FIGURE 9 | Illustration of representative features from Guided Grad-CAM that are used by the model to predict the Flavescence dorée (FD) and Non-FD classes. (**A–C**): Non-FD images, **(D–H)**: FD images. The predicted label and the source dataset are given next to each image.

To improve the readability of this analysis, we used a subset of our evaluation datasets to create several visualizations. This subset contains 360 patches extracted from the images within the six evaluation datasets. Specifically, 30 patches were extracted for each of the following 12 categories: *Pers. Chardonnay*, Healthy; *Pers. Chardonnay*, Downy Mildew; *Pers. Chardonnay*, Deficiencies; *Pers. Chardonnay*, FD; *Ext. Chardonnay*, FD; *Pers. Ugni-Blanc*, Healthy; *Pers. Ugni-Blanc*, FD; *Ext. Ugni-Blanc*, FD; *Ext. Exalta*, FD; *Ext. Merlot Blanc*, FD; *Ext. Riesling*, FD; *Ext. Semillon*, FD.

Multi-Varietal Application

We expect that the need for FD detection will appear for many grape varieties. If images have to be collected on all the targeted grape varieties, the task would become increasingly challenging and expensive. Our model, trained on a single symptomatic white grape variety, resulted in mixed performances when tested on other varieties. By applying GG-CAM to our images, the characteristics typically used by the model to associate FD and Non-FD classes are highlighted (**Figure 9**). For both classes, leaf morphology seems to be important. For the Non-FD class, it is mainly the leaf teeth that stand out (**Figures 9A–C,G**), but the orientation of the veins also appears to be relevant (**Figure 9A**). For the FD class, morphologically, it is mainly the curl, the sinus shape and the vein orientation that emerge (**Figures 9D–F,H**). Leaf texture also seems to be a discriminating feature, as an embossed, scaly aspect of the leaves appears (**Figures 9E,F,H**). Color is also a decisive feature. For the FD class there is a yellow dominance, whereas for the Non-FD class green and purple spots are frequently seen. Canes and petioles are mostly correctly associated with the Non-FD class (**Figures 9A,F,H**). The model's discriminative leaf features spotted through GG-Cam are of interest since they reflect the characteristics used by experts in the field for their diagnosis.

Such general class feature descriptions provide a better understanding about FD false negative predictions across the different grape varieties. The Ugni-Blanc, which here expresses little curling, seems to be associated with the Non-FD class because of the toothed aspect of the leaf, which remains despite FD contamination (**Figure 9G**). In some cases, including those showing partial curling, both classes have strong activations (**Figure 9F,H**), but the Non-FD class is predicted by the model. This problem may have been accentuated here by the size of the image patch as well as by the core size of the average pooling layer. Some FD symptoms do not appear in the GG-CAM maps, including vein discoloration and FD-symptomatic leaves without curling.

The unawareness of these symptoms, which are very common for some grape varieties, is also revealed through UMAP visualization (**Figure 10**). Indeed, in this visualization, two clusters are found at each end: the true predictions of Non-FD and FD. In the Non-FD cluster, healthy Ugni-Blanc and healthy Chardonnay share the same space. Downy mildew and deficiencies

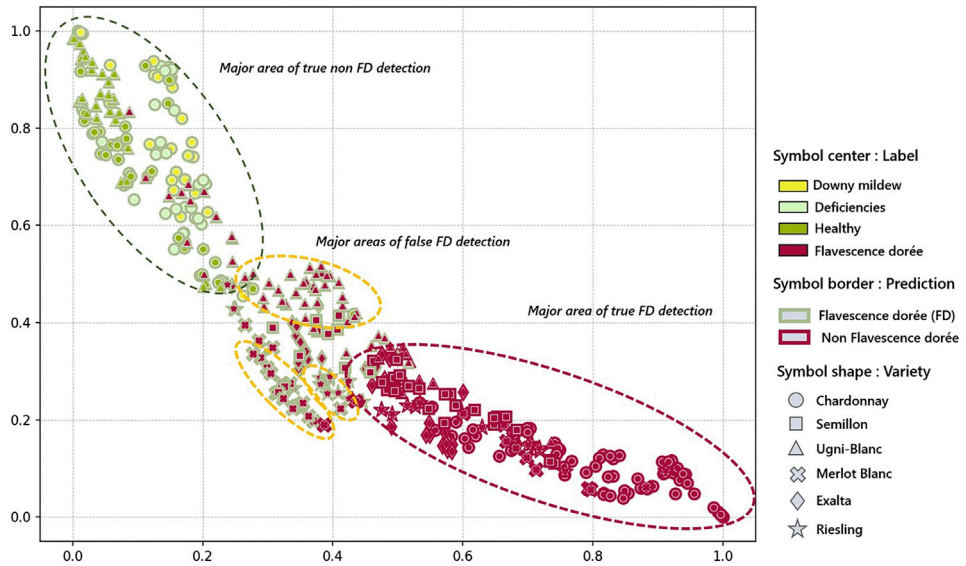

FIGURE 10 | Grape varieties in the model space: UMAP visualization of the embeddings based on an arbitrarily-set seed.

embeddings are also quite close. At one end of this cluster, we observe proximity to the true Non-FD positives that correspond to Chardonnay deficiencies and downy mildew patches with false Non-FD negatives of Ugni-Blanc.

In the true FD prediction cluster, all of the Chardonnay patches from the *Pers.* and *Ext.* datasets are found with the majority grouped at the 0.8:1 end of the *x*-axis. Elsewhere in the cluster, many Semillon, some Exalta, Riesling, Merlot Blanc and, to a lesser extent, Ugni-Blanc can be found. In between our two main clusters of true predictions is a zone of false negative FD embeddings. In this area, there are micro-clusters based on grape variety. These are certainly influenced by the closeness of the expressed symptoms among the same grape variety, but potentially also by having images from the same source. Despite this possible bias, the UMAP visualization confirms the hypothesis that the model fails to detect FD symptoms on Ugni-Blanc because of their closeness to deficiencies and to downy mildew symptoms. For false detections on the other grape varieties, a greater distance is observed from the cluster of good Non-FD predictions, which expresses the model's ambiguity toward symptom expressions not directly seen during training. To improve this model and make it multi-varietal, new acquisitions would be necessary to include images with grapevines presenting symptoms with intermediate curling intensities and vein colorings.

Toward a Robust Detection Model

Although the results obtained on Chardonnay are promising, further acquisitions are needed to increase our model's reliability. Indeed, even for only FD on a given grape variety, symptoms may be expressed differently from one year to another. In addition, the

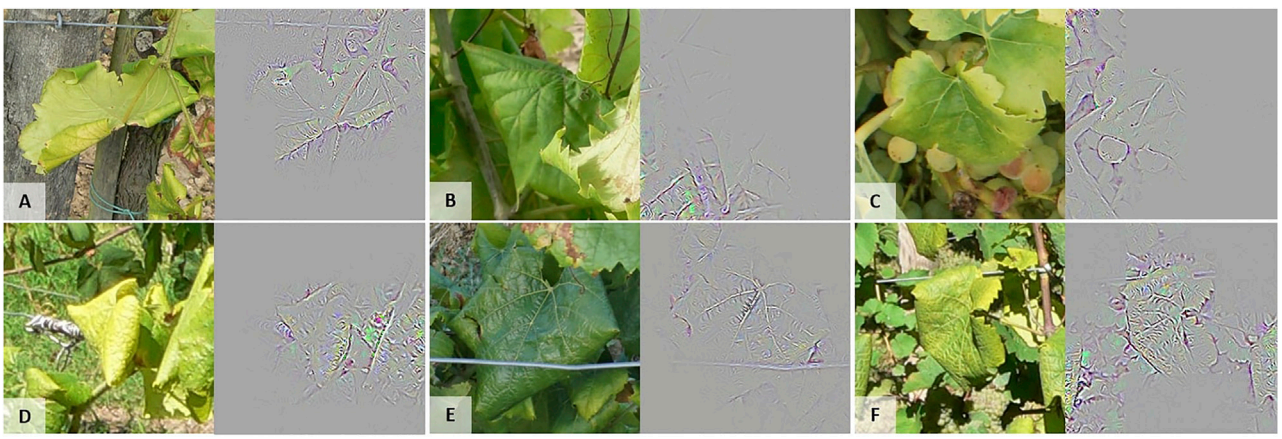

FIGURE 11 | Illustration of irrelevant features from GG-CAM that are used by the model to predict the Flavescence dorée (FD) class. **(A)** Trellising wire, petioles, canes, soil and rolled shapes, **(B)** Canes, **(C)** Canes and grapes, **(D)** Trellising wire and triangular shadow shape, **(E)** Trellising wire and triangular shadow shape, and **(F)** Trellising wire and soil.

general phytosanitary situation of a field can be very different from one year to another. Since FD is a late-season disease, grape leaves may already be damaged when FD symptoms appear. Acquisitions over several years will provide the needed data on the co-occurrence of phytosanitary problems.

The weaknesses of our model can also be targeted using the feature maps produced with GG-CAM (**Figure 11**). Indeed, some elements of the image stand out for the FD target when they should not, such as the trellising wire (**Figures 11A,D,E,F**), the soil (**Figures 11A,D,F** and **9H**), some shadows or leaf overlays with triangular or rolled shapes (**Figures 11D** and **9A,H**), as well as, in rare cases, grapes, petioles or canes (**Figures 11A–C**). To overcome this, targeted acquisitions to increase training image diversity could be conducted. Intervention at the model level could also be performed, e.g. by adding the annotation mask as an extra channel during training to act as an attention map, or with explanatory interactive learning, as in Schramowski et al. (2020) for plant phenotyping. These techniques would allow us to ensure that the predictions obtained are based on relevant features.

CONCLUSION

In this study, we sought to automatically pre-identify FD symptoms on several white grape varieties. We trained a CNN model and turned it into an FCN so that it produced segmentation of areas expressing FD symptoms. In the training dataset, all symptomatic annotations were based on Chardonnay grapevines. During evaluation, this model obtained high accuracy on Chardonnay images, both on the personal and on the external datasets. Using the GG-CAM visualization technique, weaknesses in the model were identified, such as sensitivity to background or to trellising elements. Further targeted acquisitions and re-training could solve this problem, which is highly encouraging. A larger external dataset obtained over several growing seasons would nevertheless be necessary to ensure that this model is indeed operational on Chardonnay grapevines.

For the other grape varieties, the results obtained were more nuanced, and showed how different symptoms' expressiveness could be problematic for FD pre-identification. As expected, the model performed poorly for the grape variety expressing the least amount of curling, i.e., Ugni-Blanc. We were able to confirm this hypothesis using GG-CAM visualizations. The embeddings extracted through convolutions for FD-contaminated Ugni-Blanc leaves are close to those of leaves affected by downy mildew and deficiencies. For the other grape varieties, the performance is mixed, with very high prediction rates. For a model trained with a grape variety expressing a lot of curling, good results were achieved on grape varieties expressing curling to a lesser extent. This underlines the model's relative flexibility and validates the idea that the development of a multi-varietal model is achievable.

Our next steps are to develop a comprehensive overview of the various FD symptoms and to include them in a multi-varietal dataset. Appropriate precautions will be undertaken regarding all other diseases that may express yellowing and curling. Capturing images from these misleading negatives would help to establish a more accurate boundary between FD and Non-FD images. Finally, going back and forth between model training and field testing will allow us to develop a robust detection tool. Using visualization techniques, as we have done in this study, will help to provide a more complete model analysis, and thus help to achieve the reliability needed for in-field applications.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: non-commercial use only. Requests to access these datasets should be directed to JB, justine.boulent@usherbrooke.ca

AUTHOR CONTRIBUTIONS

JB carried out the image acquisition, dataset preparation, model training, all the associated analysis and wrote the manuscript. P-LS-C and JB developed the Python code. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research project was funded by the Center de Géomatique du Québec (CGQ), Mitacs, the Ministère de l'Économie et de l'Innovation (MEI) of the province of Québec, and Microdrones Canada.

ACKNOWLEDGMENTS

Many thanks to all the people who enabled the dataset acquisition used in this study: Microdrones Canada and France, the Rémy Martin vineyards, the Chambre d'Agriculture de l'Aude, more specifically, Céline Forget, the Fédération Départementale de défense contre les Organismes Nuisibles (FEDON) de l'Aude, and Julien Sendrous in particular, as well as all the owners and wine growers who gave us access to their vineyards and/or guided us in the image acquisition process. Special thanks to Eve Laroche-Pinel for sharing this in-field data acquisition. We would also like to thank the Fédération Régionale de défense contre les Organismes Nuisibles (FREDON) of Aquitaine, especially Dominique Vergnès, who shared his photos collected on various white grape varieties, as well as his expertise in identifying flavescence dorèe symptoms. Thanks also go to Maël Moreni for his expertise on the learning rate definition technique and to Simon Morvan for his advice on FD biology and epidemiology.

REFERENCES

Al-Saddik, H., Simon, J. C., and Cointault, F. (2017). Development of spectral disease indices for 'flavescence dorée' grapevine disease identification. *Sensors (Basel)* 17, 12–32. doi:10.3390/s17122772

Albetis, J., Jacquin, A., Goulard, M., Poilvé, H., Rousseau, J., Clenet, H., et al. (2018). On the potentiality of UAV multispectral imagery to detect flavescence dorée and grapevine trunk diseases. *Rem. Sens.* 11, 23. doi:10.3390/rs11010023

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence (xai): concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf. Fusion* 58, 82–115. doi:10.1016/j.inffus.2019.12.012

Bonfils, J., and Schvester, D. (1960). Les cicadelles (homoptera auchenorhyncha) dans leurs rapports avec la vigne dans le sud- ouest de la France. *Ann. Epiphyt. (Paris)* 9, 325–336.

Boulent, J., Beaulieu, M., St-Charles, P.-L., Théau, J., and Foucher, S. (2019a). "Deep learning for in-field image-based grapevine downy mildew identification," in *Precision agriculture '19*. Editor J. V. Stafford (Wageningen: Wageningen Academic Publisher), 141–148. doi:10.3920/978-90-8686-888-9_16

Boulent, J., Foucher, S., Théau, J., and St-Charles, P.-L. (2019b). Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10, 941. doi:10.3389/fpls.2019.00941

Chuche, J., and Thiéry, D. (2014). Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: a review. *Agron. Sustain. Dev.* 34, 381–403. doi:10.1007/s13593-014-0208-72

Cruz, A., Ampatzidis, Y., Pierro, R., Materazzi, A., Panattoni, A., De Bellis, L., et al. (2019). Detection of grapevine yellows symptoms in vitis vinifera l. with artificial intelligence. *Comput. Electron. Agric.* 157, 63–76. doi:10.1016/j.compag.2018.12.028

DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., et al. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* 107, 1426–1432. doi:10.1094/PHYTO-11-16-0417-R

Doran, D., Schulz, S., and Besold, T. R. (2018). "What does explainable ai really mean? a new conceptualization of perspectives," in CEUR workshop proceedings 2071 proceedings of the first international workshop on comprehensibility and explanation in AI and ML 2017. Copyright © 2018 for this paper by its authors. Copying permitted for private and academic purposesco-located with 16th international conference of the Italian Association for Artificial Intelligence (AI*IA 2017), Bari, Italy, November 16 and 17, 2017.

Fanjul, M. J. (2017). Technical datasheets in innovative solutions to control grapevine trunk disease and Flavescence Dorée. Tech. Rep., WinetWork.

Filippin, L., Jović, J., Cvrković, T., Forte, V., Clair, D., Toševski, I., et al. (2009). Molecular characteristics of phytoplasmas associated with Flavescence doréein clematis and grapevine and preliminary results on the role of Dictyophara europaeaas a vector. *Plant Pathol.* 58, 826–837. doi:10.1111/j.1365-3059.2009.02092.x

Firrao, G., Andersen, M., Bertaccini, A., Boudon, E., Bové, J., Daire, X., et al. (2004). Candidatus phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int. J. Syst. Evol. Microbiol.* 54, 1243–1255. doi:10.1099/ijs.0.02854-0

Fuentes, A. F., Yoon, S., Lee, J., and Park, D. S. (2018). High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9, 28–44. doi:10.3389/fpls.2018.01162

Goodfellow, I. J., Bengio, Y., and Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, June 27–30, 2016 (IEEE). doi:10.1109/cvpr.2016.90.2016

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining Knowl. Discov.* 9, e1312. doi:10.1002/widm.1312

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Reynolds, J., Melnikov, A., et al. (2019). Pytorch captum. Available at: https://github.com/pytorch/captum (Accessed November 15, 2020).

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539

McInnes, L., Healy, J., and Melville, J. (2018a). Umap: Uniform manifold approximation and projection for dimension reduction. Vol. 3, 861.

McInnes, L., Healy, J., Saul, N., and Grossberger, L. (2018b). Umap: Uniform manifold approximation and projection. *Joss* 3, 861. doi:10.21105/joss.00861

Ministère Français de l'Agriculture, d. l. e. d. l. f. (2013). Arrêté du 19 décembre 2013 relatif á la lutte contre la flavescence dorée de la vigne et contre son agent vecteur. Available at: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000028409891 (Accessed Feburary, 2020).

Schramowski, P., Stammer, W., Teso, S., Brugger, A., Shao, X., Luigs, H.-G., et al. (2020). *Right for the wrong scientific reasons: Revising deep networks by interacting with their explanations*.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2019). Grad-cam: visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* 128, 336–359. doi:10.1007/s11263-019-01228-7

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. doi:10.1109/tpami.2016.2572683

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: part 1—learning rate, batch size, momentum, and weight decay. ArXiv, 21.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., et al. (2016). Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imaging* 35, 1299–1312. doi:10.1109/tmi.2016.2535302

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). "A survey on deep transfer learning," in *Artificial neural networks and Machine learning – ICANN 2018*. Editors V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis (Cham: Springer International Publishing)), 270–279.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Boulent, St-Charles, Foucher and Théau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 01 December 2020
doi: 10.3389/fpls.2020.576177

Machine Learning Unmasked Nutritional Imbalances on the Medicinal Plant *Bryophyllum* sp. Cultured *in vitro*

Pascual García-Pérez[1,2], *Eva Lozano-Milo*[1,2], *Mariana Landin*[3,4] *and Pedro Pablo Gallego*[1,2]*

[1] *Applied Plant and Soil Biology, Plant Biology and Soil Science Department, Biology Faculty, University of Vigo, Vigo, Spain,* [2] *Clúster de Investigación e Transferencia Agroalimentaria do Campus da Auga - Agri-Food Research and Transfer Cluster, University of Vigo, Ourense, Spain,* [3] *Grupo I+D Farma (GI-1645), AeMat, Pharmacology, Pharmacy and Pharmaceutical Technology Department, Pharmacy Faculty, University of Santiago, Santiago de Compostela, Spain,* [4] *Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain*

OPEN ACCESS

Edited by:
Reza Ehsani, University of California, Merced, United States

Reviewed by:
Johannes Felix Buyel, Fraunhofer Society (FHG), Germany
Manuel Martinez-Estevez, Centro de Investigación Científica de Yucatán, Mexico

***Correspondence:**
Pedro Pablo Gallego
pgallego@uvigo.es

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *25 June 2020*
Accepted: *06 November 2020*
Published: *01 December 2020*

Citation:
García-Pérez P, Lozano-Milo E, Landin M and Gallego PP (2020) Machine Learning Unmasked Nutritional Imbalances on the Medicinal Plant Bryophyllum sp. Cultured in vitro. Front. Plant Sci. 11:576177. doi: 10.3389/fpls.2020.576177

Plant nutrition is a crucial factor that is usually underestimated when designing plant *in vitro* culture protocols of unexploited plants. As a complex multifactorial process, the study of nutritional imbalances requires the use of time-consuming experimental designs and appropriate statistical and multiple regression analysis for the determination of critical parameters, whose results may be difficult to interpret when the number of variables is large. The use of machine learning (ML) supposes a cutting-edge approach to investigate multifactorial processes, with the aim of detecting non-linear relationships and critical factors affecting a determined response and their concealed interactions. Thus, in this work we applied artificial neural networks coupled to fuzzy logic, known as neurofuzzy logic, to determine the critical factors affecting the mineral nutrition of medicinal plants belonging to *Bryophyllum* subgenus cultured *in vitro*. The application of neurofuzzy logic algorithms facilitate the interpretation of the results, as the technology is able to generate useful and understandable "IF-THEN" rules, that provide information about the factor(s) involved in a certain response. In this sense, ammonium, sulfate, molybdenum, copper and sodium were the most important nutrients that explain the variation in the *in vitro* culture establishment of the medicinal plants in a species-dependent manner. Thus, our results indicate that Bryophyllum spp. display a fine-tuning regulation of mineral nutrition, that was reported for the first time under *in vitro* conditions. Overall, neurofuzzy model was able to predict and identify masked interactions among such factors, providing a source of knowledge (helpful information) from the experimental data (non-informative *per se*), in order to make the exploitation and valorization of medicinal plants with high phytochemical potential easier.

Keywords: *Kalanchoe*, ANNs, media mineral nutrition, plant *in vitro* culture, artificial intelligence algorithms, structural risk minimization (SRM), ASMOD

INTRODUCTION

Recent biotechnological reports highlighted that medicinal plants constitute the source for more than the 25% of drugs officially approved by the Food and Drug Administration (FDA; Marchev et al., 2020). Furthermore, medicinal plant-derived products are effectively used in the primary healthcare systems for around the 90% of developing countries (El Sheikha, 2017). Taking this into account, the exploitation of medicinal plants has emerged as one of the major challenges in the biotechnological and pharmacological industries for this decade. Additionally, in response to current global demands, increasing efforts are being made to satisfy the future expectations of plant-derived food and drug production worldwide, thus requiring surface maximization for agricultural and nutritional purposes (Pastor et al., 2019). Consequently, novel approaches must be undertaken in the field of medicinal plant research with the aim of meeting the requirements for the large-scale exploitation of medicinal plants. In this sense, plant tissue culture (PTC) constitutes an efficient pool of methodologies becoming a sustainable platform to achieve true-to-type products with added-value properties (Eibl et al., 2018; Chandran et al., 2020) and requires much less space for achieving the same yields than conventional open field agriculture, due to their ability for scaling-up in bioreactors. This methodology confers an absolute independence of climatic threats, plant pathogens and harsh agriculture management and large storage facilities of plant materials. Nevertheless, PTC should cope with their own difficulties, as it requires the investment for specialized equipment and consumables and the recruitment of trained staff to develop the associated methodologies (Bridgen et al., 2018; Patra et al., 2020).

One of the crucial factors that impact the success of PTC establishment is the optimization of growth conditions and the mineral composition of culture medium (Isah et al., 2018). Although a high number of publications have focused on the study of *in vitro* growth conditions for many species (Batista et al., 2016; Golkar et al., 2019; Hoang et al., 2019), culture medium composition is a paramount factor usually underestimated during the design of plant *in vitro* culture protocols (Nezami-Alanagh et al., 2014; García-Pérez et al., 2020a). Under PTC conditions, the ingredients included in the culture medium constitute the only source of nutrients available for plants and subsequent nutritional imbalances may occur discretely affecting culture development (Niedz and Evens, 2007), reflecting substantial physiological symptoms (Nezami-Alanagh et al., 2019). Thus, PTC media formulations contain a wide spectrum of mineral and organic nutrients that interact in a complex, multifactorial, nonlinear and non-deterministic way, without considering the individual susceptibilities and requirements that discrete species present, leading to the existence of nutritional imbalances, causing underlying deleterious effects that may be easily detectable (Nezami-Alanagh et al., 2019; Phillips and Garda, 2019).

As a rule, culture media components contain, as average, 18 different mineral nutrients, some required at high concentrations (macronutrients) such as nitrogen, potassium, calcium, phosphorus, sulfur and magnesium, while others are required at lower concentrations (micronutrients), such as manganese, zinc, boron, molybdenum, copper and iron, among others, being all essential for certain physiological processes (Twaij et al., 2020). Together with mineral nutrients, a source of carbon, normally sucrose, as well as other organic molecules, such as vitamins and amino acids, some plant growth regulators, are supplied to media to ensure a healthy plant growth and development (Saad and Elshahed, 2012). In addition, there are additional factors that show a significant impact on mineral nutrition, such as the genotype, because even closely related species have been shown to present differential behaviors toward certain media ingredients (Gago et al., 2011; Nezami-Alanagh et al., 2014).

Due to the high heterogeneity of ingredients that make part of culture media formulations and other additional factors, such as plant genotype or growth conditions, the study of nutritional requirements applied to unknown medicinal plants leads to the design of complex multivariate experimental designs (Nezami-Alanagh et al., 2017; Teixeira da Silva et al., 2020). In the last decade (Hesami and Jones, 2020; Niazian and Niedbala, 2020), several ML algorithms have been successfully employed as alternative to traditional statistical methods and/or response surface methodology (RSM) to identify factors and interactions on complex, non-linear and non-deterministic process such as PTC (Landin et al., 2009; Gago et al., 2010a,c; Nezami-Alanagh et al., 2017). Therefore, revealing all the information encrypted over the large amount of experimental results derived from this type of multifactorial processes becomes a highly challenging task. In such cases, machine learning (ML) offers a cutting-edge computer-based methodology with the ability of handling very complex multivariate datasets, in which there are unknown patterns between inputs and outputs or large amount of uncategorized or different kind of data relating with complex processes, being able to transform data into useful information and knowledge (Gago et al., 2010c; Ertel, 2017; Bini, 2018; Freiesleben et al., 2020). On this purpose, different ML algorithms such as artificial neural networks (ANNs); deep neural networks (DNNs); convolutional neural networks (CNNs); support vector machines (SVMs) or random forest (RF) has been used in plant biotechnology (Niazian and Niedbala, 2020) and, particularly, in PTC (Gago et al., 2010a). Among all of them, ANNs have been successfully applied with the aim of establishing robust predictive models that contribute to the optimization and characterization of multifactorial processes (Landin and Rowe, 2013; Gago et al., 2014; Arteta et al., 2018; Villarrubia et al., 2018; Nezami-Alanagh et al., 2019). In addition, the combination of ANNs with fuzzy logic, the so-called neurofuzzy logic, confers several advantages in the search of critical factors that impact plant nutrition, by providing "IF-THEN" rules that make result interpretation easier, in other words, understandable for the human brain (Landin et al., 2009; Gago et al., 2010b; Gallego et al., 2011). Successful applications of neurofuzzy logic in the field of PTC for seed germination (Ayuso et al., 2017), the identification of physiological disorders associated to nutritional imbalances (Nezami-Alanagh et al., 2018, 2019), improvement of bioactive compounds accumulation (García-Pérez et al., 2020b) and revealing the role of phytohormones on plant *in vitro*

organogenesis (García-Pérez et al., 2020a) have been already performed successfully.

Bryophyllum (genus *Kalanchoe*, Crassulaceae family) constitutes a subgenus with more than 25 plant species that have been used in the traditional medicine across both the American and African continents (Stefanowicz-Hajduk et al., 2020). Pharmacognostical and phytochemical analyses have highlighted that phenolic compounds and bufadienolides were the bioactive compounds that develop their therapeutic effects, since *Bryophyllum* spp. have been largely applied to treat infections and chronic diseases, such as diabetes, cardiovascular diseases and cancer (García-Pérez et al., 2018a). The knowledge derived from the combination of ML and PTC will be highly valuable for considering the biotechnological exploitation of *Bryophyllum* spp. in order to take advantage of their added-value properties as a potential source of bioactive compounds.

In this work, we applied the ML (ANNs algorithms), to model and provide insight about the critical factors and their interactions that drive mineral nutrition of three medicinal plants from the subgenus *Bryophyllum* cultured *in vitro*, by focusing on the content in macronutrients and micronutrients from culture media formulations, with the aim of revealing masked nutritional imbalances and interactions that may occur between nutrients that impact plant growth-related parameters.

MATERIALS AND METHODS

Plant Material and Culture Conditions

The establishment of *in vitro* culture was conducted for three different *Bryophyllum* species, namely: *Bryophyllum daigremontianum* Raym. - Hamet et Perr. (syn. *Kalanchoe daigremontinana*, BD), *Bryophyllum* × *houghtonii* D.B. Ward (*Bryophyllum daigremontianum* × *tubiflorum,* syn. *Kalanchoe* × *houghtonii*, BH) and *Bryophyllum tubiflorum* Harv. (syn. *Kalanchoe tubiflora*, BT).

Epiphyllous plantlets from these three species were used for the disinfection and transference to *in vitro* conditions as described in previous works (García-Pérez et al., 2019). After surface disinfection, plantlets were cultured by groups of three in glass culture vessels containing 25 mL of previously autoclaved MS medium (Murashige and Skoog, 1962) supplemented with 3% sucrose and solidified with 0.8% agar at pH = 5.8. Cultures were then transferred to growth chambers and placed randomly in the shelves at 25 ± 1°C under a photoperiod of 16 h light (55 μmol m^{-2} s^{-1}) and 8 h dark and subcultured every 12 weeks by using newly formed epiphyllous plantlets as the explants for next subculture.

Experimental Design

Spontaneously rooted epiphyllous plantlets from the three *Bryophyllum* species, proceeding from 12-week-old plants grown on MS medium, were subjected to nutrition experiments. Plantlets were transferred by pairs into 10 glass culture vessels, grown and subcultured under the same conditions stated above, making a total of 20 replicates per treatment.

For nutrition experiment, nine different culture media formulations, based on MS medium were used. Due to the low mineral requirements associated to Crassulaceae plants, as it is the case of *Bryophyllum* spp. (Phillips and Garda, 2019; García-Pérez et al., 2020b), each media contained proportional reduced contents of both either MS macronutrients (M) or MS micronutrients (μ). Thus, half-concentrations (1/2MSM and 1/2MSμ), quarter-concentrations (1/4MSM and 1/4MSμ), eighth-concentrations (1/8MSM and 1/8MSμ) and macronutrient and micronutrient absence (0MSM, 0MSμ) and, as control, full MS medium was tested (**Table 1**). EDTA-chelated iron, vitamins and organic molecules were supplied in all media at same concentration than in the original MS formulation. All media were also supplemented with 3% sucrose and solidified with 0.8% agar at pH = 5.8.

In order to observe the nutritional long-term impact on *Bryophyllum* growth parameters, four subcultures were performed. At the end of each subculture (every 12 weeks) two newly-formed epiphyllous plantlets per vessel were randomly selected, transferred to fresh medium of the same media (treatment) in which were cultured, and grown in the same conditions. In the treatments in which epiphyllous plantlet formation was not observed (such as 0MS), two rooted and newly-formed epiphyllous plantlets from the control treatment (MS) were selected and cultured in that media. In total, 20 replicates (10 vessels with two plantlets) were kept in each subculture per treatment.

Thus, the experimental design included 3 different genotypes (BD, BH and BT) × 9 different culture media formulations (MS control + 4 macronutrient-reduced formulations + 4 micronutrient-reduced formulations) × 4 subcultures, accounting for a total of 108 different treatments with 20 replicates each.

At the end of each subculture (12 weeks) six physiological parameters were determined in the new epiphyllous plantlets: shoot length (SL, expressed as cm), longest root length (RL, expressed as cm), plantlet number (PN), leaf number (LN), aerial parts fresh weight (AFW, expressed in g) and root fresh weight (RFW, expressed in g).

After experimental data collection from nutrition experiment, all data were merged into one large multifactorial database including 108 treatments following a factorial design for 18 factors (**Supplementary Table 1**). Salts included in culture media were split into their containing ions with the aim of avoiding ion confounding (Niedz and Evens, 2006). As a result, the eighteen factors were selected as the inputs (genotype, subculture number and 16 ions) plus the six physiological parameters as outputs (SL, RL, PN, LN, AFW, and RFW) for building the model. In all cases, results were expressed as the mean ± standard error (**Supplementary Table 1**).

Statistical Analysis

Initially, data derived from the nutrition experiments (SL, RL, PN, LN, AFW, and RFW) were analyzed statistically in order to evaluate the significance of each factor and their interactions (significance level: $\alpha = 0.01$) on the parameters studied. To that end, factorial ANOVA was performed to elucidate the effect of genotype, subculture and culture media and their interactions, followed by Tukey's HSD *post hoc* test ($\alpha = 0.01$). Data

TABLE 1 | Mineral salt composition included in cultured media formulations used in this study.

	Salt	MS (mg L^{-1})	1/2MSM (mg L^{-1})	1/4MSM (mg L^{-1})	1/8MSM (mg L^{-1})	0MSM (mg L^{-1})	1/2MSμ (mg L^{-1})	1/4MSμ (mg L^{-1})	1/8MSμ (mg L^{-1})	0MSμ (mg L^{-1})
Macronutrients	KNO_3	1,900	950	475	237.5	0	1,900			
	NH_4NO_3	1,650	825	412.5	206.3		1,650			
	$CaCl_2$ $2H_2O$	440	220	110	55		440			
	$MgSO_4$ $7H_2O$	370	185	92.5	46.3		370			
	KH_2PO_4	170	85	42.5	21.30		170			
Micronutrients	$MnSO_4$ $4H_2O$	22.3					11.2	5.6	2.8	0
	$ZnSO_4$ $7H_2O$	8.6					4.3	2.2	1.1	
	H_3BO_3	6.2					3.1	1.6	0.78	
	KI	0.83					0.42	0.21	0.10	
	Na_2MoO_4 $2H_2O$	0.25					0.13	0.063	0.031	
	$CuSO_4$ $5H_2O$	0.025					0.013	0.0063	0.0031	
	$CoCl_2$ $6H_2O$	0.025					0.013	0.0063	0.0031	
Iron source	Na_2EDTA	37.25								
	$FeSO_4$ $7H_2O$	27.85								

normality and homoscedasticity was assessed by Kolmogorov-Smirnov's and Levene's tests, respectively. Count data (PN and LN) parameters should be analyzed by Poisson Regression, but as number of replicates was large ($n = 20$) and, thus, ANOVA had the same inference than Poisson Regression ($\alpha = 0.01$; Mize et al., 1999). ANOVA was also applied to those parameters as in previous works (Ayuso et al., 2019). In both cases, the software used was STATISTICA v.12 (StatSoft Inc., 2014, Street Tulsa, OK, USA).

Modeling Tools

Data modeling was performed by using the commercial FormRules® v.4.03 software (Intelligensys LTD, UK) as described elsewhere (Nezami-Alanagh et al., 2018; García-Pérez et al., 2020a). Briefly, FormRules® performed the adaptive-spline-modeling-of-data (ASMOD) algorithm to minimize the number of relevant inputs and to reduce the model complexity and facilitating accuracy with fewer parameters (Shao et al., 2006). Several statistical fitness criteria including cross validation (CV), leave one out cross validation (LOOCV), minimum description length (MDL), Bayesian information criterion (BIC) and structural risk minimization (SRM) were investigated to obtain the model that gave the best Train Set R^2. Two of these, CV and LOOCV, split the data into subsets that are either used for training and testing (validation method), while the others (MDL, BIC and SRM) are statistical significance methods, which use all the data for training. These are designed to avoid overtraining, minimizing a criterion that contains two terms: (i) the number of parameters in the model (the variance) and (ii) the prediction errors computed on the data set (the bias). The best results were found for SRM, which ensured the highest predictability with the minimum generalization error and provided the generation of the simplest and more intelligible rules (Vapnik, 1992).

TABLE 2 | Training parameters for the construction of neurofuzzy model used by FormRules®.

Minimization Parameters (ASMOD)
Ridge regression factor: 1×10^{-6}
MODEL SELECTION CRITERIA
Structural risk minimization (SRM)
C1 = 0.95, C2 = 4.8
Number of set densities: 2
Set densities: 2
Adapt nodes: TRUE
Maximum inputs per submodel: 3
Maximum nodes per input: 15

The training process was conducted as described in detail elsewhere (Shao et al., 2006) and training parameters are summarized in **Table 2**. The quality of submodels (predictability and accuracy), independently generated for every output, were assessed according to the ANOVA f-ratio, mean square error (MSE; Equation 1) and the coefficient of determination (Train Set R^2; Equation 2):

$$MSE = \left(\frac{\sum_{i=1}^{n} (y_i - y_i^{'})^2}{n} \right) \quad (1)$$

$$R^2 = \left(1 - \frac{\sum_{i=1}^{n} (y_i - y_i^{'})^2}{\sum_{i=1}^{n} (y_i - y_i^{''})^2} \right) \times 100 \quad (2)$$

Where y_i represents the experimental value from the data set, y_i' represents the predicted value by the model, and y_i'' represents the mean of the dependent variable. MSE are calculated to provide information about the random error component of the built model, thus indicating the usefulness of model fitting

for prediction due to a smaller incidence of random error (Hesami and Jones, 2020). Models with high Train Set R^2 (>70%) and an f-ratio (>4) assess model accuracy and no statistical differences among predicted and experimental values. Models with higher values than 99.9% should be rejected due to model over-fitting (Colbourn and Rowe, 2005; Landin et al., 2009).

The results provided by the application of neurofuzzy logic were expressed as "IF-THEN" rules, thus making their interpretation easier, and they were given a range level (from low to high), combined with a corresponding membership degree value, that ranges between 0 and 1 (Gallego et al., 2011). **Supplementary Figure 1** is attached to facilitate the understanding of the linguistic expressions of the variables

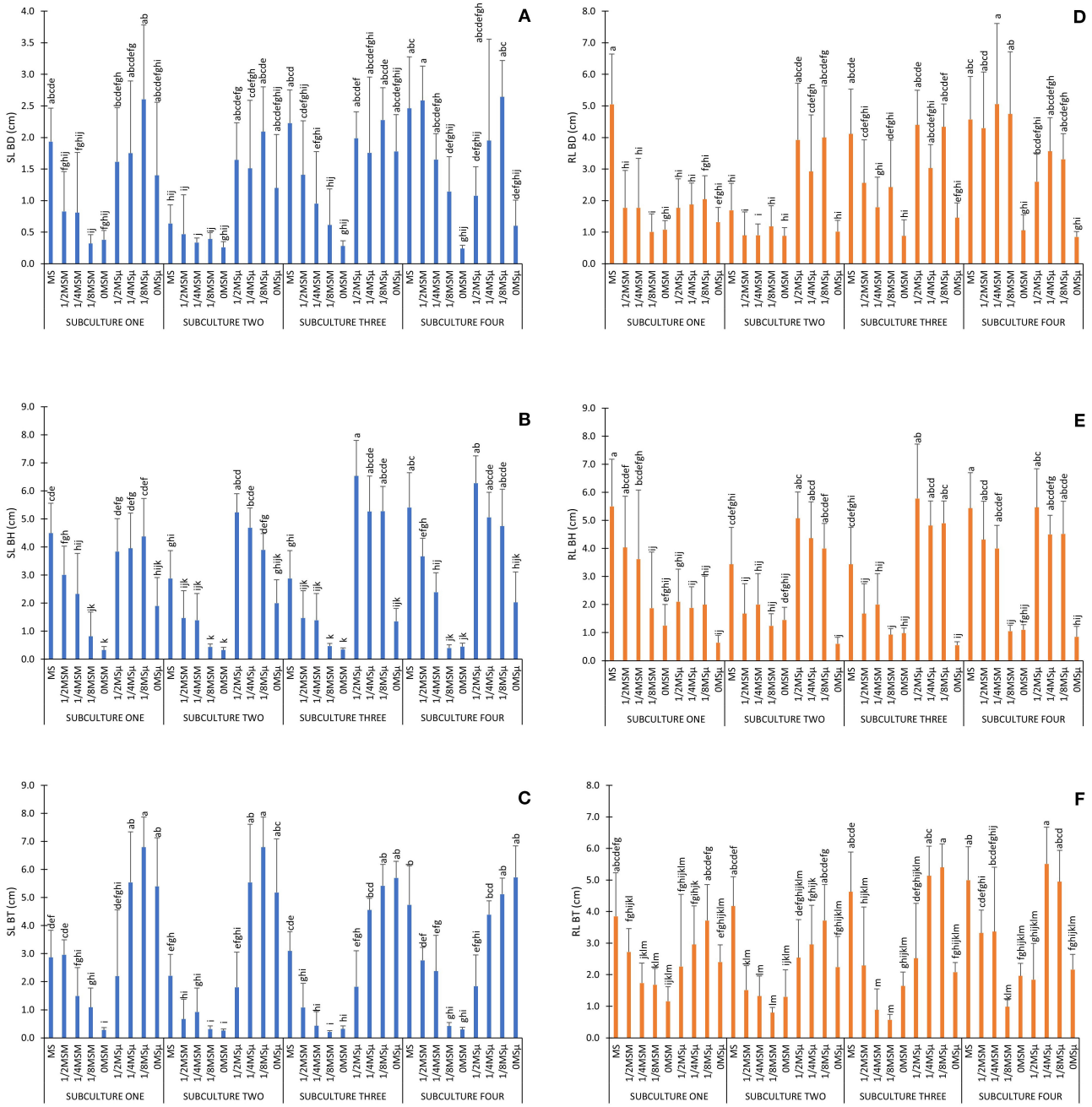

FIGURE 1 | Experimental data obtained for SL and RL. Values are expressed as the mean and vertical bars indicate standard deviation. Different letters indicate significant differences (α = 0.01). **(A)** SL BD (cm); **(B)** SL BH (cm); **(C)** SL BT (cm); **(D)** RL BD (cm); **(E)** RL BH (cm); **(F)** RL BT (cm).

obtained by the neurofuzzy logic model (Low, Medium and High).

RESULTS

Traditionally plant tissue researchers used statistical method such as factorial ANOVA to analyze data and test if two or more treatment differ significantly from each other due to the effect of some independent variables (factors), but also serves to infer cause-effect relationships. Here, data analysis using factorial ANOVA reflected that all factors studied (genotype, subculture number and culture media) and all interactions among them, caused a significance effect on all parameters studied, except for the interaction genotype × subculture on RL, PN, AFW and RFW parameters ($\alpha = 0.01$; **Supplementary Table 2**). The combined effect of the interactions between the number of subcultures and the culture medium is graphically exemplified for SL and RL in **Figure 1**, depending on each genotype studied: BD, BH and BT, respectively.

In general, MS full strength media promoted high SL and RL in all genotypes, although the final shoot and root length varied significantly depending on the genotype and subculture (**Figure 1**). Those treatments with reduced micronutrients (1/2MSμ, 1/4MSμ, and 1/8MSμ) promoted more length than those with reduced macronutrients (1/2MSM, 1/4MSM, and 1/8MSM), and the treatments with absence of minerals in the media, particularly without macronutrients (0MSM), promoted the worst results overall (**Figure 1**). The same effect, can be observed for the rest of parameters (see **Supplementary Table 1**), particularly for PN, which measures the organogenesis capacity, since the absence of mineral nutrients causes the total inhibition of the generation of new epiphyllous plantlets in the leaf margins of all species (**Supplementary Table 1**: treatments 9, 18, 27, 36 for BD; 45, 54, 63, 72 for BH, and 81, 90, 99, 108 for BT). All together, these ANOVA results showed that MS media composition can be modified dramatically by reducing the amount of macro and micronutrients and obtain exactly the same results. But little information was obtained about the effect of each mineral nutrient or its role on the effect observed.

Data modeling by neurofuzzy logic emerges as a solution to provide useful knowledge on *Bryophyllum* mineral nutrition after training and learning from the experimental data. The model showed high predictability with Train Set R^2 values higher than 70%, f-ratio >4 and low values of MSE in all submodels (**Table 3**). In addition, model accuracy was assessed by ANOVA f-critical, which proved that predicted values from the model did not show statistically significant differences with respect to the experimental values for any of the outputs ($\alpha = 0.05$; **Table 3**).

For SL, the model generated two submodels being the interaction between genotype and copper the one with the highest contribution (**Table 3**). As previously stated, neurofuzzy model had the ability of unraveling masked interactions between different factors, once the salt media composition of each treatment was formulated as their ion composition, to avoid ion confounding. An additional submodel showed that NH_4^+ concentration also caused a significant effect but with lower contributions to SL prediction, than the interaction of genotype × Cu^{2+} (**Table 3**). In the case of RL, two submodels were found, being the triple interaction between the genotype, SO_4^{2-} and MoO_4^{2-} the one presenting the major contribution to RL. Additionally, a second submodel for RL included the interaction of number of subcultures with sodium (**Table 3**). Exactly the same submodels were also predicted for PN and, interestingly, this finding suggests that sulfur and molybdenum play a crucial role not only on *Bryophyllum* nutrition, but on controlling its asexual reproduction. The second submodel spotted the interaction of number of subcultures with sodium for PN, too. Concerning LN only one model was generated, being predicted by the interaction between the genotype and NH_4^+ concentration (**Table 3**), what indicates that this output is closely related and highly influenced by this nitrogen-containing macronutrient ion. In the case of AFW, the interaction between genotype, SO_4^{2-} and MoO_4^{2-} was the only factor spotted as the most significant affecting this output, in the same way than RL and PN. Finally, RFW was predicted by two submodels, showing the interaction between the genotype, Cu^{2+} and SO_4^{2-} the major contribution (**Table 3**), being the only output that was dependent on copper besides SL. An additional submodel for RFW was predicted by molybdate concentration.

TABLE 3 | Quality parameters and critical factors detected by neurofuzzy logic model.

Output	Submodel	Significant inputs	MSE	Train Set R^2	f ratio	df_1, df_2	f critical ($\alpha = 0.05$)
SL	1	NH_4^+	0.92	74.97	42.79	7, 100	2.10
	2	**Genotype × Cu^{2+}**					
RL	**1**	**SO_4^{2-} × MoO_4^{2-} × Genotype**	0.74	76.97	10.91	25, 82	1.64
	2	Subculture × Na^+					
PN	**1**	**SO_4^{2-} × MoO_4^{2-} × Genotype**	60.60	72.85	8.76	25, 82	1.64
	2	Subculture × Na^+					
LN	**1**	**Genotype × NH_4^+**	9.00	72.90	45.29	6, 101	2.19
AFW	**1**	**SO_4^{2-} × MoO_4^{2-} × Genotype**	0.04	89.89	43.66	18, 89	1.72
RFW	**1**	**Genotype × Cu^{2+} × SO_4^{2-}**	0.001	79.69	25.75	14, 93	1.80
	2	MoO_4^{2-}					

Bold inputs indicate the strongest effect associated to each output.

In general, the application of neurofuzzy logic identified all the significant factors on all the outputs related to *Bryophyllum in vitro* growth and their concealed interactions. Nevertheless, the information conferred by this machine learning-based tool was useful, thanks to the establishment of "IF-THEN" rules, which described how these inputs influenced their corresponding outputs. The full set of rules can be found in **Supplementary Table 3**, while the rules including the highest membership degrees for each output were summarized in **Table 4**. In order to make result interpretation easier, all factors were ranged as low, mid and high at the same time, according to their effect on every output and the experimental space tested. The graphical ranging of each ion can be visualized in **Supplementary Figure 1**.

As previously stated, the observed changes in the output SL was mainly caused by the interaction between genotype and Cu^{2+} and, besides that, the model generated the corresponding rules. The rules for SL indicate that high values were obtained in the case of BT when cultured under Low Cu^{2+} concentrations (<0.05 μM) with a membership degree of 0.77 (rule 7; **Table 4**). In fact, it was the only case that presented a High SL response related to copper. In contrast, the lowest SL value (showing a membership degree of 1.00) was obtained for BD under Low Cu^{2+} concentrations (rule 5; **Table 4**). Concerning the rules associated to the other submodels, High SL values were obtained under High NH_4^+ concentrations (>10.31 mM) with high membership (0.83, rule 2; **Table 4**). These results suggest a predominant effect of genotype and copper, over the NH_4^+, on SL.

RL, PN and AFW were predicted mainly by the interaction of three factors: genotype, SO_4^{2-} and MoO_4^2. Thus, the response with the highest contribution to High RL values (membership 1.00) is the interaction between Low SO_4^{2-} concentrations (<1.01 mM) with Mid MoO_4^{2-} concentrations (0.25–0.75 μM) for BT (rule 14 for RL, rule 40 for PN, and rule 72 for AFW; **Table 4**). On the contrary, the interaction between Low SO_4^{2-} concentrations (<1.01 mM) with Low MoO_4^{2-} concentrations (<0.25 μM) for BH showed the highest contribution to Low RL, PN and AFW values (rules 9, 35, and 67, respectively; **Table 4**). In addition, RL and PN presented another submodel, based on the interaction between the number of subcultures and sodium (**Table 3**). In all cases, the model rules showed that High Na^+ concentrations (0.22 mM) caused High RL and PN in all subcultures (rules 27–34 for RL, and 53–60 for PN; **Table 4**).

LN was exclusively predicted by the interaction between the genotype and NH_4^+ concentrations. High LN values were predicted only by High NH_4^+ concentrations for BT (>10.31 mM, membership degree 0.74, rule 66; **Table 4**), showing Low values for the rest of cases, specially BD at Low NH_4^+ concentrations (membership degree 0.96, rule 63; **Table 4**). These results suggest a predominant role of genotype, as LN was exclusively favored on BT, only if High NH_4^+ concentrations were supplied into the media.

Finally, the major submodel predicting RFW included the interaction between genotype, SO_4^{2-} and Cu^{2+} concentrations. The High RFW values with the highest membership degree (1.00) were obtained for BH cultured under High SO_4^{2-} concentrations (>0.88 mM) and Low Cu^{2+} concentrations (<0.05 μM, rule 87; **Table 4**). Meanwhile, Low values were predicted by Low concentrations of both ions (membership degree 1.00, rule 85; **Table 4**), independently of the genotype used (rules 85, 89 and 93; **Table 4**). In addition, the second submodel generated for RFW pointed at molybdate concentrations as the significant output, causing Low RFW values in all cases (rules 97–99; **Table 4**). These results suggest a predominant role of genotype, favored when High SO_4^{2-} and/or Cu^{2+} concentrations were included into the media.

DISCUSSION

The combination of artificial neural networks (ANNs) with fuzzy logic, called neurofuzzy logic, constitutes ML algorithms used for predicting and identifying critical factors of multifactorial nonlinear systems (Shihabudheen and Pillai, 2017), as it is the case of plant *in vitro* nutrition (Gallego et al., 2011). Advantages of ANNs algorithms over traditional statistics have been pointed out previously (Landin et al., 2009; Gago et al., 2010a,c). In this work, the application of neurofuzzy confers a simple and efficient solution about which factors determined the effects found on each *Bryophyllum* growth parameter, by extracting the knowledge among the deep interactions learnt after data training.

Genotype was a widely distributed factor identified for the prediction of all outputs either alone or in combination with one or two additional factors. This indicates that, although the three species of the *Bryophyllum* subgenus are considered closely genetically related, each species shows different nutritional requirements, including macronutrients and micronutrients. These differences may probably be due to the transcriptional regulation of uptake systems, such as the primary response to nutrient limitation conditions, since they are highly inducible by environmental conditions (Bird, 2015). Thus, the establishment of *in vitro* culture results in an effective system to test nutritional imbalances, as it eliminates the influence of side biotic or abiotic factors that impact mineral acquisition, such as pathogen and soil-mediated interactions (Comerford, 2005; Ferrante et al., 2017). These differential patterns for *Bryophyllum* species have already been related to leaf morphology (Chernetskyy et al., 2018) and other discrepancies in physiological processes, such as the biosynthesis of phenolic compounds (Fürer et al., 2016; Bogucka-Kocka et al., 2018; García-Pérez et al., 2020a) and organogenesis (García-Pérez et al., 2020b). Furthermore, the specific growth responses predicted by the ANN model, denote that *Bryophyllum* spp. present a tight range of concentrations to achieve an efficient mineral nutrition (Shrivastav et al., 2020).

Another factor, associated to PTC technology and identified by the model as critical, was the number of subcultures. The number of subcultures was identified in combination with sodium to be significant in a secondary submodel for RL and PN. The differential number of subcultures required

TABLE 4 | Summary of "IF-THEN" rules generated by ANN modeling.

Rules		Genot	Subcult	NH_4^+	SO_4^{2-}	Na^+	Cu^{2+}	MoO_4^{2-}		SL	RL	PN	LN	AFW	RFW	Membership
1	I			Low					T	Low						0.98
2	F			High					H	High						0.83
5		**BD**					**Low**		E	**Low**						**1.00**
6		BD					High		N	Low						1.00
7		**BT**					**Low**			**High**						**0.77**
8		BT					High			Low						1.00
9	I	**BH**			**Low**			**Low**	T		**Low**					**1.00**
10	F	BD			Low			Low	H		Low					1.00
11		BT			Low			Low	E		Low					1.00
12		BH			Low			Mid	N		High					1.00
13		BD			Low			Mid			High					1.00
14		**BT**			**Low**			**Mid**			**High**					**1.00**
15		BH			Low			High			Low					1.00
16		BD			Low			High			Low					1.00
17		BT			Low			High			Low					1.00
18		BH			High			Low			High					1.00
19		BD			High			Low			High					1.00
20		BT			High			Low			High					1.00
21		BH			High			Mid			High					1.00
22		BD			High			Mid			High					1.00
23		BT			High			Mid			Low					1.00
24		BH			High			High			Low					1.00
25		BD			High			High			Low					1.00
26		BT			High			High			Low					1.00
27			ONE			Low					Low					1.00
28			ONE			High					High					1.00
29			TWO			Low					Low					1.00
30			TWO			High					High					1.00
31			THREE			Low					Low					1.00
32			THREE			High					High					1.00
33			FOUR			Low					Low					1.00
34			FOUR			High					High					1.00
35	I	**BH**			**Low**			**Low**	T			**Low**				**1.00**
36	F	BD			Low			Low	H			Low				1.00
37		BT			Low			Low	E			Low				1.00
38		BH			Low			Mid	N			Low				1.00
39		BD			Low			Mid				High				1.00
40		**BT**			**Low**			**Mid**				**High**				**1.00**
41		BH			Low			High				Low				1.00
42		BD			Low			High				Low				1.00
43		BT			Low			High				Low				1.00
44		BH			High			Low				High				1.00
45		BD			High			Low				High				1.00
46		BT			High			Low				High				1.00
47		BH			High			Mid				High				1.00
48		BD			High			Mid				Low				1.00
49		BT			High			Mid				Low				1.00
50		BH			High			High				Low				1.00
51		BD			High			High				Low				1.00
52		BT			High			High				Low				1.00
53			ONE			Low						Low				1.00
54			ONE			High						High				1.00

(Continued)

TABLE 4 | Continued

Rules		Genot	Subcult	NH_4^+	SO_4^{2-}	Na^+	Cu^{2+}	MoO_4^{2-}		SL	RL	PN	LN	AFW	RFW	Membership
55			TWO			Low						Low				1.00
56			TWO			High						High				1.00
57			THREE			Low						Low				1.00
58			THREE			High						High				1.00
59			FOUR			Low						Low				1.00
60			FOUR			High						High				1.00
63	I	**BD**		**Low**					T				**Low**			**0.96**
66	F	**BT**		**High**					H E N				**High**			**0.74**
67	I	**BH**			**Low**			**Low**	T					**Low**		**1.00**
68	F	BD			Low			Low	H E N					Low		1.00
69		BT			Low			Low						Low		1.00
70		BH			Low			Mid						Low		1.00
71		BD			Low			Mid						Low		1.00
72		**BT**			**Low**			**Mid**						**High**		**1.00**
73		BH			Low			High						Low		1.00
75		BT			Low			High						Low		1.00
76		BH			High			Low						High		1.00
77		BD			High			Low						High		1.00
78		BT			High			Low						High		1.00
79		BH			High			Mid						High		1.00
80		BD			High			Mid						High		1.00
81		BT			High			Mid						Low		1.00
85	I	**BH**			**Low**		**Low**		T						**Low**	**1.00**
86	F	BH			Low		High		H E N						High	1.00
87		**BH**			**High**		**Low**								**High**	**1.00**
88		BH			High		High								High	1.00
89		BD			Low		Low								Low	1.00
90		BD			Low		High								High	1.00
91		BD			High		Low								High	1.00
92		BD			High		High								High	1.00
93		BT			Low		Low								Low	1.00
94		BT			Low		High								High	1.00
95		BT			High		Low								High	1.00
96		BT			High		High								High	1.00
97								Low							Low	1.00
98								Mid							Low	1.00
99								High							Low	1.00

Inputs with the highest membership degree, showing the major contibutions for each response on every output, are indicated in bold. Only the rules with the highest membership degree were selected Genot: genotype; Subcult: number of subcultures. SL and RL were expressed in cm. AFW and RFW were expressed in g.

to achieve certain responses reveals that nutrient deficiencies may be sensed at different periods during the culture time. The delay in responses under nutritional deficiencies, may be explained as a consequence of the induced stress triggered by the increased synthesis of signaling molecules, such as nitric oxide and reactive oxygen species (ROS), mainly driven by macronutrient limitations and micronutrient limitation to a lesser degree (Hajiboland, 2012; Pérez-Pérez et al., 2012; Buet et al., 2019). Its importance relies on the decrease in the rate of epigenetic variation after successive subcultures (Smulders and de Klerk, 2011). This factor becomes crucial to assess the genetical stability of *in vitro*-cultured plants, making their valorization easier. Moreover, long-term subcultures constitute an efficient strategy to improve interesting phenomena under *in vitro* conditions, such as rooting (Mendonça et al., 2019; Wang and Yao, 2020), plant regeneration (Konar et al., 2019) and callus induction (Nakasha et al., 2016), very useful for biotechnological production of by-products from medicinal plants.

Among the nutrients, the model has identified as critical factors, the macronutrients NH_4^+ and SO_4^{2-}, and the micronutrients Cu^{2+}, Na^+, and MoO_4^{2-}. The effect of ammonia, the source of nitrogen along with nitrate in most culture media, is clear on the SL parameter. Furthermore, its effect varies depending on the genotype for LN (**Table 3**). Nitrogen plays a controversial role on crassulacean species, such as *Bryophyllum* spp. Differential rates of crassulacean acid metabolism (CAM) have been observed as a function of nitrogen source (Pereira et al., 2017). Thus, two groups are distinguished: nitrate-enhanced CAM species and ammonium-enhanced CAM species (Rodrigues et al., 2014). Nevertheless, there is no general rule for this classification, since different *Bryophyllum* species show particular preferences toward both nitrogen sources (Pereira and Cushman, 2019). Our results suggest that nitrogen source preferences is species-dependent. The effects caused by NH_4^+ on CAM activity are mainly negative, due to the inhibition of nocturnal transport rates of organic acids into the vacuole and the cost, in terms of energy, required for ammonium mobilization (Lüttge et al., 2000; Britto et al., 2001). However, it could be noted that such paradigm has been established for soil-grown plants and nitrogen influence may not be the same under *in vitro* conditions. In this sense, only BT presented high LN values under high NH_4^+ concentrations (rule 66), while BD and BH always showed low values, whatever the ammonium supply was within the limits of the study (rules 61–65; **Supplementary Table 3**). Secondarily, high SL values were obtained under high NH_4^+ concentrations (rule 2; **Table 4**). In addition to being an essential nutrient for plant development, NH_4^+ has recently been identified as a signal molecule that triggers both, physiological and morphological responses (Liu and von Wirén, 2017). A recent report has shown that ammonium concentration lower than 15 mM improves the biosynthesis of phenolic compounds in the aerial parts of *Bryophyllum* spp. cultured *in vitro* as a consequence of a physiological response induced by abiotic stress (García-Pérez et al., 2020a). Consequently, in order to characterize the impact of ammonium on *Bryophyllum* spp., further studies at a molecular level are required.

The model reveals a close relationship between sulfur and molybdenum, whose effects on RL, PN and AFW parameters depend on the genotype (**Table 3**). This close interaction between both nutrients has been already reported in other species, since molybdenum is a crucial component of molybdoenzymes, involved in sulfur metabolism, being both nutrients essential for the development of aerial tissues and roots (Mendel and Hänsch, 2002; Naqib and Jahan, 2017; Blasco et al., 2018; Bouranis et al., 2020). Furthermore, due to the identical chemical configuration of both ions, the uptake of molybdate and sulfate by roots, may take place via sulfur-specific receptors found in root tissues, thus enabling the co-absorption of molybdate with sulfate (Ali et al., 2020). This could be suggested according to the model results, since low RL, PN and AFW values were observed under low SO_4^{2-} concentrations and high MoO_4^{2-} concentrations, and conversely, RL, PN and AFW values were high under high SO_4^{2-} concentrations and low MoO_4^{2-} concentrations (**Table 4**). In the case of RL, sulfate plays a positive role, by promoting root biomass accumulation and nutrient uptake during root growth (Alarcón-Poblete et al., 2018), in agreement with the results describe here for RFW (**Table 3**). These observations indicate that sulfur is essential for *in vitro* root development on *Bryophyllum* spp., although molybdate at mid concentrations may assist to its function when sulfate is limited, as demonstrated for BT (rules 14, 40, and 72; **Table 4**), and reported by other authors (Alhendawi et al., 2005; Shinmachi et al., 2010). Moreover, similar results were obtained for PN.

PN was the only output associated to reproduction of *Bryophyllum* spp., since this process constitutes the mechanism developed by these species for their asexual reproduction. It combines several processes belonging to both organogenesis and embryogenesis phenomena, that have not been fully elucidated to date (Garcês et al., 2014). Such process takes place at the margin of adult leaves and is considered the major mechanism driving *Bryophyllum* clonal invasiveness, as it has been reported for BD, BH and BT (Guerra-García et al., 2015). Fully developed plantlets require the formation of their proper root systems before detaching from mother plants to form new functional clones. The development of such process may explain the close relationship between RL and PN according to their same critical factors spotted by the model (**Table 3**), since rooting should occur at both adult plants and newly-formed plantlets. Additionally, RL and PN presented a secondary submodel, indicating that high sodium concentrations were required for high values of both parameters from the first subculture (**Table 4**; >0.223 mM), which is in accordance to the sodium requirements previously reported for BT and other Crassulaceae species (George et al., 2008). In this sense, this asexual reproductive mechanism makes *Bryophyllum* spp. a suitable biological system for the establishment of *in vitro* culture, thanks to their constitutive plantlet formation. Furthermore, PN was strongly dependent on genotype, and it could be explained because of the mechanisms of plantlet formation: BH and BD develop this process along the whole leaf margins (Garcês and Sinha, 2009; Herrando-Moraira et al., 2020), whereas PN is restricted to the distal leaf end in BT (Guerra-García et al., 2018), thus potentially causing such genotype influence (**Figure 2**).

Copper was revealed as the most influential micronutrient on *Bryophyllum* spp. cultured *in vitro*, since it was selected as a critical factor for SL, interacting with genotype, and RFW, interacting with genotype and sulfate concentrations (**Table 3**). The results from the neurofuzzy logic model indicate that this nutrient affects *Bryophyllum* physiology in a species-dependent manner, as it was always found in combination with genotype. In this sense, the interaction between genotype and copper was the major factor that influenced SL. In fact, this nutrient played a differential role among the three genotypes. Only BT showed high SL values cultured under Cu^{2+} concentrations below 0.05 μM (rule 7). This indicates that BT was the genotype most affected negatively by copper, suggesting that toxicity events may occur for this species. These findings reveal that a fine-tuned control of copper homeostasis is required to prevent its toxicity due to a copper excess, since this nutrient is essential for the correct cell function, being part of highly important

FIGURE 2 | Plantlet formation in *Bryophyllum* spp. cultured *in vitro*. **(A)** Plantlets forming along the leaf margins on BD. **(B)** Plantlets forming along the leaf margins on BH. **(C)** Plantlets forming at the distal leaf end on BT.

metalloproteins as a cofactor (Printz et al., 2016; García-Pérez et al., 2018b). Its importance relies on its contribution to basic physiological processes in plants, such as early plant growth, photosynthetic efficiency, mitochondrial respiration and the impairment of oxidative stress (Schulten and Kramer, 2017; Blasco et al., 2018). This wide influence on plant physiology could aid explaining why this nutrient was crucial for SL, associated to aerial part development, SL, and linked to root formation and growth, RFW. Such hypothesis is reinforced by the sophisticated mechanism of copper distribution within plant tissues that contains preventive molecular mechanisms enabling its accumulation by preventing eventual toxic effects at root level (Castro et al., 2018; Migocka and Malas, 2018). In the case of RFW, the coordinate action of sulfate with copper, as described by the ML model (**Table 3**), may indicate that sulfur contributes to such copper-induced toxic prevention, as it was earlier stated to other metals. In addition, RFW presented a second submodel that included molybdate as a critical factor, that was shown as a negative factor on this parameter, according to the model rules (rules 97–99). This observation can be justified by the effects reported for molybdenum excess, including a severe impairment of photosynthetic efficiency and the inhibition of rooting in other species (Arif et al., 2016). Thus, our results suggest that a minimum copper supplementation ($< 0.05\,\mu M$) may efficiently contribute to *in vitro*-cultured *Bryophyllum* plant growth.

In conclusion, our results show that the lack of specific culture media forces the use of universal formulations, as MS medium. Although such formulations contain complete combinations of mineral essential nutrients, may suppose supra-optimal concentrations for the cultivation of many species (Phillips and Garda, 2019), particularly for little-studied medicinal plants with high phytochemical potential. The nutritional imbalances spotted by ML offered a source of knowledge for the prediction of critical factors affecting *Bryophyllum* spp. plant *in vitro* culture. Through the ion split approach, neurofuzzy logic model was able to shed light on the masked interactions that take place during the *in vitro* culture of three different *Bryophyllum* species, by additionally highlighting the importance of related factors, such as the genotype and the number of subcultures. The essentiality of achieving an enhanced nutritional profiling for the correct development of medicinal plants, as it is the case of *Bryophyllum* spp., is a paramount feature that should be successfully accomplished in order to get a sustainable exploitation of such species with the aim of reaching large-scale applications in several fields, such as food, cosmetics and nutraceutical industries.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PG-P and EL-M performed the experiments. PG contributed with reagents and materials. PG-P, ML, and PG contributed with modeling and analysis tools. PG-P and PG conceived the experimental design and wrote the manuscript. All authors contributed in the revision and acceptance version of the manuscript.

FUNDING

This study was funded by Xunta de Galicia through Red de Uso Sostenible de los Recursos Naturales y Agroalimentarios (REDUSO, Grant Number ED341D 2017/2018) and Cluster

of Agricultural Research and Development (CITACA Strategic Partnership, Grant Number ED431E 2018/07).

ACKNOWLEDGMENTS

The authors declare their sincere acknowledgments to the Spanish Ministry of Education for the FPU grant awarded to PG-P (Grant Number FPU15/04849) and Clinical Research Center ADICAM for kindly providing the plant material used in this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.576177/full#supplementary-material

REFERENCES

Alarcón-Poblete, E., Inostroza-Blancheteau, C., Alberdi, M., Rengel, Z., and Reyes-Díaz, M. (2018). Molecular regulation of aluminum resistance and sulfur nutrition during root growth. *Planta* 247, 27–39. doi: 10.1007/s00425-017-2805-6

Alhendawi, R. A., Kirkby, E. A., and Pilbeam, D. J. (2005). Evidence that sulfur deficiency enhances molybdenum transport in xylem sap of tomato plants. *J. Plant Nutr.* 28, 1347–1353. doi: 10.1081/PLN-200067449

Ali, A., Bhat, B. A., Rather, G. A., Malla, B. A., and Ganie, S. A. (2020). "Proteomic studies of micronutrient deficiency and toxicity," in *Plant Micronutrients*, eds T. Aftab and K. R. Hakeem (Cham: Springer Nature Switzerland AG), 257–284.

Arif, N., Yadav, V., Singh, S., Singh, S., Ahmad, P., Mishra, R. K., et al. (2016). Influence of high and low levels of plant-beneficial heavy metal ions on plant growth and development. *Front. Environ. Sci.* 4:69. doi: 10.3389/fenvs.2016.00069

Arteta, T. A., Hameg, R., Landin, M., Gallego, P. P., and Barreal, M. E. (2018). Neural networks models as decision-making tool for *in vitro* proliferation of hardy kiwi. *Eur. J. Horticult. Sci.* 83, 259–265. doi: 10.17660/eJHS.2018/83.4.6

Ayuso, M., García-Pérez, P., Ramil-Rego, E., Gallego, P. P., and Barreal, M. E. (2019). *In vitro* culture of the endangered plant *Eryngium viviparum* as dual strategy for its *ex situ* conservation and source of bioactive compounds. *Plant Cell Tissue Organ Cult.* 138, 427–435. doi: 10.1007/s11240-019-01638-y

Ayuso, M., Ramil-Rego, P., Landin, M., Gallego, P. P., and Barreal, M. E. (2017). Computer-assisted recovery of threatened plants: keys for breaking seed dormancy of *Eryngium viviparum*. *Front. Plant Sci.* 8:2092. doi: 10.3389/fpls.2017.02092

Batista, D. S., de Castro, K. M., da Silva, A. R., Teixeira, M. L., Sales, T. A., Soares, L. I., et al. (2016). Light quality affects *in vitro* growth and essential oil profile in *Lippia alba* (Verbenaceae). *In Vitro Cell. Dev. Biol. Plant* 52, 276–282. doi: 10.1007/s11627-016-9761-x

Bini, S. A. (2018). Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J. Anthroplasty* 33, 2358–2361. doi: 10.1016/j.arth.2018.02.067

Bird, A. J. (2015). Cellular sensing and transport of metal ions: implications in micronutrient homeostasis. *J. Nutr. Biochem.* 26, 1103–1115. doi: 10.1016/j.jnutbio.2015.08.002

Blasco, B., Navarro-León, E., and Ruiz, J. M. (2018). "Oxidative stress in relation with micronutrient deficiency or toxicity," in *Plant Micronutrient Use Efficiency*, eds M. A. Hossain, T. Kamiya, D.J., Burritt, L.-S. P. Tran, and T. Fujiwara (London: Academic Press, Elsevier), 181–194.

Bogucka-Kocka, A., Zidorn, C., Kasprzycka, M., Szymczak, G., and Szewczyk, K. (2018). Phenolic acid content, antioxidant and cytotoxic activities of four *Kalanchoë* species. *Saudi J. Biol. Sci.* 25, 622–630. doi: 10.1016/j.sjbs.2016.01.037

Bouranis, D. L., Malagoli, M., Avice, J. C., and Bloem, E. (2020). Advances in plant sulfur research. *Plants* 9, 1–6. doi: 10.3390/plants9020256

Bridgen, M. P., Van Houtven, W., and Eeckhaut, T. (2018). "Plant tissue culture techniques for breeding," in *Ornamental Crops*, ed J. Van Huylenbroeck (Amsterdam: Springer International Publishing AG), 127–144.

Britto, D. T., Siddiqi, M. Y., Glass, A. D. M., and Kronzucker, H. J. (2001). Futile transmembrane NH4+ cycling: a cellular hypothesis to explain ammonium toxicity in plants. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4255–4258. doi: 10.1073/pnas.061034698

Buet, A., Galatro, A., Ramos-Artuso, F., and Simontacchi, M. (2019). Nitric oxide and plant mineral nutrition: current knowledge. *J. Exp. Bot.* 70, 4461–4476. doi: 10.1093/jxb/erz129

Castro, P. H., Lilay, G. H., and Assunção, A. G. L. (2018). "Regulation of micronutrient homeostasis and deficiency response in plants," in *Plant Micronutrient Use Efficiency*, eds M.A. Hossain, T. Kamiya, D.J., Burritt, L.-S.P. Tran, and T. Fujiwara (London: Academic Press, Elsevier), 1–15.

Chandran, H., Meena, M., Barupal, T., and Sharma, K. (2020). Plant tissue culture as a perpetual source for production of industrially important bioactive compounds. *Biotechnol. Rep.* 26:e00450. doi: 10.1016/j.btre.2020.e00450

Chernetskyy, M., Wozniak, A., Skalska-Kamińska, A., Zuraw, B., Blicharska, E., Rejdak, R., et al. (2018). Structure of leaves and phenolic acids in *Kalanchoë daigremontinana* Raym.-Hamet & H. Perrier. *Acta Sci. Polonorum Hortorum Cultus* 17, 137–155. doi: 10.24326/asphc.2018.4.13

Colbourn, E., and Rowe, R. C. (2005). "Neural computer and pharmaceutical formulation," in *Encyclopedia of Pharmaceutical Technology*, eds J. Swarbrick and J. C. Boylan (New York, NY: Marcel Dekker).

Comerford, N. B. (2005). "Soil factors affecting nutrient bioavailability," in *Nutrient Acquisition by Plants: An Ecological Perspective*, ed H. Bassirirad (Berlin: Springer Verlag Berlin Heidelberg), 1–14.

Eibl, R., Meier, P., Stutz, I., Schildberger, D., Hühn, T., and Eibl, D. (2018). Plant cell culture technology in the cosmetic and food industries: current state and future trends. *Appl. Microbiol. Biotechnol.* 102, 8661–8675. doi: 10.1007/s00253-018-9279-8

El Sheikha, A. F. (2017). "Medicinal plants: ethno-uses to biotechnology era," in *Biotechnology and Production of Anti-Cancer Compounds*, ed S. Malik (Cham: Springer International Publishing AG), 1–38.

Ertel, W. (2017). "What is artificial intelligence?" in *Introduction to Artificial Intelligence*, ed I. Mackie (Cham: Springer International Publishing AG), 1–20.

Ferrante, A., Nocito, F. F., Morgutti, S., and Sacchi, G. A. (2017). "Plant breeding for improving nutrient uptake and utilization efficiency," in *Advances in Research on Fertilization Management of Vegetable Crops*, eds F. Tei, S. Nicola, and P. Benincasa (Cham: Springer International Publishing AG), 221–246.

Freiesleben, J., Keim, J., and Grutsch, M. (2020). Machine learning and design of experiments: alternative approaches or complementary methodology for quality improvement? *Qual. Reliabil. Eng. Int.* 36, 1837–1848. doi: 10.1002/qre.2579

Fürer, K., Simões-Wüst, A. P., von Mandach, U., Hamburger, M., and Potterat, O. (2016). *Bryophyllum pinnatum* and related species used in anthroposophic medicine: constituents, pharmacological activities, and clinical efficacy. *Planta Med.* 82, 930–941. doi: 10.1055/s-0042-106727

Gago, J., Landin, M., and Gallego, P. P. (2010a). Strengths of artificial neural networks in modeling complex plant processes. *Plant Signal. Behav.* 5, 743–745. doi: 10.4161/psb.5.6.11702

Gago, J., Landin, M., and Gallego, P. P. (2010b). A neurofuzzy logic approach for modeling plant processes: A practical case of *in vitro* direct rooting and acclimatization of *Vitis vinifera* L. *Plant Sci.* 179, 241–249. doi: 10.1016/j.plantsci.2010.05.009

Gago, J., Martínez-Núñez, L., Landin, M., Flexas, J., and Gallego, P. P. (2014). Modelling the effects of light and sucrose on *in vitro* propagated plants: a multiscale system analysis using artificial intelligence technology. *PLoS ONE* 9:e85989. doi: 10.1371/journal.pone.0085989

Gago, J., Martínez-Núñez, L., Landin, M., and Gallego, P. P. (2010c). Artificial neural networks as an alternative to the traditional statistical methodology in plant research. *J. Plant Physiol.* 167, 23–27. doi: 10.1016/j.jplph.2009.07.007

Gago, J., Pérez-Tornero, O., Landin, M., Burgos, L., and Gallego, P. P. (2011). Improving knowledge of plant tissue culture and media formulation by neurofuzzy logic: a practical case of data mining using apricot databases. *J. Plant Physiol.* 168, 1858–1865. doi: 10.1016/j.jplph.2011.04.008

Gallego, P. P., Gago, J., and Landin, M. (2011). "Artificial neural networks technology to model and predict plant biology process," in *Artificial Neural Networks – Methodological Advances and Biomedical Applications*, ed K. Suzuki (Rijeka: InTech Open), 197–216.

Garcês, H. M. P., Koenig, D., Townsley, B. T., Kim, M., and Sinha, N. R. (2014). Truncation of LEAFY COTYLEDON1 protein is required for asexual reproduction in *Kalanchoë daigremontinana*. *Plant Physiol.* 165, 196–206. doi: 10.1104/pp.114.237222

Garcês, H. M. P., and Sinha, N. (2009). The 'Mother of Thousands' *(Kalanchoë daigremontiana):* a plant model for asexual reproduction and CAM studies. *Cold Spring Harbor Protoc.* 4, 1–9. doi: 10.1101/pdb.emo133

García-Pérez, P., Barreal, M. E., Rojo-de Dios, L., Cameselle-Teijeiro, J. F., and Gallego, P. P. (2018a). "Bioactive natural products from the genus *Kalanchoe* as cancer chemopreventive agents: a review," in *Studies in Natural Products Chemistry*, ed A. ur-Rahman (Amsterdam: Elsevier), 49–84.

García-Pérez, P., Losada-Barreiro, S., Gallego, P. P., and Bravo-Díaz, C. (2019). Adsorption of gallic acid, propyl gallate and polyphenols from *Bryophyllum* extracts on activated carbon. *Sci. Rep.* 9:14830. doi: 10.1038/s41598-019-51322-6

García-Pérez, P., Lozano-Milo, E., Gallego, P. P., Tojo, C., Losada-Barreiro, S., and Bravo-Díaz, C. (2018b). "Plant antioxidants in food emulsions," in *Some New Aspects of Colloidal Systems in Foods*, ed J. Milani (Rijeka: InTech Open), 11–29.

García-Pérez, P., Lozano-Milo, E., Landin, M., and Gallego, P. P. (2020a). Machine learning technology reveals the concealed interactions of phytohormones on medicinal plant *in vitro* organogenesis. *Biomolecules* 10:746. doi: 10.3390/biom10050746

García-Pérez, P., Lozano-Milo, E., Landin, M., and Gallego, P. P. (2020b). Combining medicinal plant *in vitro* culture with machine learning technologies for maximizing the production of phenolic compounds. *Antioxidants* 9:210. doi: 10.3390/antiox9030210

George, E. F., Hall, M. A., and De Klerk, G.-J. (2008). "The components of plant tissue culture media I: macro- and micro-nutrients," in *Plant Propagation by Tissue Culture*, eds E. F. George, M. A. Hall, and G.-J. De Klerk (Dordrecht: Springer), 65–114.

Golkar, P., Hadian, F., and Dehkordi, M. K. (2019). Production of a new mucilage compound in *Lepidium sativum* callus by optimizing *in vitro* growth conditions. *Nat. Prod. Res.* 33, 130–135. doi: 10.1080/14786419.2018.1437426

Guerra-García, A., Barrales-Alcal,á, D., Argueta-Guzmán, M., Cruz, A., Mandujano, M. C., Arévalo-Ramírez, J. A., et al. (2018). Biomass allocation, plantlet survival, and chemical control of the invasive chandelier plant (*Kalanchoe delagoensis*) (Crassulaceae). *Invas. Plant Sci. Manag.* 11, 33–39. doi: 10.1017/inp.2018.6

Guerra-García, A., Golubov, J., and Mandujano, M. C. (2015). Invasion of *Kalanchoe* by clonal spread. *Biol. Invas.* 17, 1615–1622. doi: 10.1007/s10530-014-0820-0

Hajiboland, R. (2012). "Effect of micronutrient deficiencies on plants stress responses," in *Abiotic Stress Responses in Plants: Metabolism, Productivity and Sustainability*, eds P. Ahmad and M. N. V. Prasad (New York, NY: Springer), 283–329.

Herrando-Moraira, S., Vitales, D., Nualart, N., Gómez-Bellver, C., Ibáñez, N., Mass,ó, S., et al. (2020). Global distribution patterns and niche modelling of the invasive *Kalanchoe* × *houghtonii* (Crassulaceae). *Sci. Rep.* 10:3143. doi: 10.1038/s41598-020-60079-2

Hesami, M., and Jones, A. M. P. (2020). Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. *Appl. Microbiol. Biotechnol.* 104, 9449–9485. doi: 10.1007/s00253-020-10888-2

Hoang, N. N., Kitaya, Y., Shibuya, T., and Endo, R. (2019). Effects of suppoting materials in *in vitro* acclimatization stage on *ex vitro* growth of wasabi plants. *Sci. Horticult.* 261:109042. doi: 10.1016/j.scienta.2019.109042

Isah, T., Umar, S., Mujib, A., Sharma, M. P., Rajasekharan, P. E., Zafar, N., et al. (2018). Secondary metabolism of pharmaceuticals in the plant *in vitro* cultures: strategies, approaches, and limitations to achieving higher yield. *Plant Cell Tissue Organ Cult.* 132, 239–265. doi: 10.1007/s11240-017-1332-2

Konar, S., Adhikari, S., Karmakar, J., Ray, A., and Bandyopadhyay, T. K. (2019). Evaluation of subculture ages on organogenic response from root callus and SPAR based genetic fidelity assessment in the regenerants of *Hibiscus sabdariffa* L. *Industr. Crops Prod.* 135, 321–329. doi: 10.1016/j.indcrop.2019.04.018

Landin, M., and Rowe, R. C. (2013). "Artificial neural networks technology to model, understand, and optimize drug formulations," in *Formulation Tools for Pharmaceutical Development* (Cambridge: Woodhead Publishing), 7–37.

Landin, M., Rowe, R. C., and York, P. (2009). Advantages of neurofuzzy logic against conventional experimental design and statistical analysis in studying and developing direct compression formulations. *Eur. J. Pharm. Sci.* 38, 325–331. doi: 10.1016/j.ejps.2009.08.004

Liu, Y., and von Wirén, N. (2017). Ammonium as a signal for physiological and morphological responses in plants. *J. Exp. Bot.* 68, 2581–2592. doi: 10.1093/jxb/erx086

Lüttge, U., Pfeifer, T., Fischer-Schliebs, E., and Ratajczak, R. (2000). The role of vacuolar malate-transport capacity in crassulacean acid metabolism and nitrate nutrition. Higher malate-transport capacity in ice plant after crassulacean acid metabolism-induction and in tobacco under nitrate nutrition. *Plant Physiol.* 124, 1335–1348. doi: 10.1104/pp.124.3.1335

Marchev, A. S., Zhenya, P. Y., and Georgiev, M. I. (2020). Green (cell) factories for advanced production of plant secondary metabolites. *Crit. Rev. Biotechnol.* 40, 443–458. doi: 10.1080/07388551.2020.1731414

Mendel, R. R., and Hänsch, R. (2002). Molybdoenzymes and molybdenum. *J. Exp. Bot.* 53, 1689–1698. doi: 10.1093/jxb/erf038

Mendonça, E. G., Batista, T. R., Stein, V. C., Balieiro, F. P., de Abreu, J. R., Pires, M. F., et al. (2019). *In vitro* serial subculture to improve rooting of *Eucalyptus urophylla*. *New Forests* 51, 801–816. doi: 10.1007/s11056-019-09761-6

Migocka, M., and Malas, K. (2018). "Plant responses to copper: molecular and regulatory mechanisms of copper uptake, distribution and accumulation in plants," in *Plant Micronutrient Use Efficiency*, eds M. A. Hossain, T. Kamiya, D. J., Burritt, L.-S. P. Tran, and T. Fujiwara (London: Academic Press, Elsevier), 71–86.

Mize, C. W., Koehler, K. J., and Compton, M. E. (1999). Statistical considerations for *in vitro* research: II – Data to presentation. *In Vitro Cell. Dev. Biol. Plant* 35, 122–126. doi: 10.1007/s11627-999-0021-1

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Nakasha, J. J., Sinniah, U. R., Kemat, N., and Mallappa, K. S. (2016). Induction, subculture cycle, and regeneration of callus in safed musli (*Chlorophytum borivilianum*) using different types of phytohormones. *Pharmacogn. Mag.* 12, S460–S464. doi: 10.4103/0973-1296.191457

Naqib, S. A., and Jahan, M. S. (2017). The function of molybdenum and boron on the plants. *J. Agric. Res.* 2, 000136.

Nezami-Alanagh, E., Garoosi, G.-A., Landin, M., and Gallego, P. P. (2019). Computer-based tools provide new insight into the key factors that cause physiological disorders of pistachio rootstocks cultured *in vitro*. *Sci. Rep.* 9:9740. doi: 10.1038/s41598-019-46155-2

Nezami-Alanagh, E., Garoosi, G. A., Haddad, R., Maleki, S., Landin, M., and Gallego, P. P. (2014). Design of tissue culture media for efficient *Prunus* rootstock micropropagation using artificial intelligence models. *Plant Cell Tissue Organ Cult.* 117, 349–359. doi: 10.1007/s11240-014-0444-1

Nezami-Alanagh, E., Garoosi, G. A., Landin, M., and Gallego, P. P. (2018). Combining DOE with neurofuzzy logic for healthy mineral nutrition of pistachio rootstocks *in vitro* culture. *Front. Plant Sci.* 9:1474. doi: 10.3389/fpls.2018.01474

Nezami-Alanagh, E., Garoosi, G. A., Maleki, S., Landin, M., and Gallego, P. P. (2017). Predicting optimal *in vitro* culture medium for *Pistacia vera* micropropagation using neural networks models. *Plant Cell Tissue Organ Cult.* 129, 19–33. doi: 10.1007/s11240-016-1152-9

Niazian, M., and Niedbala, G. (2020). Machine learning for plant breeding and biotechnology. *Agriculture* 10:436. doi: 10.3390/agriculture10100436

Niedz, R. P., and Evens, T. J. (2006). A solution to the problem of ion confounding in experimental biology. *Nat. Methods* 3:417. doi: 10.1038/nmeth0606-417

Niedz, R. P., and Evens, T. J. (2007). Regulating plant tissue growth by mineral nutrition. *In Vitro Cell. Dev. Biol. Plant* 43, 370–381. doi: 10.1007/s11627-007-9062-5

Pastor, A. V., Palazzo, A., Havlik, P., Blemans, H., Wada, Y., Obersteiner, M., et al. (2019). The global nexus of food-trade-water sustaining environmental flows by 2050. *Nat. Sustain.* 2, 499–507. doi: 10.1038/s41893-019-0287-1

Patra, J. K., Das, G., Das, S. K., and Thatoi, H. (2020). "Plant tissue culture techniques and nutrient analysis," in *A Practical Guide to Environmental Biotechnology*, eds J. K. Patra, G. Das, S. K. Das, and H. Thatoi (Singapore: Springer Nature Singapore Pte Ltd.), 135–164.

Pereira, P. N., and Cushman, J. C. (2019). Exploring the relationship between crassulacean acid metabolism (cam) and mineral nutrition with a special focus on nitrogen. *Int. J. Mol. Sci.* 20:4363. doi: 10.3390/ijms20184363

Pereira, P. N., Smith, J. A. C., and Mercier, H. (2017). Nitrate enhancement of CAM activity in two *Kalanchoë* species is associated with increased vacuolar proton transport capacity. *Physiol. Plant.* 160, 361–372. doi: 10.1111/ppl.12578

Pérez-Pérez, M. E., Lemaire, S. D., and Crespo, J. L. (2012). Reactive oxygen species and autophagy in plants and algae. *Plant Physiol.* 160, 156–164. doi: 10.1104/pp.112.199992

Phillips, G. C., and Garda, M. (2019). Plant tissue culture media and practices: an overview. *In Vitro Cell. Dev. Biol. Plant* 55, 242–257. doi: 10.1007/s11627-019-09983-5

Printz, B., Lutts, S., Hausman, J.-F., and Sergeant, K. (2016). Copper trafficking in plants and its implication on cell wall dynamics. *Front. Plant Sci.* 7:601. doi: 10.3389/fpls.2016.00601

Rodrigues, M. A., Freschi, L., Pereira, P. N., and Mercier, H. (2014). "Interaction between nutrients and crassulacean acid metabolism," in *Progress in Botany*, eds U. Lüttge, W. Beyschlag, and J. Cushman (London: Springer-Verlag Berlin Heidelberg), 167–186.

Saad, A. I. M., and Elshahed, A. M. (2012). "Plant tissue culture media," in *Recent Advances in Plant in vitro Culture*, ed A. Leva (Rijeka: InTech Open), 29–40.

Schulten, A., and Kramer, U. (2017). "Interaction between copper homeostasis and metabolism in plants," in *Progress in Botany*, eds F. Cánovas, U. Lüttge, and R. Matyssek (Cham: Springer International Publishing AG), 111–146.

Shao, Q., Rowe, R. C., and York, P. (2006). Comparison of neurofuzzy logic and neural networks in modelling experimental data of an immediate release tablet formulation. *Eur. J. Pharm. Sci.* 28, 394–404. doi: 10.1016/j.ejps.2006.04.007

Shihabudheen, K. V., and Pillai, G. N. (2017). Regularized extreme learning adaptive neurofuzzy algorithm for regression and classification. *Knowled. Based Syst.* 127, 100–113. doi: 10.1016/j.knosys.2017.04.007

Shinmachi, F., Buchner, P., Stroud, J. L., Parmar, S., Zhao, F.-J., McGrath, S. P., et al. (2010). Influence of sulfur deficiency on the expression of specific sulfate transporters and the distribution of sulfur, selenium and molybdenum in wheat. *Plant Physiol.* 153, 327–336. doi: 10.1104/pp.110.153759

Shrivastav, P., Prasad, M., Singh, T. B., Yadav, A., Goyal, D., Ali, A., et al. (2020). "Role of nutrients in plant growth and development," in *Contaminants in Agriculture*, eds M. Naeem, A. Ansari, and S. Gill (Cham: Springer International Publishing AG), 43–59.

Smulders, M. J. M., and de Klerk, G. J. (2011). Epigenetics in plant tissue culture. *Plant Growth Regul.* 63, 137–146. doi: 10.1007/s10725-010-9531-4

Stefanowicz-Hajduk, J., Asztemborska, M., Krauze-Baranowska, M., Godlewska, S., Gucwa, M., Moniuszko-Szajwaj, B., et al. (2020). Identification of flavonoids and bufadienolides and cytotoxic effects of *Kalanchoe daigremontiana* extracts on human cancer cell lines. *Planta Med.* 86, 239–246. doi: 10.1055/a-1099-9786

Teixeira da Silva, J. A., Nezami-Alanagh, E., Barreal, M. E., Kher, M. M., Wicaksono, A., Gulyás, A., et al. (2020). Shoot tip necrosis of in vitro plant cultures: a reappraisal of possible causes and solutions. *Planta* 252, 1–35. doi: 10.1007/s00425-020-03449-4

Twaij, B. M., Jazar, Z. H., and Hasan, M. N. (2020). Trends in the use of tissue culture, applications and future aspects. *Int. J. Plant Biol.* 11. doi: 10.4081/pb.2020.8385

Vapnik, V. (1992). "Principles of risk minimization for learning theory," in *Proceedings of the Advances in Neural Information Processing Systems* (Denver, CO), 831–838.

Villarrubia, G., de Paz, J. F., Chamoso, P., and de la Pietra, F. (2018). Artificial neural networks used in optimization problems. *Neurocomputing* 272, 10–16. doi: 10.1016/j.neucom.2017.04.075

Wang, Y., and Yao, R. (2020). Increased endogenous gibberellin levels inhibits root growth of *Pinus massoniana* Lamb. plantlets during long-term subculture. *In Vitro Cell. Dev. Biol. Plant* 56, 470–479. doi: 10.1007/s11627-020-10067-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 García-Pérez, Lozano-Milo, Landin and Gallego. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 14 December 2020
doi: 10.3389/fpls.2020.590529

Machine Learning Techniques for Soybean Charcoal Rot Disease Prediction

*Elham Khalili[1], Samaneh Kouchaki[2], Shahin Ramazi[3] and Faezeh Ghanati[1]**

[1] Department of Plant Science, Faculty of Science, Tarbiat Modarres University, Tehran, Iran, [2] Faculty of Engineering and Physical Sciences, Centre for Vision, Speech, and Signal Processing, University of Surrey, Guildford, United Kingdom, [3] Department of Biophysics, Faculty of Biological Science, Tarbiat Modares University, Tehran, Iran

Early prediction of pathogen infestation is a key factor to reduce the disease spread in plants. *Macrophomina phaseolina* (Tassi) Goid, as one of the main causes of charcoal rot disease, suppresses the plant productivity significantly. Charcoal rot disease is one of the most severe threats to soybean productivity. Prediction of this disease in soybeans is very tedious and non-practical using traditional approaches. Machine learning (ML) techniques have recently gained substantial traction across numerous domains. ML methods can be applied to detect plant diseases, prior to the full appearance of symptoms. In this paper, several ML techniques were developed and examined for prediction of charcoal rot disease in soybean for a cohort of 2,000 healthy and infected plants. A hybrid set of physiological and morphological features were suggested as inputs to the ML models. All developed ML models were performed better than 90% in terms of accuracy. Gradient Tree Boosting (GBT) was the best performing classifier which obtained 96.25% and 97.33% in terms of sensitivity and specificity. Our findings supported the applicability of ML especially GBT for charcoal rot disease prediction in a real environment. Moreover, our analysis demonstrated the importance of including physiological featured in the learning. The collected dataset and source code can be found in https://github.com/Elham-khalili/Soybean-Charcoal-Rot-Disease-Prediction-Dataset-code.

Keywords: charcoal rot, gradient tree boosting algorithm, *Macrophomina phaseolina* (Tassi) Goid, machine learning, prediction

OPEN ACCESS

Edited by:
Spyros Fountas, Agricultural University of Athens, Greece

Reviewed by:
Qammer H. Abbasi, University of Glasgow, United Kingdom
Mahmood Safaei, National University of Malaysia, Malaysia
Xiaorong Ding, University of Oxford, United Kingdom

****Correspondence:***
Faezeh Ghanati
ghangia@modares.ac.ir

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *01 August 2020*
Accepted: *23 November 2020*
Published: *14 December 2020*

Citation:
Khalili E, Kouchaki S, Ramazi S and Ghanati F (2020) Machine Learning Techniques for Soybean Charcoal Rot Disease Prediction. Front. Plant Sci. 11:590529. doi: 10.3389/fpls.2020.590529

INTRODUCTION

The production of global crops has to be doubled by 2050 to meet the increasing needs of the world's population (Khalili et al., 2019). Plant diseases are the lead causes of extensive economic losses in the agricultural industry around the world. Recent statistics have confirmed that there is a decline of worldwide crop yields by 14% worldwide due to plant diseases, weeds and insects, and hence,

Abbreviations: AUC, Area under the ROC curve; ELISA, Enzyme-linked immunosorbent assay; FCM, Flow cytometry; FISH, Fluorescence *in situ* hybridization; FN, False negatives; FP, False positives; GBT, Gradient tree boosting; IF, Immunofluorescence; LR, Regularized logistic regression; MCC, Matthews correlation coefficient; ML, Machine learning; MLP, Multilayer perceptron; PPV, Precision or positive prediction value; R7, Yellowing of the leaves and yellow pods at 50% growing stage; RF, Random forest; ROC, Receiver operating characteristic; RT-PCR, Polymerase chain reaction; SVM, Support vector machines; TN, True negatives; TP, True positive; t-SNE, t-distributed stochastic neighbor embedding; WB, Western blotting.

early detection of diseases is of a key importance to prevent disease spread and reduce damage to crop production (Martinelli et al., 2015). *Macrophomina phaseolina* (Tassi) Goid causes rot diseases in about 700 plant species. It is an extremely robust soil-borne fungus that damages several crops i.e., cotton, grains, oilseeds, legumes, jute along with fruits and vegetable plants (Ambrosio et al., 2015; Sun et al., 2016). A wide range of physiological, morphological, and pathogenic diversity enables *M. phaseolina* to adapt across various climatic conditions (Ambrosio et al., 2015). Moreover, sclerotia and chlamydospores structures enable the fungus to survive in the soil for a longer period (Katan, 2017). Gaige et al. (2010) described that the disease is dispersed by infected plant residues, wind, and soil. The infestation of *M. phaseolina* pathogen may occur at any growth stage whereas symptoms often appear after the midseason or at maturity i.e., growth stage R7 where yellowing of the leaves and yellow pods can be observed (Hartman et al., 2016). Other symptoms may include the development of "blackleg" in infected plants which results in weaker plants and lower productivity (Santos et al., 2016). The infected plants ultimately die due to various reasons such as vascular blockages that weaken the nutrient transport (Santos et al., 2016) or exposure to phytotoxic metabolites released by *M. phaseolina*.

For decades, agricultural management strategies for controlling plant diseases were mainly based on cultural practices e.g., soil solarization, crop rotation, cultivation of tolerant cultivars, alone or combined with other techniques such as low doses of pesticides and biological agents (Holmes et al., 2020). Generally, fumigants and fungicides are used to control *M. phaseolina* infections in crops that can be ineffective and inefficient due to different environmental factors as reported in Abbas et al. (2019). In a work presented by Khalili et al. (2019), a higher dose of fungicides was suggested for an economical yield. An increased dose of these chemicals leads to concerns over the long-lasting harmful impacts of pesticides on human health and ecology as agricultural run-offs contain pesticides which pollute the water resources (Chamorro et al., 2015; Pastrana et al., 2016). Moreover, the bioaccumulation of these toxic compounds in the food chain and further ingestion by bird populations and mammals pose vital health-associated threats (Brevik et al., 2020).

The efficient detection of diseases can be a key factor in the sustainability of the agroecosystem. The developments in molecular biology and biotechnology have improved the detection of plant diseases. Reverse Transcription Polymerase Chain Reaction (RT-PCR), Enzyme-Linked Immuno-Sorbent Assay (ELISA), and Western blotting (WB) are examples of plant disease diagnostic techniques (Jeong et al., 2014; Golhani et al., 2018). However, these techniques are not able to predict the fungal disease despite their diagnostic efficiency (Sakudo et al., 2006; Thanarajoo et al., 2014). Moreover, RT-PCR, ELISA, and WB are limited in terms of cost-effectiveness, efficiency, and accuracy for the prediction of disease infestation (Eun et al., 2002).

Therefore, an automated diagnostic system is important to prevent and control diseases in soybean. It would minimize the yield and economic losses, reduce pesticide residues, and enhance product quality (Nagasubramanian et al., 2018). Effective soybean disease classification is critical to predict the disease at the early stages. Machine learning (ML) techniques have found application in several areas of research such as crop management, yield prediction (Chlingaryan et al., 2018), disease detection (Kouchaki et al., 2019), and weed detection crop quality (Liakos et al., 2018; Wang et al., 2019). These algorithms learn through examples (training data), to predict the unseen data (Ashfaq et al., 2017). Researchers have also applied learning algorithms in predicting the pest attack and disease infestation in crops (Patricio and Rieder, 2018).

In this work, a number of ML algorithms, including linear regression with L1 and L2 regularization terms (LR-L1 and LR-L2), neural network (Multilayer perceptron, MLP), random forest (RF), gradient tree boosting (GBT), and support vector machines (SVM) were developed and compared for soybean disease prediction. These algorithms have been used to classify healthy and infected plants using spectral imaging data of aerial parts of plants (Ur Rahman et al., 2017). ML methods have also been proven successful in monitoring morphological traits (Singh et al., 2017; Mochida et al., 2019). Nonetheless, variations in symptoms may lead to an improper prediction due to dynamic nature of plant changes. Consequently, the appearance-based identification of diseases is not reliable enough to accurately detect unhealthy plants especially in the early growth stages. An appropriate method is vital for detection of the causal agent as charcoal rot does not have any visible symptoms until the midseason (Sladojevic et al., 2016). Hence, we have proposed a hybrid feature set for the prediction of charcoal rot disease using physiological features and morphological characteristics (including growth attributes as well as yield-related features). As a result, ML algorithms are trained and assessed based on the hybrid feature sets of healthy and infected soybean plants. The available dataset contains both experimental setups and real cultivation conditions in the field. The work shows the application of ML techniques to detect unhealthy plants from the healthy group.

Currently, no public dataset for soybean charcoal rot disease classification is available. The applicability and success of supervised ML algorithms on predictive disease modeling have been reported but for other diseases and mainly based on image datasets. Therefore, our main focus is to suggest a set of informative features to enhance charcoal rot disease prediction as well as providing a comprehensive comparison of several ML techniques.

MATERIALS AND METHODS

Dataset Collection

Soybean (*Glycine max* L.) plants were collected from 10 different areas of Mazandaran province which is the most prolific geographical region for the production of soybean in Iran (**Supplementary Figure 1**). Soybean healthy plants were collected based on the symptomless features of leaf, stem, and root of mature during the ripeness stage. In this study, the R7 was chosen for infected plants based on the physical properties e.g.,

the existence of bright gray and sclerotia on the stem and root and suspicious of diseases. All samples were transferred to the laboratory of the Agricultural and Resource Research Center of Mazandaran (Iran) and stored at 4°C until further analysis. Overall, 2,500 plants were randomly chosen from healthy and infected plants.

Symptoms of Infected Soybean Samples

The infection of this pathogen is observed on all parts of the plant i.e., branches, leaves, pods, petioles, root, stem, and seeds on soybean (Gupta et al., 2012), however, the key indications of disease are observed after the flowering stage in infected plants, i.e., R7 stage, especially in low humidity level and high-temperature conditions (Schoving et al., 2020). Chlorosis of leaves, premature defoliation, and reduced vigor are the major symptoms observed in the infected plants (Romero Luna et al., 2017), which result in reduced productivity, sterility of pod, and formation of crinkled and tiny seeds. A brown discoloration in the vascular tissues of the taproot advanced into the stem is seen in infected plants. An appearance of powdery black sclerotia is found under the epidermis and root at the seed formation stage in the infected plants. Sometimes, the plant symptoms of this disease are confused with other plant abiotic stresses like drought or abiotic stress like cyst nematode, therefore the detection of this disease based on morphological aerial plant parts is challenging (Sanchez et al., 2019).

Laboratory Assessment

Determination of Morphological Parameters

All soybean healthy and infected samples were collected and transferred to the laboratory and then cleaned with tap water until all noticeable soil and sand spots were removed. Forceps were used to remove the remaining particles manually. The mature seedlings were observed on the 54th day after sowing while mature pods were observed on the 80th day after sowing. Specifically, the length and thickness of the stem and root as well as length, width, and thickness of the seeds were examined (Fenta et al., 2014). Length of the mature seedlings, stem and root of a soybean plant was measured and reported in cm. A pair of calipers were used to measure the length of root, pods and seeds and thickness of the seeds. Meanwhile, the thickness of the seeds was measured using a micrometer screw gauge. At each harvest, the number of seeds and pods per each plant were manually categorized based on the date of pod or flowering set to count the numbers of empty and filled pods (with or without rudimentary seeds) (Joshi et al., 2015).

Determination of Fresh Weight and Dry Weight

An electronic top pan balance was employed to calculate the fresh weights (FW) of soybean seedlings, stems and roots (Model BL-210-S, Sartorius, Germany). On the other hand, Samples were oven-dried at 70 ± 2°C for 72 h for weighting the dry weight (DW). DW and FW were stated in grams per plant (Schnyder and Baum, 1992).

Yield-Related Parameters Assessment

Seed Quality Index

Germination percentage (GP) and Seedling vigor index (VI) of the soybean seeds were measured after 2 weeks. Daily observations of seed emergence were carried out. Seed germination percentage is calculated as follows (Islam et al., 2009):

$$GP = \frac{\sum(N - i) \times Gi}{N \times GN} \times 100 \quad (1)$$

where i and N are the number of days since the day of sowing and the total number of days, respectively. Gi and GN are the number of seeds germinated on day i, and the total number of germinated seeds, respectively.

Furthermore, seedling vigor index (SI) is calculated by Islam et al. (2009):

$$VI = GP \times SL \times 100 \quad (2)$$

where, GP is germination percentage, SL is the seedling length in cm.

Thousand Seed Weight

It is highly useful for calculating the optimal seeding rate for a given crop type. A large variation was observed across measured seed weights. We grouped the weights in two groups of light (<100 g) and medium or intermediate (>100–200 g). Individual seeds were weighted by calculating the weight of 1,000 fresh seeds (empty seeds were discarded). The frequency distribution of the seed weight was determined by checking its normality using the K-S test (Brzezinski et al., 2015).

Determination of Physiological Parameters

After the harvest, the soybean seeds were taken to the laboratory for physiological quality assessments, through the following tests:

Protein Content and Seed Oil Content

The oil content was calculated and expressed in percentage of the dry matter using petroleum ether and in a Soxhlet instrument (technique 920.85, AOAC, 1990). The protein content was obtained and expressed in percentage of the dry matter by indicating the total nitrogen based on the micro Kjeldahl method (technique 920.87, AOAC, 1990) by considering a 6.25 conversion factor (Vaknin et al., 2011).

Amount of Chlorophyll and Carotenoid

Samples with 0.1 g of leaves (fresh material) was chosen randomly. Each soybean sample was grounded in 0.5% (w v^{-1}) magnesium carbonate and 10 mL of 80% acetone. Then, 10 ml of 100% acetone was added. A spectrophotometer was used to measure the absorbance (Jenway 6105 UV/VIS) in 663 nm (chlorophyll *a*—Chl *a*), 645 nm (chlorophyll *b*—Chl *b*) and 480 nm (carotenoids—C_{x+c}) wavelengths. Equations described by Hendry and Price (1993) was employed to calculate the chlorophyll concentrations. The fraction of photosynthetically active irradiance absorbed by the leaf (α) depends on the chlorophyll content (μmol m^{-2}) and it was calculated as $\alpha = Chl_{tot}/(Chl_{tot} + 76)$ by considering the

TABLE 1 | List of features for prediction of charcoal rot disease in soybean.

Features		
Morphological		
Growth attributes	**Yield- related**	**Physiological**
Stem length	Germination percentage	Seed oil content
Root length	Seedling vigor index	Amount of chlorophyll
Thickness of seed	Thousand seed weight	Amount of carotenoid
Stem bark thickness	Number of pods per plant	Protein content
Root bark thickness	Number of seeds per plant	
Stem fresh weight	Empty pods per plant	
Stem dry weight		
Root fresh weight		
Root dry weight		
Seedling fresh weight		
Seedling dry weight		

work of Evans and Poorter (2001). The data obtained was subject to a regression analysis using the SigmaPlot 8.02 package for Windows.

Morphological and Physiological Feature Extraction

An observation was conducted to check the morphological and physiological characteristics of each soybean plants. In order to carry out the ML experiments, two categories were considered; healthy and infected. Healthy (negative) and infected (positive) plants were separated based on symptoms of charcoal rot. Appropriate attributes were selected based on the differences between the healthy and infected plants. Some data within each category of healthy or infected samples had very similar feature values. Therefore, as a preprocessing step, we dropped all but one of very similar data samples as they would not add any extra information to the learning or validation of the proposed pipeline. Finally, 1,000 healthy soybean plants (negative) and 1,000 infected plants (positive) were selected for charcoal rot disease prediction (**Supplementary Table 1**).

Feature Selection

Feature selection was designed and optimized to enhance the performance and generalizability of ML models. In order to select the relevant features, analysis of variance and *F*-test were used (Elssied et al., 2014). These analyses were based on *p*-value for feature selection by skipping the irrelevant attributes from the data set (Eskandari and Javidi, 2016). *F*-test was performed to compute the statistical significance value and to calculate the *p*-value for the difference in means at the 5% level of significance. We finally ended up with a list of 21 features to be analyzed by ML techniques. Results of the *F*-test confirmed that morphological and physiological characters parameters were among the most important features for prediction of charcoal rot disease in soybean (**Table 1**).

Computational Methods for Predicting Infected Soybeans

Our pipeline for predicting infected soybean has four main steps: (1) data gathering; (2) feature extraction; (3) training the predictors; and (4) performance assessment. These steps have been described and have been schematically shown in **Figure 1**. Data gathering is the first step of the healthy soybean prediction (**Figure 1A**). After creating the positive and negative datasets, incomplete instances were removed. In order to have a balanced positive and negative dataset, a random subset of the negative dataset with an equal number of positive samples was selected. In the feature extraction step, the positive and negative samples (soybeans) are coded into numerical feature vectors to be used to learn the classifiers.

There is a variety of classifiers that can be learned and based on the performance of different classifiers, a suitable classifier can be selected (**Figure 1B**). A standard procedure for assessing the performance of a classifier is *k*-fold cross-validation. In this process, the available dataset is randomly divided into *k* subsets without an overlap. Then, $k-1$ of them is utilized as a training dataset, and the remaining as the test set for evaluating the model (Lyons et al., 2018). This process is repeated *k* times to allow every subset to be used precisely once as the test set. Finally, the average performance for all *k* test sets is calculated (**Figure 1C**). The most important performance assessment measures are used in the prediction of the healthy soybeans are described in the following subsections. All of these measures are based on the four basic elements of the confusion matrix (true positive, false positive true negative, and false negative represented as TP, FP, TN, and FN, respectively).

Machine Learning Methods

After the data collection and feature extraction steps, six ML techniques (LR-L1, LR-L2, MLP, RF, GBT, and SVM) were developed and applied to the training set. We used 10-fold cross-validation while the threshold was set based on the training data considering false positive and false negative rates. All the ML techniques were run by the open-source ML toolkit scikit-learn (version 0.20.1) in python 3.6.7. The parameters of the models (e.g., number of ensembles for RF or GBT) were optimized through an internal cross-validation on the training data. This was done by a grid search over a range of values and selecting parameters that generated the best area under the area under the receiver-operator characteristic (ROC) curve (AUC). The model with the highest performance was reported in the paper.

Regularized Logistic Regression (LR-L1 and LR-L2)

LR is a linear classification model that predicts binary outcomes based on a set of explanatory variables (i.e., features). This model is performed using LIBLINEAR library and L1 or L2 regularizations (LR-L1 and LR-L2). L1 regularization and L2 regularization are two common techniques to reduce the model over-fitting (Couronne et al., 2018).

Multilayer Perceptron (MLP)

MLP maps the input data to a non-linear latent representation. MLP contains several fully connected layers of nodes in which a

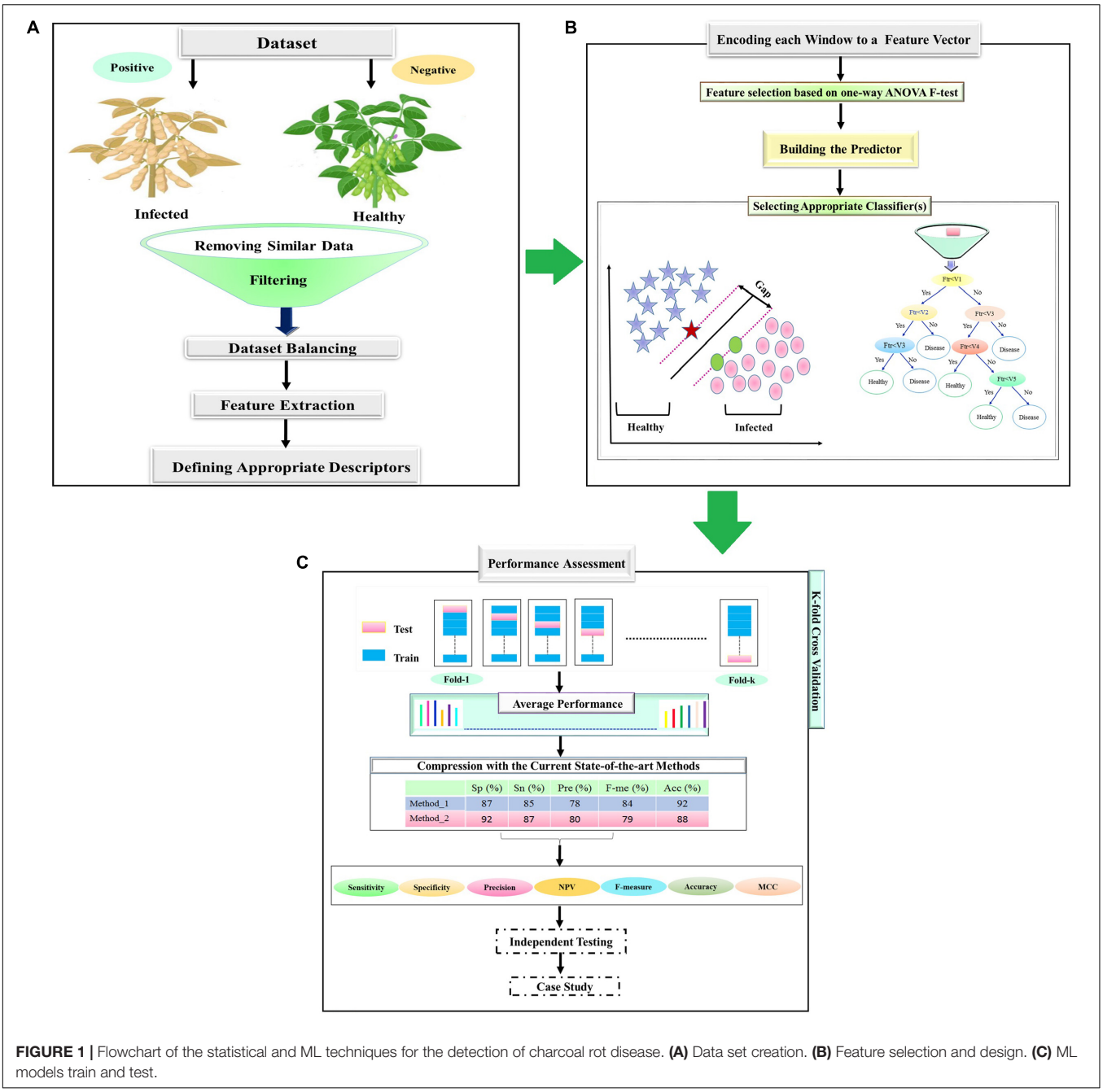

FIGURE 1 | Flowchart of the statistical and ML techniques for the detection of charcoal rot disease. **(A)** Data set creation. **(B)** Feature selection and design. **(C)** ML models train and test.

non-linear activation function is considered for each node, except at the input layer. MLP employs back-propagation for training (Breiman, 2001) and has shown to be a highly applicable network, thus a popular choice among researchers (Shan et al., 2018). Two hidden layers of size 10 and 4 and Adam optimization were considered in this work.

Random Forest (RF)

RF is a non-linear ensemble method that consists of multiple decision trees. The final prediction is determined from the results of the individual trees (Basu et al., 2018), which improves the generalization ability of the model for a better prediction. The accuracy of an individual tree and a correlation between these trees are key points in the generalization ability of RF. RF is not usually sensitive in the choice of parameter selections (Teixeira et al., 2013).

Gradient Tree Boosting (GBT)

GBT (Friedman, 2002) is another ensemble algorithm based on decision trees that can be considered for both classification and regression problems (Cheng et al., 2018). In contrast to RF, this model sequentially builds decision trees by a weighting strategy

to put more emphasis on harder samples. A weighted majority vote is then used to make the final prediction.

Support Vector Machines (SVM)

SVM aims to find a hyperplane that minimizes the structural risk (Czarnecki and Tabor, 2015) in kernel space. Gaussian radial basis function, Linear, and polynomial are several common kernel functions. SVM has two important hyperparameters, the kernel coefficient γ and the penalty parameter C. This model follows two goals of finding a low complexity model that best separates the data to have a better generalizability ability (Uddin et al., 2019). Linear kernel was considered in this work.

Model Evaluation Criteria

The considered ML classification models are evaluated by calculating several evaluation parameters, true positive (TP) that indicates the number of correctly classified infected plants, true negatives (TN) that indicates the number of correctly classified healthy plants, false positives (FP) that denotes the number of healthy plants incorrectly classified as infected plants and false negatives (FN) that represents the number of infected plants incorrectly classified as healthy plants. The classification performance is often evaluated by accuracy, specificity, sensitivity, precision, Negative Predictive Value (NPV), F1 score and, Matthews Correlation Coefficient (MCC) value as shown in **Figure 1C**. Besides, we also assessed AUC as an indicator of model performance. The threshold for reporting the classification performance on the test sets was set on the train data. All performance criteria in this work are explained as follows:

Accuracy

Accuracy (Acc) is a ratio between the correctly classified data points to the total number of samples as described by Sokolova et al. (2006):

$$\text{Acc} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \tag{3}$$

Sensitivity and Specificity

Sensitivity describes the correctly classified positive samples to the total number of positive samples:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \tag{4}$$

whereas specificity is stated as a ratio of the correctly classified negative samples to the total number of negative samples:

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \tag{5}$$

Precision

Precision or positive prediction value (PPV) shows the correctly classified positive samples to the total number of samples predicted as positive and described by Sokolova et al. (2006) as:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{6}$$

Negative Predictive Value (NPV)

Inverse precision, or true negative accuracy measures the proportion of negative samples that were correctly classified to the total number of negative predicted samples (Sokolova et al., 2006) as:

$$\text{NPV} = \frac{TN}{FN + TN} \times 100\% \tag{7}$$

F-Measure

F-measure shows the harmonic mean of recall and precision and calculated as:

$$\text{F1score} = \frac{2TP}{2TP + FP + FN} \times 100\% \tag{8}$$

Matthews Correlation Coefficient (MCC)

MCC shows the correlation between true and predicted labels and described in Boughorbel et al. (2017) as:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + TN)(TP + FN)(TN + FP)(TN + FN)}} \times 100\% \tag{9}$$

Area Under the ROC Curve (AUC)

ROC has been used over the past years within ML community to visualize and evaluate the trade-off between the true positive rates and the false-positive rates (Fawcett, 2006). In order to compare classifiers, ROC can be reduced to the single scalar value called the area under the curve (AUC) and defined as the area under the ROC curve, a measure of the quality of the classification (Marrocco et al., 2008). AUC is not impacted by the arbitrary selection of a specific classification threshold and we thus use it as the primary evaluation metric.

t-Distributed Stochastic Neighbor Embedding (t-SNE) Data Visualization

The t-Distributed Stochastic Neighbor Embedding (t-SNE) has been successfully applied to visualization problems. Schubert and Gertz (2017), described that it attempts to preserve pairwise distance distribution of points in the lower dimensions. As the prediction in the lower dimensions includes the distribution of relative distances, it needs large data points to determine an expressive depiction. The t-SNE is a new technique in ML, which has been employed in biological data analysis (Grimes et al., 2013; Irish, 2014; Dimitriadis et al., 2018). It has also been successfully applied to visualize the infected rice leaf data in Zhang et al. (2020). In our work, t-SNE was used to visualize distinctions among positive (infected) and negative (healthy) samples.

RESULTS

Model Verification and Evaluation

We employed 10-fold cross-validation to measure and relate the strength and trustworthiness of all models as a model build by

TABLE 2 | Performance comparison of various ML techniques on the full features for prediction of soybean charcoal rot disease.

Method	TP	FP	TN	FN	Accuracy	Sensitivity	Specificity	Precision	NPV	F1 score	MCC	AUC
LR-L1	1153	47	1149	51	95.92 ± 8.32	95.75 ± 8.74	96.08 ± 7.93	96.01 ± 8.05	95.84 ± 8.60	95.88 ± 8.39	91.84 ± 16.63	97.37 ± 5.82
LR-L2	1143	57	1151	49	95.58 ± 9.00	95.92 ± 8.90	95.25 ± 9.16	95.31 ± 9.11	95.87 ± 8.90	95.61 ± 8.99	91.18 ± 17.99	97.05 ± 6.58
MLP	1139	61	1138	62	94.88 ± 9.58	94.83 ± 10.66	94.92 ± 8.51	94.72 ± 8.97	95.06 ± 10.16	94.77 ± 9.83	89.76 ± 19.14	96.69 ± 7.39
RF	1143	57	1147	53	95.42 ± 9.64	95.58 ± 9.27	95.25 ± 10.01	95.34 ± 9.80	95.50 ± 9.47	95.46 ± 9.54	90.83 ± 19.28	97.20 ± 6.35
GBT	**1168**	**32**	**1155**	**45**	**96.79 ± 6.49**	**96.25 ± 7.88**	**97.33 ± 5.16**	**97.16 ± 5.55**	**96.49 ± 7.29**	**96.68 ± 6.75**	**93.62 ± 12.90**	**98.42 ± 3.42**
SVM	1150	50	1155	45	96.04 ± 7.55	96.25 ± 7.88	95.83 ± 7.26	95.81 ± 7.36	96.29 ± 7.76	96.03 ± 7.61	92.09 ± 15.10	97.86 ± 4.65

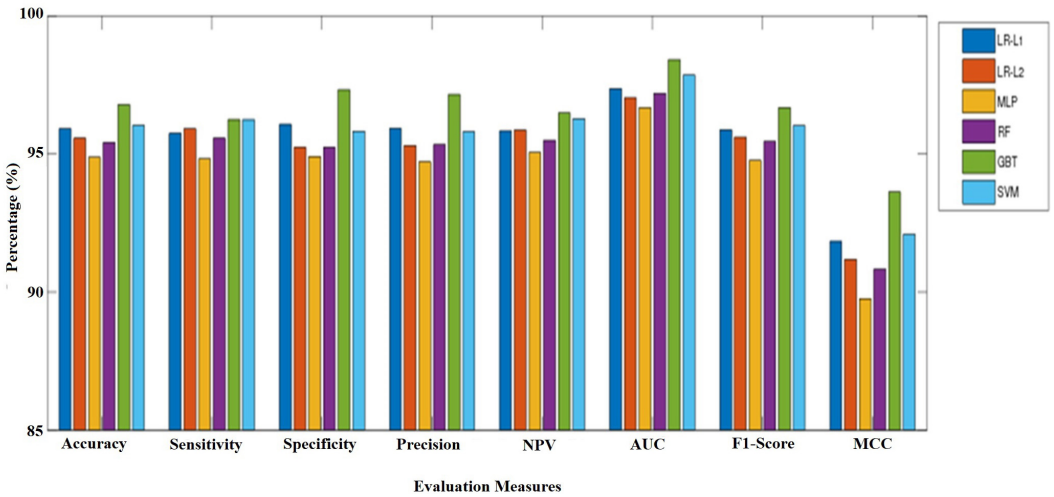

FIGURE 2 | A comparison of different evaluation criteria for the prediction of healthy and infected soybean plants with charcoal rot disease considering different ML algorithms.

only one random scale may tend to be over-fitting or occasional. The mean performance of six ML models for the test sets were shown in **Table 2** and **Figure 2**. MLP performed the worst in terms of all the evaluation criteria with the lowest accuracy (94.88%), sensitivity (94.83%), specificity (94.92%), precision (94.72%), NPV (95.06%), F1 score (94.77%), and MCC (89.76%).The final analysis shows that GBT classifier performed the best with the highest classification accuracy (96.79%), specificity (97.33%), precision (97.16%), NPV(96.49%), F1 score (96.68%), and MCC (93.62%). SVM classifier is the second best with a classification accuracy of 96.04%, with TP (1150) and specificity (95.83%), precision (95.81%), NPV (96.29%), F1 score (96.03%), and MCC (92.09%) and LR-L1, LR-L2, and RF attained an average accuracy of more than 95%. Similarly, sensitivity for GBT and SVM were almost the same and ranked the highest. LR-L1 and LR-L2 also performed quite well with only slightly lower than GBT and SVM; their sensitivity was more than 95%. It could be summarized that the GBT and SVM models outperformed the other six models for the prediction of charcoal rot disease.

Determination of the Prediction Performances

ROC curve is one of the most robust approaches for evaluating ML techniques (Bradley, 1997). Here, the ROC curve was generated by varying the output threshold of the LR-L1, LR-L2, MLP, RF, GBT, and SVM classifiers and plotting the true positive rate (sensitivity) against the false positive rate (1—specificity) for each threshold value. An accurate classifier leads to a ROC curve which is close to the left-hand and top borders of the plot and hence AUC can be used as a performance measure (Robin et al., 2011). The maximum value of AUC is 1 while weak classifiers and random guessing have AUC values close to 0.5. We plotted the ROC curves and calculated the AUC for six models based on 10-fold cross-validation for prediction of charcoal rot disease. The evaluation was performed using 2,000 data which consists of 1,000 positive and 1,000 negative samples. In **Supplementary Figure 2**, the ROC curve of the GBT model is highlighted by the red color with the highest AUC value of 98%. Results demonstrated that the average AUC values of LR-L1, LR-L2, RF, and SVM were very close (97%), which means that the four models have equal sorting or accumulation ability in prediction probability. Meanwhile, MLP model gave the lowest AUC value (96%). GBT is a robust prediction system for charcoal rot disease on soybean considering AUC as the performance measure.

GBT Model Performance

In the proposed system, we have classified healthy and infected plants of soybean dataset learning various ML classifiers on a hybrid feature set. After classification, we have calculated and

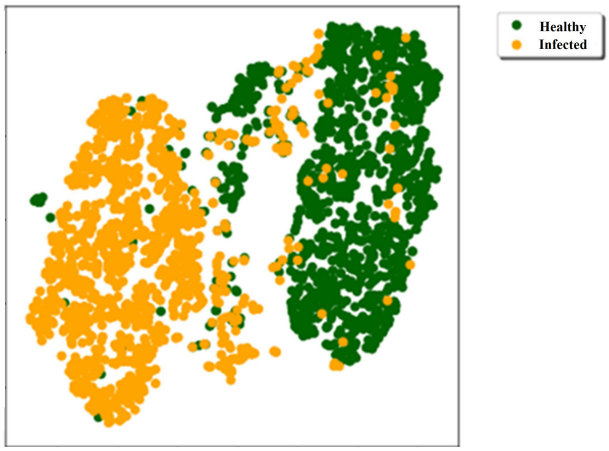

FIGURE 3 | Two-dimensional t-SNE visualization of the data.

compared their performance scores. The t-SNE was also applied to our dataset (**Figure 3**) to visualize the data in two-dimensions. As can be seen, most of the healthy and infected samples shape their clusters, although some characters of healthy and infected plants were identical which had the lowest difference in some physiological and morphological features. Consequently, having only 32 + 45 samples that were not correctly classified in our 10-fold cross-validation, demonstrates the application of ML techniques to classify most of such samples.

Effectiveness Analysis of Feature Selection

To further evaluate the effectiveness of the full features on ML performance for prediction of charcoal rot, we took 12 morphological features for classification algorithms (LR-L1, LR-L2, MLP, RF, GBT, and SVM). Then, the prediction results are evaluated on the 12 features using the 10-fold cross-validation. For accuracy, GBT reached the highest value of 96.13%, followed by SVM and LR-L1 which performed only slightly lower than GBT with an average accuracy of 95.58%. The lowest classification accuracies of 94.50%, was resulted from MLP. The averaged prediction performance is listed in **Table 3** and compared with that the full feature set. As can be seen from **Table 2**, the accuracy, sensitivity, specificity, precision, NPV, F1 score, MCC, and AUC of the full features are slightly higher to the morphological feature set. As shown in **Tables 2**, 3, the GBT algorithm has a higher performance by considering the hybrid feature set in comparison to the morphological features (96.79% vs. 96.13%).

Feature Ranking

Table 4 shows the importance of the features by considering the ML models. The features were ranked according to their importance in the classification. The incremental usefulness is important in relevance from the perspective of feature ranking where the presence of such features enhances the performance of a classification system. The top 10 features ranked by each ML algorithm in this work are represented and highlighted by different colors in **Table 4**. To further understand the importance of individual features on model predictions, SHAP analysis (SHapley Additive exPlanations) was performed on the GBT model, and the results are presented in **Figure 4**. SHAP values can be used to interpret the impact on model prediction of the value of a given feature, in comparison to a baseline value (Padarian et al., 2020). According to the results, top features were mostly among the physiological features showing their importance in comparison with the morphological features for predicting the early stage of charcoal rot disease on soybean. Observing protein content, seed oil content and amount of chlorophyll in the top 10 feature means that they are predictive features for all the ML methods. The amount of amount of carotenoid and empty pods per plant is listed in the top 10 by all of the methods except MLP and GBT. Following thousand seed weight, thickness of seed, and number of seeds per plant are selected by at least four ML methods. Root length, stem bark thickness, root bark thickness, and seedling vigor index features are examples of the least informative features. On the other hand, features that are not in this list or are just selected by one method can be categorized as the least informative features. This information is significant as the most important features can be checked first to evaluate the seeds.

DISCUSSION

Fungal diseases can be predicted through direct or indirect procedures. Direct procedures include polymerase chain reaction, immunofluorescence, fluorescence *in situ* hybridization, ELISA, flow cytometry, gas chromatography-mass spectrometry, and Western blotting. These could be used for high-throughput analysis when large numbers of samples are needed to be analyzed to get precise information (Fang and Ramasamy, 2015). Whereas, indirect methods estimate the plant diseases

TABLE 3 | Performance comparison of various ML techniques based on 12 morphological features for prediction of soybean charcoal rot disease.

Method	Accuracy	Sensitivity	Specificity	Precision	NPV	F1 score	MCC	AUC
LR-L1	95.58 ± 8.51	95.25 ± 10.27	95.92 ± 6.97	95.66 ± 7.59	95.61 ± 9.28	95.41 ± 8.97	91.21 ± 16.92	97.24 ± 5.84
LR-L2	94.96 ± 9.63	94.50 ± 12.51	95.42 ± 7.32	94.99 ± 8.32	95.15 ± 10.55	94.64 ± 10.54	90.03 ± 18.99	96.96 ± 6.40
MLP	94.50 ± 8.64	94.17 ± 12.04	94.83 ± 5.88	94.46 ± 6.57	94.90 ± 10.32	94.17 ± 9.41	89.18 ± 16.98	97.29 ± 5.45
RF	95.46 ± 9.33	95.08 ± 10.33	95.83 ± 8.41	95.64 ± 8.88	95.31 ± 9.73	95.35 ± 9.62	90.93 ± 18.63	97.12 ± 6.35
GBT	**96.13 ± 7.64**	**95.92 ± 8.34**	**96.33 ± 6.97**	**96.22 ± 7.22**	**96.05 ± 8.04**	**96.06 ± 7.78**	**92.26 ± 15.27**	**98.00 ± 4.28**
SVM	95.58 ± 7.73	95.67 ± 9.05	95.50 ± 6.53	95.34 ± 6.92	95.88 ± 8.49	95.48 ± 7.99	91.20 ± 15.41	97.46 ± 5.51

TABLE 4 | Feature ranking results for various ML techniques.

Number	Features	LR-L1	LR-L2	MLP	RF	GBT	SVM
1	Stem length	13	13	21	2	2	13
2	Root length	3	3	9	14	3	3
3	Thousand seed weight	18	18	19	16	5	19
4	Stem bark thickness	19	19	18	8	13	18
5	Root bark thickness	6	6	16	5	14	6
6	Thickness of seed	14	9	14	13	19	14
7	Stem fresh weight	17	1	6	18	21	1
8	Stem dry weight	1	14	12	11	7	16
9	Root fresh weight	16	17	13	17	17	4
10	Root dry weight	4	16	7	21	16	17
11	Seedling fresh weight	15	15	2	19	11	15
12	Seedling dry weight	21	4	3	10	10	21
13	Protein content	9	20	20	20	4	9
14	Seed oil content	12	21	17	4	1	11
15	Germination percentage	20	7	11	7	11	20
16	Amount of chlorophyll	5	12	15	1	15	7
17	Amount of carotenoid	7	11	5	3	9	5
18	Empty pods per plant	11	5	8	9	8	12
19	Number of seeds per plant	10	10	10	15	6	10
20	Seedling vigor index	8	8	1	6	20	8
21	Number of pods per plant	2	2	4	12	18	2

Colors indicate the top 10 features ranked from high (blue) to low (red).

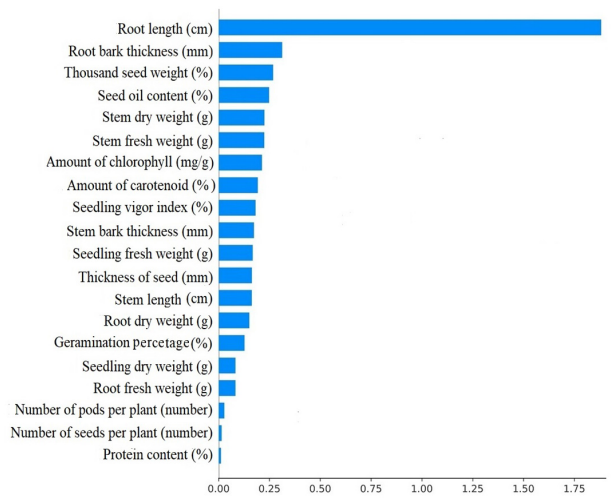

FIGURE 4 | Mean SHAP values for the GBT model.

by measuring the morphological and physiological changes or compounds released by infected plants in their defense (Golhani et al., 2018). The most popular indirect methods such as ML approaches offer a wide range of techniques for the detection of plant diseases (Golhani et al., 2018). The advantages and disadvantages of different types of detection methods for charcoal rot disease in crops are listed in **Supplementary Table 2**.

In agriculture research, ML methods are mainly used to detect, identify, and predict crop diseases and plant stress phenotyping (Yang and Guo, 2017). An efficient and precise prediction of plant diseases is a prerequisite in plant protection management. Moreover, early detection of disease minimizes the interference of humans (Golhani et al., 2018) which has been recently employed successfully (Saleem et al., 2019). However, prediction and quantification of charcoal rot disease are more crucial than the identification and classification of this disease in the future due to the implications of precise agriculture (Nagasubramanian et al., 2018). Such research works could lead to prevent the crop diseases at an early stage and cut costs of the pesticides (Barbedo, 2018).

In this work, specialized ML models were developed, for identification of charcoal rot disease by scrutinizing the symptoms of different parts of the soybean plants. In consequence of the lack of dataset for this disease, we have created our dataset; details of the dataset are provided in the dataset section. The main advantage of our proposed method is the identification of soybean charcoal rot disease at its early stage. A database of 2,000 soybean plants in natural field conditions was established. Supervised ML classifiers of LR-L1 LR-L2 MLP, RF, GBT, and SVM were trained to differentiate the healthy and infected soybean plants. Among these models, GBT classifier achieved a success rate of 96.79% through the analysis of the suggested feature set.

The occurrence of charcoal rot disease is regular, and the type and the probability of the soybean disease change during the soybean growth. Therefore, different charcoal rot disease identification techniques can be established by using the developed methods in this study. Furthermore, the automated charcoal rot disease prediction can be realized by combining identification models and domain knowledge of soybean disease. It has been previously reported that image processing and computer vision techniques can help to identify plant diseases (Golhani et al., 2018). The accuracy of the classification along with the image pre-processing could yield 90.5% recognition rate (Azlah et al., 2019). Thus far, only a few studies have been carried out to predict the charcoal rot disease development onset (Nagasubramanian et al., 2018). An algorithm such as image classification and image segmentation are mostly used for diseased charcoal rot identification (Saleem et al., 2019). These algorithms are used to classify healthy and no healthy plant leaves and stems of soybean (Saleem et al., 2019). By using the SVM approach, the highest classification accuracy was 95.76% and F1-score was 87% to identify the charcoal rot disease in soybeans (Nagasubramanian et al., 2018).

Although image processing and ML have provided significant evidence in the early prediction of disease, but different illumination conditions impact their performance (Mujika et al., 2018). Therefore, physiological evaluations can help to tackle this challenge (Khanna et al., 2019). Presented results have shown the applicability of the physiological features for the prediction of charcoal rot in soybean. As stated in **Table 2**, the result after using hybrid features, compared with only morphological features detailed in **Table 3** has slightly higher performance. Moreover, **Table 4** indicates the feature ranking based on various ML models highlighted the importance of physiological features in disease prediction.

In terms of classification performance, all methods performed well. GBT was the best preforming classifier as it tries to

sequentially improves the performance and also it includes the feature interactions in the learning. MLP had the lowest performance among others. It could be due to our small data size as neural networks usually needs larger data size to perform well. We note that the small size of the dataset and considering all features to have the same importance are the limitations of this study.

CONCLUSION

This paper investigated different ML algorithms for soybean charcoal rot disease detection and classification using morphological, physiological features. In this research effort, we presented an evaluation and comparison of six ML techniques on predicating charcoal rot disease. The results indicated that various ML techniques were slightly different in terms of their performance considering different evaluation metrics. Quantitative analysis of results indicated that GBT and SVM performed almost the same and demonstrated better performance compared with LR-L1, LR-L2, MLP, and RF approaches. Moreover, the feature ranking has shown the importance of including various features in the learning. Including other feature types such as chemical compositions and molecular structures and more data in the learning can be investigated as future work.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/Elham-khalili/Soybean-Charcoal-Rot-Disease-Prediction-Dataset-code.

AUTHOR CONTRIBUTIONS

EK, SK, SR, and FG formulated the research problem and designed the approaches. EK performed the experiments and collected the dataset. EK wrote the initial draft of the paper. All authors contributed to the final draft of the paper and approved the final version of the manuscript. EK, SK, and SR developed the processing workflow and performed the data analytics. All authors contributed to the writing and development of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Elite Foundation of Iran (INEF) from the Ministry of Higher Education, Iran.

ACKNOWLEDGMENTS

The Agricultural and Resource Research Center of Mazandaran (Iran) is acknowledged for the technical support during this study. We would also like to acknowledge valuable help and suggestions provided by our colleagues in Faculty of Science, Department of Plant Science, Tarbiat Modarres University, Tehran, Iran.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.590529/full#supplementary-material

REFERENCES

Abbas, H. K., Bellaloui, N., Accinelli, C., Smith, J. R., and Shier, W. T. (2019). Toxin production in soybean (Glycine max L.) plants with charcoal rot disease and by *Macrophomina phaseolina*, the fungus that causes the disease. *Toxins* 11:645. doi: 10.3390/toxins11110645

Ambrosio, M. M., Dantas, A. C., Martinez-Perez, E., Medeiros, A. C., Nunes, G. H., and Pico, M. B. (2015). Screening a variable germplasm collection of Cucumis melo L. for seedling resistance to *Macrophomina phaseolina*. *Euphytica* 206, 287–300. doi: 10.1007/s10681-015-1452-x

Ashfaq, R. A. R., Wang, X. Z., Huang, J. Z., Abbas, H., and He, Y. L. (2017). Fuzziness based semi-supervised learning approach for intrusion detection system. *Inform. Sci.* 378, 484–497. doi: 10.1016/j.ins.2016.04.019

Azlah, M. A. F., Chua, L. S., Rahmad, F. R., Abdullah, F. I., and Wan Alwi, S. R. (2019). Review on techniques for plant leaf classification and recognition. *Computers* 8, 4–77. doi: 10.3390/computers8040077

Barbedo, J. G. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. doi: 10.1016/j.biosystemseng.2018.05.013

Basu, S., Kumbier, K., Brown, J. B., and Yu, B. (2018). Iterative random forests to discover predictive and stable high-order interactions. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1943–1948. doi: 10.1073/pnas.1711236115

Boughorbel, S., Jarray, F., and El-Anbari, M. (2017). Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PloS One* 12:e0177678. doi: 10.1371/journal.pone.0177678

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern. Recognit.* 30, 1145–1159. doi: 10.1016/S0031-3203(96)00142-2

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.

Brevik, E. C., Slaughter, L., Singh, B. R., Steffan, J. J., Collier, D., Barnhart, P., et al. (2020). Soil and human health: current status and future needs. *Air. Soil. Water. Res.* 13, 3–21. doi: 10.1177/1178622120934441

Brzezinski, C. R., Henning, A. A., Abati, J., Henning, F. A., França-Neto, J. D. B., Krzyzanowski, F. C., et al. (2015). Seeds treatment times in the establishment and yield performance of soybean crops. *J. Seed. Sci.* 37, 147–153. doi: 10.1590/2317-1545v37n2148363

Chamorro, M., Miranda, L., Domínguez, P., Medina, J. J., Soria, C., Romero, F., et al. (2015). Evaluation of biosolarization for the control of charcoal rot disease (*Macrophomina phaseolina*) in strawberry. *Crop. Prot.* 67, 279–286. doi: 10.1016/j.cropro.2014.10.021

Cheng, J., Li, G., and Chen, X. (2018). Research on travel time prediction model of freeway based on gradient boosting decision tree. *IEEE Access* 7, 7466–7480. doi: 10.1109/ACCESS.2018.2886549

Chlingaryan, A., Sukkarieh, S., and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agric.* 151, 61–69. doi: 10.1016/j.compag.2018.05.012

Couronne, R., Probst, P., and Boulesteix, A. L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment.

BMC Bioinformatics 19, 270–284. doi: 10.1186/s12859-018-2264-5

Czarnecki, W. M., and Tabor, J. (2015). Multithreshold entropy linear classifier: theory and applications. *Expert Syst. Appl.* 42, 5591–5606. doi: 10.1016/j.eswa.2015.03.007

Dimitriadis, G., Neto, J. P., and Kampff, A. R. (2018). T-SNE visualization of large-scale neural recordings. *Neural. Comput.* 30, 1750–1774. doi: 10.1162/neco-a-01097

Elssied, N. O. F., Ibrahim, O., and Osman, A. H. (2014). A novel feature selection based on one-way anova f-test for e-mail spam classification. *J. Appl. Sci. Eng.* 7, 625–638. doi: 10.19026/rjaset.7.299

Eskandari, S., and Javidi, M. M. (2016). Online streaming feature selection using rough sets. *Int. J. Approx. Reason.* 69, 35–57. doi: 10.1016/j.ijar.2015.11.006

Eun, A. J. C., Huang, L., Chew, F. T., Li, S. F. Y., and Wong, S. M. (2002). Detection of two orchid viruses using quartz crystal microbalance (QCM) immunosensors.). *J. Virol. Methods* 99, 71–79. doi: 10.1016/S0166-0934(01)00382-2

Evans, J., and Poorter, H. (2001). Photosynthetic acclimation of plants to growth irradiance: the relative importance of specific leaf area and nitrogen partitioning in maximizing carbon gain. *Plant Cell Environ.* 24, 755–767. doi: 10.1046/j.1365-3040.2001.00724.x

Fang, Y., and Ramasamy, R. P. (2015). Current and prospective methods for plant disease detection. *Biosensors* 5, 537–561. doi: 10.3390/bios50305537

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010

Fenta, B. A., Beebe, S. E., Kunert, K. J., Burridge, J. D., Barlow, K. M., Lynch, J. P., et al. (2014). Field phenotyping of soybean roots for drought stress tolerance. *Agron* 4, 418–435. doi: 10.3390/agronomy4030418

Friedman, J. H. (2002). Stochastic gradient boosting. *Comput. Stat. Data An.* 38, 367–378. doi: 10.1016/S0167-9473(01)00065-2

Gaige, A. R., Ayella, A., and Shuai, B. (2010). Methyl jasmonate and ethylene induce partial resistance in Medicago truncatula against the charcoal rot pathogen *Macrophomina phaseolina*. *Physiol. Mol. Plants* 74, 412–418. doi: 10.1016/j.pmpp.2010.07.001

Golhani, K., Balasundram, S. K., Vadamalai, G., and Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* 5, 354–371. doi: 10.1016/j.inpa.2018.05.002

Grimes, M. L., Lee, W. J., Van der Maaten, L., and Shannon, P. (2013). Wrangling phosphoproteomic data to elucidate cancer signaling pathways. *PloS One* 3:e52884. doi: 10.1371/journal.pone.0052884

Gupta, G. K., Sharma, S. K., and Ramteke, R. (2012). Biology, epidemiology and management of the pathogenic fungus *Macrophomina phaseolina* (Tassi) Goid with special reference to charcoal rot of soybean (Glycine max (L.) *Merrill). J. Phytopathol.* 160, 167–180. doi: 10.1111/j.1439-0434.2012.01884.x

Hartman, G. L., Pawlowski, M. L., Herman, T. K., and Eastburn, D. (2016). Organically grown soybean production in the USA: constraints and management of pathogens and insect pests. *Agron* 6:16. doi: 10.3390/agronomy6010016

Hendry, G. A. F., and Price, A. H. (1993). "Stress indicators: chlorophylls and carotenoids," in *Methods in Comparative Plant Ecology Chapman and Hall*, eds G. A. F. Hendry, and J. P. Grime, (Berlin: Springer).

Holmes, G. J., Mansouripour, S. M., and Hewavitharana, S. S. (2020). Strawberries at the Crossroads: management of soilborne diseases in california without methyl bromide. *Phytopathology* 110, 956–968. doi: 10.1094/phyto-11-19-0406-ia

Irish, J. M. (2014). Beyond the age of cellular discovery. *Nat. Immunol.* 15, 1095–1097. doi: 10.1038/ni.3034

Islam, A. K. M. A., Anuar, N., and Yaakob, Z. (2009). Effect of genotypes and pre-sowing treatments on seed germination behavior of Jatropha. *Asian J. Plant Sci.* 8, 433–439. doi: 10.3923/ajps.2009.433.439

Jeong, J. J., Ju, H. J., and Noh, J. (2014). A review of detection methods for the plant viruses. *Res. Plant Dis.* 20, 173–181. doi: 10.5423/RPD.2014.20.3173

Joshi, R., Singh, J., and Vig, A. P. (2015). Vermicompost as an effective organic fertilizer and biocontrol agent: effect on growth, yield and quality of plants. *Rev. Environ. Sci. Biotechnol.* 14, 137–159. doi: 10.1007/s11157-014-9347-1

Katan, J. (2017). Diseases caused by soil borne pathogens: biology, management and challenges. *Eur. J. Plant Pathol.* 99, 305–315. doi: 10.4454/jpp.v99i2.3862

Khalili, E., Javed, M. A., Huyop, F., and Wahab, R. A. (2019). Efficacy and cost study of green fungicide formulated from crude beta-glucosidase. *Int. J. Environ. Sci. Technol.* 16, 4503–4518. doi: 10.1007/s13762-018-2084-1

Khanna, P., Chatterjee, K., Goyal, S., Pisharody, R. R., Patra, P., and Sharma, N. (2019). Psychological stress in the navy and a model for early detection. *J. Mar. Med. Soc.* 21, 116–120. doi: 10.4103/jmms.jmms-84-18

Kouchaki, S., Yang, Y., Walker, T. M., Sarah Walker, A., and Wilson, D. J. (2019). Application of machine learning techniques to tuberculosis drug resistance analysis. *Bioinformatics* 35, 2276–2282. doi: 10.1093/bioinformatics/bty949

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: a review. *Sensors* 18:2674. doi: 10.3390/s18082674

Lyons, M. B., Keith, D. A., Phinn, S. R., Mason, T. J., and Elith, J. (2018). A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sens. Environ.* 208, 145–153. doi: 10.1016/j.rse.2018.02.026

Marrocco, C., Duin, R. P., and Tortorella, F. (2008). Maximizing the area under the ROC curve by pairwise feature combination. *Pattern Recognit.* 41, 1961–1974. doi: 10.1016/j.patcog.2007.11.017

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., et al. (2015). Advanced methods of plant disease detection. *Rev. Agron. Sustain. Dev.* 35, 1–25. doi: 10.1007/s13593-014-0246-1

Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., et al. (2019). Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. *GigaScience* 8:giy153. doi: 10.1093/gigascience/giy153

Mujika, K. M., Mendez, J. A. J., and de Miguel, A. F. (2018). Advantages and disadvantages in image processing with free software in radiology. *J. Med. Syst.* 3, 36–42. doi: 10.1007/s10916-017-0888-z

Nagasubramanian, K., Jones, S., Sarkar, S., Singh, A. K., Singh, A., and Ganapathysubramanian, B. (2018). Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods* 14, 86–99. doi: 10.1186/s13007-018-0349-9

Padarian, J., McBratney, A. B., and Minasny, B. (2020). Game theory interpretation of digital soil mapping convolutional neural networks. *Soil* 2, 389–397. doi: 10.5194/soil-6-389-2020

Pastrana, A. M., Basallote-Ureba, M. J., Aguado, A., Akdi, K., and Capote, N. (2016). Biological control of strawberry soil-borne pathogens *Macrophomina phaseolina* and Fusarium solani, using Trichoderma asperellum and Bacillus spp. *Phytopathol. Mediterr.* 55, 109–120. doi: 10.14601/Phytopathol-Mediterr-16363

Patricio, D. I., and Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. *Comput. Electron. Agric.* 153, 69–81. doi: 10.1016/j.compag.2018.08.001

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. doi: 10.1186/1471-2105-12-77

Romero Luna, M. P., Mueller, D., Mengistu, A., Singh, A. K., Hartman, G. L., and Wise, K. A. (2017). Advancing our understanding of charcoal rot in soybeans. *J. Integr. Pest. Manag.* 8:1. doi: 10.1093/jipm/pmw020

Sakudo, A., Suganuma, Y., Kobayashi, T., Onodera, T., and Ikuta, K. (2006). Near-infrared spectroscopy: promising diagnostic tool for viral infections. *Biochem. Biophys. Res. Commun.* 341, 279–284. doi: 10.1016/j.bbrc.2005.12.153

Saleem, M. H., Potgieter, J., and Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants* 8, 468–488. doi: 10.3390/plants8110468

Sanchez, S., Grez, J., Contreras, E., Gil, P. M., and Gambardella, M. (2019). Physiological response and susceptibility of strawberry cultivars to the charcoal rot caused by *Macrophomina phaseolina* under drought stress conditions. *J. Berry Res.* 9, 165–177. doi: 10.3233/JBR-180329

Santos, C. A., Zanphorlin, L. M., Crucello, A., Tonoli, C. C., Ruller, R., Horta, M. A., et al. (2016). Crystal structure and biochemical characterization of the recombinant ThBgl, a GH1 β-glucosidase overexpressed in *Trichoderma harzianum* under biomass degradation conditions. *Biotechnol. Biofuels* 9:71. doi: 10.1186/s13068-016-0487-0

Schnyder, H., and Baum, U. (1992). Growth of the grain of wheat (Triticum aestivum L.). The relationship between water content and dry matter

accumulation. *Eur. J. Agron.* 1, 51–57. doi: 10.1016/S1161-0301(14)80001-4

Schoving, C., Stockle, C. O., Colombet, C., Champolivier, L., Debaeke, P., and Maury, P. (2020). Combining simple phenotyping and photothermal algorithm for the prediction of soybean phenology: application to a range of common cultivars grown in europe. *Front. Plant Sci.* 10:1755. doi: 10.3389/fpls.2019.01755

Schubert, E., and Gertz, M. (2017). Intrinsic t-stochastic neighbor embedding for visualization and outlier detection. *Similarity Search Appl. Cham.* 4, 188–203. doi: 10.1007/978-3-319-68474-1-13

Shan, J., Wang, Y., and Gao, W. (2018). Prediction of chemical exergy of organic substances using artificial neural network-multi layer perceptron. *Energ. Sour. Part A* 40, 1826–1832. doi: 10.1080/15567036.2018.1486924

Singh, A. P., Medida, S., and Duraisamy, K. (2017). Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. *AIAA J.* 55, 2215–2227. doi: 10.2514/1.J055595

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intel. Neurosci.* 2016:3289801. doi: 10.1155/2016/3289801

Sokolova, M., Japkowicz, N., and Szpakowicz, S. (2006). "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation," in *Proceedings of the Australasian Joint Conference on Artificial Intelligence*, (Berlin: Springer), 1015–1021. doi: 10.1007/11941439_114

Sun, S., Wang, X., Zhu, Z., Wang, B., and Wang, M. (2016). Occurrence of charcoal rot caused by *Macrophomina phaseolina*, an emerging disease of adzuki bean in China. *J. Phytopathol.* 164, 212–216. doi: 10.1111/jph.12413

Teixeira, A. L., Leal, J. P., and Falcao, A. O. (2013). Random forests for feature selection in QSPR Models-an application for predicting standard enthalpy of formation of hydrocarbons. *J. Cheminformatics* 1:9. doi: 10.1186/1758-2946-5-9

Thanarajoo, S. S., Kong, L. L., Kadir, J., Lau, W. H., and Vadamalai, G. (2014). Detection of Coconut cadang-cadang viroid (CCCVd) in oil palm by reverse transcription loop-mediated isothermal amplification (RT-LAMP). *J. Virol. Methods* 202, 19–23. doi: 10.1016/j.jviromet.2014.02.024

Uddin, S., Khan, A., Hossain, M. E., and Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Med. Inform. Decis. Mak.* 19:281. doi: 10.1186/s12911-019-1004-8

Ur Rahman, H., Ch, N. J., Manzoor, S., Najeeb, F., Siddique, M. Y., and Khan, R. A. (2017). A comparative analysis of machine learning approaches for plant disease identification. *Adv. Life. Sci.* 4, 120–126.

Vaknin, Y., Ghanim, M., Samra, S., Dvash, L., Hendelsman, E., Eisikowitch, D., et al. (2011). Predicting jatropha curcas seed-oil content, oil composition and protein content using near-infrared spectroscopy—a quick and non-destructive method. *Ind. Crops. Prod.* 34, 1029–1034. doi: 10.1016/j.indcrop.2011.03.011

Wang, A., Zhang, W., and Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* 158, 226–240. doi: 10.1016/j.compag.2019.02.005

Yang, X., and Guo, T. (2017). Machine learning in plant disease research. *Eur. J. Med. Res.* 3:6. doi: 10.18088/ejbmr

Zhang, J., Yang, Y., Feng, X., Xu, H., Chen, J., and He, Y. (2020). Identification of bacterial blight resistant rice seeds using terahertz imaging and hyperspectral imaging combined with convolutional neural network. *Front. Plant Sci.* 11:821. doi: 10.3389/fpls.2020.00821

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Khalili, Kouchaki, Ramazi and Ghanati. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

METHODS
published: 14 December 2020
doi: 10.3389/fpls.2020.601250

Attention-Based Recurrent Neural Network for Plant Disease Classification

Sue Han Lee[1], *Hervé Goëau*[2,3†], *Pierre Bonnet*[2,3*†] *and Alexis Joly*[4†]

[1] Swinburne University of Technology Sarawak Campus, Kuching, Malaysia, [2] AMAP, Univ Montpellier, CIRAD, CNRS, INRA, IRD, Montpellier, France, [3] CIRAD, UMR AMAP, Montpellier, France, [4] INRIA Sophia-Antipolis - ZENITH Team, LIRMM - UMR 5506, Montpellier, France

OPEN ACCESS

Edited by:
Spyros Fountas,
Agricultural University of Athens, Greece

Reviewed by:
John Louis Van Hemert,
Corteva Agriscience™, United States
Michael Gomez Selvaraj,
Consultative Group on International Agricultural Research (CGIAR), United States

***Correspondence:**
Pierre Bonnet
pierre.bonnet@cirad.fr

†ORCID:
Hervé Goëau
orcid.org/0000-0003-3296-3795
Pierre Bonnet
orcid.org/0000-0002-2828-4389
Alexis Joly
orcid.org/0000-0002-2161-9940

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: 31 August 2020
Accepted: 09 November 2020
Published: 14 December 2020

Citation:
Lee SH, Goëau H, Bonnet P and Joly A (2020) Attention-Based Recurrent Neural Network for Plant Disease Classification. Front. Plant Sci. 11:601250. doi: 10.3389/fpls.2020.601250

Plant diseases have a significant impact on global food security and the world's agricultural economy. Their early detection and classification increase the chances of setting up effective control measures, which is why the search for automatic systems that allow this is of major interest to our society. Several recent studies have reported promising results in the classification of plant diseases from RGB images on the basis of Convolutional Neural Networks (CNN). These studies have been successfully experimented on a large number of crops and symptoms, and they have shown significant advantages in the support of human expertise. However, the CNN models still have limitations. In particular, CNN models do not necessarily focus on the visible parts affected by a plant disease to allow their classification, and they can sometimes take into account irrelevant backgrounds or healthy plant parts. In this paper, we therefore develop a new technique based on a Recurrent Neural Network (RNN) to automatically locate infected regions and extract relevant features for disease classification. We show experimentally that our RNN-based approach is more robust and has a greater ability to generalize to unseen infected crop species as well as to different plant disease domain images compared to classical CNN approaches. We also analyze the focus of attention as learned by our RNN and show that our approach is capable of accurately locating infectious diseases in plants. Our approach, which has been tested on a large number of plant species, should thus contribute to the development of more effective means of detecting and classifying crop pathogens in the near future.

Keywords: plant disease classification, deep learning, recurrent neural network, automated visual crops analysis, precision agriculture technologies, crops monitoring, pests analysis, smart farming

1. INTRODUCTION

Plant diseases are a major threat to agricultural production, causing severe food recessions and affecting crop quality (Bhange and Hingoliwala, 2015). To detect plant diseases in crops, plant pathologists generally use molecular and serological methods or measurements of various parameters, such as morphological change, temperature change, change in transpiration rate, or volatile organic compound emission from infected plants (Fang et al., 2015). Although it is an effective means of controlling plant diseases, consulting experts is nonetheless a costly and time-consuming process, especially since it is not always easy to bring an expert in time

before the disease spreads to the crops. In recent years, automated classification of plant diseases has been addressed by the computer vision community to compensate for the lack of human expertise. Researchers used deep learning techniques to automatically identify diseases in individual crops, such as banana (Selvaraj et al., 2019), coffee (Kumar et al., 2020), grape (Liu et al., 2020), cassava (Ramcharan et al., 2017), tomato (Durmuş et al., 2017; Fuentes et al., 2017; Liu and Wang, 2020), and apple (Liu et al., 2017), as well as in multi-crops (Mohanty et al., 2016; Ferentinos, 2018; Too et al., 2018). In most cases, researchers fine-tune off-the-shelf Convolutional Neural Networks (CNNs) (Saleem et al., 2019).

Although the evaluated CNN methods in these publications appear to be effective and seem to learn relevant feature representations of the diseases, they unfortunately also learn irrelevant disease characteristics such as background noise (Mohanty et al., 2016; Atabay, 2017) or uninfected plant parts (Ferentinos, 2018; Toda and Okura, 2019; Lee et al., 2020). For example, (Atabay, 2017) has shown that a CNN trained on tomato plant diseases has neuron activations that fall mostly in the background. Unfortunately, it has been shown that background suppression with image segmentation does not give better results than an ordinary colored background with CNN (Mohanty et al., 2016), confirming a dependence of background characteristics for disease identification. Even worse, Ferentinos (2018) showed that a CNN tends to be confused between similar crops of different disease classes. It is thus indicated that a CNN model, which is supposed to learn the visual representation of plant diseases, tends to be biased toward irrelevant crop characteristics. Region-based deep neural networks can help to focus on contaminated parts (Fuentes et al., 2017, 2019), but such a technique involves labor-intensive annotations of disease locations and also depends heavily on prior knowledge of plant diseases.

Henceforth, these observations have motivated us to go beyond existing practices by exploring a new technique for identifying plant diseases that allows us to automatically learn the regions of interest in the plant image, which correspond to the infected regions, and then to identify the diseases. Inspired by recent work on multi-organ plant identification that has shown the ability of an attention-based Recurrent Neural Network (RNN) to locate relevant regions of plant structures without any prior human annotation (Lee et al., 2018), we have adapted this approach to learn visual representations of plant diseases and show that discriminating infected regions of a plant can be successfully located and highlighted for disease identification, as illustrated in **Figure 1**.

Our contribution in this paper is three-fold: firstly, to our knowledge, this is the first time that the RNN-based approach is being explored to learn representations of plant diseases at that scale and that a comparison of identification performance is made against the widely used CNN approaches in this field. Second, we show quantitatively that the RNN approach outperforms the CNN approaches. Finally, we also show qualitatively that the RNN approach is able to detect precisely the infected regions through the neuron activations. The source code of our computational implementation is provided on an open

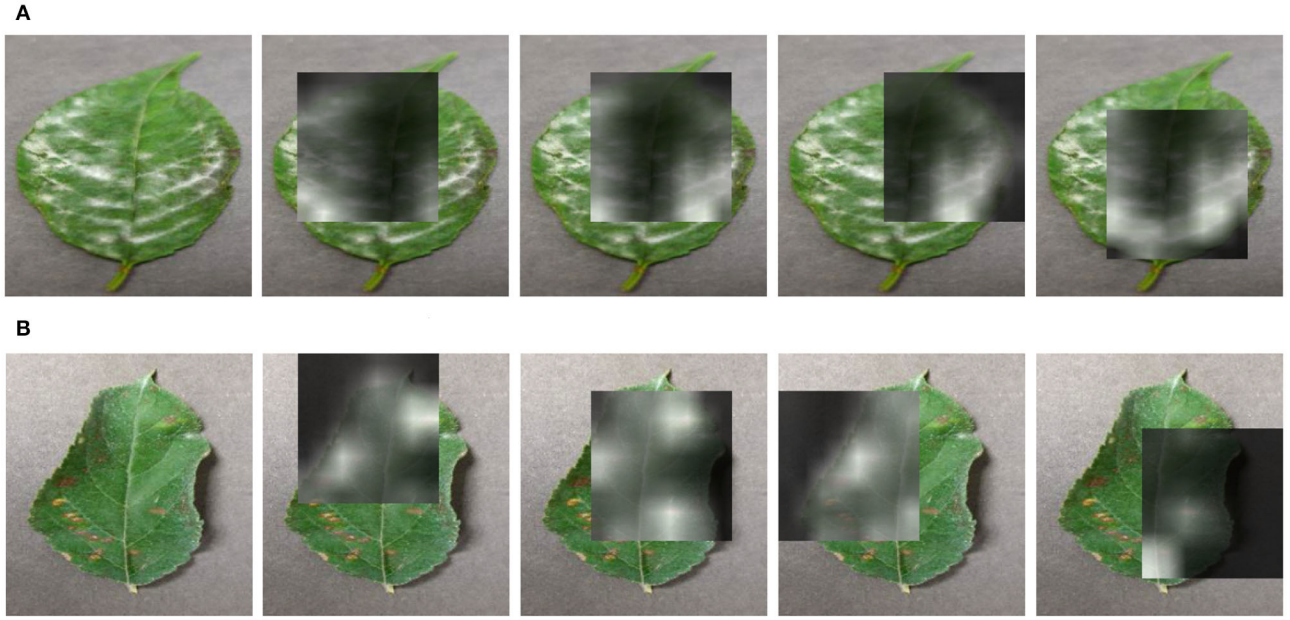

FIGURE 1 | The learned attention maps by our proposed approach on two diseases that have contaminated leaves. The visualizations highlight the infected regions. **(A)** Powdery mildew disease, on cherry plant. **(B)** Cedar apple rust disease, on apple.

online repository to facilitate its long-term accessibility and use by the scientific community[1].

2. MATERIALS AND METHODS

2.1. Attention-Based RNN Model

A recurrent neural network (RNN) is a class of neural networks where connections between nodes of a layer form a directed graph along a sequence of variables (e.g., a temporal sequence). The recurrent connections typically allow for modeling of the relationship between the current state of a variable and the previous states (similarly to a Markov chain). The RNN-based approach has received much attention because of its ability to handle sequential data to make predictions, such as in language translation (Sutskever et al., 2014) or action recognition (Du et al., 2018; Song et al., 2018). Improved RNN models, such as Long Short-Term Memory networks (LSTMs) or Gated Recurrent Units (GRU), enable training on long sequences, overcoming problems like vanishing gradients. Recently, a few publications have shown the effectiveness of RNN approaches to sequentially process variable-length data of fixed sizes, such as a picture. For example, it has been shown that a RNN architecture based on GRU can efficiently model dependencies between different images of plant observations Lee et al. (2018) or that LSTM can be used to capture discriminating regions of images for fine-grained classification (Zhao et al., 2017). Attention is a mechanism that can be combined in the RNN to allow it to focus on certain parts of the input when predicting a certain part of the output, thus enabling an easier learning and of higher quality. For instance, RNN with attention mechanism was used in Ren and Zemel (2017) to capture the spatial structure in images and produce detailed instance segmentation.

Inspired by these previous works, we combine in a RNN an attention mechanism with Gated Recurrent Units to dynamically push salient plant disease characteristics to the forefront in order to strengthen the model in learning disease characteristics for identification. **Figure 2** shows the framework of the proposed architecture. First, a CNN trained on a plant disease classification task is used as a visual features extractor: a plant image is thus encoded as *CNN features* (i.e., a tensor of feature maps at a given output of a chosen convolutional layer). These CNN features can be considered as a new smaller image of activations with as many channels as filters used in the convolution layer. This new image is then sliced into sub-parts of the same size to get local activations in many regions covering all of the image. These new local CNN features can then be used to build a sequence and feed an RNN based on GRUs, enabling an attention mechanism to locate important parts or components in the CNN features. It extends the effective pixel neighborhood in each sub-part and maximizes the information gain across several sub-parts of the CNN features. Finally, prediction error is minimized throughout the optimization process.

[1]the web link to the open online repository on which the source code is hosted will be displayed here after acceptance of the manuscript.

2.1.1. Our Overall RNN Architecture

We denote a plant disease image as $\mathbf{I}$ and the corresponding feature maps extracted by the convolutional layers of the CNN as $\boldsymbol{\delta} \in \mathbb{R}^{H \times W \times C}$, where H, W and C are, respectively, the height, width, and number of channels in the feature maps. The CNN model is initially pre-trained and optimized purely based on plant disease target classes. A sequence of T regional feature maps $\{\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \cdots, \boldsymbol{\delta}_T\} \in \boldsymbol{\delta}$ is then generated by slicing the global feature map $\boldsymbol{\delta}$ following the sliding direction shown in **Figure 2**. The resulting sequence of feature maps is then used as input of the *RNN* module displayed in **Figure 2A** and detailed in **Figure 2B**. Thanks to the RNN connections between the different feature maps, the network is able to iteratively learn the discriminant visual patterns and to model the spatial relationship between them. For instance, brown specks that spread from side to side on a part of the leaf can be distinguished from brown specks that appear randomly on the leaf.

2.1.2. Attention Module

The attention module is used to model the relative contribution of each pixel of the T regional feature maps. Specifically, it forces an explicit additional step in the reasoning process, identifying salient regions by assigning different importance to features from different image regions. The attention mechanism is introduced by the $\boldsymbol{\lambda_t}$ terms (also called *regional attention map*) that control the contribution of the pixels of the t-th state and that are trained by the neural network. A larger $\boldsymbol{\lambda_t}$ value indicates higher importance. More formally, the attention function $g : \boldsymbol{\delta_t}, \mathbf{h_{t-1}} \rightarrow \boldsymbol{\epsilon_t}$ is defined as follows:

$$\boldsymbol{\zeta_t} = \{\tanh(\boldsymbol{\delta_t}\mathbf{W_\delta} + \mathbf{h_{t-1}}\mathbf{W_h})\}\mathbf{W_a} \quad (1)$$

$$\boldsymbol{\lambda_t} = softmax(\boldsymbol{\zeta_t}) \quad (2)$$

$$\boldsymbol{\epsilon_t} = g(\boldsymbol{\delta_t}, \mathbf{h_{t-1}}) = \sum_{i,j} \boldsymbol{\lambda_{t,ij}}.\boldsymbol{\delta_{t,ij}} \quad (3)$$

where $\mathbf{W_\delta} \in \mathbb{R}^{C \times C}$, $\mathbf{W_h} \in \mathbb{R}^{E \times C}$, $\mathbf{W_a} \in \mathbb{R}^{C \times 1}$ are the embedding matrices, E is the dimensionality of GRU cell, and $\boldsymbol{\delta_{t,ij}}$ denotes the value of the t-th regional feature map at position $(i, j) \in H' \times W'$. Note that $\boldsymbol{\epsilon_t}$ is the output representation (feature vector) for the t-th regional feature map.

2.1.3. Bi-Directional Training

Inspired by previous experiments (Lee et al., 2018) that show that bidirectional states modeling performs better compared to uni-state modeling in plant-view correlation learning, we built a bidirectional states modeling mechanism where the forward neuron activations $\overrightarrow{\mathbf{h_T}}$ and the backward neuron activations $\overleftarrow{\mathbf{h_0}}$ model $P(\mathbf{h_t}|\boldsymbol{\delta_t}, \mathbf{h_0}, \cdots, \mathbf{h_{T-1}})$ and $P(\mathbf{h_t}|\boldsymbol{\delta_t}, \mathbf{h_T}, \cdots, \mathbf{h_1})$, respectively. In order to correlate between both states, the final output activations of the forward and backward GRU are cascaded as follows: $\mathbf{h} = [\overrightarrow{\mathbf{h_T}}, \overleftarrow{\mathbf{h_0}}]$. We then multiply $\mathbf{h}$ with a class embedding matrix, $\mathbf{W_{em}}$, which is $\mathbf{s(I)} = \mathbf{W_{em}h}$ before normalizing it with a softmax function: $P(r|\mathbf{I}) = \frac{e^{\mathbf{s_r(I)}}}{\sum_{m=1}^{M} e^{\mathbf{s_m(I)}}}$ where M and r stand for the total number of classes and the target

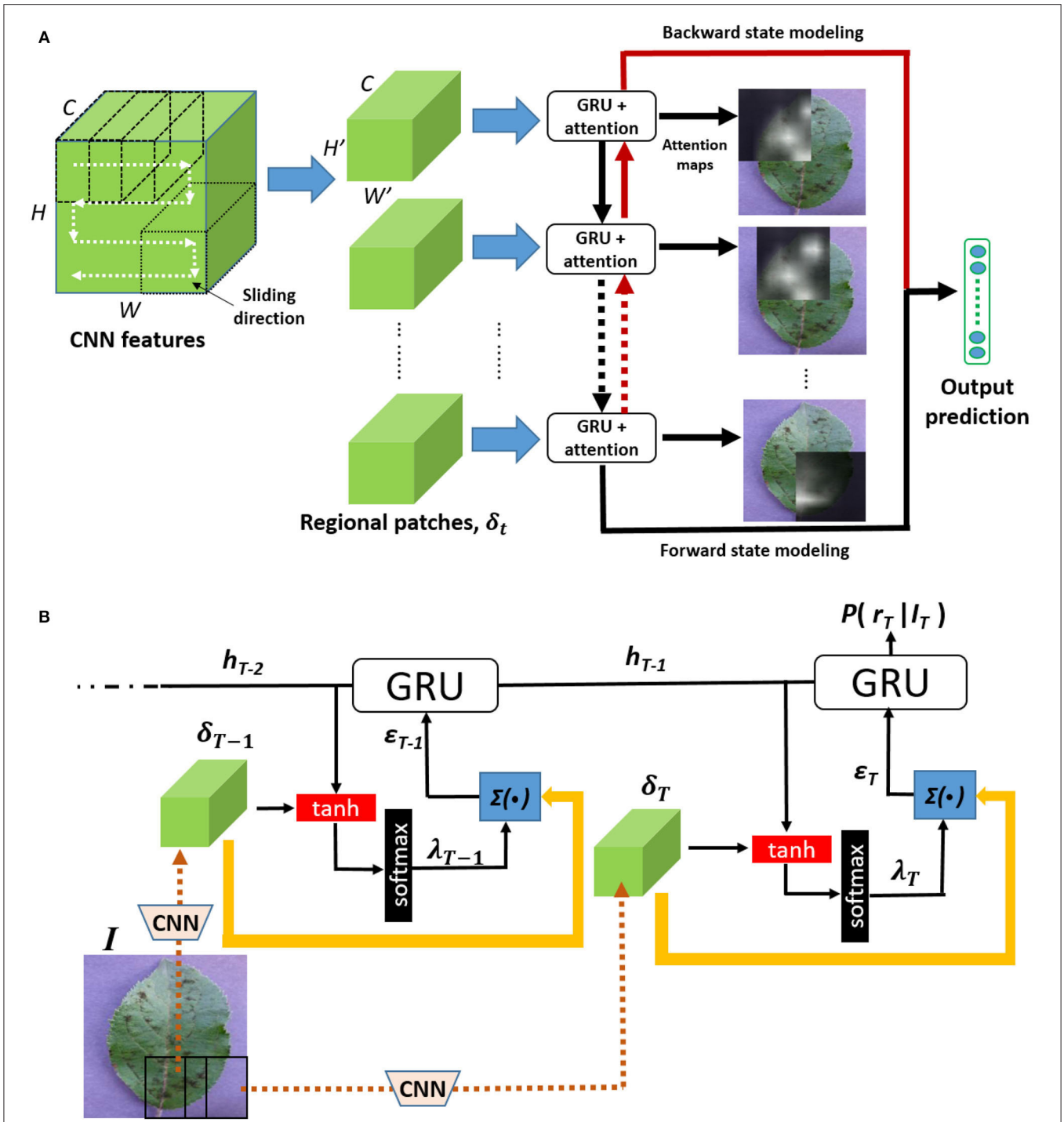

FIGURE 2 | The proposed architecture: from an image of a contaminated plant, feature maps are first extracted from a given convolutional layer of a pre-trained CNN. They are then sliced into several patches following a "snaking" sliding direction. The patches then feed into Gated Recurrent Units that share, combine, and retain relevant information in a bidirectional way to update an internal representation of plant disease. Soft attention mechanism is used to infer discriminating local features. **(A)** Overall architecture. **(B)** Soft attention mechanism.

class, respectively. After performing the softmax operation, we find the maximum likelihood of the sample by applying the objective function, $L = -logP(r|\mathbf{I})$.

2.1.4. Implementation Details

To extract the CNN features, we used an extension of a GoogleNet architecture Szegedy et al. (2015) with modified

FIGURE 3 | Visualization of activations of the learned features with the CNN model (GoogleNet). Best viewed in color.

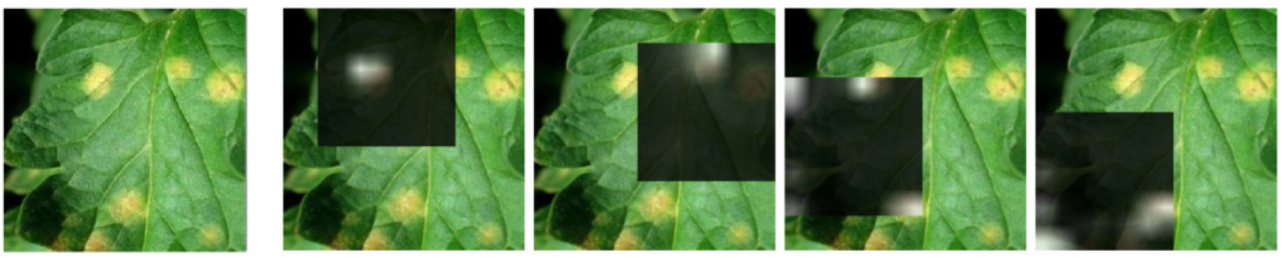

FIGURE 4 | Visualization of attention maps learned using our attention-based RNN on a leaf infected by the *Leaf mold* disease (to be compared with **Figure 3**). Best view in color.

convolutional layers and additional batch normalization to increase accuracy and reduce computational complexity[2]. After training the CNN on the disease classification task, we extracted CNN features with a size of $14 \times 14 \times 576$ from the convolutional layer *Inception_4d* and sliced them with a stride of 1 into regional patches with a size of $H' = W' = 8$ each to finally feed the RNN. The RNN is trained using the Tensorflow library (Abadi et al., 2016). We use the ADAM optimizer (Kingma and Ba, 2014) with the parameters $\alpha = 1e - 08$, $\beta 1 = 0.9$, and $\beta 2 = 0.999$. We applied the weight decay L_2 with the penalty multiplier set to 1×10^{-4} and dropout ratio set to 0.5, respectively. We set the learning rate to 1×10^{-4}, and the mini batch size was set to 30.

2.1.5. Features Visualization Method

We describe here the methodology used to visualize the visual features captured by the CNN and the RNN model (see **Figures 3–5** of the results section). For the CNN model (GoogleNet), we first tracked the position of the highest activation across all the feature maps extracted from the last convolutional layer. From this, we accumulated the first 30 dominant activations and assessed them according to the original image. For the RNN model, we simply displayed a subset of the regional attention maps $\boldsymbol{\lambda_t}$.

[2]https://github.com/AdelineMomo/CNN-plant-disease (Caffe).

2.2. Experimental Dataset

Plant Village (PV) (Hughes and Salathé, 2015) is a popular dataset dedicated to the evaluation of automated identification of plant diseases under controlled environments. It has 38 crop-disease pairs, with 26 crop-disease categories concerning 14 crop plants. The dataset was provided with predefined training and test sets, and a configuration with a percentage ratio of 80 and 20% is used, giving a total of 10,495 training images and 4,310 test images. Since plant diseases, named by vernacular names, share the same visual characteristics for different species, we categorized leaf samples from the PV dataset into 21 classes (20 diseases and one healthy class) and trained a classifier based on these classes. Note that, in practice, it is impossible to collect all disease samples from different crops under different environments to train a deep model. In fact, what we intended to achieve is a model that is general enough to represent knowledge in a way that can be transferred between different plant disease tasks. We therefore explored the generalization of models to identify disease of unseen crops and also plants that are captured under different contexts (typically in the field).

To evaluate the ability of the model to generalize to unseen crops, we excluded one crop from the training set and use it only for testing. More specifically, we removed all the leaf samples from the pepper crop, i.e., the ones from the *Pepper_bell Bacterial_spot* class and the ones from the *Pepper_bell healthy*, so as to treat the pepper class as an unseen crop in this experiment

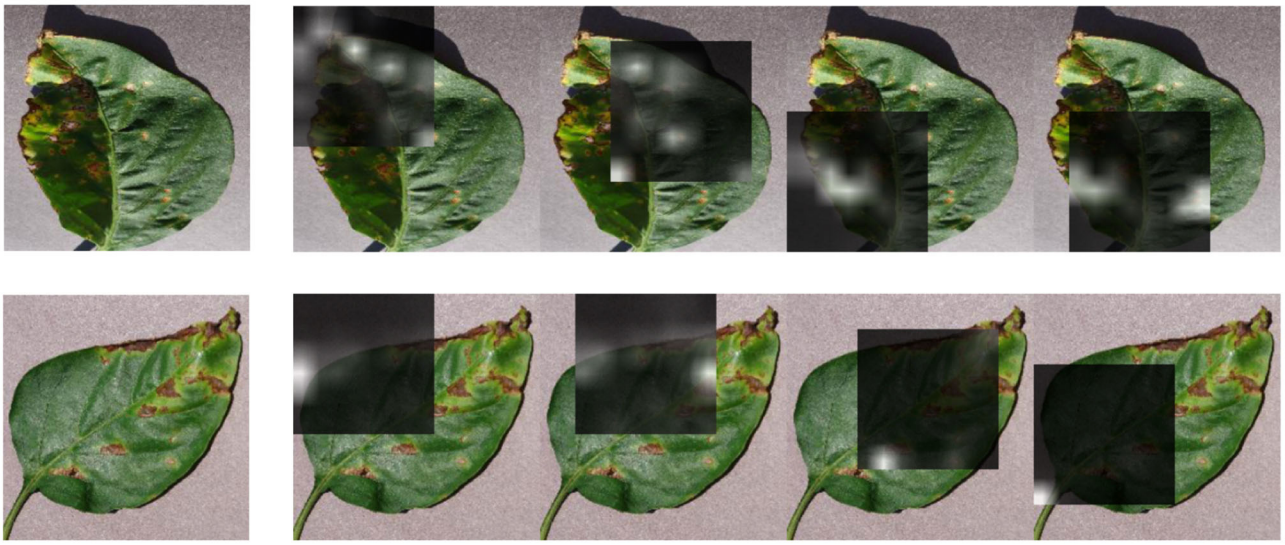

FIGURE 5 | Visualization of attention maps learned using our new attention-based RNN model on two correctly classified images of the Unseen Crop of *pepper_bell Bacterial_spot*.

TABLE 1 | Top-1 accuracy (%) comparison on Plant Village (PV), IPM, and Bing.

Method	PV-SC	PV-UCB	IPM-SC	Bing-SC
CNN InceptionV3 of Brahimi et al. (2018)	98.05	29.63	29.06	28.57
Our CNN (GoogleNet)	**99.17**	18.98	37.61	36.51
Our new model Seq-RNN	98.17	**58.80**	**40.17**	**39.68**

Note that the performance of IPM and Bing are based only on Seen Crop (SC) images, whereas the performance on PV is measured for both Seen Crop (PV-SC) and Unseen Crop (PV-UCB), i.e., for the pepper_bell Bacterial_spot.
The best accuracy for each dataset is shown in bold.

while the *Bacterial_spot* can be learnt through other crops. Besides the PV dataset, we also assessed the robustness of our models using pictures related to the same disease categorization and from reliable online sources that are not restricted to a controlled environment. We used 119 and 64 images from IPM and Bing, respectively, collected by Mohanty et al. (2016) (121 number of Bing images were expected, but half of them are no longer available).

3. RESULTS AND DISCUSSION

The experimental results are presented in the **Table 1**. As there are only few images of pepper crops in IPM and Bing (two images in IPM and one in Bing), we considered the number of collections insufficient to infer the performance of the models in recognizing the disease of an unseen crop. We therefore tested the model using only the seen crop images in IPM and Bing. We compared the performance of the models with the Inception-V3 model employed by Brahimi et al. (2018) since it was reported to be the best approach for disease identification on the PV dataset. The model is trained on the basis of the aforementioned 21 target classes with the pepper crop excluded from the training set.

First, by examining the accuracy of the PV-SC test set (seen crops from the PlantVillage dataset), we can see that both CNN models and our new attention-based RNN achieved very high accuracy values. This is mainly due to the fact that the images in this test set were acquired under exactly the same conditions as the training set so that any method can exhibit a high performance. Secondly, we can observe that both CNN models as well as our new attention-based RNN achieved lower accuracy values for the IPM-SC and Bing-SC test sets as well as the PV-UCB (unseen crop from the PlantVillage dataset) compared to the PV-SC. We believe that this is due to a change in the data distribution between the PV data and the IPM-SC and BING-SC test sets, as the two data are collected under different conditions, and the disease training data are not sufficiently diverse to cover the ranges of visual appearance of the disease found in the unseen crop. To cope with such difficult datasets, CNN models achieved a much lower performance, which means that the models formed have some difficulty in generalizing to the unseen crop of PlantVillage (PV-USC) or to images acquired in a different domain (IPM-SC and Bing-SC). In comparison, our attention-based RNN model was much more accurate over the three series of tests. Although the accuracy is far from perfect, it is still quite reasonable considering the difficulty of the problem.

We further compared the generalization ability of our new model and that of a classical CNN thanks to visualization experiments presented in **Figures 3**, **4**. As shown in **Figure 3**, the regions of the image leading to a strong activation of the CNN neurons do not really seem to correspond to the visual patterns characteristic of the disease. Indeed, they largely correspond to healthy leaf features, specifically the venation. On the contrary, **Figure 4** shows that the activated regions of the attention maps

of our new RNN model do match the disease spots much more precisely and accurately. **Figure 5** shows a similar visualization on two leaves of the *Pepper_bell* crop affected by *Bacterial_spot*. Here again, we can observe that most of the activated regions do correspond to altered parts of the leaf, whereas the *Pepper_bell* was not even present in the training set.

In line with our results, we deduced that the transferability of Seq-RNN knowledge is more useful than that of CNN in differentiating data, especially those taken in the field. This could be due to the fact that CNN is learned on the basis that its spatial information collapses in the final convolutional layer, resulting in the relativity of local features that are important for representing diseases not retained. This might be also the reason the CNN searches for global characteristics that are not relevant for the infected area but are relevant for leaf characteristics, such as shape and venation.

On the other hand, Seq-RNN takes the convolutional map, where substantial spatial information is retained as input and is formulated in such a way as to learn the relationship between the neighboring regions of an image. Because of this formulation, the Seq-RNN's attention mechanism will be forced to detect salient regions, where, in this case, there are the obvious infected regions that appear in the local areas of an image. This disease-focused knowledge could be effectively transferred to field images where the plant structure may not be easily visible (due to organ deformation or clutter) but the visual aspect of the disease is evident. Thus, in field images where the plant structure is not visible or may be difficult to distinguish, the disease characteristics become more determinant, allowing Seq-RNN, which has a greater knowledge of the disease, to better differentiate the data than the CNN.

4. CONCLUSIONS AND FUTURE DIRECTIONS

We presented in this study a new efficient computational architecture that opens up new perspectives for the automated classification of plant diseases. We showed that our RNN approach has a higher generalization ability than the classical CNN approach, especially in distinguishing disease samples that are different from the training set. This is a major critical point in plant pathology, as it is highly difficult (due to required expertise level and time consumed) to produce a complete and diversified visual dataset of all the symptoms of any crop disease at a global scale. Our approach could thus overcome the problems related to the lack of training data usually necessary for the development of performing recognition deep learning models.

In this study we also analyzed the attention maps learned by our RNN and showed that it meets our expectation of localization of infectious diseases in plants images. We believe that our RNN-based approach, which captures the context of the relationship between local features, could provide a better insight into where the machine actually detects relevant information. It is important to note that by assessing the machine's perspective in disease differentiation, humans will certainly benefit, as we all know that human capabilities are limited when it comes to identifying the thousands of diseases for all the plants in the world.

In this paper, we have shown that the integration of our RNN approach in the design of the deep learning architecture for learning disease representation can not only provide a better classification performance but also contribute to the knowledge of plant diseases, which could potentially be useful to address doubts that have not yet been resolved. Future research directions should focus on the integration of CNN and RNN based models to simultaneously address the learning of rich global and local visual representations within a deep end-to-end network. Furthermore, alternative slicing patterns on convolutional maps before transmitting the extracted patches to RNN should be studied for a better modeling of the local characteristics of the plant disease. In addition, complementary statistical analyses should be performed on the evaluation of the RNN approach for field images that are of main interest to farmers and all field agricultural actors. We believe that these findings could encourage future research to rethink the current de facto paradigm of purely relying on the CNN in plant disease identification.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://doi.org/10.3389/fpls.2016.01419 (Mohanty et al., 2016) and https://arxiv.org/abs/1511.08060 (Hughes and Salathé, 2015).

AUTHOR CONTRIBUTIONS

SL and HG conducted experiments and data analysis. SL wrote the first draft of the manuscript. HG, PB, and AJ revised the manuscript. All authors designed the research, contributed to the article and approved the submitted version.

FUNDING

This project was supported by Agropolis Fondation, Numev, Cemeb, #DigitAG, under the reference ID 1604-019 through the Investissements d'avenir programme (Labex Agro: ANR-10-LABX-0001-01). This work was also supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "Tensorflow: a system for large-scale machine learning," in *OSDI* (Savannah, GA), 265–283.

Atabay, H. A. (2017). Deep residual learning for tomato plant leaf disease identification. *J. Theoret. Appl. Inform. Technol.* 95, 6800–6808.

Bhange, M., and Hingoliwala, H. (2015). Smart farming: pomegranate disease detection using image processing. *Proc. Comput. Sci.* 58, 280–288. doi: 10.1016/j.procs.2015.08.022

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., and Moussaoui, A. (2018). "Deep learning for plant diseases: detection and saliency map visualisation," in *Human and Machine Learning*, eds J. Zhou and F. Chen (Heidelberg: Springer), 93–117. doi: 10.1007/978-3-319-90403-0_6

Du, W., Wang, Y., and Qiao, Y. (2018). Recurrent spatial-temporal attention network for action recognition in videos. *IEEE Trans. Image Process.* 27, 1347–1360. doi: 10.1109/TIP.2017.2778563

Durmuş, H., Güneş, E. O., and Kırcı, M. (2017). "Disease detection on the leaves of the tomato plants by using deep 1learning," in *2017 6th International Conference on Agro-Geoinformatics* (Fairfax, VA: IEEE), 1–5. doi: 10.1109/Agro-Geoinformatics.2017.8047016

Fang, Y., Ramasamy, R. P., and Rains, G. C. (2015). "Current and prospective methods for plant disease detection," in *Biosensors Vol. 5,* ed G. C. Rains (Basel: MDPI), 537–561. doi: 10.3390/bios5030537

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009

Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17:2022. doi: 10.3390/s17092022

Fuentes, A., Yoon, S., and Park, D. S. (2019). Deep learning-based phenotyping system with glocal description of plant anomalies and symptoms. *Front. Plant Sci.* 10:1321. doi: 10.3389/fpls.2019.01321

Hughes, D., and Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv [preprint]. arXiv:1511.08060*. Available online at: https://arxiv.org/abs/1511.08060

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv [preprint]. arXiv:1412.6980*. Available online at: https://arxiv.org/abs/1412.6980

Kumar, M., Gupta, P., and Madhav, P. (2020). "Disease detection in coffee plants using convolutional neural network," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (Coimbatore: IEEE), 755–760. doi: 10.1109/ICCES48766.2020.9138000

Lee, S. H., Chan, C. S., and Remagnino, P. (2018). Multi-organ plant classification based on convolutional and recurrent neural networks. *IEEE Trans. Image Process.* 27, 4287–4301. doi: 10.1109/TIP.2018.2836321

Lee, S. H., Goëau, H., Bonnet, P., and Joly, A. (2020). New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* 170:105220. doi: 10.1016/j.compag.2020.105220

Liu, B., Ding, Z., Tian, L., He, D., Li, S., and Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Front. Plant Sci.* 11:1082. doi: 10.3389/fpls.2020.01082

Liu, B., Zhang, Y., He, D., and Li, Y. (2017). Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* 10:11. doi: 10.3390/sym10010011

Liu, J., and Wang, X. (2020). Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Front. Plant Sci.* 11:898. doi: 10.3389/fpls.2020.00898

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls.2016.01419

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Front. Plant Sci.* 8:1852. doi: 10.3389/fpls.2017.01852

Ren, M., and Zemel, R. S. (2017). "End-to-end instance segmentation with recurrent attention," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI), 21–26. doi: 10.1109/CVPR.2017.39

Saleem, M. H., Potgieter, J., and Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants* 8:468. doi: 10.3390/plants8110468

Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., et al. (2019). Ai-powered banana diseases and pest detection. *Plant Methods* 15:92. doi: 10.1186/s13007-019-0475-z

Song, S., Lan, C., Xing, J., Zeng, W., and Liu, J. (2018). Spatio-temporal attention-based LSTM networks for 3d action recognition and detection. *IEEE Trans. Image Process.* 27, 3459–3471. doi: 10.1109/TIP.2018.2818328

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems* (Montreal, QC), 3104–3112.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA), 1–9. doi: 10.1109/CVPR.2015.7298594

Toda, Y., and Okura, F. (2019). How convolutional neural networks diagnose plant disease. *Plant Phenomics* 2019:9237136. doi: 10.1155/2019/9237136

Too, E. C., Yujian, L., Njuki, S., and Yingchun, L. (2018). A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161, 272–279. doi: 10.1016/j.compag.2018.03.032

Zhao, B., Wu, X., Feng, J., Peng, Q., and Yan, S. (2017). Diversified visual attention networks for fine-grained object classification. *IEEE Trans. Multimedia* 19, 1245–1256. doi: 10.1109/TMM.2017.2648498

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lee, Goëau, Bonnet and Joly. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 December 2020
doi: 10.3389/fpls.2020.599616

Application of Laser-Induced Breakdown Spectroscopy in Detection of Cadmium Content in Rice Stems

Wei Wang[1†], Wenwen Kong[1,2†], Tingting Shen[1], Zun Man[1], Wenjing Zhu[3], Yong He[1,4], Fei Liu[1,4] and Yufei Liu[1,4]*

[1] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2] School of Information Engineering, Zhejiang A & F University, Hangzhou, China, [3] School of Agricultural Equipment Engineering, Jiangsu University, Zhenjiang, China, [4] Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou, China

OPEN ACCESS

Edited by:
Spyros Fountas,
Agricultural University of Athens,
Greece

Reviewed by:
Panmanas Sirisomboon,
King Mongkut's Institute
of Technology Ladkrabang, Thailand
Daming Fan,
Jiangnan University, China

***Correspondence:**
Yufei Liu
yufeiliu@zju.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: 27 August 2020
Accepted: 30 November 2020
Published: 18 December 2020

Citation:
Wang W, Kong W, Shen T, Man Z, Zhu W, He Y, Liu F and Liu Y (2020) Application of Laser-Induced Breakdown Spectroscopy in Detection of Cadmium Content in Rice Stems. Front. Plant Sci. 11:599616. doi: 10.3389/fpls.2020.599616

The presence of cadmium in rice stems is a limiting factor that restricts its function as biomass. In order to prevent potential risks of heavy metals in rice straws, this study introduced a fast detection method of cadmium in rice stems based on laser induced breakdown spectroscopy (LIBS) and chemometrics. The wavelet transform (WT), area normalization and median absolute deviation (MAD) were used to preprocess raw spectra to improve spectral stability. Principal component analysis (PCA) was used for cluster analysis. The classification models were established to distinguish cadmium stress degree of stems, of which extreme learning machine (ELM) had the best effect, with 91.11% of calibration accuracy and 93.33% of prediction accuracy. In addition, multivariate models were established for quantitative detection of cadmium. It can be found that ELM model had the best prediction effects with prediction correlation coefficient of 0.995. The results show that LIBS provides an effective method for detection of cadmium in rice stems. The combination of LIBS technology and chemometrics can quickly detect the presence of cadmium in rice stems, and accurately realize qualitative and quantitative analysis of cadmium, which could be of great significance to promote the development of new energy industry.

Keywords: rice stem, cadmium, laser induced breakdown spectroscopy, chemometrics method, biomass resource

INTRODUCTION

Rice straw is an important secondary product of rice (Sepaskhah and Yousofi-Falakdehi, 2009; Zaima et al., 2010). For every kilogram of grain harvested, the yield of straw will increase by 1.0–1.5 kilograms. Straw has the advantages of abundant raw materials, low price, high combustion calorific value (about 50% of standard coal) (Jenkins et al., 1998; Lim et al., 2012; Yin et al., 2013). It is rich in nutrients such as mineral elements (N, P, K, Ca, Mg, and Si), plant fiber (cellulose, hemicellulose, and lignin) and protein, which can be used as fuel, feed, fertilizer, base material and industrial raw materials (Buzarovska et al., 2008; Abraham et al., 2016; Ostos-Garrido et al., 2019; Logeswaran et al., 2020). The scientific use of straw can not only effectively alleviate the current

situation of energy supply shortage, but also solve environmental pollution problems (Matsumura et al., 2005; Hernández et al., 2019). However, rice plants have strong absorption properties for cadmium (Liu et al., 2019). When environment is contaminated with heavy metals, it will harm growth of plants and accumulate in the plants (Xue et al., 2017; Cheng et al., 2018; Han et al., 2018). Studies have shown that in the cadmium absorption process of rice, cadmium is more easily transferred from roots into stem (Xie et al., 2015). Most of the cadmium absorbed by rice roots is concentrated in the stem, which is an important factor causing heavy metal pollution of straw biomass resources. Previous studies have shown that the presence of cadmium can affect the performance of straw as biomass (Narodoslawsky and Obernberger, 1996; Kirkelund et al., 2013; Yang et al., 2015). During the anaerobic digestion of straw, cadmium will be released into biogas slurry, which inhibits microbial activity and thus hinders the biogas production (Nzihou and Stanmore, 2013). During the combustion of biomass, organic matter is decomposed, while heavy metals only be partially or fully gasified at high temperatures, accumulated in ash residue or released into the environment with flying dust, forming secondary pollution (Fernandez et al., 1992; Lu et al., 2012; Delplanque et al., 2013; Sánchez et al., 2015). The presence of heavy metals in straw is likely to be a factor limiting the function as biomass (Sas-Nowosielska et al., 2004; Laval-Gilly et al., 2017). Therefore, rapid detection of cadmium in rice straw is of great significance to promote the development of straw industry and new energy industry.

Commonly used heavy metal detection methods include atomic absorption spectroscopy (AAS), atomic fluorescence spectroscopy (AFS), X-ray fluorescence spectroscopy (XRFS), and inductively coupled plasma emission spectroscopy (ICP-OES) (Fortes et al., 2013; Kim et al., 2013). Traditional heavy metal methods require sampling, pretreatment and laboratory chemical analysis, which are tedious, costly and time-consuming (Rai and Rai, 2008; Gondal et al., 2010). Laser induced breakdown spectroscopy (LIBS) has been widely applied in the fields of solid, liquid and gas as a new method of material element analysis since it was proposed in 1962 (Kim et al., 2013; van Maarschalkerweerd and Husted, 2015). LIBS can quickly obtain information on sample composition and elements content in a short time. Compared with other detection technologies, LIBS has many advantages, such as less sample required, no complex pretreatment, multi-element joint measurement, and fast implementation (Gondal et al., 2009). At present, some studies have focused on the application of LIBS to detect pollution of heavy metals. Zhao et al. (2019) studied the quantitative analysis of Pb in soil and demonstrated that dual-pulse laser induced breakdown spectroscopy (DP-LIBS) can be applied as an efficient spectroscopic tool to improve the quantitative analysis of Pb heavy metal in soil. Rehan et al. (2020) applied LIBS to estimate the amount of toxic heavy metals (Pb, Cr, Ni) in different brands of face foundation powders. Peng et al. (2019a) used collinear DP-LIBS to achieve determination of chromium content in rice leaves. However, as far as we know, LIBS has not been used to detect heavy metals in rice straws. Therefore, the detection of heavy metals in straw based on LIBS technology is of great significance for the utilization of straw as a biomass.

Based on the purpose of preventing potential risk of heavy metals in biomass, this paper selected rice stem as the research object, and introduced a rapid detection method of cadmium in rice stem of LIBS technology. The specific objectives of this paper are as follows: (1) to improve spectral stability with pretreatment methods of wavelet transform (WT) and median absolute deviation (MAD); (2) to visualize the distribution of stems with different cadmium by principal component analysis (PCA); (3) to establish classification models for fast discrimination of cadmium stress degree in rice stems; and (4) to establish multivariate models for quantitative detection of cadmium in rice stems.

MATERIALS AND METHODS

Materials

In the experiment, 10 pots of rice were cultivated to obtain rice stems with different cadmium concentrations. The rice variety selected in the experiment was Xiushui 134, which was a single-season conventional late japonica rice widely cultivated in Zhejiang Province, China. During rice cultivation, the international rice nutrient solution formula was adopted, and the pH value was set as 5.3–5.6. The rice was placed in an artificial climate chamber. The cultivation parameters were as follows: the duration of day mode was 14 h, the temperature was 30°C, the relative humidity was 85%, and the light intensity was 225 $\mu mol \cdot m^{-2} \cdot s^{-1}$; the duration of night mode was 10 h, the temperature was 22°C, and the relative humidity was 85%. When entering the tillering stage, 10 pots of rice were divided into five groups of two pots each. $CdCl_2$ was used to prepare cadmium solutions with different concentrations. Based on references and experimental experience, Cd^{2+} concentrations were, respectively, set to 0 (CK), 5, 25, 50, and 100 μM. $CdCl_2$ solution with corresponding concentration was added to nutrient solution (Peng et al., 2019b).

Rice plants with similar growth were selected from each group, at 10, 20, and 30 days under heavy metal, and stems were collected as samples. A total of 15 rice stem samples with different cadmium concentrations were obtained, namely Day 10-CK, Day 10–5, Day 10–25, Day 10–50, Day 10–100, Day 20–CK, Day 20–5, Day 20–25, Day 20–50, Day 20–100, Day 30-CK, Day 30–5, Day 30–25, Day 30–50, and Day 30–100. To remove impurities and metal ions attached to the surface of sample, washed the collected sample several times with distilled water, then immersed with EDTA-2Na solution for about 60 min, and finally washed again with distilled water. Drained the water, and dried the sample in a 60°C oven for a period of time to constant weight, which was about 72 h. Put the dried stems into 5 ml centrifuge tube and add five grinding beads with the diameter of 2.8 mm. Then, put the above centrifuge tube into the grinding tool, and use a grinder (JXFSTPRP-48, Shanghaijingxin, Shanghai, China) to shake and grind for 2 min at a frequency of 60 Hz to obtain uniform stem powder. 200 mg of powder was weighed from each sample for tableting. The powder was pressed into a tablet of 10 mm × 10 mm × 1 mm by a tablet machine (YLJ-20T, Guoyan,

Hebi, China) at a pressure of 60,000 N for 20 s. Four tablets were prepared from stems of each cadmium concentration to collect LIBS spectral data.

Experimental Setup

The LIBS equipment schematic diagram in the experiment is described in **Figure 1**. Q-switched Nd:YAG nanosecond pulsed solid-state laser (Vlite-200, Beamtech Optronics, Beijing, China) was applied as laser sources to produce laser pulses with a wavelength of 532 nm, an energy of 60 mJ, a pulse frequency of 1 Hz, a pulse width of 8 ns. The energy meter (StarLite, Ophir, Jerusalem, Israel) was applied to calibrate laser energy to ensure accuracy. In the optical path system, the laser pulses passed through half-wave plate, polaroid, mirrors and lens successively, and converged into a light spot with a diameter of 7 mm. To obtain better ablation, the sample position should be located 98 mm below focusing lens, where the focal length is 100 mm. Adopt the X–Y–Z movable sample stage (Zolix, Beijing, China) to realize real-time movement of samples. The laser pulses converge on the sample surface under the action of a lens, ablating the sample to generate plasma. Taking monochromator (SR-500i-A-R, Andor, Belfast, United Kingdom) as a spectrometer, which enable separate the optical signal during plasma transition process, so as to obtain high-resolution spectral information in the short wavelength band of 210.0107∼230.9990 nm. The wavelength of atomic and ion spectra is in one-to-one correspondence with specific elements, and the intensity of the spectral signal has a quantitative relationship with the content of the corresponding element (Rodrigues Romera et al., 2016). ICCD detector (iStar DH334T-18F-03, Andor, Belfast, United Kingdom) converted optical signal into electrical signal with optimized parameters (Delay = 1 μs, Gate width = 10 μs, Gain setting = 1000). The digital delay generator (DG645, Stanford Research Systems, California, United States) enables realize time series control between lasers and ICCD detectors. Digital delay generator has a delay range of 0–2,000 ns and a resolution of 5 ps. In order to obtain stable spectra, spectra were collected at 16 different locations on each sample. Use X–Y–Z movable sample stage to realize the path setting of 16 acquisition positions, in which the path is a square matrix with a side length of 6 mm. During the data acquisition process, each of the 16 locations was collected for 5 times and accumulated as spectral data, where the repetition frequency is 1 Hz, in an attempt to avoid spectral fluctuations. As a result, the spectral data of each rice stem was the average of 80 spectra at 16 locations. It takes less than a minute to collect 80 spectral data from a sample.

Standard Method for Cadmium Content Determination

Taking inductively coupled plasma optical emission spectrometry (ICP-OES) as the standard method to detect the reference content of cadmium in rice stem. The stem tablets from which spectra have been collected were reground into powder. Weigh 100 mg samples from each tablet into the digestion tube, add 5 mL of nitric acid and 1 mL of hydrogen peroxide, tighten the digestion tube, and place them in a microwave digestion apparatus (MARS 6, CEM, Matthews, United States) for digestion. After digestion, transfer the digestive solution to a centrifuge tube, dilute with deionized water to the scale line, and shake it evenly. Finally, cadmium content of the solution was measured by ICP-OES (Optima 8000, PerkinElmer, Waltham, United States). In addition, the solution of standard sample of citrus leaves (GBW10020, Beijing, China) and blank sample were prepared simultaneously according to the above conditions. In the experiment, it took about 150 min to detect the cadmium content in a stem sample using ICP-OES. As described in **Figure 2** that under the same time conditions, as the cadmium concentration increased, the average cadmium content in the corresponding group also increased successively. Among them, the cadmium content of two groups of stem samples was very close, which were group of Day 20–5 and Day 30–5, and group of Day 20–25 and Day 30–25, respectively. The average cadmium content of rice stems from Day 30–100 was 702.25 mg/kg, reaching the highest accumulation of cadmium under experimental conditions. ICP-OES detection value was used as a reference for fast detection of heavy metal content in quantitative analysis.

Data Processing

Rice stems are heterogeneous substance, and their spectra are extremely complex, containing a lot of redundant information (Lu et al., 2019). The LIBS detection of stem elements is usually accompanied with matrix effect (Markiewicz-Keszycka et al., 2017). To improve the analysis accuracy, it is necessary to adopt suitable methods to deal with the spectrum. The data processing process in this paper consists of data pretreatment, qualitative analysis and quantitative analysis, as described in **Figure 3**.

Data Pretreatment

During the laser ablation process, the parameters of the instrument and the physical characteristics of the sample itself will cause spectral fluctuations and outliers. Pretreatment methods were used to improve the stability of detection signal. WT was adopted to denoise LIBS spectral data, in which db4 of Daubechies wavelet was used to decompose signals in three layers. Area normalization was used to reduce point-to-point fluctuations between single spectra. The MAD was used to eliminate abnormal spectra due to the randomness of the plasma transition (Peng et al., 2019a).

Chemometrics Method

Chemometrics is a new branch of chemistry. It applies knowledge such as mathematics, statistics and computer to scientifically design experiments, select the optimal measurement method, obtain the most effective characteristic data, and extract the information about substances to the maximum extent (Zhang et al., 2018). Due to the influence of laser energy fluctuation, sample non-uniformity and matrix effects, complex LIBS spectra are generated. Considering the complexity of data, chemometrics methods are often used in combination with LIBS spectroscopy to improve the stability and reliability of real-time analysis (Lang et al., 2018). In order to accurately detect cadmium in

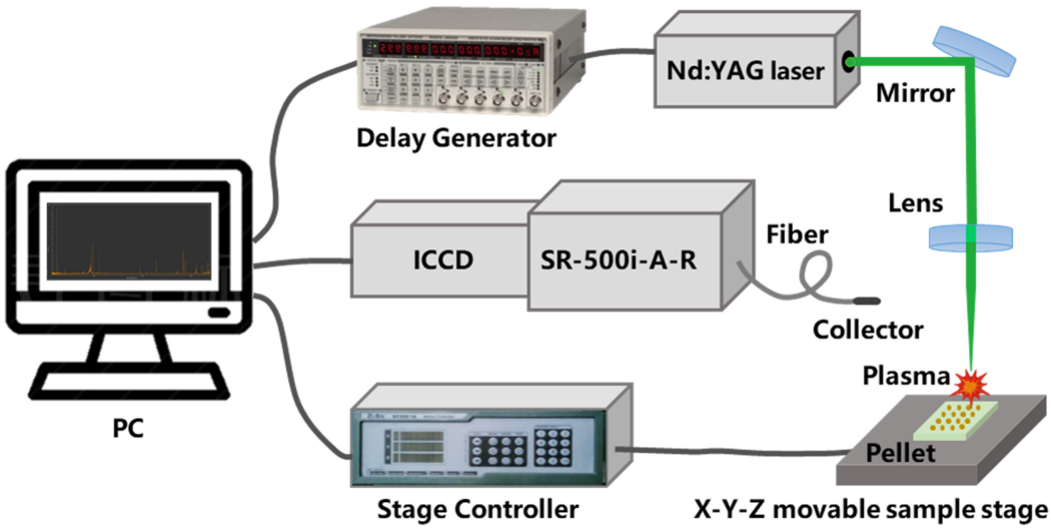

FIGURE 1 | Structural diagram of LIBS equipment.

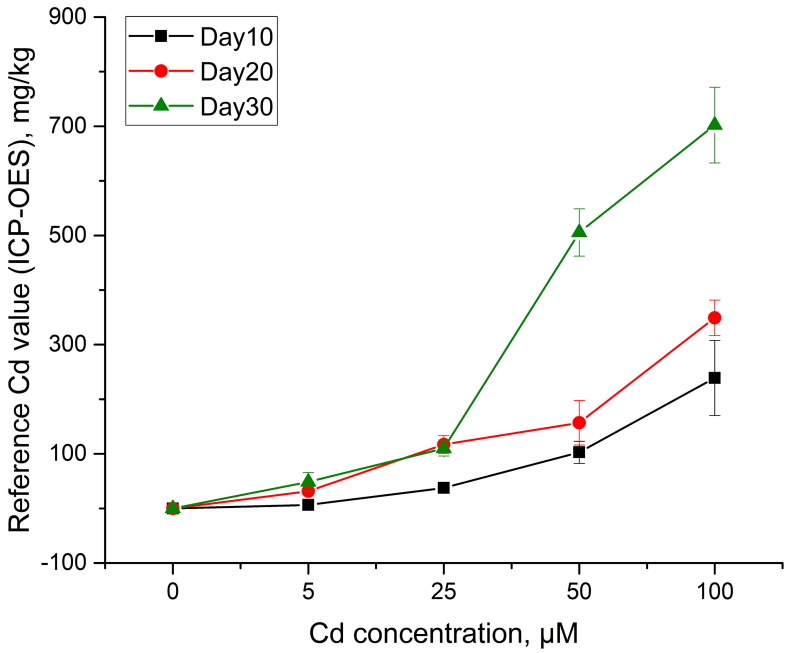

FIGURE 2 | The reference content of cadmium in rice stem obtained by ICP-OES.

rice stem, the chemometrics method and LIBS technology were combined for qualitative and quantitative analysis. In order to determine the optimal effect of discriminant analysis and quantitative detection, a variety of modeling methods were used for comparison in the paper.

Principal component analysis is a data dimension reduction method that converts all raw variables into a few unrelated variables. Each principal component is a linear combination of raw variables, which can reduce the complexity of the data and find the most useful functions, with low information loss (Zhan et al., 2015). When using PCA for qualitative analysis, calculate the contribution rate of principal components, and the distribution map of the principal component in the sample set is obtained for cluster analysis (Moncayo et al., 2017). K-nearest neighbours (KNN) is a pattern recognition algorithm (Godoi et al., 2011). The distance between the data to be tested and the data in the training set are sorted in increasing order. Then the K samples closest to the sample to be tested are selected to determine the frequency of occurrence of the category of the K samples. The category with the highest frequency among

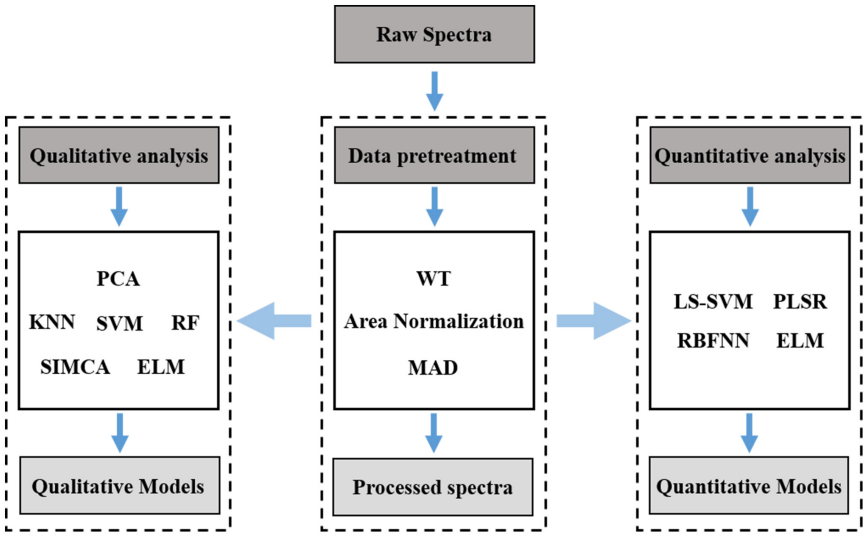

FIGURE 3 | Data processing flow.

the samples is taken as the category of test data (Ghasemi-Varnamkhasti and Forina, 2014). Support vector machine (SVM) has advantages in solving the classification of small samples, non-linear and high-dimensional data, and it belongs to a supervised learning method (Fei, 2010). Due to the simplicity of RBF function and its ability to solve complex problems, RBF function was chosen as the kernel function of SVM model for classification. Soft independent modeling of class analogy (SIMCA) is a supervised pattern recognition method based on PCA (Basri et al., 2017). It establishes a PCA model for each sample category, and projects unknown samples according to each principal component model for discriminant analysis (Ye et al., 2008). SIMCA was used for classification of rice stems. Random forest can learn the mapping relationship between features and labels from the data, which belongs to the category of supervised learning. It is a kind of classifier composed of several decision trees, the output category of which depends on the category output mode of an individual tree, and it is a typical representative of the strong classifier composed of several weak classifiers (Xin et al., 2012). Extreme learning machine (ELM) is an effective learning algorithm for single-hidden layer feedforward neural network (SLFNN) (Feng et al., 2019). ELM can randomly initialize the input weight and biases of SLFNN and obtain the corresponding output weight (Chen et al., 2012). Traditional feedforward neural network takes a long time to train and often gets into the local minimum point, while ELM method can solve the above problems with its rapid learning speed and good generalization performance. We used ELM to distinguish the cadmium pollution and to detect the cadmium content. As an improved algorithm of SVM, least square support vector machine (LS-SVM) replaces the non-equal constraints of SVM optimization problem with equality constraints, transforms the solution of SVM into the solution of linear equations, which improves the efficiency and reduces the difficulty (Pierna et al., 2011). RBF function is used as the kernel function of LS-SVM, and two parameters, sig2 and gam, are involved in the model training process. Among them, sig2 is the parameter kernel function, here is the bandwidth in the case of RBF, and gam is the regularization parameter, which determines the trade-off between the minimum model complexity and the minimum training error. Partial least squares (PLS) provides a multivariable modeling method (Maquina et al., 2019). It can effectively deal with multicollinear problems and establish linear regression models. The main idea of partial least square method is to project the original variables into mutually orthogonal dimensions through linear variables, producing latent variable (LV) (Musingarabwi et al., 2016). In the PLS model, latent variables (LVs) can explain most of the variables in the sample. It is necessary to select the appropriate number of LVs to obtain the best effect. In this paper, cross-validation method was used to select the optimal number of LVs. In order to avoid overfitting, the maximum number of variables was determined in one-tenth of the total number of samples (Gowen et al., 2011; Goto et al., 2015). Using the measured values of the sample to assign values to the Y matrix, PLS can be used as partial least squares regression (PLSR) for regression analysis (Peng et al., 2019b). Radial basis function neural network (RBFNN) is a feedforward network with a single hidden layer based on function approximation, in which the first layer is the input layer, the second is the hidden layer, the third is the output layer (Feng et al., 2018). RBFNN has been widely used for its simple training structure, fast convergence speed, strong generalization ability and arbitrary approximation.

RESULTS AND DISCUSSION

LIBS Spectral Analysis

The representative LIBS spectra of rice stems under different cadmium stress are illustrated in **Figure 4**, where (A) is the average raw spectra and (B) is the average spectra after

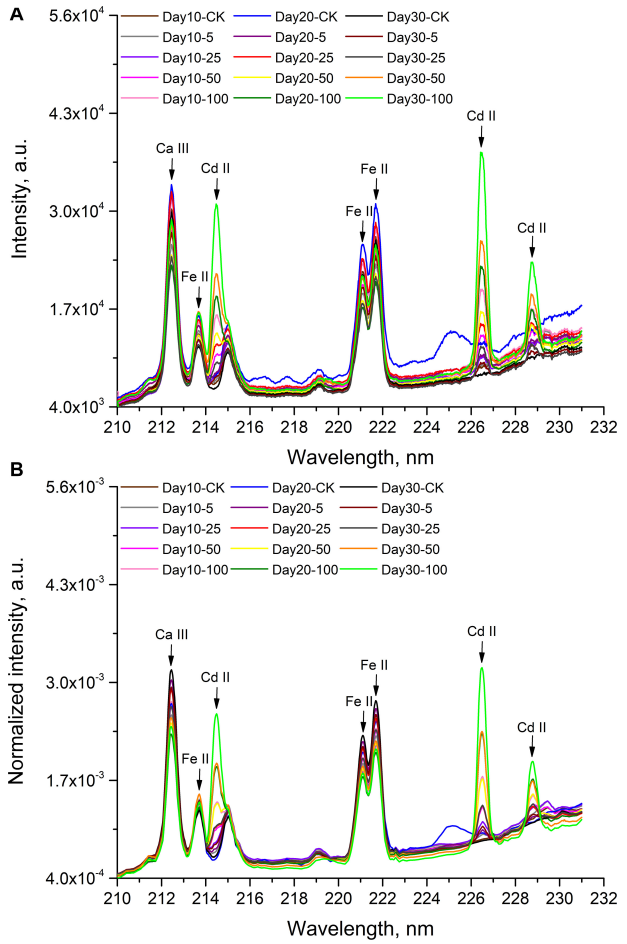

FIGURE 4 | LIBS spectra of rice stem tablets with different Cd concentrations in the range of 210.01∼231.00 nm. **(A)** average raw spectra; **(B)** average spectra after pretreatment.

pretreatment. It can be seen from the **Figure 4A** that rice stems with different cadmium concentrations had similar emission lines, which are mainly related to organic compounds, nutrients and cadmium. The pretreatment methods were used to remove noise, fluctuations and outliers from the raw spectrum. In the spectral range of 210.01∼231.00 nm, three cadmium emission lines can be clearly observed with reference to the atomic spectra database of National Institute of Standards and Technology (NIST), namely ionic spectral lines Cd II 214.44 and Cd II 226.50 nm, and atomic spectral lines Cd I 228.80 nm. We also found that emission lines Ca III 212.30, Fe II 213.77, Fe II 221.10, and Fe II 221.71 nm emerged near the emission lines of cadmium. The spectral characteristics of Ca, Fe, and Cd in different samples are significantly different, as described in **Figure 4B**. In all samples, Day 30–100 had the highest spectral line intensity at three characteristic lines of cadmium including 214.44, 226.50, and 228.80 nm. It was consistent with the results in ICP-OES, which showed that the sample with the largest accumulation of cadmium was Day 30–100. Although it can be observed that cadmium-contaminated stem samples have different spectral line characteristics, it is not a straightforward process to complete accurate and rapid sample classification based on the spectral peak position or peak intensity of the elements. For specific elements, the characteristic spectral lines of the detected element may be different due to the matrix effect. In addition, the cadmium content of some samples from 15 different stems (Day 10–5, Day 20–5, and Day 30–5, Day 20–25 and Day 30–25) is very close, which will interfere with the classification effect. Multiple emission lines need to be analyzed in a more efficient way to accurately classify stem samples. Therefore, it was necessary to use chemometrics to find differences and achieve good identification.

Cluster Analysis

We used PCA to classify different stems, and visualize the distribution of them in the principal component (PC) scores scatter plots. PCA analysis was performed on the four group samples of Day 10, Day 20, and Day 30 and all samples (Day 10, Day 20, and Day 30). **Figure 5** shows the PCA visualization of rice stems at CK, 5, 25, 50, and 100 μM concentrations under different stress days. Each point in the 3D scatter plot represented a sample, which can visually show the clustering effect. For the samples of Day 10, Day 20, and Day 30 and all samples, the total variance of the first three principal components was, respectively, 99.53, 99.66, 99.89, and 99.07%.

It can be seen that the clustering effects of rice stem scatters vary in different days. For Day 10 samples, five clusters were easily found in the PC1-PC2-PC3 space. However, the clustering effect of stems with different cadmium concentrations was different. As presented in **Figure 5A**, for CK, 5, 25 and 100 μM stem samples, each type tended to cluster together, which made these type of samples easier to distinguish from other samples. However, the 50 μM samples was close to the adjacent concentration gradient samples (25, 100 μM), which made it difficult to get good seperation between three concentration samples (25, 50, and 100 μM). For Day 20 and Day 30 samples, as depicted in **Figures 5B,C**, it was clearly that five clusters were found in the spaces. Rice stems were more likely to cluster together, under the same stress concentration. The five rice stem samples can be well separated from each other. In **Figure 5C**, a sample of 25 μM tended to cluster with 5 μM samples. For all the samples (Day 10, Day 20, and Day 30), PCA analysis was performed on fifteen different rice stems, as shown in **Figure 5D**. Fifteen clusters were obviously found in the spaces. However, the 15 types of samples tend to cluster together more than in previous cases. This is because under the combined effect of time and concentration, the cadmium content of two groups of stem samples was very close, which were Day 20–5 and Day 30–5, and Day 20–25 and Day 30–25, respectively. We can see from **Figure 5D**, 5 μM of 20 day and 5 μM of 30 day were clustered together, and 25 μM of 20 day and 25 μM of 30 day were also clustered, although the two did not belong to the same type of stress gradient.

In this paper, PCA was carried out on the stems of different rice stems, and sample scatter plots were made in three-dimensional space to visualize clustering effect of stems contaminated with different cadmium. For the stems of Day

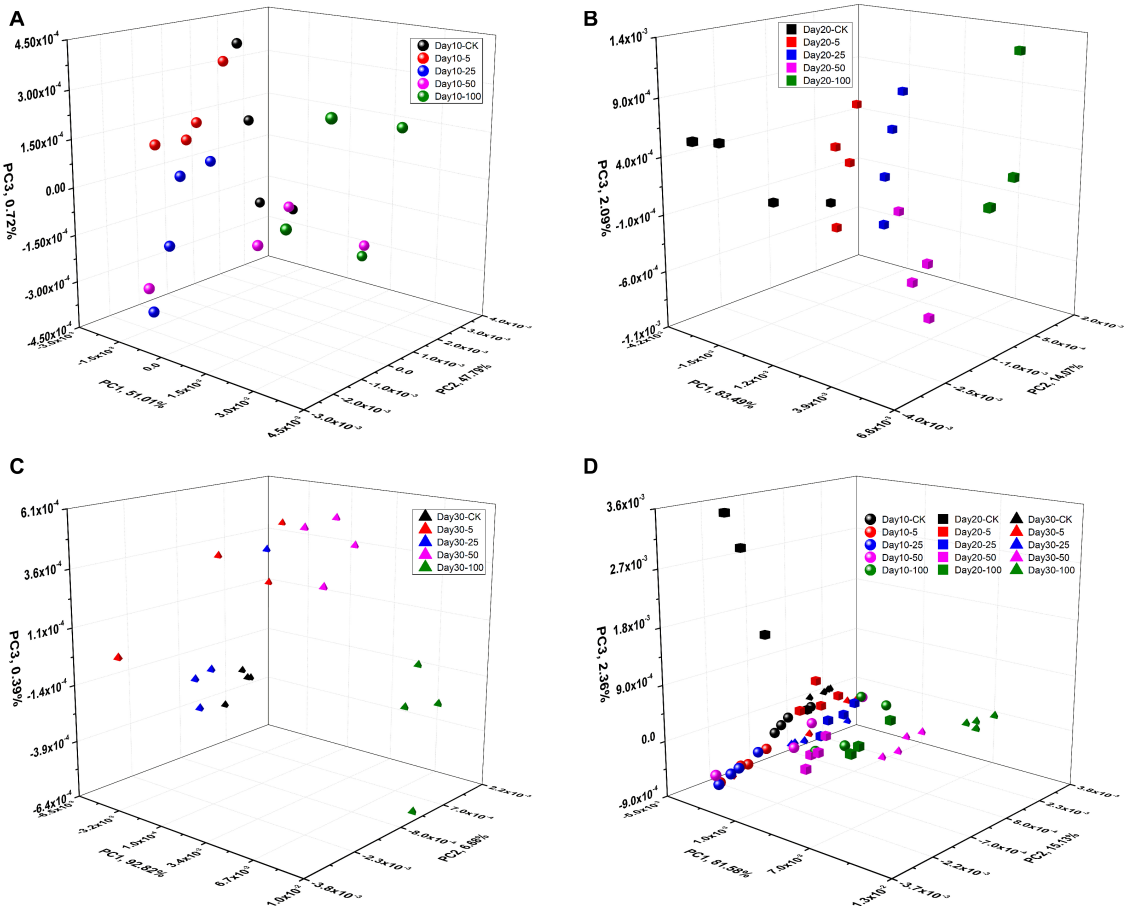

FIGURE 5 | PC scores scatter plots of rice stem tablets. **(A)** PC1-PC2-PC3 for Day 10; **(B)** PC1-PC2-PC3 for Day 20; **(C)** PC1-PC2-PC3 for Day 30; **(D)** PC1-PC2-PC3 for all samples.

10, Day 20, and Day 30, PCA analysis can achieve better clustering effect. But for all samples (Day 10, Day 20, and Day 30), 15 different stems cannot be distinguished well because samples of different gradients had similar cadmium concentrations. In addition, PCA is a clustering analysis method, which can clearly visualize the clustering effect of samples and cannot provide accurate discriminant results. Therefore, supervised multivariate analysis was used to predict contamination degree and content of cadmium in the next section.

Rapid Discrimination of 15 Cadmium Stress

The spectra of different rice stems after pretreatment were used to establish classification models including KNN, SVM, SIMCA, RF, and ELM to predict cadmium content. Before modeling, sample division was performed. For every four samples, the first three were used for calibration and the remaining one for prediction. Thus, 45 tablets were selected to calibrate model and 15 tablets were used to predict sample properties, with a ratio of 3:1. For the discriminant analysis model, accuracy was used as a performance indicator to evaluate the model, that was, the proportion of the number of correctly classified samples to the total number of samples.

The discriminant models results of rice stems with 15 different cadmium concentrations are shown in the **Table 1**. The ability of different chemometrics models to discriminate the cadmium

TABLE 1 | Results for cadmium stress discrimination in rice stems.

Model	Parameter[a]	Accuracy of calibration set	Accuracy of prediction set
KNN	3	75.56%	53.33%
SVM	(1.000, 1,000)	91.11%	73.33%
SIMCA	(2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2)	95.56%	73.33%
RF	(151,7)	100%	73.33%
ELM	12	91.11%	93.33%

[a] *Parameters of different models: the k-value for KNN, different penalty parameters (c) and kernel function parameters (g) for SVM, the number of principal components (PCs) of SIMCA, number of trees in the forest and nodes per tree for RF, and the number of hidden nodes of ELM.*

stress of stems was compared. The discrimination effect of KNN is poor, and the accuracy of the calibration set and the prediction set were 75.56 and 53.33%, respectively. Compared with KNN, the discriminative effect of SVM, SIMCA and RF has improved. The accuracy of calibration set was 91.11, 95.56, and 100%, respectively, and the accuracy of their prediction set was 73.33%. ELM has the best discriminating effect, with 91.11% of calibration accuracy and 93.33% of prediction accuracy. Comparing the ability of different models to discriminate the cadmium concentration of rice stems, we can find that ELM had the best effect, in which the accuracy of calibration and prediction were both more than 90%, followed by SVM, SIMCA and RF models, in which the calibration accuracy of models all exceeded 90% while the prediction efficiency was only 73.33%, and KNN had the worst discrimination effect, in which the accuracy of correction and prediction did not exceed 80%. Therefore, compared with other methods, ELM has higher discriminant accuracy, which is more advantageous in the discriminant analysis of cadmium pollution degree in rice stems. It takes no more than 5 min to classify a stem sample based on the above method. LIBS method can quickly distinguish 15 cadmium-contaminated stems. These results indicate that LIBS combined

TABLE 2 | Multivariate models results for cadmium concentrations in stems.

Model	Parameter[a]	Accuracy of calibration set		Accuracy of prediction set	
		R_c	RMSEC (mg/kg)	R_p	RMSEP (mg/kg)
LS-SVM	$(5.492 \times 10^3, 3.712 \times 10^3)$	0.999	9.42	0.976	46.47
PLSR	4	0.972	47.80	0.991	35.60
RBFNN	938	0.985	35.31	0.990	37.05
ELM	11	0.981	39.80	0.995	28.96

[a]*Parameters for models: the bandwidth of kernel function (sig2) and the trade-off between the minimum model complexity and the minimum training error (gam) of LS-SVM, optimal number of latent variables (LVs) of PLSR, spread coefficient of RBFNN, the number of hidden nodes of ELM.*

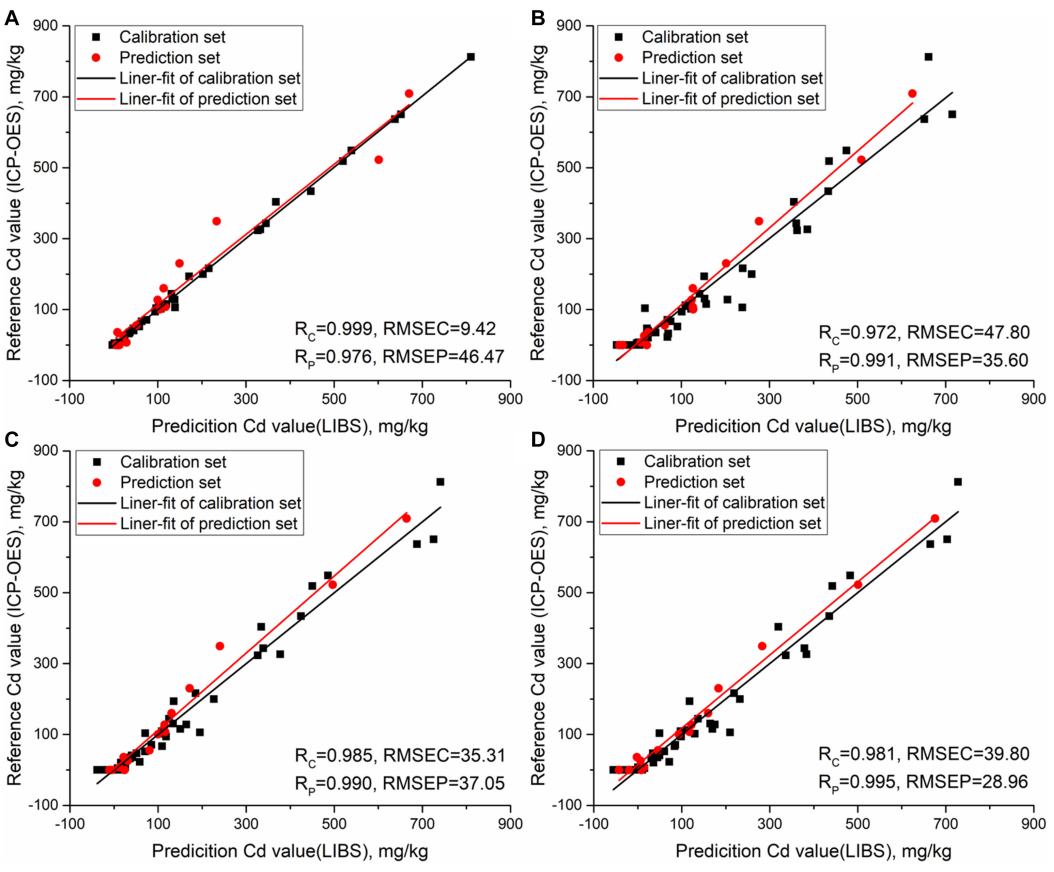

FIGURE 6 | Relationship between reference cadmium value and prediction cadmium value based on LIBS in rice stems under different models. **(A)** LS-SVM model; **(B)** PLSR model; **(C)** RBFNN model; **(D)** ELM model.

with ELM can quickly detect cadmium in stems, and accurately distinguish the degree of cadmium pollution. The combination of chemometrics and LIBS provides a fast method for discriminant analysis of cadmium content in rice stems.

Quantitative Detection of Cadmium Content

Multivariate models such as LS-SVM, PLSR, RBFNN, and ELM were established to complete the quantitative detection of cadmium in rice stems based on LIBS. The LIBS spectra ranged from 210.01 to 231.00 nm and contained 1,024 variables. 45 samples were selected as the calibration set, and 15 samples were used as the prediction set. Using correlation coefficient (R) and root mean square error (RMSE) as indicators to evaluate model performance. Multivariate models results of cadmium concentration in rice stems based on PLSR, LS-SVM, RBFNN, and ELM are shown in **Table 2**. Obviously, the four models all showed good results in predicting cadmium content in rice stems. For LS-SVM models, the optimal parameters were determined based on grid-search procedure, with sig2 and gam of 5.492×10^3 and 3.712×10^3. LS-SVM models achieved good performance, with R_c of 0.999, RMSEC of 9.42 mg/kg, R_p of 0.976 and RMSEP of 46.47 mg/kg. In PLSR modeling process, leave-one-out cross validation was used to select LVs to determine the model, and LVs was 4. R_c reached 0.972 and RMSEC was 47.80 mg/kg, R_p reached 0.991 and RMSEP was 35.60 mg/kg. The spread coefficient of RBFNN was optimized, which ranges from 1 to 1000, with a step size of 1. The number of hidden nodes of ELM was optimized, ranging from 1 to 45, with a step size of 1. Compared with LS-SVM and PLSR, RBFNN and ELM had better prediction effects, in which R_c and R_p of the above two models both exceeded 0.98.

Figure 6 shows the relationship between reference cadmium value and prediction cadmium value based on LIBS in rice stems under different models. We can observe that the data of calibration and prediction set in the four models are well fitted. For ELM, the model had the best correlation and lowest prediction error, where R_p is 0.995 and the RMSEP is 28.96 mg/kg in prediction set. The quantitative determination of cadmium in a stem sample takes no more than 5 min. The results showed that multivariate analysis with ELM method can realize the fast detection of cadmium in stem more effectively.

CONCLUSION

In this paper, the combination of LIBS and chemometrics method successfully achieved the discriminant analysis and quantitative detection of rice stems with different cadmium concentrations. With the help of WT, area normalization and MAD, the noise, fluctuation and outliers in the raw spectrum were improved. PCA method was applied for cluster analysis of 15 different cadmium stresses rice stems, and it visualized the distribution of different samples in scores scatter plots. The classification models including KNN, SVM, SIMCA, RF, and ELM were established to distinguish stems with different cadmium stress degree. Compared to other models, ELM had the best discriminating effect, with 91.11% of calibration accuracy and 93.33% of prediction accuracy. The results indicate that LIBS combined with ELM can quickly detect cadmium in rice stems, and accurately distinguish different degree of cadmium pollution. To complete the evaluation of LIBS on the ability of quantitative detection of cadmium in rice stems, multivariate analysis methods such as PLSR, LS-SVM, RBFNN, and ELM were used for modeling. The LIBS spectra ranged from 210.01 to 231.00 nm and contained 1,024 variables. It can be found that, the ELM model had the best correlation coefficient and lowest prediction error, with R_p of 0.995 and RMSEP of 28.96 mg/kg in prediction set. It indicated that multivariate analysis with ELM method can realize the fast and accurate detection of cadmium content in stem more effectively. Compared with traditional detection methods (more than 150 min), the combination of LIBS technology and ELM method (less than 5 min) greatly reduces the time required to detect heavy metals on a sample. The results show that the combination of LIBS technology and chemometrics provides significant advantages for fast and accurate detection of cadmium contamination degree and cadmium content in rice stems. The method can timely diagnose the straw containing cadmium, prevent the risk of heavy metals in the straw, ensure the safety of the straw as a clean energy, and improve utilization rate of energy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

WW, WK, and FL conceived and designed the experiments. WW, TS, ZM, and YL performed the experiments. WW, TS, WZ, and YL contributed to data analysis. YH, FL, and YL contributed to reagents, materials, and analysis tools. WW, TS, and FL wrote the whole manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research was supported by Natural Science Foundation of China (31701320), Science and Technology Department of Zhejiang Province (2020C02016), and Synergistic Innovation Center of Jiangsu Modern Agricultural Equipment and Technology (4091600021).

ACKNOWLEDGMENTS

We thank Jiming Xu in College of Life Science, Zhejiang University for providing ICP-OES assay.

REFERENCES

Abraham, A., Mathew, A. K., Sindhu, R., Pandey, A., and Binod, P. (2016). Potential of rice straw for bio-refining: an overview. *Bioresour. Technol.* 215, 29–36. doi: 10.1016/j.biortech.2016.04.011

Basri, K. N., Hussain, M. N., Bakar, J., Sharif, Z., Khir, M. F. A., and Zoolfakar, A. S. (2017). Classification and quantification of palm oil adulteration via portable NIR spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectroscopy* 173, 335–342. doi: 10.1016/j.saa.2016.09.028

Buzarovska, A., Bogoeva-Gaceva, G., Grozdanov, A., Avella, M., Gentile, G., and Errico, M. (2008). Potential use of rice straw as filler in eco-composite materials. *Aust. J. Crop Sci.* 1, 37–42.

Chen, Q. S., Ding, J., Cai, J. R., and Zhao, J. W. (2012). Rapid measurement of total acid content (TAC) in vinegar using near infrared spectroscopy based on efficient variables selection algorithm and nonlinear regression tools. *Food Chem.* 135, 590–595. doi: 10.1016/j.foodchem.2012.05.011

Cheng, S., Yu, H., Hu, M., Wu, Y., Cheng, L., Cai, Q., et al. (2018). Miscanthus accessions distinctively accumulate cadmium for largely enhanced biomass enzymatic saccharification by increasing hemicellulose and pectin and reducing cellulose CrI and DP. *Bioresour. Technol.* 263, 67–74. doi: 10.1016/j.biortech.2018.04.031

Delplanque, M., Collet, S., Del Gratta, F., Schnuriger, B., Gaucher, R., Robinson, B., et al. (2013). Combustion of Salix used for phytoextraction: the fate of metals and viability of the processes. *Biomass Bioenergy* 49, 160–170. doi: 10.1016/j.biombioe.2012.12.026

Fei, S. W. (2010). Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine. *Expert Syst. Applications* 37, 6748–6752. doi: 10.1016/j.eswa.2010.02.126

Feng, L., Zhu, S. S., Chen, S. S., Bao, Y. D., and He, Y. (2019). Combining fourier transform mid-infrared spectroscopy with chemometric methods to detect adulterations in milk powder. *Sensors* 19:14. doi: 10.3390/s19132934

Feng, L., Zhu, S. S., Zhang, C., Bao, Y. D., Gao, P., and He, Y. (2018). Variety identification of raisins using near-infrared hyperspectral imaging. *Molecules* 23:15. doi: 10.3390/molecules23112907

Fernandez, M. A., Martinez, L., Segarra, M., Garcia, J. C., and Espiell, F. (1992). Behavior of heavy metals in the combustion gases of urban waste incinerators. *Environ. Sci. Technol.* 26, 1040–1047. doi: 10.1021/es00029a026

Fortes, F. J., Moros, J., Lucena, P., Cabalin, L. M., and Laserna, J. J. (2013). Laser-induced breakdown spectroscopy. *Anal. Chem.* 85, 640–669. doi: 10.1021/ac303220r

Ghasemi-Varnamkhasti, M., and Forina, M. (2014). NIR spectroscopy coupled with multivariate computational tools for qualitative characterization of the aging of beer. *Comp. Electronics Agric.* 100, 34–40. doi: 10.1016/j.compag.2013.10.001

Godoi, Q., Leme, F. O., Trevizan, L. C., Pereira, E. R., Rufini, I. A., Santos, D., et al. (2011). Laser-induced breakdown spectroscopy and chemometrics for classification of toys relying on toxic elements. *Spectrochim. Acta Part B Atomic Spectroscopy* 66, 138–143. doi: 10.1016/j.sab.2011.01.001

Gondal, M. A., Hussain, T., Yamani, Z. H., and Baig, M. A. (2009). On-line monitoring of remediation process of chromium polluted soil using LIBS. *J. Hazard. Mater.* 163, 1265–1271. doi: 10.1016/j.jhazmat.2008.07.127

Gondal, M. A., Seddigi, Z. S., Nasr, M. M., and Gondal, B. (2010). Spectroscopic detection of health hazardous contaminants in lipstick using laser induced breakdown spectroscopy. *J. Hazard. Mater.* 175, 726–732. doi: 10.1016/j.jhazmat.2009.10.069

Goto, N., Bazar, G., Kovacs, Z., Kunisada, M., Morita, H., Kizaki, S., et al. (2015). Detection of UV-induced cyclobutane pyrimidine dimers by near-infrared spectroscopy and aquaphotomics. *Sci. Rep.* 5:11808. doi: 10.1038/srep11808

Gowen, A. A., Downey, G., Esquerre, C., and O'donnell, C. P. (2011). Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. *J. Chemometrics* 25, 375–381. doi: 10.1002/cem.1349

Han, X. Q., Xiao, X. Y., Guo, Z. H., Xie, Y. H., Zhu, H. W., Peng, C., et al. (2018). Release of cadmium in contaminated paddy soil amended with NPK fertilizer and lime under water management. *Ecotoxicol. Environ. Safety* 159, 38–45. doi: 10.1016/j.ecoenv.2018.04.049

Hernández, C., Escamilla-Alvarado, C., Sánchez, A., Alarcón, E., Ziarelli, F., Musule, R., et al. (2019). Wheat straw, corn stover, sugarcane, and agave biomasses: chemical properties, availability, and cellulosic-bioethanol production potential in Mexico. *Biofuels Bioproducts Biorefining Biofpr* 13, 1143–1159. doi: 10.1002/bbb.2017

Jenkins, B. M., Baxter, L. L., Miles, T. R. Jr., and Miles, T. R. (1998). Combustion properties of biomass. *Fuel Processing Technol.* 54, 17–46.

Kim, G., Kwak, J., Kim, K. R., Lee, H., Kim, K. W., Yang, H., et al. (2013). Rapid detection of soils contaminated with heavy metals and oils. by laser induced breakdown spectroscopy (LIBS). *J. Hazard. Mater.* 263, 754–760. doi: 10.1016/j.jhazmat.2013.10.041

Kirkelund, G. M., Damoe, A. J., and Ottosen, L. M. (2013). Electrodialytic removal of Cd from biomass combustion fly ash suspensions. *J. Hazard. Mater.* 25, 212–219. doi: 10.1016/j.jhazmat.2013.02.004

Lang, A., Engelberg, D., Smith, N. T., Trivedi, D., Horsfall, O., Banford, A., et al. (2018). Analysis of contaminated nuclear plant steel by laser-induced breakdown spectroscopy. *J. Hazard. Mater.* 345, 114–122. doi: 10.1016/j.jhazmat.2017.10.064

Laval-Gilly, P., Henry, S., Mazziotti, M., Bonnefoy, A., Comel, A., and Falla, J. (2017). Miscanthus x giganteus composition in metals and potassium after culture on polluted soil and its use as biofuel. *BioEnergy Res.* 10, 846–852. doi: 10.1007/s12155-017-9846-3

Lim, J. S., Manan, Z. A., Alwi, S. R. W., and Hashim, H. (2012). A review on utilisation of biomass from rice industry as a source of renewable energy. *Renewable Sustain. Energy Rev.* 16, 3084–3094. doi: 10.1016/j.rser.2012.02.051

Liu, X. Y., Chen, S. L., Chen, M. X., Zheng, G. Y., Peng, Y., Shi, X. L., et al. (2019). Association study reveals genetic loci responsible for arsenic, cadmium and lead accumulation in rice grain in contaminated farmlands. *Front. Plant Sci.* 10:16. doi: 10.3389/fpls.2019.00061

Logeswaran, J., Shamsuddin, A. H., Silitonga, A. S., and Mahlia, T. M. I. (2020). Prospect of using rice straw for power generation: a review. *Environ. Sci. Pollut. Res. Int.* 27, 25956–25969. doi: 10.1007/s11356-020-09102-7

Lu, S. Y., Du, Y. Z., Zhong, D. X., Zhao, B., Li, X. D., Xu, M. X., et al. (2012). Comparison of trace element emissions from thermal treatments of heavy metal hyperaccumulators. *Environ. Sci. Technol.* 46, 5025–5031. doi: 10.1021/es202616v

Lu, Z., Chen, X., Yao, S., Qin, H., Zhang, L., Yao, X., et al. (2019). Feasibility study of gross calorific value, carbon content, volatile matter content and ash content of solid biomass fuel using laser-induced breakdown spectroscopy. *Fuel* 258:116150. doi: 10.1016/j.fuel.2019.116150

Maquina, A. D. V., Sitoe, B. V., Buiatte, J. E., Santos, D. Q., and Neto, W. B. (2019). Quantification and classification of cotton biodiesel content in diesel blends, using mid-infrared spectroscopy and chemometric methods. *Fuel* 237, 373–379. doi: 10.1016/j.fuel.2018.10.011

Markiewicz-Keszycka, M., Cama-Moncunill, X., Casado-Gavalda, M. P., Dixit, Y., Cama-Moncunill, R., Cullen, P. J., et al. (2017). Laser-induced breakdown spectroscopy (LIBS) for food analysis: a review. *Trends Food Sci. Technol.* 65, 80–93. doi: 10.1016/j.tifs.2017.05.005

Matsumura, Y., Minowa, T., and Yamamoto, H. (2005). Amount, availability, and potential use of rice straw (agricultural residue) biomass as an energy resource in Japan. *Biomass Bioenergy* 29, 347–354. doi: 10.1016/j.biombioe.2004.06.015

Moncayo, S., Manzoor, S., Rosales, J. D., Anzano, J., and Caceres, J. O. (2017). Qualitative and quantitative analysis of milk for the detection of adulteration by Laser Induced Breakdown Spectroscopy (LIBS). *Food Chem.* 232, 322–328. doi: 10.1016/j.foodchem.2017.04.017

Musingarabwi, D. M., Nieuwoudt, H. H., Young, P. R., Eyeghe-Bickong, H. A., and Vivier, M. A. (2016). A rapid qualitative and quantitative evaluation of grape berries at various stages of development using Fourier-transform infrared spectroscopy and multivariate data analysis. *Food Chem.* 190, 253–262. doi: 10.1016/j.foodchem.2015.05.080

Narodoslawsky, M., and Obernberger, I. (1996). From waste to raw material - the route from biomass to wood ash for cadmium and other heavy metals. *J. Hazard. Mater.* 50, 157–168. doi: 10.1016/0304-3894(96)01785-2

Nzihou, A., and Stanmore, B. (2013). The fate of heavy metals during combustion and gasification of contaminated biomass-a brief review. *J. Hazard. Mater.* 25, 56–66. doi: 10.1016/j.jhazmat.2013.02.050

Ostos-Garrido, F. J., De Castro, A. I., Torres-Sanchez, J., Piston, F., and Pena, J. M. (2019). High-throughput phenotyping of bioethanol potential in cereals using UAV-based multi-spectral imagery. *Front. Plant Sci.* 10:15. doi: 10.3389/fpls.2019.00948

Peng, J., He, Y., Jiang, J., Zhao, Z., Zhou, F., and Liu, F. (2019a). High-accuracy and fast determination of chromium content in rice leaves based on collinear dual-pulse laser-induced breakdown spectroscopy and chemometric methods. *Food Chem.* 295, 327–333. doi: 10.1016/j.foodchem.2019.05.119

Peng, J., He, Y., Zhao, Z., Jiang, J., Zhou, F., Liu, F., et al. (2019b). Fast visualization of distribution of chromium in rice leaves by re-heating dual-pulse laser-induced breakdown spectroscopy and chemometric methods. *Environ. Pollut.* 252, 1125–1132. doi: 10.1016/j.envpol.2019.06.027

Pierna, J. A. F., Lecler, B., Conzen, J. P., Niemoeller, A., Baeten, V., and Dardenne, P. (2011). Comparison of various chemometric approaches for large near infrared spectroscopic data of feed and feed products. *Anal. Chim. Acta* 705, 30–34. doi: 10.1016/j.aca.2011.03.023

Rai, N. K., and Rai, A. K. (2008). LIBS - an efficient approach for the determination of Cr in industrial wastewater. *J. Hazard. Mater.* 150, 835–838. doi: 10.1016/j.jhazmat.2007.10.044

Rehan, I., Gondal, M. A., Rehan, K., and Sultana, S. (2020). Spectral diagnosis of health hazardous toxins in face foundation powders using laser induced breakdown spectroscopy and inductively coupled plasma-optical emission spectroscopy (ICP-OES). *Talanta* 217:121007. doi: 10.1016/j.talanta.2020.121007

Rodrigues Romera, J. P., Barsanelli, P. L., and Verbi Pereira, F. M. (2016). Expeditious prediction of fiber content in sugar cane: an analytical possibility with LIBS and chemometrics. *Fuel* 166, 473–476. doi: 10.1016/j.fuel.2015.11.029

Sánchez, R., Sánchez, C., Lienemann, C.-P., and Todolí, J.-L. (2015). Metal and metalloid determination in biodiesel and bioethanol. *J. Anal. Atomic Spectrometry* 30, 64–101. doi: 10.1039/c4ja00202d

Sas-Nowosielska, A., Kucharski, R., Malkowski, E., Pogrzeba, M., Kuperberg, J. M., and Krynski, K. (2004). Phytoextraction crop disposal - an unsolved problem. *Environ. Pollut.* 128, 373–379. doi: 10.1016/j.envpol.2003.09.012

Sepaskhah, A. R., and Yousofi-Falakdehi, A. (2009). Interaction between the effects of deficit irrigation and water salinity on yield and yield components of rice in pot experiment. *Plant Production Sci.* 12, 168–175. doi: 10.1626/pps.12.168

van Maarschalkerweerd, M., and Husted, S. (2015). Recent developments in fast spectroscopy for plant mineral analysis. *Front. Plant Sci.* 6:14. doi: 10.3389/fpls.2015.00169

Xie, P., Deng, J., Zhang, H., Ma, Y., Cao, D., Ma, R., et al. (2015). Effects of cadmium on bioaccumulation and biochemical stress response in rice (*Oryza sativa* L.). *Ecotoxicol. Environ. Safety* 122, 392–398. doi: 10.1016/j.ecoenv.2015.09.007

Xin, N., Gu, X. F., Wu, H., Hu, Y. Z., and Yang, Z. L. (2012). Discrimination of raw and processed *Dipsacus asperoides* by near infrared spectroscopy combined with least squares-support vector machine and random forests. *Spectrochim. Acta Part A: Mol. Biomol. Spectroscopy* 89, 18–24. doi: 10.1016/j.saa.2011.12.006

Xue, S. G., Shi, L. Z., Wu, C., Wu, H., Qin, Y. Y., Pan, W. S., et al. (2017). Cadmium, lead, and arsenic contamination in paddy soils of a mining area and their exposure effects on human HEPG2 and keratinocyte cell-lines. *Environ. Res.* 156, 23–30. doi: 10.1016/j.envres.2017.03.014

Yang, S., Zhang, C., Wang, J., and Liu, Z. (2015). Study on enzymatic hydrolysis efficiency of cellulose from rice straw with high content of cadmium. *Chem. Bioeng.* 32, 28–31. doi: 10.3969/j.issn.1672-5425.2015.06.007

Ye, X. P., Liu, L., Hayes, D., Womac, A., Hong, K. L., and Sokhansanj, S. (2008). Fast classification and compositional analysis of cornstover fractions using Fourier transform near-infrared techniques. *Bioresour. Technol.* 99, 7323–7332. doi: 10.1016/j.biortech.2007.12.063

Yin, L., Hu, T., Zeng, G., Shen, Y., Liu, Y., Wu, J., et al. (2013). Optimum conditions and effect of inducer of *Penicillium simplicissimum* biodegrading rice straw. *Chin. J. Environ. Eng.* 7, 3642–3648.

Zaima, N., Goto-Inoue, N., Hayasaka, T., and Setou, M. (2010). Application of imaging mass spectrometry for the analysis of *Oryza sativa* rice. *Rapid Commun. Mass Spectrometry* 24, 2723–2729. doi: 10.1002/rcm.4693

Zhan, H., Wu, S., Bao, R., Ge, L., and Zhao, K. (2015). Qualitative identification of crude oils from different oil fields using terahertz time-domain spectroscopy. *Fuel* 143, 189–193. doi: 10.1016/j.fuel.2014.11.047

Zhang, T. L., Tang, H. S., and Li, H. (2018). Chemometrics in laser-induced breakdown spectroscopy. *J. Chemometrics* 32, 18. doi: 10.1002/cem.2983

Zhao, S., Song, C., Gao, X., and Lin, J. (2019). Quantitative analysis of Pb in soil by femtosecond-nanosecond double-pulse laser-induced breakdown spectroscopy. *Results Phys.* 15:102736. doi: 10.1016/j.rinp.2019.102736

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Kong, Shen, Man, Zhu, He, Liu and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 December 2020
doi: 10.3389/fpls.2020.590889

Accuracy of a Smartphone-Based Object Detection Model, PlantVillage Nuru, in Identifying the Foliar Symptoms of the Viral Diseases of Cassava–CMD and CBSD

Latifa M. Mrisho[1,2], Neema A. Mbilinyi[1], Mathias Ndalahwa[1], Amanda M. Ramcharan[3], Annalyse K. Kehs[4], Peter C. McCloskey[4], Harun Murithi[5], David P. Hughes[3,6] and James P. Legg[1]**

[1] Virus and Vector Ecology Group, International Institute of Tropical Agriculture (IITA), Dar es Salaam, Tanzania, [2] Department of Molecular Biology and Biotechnology, University of Dar es Salaam, Dar es Salaam, Tanzania, [3] Bayer Crop Science, Chesterfield, MO, United States, [4] Department of Entomology, The Pennsylvania State University, University Park, PA, United States, [5] Agricultural Research Service (ARS) Research Participation Program, Oak Ridge Institute for Science and Education, Oak Ridge, TN, United States, [6] Department of Biology, The Pennsylvania State University, University Park, PA, United States

OPEN ACCESS

Edited by:
Yiannis Ampatzidis,
University of Florida, United States

Reviewed by:
Lea Hallik,
University of Tartu, Estonia
Julia Christine Meitz-Hopkins,
Stellenbosch University, South Africa
Titus Alicai,
National Agricultural Research Organisation, Uganda

***Correspondence:**
Latifa M. Mrisho
l.mrisho@cgiar.org;
lmrisho@gmail.com
James P. Legg
j.legg@cgiar.org

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: 03 August 2020
Accepted: 17 November 2020
Published: 18 December 2020

Citation:
Mrisho LM, Mbilinyi NA, Ndalahwa M, Ramcharan AM, Kehs AK, McCloskey PC, Murithi H, Hughes DP and Legg JP (2020) Accuracy of a Smartphone-Based Object Detection Model, PlantVillage Nuru, in Identifying the Foliar Symptoms of the Viral Diseases of Cassava–CMD and CBSD. Front. Plant Sci. 11:590889. doi: 10.3389/fpls.2020.590889

Nuru is a deep learning object detection model for diagnosing plant diseases and pests developed as a public good by PlantVillage (Penn State University), FAO, IITA, CIMMYT, and others. It provides a simple, inexpensive and robust means of conducting in-field diagnosis without requiring an internet connection. Diagnostic tools that do not require the internet are critical for rural settings, especially in Africa where internet penetration is very low. An investigation was conducted in East Africa to evaluate the effectiveness of Nuru as a diagnostic tool by comparing the ability of Nuru, cassava experts (researchers trained on cassava pests and diseases), agricultural extension officers and farmers to correctly identify symptoms of cassava mosaic disease (CMD), cassava brown streak disease (CBSD) and the damage caused by cassava green mites (CGM). The diagnosis capability of Nuru and that of the assessed individuals was determined by inspecting cassava plants and by using the cassava symptom recognition assessment tool (CaSRAT) to score images of cassava leaves, based on the symptoms present. Nuru could diagnose symptoms of cassava diseases at a higher accuracy (65% in 2020) than the agricultural extension agents (40–58%) and farmers (18–31%). Nuru's accuracy in diagnosing cassava disease and pest symptoms, in the field, was enhanced significantly by increasing the number of leaves assessed to six leaves per plant (74–88%). Two weeks of Nuru practical use provided a slight increase in the diagnostic skill of extension workers, suggesting that a longer duration of field experience with Nuru might result in significant improvements. Overall, these findings suggest that Nuru can be an effective tool for in-field diagnosis of cassava diseases and has the potential to be a quick and cost-effective means of disseminating knowledge from researchers to agricultural extension agents and farmers, particularly on the identification of disease symptoms and their management practices.

Keywords: cassava mosaic disease, cassava brown streak disease, e-extension services, Africa, Kenya, Tanzania, image recognition systems, mobile applications for agriculture

INTRODUCTION

The steady increase in the world population and changes in climate are adding pressure to agriculture as the need to produce more food intensifies, and pests and diseases exacerbate these threats to food production. Enhancing the management of emerging pests and diseases along with the production of climate-resilient and disease-resistant crops are some of the efforts that are being put in place to prevent the risk of hunger (Fao et al., 2018). Tools and technologies that can be used for early detection and diagnosis of crop diseases and pests are being encouraged to facilitate their management.

Information and Communication Technology (ICT) platforms in the form of the internet, call-centers and SMS have been adopted to disseminate agricultural information to farmers in several regions including Latin America, Asia, and Africa (Qiang et al., 2012; Tsan et al., 2019). These technologies have shown promising results in reducing the knowledge gap between experts and farmers by enabling transfer of information about basic skills, new technologies and production techniques (Furuholt and Matotay, 2011; Qiang et al., 2012; Misaki et al., 2018).

Developing ICT tools that are capable of identifying crop disease and pest damage poses a greater challenge due to the variability of symptoms. However, several novel techniques for disease and pest identification that use image recognition systems have been developed (Barbedo, 2014; Tuhaise et al., 2014; Pethybridge and Nelson, 2015; Prasad et al., 2016; Qin et al., 2016; Sladojevic et al., 2016; Fuentes et al., 2017; Johannes et al., 2017; Ramcharan et al., 2017). The image recognition system developed by Tuhaise et al. (2014) is one of the first systems developed for cassava. It estimates disease severity by evaluating the percentage of root necrosis attributed to cassava brown streak disease (CBSD). Even though this method is effective and might be useful for research purposes it is not practical for farmers as it requires plants to be uprooted.

Most of the image recognition systems developed for diagnosis of plant diseases provide remote, indirect detection where images of diseased plants are uploaded and analyzed prior to sending feedback to the users. Such technologies are useful to researchers but may be less effective for agricultural extension officers and farmers in areas with limited phone networks. PlantVillage Nuru and Leaf Doctor are currently the only publicly available mobile-based applications that can be used for in-field diagnosis of plant diseases (Pethybridge and Nelson, 2015; Ramcharan et al., 2017, 2019). Both applications can be downloaded, free of charge, from the Apple Store and Android PlayStore. However, PlantVillage Nuru is more geared to application in Africa where Android has a very large market share (>85%; StatCounter Global Stats, 2020). PlantVillage Nuru also provides real-time diagnosis and management advice in the absence of a mobile network, making it ideal for use in remote areas.

PlantVillage Nuru was created from a deep learning object detection model that can determine the presence of diseases and pests in plants based on foliar symptoms (Ramcharan et al., 2017, 2019). The model was trained using 2,756 images of cassava leaves with symptoms of cassava pests and diseases, including CBSD, CMD, brown leaf spot (BLS), cassava green mite (CGM), red mite (RM) as well as asymptomatic leaves. Currently Nuru is trained to identify the presence of the most important pest/disease constraints of cassava. These include the viral diseases of cassava (CMD and CBSD) and the damage caused by CGM (sample images are illustrated in **Supplementary Material A**). In addition, Nuru has been trained to detect damage caused by fall armyworm (FAW), and new extensions will provide identification for maize lethal necrosis disease and early and late blight in potato. All of these will be delivered through one integrated objected detection model that works offline. Together, CMD and CBSD are arguably the greatest causes of economic losses in cassava production in Africa (Legg et al., 2011), while FAW has been reported to cause damage in several major crops including maize, sorghum, rice and sugarcane (Abrahams et al., 2017). These diseases and pests pose a great challenge to food production since cassava and maize are two of the major staple food crops in Africa, South America, and Asia (Pariona, 2019). Moreover, cassava is also used as a source of food for animals and as a raw material for the production of starch (Hillocks et al., 2002). Africa is the largest producer of cassava in the world, accounting for more than 50% of global production (FAOSTAT, 2018), hence management of CMD and CBSD is crucial for the survival of the crop.

Both CMD and CBSD are controlled through the development and deployment of resistant varieties, as well as through the application of phytosanitary measures, including the removal of infected plants during early growth stages and the selection of healthy stems for use as planting material (Dixon et al., 2003; Thresh and Cooter, 2005; Legg et al., 2015; Kawuki et al., 2016). Effective diagnostic methods are vital if these measures are to be successfully implemented. Laboratory methods can be used to test for the presence of the viruses that cause CMD and CBSD (Hong et al., 1993; Monger et al., 2001), however these tests cannot be used at community level, and hence symptom recognition continues to be the most common way of diagnosing CMD, CBSD and CGM-damage (Sseruwagi et al., 2004). Although visual assessment of symptoms is a valuable first-line diagnostic approach for each of these major diseases/pests of cassava, it has not been widely nor effectively applied. Ineffectiveness of the visual assessment of CMD and CBSD is due to the weakly resourced nature of extension systems in most African countries which hinders knowledge transfer from researchers (via extension officers) to farmers, thereby depriving them of relevant information required for accurate diagnosis and management of diseases and pests (Daniel et al., 2013). Mobile-based apps like PlantVillage Nuru have an important potential role to play within extension systems by serving as diagnosis and training tools for extension officers and farmers. In view of the rapid penetration of mobile phone technology in developing countries (GSMA, 2019), such tools will become accessible to the majority of farmers in the near future, which means that their reach will greatly exceed that of existing extension systems.

The mobile-based technologies available for diagnosis of plant diseases are relatively new and have not been around long enough to evaluate their effectiveness in enabling farmers to improve

their disease diagnosis capability. The present study evaluates the effectiveness of PlantVillage Nuru for in-field diagnosis of the viral diseases of cassava and compares its accuracy to that of researchers, agricultural extension officers and farmers. Although the primary role of PlantVillage Nuru is for rapid pest/disease diagnosis rather than training, the teaching capability was also evaluated to provide information about the potential of the app to be used as a tool for transferring knowledge on pest and disease management from researchers to agricultural extension officers and farmers.

MATERIALS AND METHODS

PlantVillage Nuru–Cassava Model

The cassava model of PlantVillage Nuru was developed using deep convolutional neural network (CNN) algorithms trained to identify symptoms of CMD and CBSD as well as damage caused by CGM on cassava leaf images (Ramcharan et al., 2017). A total of 2,756 images of asymptomatic and symptomatic cassava leaves from different varieties of cassava plants grown at field sites in coastal Tanzania (Chambezi, Bagamoyo) were selected and annotated by cassava disease experts at the International Institute of Tropical Agriculture (IITA) (Ramcharan et al., 2017). The annotated images were for training the cassava diagnostic model and these included images with symptoms of CBSD (398 images), CMD (388 images), BLS (386 images), CGM (309 images), and RM (415 images) as well as images of asymptomatic leaves (860 images) (Ramcharan et al., 2017). The model was subsequently deployed as a mobile app (PlantVillage Nuru) in Android smartphones and then tested in fields to account for environmental factors and hence determine the best conditions that will enable the model to perform with high accuracy (Ramcharan et al., 2019).

PlantVillage Nuru works by performing real-time analysis of the image displayed on the screen when using the app. The user is directed to point the phone's camera onto a leaf that does not look healthy and ensure that the image is in focus prior to analysis. Once the image is in focus the user can start the analysis and the app will display boxes indicating the condition it has identified on individual leaflets. Once the user has finished inspecting the whole plant the app provides the user with an overall diagnosis of the condition of the plant followed by advice on management of the disease or pest it has detected. Both the disease diagnosis and advice capabilities of PlantVillage Nuru are available while offline, enabling users to get results even when in remote areas with no network services. PlantVillage Nuru is programmed in multiple languages including English and Swahili; Swahili is widely spoken throughout East Africa as well as in parts of central Africa. On-going efforts are being made to provide access in other languages through a crowd-sourcing translation tool, and a voice command functionality is being integrated into the app to give access to users who are not able to read and write.

PlantVillage Nuru was evaluated by comparing its capability to diagnose cassava diseases to that of expert researchers, other researchers, agricultural extension officers and farmers. This required experts to be highly skilled in determining the presence of cassava diseases based on foliar symptoms. The capability of expert researchers (cassava experts from IITA, Dar es Salaam, Tanzania) to identify foliar symptoms of cassava diseases was therefore determined at the outset.

Determination of the Capability of Cassava Experts to Diagnose Cassava Diseases Based on Foliar Symptoms

The ability of cassava experts to accurately identify the symptoms of CMD, CBSD and CGM damage, based on symptoms observed from the leaves, was assessed by comparing visual and molecular diagnosis (through conventional PCR or qPCR amplification of the disease-causing viruses). Two cassava experts, each with at least 3 years of work experience on cassava virus diseases, visually diagnosed 75 cassava plants of five different varieties by assessing the presence or absence of symptoms of CMD and CBSD. The experts randomly selected two leaves (one from the top and the other from the bottom part of each cassava plant) and scored the leaves based on the observed condition. The inspected leaves were sampled and used to test for the presence of the disease-causing virus by molecular diagnosis. The accuracy of expert diagnosis was determined by calculating the proportion of plants for which the symptom diagnoses of experts matched those of the virus test results.

Molecular Identification of the Viruses That Cause CMD and CBSD From Cassava Leaves

Standard protocols for laboratory detection of CMD-causing viruses (Abarshi et al., 2012) and CBSD-causing viruses (Adams et al., 2012) were used. These involved extraction of nucleic acid from the investigated cassava leaves and PCR amplification of the viruses that cause CMD and CBSD. Two leaves, one from the top and the other from the bottom of the cassava plants were sampled from the 75 cassava plants that were investigated. Symptomatic leaflets were labeled, and the leaves were dried prior to analysis.

Extraction of Nucleic Acid From Investigated Leaves

Prior the PCR amplification the cassava leaves were dried, ground and total nucleic acid was extracted using a cetyl trimethyl ammonium bromide (CTAB) extraction procedure. Briefly, 1 mL of CTAB buffer (containing 2.0% w/v CTAB, 2.0% PVP, 25 mM EDTA, 2.0 M NaCl, 100 mM Tris HCL pH 8.0, and 0.2% β-mercaptoethanol which was added immediately before extraction) was added to the dried cassava leaves. The leaves were then ground and incubated at 65°C for 15 min to lyse the cells to facilitate the separation of polysaccharides and polyphenols from the cellular material. An equal amount of chloroform:isoamyl alcohol (24:1) was added to the cell lysate to separate the nucleic acids from the cell lysate, the nucleic acid was precipitated from the solution by adding 0.6x volume of cold isopropanol prior to incubation at −20°C for 30 min followed by centrifugation at 13,000 rpm and 4°C for 10 min. The supernatant was discarded and the nucleic acid pellet was washed twice by adding 700 μL of 70% ethanol followed by vortexing and incubating at −20°C

prior to centrifugation at 13,000 rpm for 5 min. Afterwards, the ethanol solution was removed, the nucleic acid pellet was air-dried and then resuspended in 100 μL Tris-EDTA buffer (1x). The quantity and quality of the nucleic acid extracts were determined by spectrometry at 260 nm.

PCR Amplification of Viruses That Cause CMD

Nucleic acid of the virus that causes CMD in the coastal region (East African cassava mosaic virus – EACMV) was amplified using the primer pair EAB555F and EAB555R, designed to amplify a 560 bp DNA fragment as described by Ndunguru et al. (2005). About 20 ng of total nucleic acid was added to the PCR master mix containing One Taq 2x Master mix with standard buffer (M0482S, New England Biolabs) and 200 nM each of the forward and reverse primers. A negative control (no-template) and positive controls (samples obtained from symptomatic plants maintained in the screen house that had previously been tested and shown to have the virus of interest) were also included in the analysis.

PCR amplification was done using a Veriti thermocycler (Applied Biosystems) with the following cycling conditions: initial DNA denaturation at 94°C for 2 min, followed by 30 cycles of denaturation (at 94°C for 30 s), annealing (at 55°C for 30 s) and extension (at 68°C 40 s), then a final extension at 68°C for 10 min. The PCR products were analyzed by gel electrophoresis using 1% (w/v) agarose gels and 1X TAE buffer. The DNA products were stained with GelRed nucleic acid stain (Biotium, California, United States) and the gels were viewed and photographed using the Syngene GBox system (Syngene, Cambridge, United Kingdom). Samples containing DNA bands of about 560 bp were considered as EACMV positive results.

PCR Amplification of Viruses That Cause CBSD

Detection of the viruses that cause CBSD (Cassava brown streak virus and Ugandan cassava brown streak virus) was done by real-time RT-PCR (qPCR) using TaqMan chemistry and primers described by Adams et al. (2012). Four microlitres of the template nucleic acid was added into the PCR reaction mixtures containing 1x PCR buffer, 5.5 mM $MgCl_2$, 0.5 mM dNTPs, 300 nM primer, 100 nM probe, 30 nM Rox reference dye, 0.625 Units of Taq DNA polymerase and 0.4 Units of M-MLV- reverse transcriptase into a 25 μL reaction. The Taq DNA polymerase and reverse transcriptase were obtained from Life Technologies (California, United States) while all the reagents in the PCR master mix were obtained from IDT (Iowa, United States). An internal control (Cytochrome oxidase 1), a negative control (no-template), and positive controls (samples obtained from symptomatic plants maintained in screen houses that had previously been tested and shown to have the virus of interest) were also included in the analysis.

The amplification reactions were done using a Stratagene MX3000P qPCR machine (Agilent Technologies, New Jersey, United States) with the following thermo-cycling conditions: 30 min incubation at 48° for reverse transcription, initial denaturation of the cDNA at 95°C for 10 min, 40 cycles of denaturation at 95°C for 15 s and annealing and extension at 60°C for 1 min. Fluorescence data were collected during the 60°C step using Stratagene MxPro Real-time qPCR software version 4 (Agilent Technologies, New Jersey, United States). Based on the amplification curves, samples with cycle threshold (Ct) values below 36 were considered as positive results.

Evaluation of In-Field Accuracy of PlantVillage Nuru for Diagnosing the Viral Diseases of Cassava Based on Foliar Symptoms

The ability of PlantVillage Nuru to identify symptoms of CMD, CBSD, and CGM damage was tested in the field (using a Huawei P10 smartphone) by selecting 15 plants for each condition, as identified by experts, and five asymptomatic plants, making a total of 50 plants. Of the 15 plants with each disease/pest condition, five plants had intermediate symptoms of the condition, five plants had mild symptoms of the condition and five plants had unclear symptoms of the condition. This last group comprised plants in which symptoms were not typical of symptoms for that condition. Plants in each of these groups were selected by two researchers with at least three years of experience on cassava pests and diseases. For each of the sampled plants, six leaves were assessed, three from the top and three from the bottom part of the plant. During diagnosis with PlantVillage Nuru, the app was pointed at the leaves for a period of 10 s and the symptoms that were detected were identified by boxes that popped up on the diagnosis screen. The degree of congruence between PlantVillage Nuru's diagnoses and those of the experts, for each of the symptom categories, was determined by calculating the percentage of leaves for which diagnoses matched.

The value of making diagnoses using more than one leaf was determined and the degree to which results from experts and PlantVillage Nuru matched for upper leaves was assessed by comparing results first for leaf 1 (top upper leaf), then for leaves 1 and 2 (two upper leaves), then for leaves 1, 2, and 3 (three upper leaves). The same process was repeated for lower leaves where the one-leaf comparison was done using the first of the lower leaves, the two-leaf comparison using two lower leaves, and the three-leaf comparison using three lower leaves. Finally, for each of the three conditions (CMD, CBSD, and CGM), percentage matches between expert and PlantVillage Nuru were calculated when using between one and six leaves. The one-leaf comparison used the top upper leaf only. Subsequent comparisons used the following sets of leaves: two-leaf comparison (one upper leaf and one lower leaf); three-leaf comparison (two upper leaves and one lower leaf); four-leaf comparison (two upper and two lower leaves), five-leaf comparison (three upper and two lower leaves), and six-leaf comparison (three upper and three lower leaves). Results for all of the conditions were then combined to give an overall matching percentage value between expert and PlantVillage Nuru when using different numbers of leaves for diagnosis. Examples of the leaves that were used for investigating the use of multiple

leaves for the diagnosis of the whole plant are illustrated in **Supplementary Material B**.

In-field and on-screen symptom recognition accuracy of PlantVillage Nuru were compared to determine if there was a difference in the performance when assessing leaves "on-screen" or "in-field." To do this, pictures were taken of each of the 300 leaves used for the in-field assessment described above (90 leaves from CMD-affected plants, 90 from CBSD-infected plants, 90 from plants with CGM-damage and 30 leaves that were asymptomatic). PlantVillage Nuru was then used to diagnose these pictures on a laptop screen and the accuracy of PlantVillage Nuru in diagnosing each of the four conditions in-field and on-screen was determined by comparing the diagnoses of the AI system with those of the experts.

Development of the Cassava Symptom Recognition Assessment Tool for Validation of the Cassava Model, on PlantVillage Nuru

The cassava symptom recognition assessment tool (CaSRAT) was developed at IITA based on a scoring matrix for assessing the condition of 170 images that were randomly selected from a local cassava farm in Mkuranga region. These images include cassava leaves that were symptomatic for CMD (30 images), CBSD (69 images), CGM (51 images), and co-infection (8 images) as well as images with symptoms of other conditions (such as mineral deficiency–8 images) and images that could not be properly diagnosed (4 images). The 170 images were reviewed by 10 cassava experts at IITA-Tanzania, with more than 3 years' experience of working on cassava pests and diseases, and the group made consensus diagnoses for each of the images. These consensus identifications were then used for the image set as a baseline against which to judge the performance of test groups. During the assessment with CaSRAT, the images were presented to individuals being tested using a projector at an interval of 15 s per image and each individual filled in a scoring sheet (**Supplementary Material C**) with pre-coded scores (i.e., "1" for CMD, "2" for CBSD etc.) indicating the symptoms observed on the cassava leaf images. The scores filled in by the individuals being assessed were then compared to the expert score and the accuracy for symptom recognition for each individual was then calculated as a percentage of leaves that were correctly scored. Detailed information on how the symptom recognition assessment tool was developed can be found in **Supplementary Material D**.

Comparison of the Diagnostic Capabilities of Researchers, Extension Officers, Farmers, and PlantVillage Nuru by Using CaSRAT

The diagnostic capabilities of 60 people in three major categories (researchers, extension agents and farmers) were compared. For each major category, there were 10 who had been trained on cassava pests and diseases and 10 who had not. Both sets of researchers were from IITA-Tanzania, whilst the sets of extension officers and farmers (both trained and untrained) were from Mkuranga District, south of Dar es Salaam, Tanzania. The farmers evaluated in this study were all cassava growers and they were from one of the main cassava growing regions in Tanzania. The trained farmers were selected from farmers that had been involved in a study between 2014 and 2016 investigating the effectiveness of community phytosanitation in reducing the effects of cassava diseases (Legg et al., 2017). This group had received training on each of the major pests and diseases of cassava. The untrained farmers were selected from farmers that had not participated in the phytosanitation study. All farmers selected had been growing cassava for at least 5 years. During the assessments of accuracy in symptom recognition for the various groups tested, each individual was shown the CaSRAT images and was requested to score the images based on the symptom/condition that they observed. Average scores were obtained for each set.

To evaluate PlantVillage Nuru using CaSRAT, the app was pointed at a laptop screen showing the CaSRAT images at 15 s intervals. Diagnoses provided by PlantVillage Nuru were used to fill in the CaSRAT scoring sheet for determining the app's accuracy score by comparing its diagnoses with the consensus diagnoses of the experts. Evaluation of PlantVillage Nuru by CaSRAT was done using four different phones (Huawei P10, Samsung Galaxy 4, Tecno Camon CM and Infinix Hot 5) to determine if there was variation in Nuru's accuracy based on the type of phone used. Additionally, the CaSRAT assessment was conducted twice in order to evaluate potential improvements over time, firstly using PlantVillage Nuru V1.05 (released in March 2018) and secondly with PlantVillage Nuru V2.6.0-39 (released in June 2020).

Evaluating the Teaching Capability of the Cassava Model in PlantVillage Nuru

The teaching capability of PlantVillage Nuru was evaluated by determining the ability of agricultural extension agents to identify symptoms of cassava diseases prior to and after using the app. CaSRAT was used to determine the disease diagnosis ability of 30 agricultural extension agents and 50 farmers from Busia County (western Kenya) prior to the introduction and use of PlantVillage Nuru. After the assessment, all the agricultural extension officers and farmers were trained on diseases and pests of cassava and then introduced to the app. Afterwards, both the agricultural extension agents and farmers were divided into two groups, half of whom were given phones with access to PlantVillage Nuru and requested to inspect and collect data from 100 healthy and diseased cassava plants using app, within two weeks. The other half were not given access to PlantVillage Nuru nor asked to collect data from cassava plants. After the two weeks, the symptom recognition ability of extension agents and farmers, with and without PlantVillage Nuru, was assessed using CaSRAT and their symptom recognition accuracy scores were compared.

Statistical Analysis

Boxplots, single factor ANOVA and Tukey HDS statistics were used to evaluate and compare the accuracy scores obtained from the researchers, agricultural extension agents and farmers. The

statistical analyses were done using R Studio Version 1.1.456 (RStudio Inc.). The script used for the statistical analysis is found in **Supplementary Material D**.

RESULTS

Accuracy of Cassava Experts in Diagnosing Symptoms of CMD and CBSD

Comparison of the molecular and visual diagnoses showed that the experts in cassava pests and diseases could determine the presence or absence of either CMD and CBSD in a plant with a high degree of accuracy (95% for CMD and 81% for CBSD), when only one condition is considered (**Table 1**). The majority of the cases where the experts failed to identify symptoms of diseases were due to the presence of CBSD-causing viruses that produced few or no symptoms, a condition known as latent infection.

Ability of PlantVillage Nuru and Cassava Experts to Diagnose Symptoms of the Viral Diseases of Cassava Based on the Examination of Single Leaves

The ability of PlantVillage Nuru to accurately identify symptoms of CMD, CBSD, and CGM-damage, and therefore diagnose the presence of these diseases/conditions, was examined using 90 cassava leaves (from 15 cassava plants) that were either asymptomatic or symptomatic for CMD, CBSD, and CGM-damage. Importantly, although plants were classified by experts as affected by CMD, CBSD, or CGM, not all inspected leaves of those plants expressed symptoms of those diseases/pests. PlantVillage Nuru could identify asymptomatic leaves with an accuracy above 90%, which is higher than the accuracy for symptomatic leaves (21–59%); it could also identify symptoms of CGM-damage (40–56%) and CMD (52–59%) better than CBSD (21% accuracy) (**Table 2**). Furthermore, PlantVillage Nuru's accuracy for identification of CBSD symptoms (21%) was significantly lower than that of CMD (59%) and CGM (56%).

TABLE 1 | Comparison of visual and molecular diagnosis of CMD and CBSD.

Condition	Visual	Molecular	Number of plants	Overall accuracy score	Overall condition
CMD	Present	Present	19	95%	Accurately diagnosed
	Absent	Absent	52		
	Absent	Present	3	5%	Misdiagnosed
	Present	Absent	1		
CBSD	Present	Present	35	81%	Accurately diagnosed
	Absent	Absent	26		
	Absent	Present	9	19%	Misdiagnosed
	Present	Absent	5		

TABLE 2 | Comparison of in-field and on-screen capacity of the PlantVillage Nuru app to identify symptoms of CMD, CBSD, and CGM-damage as well as no symptoms on plants identified by experts as affected by CMD, CBSD, CGM or asymptomatic.

	Level of identification (%)			
Condition of the plants	Nuru in-field	Nuru on-screen	Expert on-screen	Expert-Nuru agreement
CMD	59%	52%	100%	52%
CBSD	21%	21%	74%	44%
CGM-damage	56%	40%	76%	51%
Asymptomatic	90%	97%	97%	100%
Overall	49%	53%	87%	62%

For all disease/pest damage conditions, mild symptoms were correctly identified by PlantVillage Nuru with much lower accuracy (13–37%) than intermediate symptoms (27–83%), whilst symptoms that were considered to be unclear were identified with moderate accuracy (23–63%) (**Table 3**). Examples of leaf images showing intermediate, mild and unclear symptoms are illustrated in **Supplementary Material A**. Plants with "intermediate" symptoms had more distinctive and uniformly distributed patterns of disease expression than those with mild symptoms, where symptoms were less prominent on a smaller proportion of leaves. Plants with "unclear" symptoms had symptoms that could be confused with other diseases or conditions, such as mineral deficiency, or brown leaf spot.

The on-screen accuracy of PlantVillage Nuru for identifying symptoms of CMD, CBSD, and CGM-damage as well as asymptomatic leaves was similar to the in-field accuracy (**Table 2**). PlantVillage Nuru's accuracy for identifying symptoms of CMD and CGM-damage in-field (59 and 56%, respectively) was slightly higher for than that obtained from on-screen diagnosis in CaSRAT (52 and 40%, respectively) while accuracy for identifying CBSD symptoms was the same in both cases (21%). On the other hand, the on-screen accuracy of PlantVillage Nuru for identification of asymptomatic leaves (97%) was slightly higher than that of in-field diagnosis (90%). Overall, accuracy

TABLE 3 | Capability of the cassava AI Nuru app to identify symptoms of cassava pests and diseases in the leaves of plants identified by experts as affected by CMD, CBSD, CGM based on the severity of the symptoms.

Condition	Severity	No. of leaves	No. of leaves accurately diagnosed	% Accuracy of symptom recognition
CMD	Intermediate	30	25	83%
	Mild	30	11	37%
	Unclear	30	17	57%
CBSD	Intermediate	30	8	27%
	Mild	30	4	13%
	Unclear	30	7	23%
CGM	Intermediate	30	24	80%
	Mild	30	7	23%
	Unclear	30	19	63%

scores for PlantVillage Nuru "in-field" and "on-screen" were very similar (**Table 2**), which demonstrates the validity of using CaSRAT to compare the "on-screen" accuracy of PlantVillage Nuru with that of humans.

The ability of the experts to diagnose CBSD and CGM-damage on-screen (**Table 2**), using the images obtained from the 300 leaves (of 50 plants: 15 CMD, 15 CBSD, 15 CGM, 5 asymptomatic), was less than for the in-field analysis of the 50 whole plants. Although humans correctly identified CMD symptoms from 100% of the images obtained from plants tagged in the field as affected by CMD and 97% of images were correctly identified as asymptomatic, levels of image recognition were lower for CBSD (74%) and CGM (76%). The reason for this is that several of the leaves from plants tagged as affected by CBSD or CGM did not express symptoms of these conditions. CBSD-infected plants typically express weak symptoms or no symptoms at all on upper leaves whilst CGM symptoms are only prominent on upper leaves. For this reason, the use of multiple leaves for diagnosis of the whole plant was hypothesized as a means to improve the diagnostic accuracy of PlantVillage Nuru since cassava experts usually inspect the whole plant instead of a few leaves.

When the number of leaves examined by PlantVillage Nuru was increased, the diagnosis capacity for all three conditions (CMD, CBSD, and CGM) improved greatly (**Figure 1**). The use of two leaves per plant improved PlantVillage Nuru's diagnosis capacity for CMD from 53 to 93% for upper leaves and from 73 to 93% for lower leaves, CGM-damage from 67 to 93% for upper leaves and from 47 to 53% for lower leaves, while at least three leaves were required to achieve a similar level of improvement for diagnosis of CBSD (from 0 to 27% for upper leaves and 33 to 53% for lower leaves) (**Figure 2**).

The location of leaves had an impact on PlantVillage Nuru's accuracy for identifying symptoms of CGM-damage and CBSD but not CMD. CGM-damage was more accurately identified using upper leaves while CBSD symptoms were more accurately identified using lower leaves. However, the use of six leaves (3-upper and 3-lower) provided the highest likelihood of diagnosis of all conditions, at which the diagnosis of CMD, CBSD and CGM-damage were diagnosed with 93, 73 and 93% accuracy (**Figure 1**). For this reason, the use of three upper and three lower leaves was the recommended number of leaves suggested for improving the accuracy of the cassava model in Nuru. This is the model that is now implemented in the currently available PlantVillage Nuru app in the Android PlayStore.

Ability of PlantVillage Nuru, Agricultural Extension Agents and Farmers to Diagnose Symptoms of CMD, CBSD, and CGM on Cassava Leaves

The CaSRAT was used to determine the ability of individuals and groups to accurately diagnose foliar symptoms. The researchers trained on cassava pests and diseases were able to identify symptoms of CMD, CBSD, and CGM-damage with higher accuracy than untrained researchers, agricultural extension officers and farmers (**Figure 3**). The mean accuracy score of symptom recognition by trained researchers (86%) was about four times higher than that of untrained researchers (21%) (Welch two-sample t-test $p = 4.30E{-}11$), almost twice as high as that of trained agricultural extension officers (49%) (Tukey HSD $p < 0.001$) and three times higher than that of trained farmers (23%) (Tukey HSD $p < 0.001$). However, there were small differences between the mean accuracy scores of trained and untrained agricultural extension officers (49 and 32%, respectively, Tukey HSD $p < 0.05$) and trained and untrained farmers (23 and 12%, respectively, Tukey HSD $p > 0.5$).

The majority of images that were misdiagnosed by farmers, trained and untrained, had symptoms of CBSD and CGM-damage; these accounted for 37–42% of all misdiagnoses (**Supplementary Material E1**). The untrained farmers misdiagnosed leaves with CBSD symptoms as healthy, i.e., asymptomatic (24% of all misdiagnoses), and CGM-damage as CBSD and healthy (14% of all misdiagnoses for each condition). On the other hand, trained farmers most commonly confused symptoms of CBSD with those of CGM-damage (12% of all misdiagnoses) and symptoms of CGM-damage with those of CBSD (20% of all misdiagnoses).

Similarly, the majority of misdiagnoses by agricultural extension officers, trained and untrained, were also due to symptoms of CBSD (30%) and CGM-damage (45%) (**Supplementary Material E2**). Untrained agricultural extension officers confused symptoms of CBSD with those of CMD (22% of all misdiagnoses) and symptoms of CGM-damage with those of CBSD (23% of all misdiagnoses). Trained agricultural officers also confused symptoms of CGM-damage with those of CBSD (28% of all misdiagnoses), however, they seemed to be less clear with symptoms of CBSD as they misdiagnosed some of the leaves with CBSD symptoms as CMD-infected (11% of all misdiagnoses) and CGM-damage (10% of all misdiagnoses).

Untrained researchers also confused symptoms of CBSD (41% of misdiagnoses) and CGM-damage (39%) more than CMD, while the majority of the misdiagnoses obtained from trained researchers were due to CGM-damage and co-infection (34 and 37% of all misdiagnoses; **Supplementary Material E3**). The untrained researchers seemed to have confused symptoms of CBSD with those of CMD (11% of all misdiagnoses), CGM-damage (10% of all misdiagnoses) as well as healthy (7% of all misdiagnoses) and other conditions (14% of all misdiagnoses). On the other hand, trained researchers confused symptoms of CGM-damage with those of CMD and CBSD (12% of all misdiagnoses for each condition).

Most of the misdiagnoses due to symptoms of co-infection were obtained from leaves that had symptoms of both, CBSD and CGM-damage (**Supplementary Material E**), and most of these were identified by researchers and agricultural extension officers to have either CBSD or CGM only. However, the number of images with symptoms of co-infection were few hence more images are required for a proper analysis.

Nuru was able to identify disease symptoms with 54% accuracy when tested on the 170 images used with the CaSRAT, using the same phone that was used for the in-field assessment. However, Nuru's accuracy for symptom identification was lower (30–41%) for the other three phones assessed using the cassava

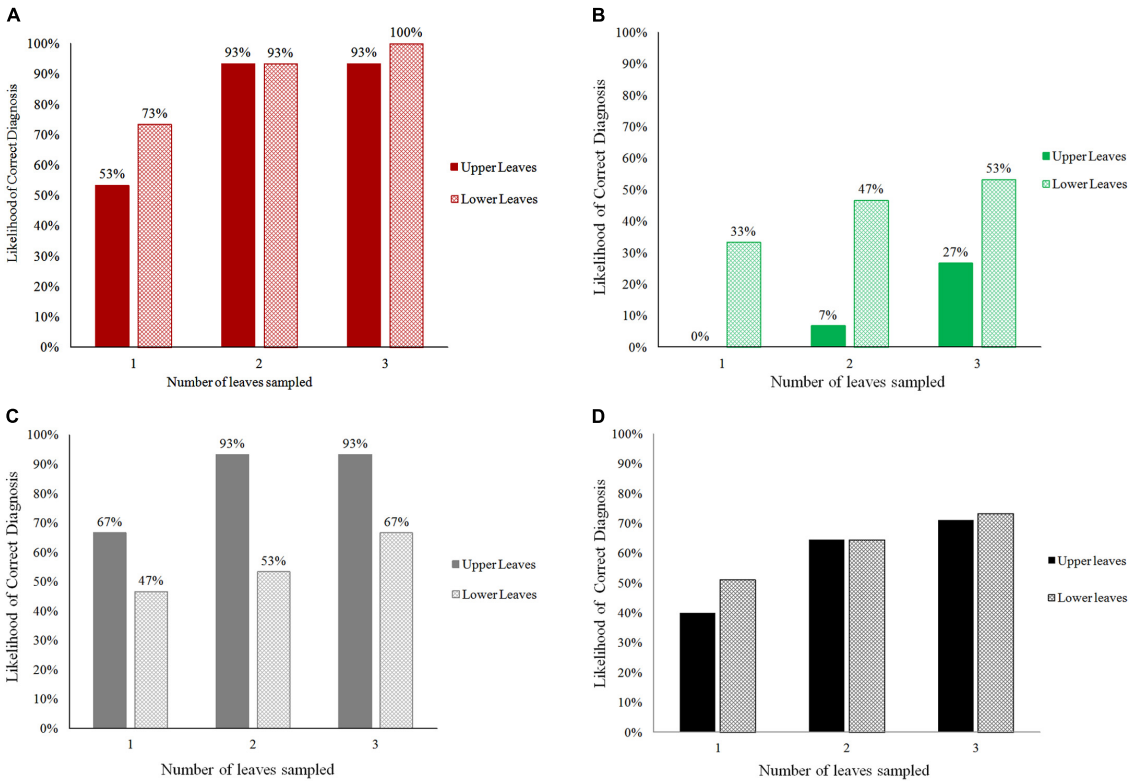

FIGURE 1 | The overall likelihood of obtaining correct diagnoses for CMD **(A)**, CBSD **(B)**, CGM-damage **(C)**, and all conditions **(D)** when using the cassava AI Nuru based on the number and position (upper or lower) of leaves sampled from 90 cassava plants.

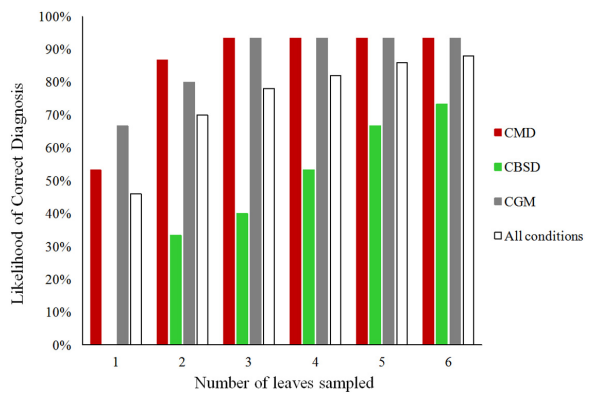

FIGURE 2 | The likelihood of obtaining a correct diagnosis for CMD, CBSD, CGM-damage, and all the three conditions from cassava leaves obtained from 90 plants by using the PlantVillage Nuru app for symptoms based on the number of leaves sampled.

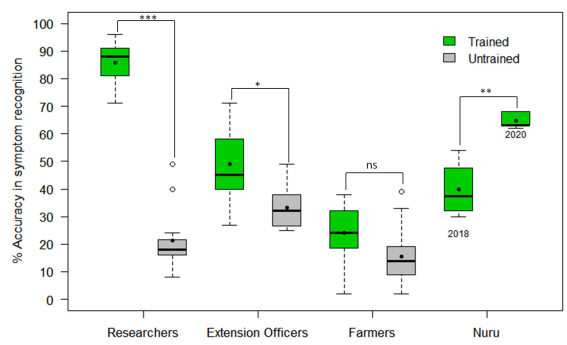

FIGURE 3 | Comparison of accuracy in recognizing symptoms of CMD, CBSD and CGM damage by PlantVillage Nuru, researchers, agricultural extension officers and farmers who have and those who have not previously received training on cassava diseases and pests ($n = 22$ for each paired comparison). The black line within the boxes represents the median, the top and bottom of the box represent the 75th and the 25th percentiles, the whiskers represent the maximum and minimum values while the symbols (●) and (◯) represent the mean and the outlier.

model V1.05 (Nuru, 2018) (**Table 4**). Although the average accuracy score of Nuru, 2018 with four phones (40 ± 10%) was slightly lower than that of trained agricultural extension officers (49%) (Tukey HSD $p > 0.5$), its score in 2020 (65 ± 3%) was significantly higher (Tukey HSD $p < 0.05$). The accuracy of the cassava model V.2.6 in Nuru, 2020 was higher for all four different phones that were evaluated (62–68% overall accuracy score) suggesting that PlantVillage Nuru's improvements also enabled the app to perform better in different phones.

TABLE 4 | Comparison of overall accuracy of Nuru for identification of foliar symptoms of cassava mosaic disease (CMD), cassava brown streak disease (CBSD), and damage caused by cassava green mites (CGM) on cassava leaves based on the cassava symptom recognition assessment tool (CaSRAT).

Year	March 2018				June 2020				
Phone type	**Samsung Galaxy 4**	**Infinix Hot 5**	**Tecno Camon CM**	**Huawei P1O**	**Samsung Galaxy A10**	**Infinix Hot 7**	**Tecno Camon CM**	**Tecno Spark3 Pro**	**Huawei P30**
Overall accuracy score	30%	41%	34%	54%	62%	63%	68%	63%	68%
Average accuracy score	40 ± 10%				65 ± 3%				

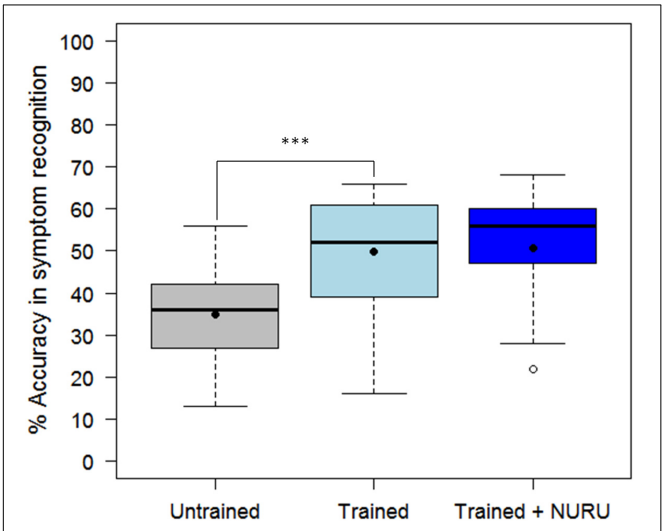

FIGURE 4 | Comparison of the ability of 30 agricultural extension officers and 50 farmers to accurately recognize symptoms of CMD, CBSD, and CGM-damage before and after receiving training on cassava pests and diseases. The black line within the boxes represents the median, the top and bottom of the box represent the 75th and the 25th percentiles, the whiskers represent the maximum and minimum values while the symbols (●) and (◯) represent the mean and the outlier.

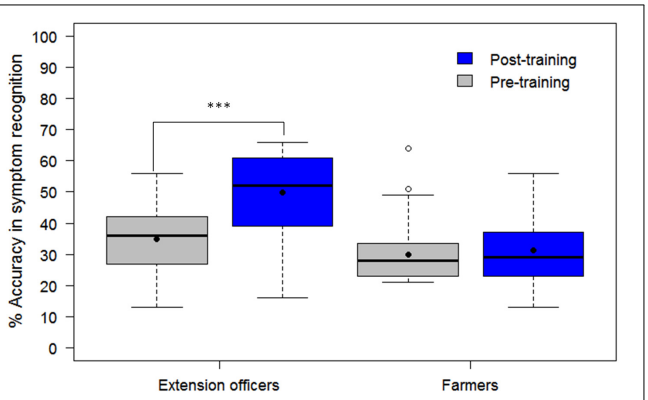

FIGURE 5 | Comparison of the ability of 30 agricultural extension officers to accurately recognize symptoms of CMD, CBSD, and CGM-damage before and after receiving training on cassava pests and diseases as well as after using the PlantVillage Nuru app for diagnosing cassava diseases. The black line within the boxes represents the median, the top and bottom of the box represent the 75th and the 25th percentile, the whiskers represent the maximum and minimum values while the symbols (●) and (◯) represent the mean and the outlier.

The Teaching Capability of Nuru

The assessment tool (CaSRAT) was also used to evaluate the teaching capability of Nuru by determining the ability of agricultural extension officers and farmers trained on cassava pests and diseases to correctly identify symptoms of diseases after using PlantVillage Nuru for about two weeks. There was a significant increase in the ability of the extension agents to identify symptoms of CMD, CBSD, and CGM after the training as evidenced by the increase in their mean accuracy score from 34.9 ± 10.9 to 49.9 ± 13.5% (Tukey HSD $p < 0.005$) (**Figure 4**). However, this was not the case for the farmers where there was only a marginal change in their accuracy score after training, from 29.9 ± 9.7 to 31.2 ± 10.2% (Tukey HSD $p > 0.5$). There was no significant difference in the mean accuracy scores of the agricultural extension officers before and after using PlantVillage Nuru, from 49.9 ± 13.5 to 50.5 ± 14.5% (Tukey HSD $p > 0.5$). However, the range of the accuracy scores obtained after training and Nuru usage (46–60%) was slightly narrower than that obtained from agricultural extension officers who had only received training without using Nuru (38–61%) (**Figure 5**).

The extension officers in this group also seemed to confuse symptoms of CBSD with those of CMD (14–19% of all misdiagnoses) and CGM-damage (19–27% of all misdiagnoses) as well as symptoms of CGM-damage with those of CMD (8–12% of all misdiagnoses) and CBSD (16– 20% of all misdiagnoses) (**Supplementary Material E5**). The percent of images that were misdiagnoses did not differ much after training and the use of PlantVillage Nuru for two weeks.

DISCUSSION

Development of smartphone-based technologies for diagnosis of disease and pest damage on plants requires input from experts who understand the phenotypes of the diseases and pests. Therefore, it is important to evaluate the expertise of these experts to ensure that the information used to develop such technologies is as accurate as possible. Quantifying the knowledge of the experts also provides a baseline that can be used to evaluate the effectiveness of the developed technologies. The present study examined the expertise of the experts who generated datasets used for the development of an

object-detection model for cassava disease diagnosis based on foliar symptoms, known as PlantVillage Nuru. The symptom recognition capability of PlantVillage Nuru was compared to that of the experts and its intended users so as to determine the effectiveness of the app. Since PlantVillage Nuru was developed as a diagnosis and training tool, its capacity to train users was also evaluated to determine if the app can be used to train its users to recognize the symptoms of diseases and pests affecting their plants and hence improve their diagnosis capability.

Accuracy of Cassava Experts in Diagnosing Symptoms of CMD and CBSD

Cassava experts were able to achieve a high level of accuracy in correctly identifying the disease affecting cassava plants (CMD or CBSD) using visual-based symptom recognition on leaves which were confirmed to be healthy or contain the disease-causing viruses by molecular diagnostic methods. This may represent the first time that such a visual inspection vs. molecular diagnostics comparison has been made for CMD, and it confirms the otherwise widely held view that CMD is relatively easy to identify based on visual symptom assessments (Abarshi et al., 2012). The small number of false negatives associated with CMD infection, identified in the present study, were thought to be due to latent infection that had not resulted in disease symptom expression.

CBSD is known to have more cryptic symptoms than CMD (Nichols, 1950) and more frequent latent infection (Adams et al., 2012). Furthermore, the expression of foliar symptoms in CBSD-infected plants has been reported to vary between leaves on the plants, cassava variety, growing conditions (temperature, rainfall, and altitude), age of the plant and the virus isolate involved in causing the disease symptoms (Hillocks and Jennings, 2003; Mohammed et al., 2012; Shirima et al., 2019, 2020). It is therefore unsurprising that there was a lower level of congruence between symptom vs. virus testing identifications for CBSD than was observed for CMD.

However, the overall high level of accuracy in the visual assessment of CMD and CBSD infection achieved by the experts provided a strong basis both for the previous development of the PlantVillage Nuru app (Ramcharan et al., 2019), as well as the use of expert diagnoses as a benchmark for comparison of diagnoses made by other groups as well as the PlantVillage Nuru app itself.

Comparison of the Ability of PlantVillage Nuru and Cassava Experts in Diagnosing Symptoms of the Viral Diseases of Cassava Based on Examination of One or More Leaves

When PlantVillage Nuru was used to identify the symptoms of CMD, CBSD, and CGM-damage (using single leaves), it was partially accurate for CMD (58% accuracy score) and CGM (56% accuracy score), but mostly inaccurate for CBSD (21% accuracy score). This highlights the fact that single-leaf diagnoses using a smartphone app such as PlantVillage Nuru are unreliable, partly due to the difference in the severity of the symptoms PlantVillage Nuru can recognize, but also resulting from the uneven distribution of symptoms within cassava plants, particularly for CBSD (Nichols, 1950).

The distinction of sampled plants into those with "mild," "intermediate," and "unclear" symptoms gave rise to large differences in the accuracy of single leaf diagnoses by PlantVillage Nuru. The accuracy for diagnosis of intermediate symptoms of CMD and CGM-damage was about 20% higher than that of unclear symptoms and the overall diagnosis accuracy. The diagnostic capacity of Nuru for leaves with mild symptoms was low for all the three conditions (CMD, CBSD, and CGM-damage), probably because images selected for training the app were mostly of leaves with clear symptoms. The overall lower accuracy in identifying CBSD symptoms has been reported previously for PlantVillage Nuru (Ramcharan et al., 2019), and it is a widely published fact that CBSD symptoms are cryptic, seasonally variable and can be difficult to identify, even for experienced researchers (Nichols, 1950; Hillocks and Thresh, 2000).

All of the pest and disease conditions of cassava (especially CMD, CBSD, and CGM) have patterns of symptom expression that vary greatly within plants, from plant to plant, and between varieties and they are affected also by virus strain variation, weather conditions and other environmental factors (Hahn et al., 1980; Thresh et al., 1994; Owor et al., 2004a,b; Mohammed et al., 2012). A plant with a recent vector-borne infection of CMD only expresses symptoms in upper leaves (Sseruwagi et al., 2004) and CGM damage is always most prominent on upper leaves (Onzo et al., 2005) while CBSD symptoms are usually present on leaves toward the bottom of the plant (Nichols, 1950). By definition, asymptomatic plants are uniformly symptom-free, which explains the much higher level of accuracy (90%) achieved by PlantVillage Nuru in identifying this condition from a single leaf. These factors highlight the importance of applying multi-leaf assessments when using PlantVillage Nuru. The same principle would also apply to any other app attempting to deliver phone-based diagnoses of cassava diseases and pest damage.

The use of two leaves (one upper and one lower) improved PlantVillage Nuru's diagnostic capability for CMD and CGM-damage to a similar accuracy to that of the cassava experts, however, this was not the case for CBSD which required six leaves to approach the expert's accuracy (accuracy score $> 74\%$). Hence the use of six leaves was adopted as the means for diagnosis of the whole plant as it provided PlantVillage Nuru with a better chance of identifying disease symptoms and therefore improved its performance, even when the severity of the symptoms varied. The use of six leaves mimics the approach that a researcher would take in the field, in which both upper and lower parts of the plants would be inspected in order to confirm the presence of CMD, CBSD or CGM symptoms (Hillocks and Thresh, 2000; Sseruwagi et al., 2004).

Comparison of the Ability of PlantVillage Nuru, Agricultural Extension Agents and Farmers to Correctly Diagnose Symptoms of CMD, CBSD, and CGM on Cassava Leaves

We compared the diagnostic capability of PlantVillage Nuru (for CMD, CBSD and CGM-damage) to that of its intended users—agricultural extension officers and farmers—to assess the potential beneficial effects of PlantVillage Nuru for these groups. PlantVillage Nuru had the same overall accuracy for identifying symptoms of CMD, CBSD, and CGM-damage as trained agricultural extension officers, who could identify disease symptoms better than untrained researchers, untrained agricultural extension officers and trained or untrained farmers.

The large difference in the symptom identification ability of the trained researchers compared to untrained researchers highlights the difficulty of diagnosing symptoms of CMD, CBSD, and CGM-damage such that only well-trained individuals can do so effectively. By contrast, the difference between the symptom recognition accuracy scores of the trained farmers and agricultural extension officers suggested that the vertical transfer of knowledge from researchers to extension agents is currently inefficient. In order to address this problem, a rigorous training programme might be needed to improve the ability of the agricultural extension agent and farmers to identify and differentiate symptoms of the diseases and conditions that they encounter, and efforts should be made to improve the efficiency with which agricultural extension information flows down to farmers. In order to achieve this, it will also be necessary to tackle the resource constraints which are common features of agricultural knowledge transfer systems in sub-Saharan Africa (Azumah et al., 2018). The ability of PlantVillage Nuru to diagnose disease symptoms with a similar level of accuracy to that achieved by experts, when using multiple leaves, indicated that the app might be able fill the knowledge gap between researchers and agricultural extension officers as well as farmers. Digital tools such as PlantVillage Nuru offer great potential for extending the reach and improving the efficiency of agricultural knowledge transfer systems in sub-Saharan Africa, as telecommunications networks continue their rapid expansion through the continent.

Evaluation of the Teaching Capability of PlantVillage Nuru

We demonstrated that training alone delivered significant improvements in accuracy of disease diagnosis for extension officers but not farmers, indicating that the applied training method might not have been suitable for farmers. Hence further investigations are recommended to determine the most suitable methods for training farmers and how to evaluate the effectiveness of such trainings. Training with PlantVillage Nuru resulted in a slight improvement in the symptom recognition capacity of both agricultural extension officers and farmers, suggesting that a longer period of time might be required to observe more substantial changes in the symptom recognition capacity of its users.

The training potential of PlantVillage Nuru is based on its ability to show users the symptoms of the diseases/conditions present in the leaves as it is being used (as illustrated in the Supplementary Videos for Healthy-diagnosis, CMD-diagnosis, CBSD-diagnosis, and CGM-diagnosis). This helps users to become familiar with the characteristic symptoms of each of the disease/pest damage types, which over time should enable them to recognize each of these conditions without the aid of PlantVillage Nuru. This learning function is reinforced by PlantVillage Nuru's library containing images of disease symptoms that the user can access and learn from, in the absence of a mobile network. Furthermore, if the user has access to a mobile network, PlantVillage Nuru can connect the users to researcher experts who can assist in diagnosis of the condition of the plant, through a platform where the user is able to ask questions and share images of their plants (as seen at https://plantvillage.psu.edu/posts). Furthermore, PlantVillage Nuru can link users with the PlantVillage database where users can get information on agricultural practices, diseases, and pests as well as their management techniques for different crops.

Hence, with time, PlantVillage Nuru can provide a quick, cost-effective and easily accessible means for disseminating knowledge and ensuring continuous training of agricultural extension agents and farmers, thereby improving their skills in pest/disease identification and management. On-going work in western Kenya indicates that Nuru is improving disease diagnosis skills of cassava farmers suggesting that efforts to scale out the use of Nuru across the cassava-growing regions of Africa will improve farmers' recognition and knowledge of cassava diseases/pests, which will contribute to improved disease/pest control and greater productivity.

Improvements Made on PlantVillage Nuru

The model that we tested here has already been updated and continuous improvements are being made as more data become available, so the diagnosis capacity of PlantVillage Nuru is expected to increase over time. Improvements in smartphone technology will also contribute to better performance of the app, since newer smartphone models tend to have better cameras which enhance sensitivity and accuracy of PlantVillage Nuru.

The diagnostic models used to develop Nuru are publicly available for independent validation, which allows their reliability to be assessed in diverse geographic regions where different varieties of the investigated plants are grown. This may enable modification of the disease diagnostic models to improve performance based on the condition and variety of the host plant as well as identification of conditions that may need to be included in the training datasets. "Open science" models from other sectors, like genome biology and genetic diseases, have led to rapid advances in the application of machine learning approaches to the development of diagnostics.

Additional symptom types that have been proposed for inclusion within the cassava PlantVillage Nuru model include nutrient-deficiency and fungal infection (Howeler, 2002; Hillocks et al., 2002), although these are currently of much

lower importance in sub-Saharan Africa than the cassava viruses and CGM. However, the accuracy of the model might be reduced by increasing the number of different conditions with similar symptoms. Therefore other tools which could provide complementary simple, cheap and rapid means for in-field diagnosis of plant diseases (such as pathogen-based biosensors) could be used to enhance the capability of PlantVillage Nuru. Pathogen-based biosensors use the pathogen's antibodies or nucleic acid for diagnosis (Khater et al., 2017) and hence could act as a confirmatory test for cases where disease symptoms cannot be determined or if there is latent infection. Since pathogen-based biosensors have not yet been developed for in-field diagnosis of the viral diseases of cassava, symptom recognition and lab-based assays will continue to remain the main methods for diagnosis of CMD and CBSD in the immediate future. Therefore, Nuru has an important role to play in building disease diagnostic capacity of agricultural extension officers and farmers.

CONCLUSION

The present study has shown that PlantVillage Nuru can be as effective as experts in identifying the symptoms of the viral diseases of cassava (CMD and CBSD) and CGM-damage. PlantVillage Nuru gave a high level of symptom recognition accuracy, which was better than that achieved by agricultural extension officers and farmers, suggesting that it can be used for increasing their diagnostic capacity for the viral diseases of cassava as well as CGM. Since the disease diagnostic models in PlantVillage Nuru are continuously being improved and more knowledge is added as it becomes available, the app provides agricultural extension officers and farmers with an ever-improving direct link to experts and expert knowledge. These features give PlantVillage Nuru and mobile-based apps that can effectively diagnose symptoms of disease and pest damage the ability to revolutionize disease and pest management in agriculture.

The rapid penetration of affordable smartphone technology throughout rural Africa will certainly ensure that the platform for widespread access to such apps will be in place within the near future. Raising awareness amongst farming communities about the availability and utility of apps such as PlantVillage Nuru is an important next step in promoting this process and will require determined and innovative efforts from stakeholders in agricultural development supported by teams of IT specialists. Crucially, however, the success of these endeavours will ultimately depend on farmers' access to affordable control measures. Although some of this, such as advice on cultural control techniques, can be delivered through apps, elements such as the provision of high-quality seed/planting material of pest/disease resistant varieties will require investment in more traditional on-the-ground extension approaches. As governments increasingly seek to promote the application of ICT solutions to agricultural development in Africa, strong parallel efforts will be required to strengthen variety development, deployment and dissemination systems as well as sustainable approaches for the delivery of other inputs required for effective pest and disease management.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LM, NM, MN, AR, AK, HM, PM, and JL collected the data. LM and JL prepared the manuscript. JL and DH supervised the projects. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financed through an "Inspire Challenge" grant from the BigData Platform of the Consultative Group on International Agricultural Research (CGIAR), Pennsylvania State University, the Bill and Melinda Gates Foundation (to DH), Schmidt Futures (to DH), and Self Help Africa (SHA–Kenya). The contributions of JL were supported through the Roots, Tubers and Bananas (RTB) program of the CGIAR. The support of the Ministries of Agriculture in Tanzania and Kenya is recognized, as are the contributions of farmers and extension staff in Mkuranga District, Tanzania and Busia County, Kenya.

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at Biorxiv.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.590889/full#supplementary-material

REFERENCES

Abarshi, M., Mohammed, I., Jeremiah, S., Legg, J., Kumar, L., Hillocks, R., et al. (2012). Multiplex RT-PCR assays for the simultaneous detection of both RNA and DNA viruses infecting cassava and the common occurrence of mixed infections by two cassava brown streak viruses in East Africa. *J. Virol. Methods* 179, 176–184. doi: 10.1016/j.jviromet.2011.10.020

Abrahams, P., Beale, T., Cock, M., Corniani, N., Day, R., and Godwin, J. (2017). *Fall Armyworm Status: Impacts and Control Options in Africa: Preliminary Evidence Note*. Oxfordshire: Centre for Agriculture and Biosciences International.

Adams, P., Abidrabo, P., Miano, W., Alicai, T., Kinyua, M., Clarke, J., et al. (2012). High throughput real-time RT-PCR assays for specific detection of cassava brown streak disease causal viruses and their application to testing planting material. *Plant Pathol.* 62, 233–242. doi: 10.1111/j.1365-3059.2012.02622.x

Azumah, B., Donkoh, S., and Awuni, A. (2018). The perceived effectiveness of agricultural technology transfer methods: evidence from rice farmers in Northern Ghana. *Cogent Food Agric.* 4:1. doi: 10.1080/23311932.2018.1503798

Barbedo, J. (2014). An automatic method to detect and measure leaf disease symptoms using digital image processing. *Plant Dis.* 98, 1709–1716. doi: 10.1094/PDIS-03-14-0290-RE

Daniel, E., Bastiaans, L., Rodenburg, J., Center, A. R., Schut, M., and Mohamed, J. K. (2013). *Assessment of Agricultural Extension Services in Tanzania. A Case Study of Kyela, Songea Rural, and Morogoro Rural Districts. Internship Report in Plant Sciences. Crops Systems Analysis Group*. Wageningen: University and Research Centers.

Dixon, A., Bandyopadhyay, R., Coyne, D., Ferguson, M., Ferris, R., Hanna, R., et al. (2003). Cassava: from poor farmers' crop to pacesetter of African rural development. *Chron. Horticult.* 43, 8–15.

Fao, I. F. A. D., Unicef, W. F. P., and Who. (2018). *The State of Food Security and Nutrition in the World. Building Climate Resilience for Food Security and Nutrition*. Rome: *FAO*.

FAOSTAT (2018). *Food and Agriculture Organisation of the United Nations Database*. Paris: FAOSTAT.

Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17:2022. doi: 10.3390/s17092022

Furuholt, B., and Matotay, E. (2011). The developmental contribution from mobile phones across the agricultural value chain in rural Africa. *Electr. J. Inform. Syst. Dev. Countries* 48, 1–16. doi: 10.1002/j.1681-4835.2011.tb00343.x

GSMA (2019). *The Mobile Economy-Sub-Saharan Africa*. London: Global System for Mobile Communications Association.

Hahn, S., Leuschner, K., Ezeilo, W., Carpenter, A., Khatibu, A., and Constantin, C. (1980). Resistance of Cassava Clones to Cassava Green Mite, *Mononychellus tanajoa*. *Trop. Pest Manag.* 26, 265–267. doi: 10.1080/09670878009414410

Hillocks, R., and Jennings, D. (2003). Cassava brown streak disease: a review of present knowledge and research needs. *Int. J. Pest Manag.* 49, 225–234. doi: 10.1080/0967087031000101061

Hillocks, R., and Thresh, J. (2000). Cassava mosaic and cassava brown streak virus diseases: a comparative guide to symptoms and aetiologies. *Roots* 7, 1–8. doi: 10.9734/arrb/2016/26879

Hillocks, R., Thresh, J., and Bellotti, A. (2002). *Cassava: Biology, Production and Utilization*. Wallingford: CABI publishing, 301–318.

Hong, Y., Robinson, D., and Harrison, B. (1993). Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. *J. Gen. Virol.* 74, 2437–2443. doi: 10.1099/0022-1317-74-11-2437

Howeler, R. (2002). "Cassava mineral nutrition and fertilization," *Cassava: Biology, Production and Utilization*, eds R. J. Hillocks, J. M. Thresh, and A. C. Bellotti (CAB International), 115–147.

Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A. D., et al. (2017). Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comp. Electr. Agric.* 138, 200–209. doi: 10.1016/j.compag.2017.04.013

Kawuki, R., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I., Ozimati, A., et al. (2016). Eleven years of breeding efforts to combat cassava brown streak disease. *Breed. Sci.* 66, 560–571. doi: 10.1270/jsbbs.16005

Khater, M., de la Escosura-MuÑiz, A., and Merkoçi, A. (2017). Biosensors for plant pathogen detection. *Biosens. Bioelectron.* 93, 72–86. doi: 10.1016/j.bios.2016.09.091

Legg, J., Jeremiah, S., Obiero, H., Maruthi, M., Ndyetabula, I., Okao-Okuja, G., et al. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak pandemics in Africa. *Virus Res.* 159, 161–170. doi: 10.1016/j.virusres.2011.04.018

Legg, J., Lava, P., Makeshkumar, T., Ferguson, M., Kanju, E., and Ntawuruhunga, P. (2015). Cassava virus diseases: biology, epidemiology and management. *Adv. Virus Res.* 91, 85–142. doi: 10.1016/bs.aivir.2014.10.001

Legg, J., Ndalahwa, M., Yabeja, J., Ndyetabula, I., Bouwmeester, H., Shirima, R., et al. (2017). Community phytosanitation to manage cassava brown streak disease. *Virus Res.* 241, 236–253. doi: 10.1016/j.virusres.2017.04.020

Misaki, E., Apiola, M., Gaiani, S., and Tedre, M. (2018). Challenges facing sub-Saharan small-scale farmers in accessing farming information through mobile phones: a systematic literature review. *Electr. J. Inform. Syst. in Dev. Countries* 84:e12034. doi: 10.1002/isd2.12034

Mohammed, I., Abarshi, M., Muli, B., Hillocks, R., and Maruthi, M. (2012). The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. *Adv. Virol.* 2012:795697. doi: 10.1155/2012/795697

Monger, W., Seal, S., Isaac, A., and Foster, G. (2001). Molecular characterization of the Cassava brown streak virus coat protein. *Plant Pathol.* 50, 527–534. doi: 10.1046/j.1365-3059.2001.00589.x

Ndunguru, J., Legg, J. P., Aveling, T. A. S., Thompson, G., and Fauquet, C. M. (2005). Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a canter of diversity of cassava geminiviruses. *Virol. J.* 2:21.

Nichols, J. (1950). The brown streak disease of cassava: distribution, climatic effects and diagnostic symptoms. *East African Agric. J.* 15, 154–160. doi: 10.1080/03670074.1950.11664727

Onzo, A., Hanna, R., and Sabelis, M. W. (2005). Biological control of cassava green mites in Africa: impact of the predatory mite Typhlodromalus aripo. *Entomol. Berichten.* 65, 2–7.

Owor, B., Legg, J., Obonyo, R., Okao-Okuja, G., Kyamanywa, S., and Ogenga-Latigo, M. W. (2004a). Field studies of cross protection with cassava in Uganda. *J. Phytopathol.* 152, 243–249. doi: 10.1111/j.1439-0434.2004.00837.x

Owor, B., Legg, J., Okao-Okuja, G., Obonyo, R., and Ogenga-Latigo, M. (2004b). The effect of cassava mosaic geminiviruses on symptom severity, growth and root yield of a cassava mosaic virus disease-susceptible cultivar in Uganda. *Ann. Appl. Biol.* 145, 331–337. doi: 10.1111/j.1744-7348.2004.tb00390.x

Pariona, A. (2019). *Most Important Staple Foods in the world*. Available online at: https://www.worldatlas.com/articles/most-important-staple-foods-in-the-world.html (accessed on 25 May 2020).

Pethybridge, S., and Nelson, S. (2015). Leaf doctor: a new portable application for quantifying plant disease severity. *Plant Dis.* 99, 1310–1316. doi: 10.1094/PDIS-03-15-0319-RE

Prasad, S., Peddoju, S. K., and Ghosh, D. (2016). Multi-resolution mobile vision system for plant leaf disease diagnosis. *Signal Image Video Process.* 10, 379–388. doi: 10.1007/s11760-015-0751-y

Qiang, C., Kuek, S., Dymond, A., and Esselaar, S. (2012). *Mobile Applications for Agriculture and Rural Development*. Washington, DC: World Bank.

Qin, F., Liu, D., Sun, B., Ruan, L., Ma, Z., and Wang, H. (2016). Identification of alfalfa leaf diseases using image recognition technology. *PLoS One* 11:e0168274. doi: 10.1371/journal.pone.0168274

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. (2017). Deep learning for image-based cassava disease detection. *Front. Plant Sci.* 8:1852. doi: 10.3389/fpls.2017.01852

Ramcharan, A., McCloskey, P., Baranowski, K., Mbilinyi, N., Mrisho, L., Ndalahwa, M., et al. (2019). A mobile-based deep learning model for cassava disease diagnosis. *Front. Plants Sci.* 10:272. doi: 10.3389/fpls.2019.00272

Shirima, R., Legg, J., Maeda, D., Tumwegamire, S., Mkamilo, G., Mtunda, K., et al. (2020). Genotype by environment cultivar evaluation for cassava brown streak disease resistance in Tanzania. *Virus Res.* 286:198017. doi: 10.1016/j.virusres.2020.198017

Shirima, R., Maeda, D., Kanju, E., Tumwegamire, S., Ceasar, G., Mushi, E., et al. (2019). Assessing the degeneration of cassava under high virus inoculum

conditions in Coastal Tanzania. *Plant Dis.* 103, 2652–2664. doi: 10.1094/PDIS-05-18-0750-RE

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* 2016:3289801. doi: 10.1155/2016/3289801

Sseruwagi, P., Sserubombwe, W., Legg, J., Ndunguru, J., and Thresh, J. M. (2004). Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. *Virus Res.* 100, 129–142. doi: 10.1016/j.virusres.2003.12.021

StatCounter Global Stats (2020). *Mobile Operating System Market Share Africa*. Available online at: https://gs.statcounter.com/os-market-share/mobile/africa (accessed on 18 May 2020).

Thresh, J., and Cooter, J. (2005). Strategies for controlling cassava mosaic virus disease in Africa. *Plant Pathol.* 54, 587–614. doi: 10.1111/j.1365-3059.2005.01282.x

Thresh, J., Otim-Nape, G., and Jennings, D. (1994). Exploiting resistance to African cassava mosaic virus. *Aspects Appl. Biol.* 39, 51–60.

Tsan, M., Totapally, S., Hailu, M., and Addom, B. K. (2019). *The Digitalisation of African Agriculture Report 2018-2019*. Chicago, IL: CTA.

Tuhaise, J., Quinn, J. A., and Mwebaze, E. (2014). "Pixel classification methods for automatic symptom measurement of cassava brown streak disease," in *Proceding of the 1st International Conference on the Use of Mobile ICT in Africa*, Stellenbosch.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Mrisho, Mbilinyi, Ndalahwa, Ramcharan, Kehs, McCloskey, Murithi, Hughes and Legg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 22 December 2020
doi: 10.3389/fpls.2020.600854

Fine-Grained Image Classification for Crop Disease Based on Attention Mechanism

Guofeng Yang[1,2,3], *Yong He*[1]*, *Yong Yang*[2,3] *and Beibei Xu*[2,3]

[1] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2] Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China, [3] Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, China

Fine-grained image classification is a challenging task because of the difficulty in identifying discriminant features, it is not easy to find the subtle features that fully represent the object. In the fine-grained classification of crop disease, visual disturbances such as light, fog, overlap, and jitter are frequently encountered. To explore the influence of the features of crop leaf images on the classification results, a classification model should focus on the more discriminative regions of the image while improving the classification accuracy of the model in complex scenes. This paper proposes a novel attention mechanism that effectively utilizes the informative regions of an image, and describes the use of transfer learning to quickly construct several fine-grained image classification models of crop disease based on this attention mechanism. This study uses 58,200 crop leaf images as a dataset, including 14 different crops and 37 different categories of healthy/diseased crops. Among them, different diseases of the same crop have strong similarities. The NASNetLarge fine-grained classification model based on the proposed attention mechanism achieves the best classification effect, with an F_1 score of up to 93.05%. The results show that the proposed attention mechanism effectively improves the fine-grained classification of crop disease images.

Keywords: crop disease, fine-grained, image classification, attention mechanism, fine-tuning

OPEN ACCESS

Edited by:
Spyros Fountas, Agricultural University of Athens, Greece

Reviewed by:
Shitala Prasad, NTU Institute of Structural Biology (NISB), Singapore
Ning Yang, Jiangsu University, China

***Correspondence:**
Yong He
yhe@zju.edu.cn

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *31 August 2020*
Accepted: *30 November 2020*
Published: *22 December 2020*

Citation:
Yang G, He Y, Yang Y and Xu B (2020) Fine-Grained Image Classification for Crop Disease Based on Attention Mechanism. Front. Plant Sci. 11:600854. doi: 10.3389/fpls.2020.600854

INTRODUCTION

Outbreaks of crop disease have a significant impact on the yield of agricultural production. Often, large-scale disease outbreaks destroy crops that have taken considerable efforts to grow, causing irreparable damage. Even without large-scale disease outbreaks, small-scale emergence can cause serious losses to crop yield and quality (Mutka and Bart, 2015). Therefore, developing techniques to accurately classify crop leaf disease categories is critical for disease prevention. With advances in image classification technology, researchers in the field of crop disease have gradually come to use deep learning approaches (Ramcharan et al., 2017; Fuentes et al., 2018; Liu B. et al., 2020). To date, research on the general classification of crop diseases has made several remarkable achievements in terms of better classification. However, for some fine-grained crop leaf diseases, there are still many difficulties.

Fine-grained image classification aims to classify sub-categories of a single larger category through fine-grained images (Peng et al., 2018). Examples include Stanford Cars (Yu et al., 2018; Tan and Le, 2019), CUB-200-2011 (Chen et al., 2019; Zhuang et al., 2020), FGVC Aircrafts

(Ding et al., 2019; Sun et al., 2020), and Oxford 102 Flowers (Dubey et al., 2018; Touvron et al., 2019). Fine-grained image classification models can be divided into algorithms based on strong supervision and algorithms based on weak supervision, which depends on how much supervision information can be used. For classification models based on strong supervision information, superior classification accuracy during model training requires artificial annotation information, such as object bounding boxes and part annotation, in addition to image-level category labels. Fine-grained image classification models based on weakly supervised information are similar, but also require the use of global and local information. Weakly supervised fine-grained classification attempts to achieve better local information capture without resorting to the key point information of object parts. As our goal is fine-grained image classification, we need to build a model that can identify the most discriminating image features. Therefore, it is vital to detect subtle discriminatory features from similar regions (Ou et al., 2016; Zhang et al., 2016). Because the occurrence of crop diseases is often not controlled by humans, the fine-grained classification of crop diseases is common, but remains challenging. In general, different sub-categories have very similar appearance, although occasionally the different sub-categories are completely inconsistent. More seriously, the many visual disturbances (such as reflection, dispersion, and blur) caused by dew, shooting jitter, and light intensity seriously reduce the classification accuracy of crop disease images (Lu et al., 2017).

In terms of both theoretical research and practical applications, the fine-grained image classification of crop leaf diseases is of great importance, and is thus the focus of this study. Many researchers have studied the classification of crop diseases based on pattern recognition and machine learning. Guo et al. (2014) utilized texture and color features using a Bayesian approach for recognizing downy mildew, anthracnose, powdery, and gray mold infection with respective accuracy levels of 94.0, 86.7, 88.8, and 84.4%. Zhang et al. (2017) developed a leaf disease identification application in cucumber plants. This application isolates the infected part of the leaf through k-means clustering before extracting the color and shape, resulting in an accuracy level of 85.7%.

Although the above methods have made some progress, the identification and classification of diseases of different crops under actual field conditions can be further improved. For example, although some models can achieve extremely high accuracy on datasets under laboratory conditions, they often have poor identification effects when faced with actual field conditions. We think this is because insufficient disease features are extracted, resulting in a lack of disease details. In summary, the main challenge of fine-grained image classification of crop leaf diseases is undoubtedly the subtle discrimination between different sub-categories. The primary difficulties can be roughly divided into three aspects: (1) the similarity between the sub-categories under the same disease category is very strong; (2) the field environment has significant background interference; and (3) the location of different crop diseases is inconsistent.

In an attempt to overcome these difficulties, many researchers have applied convolutional neural network (CNN) to crop disease classification. To investigate the impact of dataset size and species on the effectiveness of crop disease classification based on deep learning and transfer learning, Barbedo (2018) showed that, although CNNs can largely overcome the technical limitations associated with automated crop disease classification, training with a limited set of image data can have many negative consequences. Kaya et al. (2019) studied and demonstrated that the transfer learning model can help crop classification identification and improve the low-performance classification model. Too et al. (2019) fine-tuned and evaluated the most advanced deep CNN for image-based crop disease classification. The data used in their experiments covered 38 different categories, including disease and health images of the leaves of 14 crops from PlantVillage. The accuracy of DenseNet reached 99.75%, better than that of other models. Cruz et al. (2019) used CNNs to detect leaf images of Grapevine Yellows (GY) disease in red vines (cv. Sangiovese). ResNet-50 was found to be the best compromise network in terms of accuracy and training cost. Turkoglu et al. (2019) proposed a multi-model pre-trained CNN (MLP-CNN) based on long short-term memory for detecting apple diseases and insect pests. Their results were comparable to or better than those of pre-trained CNN models. Deep learning has been widely applied to various crop categories and crop disease classification studies, and deep learning models based on transfer learning can accelerate the training stage. At the same time, to cope with the impact of complex scenes on model classification performance, it is necessary to enhance the performance of CNNs to better handle fine-grained image classification tasks.

In recent years, it has been found that human cognitive processes do not focus attention on the entire scene at one time. On the contrary, they pay more attention to local regions in the scene while extracting relevant information. Models based on attention mechanisms have achieved good results on many challenging tasks, such as visual question answering (Malinowski et al., 2018), object detection (Li et al., 2019), and scene segmentation (Fu et al., 2019). Although the attention mechanism has been applied to different tasks, it has not been used for the fine-grained classification of crop disease images.

In this research, we propose a novel attention mechanism and use transfer learning to quickly build several fine-grained image classification models of crop diseases based on the attention mechanism, so as to solve the problem that the accuracy of CNN model in complex scenes is low due to visual interference in practical applications. Therefore, the contributions of this paper are as follows:

According to the characteristics of crop disease images in real scenes, a fine-grained fine-tuning classification algorithm based on attention mechanism is constructed on the basis of using pre-trained CNN to extract convolutional features of fine-grained images as the input of the network. The attention mechanism makes the classification algorithm pay more attention to some more discriminative local regions of the image, thereby improving the classification accuracy of the model in complex scenes.

We collect crop disease images in real scenes and added these images to the PlantVillage dataset to form a new

hybrid dataset for training the CNN model. We verify the effectiveness of our proposed method by designing multiple comparative experiments.

In addition, we also explore and prove the importance of the image source to the classification results, as well as the impact of problem scenes and special interference on the classification results.

The rest of this paper is organized as follows. Section Materials and methods introduces the main experimental dataset and our proposed fine-grained fine-tuning classification algorithm based on the attention mechanism. The experimental results are described in section Results. In section Discussion, we discuss the importance of training image sources, the impact of problem scenes and special interference on the classification results. Finally, this paper concludes in Section conclusion.

MATERIALS AND METHODS

Image Datasets

This study considered the PlantVillage public dataset of 52,629 images (except for images from the Tomato Two Spotted Spider Mite, Tetranychus urticae, category) (Mohanty et al., 2016), which covers a total of 37 categories including images of 14 different healthy or diseased crops. Using the Scrapy web crawler on the Internet's agricultural technology and consulting platforms, we then extracted a total of 5,571 images uploaded by users in the abovementioned 37 categories (images of crop diseases under actual field conditions). Finally, CNN models were trained and tested using the full dataset of 58,200 healthy and diseased crop disease images (**Figure 1**).

Table 1 provides statistical data for the 37 categories of the dataset, such as the number of images for each category and the percentage of images taken under laboratory or field conditions. It is well-known that there is a single color or no background in the image of laboratory conditions, while the background in the image of field conditions is relatively complex and changeable. As shown in **Table 1**, nearly 10% of the available images were taken under field conditions.

Figure 2 shows disease images of potato late blight, including four disease images obtained under field conditions and four disease images obtained under laboratory conditions. The increase in complexity of the four diseased images under field conditions is obvious (e.g., there are many leaves and other parts in the images, different backgrounds, shadow effects, and so on).

The dataset includes images taken under laboratory conditions and under field conditions (see **Figure 2**); the percentages of each are presented in **Table 1**. The whole dataset was randomly divided into a training set (80%) and a test set (20%). Therefore, 46,560 images were used for CNN model training, while the remaining 11,640 images were used to test the performance of the model. The training set and the test set were preprocessed to satisfy the model's input size requirements, and the image sizes were reduced and cropped to 256×256, 299×299, and 331×331 pixels.

We conducted several experiments to evaluate the importance of the conditions under which the leaf images were captured. Namely, we first conducted the training using only laboratory conditions images (PlantVillage, 52,629 photos) and the testing using images of actual field conditions (Internet, 5,571 photos), and then performed training using only images of actual field conditions (Internet, 5,571 photos) and testing with images taken under laboratory conditions (PlantVillage, 52,629 photos).

Experimental Methods and Parameters

Transfer Learning

VGG, ResNet, and other deep CNN models have achieved great success in image classification. The pre-trained deep CNN model has been fully trained on a large image dataset (ImageNet), allowing many features required for image classification to be learned. Therefore, we can use the idea of transfer learning to fully utilize the large amount of knowledge learned by pre-training the CNN model on the ImageNet dataset, and apply it to crop disease image classification. This paper describes how the transfer learning method of parameter transfer was adopted to remove the maximum pooling and fully connected layers after the final convolution, and introduces a new fine-grained classification model based on the attention mechanism. Compared with the random initialization of the weight parameters of each layer of the network, the fine-tuning method helps accelerate the convergence of the network.

For image classification, there are several CNN baseline models that have been successfully applied to specific tasks. Regarding the task of image recognition and classification of crop diseases, six CNN models that were pre-trained using ImageNet have been applied: (1) VGG16 and VGG19 (Simonyan and Zisserman, 2015), (2) ResNet50 (He et al., 2016), (3) InceptionV3 (Szegedy et al., 2016), (4) Xception (Chollet, 2017), and (5) NASNetLarge (Zoph et al., 2018). The training and testing processes of these pre-trained models and of the proposed fine-grained pre-trained model based on the attention mechanism were implemented using the TensorFlow machine learning computing framework. Model training and testing was conducted with four NVIDIA Tesla V100 GPUs.

Attention Mechanism

The attention mechanism was first applied to natural language processing. It is often combined with recurrent neural networks, resulting in good prediction and processing ability for text sequences. In recent years, the attention mechanism has also been widely used for image classification (Meng and Zhang, 2019; Xiang et al., 2020), object detection (Chen and Li, 2019; Xiao et al., 2020), and image description generation (Liu M. et al., 2020; Zhang et al., 2020). In the field of crop disease classification, most researchers have tended to use transfer learning technology. There has also been some research on crop disease identification based on the attention mechanism (Nie et al., 2019; Karthik et al., 2020). These previous studies have focused on a certain crop, and so the disease category and scale of the dataset are limited. Therefore, we conducted related experiments to verify that increasing the attention mechanism can improve the effect of crop disease classification based on transfer learning technology.

The attention of an image refers to the process of obtaining a target region that requires attention as the human eye rapidly scans the global image. This target region is assigned more

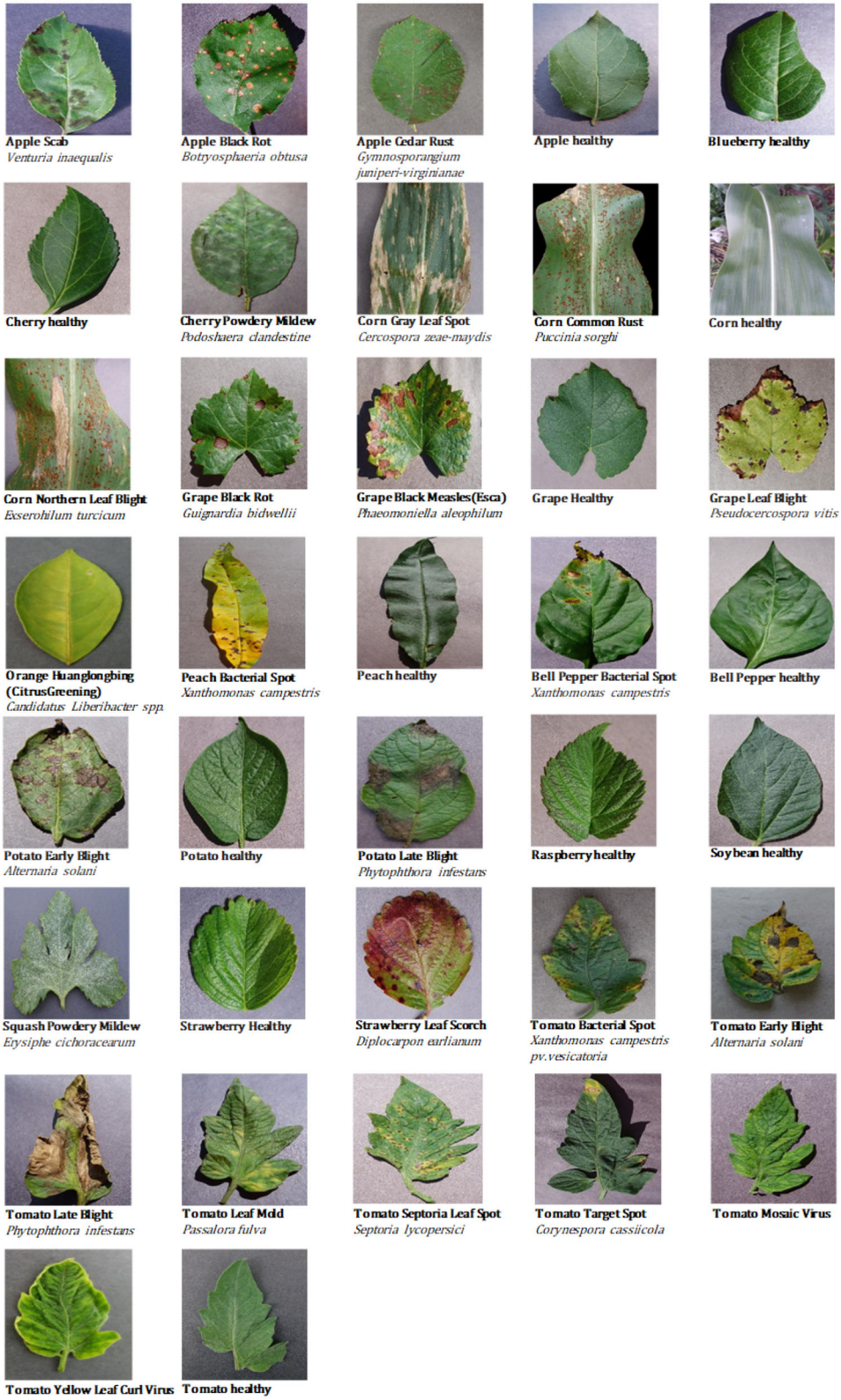

FIGURE 1 | Examples of crop leaf images in the dataset.

TABLE 1 | Statistics of crop healthy/diseased images and related data.

Class	Plant common name	Disease common name	Disease scientific name	Images (PlantVillage)	Laboratory conditions (%)	Field conditions (%)
1	Apple	Apple scab	*Venturia inaequalis*	800 (630)	78.75%	21.25%
2	Apple	Black rot	*Botryosphaeria obtuse*	800 (621)	77.63%	22.38%
3	Apple	Cedar apple rust	*Gymnosporangium juniperi-virginianae*	400 (275)	68.75%	31.25%
4	Apple	—	—	1,800 (1,645)	91.39%	8.61%
5	Blueberry	—	—	1,700 (1,502)	88.35%	11.65%
6	Cherry (and sour)	—	—	1,000 (854)	85.40%	14.60%
7	Cherry (and sour)	Powdery mildew	*Podosphaera* spp.	1,200 (1,052)	87.67%	12.33%
8	Com (maize)	Cercospora leaf spot	*Cercospora zeae-maydis*	700 (513)	73.29%	26.71%
9	Com (maize)	Common rust	*Puccinia sorghi*	1,300 (1,192)	91.69%	8.31%
10	Com (maize)	—	—	1,300 (1,162)	89.38%	10.62%
11	Com (maize)	Northern Leaf Blight	*Exserohilum turcicum*	1,100 (985)	89.55%	10.45%
12	Grape	Black rot	*Guignardia bidwellii*	1,300 (1,180)	90.77%	9.23%
13	Grape	Esca (Black measles)	*Phaeomoniella chlamydospora*	1,500 (1,383)	92.20%	7.80%
14	Grape	—	—	600 (423)	70.50%	29.50%
15	Grape	Leaf blight	*Pseudocercospora vitis*	1,200 (1,076)	89.67%	10.33%
16	Orange	Huanglongbing	Candidatus Liberibacter	5,700 (5,507)	96.61%	3.39%
17	Peach	Bacterial sport	*Xanthomonas campestris*	2,400 (2,297)	95.71%	4.29%
18	Peach	—	—	500 (360)	72.00%	28.00%
19	Pepper, bell	Bacterial spot	*Xanthomonas campestris*	1,100 (997)	90.64%	9.36%
20	Pepper, bell	—	—	1,600 (1,478)	92.38%	7.63%
21	Potato	Early blight	*Altemaria solani*	1,200 (1,000)	83.33%	16.67%
22	Potato	—	—	300 (152)	50.67%	49.33%
23	Potato	Late blight	*Phytophthora infestans*	1,200 (1,000)	83.33%	16.67%
24	Raspberry	—	—	500 (371)	74.20%	25.80%
25	Soybean	—	—	5,200 (5,090)	97.88%	2.12%
26	Squash	Powdery mildew	*Erysiphe cichoracearum*, *Sphaerotheca fuliginea*	2,000 (1,835)	91.75%	8.25%
27	Strawberry	—	—	600 (456)	76.00%	24.00%
28	Strawberry	Leaf scorch	*Diplocarpon earlianum*	1,300 (1,109)	85.31%	14.69%
29	Tomato	Bacterial spot	*Xanthomonas campestris* pv. Vesicatoria	2,300 (2,127)	92.48%	7.52%
30	Tomato	Early blight	*Altemaria solani*	1,200 (1,000)	83.33%	16.67%
31	Tomato	Late blight	*Phytophthora infestans*	2,100 (1,909)	90.90%	9.10%
32	Tomato	Leaf Mold	*Fulvia fulva*	1,100 (952)	86.55%	13.45%
33	Tomato	Septoria leaf spot	*Septoria lycopersici*	1,900 (1,771)	93.21%	6.79%
34	Tomato	Target spot	*Corynespora cassiicola*	1,600 (1,404)	87.75%	12.25%
35	Tomato	Tomato mosaic virus	*Tomato mosaic virus* (ToMV)	500 (373)	74.60%	25.40%
36	Tomato	TYLCV	Begomovirus (Fam. Geminiviridae)	5,500 (5,357)	97.40%	2.60%
37	Tomato	—	—	1,700 (1,591)	93.59%	6.41%
TOTAL:				58,200 (52,629)	90.43%	9.57%

Images column gives the number of images in each category, where the figures in parentheses are from the public dataset.

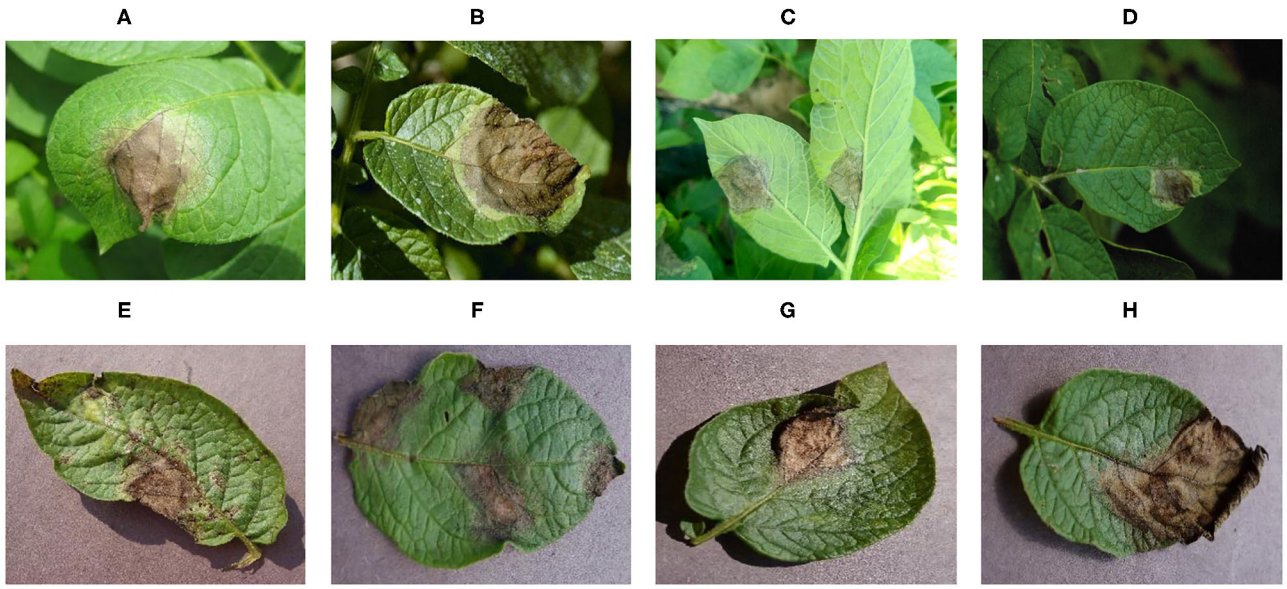

FIGURE 2 | Image sample of potato late blight. **(A–D)** Field conditions, **(E–H)** laboratory conditions.

attention (weight distribution) to obtain the required detailed information about the target, with other useless information suppressed. Soft attention is most commonly used, because this is a completely differentiable process that can realize end-to-end learning in CNN models. Most soft attention models learn an attention template to align the weights of different regions in a sequence or an image and use this template to locate the distinguishable regions. Different from soft attention, the hard attention mechanism is a random, non-differentiable process that determines the importance of individual regions one at a time, rather than identifying the important regions within the whole image.

For image classification, the weight of the arithmetic mean of attention can be extracted through attention learning to form the attention spectrum of the image. Similar to traditional natural language processing, the image-based attention can be obtained through the model illustrated in **Figure 3**.

In **Figure 3**, I is the input image. The attention model has n parameters, $a_1, a_2, \cdots, a_i, a_n$, which respectively represent a description of each part of the image. O is the return value of the model's attention spectrum (more specifically, the weight values of the n parameters), which is determined from the importance of each a_i relative to the input I. By filtering the input image through this output, the region that requires most attention can be identified.

Proposed Model

Based on the pre-trained model described in Section transfer learning and the attention mechanism introduced in Section attention mechanism, this paper proposes a fine-grained classification model based on the attention mechanism (**Figure 4**). By learning the attention of the CNN feature spectrum, the attention model calculates and identifies the most important region of the feature spectrum for the final classification task, and provides the maximum attention input (weight distribution). However, adding the attention weight to the last layer of the CNN features will cause different degrees of suppression of the original features. To overcome this suppression, the weighted feature spectrum is added to the original feature spectrum. The fusion spectrum is then input into the fully connected layer. In the second fully connected layer, the attention feature spectrum transformed by the global average pooling dimension is connected with the fully connected feature spectrum in the channel direction, before being sent to the classification layer for classification.

The attention model proposed in this paper adopts an unsupervised training mode. There is no pre-labeled ground truth to constrain the attention spectrum, so there is no separate loss calculation. Instead, a backpropagation adaptive mode is used to constrain the weight distribution of attention. The loss function defined in this paper is expressed as:

$$loss = \frac{1}{2n} \sum ||y_{truth}(x) - y_{pred}(x)||^2 \quad (1)$$

where $n = 37$ is the number of input samples, x is the input sample, y_{truth} is the actual category, and y_{pred} is the predicted category output by the final layer of the network. In the process of backpropagation, the output error of the *Softmax* layer is backpropagated, and the parameters are updated using the random gradient descent method, so that the final loss function value decreases and the network converges.

Through the soft attention mechanism, the output of the final convolutional layer of the CNN is obtained. This is taken as the input of the attention model, and the corresponding attention

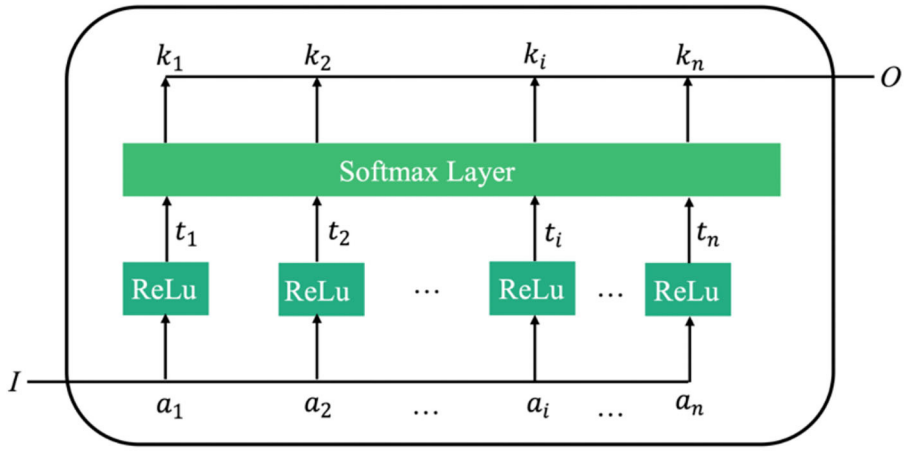

FIGURE 3 | Schematic diagram of learning mechanism of attention.

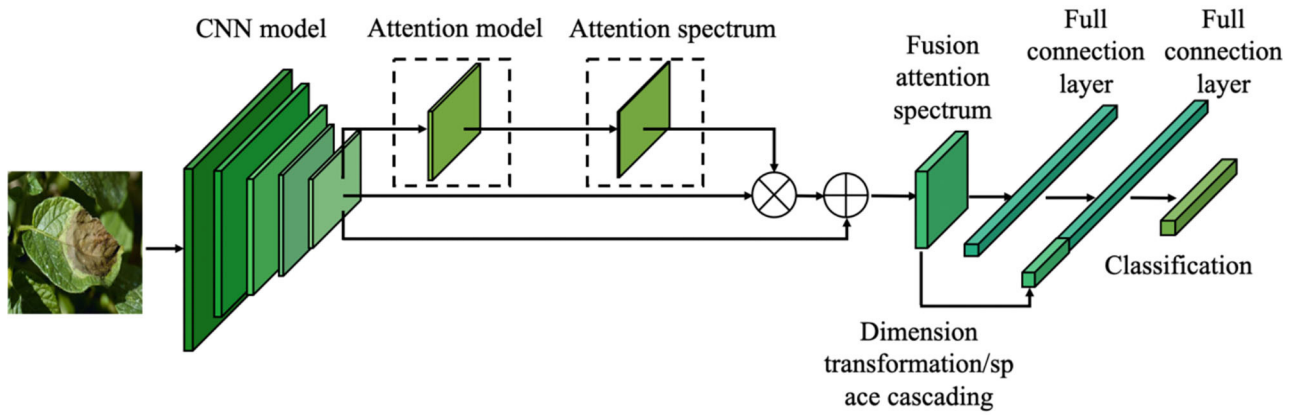

FIGURE 4 | Fine-grained image classification network structure of crop diseases based on attention mechanism.

spectrum is calculated. The original feature spectrum is then weighted by the attention spectrum, and the output attention feature spectrum is provided as the input for the subsequent network. Let the feature spectrum of the output of the final convolution spectrum after the pooling operation be expressed as $f \in \mathbb{R}^{H\times W\times C}$, where H and W refer to the height and width of the feature spectrum of this layer, C refers to the number of channels of the feature spectrum of this layer, and for each position (m, n) on the spectrum, its feature value is expressed as $f_{m,n}\epsilon\mathbb{R}^C$. The corresponding attention weight $W_{m,n}$ can then be obtained as:

$$W_{m,n} = ATT(f_{m,n};\ W_{att}) \tag{2}$$

where ATT is a mapping function learned by the attention model and W_{att} is the weight parameter of the attention model. Through $Softmax$ regression of $w_{m,n}$, the final attention spectrum $M = [M_{m,n}]$ is obtained as a normalized probability matrix, where $M_{m,n}$ is expressed as:

$$M_{m,n} = Softmax(W_{m,n}) \tag{3}$$

As can be seen from **Figure 5**, the attention model proposed in this paper takes the output of the final convolution spectrum in the neural network as its input. The attention model includes two convolutional layers and one *Softmax* layer. The kernel sizes of the convolutional layers are 3×3 and 1×1. The attention feature spectrum $f^{att} = [f^{att}_{m,n}]$ is obtained by multiplying the attention spectrum M by the CNN feature spectrum f, and is expressed as

$$f^{att} = \left[f^{att}_{m,n}\right] = [M_{m,n} \cdot f_{m,n}] \tag{4}$$

The 3×3 convolution kernel further extracts the CNN feature of the final convolution spectrum. In order not to reduce the feature receptive field and feature information, a convolution kernel with the same size as the original network is selected. Compared with the 3×3 convolution kernel, the 1×1 convolution kernel enables information interaction and integration across channels. By connecting features in the channel direction, nonlinear components can be added to features to improve the feature expression ability of the attention model.

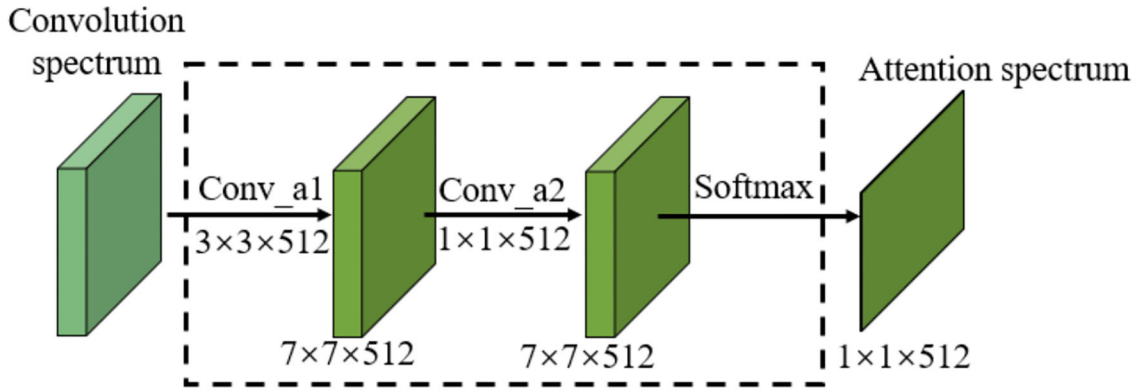

FIGURE 5 | Attention calculation model for images.

The attention spectrum of the final convolution spectrum of the CNN is obtained through the attention model, and the attention spectrum and the original CNN feature spectrum are then multiplied to obtain the attention feature spectrum. The attention spectrum is the spectrum obtained after normalizing the weights of the features. According to this definition, the attention feature spectrum obtained after multiplication has a certain attenuation compared with the original CNN feature spectrum. Additionally, during the convolution and probability calculation, the spatial transformation of the CNN feature spectrum and noise addition means that the calculated attention spectrum may be distorted. In this case, the obtained attention spectrum has no guiding significance for the original image spectrum. To overcome this problem, once the attention feature spectrum has been obtained, the original CNN feature spectrum is added and fused to obtain the final attention f_{all}^{att}, which is input into the subsequent fully connected layer, as shown in Equation (5).

$$f_{all}^{att} = f + M \cdot f \tag{5}$$

By adding the attention spectrum to the CNN feature spectrum, the distortion of the attention spectrum is overcome and the original feature spectrum before the fully connected layer can be effectively utilized.

By extracting and merging the attention spectrum, a spectrum of features is obtained that is well-located and noticeable in space. This spectrum is then input to the subsequent fully connected layer. As the connection operation maps the convolutional spectra of all channels to one point in the fully connected layer, the spatial information is destroyed by the operation of the fully connected layer. The original intention of introducing the attention model is to extract and improve the significant regions of the CNN feature spectrum in space. However, after the fully connected layer, the spatial information of the extracted attention feature has also been destroyed. Therefore, the attention space feature is reused by connecting the attention feature spectrum in the final fully connected layer, as shown in **Figure 6**.

Evaluation of the Model

The accuracy, precision (P), recall (R), and comprehensive F_1 evaluation index were used to evaluate the crop disease image classification model. The F_1 value is the harmonic average of the precision and recall, and has a maximum of 1 and a minimum of 0. It is calculated as follows:

$$F_1 = \frac{2PR}{P+R} \times 100\% \tag{6}$$

Experimental Details

In all our experiments, we preprocessed the images to sizes of 256×256, 299×299, and 331×331 pixels, conducted a total of 1,000 training epochs, and used a batch size of 32. We used a momentum SGD initial learning rate of 0.001. When the standard evaluation stopped increasing, the learning rate was multiplied by 0.1 until it had dropped to 0.0001. After lowering the learning rate, we waited for five epochs before returning to normal operation. If the loss of the test set did not improve after 20 epochs, the learning rate was reduced. We conducted experiments using multiple pre-trained models, all of which are robust to the selection of hyper-parameters.

RESULTS

Compared With the Pre-training Model and the Effect of the Attention Mechanism

Tables 2, **3** present the classification accuracy, precision, recall, and F_1 values of various models on the test set. The results indicate that the fine-grained fine-tuning classification models based on the attention mechanism outperform the original pre-trained models by 1–2% in terms of accuracy, precision, recall, and F_1 value. This demonstrates that the attention mechanism improves the classification performance of the models and allows them to focus on key regions in the image. The fine-grained NASNetLarge model based on the fine-grained attention mechanism achieves the highest accuracy, precision, recall, and F_1 values, and thus provides the best classification performance. These 12 models were further trained using only the original image to record the training period for the best performance. As

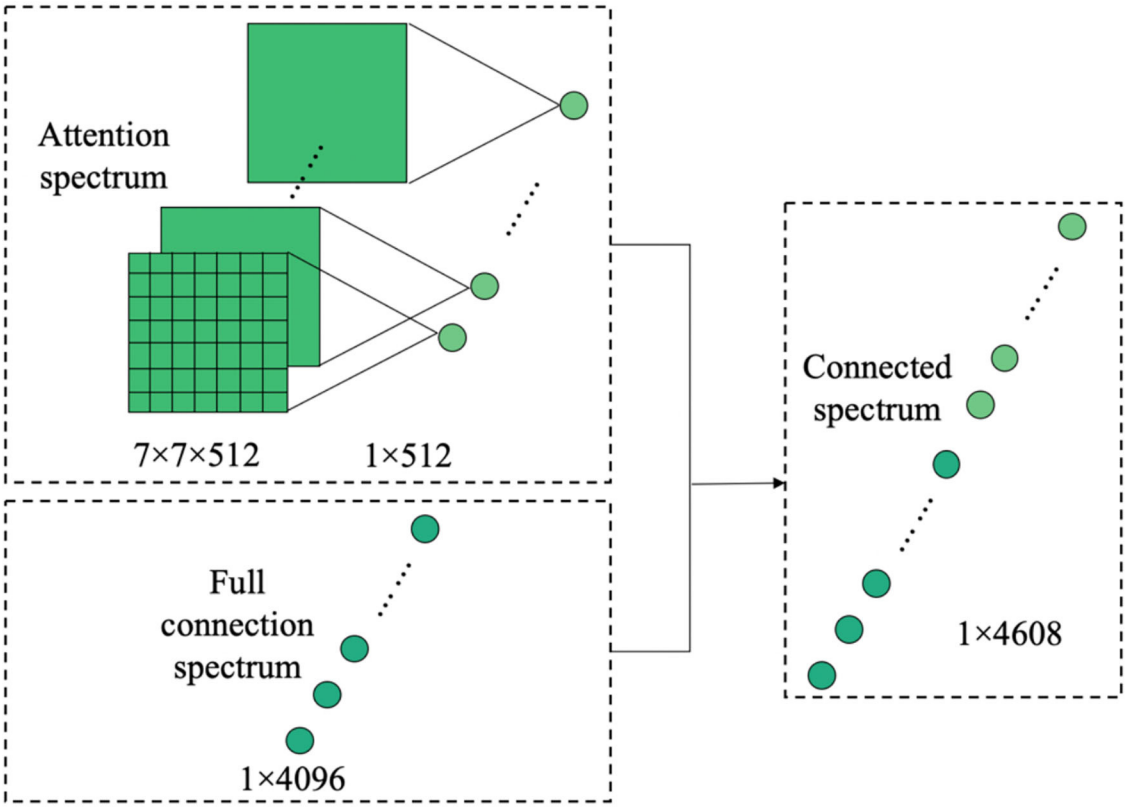

FIGURE 6 | Dimension transformation and connection diagram of attention feature spectrum.

shown in **Table 3**, the fine-grained NASNetLarge model based on the attention mechanism achieves the highest classification accuracy of 95.62%. Thus, this model was used in subsequent experiments for crop disease image classification.

Figure 7 provides a visual representation of some random images from the test set. The table on the left of the original image shows the predicted classification. The image to the right of the original image is a visual representation of the attention mechanism using the fine-grained fine-tuning NASNetLarge model based on the attention mechanism. The highest-ranked classification result for each image was considered as the final classification result predicted by the model. The images of the crop leaves shown in **Figure 7** are correctly classified. In most cases, the degree of certainty for the correct classification is close to 100%, so there is no actual ranking.

Testing on Different Dataset

We also comprehensively evaluated our algorithm on public plant datasets of Flavia (Wu et al., 2007), Swedish Leaf (Söderkvist, 2001), and UCI Leaf (Silva et al., 2013). These datasets contain clear images, and they are widely used datasets in this field, often used for algorithm development and comparison. The statistics of three datasets are shown in **Table 4**. We follow the same training/test split as in Section image datasets.

The Flavia dataset contains 1,907 images of 32 species of plants. All images in the dataset have a white background, and the number of each category varies from dozens of images and is relatively unbalanced.

The Swedish Leaf dataset contains 15 plant species, with 75 images in each category. All plant leaf images are images with white background, and the quality and resolution of each image is high.

The UCI Leaf dataset contains 40 different plants and a total of 443 images. The background colors of the images in this dataset are all pink. The number of images in each category ranges from a few to a dozen.

As seen from **Table 5** the NASNetLarge model based on the attention mechanism constructed by our proposed method can still get the best classification accuracy on the three public plant datasets. Therefore, it can be proved that our model has better performance across datasets and can achieve efficient classification on datasets of different sizes.

Validation and Comparison of Proposed CNN With Traditional Machine Learning Models

The traditional machine learning methods used for comparison in this paper are SVM, Decision tree, k-NN, and Naive Bayes. Features such as Hu-moments, Haralick features, LBP features,

TABLE 2 | Results of pre-trained model on test set.

Pre-trained model	Size(MB)	Accuracy (%)	Precision (%)	Recall (%)	F1-measure (%)	Parameters
VGG16 (Simonyan and Zisserman, 2015)	56.16	83.07	82.52	80.13	81.31	15,360,589
VGG19 (Simonyan and Zisserman, 2015)	76.42	85.79	83.45	81.75	82.59	20,670,285
ResNet50 (He et al., 2016)	90.38	85.82	85.21	83.63	84.41	25,769,613
InceptionV3 (Szegedy et al., 2016)	83.84	88.47	87.78	85.47	86.61	23,984,685
Xception (Chollet, 2017)	79.81	91.22	90.24	87.05	88.62	23,043,381
NASNetLarge (Zoph et al., 2018)	327.69	92.78	92.16	90.83	91.49	89,082,719

TABLE 3 | Results of fine-grained classification model based on attention on test set.

Model based on attention mechanism	Size(MB)	Accuracy (%)	Precision (%)	Recall (%)	F1-measure (%)	Parameters
VGG16*	59.23	85.53	84.68	81.32	82.97	15,514,783
VGG19*	79.75	86.40	84.93	82.05	83.47	20,823,653
ResNet50*	93.91	87.03	86.27	85.32	85.79	25,942,746
InceptionV3*	87.53	90.64	88.51	88.48	88.49	24,173,257
Xception*	83.28	92.89	91.82	90.95	91.38	23,218,425
NASNetLarge*	331.75	95.62	94.35	91.79	93.05	89,286,814

*The * indicates that the pre-training model is used.*

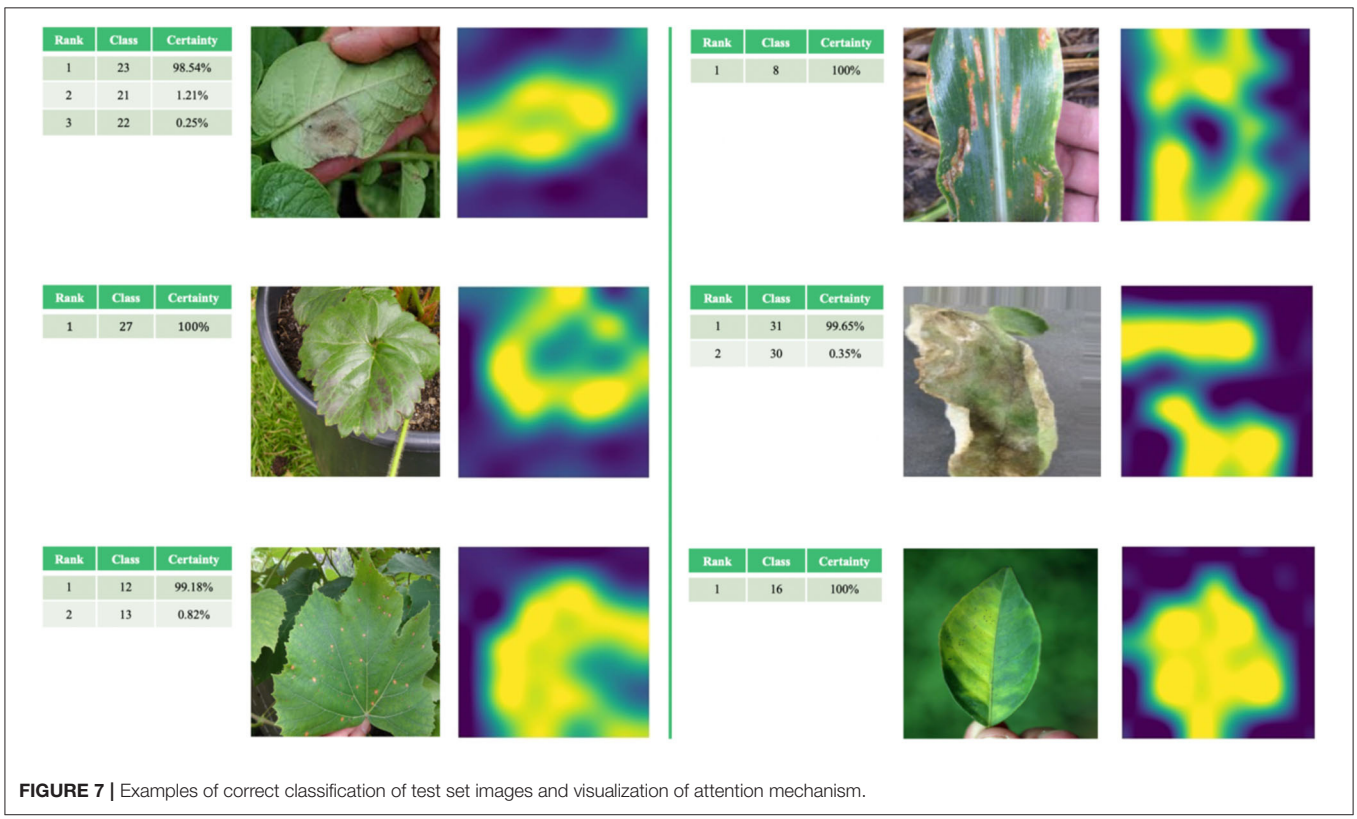

FIGURE 7 | Examples of correct classification of test set images and visualization of attention mechanism.

and HSV features have been used to evaluate the performance of all traditional Machine Learning algorithms. The results are given in **Table 6**.

Table 6 shows that the accuracy, precision, recall and F_1 of our proposed model are much higher than those obtained using other machine learning algorithms.

TABLE 4 | Statistics of benchmark datasets.

Datasets	Class	Train	Test
Flavia	32	1,526	381
Swedish leaf	15	900	225
UCI leaf	40	356	87

TABLE 5 | Comparison of methods for image classification in three datasets.

References	Datasets	Accuracy/%	Method
Lee et al. (2017)	Flavia	99.40	CNN, Fine-tuning
Yousefi et al. (2017)	Flavia	97.50	Fourier and Wavelet Descriptors, MLP
Murat et al. (2017)	Flavia	95.25	HOG, Moments, ANN,
	Swedish	99.89	RF, and SVM
Kaya et al. (2019)	Flavia	99.00	DF – VGG16/LDA
	Swedish	98.80	CNN – RNN
	UCI Leaf	96.20	DF – Alexnet/LDA
Our	Flavia	99.72	NASNetLarge –
	Swedish	99.90	Attention
	UCI Leaf	98.74	

DISCUSSION

Importance of Training Image Type

The fine-grained NASNetLarge model based on the attention mechanism produced the best classification effect, and was therefore further tested to study the importance of assessing the conditions under which the leaf image was captured. The corresponding results are presented in **Table 7**.

When only laboratory or field condition images are used for training, the accuracy on the test set is significantly lower than when both laboratory and field condition images are used for training. The results show that the model can obtain better performance when using images obtained under field conditions for training and requiring classification of laboratory condition images (F_1 value is nearly 60.47%). In contrast, when the laboratory condition images are used for training and the field condition images are classified, the classification performance is obviously reduced (F_1 value is about 34.11%). This shows that image classification under field conditions is a more difficult and complicated task than the classification of images obtained under laboratory conditions, and proves that the construction of an efficient automatic detection and diagnosis system for crop diseases using images obtained under field conditions is of great significance.

Problematic Situations and Indicative Cases

The fine-grained NASNetLarge model based on the attention mechanism reached an accuracy level of 95.62% on the test set of 11,640 images, of which 11,130 images were correctly classified. Among the 4.38% of misclassified images, there are some "problematic" images that do not contain crop leaves at all (as shown in **Figures 8A,B**). These images should be classified into category 31 (Tomato late blight, phytophthora infestans), but the model classifies their predictions into category 10 (Corn healthy), as shown in the classification table in **Figure 8**. The classification table shows the predicted classification ranking output of the final model on the original image. These images are misclassified by the model (the "correct" classification would be category 31). In fact, they do not belong to any category, because there are no crop leaves in the image. However, they are all classified as category 10. We infer that the images in category 10 (**Figures 8C,D**) contain similar soils, while the corn leaves are very slender and occupy a small portion of the image. If such problematic examples are excluded, the accuracy of the final model will be higher than 95.62%.

There are several other problems with the images obtained under field conditions, including: (1) shadows on the leaves in the images, with some images appearing dark and shaky; (2) other objects in the image that are not related to the leaf itself, such as a trunk, fruit, or fence. Note that these problematic images occupy a very small portion of the dataset. In short, according to the certainty levels provided by the final model, the attention mechanism-based approach proposed in this paper overcomes these problems in most cases.

A typical case is category 8 (Corn cercospora leaf spot, cercospora zeae-maydis). **Figure 9** shows the classification results of the model for eight representative images of category 8, including four incorrectly classified images (the lower four images in **Figure 9**) and four correctly classified images (the upper four images in **Figure 9**). The first three upper images were correctly classified, with a certainty of ∼100%, while the fourth image was correctly classified with a certainty of ∼79% (the second ranking, with a certainty level of 16% for category 10, corresponds to corn crops with different diseases). The four misclassified images (the lower four images in **Figure 9**) include a wide range of partial shadows or complex backgrounds, which increase the misclassification rate of the model. For the middle two lower images, the correct classification is ranked second, while the first ranking is category 10 or 11 (corn crop diseases). Therefore, the model correctly identifies the crop species, but does not accurately detect the particular crop disease.

CONCLUSIONS

This study constructed, trained, and tested a fine-grained neural network model based on the attention mechanism for the classification of simple leaves of healthy or diseased crops. The model was trained using 58,200 publicly available images obtained under both laboratory conditions and field conditions. The data include 14 crop species in 37 different categories of [crop, disease] combinations, including some healthy crops. The optimal model was found to be a fine-grained NASNetLarge neural network based on the attention mechanism, which achieved an accuracy level of 95.62% (precision 94.35%, recall 91.79%, F_1 value 93.05%) in the classification of the 11,640 images

TABLE 6 | Comparison of methods for image classification in three datasets.

Models	Features	Accuracy/%	Precision/%	Recall/%	F1-measure/%
Naïve Bayes	Haralick, Hu	31.67	17.11	12.23	14.26
	Hu-moments, HSV	35.61	20.92	15.56	17.84
	Haralick, Hu, HSV	39.08	29.33	20.55	24.17
	Haralick, Hu, HSV, LBP	43.24	36.76	25.64	30.21
Decision tree	Haralick, Hu	41.33	23.24	18.85	20.82
	Hu-moments, HSV	47.18	30.52	25.21	27.61
	Haralick, Hu, HSV	53.91	43.45	34.69	38.58
	Haralick, Hu, HSV, LBP	60.23	48.32	41.36	44.57
SVM	Haralick, Hu	43.33	30.42	23.28	26.38
	Hu-moments, HSV	52.67	41.23	33.01	36.66
	Haralick, Hu, HSV	55.82	45.35	36.82	40.64
	Haralick, Hu, HSV, LBP	61.45	54.51	43.24	48.23
k-NN	Haralick, Hu	69.66	65.23	53.72	58.92
	Hu-moments, HSV	75.38	73.57	61.34	66.90
	Haralick, Hu, HSV	79.52	77.41	66.38	71.47
	Haralick, Hu, HSV, LBP	84.45	80.65	74.77	77.60
Our model (NASNetLarge*)		95.62	94.35	91.79	93.05

*The * indicates that the pre-training model is used.*

TABLE 7 | Results of using different training sets and test sets with the optimal model under laboratory conditions and field conditions.

Model based on attention mechanism	Training: laboratory Testing: actual field conditions				Training: actual field conditions Testing: laboratory			
	Accuracy (%)	Precision (%)	Recall (%)	F1-measure (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-measure (%)
NASNetLarge*	38.52	34.28	33.95	34.11	66.85	61.32	59.64	60.47

*The * indicates that the pre-training model is used.*

in the test set. The fine-grained NASNetLarge neural network model based on the attention mechanism achieves excellent classification performance by analyzing simple leaf images, so it is highly suitable for the automatic detection and diagnosis of crop diseases. In addition, the experimental results show that the images taken under field conditions in the training set are of high importance, indicating that when training such models, the proportion of images obtained under field conditions in the training set should be carefully considered.

For the backbone network of NASNetLarge, the results show that the NASNetMobile neural network model, similar to NASNetLarge, achieves state-of-the-art classification results on related datasets, surpassing the performance of previous lightweight networks such as MobileNet (Sandler et al., 2018) and ShuffleNet (Ma et al., 2018). As NASNetMobile requires little computing power to classify the given images, it can run on mobile devices such as smartphones, drones, or automatic agricultural vehicles for real-time monitoring and disease identification of large open-air crops. At present, due to the large-scale application of 5G, high-efficiency transmission, and improvements to the hardware configuration of mobile terminal equipment, it is possible to upload images locally to the cloud server for processing, and then return the identification and classification results to the terminal (Johannes et al., 2017; Toseef and Khan, 2018; Picon et al., 2019), or to use a GPU/CPU at the terminal to process and display the results (Barman et al., 2020). For growers in remote areas, real-time detection and diagnosis can be carried out through mobile terminals, thus solving the practical problems of obtaining technical crop disease diagnosis and finding experts in the production process. For agricultural technicians, this is equivalent to having a valuable auxiliary consultation tool. In the future, an intelligent crop disease prevention and control recommendation system will be developed based on the results of real-time diagnosis, allowing growers to select different prevention and control methods (e.g., physical or chemical methods) according to the specific conditions. The process and dosage of the methods will also be described in detail. Intelligent crop disease identification and diagnosis, as well as intelligent crop disease prevention and recommendation, will greatly improve production efficiency, realize agricultural, scientific, and technological progress, and push agriculture into the intelligent era.

Although the system developed in this study achieved a high success rate, it is far from becoming a universal tool under actual field conditions (Boulent et al., 2019). At present, the existing research has only considered dozens of [crop, disease] combinations (Ferentinos, 2018), so it is vital to expand the existing database to include more crop species

A

Rank	Class	Certainty
1	10	72.06%
2	11	8.21%
3	8	6.47%
4	26	5.92%
5	32	3.69%
6	31	2.63%

B

C

Rank	Class	Certainty
1	10	86.31%
2	11	3.04%
3	8	2.95%
4	30	1.36%
5	31	~0%

D

FIGURE 8 | (A,B): Images in category 31 and their corresponding "classification" results. **(C,D)**: representative images in category 10.

FIGURE 9 | Examples of correct (upper four images) and incorrect (lower four images) classification of category eight images in the test set (*corn cercospora leaf spot*).

and corresponding diseases. The test set used to evaluate the model was part of the dataset from which the training set was extracted, which is a potential source of bias. This is a common method for training and testing machine learning models. However, to develop a system that can be used effectively in field scenes, data from various sources should

be used for testing to ensure that future users can obtain effective classification results in different scenes (Barbedo, 2018). At present, some preliminary experiments carried out with limited data show that when testing images different from those used in training, the classification performance of the model is significantly reduced to the range of 25–35%. The experimental results show that the classification effect depends on the data source. To improve this, more extensive image datasets should be collected from different geographical areas, field conditions, image capture modes, and multiple sources. Improving the model by increasing the size of the dataset would allow more effective and widespread identification of crop categories and diseases under field conditions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

GY and YH: conceptualization. GY: methodology, software, formal analysis, data curation, writing—original draft preparation, and visualization. GY and BX: validation. GY and YY: investigation. YH and YY: resources, supervision, and project administration. BX: writing—review and editing. All authors contributed to the article and approved the submitted version.

REFERENCES

Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53. doi: 10.1016/j.compag.2018.08.013

Barman, U., Choudhury, R. D., Sahu, D., and Barman, G. G. (2020). Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease. *Comput. Electron. Agric.* 177:105661. doi: 10.1016/j.compag.2020.105661

Boulent, J., Foucher, S., Théau, J., and St-Charles, P.-L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10:941. doi: 10.3389/fpls.2019.00941

Chen, H., and Li, Y. (2019). Three-stream attention-aware network for RGB-D salient object detection. *IEEE Trans. Image Process.* 28, 2825–2835. doi: 10.1109/TIP.2019.2891104

Chen, Y., Bai, Y., Zhang, W., and Mei, T. (2019). "Destruction and construction learning for fine-grained image recognition," in *Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA: IEEE), 5152–5161. doi: 10.1109/CVPR.2019.00530

Chollet, F. (2017). "Xception: deep learning with depthwise separable convolutions," in *Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: IEEE), 1800–1807. doi: 10.1109/CVPR.2017.195

Cruz, A., Ampatzidis, Y., Pierro, R., Materazzi, A., Panattoni, A., Bellis, L. D., et al. (2019). Detection of grapevine yellows symptoms in *Vitis vinifera* L. with artificial intelligence. *Comput. Electron. Agric.* 157, 63–76. doi: 10.1016/j.compag.2018.12.028

Ding, Y., Zhou, Y., Zhu, Y., Ye, Q., and Jiao, J. (2019). "Selective sparse sampling for fine-grained image recognition," *Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA: IEEE), 6599-6608. doi: 10.1109/ICCV.2019.00670

Dubey, A., Gupta, O., Guo, P., Raskar, R., Farrell, R., and Naik, N. (2018). "Pairwise confusion for fine-grained visual classification," in *European Conference on Computer Vision (ECCV)* Lecture Notes in Computer Science, Vol. 11216 (Munich: Springer), 71–88. doi: 10.1007/978-3-030-01258-8_5

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., et al. (2019). Dual attention network for scene segmentation," in *Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA: IEEE), 3146–3154. doi: 10.1109/CVPR.2019.00326

Fuentes, A. F., Yoon, S., Lee, J., and Park, D. S. (2018). High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9:1162. doi: 10.3389/fpls.2018.01162

Guo, P., Liu, T., and Li, N. (2014). Design of automatic recognition of cucumber disease image. *Inform. Technol. J.* 13, 2129–2136. doi: 10.3923/itj.2014.2129.2136

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE), 770–778. doi: 10.1109/CVPR.2016.90

Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A. D., et al. (2017). Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput. Electron. Agric.* 138, 200–209. doi: 10.1016/j.compag.2017.04.013

Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., and Menaka, R. (2020). Attention embedded residual CNN for disease detection in tomato leaves. *Appl. Soft Comput.* 86:105933. doi: 10.1016/j.asoc.2019.105933

Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., and Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. *Comput. Electron. Agric.* 158, 20–29. doi: 10.1016/j.compag.2019.01.041

Lee, S. H., Chan, C. S., Mayo, S. J., and Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recogn.* 71, 1–13. doi: 10.1016/j.patcog.2017.05.015

Li, L., Xu, M., Wang, X., Jiang, L., and Liu, H. (2019). "Attention based glaucoma detection: a large-scale database and CNN model," *Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA: IEEE), 10571–10580. doi: 10.1109/CVPR.2019.01082

Liu, B., Ding, Z., Tian, L., He, D., Li, S., and Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Front. Plant Sci.* 11:1082. doi: 10.3389/fpls.2020.01082

Liu, M., Li, L., Hu, H., Guan, W., and Tian, J. (2020). Image caption generation with dual attention mechanism. *Inform. Process. Manage.* 57:102178. doi: 10.1016/j.ipm.2019.102178

Lu, Y., Yi, S., Zeng, N., Liu, Y., and Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. doi: 10.1016/j.neucom.2017.06.023

Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). ShuffleNet V2: practical guidelines for efficient CNN architecture design," in *Proceedings of the 15th European Conference* (Munich, Germany), 122–138. doi: 10.1007/978-3-030-01264-9_8

Malinowski, M., Doersch, C., Santoro, A., and Battaglia, P. W. (2018). "Learning visual question answering by bootstrapping hard attention," in *European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, Vol. 11210 (Munich: Springer), 3–20. doi: 10.1007/978-3-030-01231-1_1

Meng, Q., and Zhang, W. (2019). "Multi-label image classification with attention mechanism and graph convolutional networks," in *Proceedings of the ACM Multimedia Asia* (Beijing, China). doi: 10.1145/3338533.3366589

Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant. Sci.* 7, 1419–1419. doi: 10.3389/fpls.2016.01419

Murat, M., Chang, S.-W., Abu, A., Yap, H. J., and Yong, K.-T. (2017). Automated classification of tropical shrub species: a hybrid of leaf shape and machine learning approach. *Peer J.* 5:e3792. doi: 10.7717/peerj.3792

Mutka, A. M., and Bart, R. S. (2015). Image-based phenotyping of plant disease symptoms. *Front. Plant Sci.* 5:734. doi: 10.3389/fpls.2014.00734

Nie, X., Wang, L., Ding, H., and Xu, M. (2019). Strawberry verticillium wilt detection network based on multi-task learning and attention. *IEEE Access* 7, 170003–170011. doi: 10.1109/ACCESS.2019.2954845

Ou, X., Wei, Z., Ling, H., Liu, S., and Cao, X. (2016). "Deep multi-context network for fine-grained visual recognition," in *International Conference on Multimedia and ExpoWorkshops (ICMEW)* (Seattle, WA: IEEE), 1–4. doi: 10.1109/ICMEW.2016.7574666

Peng, Y., He, X., and Zhao, J. (2018). Object-part attention model for fine-grained image classification. *IEEE Trans. Image Process.* 27, 1487–1500. doi: 10.1109/TIP.2017.2774041

Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., and Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* 161, 280–290. doi: 10.1016/j.compag.2018.04.002

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Front. Plant Sci.* 8:1852. doi: 10.3389/fpls.2017.01852

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). "MobileNetV2: inverted residuals and linear bottlenecks," in *Computer Vision and Pattern Recognition (CVPR)* (Salt Lake City, UT: IEEE), 4510–4520. doi: 10.1109/CVPR.2018.00474

Silva, P. F. B., Marçal, A. R. S., and da Silva, R. M. A. (2013). "Evaluation of features for leaf discrimination," *in International Conference Image Analysis and Recognition*, Lecture Notes in Computer Science, Vol. 7950 (Póvoa de Varzim: Springer), 197–204. doi: 10.1007/978-3-642-39094-4_23

Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)* (San Diego, CA).

Söderkvist, O. (2001). Computer vision classification of leaves from swedish trees (Master's Thesis). Linkoping University, Linkoping, Sweden, p. 74.

Sun, G., Cholakkal, H., Khan, S., Khan, F., and Shao, L. (2020). "Fine-grained recognition: accounting for subtle differences between similar classes," in *AAAI Conference on Artificial Intelligence (AAAI)* (New York, NY: AAAI) 34, 12047–12054. doi: 10.1609/aaai.v34i07.6882

Szegedy, C., vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE), 2818–2826. doi: 10.1109/CVPR.2016.308

Tan, M., and Le, Q. V. (2019). "Efficientnet: rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning (ICML)*, Vol. 97 (Long Beach, CA: ACM), 6105–6114.

Too, E. C., Yujian, L., Njuki, S., and Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161, 272–279. doi: 10.1016/j.compag.2018.03.032

Toseef, M., and Khan, M. J. (2018). An intelligent mobile application for diagnosis of crop diseases in Pakistan using fuzzy inference system. *Comput. Electron. Agric.* 153, 1–11. doi: 10.1016/j.compag.2018.07.034

Touvron, H., Vedaldi, A., Douze, M., and Jegou, H. (2019). "Fixing the train-test resolution discrepancy," in *Neural Information Processing Systems (NeurIPS)*, MIT Press, Vancouver, Canada, 32, pp. 8252–8262.

Turkoglu, M., Hanbay, D., and Sengur, A. (2019). Multi-model LSTM-based convolutional neural networks for detection of apple diseases and pests. *J. Ambient Intell. Humaniz. Comput.* doi: 10.1007/s12652-019-01591-w. [Epub ahead of print].

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y.-X., Chang, Y.-F., and Xiang, Q.-L. (2007). "A leaf recognition algorithm for plant classification using probabilistic neural network," *International Symposium on Signal Processing and Information Technology* (Giza: IEEE), 11–16. doi: 10.1109/ISSPIT.2007.4458016

Xiang, X., Yu, Z., Lv, N., Kong, X., and Saddik, A. E. (2020). "Semi-supervised image classification via attention mechanism and generative adversarial network," in *Proceedings Volume 11373, Eleventh International Conference on Graphics and Image Processing* (Hangzhou, China). doi: 10.1117/12.2557747

Xiao, F., Liu, B., and Li, R. (2020). Pedestrian object detection with fusion of visual attention mechanism and semantic computation. *Multimed. Tools Appl.* 79, 14593–14607. doi: 10.1007/s11042-018-7143-6

Yousefi, E., Baleghi, Y., and Sakhaei, S. M. (2017). Rotation invariant wavelet descriptors, a new set of features to enhance plant leaves classification. *Comput. Electron. Agric.* 140, 70–76. doi: 10.1016/j.compag.2017.05.031

Yu, Y., Jin, Q., and Chen, C. W. (2018). "FF-CMnet: a CNN-based model for fine-grained classification of car models based on feature fusion," in *International Conference on Multimedia and Expo (ICME)* (San Diego, CA: IEEE), 1–6. doi: 10.1109/ICME.2018.8486443

Zhang, L., Yang, Y., Wang, M., Hong, R., Nie, L., and Li, X. (2016). Detecting densely distributed graph patterns for fine-grained image categorization. *IEEE Trans. Image Process.* 25, 553–565. doi: 10.1109/TIP.2015.2502147

Zhang, S., Wu, X., You, Z., and Zhang, L. (2017). Leaf image based cucumber disease recognition using sparse representation classification. *Comput. Electron. Agric.* 134, 135–141. doi: 10.1016/j.compag.2017.01.014

Zhang, W., Tang, S., Su, J., Xiao, J., and Zhuang, Y. (2020). Tell and guess: cooperative learning for natural image caption generation with hierarchical refined attention. *Multimed. Tools Appl.* doi: 10.1007/s11042-020-08832-7. [Epub ahead of print].

Zhuang, P., Wang, Y., and Qiao, Y. (2020). Learning attentive pairwise interaction for fine-grained classification," in *AAAI Conference on Artificial Intelligence (AAAI)* (New York, NY: AAAI), 34, 13130–13137. doi: 10.1609/aaai.v34i07.7016

Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018). "Learning transferable architectures for scalable image recognition," in *Computer Vision and Pattern Recognition (CVPR)* (Salt Lake City, UT: IEEE), 8697–8710. doi: 10.1109/CVPR.2018.00907

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yang, He, Yang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/fpls.2020.554905

Modeling and Optimizing Culture Medium Mineral Composition for *in vitro* Propagation of *Actinidia arguta*

Radhia Hameg[1,2], Tomás A. Arteta[1,2], Mariana Landin[3,4], Pedro P. Gallego[1,2] and M. Esther Barreal[1,2]*

[1] Applied Plant & Soil Biology, Department of Plant Biology and Soil Sciences, Faculty of Biology, University of Vigo, Vigo, Spain, [2] CITACA - Agri-Food Research and Transfer Cluster, University of Vigo, Ourense, Spain, [3] Departamento de Farmacología, Farmacia y Tecnología Farmacéutica, R+D Pharma Group (GI−1645), Facultad de Farmacia y Agrupación Estratégica en Materiales (AeMat), Universidade de Santiago de Compostela, Santiago, Spain, [4] Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain

The design of plant tissue culture media remains a complicated task due to the interactions of many factors. The use of computer-based tools is still very scarce, although they have demonstrated great advantages when used in large dataset analysis. In this study, design of experiments (DOE) and three machine learning (ML) algorithms, artificial neural networks (ANNs), fuzzy logic, and genetic algorithms (GA), were combined to decipher the key minerals and predict the optimal combination of salts for hardy kiwi (*Actinidia arguta*) *in vitro* micropropagation. A five-factor experimental design of 33 salt treatments was defined using DOE. Later, the effect of the ionic variations generated by these five factors on three morpho-physiological growth responses – shoot number (SN), shoot length (SL), and leaves area (LA) – and on three quality responses - shoots quality (SQ), basal callus (BC), and hyperhydricity (H) – were modeled and analyzed simultaneously. Neurofuzzy logic models demonstrated that just 11 ions (five macronutrients (N, K, P, Mg, and S) and six micronutrients (Cl, Fe, B, Mo, Na, and I)) out of the 18 tested explained the results obtained. The rules "IF – THEN" allow for easy deduction of the concentration range of each ion that causes a positive effect on growth responses and guarantees healthy shoots. Secondly, using a combination of ANNs-GA, a new optimized medium was designed and the desired values for each response parameter were accurately predicted. Finally, the experimental validation of the model showed that the optimized medium significantly promotes SQ and reduces BC and H compared to standard media generally used in plant tissue culture. This study demonstrated the suitability of computer-based tools for improving plant *in vitro* micropropagation: (i) DOE to design more efficient experiments, saving time and cost; (ii) ANNs combined with fuzzy logic to understand the cause-effect of several factors on the response parameters; and (iii) ANNs-GA to predict new mineral media formulation, which improve growth response, avoiding morpho-physiological abnormalities. The lack of predictability on some response parameters can be due to other key media components, such as vitamins, PGRs, or organic compounds, particularly glycine, which could modulate the effect of the ions and needs further research for confirmation.

Keywords: algorithms, artificial intelligence, kiwiberry, modeling, mineral nutrition, plant tissue culture, physiological disorders

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Hamed Ahmadi, Tarbiat Modares University, Iran
Mohsen Niazian, Agricultural Research, Education and Extension Organization (AREEO), Iran

****Correspondence:***
Pedro P. Gallego
pgallego@uvigo.es

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *23 April 2020*
Accepted: *01 December 2020*
Published: *23 December 2020*

Citation:
Hameg R, Arteta TA, Landin M, Gallego PP and Barreal ME (2020) Modeling and Optimizing Culture Medium Mineral Composition for in vitro Propagation of Actinidia arguta. Front. Plant Sci. 11:554905. doi: 10.3389/fpls.2020.554905

INTRODUCTION

The process of designing protocols for successful plant tissue culture is a very complex task, since there are many potential interacting factors in this process (**Figure 1**). Plant materials, culture conditions, and culture media ingredients (inorganic and organic nutrients such as carbohydrates, vitamins, and plant growth regulators) are determining factors in the quality of the final product obtained in any plant cell culture protocol (micropropagated seedlings, somatic embryos, doubled haploids, etc.) (**Figure 1**).

Achieving quality products during *in vitro* plant tissue culture, rather than low survival rates and/or occurrence of physiological disorders, is highly dependent on the mineral nutritional composition of the media, as they are essential for optimal morphogenesis and organogenesis (Ramage and Williams, 2002; George et al., 2008; Sonnewald, 2013). In fact, physiological disorders and/or toxicity due to their deficiency or excess in the culture media inorganic composition has been reported (Bresinsky et al., 2013; Nezami-Alanagh et al., 2019). The inorganic nutrients added into the plant tissue culture media can be differentiated in two groups (**Figure 1**): macronutrients, taken up in large amounts (>0.5 mM L^{-1}) including nitrogen (N), potassium (K), calcium (Ca), phosphorus (P), magnesium (Mg) and sulfur (S); and those used in small quantities or micronutrients (<0.5 mM L^{-1}) such as iron (Fe), chlorine (Cl), manganese (Mn), zinc (Zn), boron (B), copper (Cu), and molybdenum (Mo) (Epstein, 1972; George et al., 2008).

The most widely used basal medium, MS (Murashige and Skoog, 1962), although it constitutes a good starting point for the development of new protocols (Niedz and Evens, 2007), is often inadequate because it generates physiological disorders such as shoot tip necrosis and/or hyperhydricity (Nowak et al., 2007; Bhojwani and Dantu, 2013; Nezami-Alanagh et al., 2018). MS mineral composition has been considered as unideal for many fruit species and cultivars (Reed and Hummer, 1995) and even supra-optimal for *in vitro* culture of kiwifruit *Actinidia* sp. (Moncaleán et al., 1999, 2003).

Many strategies have been carried out to improve plant-specific genotypes' tissue culture protocols by modifying the mineral composition of previously designed media. Initially, a trial and error strategy was employed by the pioneers of plant cell tissue culture (Gautheret, Heller and White's media) by changing the levels of each factor (independent variable) at a time, named "one factor at time" (OFAT), keeping the rest of the factors constant. Later, Hildebrandt et al. (1946) used the "triangulation method" (three elements varied at time). Finally, Murashige and Skoog (1962) tested the effect of one single element on several concentrations (1, 2, 4, and 8× of the basal medium based on White's nutrient solution) in the presence of several levels (3×, 8×, 16×, or even 32×) of the remaining elements. However, this strategy presented several disadvantages: (i) it does not give accurate information about the overall optimum, just partial optima for each factor; (ii) it ignores interactions between factors; and (iii) it increases the number of experiments (runs).

Currently, computer-based technologies are able to dramatically reduce the number of experiments and the associated cost (Nezami-Alanagh et al., 2018). As an example, the use of design of experiments (DOE) software facilitates the reduction of the optimal number of treatments to be performed, ensuring adequate sampling of the design space (Niedz and Evens, 2016; Nezami-Alanagh et al., 2019). The analysis of results with computer tools such as the response surface methodology (RSM) have been previously applied to study the composition of plant tissue culture media (Niedz and Evens, 2007; Poothong and Reed, 2016). However, the advantages of artificial intelligence tools, such as neurofuzzy logic, over some statistical analysis, including multiple regression analysis, has been described elsewhere (Landín et al., 2009; Gago et al., 2010a). Advantageously, algorithm-based machine learning (ML) tools provide the ability for autonomous learning and prediction of results without being explicitly programmed or with little human intervention (Gallego et al., 2011). In other words, algorithms can be trained to learn by themselves, generating a model that allows integrating and predicting results. ML approaches, such as artificial neural networks, fuzzy logic, and genetic algorithms, have been proposed as the most up-to-date methodology in the design of culture media (Gago et al., 2010a,c, 2011; Arteta et al., 2018; Nezami-Alanagh et al., 2018) to detect and understand the effect of several factors and their interactions (non-linear and multifactorial) and to predict the optimal combination of salts for the *in vitro* culture of plants (Gago et al., 2011; Gallego et al., 2011).

Due to the level of interest in the kiwi industry to introduce new kiwi genotypes (Kabaluk et al., 1997), our research group has pioneered the establishment of an *in vitro* culture protocol for kiwiberry, particularly for *Actinidia arguta* cv. *Issai* (Hameg et al., 2017). Firstly, based on successful results obtained for kiwifruit tissue culture (Revilla et al., 1992; Paradela et al., 2001), the media Cheng (1975) was used to establish *in vitro* kiwiberry explants (Hameg et al., 2018). Although good performance was achieved, the results suggested that additional research should be done to improve growth responses. To that end, several micropropagation media, previously used for *in vitro* kiwifruit culture, including B5 (Gamborg et al., 1968), Ha (Harada, 1975), Cheng (Cheng, 1975), Kh (Revilla et al., 1992), St (Standardi, 1981), and MS (Murashige and Skoog, 1962), were compared. The most appropriate medium for kiwiberry shoot proliferation was the St medium, but it caused some unwanted physiological disorders, such as basal callus formation (Hameg et al., 2018). For this reason, the challenge of designing a new basal medium that could avoid all these physiological disorders was considered (Hameg, 2019). In this work, we described how the combination of DOE and ML approaches were very useful, as a new strategy, in identifying the multifactorial and non-linear interactions between the culture media mineral nutrients and plant growth responses, and how it is possible to predict its optimal combination for the healthy *in vitro*

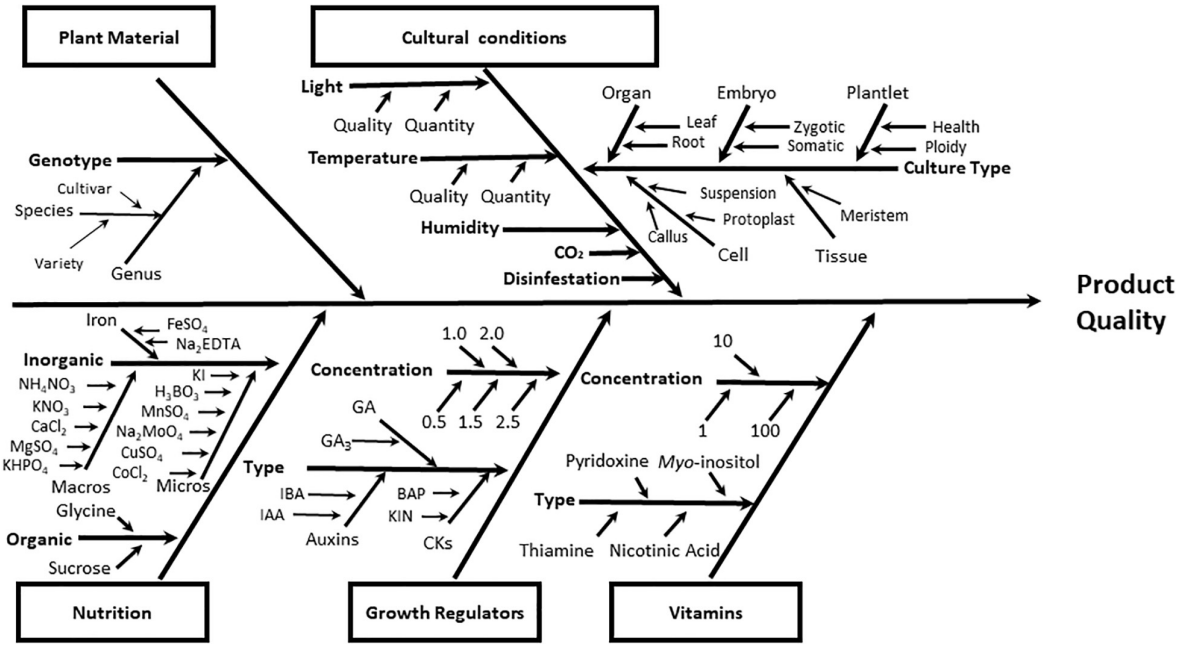

FIGURE 1 | Ishikawa diagram for quick visualization of the main categories of causes (plant material, cultural conditions, nutrition, plant growth regulators, and vitamins) that affect the plant cell tissue culture. Each category grouped factors representing the root causes of variation on the final quality of the process. As example of plant growth regulators. the next auxins (IAA, indole acetic acid; IBA, indole-butiric acid); CKs, cytokinins (BAP, Benzyl adenine purine; KIN, kinetin), and GA, gibberellins (GA_3, gibberellic acid) were included.

proliferation of any plant, particularly kiwiberry, using these promising approaches.

MATERIALS AND METHODS

Plant Material and Culture Condition

Nodal segments from a stock culture of *A. arguta* (Sieb. and Zucc.) Planch. ex Miq. cv. 'Issai' were maintained in Cheng basal medium (Cheng, 1975) supplemented with 1 mg L^{-1} 6-benzylaminopurine (BAP) and 1 mg L^{-1} gibberellic acid (GA_3), 8 g L^{-1} agar, and 30 g L^{-1} sucrose. Media pH was set to 5.7 before autoclaving at 121°C for 15 min at 105 KPa (Hameg et al., 2017; Hameg, 2019).

Experimental Design and Data Acquisition

Salts of MS medium (Murashige and Skoog, 1962) were classified into five independent factors (single salt or group of salts): (i) NH_4NO_3, (ii) KNO_3, (iii) mesos, (iv) micros, and (v) iron. Each factor has several levels corresponding to different concentrations of the MS medium (**Table 1**). These levels were defined over a range (minimum and maximum) of concentrations expressed as × MS level (1× correspond to MS concentration). The experimental space was designed to decipher the effect of extreme concentrations (very low and high) of ions with levels from 0.1 to 5× MS levels on the morpho-physiological shoots growth and quality responses.

A five-dimensional experimental design (Niedz and Evens, 2007) was developed using the software Design-Expert®8 (Design-Expert, 2010). The generated database included 36 treatments. 33 were generated by the software using modified D-optimal criteria (Reed et al., 2013b), while three additional points of MS media were used as controls (34–36; **Table 2**). All treatments contained MS medium vitamin composition and were supplemented with 2 mg L^{-1} glycine, 30 g L^{-1} sucrose, 8 g L^{-1} agar, ad 1 mg L^{-1} BAP, and 1 mg L^{-1} GA_3.

TABLE 1 | Five factors used to define the five-dimensional design space based on MS medium salts and concentration range expressed as (× MS levels).

Factors	Media salts	Range
Factor 1	NH_4NO_3	0.2–1×
Factor 2	KNO_3	0.1–1×
Factor 3 (Mesos)	$CaCl_2{\cdot}2H_2O$	0.25–3×
	$MgSO_4{\cdot}7H_2O$	
	KH_2PO_4	
Factor 4 (Micros)	$MnSO_4{\cdot}4H_2O$	0.1–1.5×
	$ZnSO_4{\cdot}7H_2O$	
	H_3BO_3	
	KI	
	$CuSO_4{\cdot}5H_2O$	
	$Na_2MoO_4{\cdot}2H_2O$	
	$CoCl_2{\cdot}6H_2O$	
Factor 5 (Iron)	$FeSO_4{\cdot}7H_2O$	1–5×
	$Na_2{\cdot}EDTA$	

TABLE 2 | Five-factor design with 33 treatments, including three replicates points (6–7, 22–23, and 27–28) plus another three replicates of MS medium (34–36), that are bolded.

Treatments (media)	Factor 1 NH_4NO_3	Factor 2 KNO_3	Factor 3 Mesos	Factor 4 Micros	Factor 5 Iron
1	0.51	0.66	1.65	1.50	3.04
2	1.00	0.66	2.59	0.10	1.00
3	1.00	1.00	2.37	1.50	1.08
4	0.20	0.24	3.00	0.97	1.84
5	0.90	0.10	0.25	0.10	1.60
6	**0.20**	**1.00**	**0.94**	**0.10**	**5.00**
7	**0.20**	**1.00**	**0.94**	**0.10**	**5.00**
8	0.40	1.00	0.25	1.50	1.00
9	0.81	0.10	0.25	1.40	5.00
10	1.00	1.00	3.00	0.39	5.00
11	0.20	0.33	3.00	0.10	5.00
12	0.81	0.10	0.25	1.40	5.00
13	0.80	0.10	2.99	1.29	1.00
14	0.40	0.10	0.25	0.10	5.00
15	0.20	0.10	2.67	1.50	4.20
16	1.00	0.33	0.25	1.50	1.00
17	0.61	0.64	1.90	0.14	3.54
18	0.20	0.76	0.25	0.10	1.00
19	0.30	1.00	3.00	1.22	5.00
20	1.00	0.10	2.38	0.10	4.46
21	1.00	1.00	0.90	1.50	4.78
22	**0.20**	**0.76**	**0.25**	**1.50**	**5.00**
23	**0.20**	**0.76**	**0.25**	**1.50**	**5.00**
24	1.00	1.00	0.25	0.10	3.30
25	1.00	0.35	1.52	0.94	3.10
26	0.20	0.76	3.00	1.50	1.00
27	**0.24**	**1.00**	**3.00**	**0.10**	**2.00**
28	**0.24**	**1.00**	**3.00**	**0.10**	**2.00**
29	0.20	0.76	3.00	1.50	1.00
30	0.32	0.10	2.34	0.10	1.00
31	0.72	0.72	0.90	0.69	1.00
32	0.20	0.10	0.80	1.36	1.00
33	1.00	0.34	3.00	1.50	5.00
34	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
35	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
36	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

Explants about 2 cm were cultured in 200 mL culture vessels containing 30 mL of each medium for 50 days. The cultures were maintained in a growth chamber at 25 ± 1°C under a 16 h photoperiod at 40 μmol m^{-2} s^{-1} irradiation provided by cool white fluorescent tubes (Hameg et al., 2017).

Each treatment included five glass culture vessels (used as replicates) containing three explants each, sealed with plastic caps. The experiments were carried out in triplicate. The explants were harvested after 50 days of culture, all followed, and the next six growth responses were evaluated (Hameg, 2019):

(1) Shoot number (SN), number of new regenerated shoots per explant.
(2) Shoot length (SL), length from the base of the shoot to the tip, per explant (cm).
(3) Leaf area (LA), the sum of areas (cm^2) of the leaves > 1.5 cm. Leaf area per explant was measured using a portable laser leaf area meter (Meter CI-202, CID biosciences, WA, United States).

As the MS mineral salts have been reported for promoting physiological disorders in some plants, the next three morpho-physiological quality responses were also evaluated:

(1) Shoot quality (SQ), as indicative of shoot vigor, was visually assessed and scored from 1 to 5 (1 very poor, 2 poor, 3 moderate, 4 good and 5 very good; **Figure 2A**).
(2) Basal callus (BC), callus formation at the cut edge of shoots was visually assessed and scored from 1 to 4 (1 necrotic, 2 big, 3 moderate, 4 absent; **Figure 2B**).
(3) Hyperhydricity (H), was visually assessed and scored from 1 to 3 (1 high, 2 low, 3 none; **Figure 2C**).

Machine Learning Tools for Modeling

Machine learning uses a wide range of algorithms to build mathematical models using databases as training data, helping humans to make predictions and decisions. Here, three artificial intelligence tools were used to build the mathematical models: neural networks (ANNs), fuzzy logic, and genetic algorithms (GA). The commercial neurofuzzy logic software, FormRules®v4.03, (Intelligensys Ltd., United Kingdom) which combines ANNs with fuzzy logic (Colbourn and Rowe, 2005; Landin and Rowe, 2013) was selected to model and decipher the effect of the mineral media composition on plant growth responses, while the commercial software INForm®v5.01 (Intelligensys Ltd., United Kingdom) that combines ANNs with GA (ANNs-GA), was used for the optimization of the mineral nutrition. Advantageously, those artificial intelligence tools allow for the modeling of large databases with an important number of inputs (factors studied) and outputs (plant response parameters determined), independently of the type of data or even if the data set is incomplete, vague, or noisy (Gago et al., 2011; Gallego et al., 2011).

The neurofuzzy logic model was built using 18 inputs (ion concentrations of each treatment) and six outputs (SN, SL, LA, SQ, BC, and H). The ion composition of each treatment (**Table 3**) was calculated from each salt concentration in the media (**Supplementary Table 1**). Each ionic concentration was used as an input for the model (**Table 3**). This procedure deeply facilitates the understanding of the specific effects of mineral elements (ions), avoiding the "ion confounding effect" as described elsewhere (Niedz and Evens, 2006, 2007; Nezami-Alanagh et al., 2017). Instead, the ANNs-GA model was built using 14 inputs corresponding to the MS salts and the same six outputs used for the neurofuzzy logic model (**Supplementary Table 1**) in order to optimize the salt composition and define a new optimal culture media for kiwiberry.

Machine learning algorithms were able to build empirical models using the training parameters presented in **Table 4**. The Adaptive Spline Modeling of Data (ASMOD algorithm) was

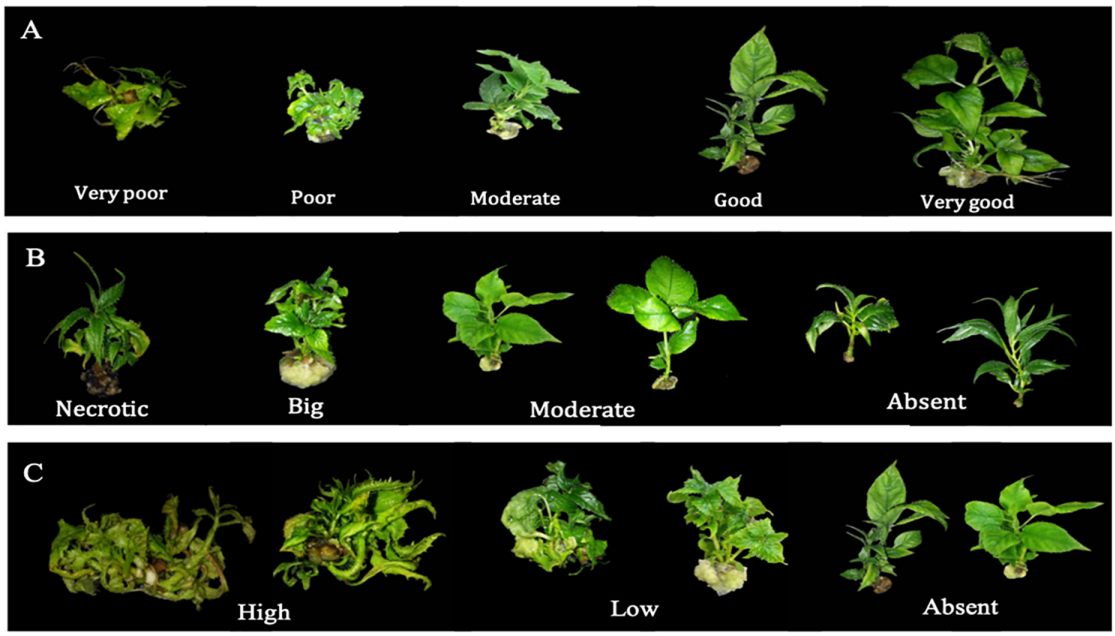

FIGURE 2 | Shoot quality rating **(A)**: 1 (very poor), 2 (poor), 3 (moderate), 4 (good), and 5 (very good); basal callus formation rating **(B)**: 1 (necrotic), 2 (big), 3 (moderate), and 4 (absent) and hyperhydricity rating; **(C)**: 1 (high), 2 (low), and 3 (absent).

used by FormRules for the parameter minimization, including in the model the relevant inputs, facilitating a more parsimonious and transparent model for users. Compared to other models of a general structure, ASMOD reduces the model complexity but improves its accuracy even with fewer parameters (Kavli and Weyer, 1994). Finally, ASMOD allows for dividing of the model obtained into submodels to easily interpret the results by generating a set of rules. Once the accuracy of the model was ensured, structural risk minimization (SRM) was selected to obtain models with the highest predictability along with the simplest rules. FormRules® presents the results obtained as a set of linguistic labels or IF-THEN rules with a degree of membership, which greatly facilitates their interpretation. The antecedent part (IF) expresses the conditions at the inputs, and the consequent part (THEN) describes the values of the outputs. The degree of membership represents a degree of truth, ranging from 0 to 1, with 1 meaning that the expected output value is always a complete member of the fuzzy set "low," "medium," or "high" (Shao et al., 2006; Gago et al., 2011; Gallego et al., 2011; Nezami-Alanagh et al., 2018).

The Back-Propagation (BP) for Multi-Layer Perceptron (MLP) was used as the training algorithm for InForm® software. To avoid overfitting during MLP training, the data set was split into two groups of data randomly: 80% for training and 20% for testing. Both training error and testing error were checked at every step to prevent overfitting, as described previously (Nezami-Alanagh et al., 2017).

For the optimization process, the software requires the definition of the desirability function for each output, together with their relative importance or weight, expressed on a scale of 0 to 10, 10 being the most important. For media optimization, only the measurable three growth parameters and shoot quality were included. The priority established was: SQ = 10, SN = 9, SL = 8, and LA = 7. As desired values to be achieved, SQ > 4.00, SN > 4.4, SL > 1.6 cm, LA > 28 cm^2 were included. Finally, the model reveals the percentage of agreement between the predicted values with those desired by the researchers on a scale of 0–100%.

The predictability and accuracy for each parameter developed by both software was assessed using the Train Set R^2 and the ANOVA f-ratios. Train Set R^2 values are calculated by the following equation (Shao et al., 2006).

$$R^2 = \left(1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i^{'}\right)^2}{\sum_{i=1}^{n}\left(y_i - y_i^{''}\right)^2}\right) \times 100\%$$

Where y_i is the experimental point in the data set, y_i' is the predicted point calculated by the model, and y_i'' is the mean of the dependent variable. The higher the Train Set R^2 value, the better the predictability of the model. In previous works, R^2 values higher than 70% have shown good model prediction capacity. It is necessary to avoid R^2 that are too high (>99%), which is indicative of over fitted models of low prediction capacity and should be readjusted as described elsewhere (Colbourn and Rowe, 2005; Landín et al., 2009; Nezami-Alanagh et al., 2017).

To assess statistically significant differences between experimental and predicted values from the model an analysis of variance (one-way ANOVA) was carried out. If the ANOVA f-ratio is higher than f-critical values for the degrees of freedom of the model, then there are no statistically significant differences

TABLE 3 | Mineral nutrients' (expressed as ions) composition of the different culture media based on the five-factor experimental design (0–33) and response values of the parameters (mean and standard deviation) used to characterize plant growth.

	Ions (mM)																	Growth parameters			Quality parameters		
Media	NH_4^+	NO_3^-	K^+	Ca^{2+}	Mg^{2+}	PO_4^{3-}	SO_4^{-2}	Cl^-	Fe^{2+}	BO_3^-	Mn^{2+}	Zn^{2+}	Cu^{2+}	MoO_2^{2-}	Na^+	Co^{2+}	I^-	$EDTA^-$	SN	SL (cm)	LA (cm^2)	SQ	BC	H
1	10.47	22.84	14.44	4.95	2.48	2.06	2.98	9.89	0.30	0.15	0.15	0.045	0.00015	0.0015	0.61	0.00016	0.0075	0.30	5.0 ± 0.8	1.6 ± 0.4	20.6 ± 6.3	3.8 ± 0.7	4.0 ± 0.0	2.9 ± 0.5
2	20.61	33.06	15.68	7.74	3.88	3.23	4.00	15.49	0.10	0.01	0.01	0.003	0.00001	0.0001	0.20	0.00001	0.0005	0.10	5.4 ± 2.7	1.7 ± 0.3	40.7 ± 10.9	3.2 ± 1.2	4.0 ± 0.0	1.8 ± 0.9
3	20.61	39.41	21.76	7.09	3.55	2.96	3.86	14.17	0.11	0.15	0.15	0.045	0.00015	0.0015	0.22	0.00016	0.0075	0.11	2.7 ± 0.8	2.3 ± 0.8	26.1 ± 12.8	4.4 ± 0.7	4.0 ± 0.0	2.8 ± 0.5
4	4.21	8.62	8.17	8.98	4.50	3.75	4.81	17.96	0.18	0.10	0.10	0.029	0.00010	0.0010	0.37	0.00010	0.0048	0.18	5.8 ± 2.6	1.9 ± 0.6	12.0 ± 4.0	3.5 ± 0.5	4.0 ± 0.0	1.8 ± 1.0
5	18.55	20.43	2.19	0.75	0.38	0.31	0.55	1.50	0.16	0.01	0.01	0.003	0.00001	0.0001	0.32	0.00001	0.0005	0.16	5.8 ± 1.6	1.2 ± 0.3	7.1 ± 3.4	3.1 ± 0.5	4.0 ± 0.0	3.0 ± 0.0
6	4.12	22.92	19.96	2.81	1.41	1.17	1.92	5.61	0.50	0.01	0.01	0.003	0.00001	0.0001	1.00	0.00001	0.0005	0.50	2.6 ± 1.1	1.2 ± 0.4	6.8 ± 3.2	2.5 ± 0.9	4.0 ± 0.0	3.0 ± 0.0
7	4.12	22.92	19.96	2.81	1.41	1.17	1.92	5.61	0.50	0.01	0.01	0.003	0.00001	0.0001	1.00	0.00001	0.0005	0.50	2.2 ± 1.2	1.4 ± 0.5	7.8 ± 4.4	2.2 ± 0.9	4.0 ± 0.0	3.0 ± 0.0
8	8.25	27.04	19.11	0.75	0.38	0.31	0.67	1.50	0.10	0.15	0.15	0.045	0.00015	0.0015	0.20	0.00016	0.0075	0.10	7.9 ± 2.7	1.1 ± 0.2	8.7 ± 3.9	2.8 ± 0.5	3.5 ± 0.5	2.9 ± 0.3
9	16.73	18.61	2.20	0.75	0.38	0.31	1.06	1.50	0.50	0.14	0.14	0.042	0.00014	0.0014	1.00	0.00015	0.0070	0.50	4.1 ± 1.5	0.7 ± 0.2	3.1 ± 1.7	1.7 ± 0.5	4.0 ± 0.0	3.0 ± 0.0
10	20.61	39.41	22.54	8.98	4.50	3.75	5.05	17.96	0.50	0.04	0.04	0.012	0.00004	0.0004	1.00	0.00004	0.0019	0.50	2.2 ± 1.1	1.3 ± 0.5	9.2 ± 3.6	2.3 ± 1.0	4.0 ± 0.0	2.2 ± 1.0
11	4.21	10.31	9.86	8.98	4.50	3.75	5.02	17.96	0.50	0.01	0.01	0.003	0.00001	0.0001	1.00	0.00001	0.0005	0.50	3.7 ± 1.1	1.0 ± 0.3	6.0 ± 2.1	1.7 ± 0.7	4.0 ± 0.0	1.9 ± 1.0
12	16.73	18.61	2.20	0.75	0.38	0.31	1.06	1.50	0.50	0.14	0.14	0.042	0.00014	0.0014	1.00	0.00015	0.0070	0.50	4.2 ± 1.5	0.8 ± 0.2	3.8 ± 1.2	1.5 ± 0.5	3.1 ± 0.2	3.0 ± 0.0
13	16.57	18.45	5.62	8.94	4.48	3.73	4.75	17.88	0.10	0.13	0.13	0.039	0.00013	0.0013	0.20	0.00014	0.0064	0.10	4.8 ± 1.0	1.7 ± 0.4	17.0 ± 5.5	3.6 ± 0.7	4.0 ± 0.0	1.9 ± 0.7
14	8.16	10.04	2.19	0.75	0.38	0.31	0.89	1.50	0.50	0.01	0.01	0.003	0.00001	0.0001	1.00	0.00001	0.0005	0.50	3.1 ± 1.4	0.7 ± 0.3	3.1 ± 1.3	1.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0
15	4.12	6.00	5.22	7.99	4.01	3.34	4.62	15.98	0.42	0.15	0.15	0.045	0.00015	0.0015	0.84	0.00016	0.0075	0.42	1.9 ± 0.8	1.1 ± 0.3	4.1 ± 1.8	1.3 ± 0.6	4.0 ± 0.0	2.9 ± 0.3
16	20.61	26.89	6.60	0.75	0.38	0.31	0.67	1.50	0.10	0.15	0.15	0.045	0.00015	0.0015	0.20	0.00016	0.0075	0.10	6.0 ± 1.2	1.5 ± 0.4	14.0 ± 3.4	3.1 ± 0.6	3.9 ± 0.2	2.6 ± 0.7
17	12.66	24.76	14.47	5.69	2.85	2.37	3.23	11.37	0.35	0.01	0.01	0.004	0.00001	0.0001	0.71	0.00001	0.0007	0.35	1.9 ± 1.1	1.2 ± 0.4	8.7 ± 2.3	2.3 ± 0.9	4.0 ± 0.0	3.0 ± 0.0
18	4.12	18.35	14.54	0.75	0.38	0.31	0.49	1.50	0.10	0.01	0.01	0.003	0.00001	0.0001	0.20	0.00001	0.0005	0.10	6.1 ± 1.4	1.3 ± 0.3	8.8 ± 2.8	2.9 ± 0.5	2.0 ± 0.0	2.9 ± 0.3
19	6.26	25.05	22.55	8.98	4.50	3.75	5.16	17.96	0.50	0.12	0.12	0.036	0.00012	0.0013	1.00	0.00013	0.0061	0.50	2.9 ± 1.5	1.7 ± 0.7	14.5 ± 6.4	3.4 ± 1.0	4.0 ± 0.0	2.7 ± 0.8
20	20.61	22.49	4.85	7.13	3.57	2.97	4.03	14.25	0.45	0.01	0.01	0.003	0.00001	0.0001	0.89	0.00001	0.0005	0.45	1.9 ± 0.9	1.5 ± 0.6	6.1 ± 3.7	2.1 ± 0.7	4.0 ± 0.0	2.9 ± 0.3
21	20.61	39.41	19.92	2.68	1.35	1.12	2.02	5.37	0.48	0.15	0.15	0.045	0.00015	0.0015	0.96	0.00016	0.0075	0.48	1.1 ± 0.3	1.7 ± 0.3	5.7 ± 3.4	2.3 ± 1.0	4.0 ± 0.0	3.0 ± 0.0
22	4.12	18.35	14.55	0.75	0.38	0.31	1.07	1.50	0.50	0.15	0.15	0.045	0.00015	0.0015	1.00	0.00016	0.0075	0.50	4.3 ± 0.9	1.0 ± 0.2	4.1 ± 1.4	1.7 ± 0.5	1.0 ± 0.0	3.0 ± 0.0
23	4.12	18.35	14.55	0.75	0.38	0.31	1.07	1.50	0.50	0.15	0.15	0.045	0.00015	0.0015	1.00	0.00016	0.0075	0.50	3.2 ± 1.3	1.0 ± 0.3	4.0 ± 1.6	1.7 ± 0.5	1.2 ± 0.7	3.0 ± 0.0
24	20.61	39.41	19.11	0.75	0.38	0.31	0.72	1.50	0.33	0.01	0.01	0.003	0.00001	0.0001	0.66	0.00001	0.0005	0.33	3.1 ± 1.5	1.3 ± 0.4	4.7 ± 2.0	1.9 ± 0.8	4.0 ± 0.0	2.6 ± 0.9
25	20.61	27.14	8.43	4.53	2.27	1.89	2.71	9.07	0.31	0.09	0.09	0.028	0.00009	0.0010	0.62	0.00010	0.0047	0.31	3.5 ± 0.9	2.2 ± 0.9	16.2 ± 5.3	3.7 ± 0.7	4.0 ± 0.0	2.9 ± 0.5
26	4.12	18.43	18.07	8.98	4.50	3.75	4.80	17.96	0.10	0.15	0.15	0.045	0.00015	0.0015	0.20	0.00016	0.0075	0.10	2.3 ± 1.0	1.9 ± 0.8	21.7 ± 10.3	4.2 ± 0.8	3.6 ± 0.5	2.8 ± 0.4
27	4.95	23.74	22.54	8.98	4.50	3.75	4.72	17.96	0.20	0.01	0.01	0.003	0.00001	0.0001	0.40	0.00001	0.0005	0.20	4.1 ± 1.8	2.0 ± 0.6	28.5 ± 8.5	3.9 ± 0.9	4.0 ± 0.0	1.4 ± 0.7
28	4.95	23.74	22.54	8.98	4.50	3.75	4.72	17.96	0.20	0.01	0.01	0.003	0.00001	0.0001	0.40	0.00001	0.0005	0.20	4.3 ± 1.5	1.9 ± 0.6	29.9 ± 13.8	4.1 ± 0.8	4.0 ± 0.0	1.8 ± 0.7
29	4.12	18.43	18.07	8.98	4.50	3.75	4.80	17.96	0.10	0.15	0.15	0.045	0.00015	0.0015	0.20	0.00016	0.0075	0.10	2.7 ± 1.0	1.8 ± 1.1	18.9 ± 9.3	3.5 ± 1.2	3.7 ± 0.5	2.4 ± 0.8
30	6.51	8.39	4.80	7.00	3.51	2.92	3.63	14.01	0.10	0.01	0.01	0.003	0.00001	0.0001	0.20	0.00001	0.0005	0.10	6.6 ± 2.4	1.1 ± 0.3	9.3 ± 3.9	1.4 ± 0.5	4.0 ± 0.0	1.0 ± 0.0
31	14.92	28.46	14.67	2.68	1.35	1.12	1.54	5.36	0.10	0.07	0.07	0.021	0.00007	0.0007	0.20	0.00007	0.0034	0.10	3.9 ± 1.4	2.6 ± 0.9	28.3 ± 9.3	4.7 ± 0.5	4.0 ± 0.0	2.7 ± 0.5
32	4.12	6.00	2.89	2.39	1.20	1.00	1.48	4.79	0.10	0.14	0.14	0.041	0.00014	0.0014	0.20	0.00014	0.0068	0.10	4.1 ± 1.7	1.2 ± 0.3	5.7 ± 2.9	2.4 ± 0.8	4.0 ± 0.0	2.9 ± 0.3
33	20.61	26.98	10.12	8.98	4.50	3.75	5.20	17.96	0.50	0.15	0.15	0.045	0.00015	0.0015	1.00	0.00016	0.0075	0.50	2.6 ± 2.1	1.5 ± 0.5	10.4 ± 5.5	2.6 ± 0.8	4.0 ± 0.0	2.4 ± 0.9
MS	**20.61**	**39.41**	**20.05**	**2.99**	**1.50**	**1.25**	**1.73**	**5.99**	**0.10**	**0.10**	**0.10**	**0.030**	**0.00010**	**0.0010**	**0.20**	**0.00011**	**0.0050**	**0.10**	**3.9 ± 1.3**	**1.7 ± 0.4**	**28.7 ± 9.3**	**4.1 ± 0.4**	**4.0 ± 0.0**	**2.6 ± 0.6**

Original medium composition (bold) used as control. SN, shoot number; SL, shoot length; LA, leaf area; SQ, shoot quality; BC, basal callus; and H, hyperhydricity.

TABLE 4 | Train parameter settings for neurofuzzy logic (FormRules® v4.03) and artificial neural networks (INForm® v5.01) software.

FormRules® v4.03	INForm® v5.01
Minimization parameters (ASMOD)	Number of inputs: 14
Ridge Regression Factor: $1e^{-6}$	Number of hidden layers: 1
Model Selection Criteria	Number of nodes: 2–4
Structural Risk Minimization (SRM)	Transfer function in hidden layer:
C1 = 0.868 (except SN: 0.8); C2 = 4.8	asymmetric sigmoid
Number of Set Densities: 2	Output transfer type: linear
Set Densities: 2. 3	Back propagation type: RPROP
Adapt Nodes: TRUE	Targets
Max. Inputs Per SubModel: 4	Target interactions: 1000
Max. Nodes Per Input: 15	Target MS error: 0.0001
	Random seed: 10000
	Test data
	Smart stop: enabled
	Minimum interactions: 20
	Test error weighting: 0.1
	Auto weight
	Interaction overshoot: 200

between those groups (predicted and experimental values) and the model is accurate.

Experimental Validation of ANNs-GA Model

In order to validate the model, a new experiment including optimal predicted *R* medium was carried out in the laboratory. As controls, another six media generally used in kiwifruit tissue culture, such as MS, B5, St, and Ha, were also tested. A total of 45 explants per medium were cultivated in the same conditions as above. After 50 days, the growth parameters (SN, SL, and LA) and quality parameters (SQ, BC, and H) were recorded.

Statistical analysis was used to check the validity of the new optimized *R* medium and search for the significant differences between the new *R* medium and MS. Continuous data (SL and LA) were analyzed using ANOVA with Tukey's Studentized range, (HSD) *post hoc* test at $\alpha = 0.001$. Discrete (SQ, BC, and H) and categorical (SN) data were analyzed by the non-parametric Kruskal–Wallis test at $\alpha = 0.001$. All the statistical analyses were performed using Statistica v.12 (StatSoft-Inc., 2014).

RESULTS

Effect of Mineral Nutrition on Shoot Growth and Quality Responses

The computer-based software for optimal design was used to guarantee a well sampled design space, giving a database of 34 treatments combining all 18 mineral nutrients based on the MS medium composition (**Table 3**). To focus on the effect of the mineral composition, the other media ingredients were kept constant. The MS based treatments caused a great variety of physiological responses compared with the MS control (**Table 3**). As example, very good quality (#31: 4.7), large SN (#8: 7.9), long SL (#31: 2.6 cm), and very large LA (#2: 40.7 cm^2) versus the MS (4.1, 3.9; 1.7 cm and 28.7 cm^2, respectively) were obtained. On the contrary, very poor quality (#14: 1.0), few SN (#21: 1.1), short SL (#9: 0.7 cm), or LA (#9 and 14: 3.1 cm^2) were also achieved.

Physiological disorders such as necrotic, big, or even moderate basal callus and hyperhydricity were also found (**Figure 3**). As an example, the treatments #18, 22, and 23 (**Figures 3A1–A3**) promoted the formation of big and necrotic callus at the cut edge of shoots, whereas the treatments #2, 3, and 30 (**Figures 3B1–B3**) promoted hyperhydricity. The MS medium, used as control, did not show basal callus formation (4.0) on kiwiberry, but some hyperhydricity (1.4) was detected.

Classical statistical techniques (ANOVA) are very useful in data analysis, but they do not help much in extracting valuable information about key mineral nutrients in these complex processes, or the right combination of minerals to promote ever-healthy plantlets. Currently, other advanced algorithm-based technologies like decision trees, SRM, or machine learning tools look promising for deciphering key minerals. Here, the latest technologies were used.

Neurofuzzy Logic Models

The neurofuzzy logic tool was able to successfully model the dataset (**Table 3**) with the training parameters described in **Table 4**. The results of the Train Set R^2, higher than 70% for all parameters, indicated good performance and a high predictability of the neurofuzzy logic models (**Table 5**). Moreover, the ANOVA f ratio was always higher than the f critical value showing the quality and accuracy for prediction, since no statistically significant differences ($\alpha < 0.001$) between experimental and predicted values were found (**Table 5** and **Supplementary Figure 1**).

Neurofuzzy logic approach also succeeded in identifying the key factors (inputs) for each parameter (outputs) studied. Among the 18 evaluated ions, just 11 were critical and explain variations among treatments (**Table 5**). Of these, five belong to macronutrients (N, K, P, Mg, and S) and six to micronutrients (Cl, Fe, B, Mo, Na, and I). While some ions caused independent effects, others interacted with each other. As an example, the variability of the SN parameter was explained as a function of the interactions of seven ions, $Na^+ \times MoO_2^{2-}$; $SO_4^{2-} \times I^-$ and $NH_4^+ \times NO_3^- \times BO_3^-$, while the SL variability was independently affected by the following 5 ions: Fe^{2+}, K^+, Mg^{2+}, BO_3^-, and PO_4^{3-} (**Table 5**).

Morpho-Physiological Growth Responses

The variability on the new regenerated shoots per explant (SN parameter) is explained by three submodels showing three interactions of Na^+ and MoO_2^{2-} (strongest effect; submodel 1), the $SO_4^{2-} \times I^-$ interaction (submodel 2), and $NH_4^+ \times NO_3^- \times BO_3^-$ (submodel 3; **Table 5**).

Through simple rules IF THEN the model pinpoints the negative effect of both Na^+ and MoO_2^{2-} ions at high concentrations on shoot proliferation (**Table 6**; rule 4), recommending their use at low concentrations to obtain the highest number of regenerated shoots (**Table 6**; rule 1). In both cases, the membership was 1.00 which means that if we combine

FIGURE 3 | Morpho-physiological abnormalities responses to some media treatments as basal callus **(A)**: treatments 18 **(A1)**, 22 **(A2)**, and 23 **(A3)** and hyperhydricity **(B)**: treatments 2 **(B1)** and 30 **(B2,B3)** versus high quality plants without symptoms **(A4,B4)**.

TABLE 5 | Neurofuzzy logic model Mean Square Error (MSE), train set R^2, ANOVA parameters for training [*f*-ratio, degrees of freedom (df1: model and df2: total), and *f*-critical value for $\alpha = 0.05$].

Outputs	MSE	Train set R^2 (%)	*f-ratio*	df1	df2	*f*-critical ($\alpha = 0.05$)	Critical factors
SN	0.0055	89.495	9.030	16	33	3.558	**Na^+ x MoO_2^{2-}** SO_4^{2-} × I^- NH_4^+ × NO_3^- × BO_3^-
SL	0.0107	80.063	10.350	9	33	4.255	Fe^{2+} K^+ **Mg^{2+}** BO_3^- PO_4^{3-}
LA	0.0159	75.629	17.378	5	33	5.382	**NO_3^-** × **Na^+** Cl^-
SQ	0.0050	92.998	26.563	11	33	3.975	Fe^{2+} **K^+** × **SO_4^{2-}** BO_3^- NO_3^-
BC	0.0069	89.346	16.771	11	33	3.975	**PO_4^{3-}** × **NH_4^+** SO_4^{-2}
H	0.0173	75.250	11.289	7	33	4.675	**Cl^-** × **$I^-SO_4^{2-}$** × **BO_3^-**

Critical factors (significant inputs) computed for each output are bolded. SN, shoot number; SL, shoot length; LA, leaf area; SQ, shoot quality; BC, basal callus; and H, hyperhydricity.

low concentrations of both ions, we will obtain a truly high number of new shoots.

Also, the presence of High SO_4^{2-} concentrations in the medium in combination with mid or high I^- concentration significantly increases the number of shoots (**Table 6**; rules 9–10). Finally, the model also pinpoints the key role of High concentrations of BO_3^- combined with either low or high concentrations of NH_4^+ × NO_3^- in the culture medium (**Table 6**; rules 12, 14, 16, and 18).

Shoots length (SL) of *A. arguta* is independently influenced by the effect of five ions: Fe^{2+}, K^+, Mg^{2+}, BO_3^-, and PO_4^{3-} (**Table 5**). The model selects Mid Mg^{2+} as the adequate concentration to obtain large (high) shoot length (**Table 6**; rules 24). When low (rule 23) or high (rule 25) concentrations are added to the media (membership 1.00) a truly short SL is achieved. Additionally, Low Fe^{2+}, High K^+, Mid BO_3^-, and Low or High PO_4^{3-} concentrations in the culture media increases SL (**Table 6**; rules 19, 22, 27, 29, and 31).

Area of leaves in *A. arguta* were mainly influenced by the interaction of NO_3^- and Na^+ concentrations and the single effect of Cl^- concentration (**Table 5**). The greatest leaf area is achieved with the combination of High NO_3^- and Low Na^+ concentrations (**Table 6**; rule 33). Other combinations promoted worse results for leaf area (**Table 6**, rules 32, 34–36). High values of LA are also obtained when high concentrations of Cl^- were used (**Table 6**, rule 37).

Morpho-Physiological Quality Responses

Mineral nutrients had a great impact on the appearance of morpho-physiological abnormalities such as basal callus and hyperhydricity (**Table 3** and **Figure 3**). Therefore, a parameter to establish the quality of the shoots was included as output and determined as explained previously (**Figure 2A**).

The neurofuzzy logic software selected just five ions as critical factors for the quality of the shoots: the interaction between K^+ and SO_4^{2-} as the stronger, and the independent effect of Fe^{2+}, BO_3^-, and NO_3 (**Table 5**). The best shoot quality is promoted (membership 1.00), if High K^+ and Mid SO_4^{2-} concentrations are added into the media (**Table 7**; rule 44), although High SO_4^{2-} concentration also improves shoot quality with high membership (0.83; **Table 7**; rule 45). On the contrary, the high concentration of Fe^{2+} ion promotes a truly low quality of shoots (**Table 7**;

TABLE 6 | Rules for morpho-physiological growth responses (SN, shoot number; SL, shoot length and LA, leaf area) with their membership degree (MD) generated by neurofuzzy logic.

Rules		$[NH_4^+]$	$[NO_3^-]$	$[K^+]$	$[Mg^{2+}]$	$[PO_4^{3-}]$	$[SO_4^{2-}]$	$[Cl^-]$	$[Fe^{2+}]$	$[BO_3^-]$	$[MoO_2^{2-}]$	$[Na^+]$	$[I^-]$		SN	SL	LA	MD
1	**IF**										**Low**	**Low**		**THEN**	**High**			**1.00**
2											High	Low			Low			1.00
3											Low	High			High			1.00
4											**High**	**High**			**Low**			**1.00**
5							Low						Low		Low			1.00
6							Low						Mid		Low			1.00
7							Low						High		High			1.00
8							High						Low		Low			1.00
9							High						Mid		High			1.00
10							High						High		High			1.00
11		Low	Low							Low					Low			1.00
12		Low	Low							High					High			1.00
13		High	Low							Low					Low			1.00
14		High	Low							High					High			1.00
15		Low	High							Low					Low			1.00
16		Low	High							High					High			1.00
17		High	High							Low					Low			1.00
18		High	High							High					High			1.00
19	**IF**								Low					**THEN**		High		1.00
20									High							Low		0.94
21				Low												Low		0.74
22				High												High		1.00
23					**Low**											**Low**		**1.00**
24					**Mid**											**High**		**1.00**
25					High											Low		1.00
26										Low						Low		1.00
27										Mid						High		1.00
28										High						Low		0.99
29						Low										High		1.00
30						Mid										Low		1.00
31						High										High		1.00
32	**IF**		**Low**									**Low**		**THEN**			**Low**	**1.00**
33			**High**									**Low**					**High**	**1.00**
34			Low									High					Low	1.00
35			High									High					Low	1.00
36								Low									Low	0.93
37								High									High	0.62

The inputs with the strongest effect indicated by the model are bolded. SN, shoot number; SL, shoot length; LA, leaf area; and MD, membership degree.

TABLE 7 | Rules for morpho-physiological quality responses (SQ, shoot quality; BC, basal callus; H, hyperhydricity) with their membership degree (MD) generated by neurofuzzy logic.

Rules		$[NH_4^+]$	$[NO_3^-]$	$[K^+]$	$[PO_4^{3-}]$	$[SO_4^{-2}]$	$[Cl^-]$	$[Fe^{2+}]$	$[BO_3^-]$	$[I^-]$		SQ	BC	H	MD
38	**IF**							Low			THEN	High			1.00
39								High				Low			1.00
40				Low		Low						Low			0.84
41				**Low**		**Mid**						**Low**			**1.00**
42				Low		High						Low			0.62
43				High		Low						Low			1.00
44				**High**		**Mid**						**High**			**1.00**
45				High		High						High			0.83
46									Low			Low			1.00
47									Mid			High			1.00
48									High			Low			0.93
49			Low									Low			1.00
50			Mid									High			1.00
51			High									Low			0.58
52	**IF**	**Low**			**Low**						**THEN**		**Low**		**1.00**
53		Mid			Low								Low		1.00
54		High			Low								Low		1.00
55		**Low**			**Mid**								**High**		**1.00**
56		Mid			Mid								High		0.92
57		High			Mid								High		1.00
58		Low			High								High		0.79
59		Mid			High								High		1.00
60		High			High								High		0.89
61						Low							High		1.00
62						Mid							High		0.75
63						High							High		1.00
64	**IF**						Low			Low	**THEN**			Low	0.60
65							**Low**			**High**				**High**	**1.00**
66							**High**			**Low**				**Low**	**1.00**
67							High			High				High	1.00
68						Low			Low					High	1.00
69						Low			High					Low	1.00
70						High			Low					High	1.00
71						High			High					Low	1.00

The inputs with the strongest effect indicated by the model are bolded. SQ, shoot quality; BC, basal callus; H, hyperhydricity; and MD, membership degree.

rule 39). The same occurs when the concentrations of BO_3^- and NO_3^- are high (**Table 7**; rules 48 and 51). However, if mid concentration of BO_3^- and NO_3^- ion are used, a truly high shoot quality will be obtained (**Table 7**; rules 47 and 50) with the highest membership degree (1.00). As an example, these results agree with the treatments #3 and 26 (**Table 3**).

The neurofuzzy model was also able to identify the key ions causing the physiological abnormalities described here (**Table 5**) and explain their effect using simple rules (**Table 7**). Thus, the model pinpoints the effect of only three ions on the development of basal callus: PO_4^{3-} × NH_4^+ and SO_4^{2-} (**Table 5**). The model highlights the positive effect of using Mid to High PO_4^{3-} concentration (**Table 7**; Rules 55–60), which favor the production of healthy shoots that show moderate or absent callus formation (**Figure 2B**), particularly at Low NH_4^+ concentration (**Table 7**, rule 55; membership 1.00). If Low PO_4^{3-} is added to the media, (membership 1.00) the formation of big and truly necrotic basal callus are promoted (**Table 7**; rules 52–54), independently of the NH_4^+ concentration. Finally, the rules 61–63 generated by the model shows that the presence of SO_4^{2-} at any concentration within the design space reduces the basal callus formation, generating healthy shoots (**Table 7**).

TABLE 8 | Composition of the standard media (MS, Murashige and Skoog; B5, Gamborg; St, Standardi; Ha, -Harada) and the predicted optimized *R* medium together with output values obtained at validation experiment (mean ± standard deviation).

	R	MS	B5	St	Ha
			Macronutrients (mg L^{-1})		
KNO_3	**1604.18**	1900	2500	1800	1900
NH_4NO_3	**1297.09**	1650	–	400	1650
$(NH_4)_2SO_4$	–	–	134	–	–
$Ca(NO_3)_2$ $4H_2O$	–	–	–	1200	–
$CaCl_2$ $2H_2O$	**545.61**	440	150	–	440
$MgSO_4$ $7H_2O$	**926.54**	370	250	360	370
KH_2PO_4	**375.63**	170	–	270	170
NaH2PO4 H2O	–	–	150	–	–
			Micronutrients (mg L^{-1})		
$MnSO_4$ $4H_2O$	**5.63**	22.30	13.20	1.0	25.00
$ZnSO_4$ $7H_2O$	**2.97**	8.60	2.00	8.60	10.00
H_3BO_3	**5.07**	6.20	3.00	6.20	10.00
KI	**1.43**	0.83	0.75	0.08	–
$CuSO_4$ $5H_2O$	**0.05**	0.025	0.025	0.025	0.025
Na_2MoO_4 $2H_2O$	**0.28**	0.25	0.25	0.25	0.25
$CoCl_2$ $6H_2O$	**0.04**	0.025	0.025	0.025	–
$FeSO_4$ $7H_2O$	**49.54**	27.85	27.85	27.85	27.85
Na_2 EDTA $2H_2O$	**53.88**	37.25	37.25	37.25	37.25
			Vitamins (mg L^{-1})		
Myo-inositol	100	100	100	100	500
Thiamine-HCl	0.1	0.1	10.0	4.0	0.5
Nicotinic-acid	0.5	0.5	1.0	–	5.0
Pyridoxine-HCl	0.5	0.5	1.0	–	0.5
Biotine	–	–	–	–	0.05
Folic acid	–	–	–	–	0.5
Glycine	2.0	2.0	–	–	2.0
			Outputs values for the validation experiment		
SN	**3.6 ± 1.12**	3.60 ± 1.01	3.51 ± 1.10	4.10 ± 1.16	4.04 ± 0.82
SL (cm)	**2.69 ± 0.55^b**	1.72 ± 0.40^d	1.88 ± 0.35^d	3.41 ± 1.20^a	2.10 ± 0.60^{cd}
LA (cm^2)	**34.06 ± 10.29^{bc}**	29.58 ± 7.62^{cd}	21.43 ± 6.86^e	43.05 ± 8.57^a	38.39 ± 11.04^{ab}
SQ	**4.51 ± 0.51^a**	4.13 ± 0.7^b	3.22 ± 0.60^c	4.02 ± 0.27^{ab}	3.87 ± 0.46^b
BC	**4.00 ± 0.00^a**	4.00 ± 0.00^a	3.00 ± 0.00^c	3.60 ± 0.50^b	4.00 ± 0.00^a
H	**2.80 ± 0.46**	2.64 ± 0.68	2.80 ± 0.55	2.67 ± 0.53	2.60 ± 0.69

Values with the same or without a letter are not significantly different at α = 0.001 (Tukey's HSD Test). Predicted optimized R media composition and validated results are bolded. SN, shoot number; SL, shoot length; LA, leaf area; SQ, shoot quality; BC, basal callus; and H, hyperhydricity.

The hyperhydricity, a well-described physiological disorder in plant tissue culture, is associated in *A. arguta* to the combined effect of four ions: $Cl^- \times I^-$ and $SO_4^{2-} \times BO_3^-$ (**Table 5**). High I^- concentration in the medium promoted no hyperhydricity (**Table 7**; rule 65 and 67), with a stronger effect if combined with low concentration of Cl^- (rule 65). Interestingly, Low BO_3^- concentration also minimized the hyperhydricity (**Table 7**; rules 69 and 71), in interaction with any concentration of SO_4^{2-}.

Designing New Optimized Medium and Experimental Validation

A database including as inputs the salt concentrations in the different culture media and as outputs all growth and quality parameter results (**Supplementary Table 1**) was modeled using INForm® software, achieving high predictability. Both the Train Set and Test Set R^2 were above 70% ($76.66 < R^2 < 96.59$) for all the parameters studied (**Supplementary Table 2**).

Genetic algorithms predicted the best combination of salts that would provide, simultaneously, the highest values for all parameters: 4.8, 2.6 cm, 39.5 cm^2 and 4.4 for the SN, SL, LA, and SQ, respectively. The new optimized medium, named "R medium," predicted higher values than MS medium used as control (**Table 3**).

The results from the experiment carried out in order to validate "R medium" are shown in **Table 8**. As can be seen, predicted and experimental (**Table 8**) are close for both *R* and MS medium. In fact, no significant differences ($\alpha < 0.001$) between the data obtained in both experiments for MS media (**Tables 3**, **8**), used as internal control, was detected.

Compared to the MS medium, the optimized *R* medium significantly improved ($\alpha < 0.001$) the two responses selected as important in the optimization process (SQ and SL) with a weight of 10 and 9, respectively (100% desirability), but not the other outputs (**Table 8**).

Regarding the other media used, the optimized *R* medium was surpassed by the St medium in promoting the growth parameters SL and LA ($\alpha < 0.001$; **Table 8**), however, *R* medium obtained a better value for the SQ ($\alpha < 0.001$) with less formation of basal calluses and hyperhydricity. On the contrary, medium B5 promoted the lowest values of SQ, BC, SL, and LA ($\alpha < 0.001$). Finally, the Ha medium promoted lower SQ and SL than *R* medium ($\alpha < 0.001$; **Table 8**).

DISCUSSION

ML as New Strategy to Predict Optimal Mineral Nutrition

Recently, algorithm-based approaches were introduced in plant tissue culture studies. Decision Trees, Chi-square Automatic Interaction Detector (CHAID), and adaptive regression splines were preferred to other ML methods such as ANNs, considering that ANNs generate "black box" models that are difficult to interpret and use (Olden et al., 2008; Akin et al., 2016, 2020). However, ANNs are powerful ML tools for plant tissue researchers, particularly when they are combined with other techniques that help in the interpretation of results or the use of the models (Gago et al., 2010b; Gallego et al., 2011). Some of their strengths are: (i) ANNs do not require a specific experimental design, so they can deal with incomplete factorial designs, trial-error series, or even historical data; (ii) they do not entail orthogonality or uniformity in the data; and (iii) subtle non-linear relationships in the data can be elucidated. The two major weakness of ANNs are the possibility of "overtraining" and the generation of "black box" models. The first limits the predictability and the second limits the possibility of using the model out of the computer used to generate it. Those limitations can be overcome by validating the model with unseen data before using it and combining the ANN model with other technology that allows knowledge to be extracted. Thus, systems that combine ANN with fuzzy logic or "neurofuzzy logic" systems allow for the obtaining of "gray box" models, providing sets of linguistic rules that help to generate knowledge about the process studied; and the combination of ANNs with GA is able to answer "How to get" questions to find the compromised solution to obtain simultaneously a set of desirable outputs. In the present work we take advantage of the use of two ML techniques, neurofuzzy logic and ANN-GA, to study the mineral nutrition of kiwiberry and address the development of an efficient *in vitro* protocol for this cultivar, as it has been carried out for other species by further innovative research groups (Zielińska and Kêpczyńska, 2013; Nezami-Alanagh et al., 2019).

The MS medium was the first and only medium described in the literature for kiwiberry *in vitro* culture (Seelye and Butcher, 1991; Matkowski and Przywara, 1995; Han et al., 2010). However, several authors have reported side effects and physiological disorders, such as shoot tip necrosis and hyperhydricity, when using it (Nowak et al., 2007; Bhojwani and Dantu, 2013; Nezami-Alanagh et al., 2018 and references therein). MS medium has also been considered supra optimal for other species of kiwifruit (Moncaleán et al., 1999, 2003), being necessary to reduce its composition by half or even more to improve performance (Monette, 1986; Akbas et al., 2007; Nasib et al., 2008).

Based on the salt composition of the MS medium, a reduced experimental design has been established (**Tables 1**, **2**). MS medium includes a combination of 14 salts. It is almost impossible to develop a complete factorial design for all of them because of the number of treatments to be assayed (e.g., 14 factors at three levels give 4,782,969 treatments). Instead, MS media salts were grouped in to five factors, which were used to define a five-dimensional design space, including just 33 treatments.

The factors selection followed the strategy used by other authors to study the shoot quality of the hazelnut, raspberry, or apricot (Wada et al., 2013; Hand et al., 2014; Poothong and Reed, 2014; Kovalchuk et al., 2017). Factors were: (i) NH_4NO_3, (ii) KNO_3, (iii) mesos, (iv) micros, and (v) iron (Reed et al., 2013a,b; Akin et al., 2017). The levels were chosen according to the maxima and minima found in the literature, in-house experience (Nezami-Alanagh et al., 2018, 2019), and the biological actions of the different nutrients (Hameg, 2019). This approach leads to more manageable and feasible research, ensuring adequate sampling of the design space (Niedz and Evens, 2006; Niedz and Evens, 2007).

The use of salts as factors creates the ion confounding effect, as has been demonstrated by some authors (Niedz and Evens, 2006), making it difficult to detect and explain the effects of individual ions on growth responses. However, it has been postulated that the specific control of some ions, such as K^+, NH_4^+, and NO_3^-, is a critical aspect in the optimization of culture media (Akin et al., 2017). In order to study the effects of the specific ions on growth parameters and/or physiological disorders of *in vitro* kiwiberry culture, the ionic composition for the different treatments was calculated (**Table 3**) and modeled by a neurofuzzy logic software. For all parameters studied, the models showed high predictability (Train Set $R^2 > 70\%$) and good accuracy. Reading the simple "IF-THEN" rules allows the acquisition of knowledge about the ions that are critical for the response parameters of kiwiberry explants. Subsequently, the ANNs-GA modeling was also carried out using the salt database, which allowed for predicting of the mineral nutrients of an optimal *R* medium, specifically adapted to kiwiberry tissue culture.

ML to Understand Shoot Growth Responses

Seven ions explain the variations in shoot number: NH_4^+, NO_3^-, SO_4^{2-}, BO_3^-, MoO_2^{2-}, Na^+, and I^-. Neurofuzzy rules (**Table 6**) state that a low MoO_2^{2-} supplement (<0.0008 mM, **Table 9**) is necessary with all Na^+ concentrations to promote a high number of shoots. Molybdate is generally added to the culture media up to 1 mM, and participates in NO_3^- catabolism (George et al., 2008). The *R* medium designed here is also proposed as 0.001 mM.

Na^+ is considered a functional but inessential element, except for C4 plants, due to its relationship with CO_2 fixation (George et al., 2008). Sotiropoulos and Dimassi (2004) have reported the beneficial effect of NaCl in the range of 10–20 mM on the *in vitro* proliferation of *Actinidia deliciosa*. Sodium is part of different salts added to the culture media, so it is usually difficult to establish its specific needs since it is affected by the ion confounding effect. However, the neurofuzzy logic model established that the Na^+ concentration within the limits of the study (0.2–1 mM) is adequate to obtain a high number of shoots for kiwiberry, especially when the molybdate is in low concentration (<0.0008 mM, **Table 9**).

Iodine has not been recognized as an essential mineral for plant nutrition, but it is included in almost 65% of the media (George et al., 2008). Chée (1986) has suggested that I^- interferes with lateral auxin transport and/or facilitates its catabolism. Generally, iodine is added to the culture media through the KI salt. The effect of adding KI to the media is controversial in the literature. Some authors (Da Silva et al., 2017) have pointed out that KI accelerates the development of plants. However, KI is not added to various media specifically designed for woody plants, such as DKW (Driver and Kuniyuki, 1984) or WPM (Lloyd and McCown, 1980). It is also not incorporated into the Ha medium, which is designed for kiwi. Neurofuzzy logic suggests that medium-high concentrations of iodine (>0.006 mM), whatever the concentration of SO_4^{2-}, contributes to improving the number of shoots.

Sulfur is an essential component and plays an important role as ligand through the -SH groups. Most culture media contain this ion in the range of 1–2.5 mM. The available form of sulfur in plants is as SO_4^{2-} ion and it is supplied in plant tissue cultures combined with other essential elements (Mg, Mn, Zn, Cu, and Fe) in at least five different salts, therefore its own role in mineral nutrition is still poorly understood due to the ion confounding effect. The neurofuzzy ML tool allows us to deduce its key role not only in the number of shoots, but in all the quality parameters studied (see below). Interestingly, the two most frequently cited media for woody plants, WPM and DKW, contain a high concentration of sulfate (7.45 and 12.39 mM, respectively). Our results demonstrate that the SO_4^{2-}, *per se*, in concentrations higher than 2.85 mM (**Table 9**), is essential to improve the number of shoots in woody fruit tree plants, which is in agreement with other authors (Poothong and Reed, 2014) who have shown that the high level of mesos and micros, including $MgSO_4$, $MnSO_4$, $ZnSO_4$, and $CuSO_4$, promotes this effect on red raspberries.

Finally, the neurofuzzy model established the positive effect of the interaction among $NH_4^+ \times NO_3^- \times BO_3^-$ on the regeneration of new shoots (SN). BO_3^- is generally supplemented in a range of 0.05–0.1 mM, being toxic at higher concentrations (0.185 mM; Bowen et al., 1979). As I^-, BO_3^- also stimulates auxin catabolism and increases its translocation (George et al., 2008), the excess of boron counteracts these important morphogenetic PGRs. Other studies showed that boron deficiency affects cell elongation more than cell division in plant growth (Martín-Rejano et al., 2011). High concentrations of BO_3^-, within the range of the study (>0.08 mM; to avoid possible toxicity), should be added to the media for obtaining high SN, independently of the levels of both nitrogen and ammonium ions.

Nitrogen sources, mainly in form of NH_4^+ and NO_3^-, are constituents of proteins, nucleic acid, and chlorophyll, being essential to plant life (George et al., 2008). In general, in media design, NO_3^- and NH_4^+ are combined as the latter acidifies the medium. The addition of NO_3^- counteracts this effect buffering the pH (George et al., 2008). Although different NO_3^-/NH_4^+ ratios have been tested, most media, as MS, have twice as much nitrate as ammonium, as a useful control of media pH. Furthermore, while high NO_3^- levels are non-toxic, high NH_4^+ levels promote physiological abnormalities such as hyperhydricity (Lips et al., 1990; Nezami-Alanagh et al., 2019). The neurofuzzy logic model points out the importance of both nitrogen ions, in combination with BO_3^- whatever its concentration, suggesting that the main role of NH_4^+ and NO_3^- could also be related with their function as pH media control, rather than only nutrients.

The shoot elongation of *A. arguta* was significantly affected by five ions (**Table 9**). With the exception of BO_3^-, the ions that explain the variability in the

TABLE 9 | Ranges (mM) and meaning of the levels (low, mid, and high) after fuzzification process by neurofuzzy logic software according to the rules **Tables 6**, **7** for each response parameter.

Inputs	Level	SN	SL	LA	SQ	BC	H	R	MS	St	Ha	B5
NH_4^+	Low	4.12 < x < 12.37				4.12 < x < 8.25						
	Mid		–	–	–	8.25 < x < 16.49	–	16.20	20.61	5.00	20.61	1.01
	High	12.37 < x < 20.61				16.49 < x < 20.61						
NO_3^-	Low	6.00 < x < 22.71		6.00 < x < 22.71	6.00 < x < 14.35							
	Mid				14.35 < x < 31.06	–	–	32.07	39.41	32.96	39.41	24.73
	High	22.71 < x < 39.41		22.71 < x < 39.41	31.06 < x < 39.41							
K^+	Low		2.19 < x < 12.37		2.19 < x < 12.37							
	Mid	–	–	–		–	–	18.64	20.05	19.79	20.05	24.73
	High		12.37 < x < 22.55	–	12.37 < x < 22.55	–	–					
Ca^{2+}	–	–	–	–	–	–	–	3.71	2.99	5.08	2.99	1.02
Mg^{2}+	Low		0.38 < x < 1.41									
	Mid	–	1.41 < x < 3.47	–	–	–	–	3.76	1.50	1.46	1.50	1.01
	High		3.47 < x < 4.50									
PO_4^{3-}	Low		0.31 < x < 1.17			0.31 < x < 1.17						
	Mid	–	1.17 < x < 2.89	–	–	1.17 < x < 2.89	–	2.76	1.25	1.98	1.25	1.09
	High		2.89 < x < 3.75			2.89 < x < 3.75						
SO_4^{2-}	Low	0.49 < x < 2.85			0.49 < x < 1.67	0.49 < x < 1.67	0.49 < x < 2.85					
	Mid		–	–	2.85 < x < 4.02	2.85 < x < 4.02		3.97	1.73	1.60	1.73	2.21
	High	2.85 < x < 5.20			4.02 < x < 5.20	4.02 < x < 5.20	2.85 < x < 5.20					
Cl^-	Low			1.50 < x < 9.73			1.50 < x < 9.73					
	Mid	–	–		–	–		7.42	5.99	0.00	5.99	2.04
	High			9.73 < x < 17.96			9.73 < x < 17.96					
Fe^{2+}	Low		0.10 < x < 0.30		0.10 < x < 0.30							
	Mid	–		–		–	–	0.18	0.10	0.10	0.10	0.10
	High		0.30 < x < 0.50		0.30 < x < 0.50							
BO^{3-}	Low	0.01 < x < 0.08	0.01 < x < 0.05		0.01 < x < 0.05		0.01 < x < 0.08					
	Mid		0.05 < x < 0.12	–	0.05 < x < 0.12	–		0.08	0.10	0.10	0.16	0.05
	High	0.08 < x < 0.15	0.12 < x < 0.15		0.12 < x < 0.15		0.08 < x < 0.15					
Mn^{2+}		–	–	–	–	–	–	0.03	0.13	0.01	0.15	0.08
Zn^{2+}		–	–	–	–	–	–	0.01	0.03	0.03	0.03	0.01
Cu^{2+}		–	–	–	–	–	–	0.0002	0.0001	0.0001	0.0001	0.0001
MoO_2^{2-}	Low	0.0001 < x < 0.0008										
	Mid		–	–	–	–	–	0.001	0.001	0.001	0.001	0.001
	High	0.0008 < x < 0.0016										
Na^+	Low	0.20 < x < 0.60		0.20 < x < 0.60								
	Mid		–		–	–	–	0.29	0.10	0.20	0.20	1.29
	High	0.60 < x < 1.00		0.60 < x < 1.00								

(Continued)

TABLE 9 | Continued

Inputs	Level	SN	SL	LA	SQ	BC	H	R	MS	St	Ha	B5
Co^{2+}		–	–	–	–	–	–	0.0002	0.0001	0.0019	0.0	0.0001
I^-	Low	0.001 < × < 0.002					0.001 < × < 0.004					
	Mid	0.002 < × < 0.06	–	–	–	–		0.008	0.005	0.0	0.0	0.005
	High	0.006 < × < 0.008					0.004 < × < 0.008					
$EDTA^-$		–	–	–	–	–	–	0.14	0.1	0.1	0.1	0.1
Na^+	Low	0.20 < × < 0.60		0.20 < × < 0.60								
	Mid		–		–	–	–	0.29	0.10	0.20	0.20	1.29
	High	0.60 < × < 1.00		0.60 < × < 1.00								
Co^{2+}		–	–	–	–	–	–	0.0002	0.0001	0.0019	0.0	0.0001
I^-	Low	0.001 < × < 0.002					0.001 < × < 0.004					
	Mid	0.002 < × < 0.06	–	–	–	–		0.008	0.005	0.0	0.0	0.005
	High	0.006 < × < 0.008					0.004 < × < 0.008					
$EDTA^-$		–	–	–	–	–	–	0.14	0.1	0.1	0.1	0.1

No ranges (–) are included in those ions not modeled. Also, the optimized R designed by ANNs-GA and MS (Murashige and Skoog), St (Standardi), Ha (Harada), and B5 (Gamborg) medium ion composition (mM) are included for easy comparison. SN, shoot number; SL, shoot length; LA, leaf area; SQ, shoot quality; BC, basal callus; and H, hyperhydricity.

elongation of the shoots are different from those that intervene in the appearance of new shoots. The model pinpointed the key positive effect of Mid Mg^{2+} ($1.41 < X < 3.47$ mM), Low Fe^{2+} (<0.3 mM), High K^+ (>12.37 mM), and Mid BO_3^- ($0.05 < X < 0.12$ mM) (**Table 9**) on shoot length growth. Finally, Mid PO_4^{3-} ($1.17 < X < 2.89$) reduced the SL.

Several studies have shown the importance of mesos salts on shoot length. Thus, it has been described that for apricot this parameter is affected by K_2SO_4 levels (Kovalchuk et al., 2017). Moreover, High mesos ($MgSO_4{\cdot}7H_2O$, and KH_2PO_4) and Low iron are required to enhance red raspberries' shoot length (Poothong and Reed, 2014, 2015). Accordingly, our results (**Table 8**), support the same pattern for another fruit tree species, *A. arguta*. The cause-effect of the specific ions is thus demonstrated, avoiding the ion confounding effect.

Within the limits of the study, to obtain a large leaf area, a combination of High NO_3^- (>22.71 mM) and Low Na^+ (<0.60 mM) concentrations (**Table 6**, rule 33) should be added to the media. Some authors have demonstrated that High KNO_3 and NH_4NO_3 levels improve the number and size of leaves in pear genotypes (Ibrahim et al., 2008; Reed et al., 2013a). Also, leaf area can be affected by NO_3^- factor as has been shown for red raspberry (Poothong and Reed, 2016). All these findings agree with our results (**Table 8**).

ML to Understand Morphophysiological Disorders and Shoot Quality Responses

Morphophysiological disorders in *in vitro* cultures are caused by a wide variety of factors (Hazarika, 2006) and avoiding them is one of the greatest challenges in shoots micropropagation. In this study, important abnormalities such as BC and hyperhydricity were observed (**Figure 3**) and neurofuzzy logic was able to determine the ions that are directly related to the appearance of them.

The effect of three ions on the development of basal callus were pointed out; $PO_4^{3-} \times NH_4^+$ and SO_4^{2-} were included in the factors nitrogen (NH_4NO_3) and mesos ($MgSO_4$). The model highlighted the positive effect of Mid-High PO_4^{3-} (>1.17 mM) to favor the production of healthy shoots with moderate or absent callus formation (**Figure 2B**), particularly when Low NH_4^+ (<8.25 mM) are used, in agreement with other reports (Niedz and Evens, 2007; Kovalchuk et al., 2018). If Low PO_4^{3-} is added to the media, necrosis and the formation of big basal callus are significantly promoted, whatever the NH_4^+ concentration. Finally, SO_4^{2-} at any concentration tested in this space of design reduces the basal callus formation, generating healthy shoots (High BC). The beneficial effect of high concentrations of these ions on callus formation has been previously demonstrated for other species, such as pear (Reed et al., 2013b), raspberries (Poothong and Reed, 2014), or hazelnut (Akin et al., 2017, 2020). MS medium promotes higher callus formation than WPM and DKV in pistachio culture (Nezami-Alanagh et al., 2019),

probably due to its lower and higher levels of NH_4^+ and SO_4^{2-}, respectively.

Hyperhydricity is one of the main morpho-physiological disorders in the micropropagation of plants that has been associated with alterations in mineral composition, hormonal imbalances, or the use of gelling agents (Pâques, 1991; Hazarika, 2006). Within the range of the study, kiwiberry also showed hyperhydricity for some treatments (**Figure 3B**). The hyperhydricity was caused by the interaction of four ions: Cl^- × I^- and SO_4^{2-} × BO_3^-. The combination of High I^- (>0.004 mM) and Low BO_3^- (<0.08 mM) concentrations in the medium avoided hyperhydricity formation, whatever the concentration of Cl^- and SO_4^{2-}, respectively. High percentages of hyperhydricity in *Prunus* and *Dianthus caryophyllus* cultures have been associated with high concentrations of Cl^- in the culture medium (Quoirin and Lepoivre, 1977; Dantas et al., 2001). In contrast, high concentrations of mesos ($CaCl_2$, KH_2PO_4, and $MgSO_4$) in pear cultures reduce hyperhydricity (Reed et al., 2013b).

The shoot quality parameter integrates, in some way, both the growth parameters, and the absence of abnormalities and physiological disorders. The neurofuzzy logic model selected five ions as critical factors for the quality of the shoots: K^+, SO_4^{2-}, Fe^{2+}, BO_3^-, and NO_3^-. All of them contribute in a way to the growth parameters (SN, SL, or LA), but some are also involved in the appearance of abnormalities as SO_4^{2-} (BC and H) or BO_3^- (H).

Rules reveal that good shoot quality is achieved when a High K^+ (>12.35 mM) concentration is supplied to the medium along with concentrations of Fe^{2+} and NO_3^- within the ranges of 0.10–0.30 mM and 14.35–31.06 mM, respectively, in agreement with other authors for other cultivars (Akin et al., 2017; Kovalchuk et al., 2017) and also with the predictions for the growth parameters of kiwiberry. Medium SO_4^{2-} (2.85–4.02 mM) concentration is also necessary, since it promotes a high number of shoots and plantlets with low callus formation. Finally, BO_3^- must be in 0.05–0.12 mM due to its effect on the number of shoots and hyperhydricity (**Table 9**).

ML to Predict Optimal Salt Composition and Experimental Validation

Different ML models based on artificial neural networks, fuzzy logic, genetic algorithms, or gene expression programming algorithms has been previously employed with success for predicting optimal *in vitro* culture media of fruit tree species such as *Prunus* (Nezami-Alanagh et al., 2014; Arab et al., 2016, 2018), pear (Jamshidi et al., 2016, 2019), or pistachio rootstocks (Nezami-Alanagh et al., 2017). In this study, ML tools, including ANNs-GA were selected to build a model based on salt composition of culture media for each growth parameter. All of them have good predictability (Train and Test Set R > 70%). The utility of GA allows the estimation of the best combination of salts to obtain a set of desired values for each parameter (maximal growth parameters). The model predicts for a medium of optimal composition (*R* medium) values of SN, SL, LA, and SQ of 4.8 shoots, 2.6 cm, 39.5 cm^2, and 4.4, respectively.

The experimental values obtained for kiwiberry culture using the *R* medium composition validate those predicted by the model. Even more, the comparison of the results obtained with the *R* medium and the four media (MS, St, Ha, and B5), used as controls in the validation experiment, shows that the optimized *R* medium outperforms the others in terms of SL, LA, and SQ. However, the SN parameter appears to be a bit overestimated (4.8 versus 3.6). Only, St basal medium promoted statistically significant ($\alpha < 0.001$) larger SL and LA.

Differences in mineral composition among all media (**Table 9**) reveals that St is the poorest media in NH_4^+ and Cl^- but rich in Ca^{2+}, as is other media used for woody plants such as DKW (Driver and Kuniyuki, 1984) and WPM (Lloyd and McCown, 1980), while B5 included low nitrogen but also the lowest mesos (Ca^{2+}, Mg^{2+}, and PO_4^{3-}) concentration among the media tested, but the highest Na^+ (**Table 9**; Hameg et al., 2018). However, the optimized *R* medium adjusted at $\frac{3}{4}$ total nitrogen of MS content, but increased 2× all mesos (Ca^{2+}, Mg^{2+}, PO_4^{3-}, and SO_4^{2-}) and also increased almost all micros (Cl^-, Cu^{2+}, Na^+, Co^{2+}, and I^-) and iron (Fe^{2+} and $EDTA^-$) with respect to MS (**Table 9**). The increasing levels of micronutrients over the level in MS promoted cell growth and morphogenesis in some species (George et al., 2008). With those adjustments, the medium *R* promoted better results than MS, particularly shoot quality. In conclusion, it is clear that it is the interactions among the ions, rather than their independent effect, that caused the described results. Thus, it is multivariable analysis, rather than single-factor analysis, that is required to really understand media component relationships. Finally, another important fact is that those media (St, Ha, and B5) included different vitamin contents and glycine, not included in this optimization. Furthermore, the PGR effect on organogenesis and growth was not studied, because all media were supplemented with the same PGRs.

CONCLUSION

The suitability of computer-based tools, such as DOE and ML, as a new strategy to design tissue culture media for kiwiberry has been stated. DOE allowed the plant cell tissue researchers to perform well sampled and efficient experiments in order to save time and plant material. ML tools allowed for the extraction of information to clarify the complex non-linear interactions between variables and understand the effects of single ions on growth parameters and morpho-physiological disorders. A new medium, named *R* medium, was established with excellent results. The designed *R* medium differs from MS by reducing up to 20% nitrogen, increasing almost 200% mesos, 100% micros, and 50% iron factor concentrations and performs better for kiwiberry. The *R* medium also performs better than the B5, Ha, and St media, since although some of them have slight advantages in terms of growth parameters, they also promote more physiological disorders. The *R* medium could be improved considering the effects of other key components of the media that have not been studied in this work, such as vitamins, PGR, or

organic compounds, particularly glycine, that can modulate the effect of ions. They need further additional research.

DATA AVAILABILITY STATEMENT

The original contributions generated for this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

RH and TA performed micropropagation experiments. RH, TA, and MB performed the statistical analysis. ML and PG contributed to machine learning models supervision. MB and PG conceived and designed the experiments. All authors equally contributed to the discussions for data interpretation and drawing the conclusions. All authors contributed to the writing and agreed to the published version of the manuscript.

FUNDING

The funding for this work was provided by Xunta de Galicia through "Red de Uso Sostenible de los Recursos Naturales y Agroalimentarios" (REDUSO, Grant number ED431D 2017/18) and CITACA Strategic Partnership, Grant number ED431E 2018/07).

ACKNOWLEDGMENTS

The authors acknowledge the Erasmus Mundus GREENIT grant from the European Commission (reference 2012-2625/001-001-EMA) to RH and the Ph.D. grant from University of Vigo to TA.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.554905/full#supplementary-material

Supplementary Figure 1 | Determination coefficient (R^2) between the predicted values from the neurofuzzy logic versus experimental values for each factor expressed as determination coefficient (R^2): **(A)** SN, shoot number; **(B)** SL, shoot length; **(C)** LA, leaf area; **(D)** SQ, shoot quality; **(E)** BC, basal callus formation; and **(F)** H, hyperhydricity.

Supplementary Figure 2 | Examples of graphical representation of the fuzzyfication process developed by neurofuzzy logic per NO_3^- **(A)** and K^{++} **(B)** and their respective domains Low–Mid–High and Low–High.

Supplementary Table 1 | Mineral nutrients' (expressed as salt concentrations) composition of the different culture media based on the five-factor experimental design (0–33) and response values of the parameters (mean and standard deviation) used to characterize plant growth. Original medium composition (bold) used as control. SN, shoot number; SL, shoot length; LA, leaf area; SQ, shoot quality; BC, basal callus; and H, hyperhydricity.

Supplementary Table 2 | Artificial neural network model train set R^2 and test set R^2. MSE (Mean Squared Error).

REFERENCES

Akbas, F. A., Isikalan, C., Namli, S., and Basaran, D. (2007). Micropropagation of kiwifruit (*Actinidia deliciosa*). *Int. J. Agric. Biol.* 9, 489–493.

Akin, M., Eyduran, E., Niedz, R. P., and Reed, B. M. (2017). Developing hazelnut tissue culture medium free of ion confounding. *Plant Cell Tissue Organ Cult.* 130, 483–494.doi: 10.1007/s11240-017-1238-z

Akin, M., Eyduran, E., and Reed, B. M. (2016). Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut. *Plant Cell Tissue Organ Cult.* 128, 303–316.doi: 10.1007/s11240-016-1110-6

Akin, M., Eyduran, S. P., Eyduran, E., and Reed, B. M. (2020). Analysis of macro nutrient related growth responses using multivariate adaptive regression splines. *Plant Cell Tissue Organ Cult.* 140, 661–670.doi: 10.1007/s11240-019-01763-8

Arab, M. M., Yadollahi, A., Eftekhari, M., Ahmadi, H., Akbari, M., and Khorami, S. S. (2018). Modeling and optimizing a new culture medium for *in vitro* rooting of G×N15 *Prunus* rootstock using artificial neural network-genetic algorithm. *Sci. Rep.* 8:9977. doi: 10.1038/s41598-018-27858-4

Arab, M. M., Yadollahi, A., Shojaeiyan, A., and Ahmadi, H. (2016). Artificial neural network genetic algorithm as powerful tool to predict and optimize *in vitro* proliferation mineral medium for G × N15 rootstock. *Front. Plant Sci.* 7:1526. doi: 10.3389/fpls.2016.01526

Arteta, T. A., Hameg, R., Landin, M., Gallego, P. P., and Barreal, M. E. (2018). Neural networks models as decision-making tool for *in vitro* proliferation of hardy kiwi. *Eur. J. Hortic. Sci.* 83, 259–265. doi: 10.17660/eJHS.2018/83.4.6

Bhojwani, S. S., and Dantu, P. K. (2013). "Micropropagation," in *Plant Tissue Culture: An Introductory Text*, eds S. S. Bhojwani and K. P. Dantu (New Delhi: Springer), 245–274. doi: 10.1007/978-81-322-1026-9_17

Bowen, R. D., Williams, D. H., and Schwarz, H. (1979). The chemistry of isolated cations. *Angew Chem. Int. Edit.* 18, 451–461. doi: 10.1002/anie.197904511

Bresinsky, A., Körner, C., Kadereit, J. W., Neuhaus, G., and Sonnewald, U. (2013). *Strasburger's Plant Sciences: Including Prokaryotes and Fungi*. Berlin: Springer.

Chée, R. (1986). *In vitro* culture of *Vitis*: the effects of light spectrum, manganese sulfate and potassium iodide on morphogenesis. *Plant Cell Tiss Org Cult.* 7, 121–134.doi: 10.1007/BF00043036

Cheng, T. Y. (1975). Adventitious bud formation in culture of Douglas fir (*Pseudotsuga menziesii* (MIRB.) Franco). *Plant Sci. Lett.* 5, 97–102.doi: 10.1016/0304-4211(75)90049-8

Colbourn, E., and Rowe, R. (2005). "Neural computing and pharmaceutical formulation," in *Encyclopaedia of Pharmaceutical Technology*, eds J. Swarbrick and J. Boylan (New York, NY: Marcel Dekker), 145–157.

Da Silva, F., Nascimento, A., Barbosa, L., and Magalhães, H. (2017). *In vitro* cultivation of purple basil *Ocimum basilicum* L. 'red rubin' at different levels of salts, charcoal, sucrose and potassium iodine. *Aust. J. Crop. Sci.* 11, 1137–1145.doi: 10.21475/ajcs.17.11.09.pne624

Dantas, A. K., Majada, J. P., Fernández, B., and Cañal, M. J. (2001). Mineral nutrition in carnation tissue cultures under different ventilation conditions. *Plant Growth Regul.* 33, 237–243. doi: 10.1023/A:1017542106252

Design-Expert (2010). *Design-Expert*. Minneapolis, MN: Stat-Ease, Inc.

Driver, J. A., and Kuniyuki, A. H. (1984). In vitro propagation of Paradox walnut rootstock. *HortScience* 19, 507–509.

Epstein, E. (1972). *Mineral Nutrition of Plants: Principles and Perspectives*. London: John Wiley and Sons, Inc.

Gago, J., Landín, M., and Gallego, P. P. (2010a). Artificial neural networks modeling the *in vitro* rhizogenesis and acclimatization of *Vitis vinifera* L. *J. Plant Physiol.* 167, 1226–1231.doi: 10.1016/j.jplph.2010.04.008

Gago, J., Landín, M., and Gallego, P. P. (2010b). Strengths of artificial neural networks in modeling complex plant processes. *Plant Signal Behav.* 5, 743–745.doi: 10.4161/psb.5.6.11702

Gago, J., Martínez-Núñez, L., Landín, M., and Gallego, P. P. (2010c). Artificial neural networks as an alternative to the traditional statistical methodology in plant research. *J. Plant Physiol.* 167, 23–27.doi: 10.1016/j.jplph.2009.07.007

Gago, J., Pérez-Tornero, O., Landín, M., Burgos, L., and Gallego, P. P. (2011). Improving knowledge of plant tissue culture and media formulation by

neurofuzzy logic: a practical case of data mining using apricot databases. *J. Plant Physiol.* 168, 1858–1865.doi: 10.1016/j.jplph.2011.04.008

Gallego, P. P., Gago, J., and Landín, M. (2011). "Artificial neural networks technology to model and predict plant biology process," in *Artificial Neural Networks-Methodological Advances and Biomedical Applications*, ed. K. Suzuki (London: IntechOpen), 197–216. doi: 10.5772/14945

Gamborg, O. L., Miller, R. A., and Ojima, K. (1968). Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell Res.* 50, 151–158.doi: 10.1016/0014-4827(68)90403-5

George, E. F., Hall, M. A., and Klerk, G.-J. D. (2008). "The components of plant tissue culture media I: macro- and micro-nutrients," in *Plant Propagation by Tissue Culture*, eds E. F. George, M. A. Hall, and G. J. D. Klerk (Dordrecht: Springer Netherlands), 65–113. doi: 10.1007/978-1-4020-5005-3_3

Hameg, R. (2019). *Computer-Based Tools for Designing Plant Tissue Culture Media*. Ph.D. thesis, University of Vigo, Vigo.

Hameg, R., Arteta, T., Gallego, P. P., and Barreal, M. E. (2018). Selecting an efficient proliferation medium for *Actinidia arguta* 'Issai' explants. *Acta Hortic.* 1218, 565–572.doi: 10.17660/ActaHortic.2018.1218.77

Hameg, R., Gallego, P. P., and Barreal, M. E. (2017). *In vitro* establishment and multiplication of hardy kiwi (*Actinidia arguta* 'Issai'). *Acta Hortic.* 1187, 51–58.doi: 10.17660/ActaHortic.2017.1187.6

Han, M., Gleave, A. P., and Wang, T. (2010). Efficient transformation of *Actinidia arguta* by reducing the strength of basal salts in the medium to alleviate callus browning. *Plant Biotechnol. Rep.* 4, 129–138.doi: 10.1007/s11816-010-0128-1

Hand, C., Maki, S., and Reed, B. M. (2014). Modeling optimal mineral nutrition for hazelnut micropropagation. *Plant Cell Tissue Organ Cult.* 119, 411–425.doi: 10.1007/s11240-014-0544-y

Harada, H. (1975). *In vitro* organ culture of *Actinidia chinensis* PL as a technique for vegetative multiplication. *J. Hortic. Sci.* 50, 81–83.doi: 10.1080/00221589.1975.11514606

Hazarika, B. N. (2006). Morpho-physiological disorders in *in vitro* culture of plants. *Sci. Hortic.* 108, 105–120.doi: 10.1016/j.scienta.2006.01.038

Hildebrandt, A. C., Riker, A. J., and Duggar, B. M. (1946). The influence of the composition of the medium on growth *in vitro* of excised tobacco and sunflower tissue cultures. *Am. J. Bot.* 33, 591–597.doi: 10.2307/2437399

Ibrahim, I. A., Nasr, M. I., Mohammedm, B. R., and El-Zefzafi, M. M. (2008). Nutrient factors affecting *in vitro* cultivation of *Stevia rebaudiana*. *Sugar Tech.* 10, 248–253.doi: 10.1007/s12355-008-0044-7

Jamshidi, S., Yadollahi, A., Ahmadi, H., Arab, M. M., and Eftekhari, M. (2016). Predicting *in vitro* culture medium macro-nutrients composition for pear rootstocks using regression analysis and neural network models. *Front. Plant Sci.* 7:274. doi: 10.3389/fpls.2016.00274

Jamshidi, S., Yadollahi, A., Arab, M. M., Soltani, M., Eftekhari, M., Sabzalipoor, H., et al. (2019). Combining gene expression programming and genetic algorithm as a powerful hybrid modeling approach for pear rootstocks tissue culture media formulation. *Plant Methods* 15:136. doi: 10.1186/s13007-019-0520-y

Kabaluk, J., Kempler, C., and Toivonen, P. (1997). *Actinidia arguta*: Characteristics relevant to commercial production. *Fruit Var J.* 51, 117–122.

Kavli, T., and Weyer, E. (1994). ASMOD (Adaptive Spline Modelling of Observation Data): some theoretical and experimental results. *IEE Colloq (Digest)* 136, 1–7.

Kovalchuk, I. Y., Mukhitdinova, Z., Turdiyev, T., Madiyeva, G., Akin, M., Eyduran, E., et al. (2017). Modeling some mineral nutrient requirements for micropropagated wild apricot shoot cultures. *Plant Cell Tissue Organ Cult.* 129, 325–335.doi: 10.1007/s11240-017-1180-0

Kovalchuk, I. Y., Mukhitdinova, Z., Turdiyev, T., Madiyeva, G., Akin, M., Eyduran, E., et al. (2018). Nitrogen ions and nitrogen ion proportions impact the growth of apricot (*Prunus armeniaca*) shoot cultures. *Plant Cell Tissue Organ Cult. (PCTOC)* 133, 263–273.doi: 10.1007/s11240-018-1379-8

Landin, M., and Rowe, R. C. (2013). "Artificial neural networks technology to model, understand, and optimize drug formulations," in *Formulation Tools for Pharmaceutical Development*, ed. J. E. Aguilar (Cambridge: Woodhead Publishing), 7–37. doi: 10.1533/9781908818508.7

Landín, M., Rowe, R. C., and York, P. (2009). Advantages of neurofuzzy logic against conventional experimental design and statistical analysis in studying and developing direct compression formulations. *Eur. J. Pharm. Sci.* 38, 325–331.doi: 10.1016/j.ejps.2009.08.004

Lips, S. H., Leidi, E. O., Silberbush, M., Soares, M. I. M., and Lewis, O. E. M. (1990). Physiological aspects of ammonium and nitrate fertilization. *J. Plant. Nutr.* 13, 1271–1289.doi: 10.1080/01904169009364151

Lloyd, G., and McCown, B. (1980). Commercially-feasible micropropagation of mountain laurel, *kalmia latifolia*, by use of shoot-tip culture. *Proc. Int. Plant Prop. Soc.* 30, 421–427.

Martín-Rejano, E. M., Camacho-Cristóbal, J. J., Herrera-Rodríguez, M. B., Rexach, J., Navarro-Gochicoa, M. T., and González-Fontes, A. (2011). Auxin and ethylene are involved in the responses of root system architecture to low boron supply in *Arabidopsis* seedlings. *Physiol. Plant.* 142, 170–178.doi: 10.1111/j.1399-3054.2011.01459.x

Matkowski, A., and Przywara, L. (1995). Callus induction and plant regeneration *in vitro* in *Actinidia*. *Acta Soc. Bot. Pol.* 64, 131–138. doi: 10.5586/asbp.1995.017

Moncaleán, P., Cañal, M. J., Feito, I., Rodríguez, A., and Fernández, B. (1999). Cytokinins and mineral nutrition in *Actinidia deliciosa* (Kiwi) shoots cultured *in vitro*. *J. Plant Physiol.* 155, 606–612. doi: 10.1016/S0176-1617(99)80061-3

Moncaleán, P., Cañal, M. J., Fernández, H., Fernández, B., and Rodríguez, A. (2003). Nutritional and gibberellic acid requirements in kiwifruit vitroponic cultures. *In Vitro Cell Dev. Biol. Plant* 39, 49–55.doi: 10.1079/IVP2002371

Monette, P. L. (1986). Micropropagation of kiwifruit using non-axenic shoot tips. *Plant Cell Tissue Organ Cult.* 6, 73–82.doi: 10.1007/BF00037760

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Nasib, A., Ali, K., and Khan, S. (2008). An optimized and improved method for the *in vitro* propagation of kiwifruit (*Actinidia deliciosa*) using coconut water. *Pak J. Bot.* 40, 2355–2360.

Nezami-Alanagh, E., Garoosi, G. A., Haddad, R., Maleki, S., Landín, M., and Gallego, P. P. (2014). Design of tissue culture media for efficient *Prunus* rootstock micropropagation using artificial intelligence models. *Plant Cell Tissue Organ Cult.* 117, 349–359.doi: 10.1007/s11240-014-0444-1

Nezami-Alanagh, E., Garoosi, G. A., Landín, M., and Gallego, P. P. (2018). Combining DOE with neurofuzzy logic for healthy mineral nutrition of pistachio rootstocks *in vitro* culture. *Front. Plant Sci.* 9:1474. doi: 10.3389/fpls.2018.01474

Nezami-Alanagh, E., Garoosi, G.-A., Landín, M., and Gallego, P. P. (2019). Computer-based tools provide new insight into the key factors that cause physiological disorders of pistachio rootstocks cultured *in vitro*. *Sci. Rep.* 9:9740. doi: 10.1038/s41598-019-46155-2

Nezami-Alanagh, E., Garoosi, G. A., Maleki, S., Landín, M., and Gallego, P. P. (2017). Predicting optimal *in vitro* culture medium for *Pistacia vera* micropropagation using neural networks models. *Plant Cell Tissue Organ Cult.* 129, 19–33.doi: 10.1007/s11240-016-1152-9

Niedz, R. P., and Evens, T. J. (2006). A solution to the problem of ion confounding in experimental biology. *Nat. Methods* 3, 417–417.doi: 10.1038/nmeth0606-417

Niedz, R. P., and Evens, T. J. (2007). Regulating plant tissue growth by mineral nutrition. *In Vitro Cell Dev. Biol. Plant* 43, 370–381.doi: 10.1007/s11627-007-9062-5

Niedz, R. P., and Evens, T. J. (2016). Design of experiments (DOE)—history, concepts, and relevance to *in vitro* culture. *In Vitro Cell Dev. Biol Plant* 52, 547–562.doi: 10.1007/s11627-016-9786-1

Nowak, B., Miczyński, K., and Hudy, L. (2007). The effect of total inorganic nitrogen and the balance between its ionic forms on adventitious bud formation and callus growth of 'Wêgierka Zwykła' plum (*Prunus domestica* L.). *Acta Physiol. Plant* 29, 479–484.doi: 10.1007/s11738-007-0058-x

Olden, J. D., Lawler, J. J., and Poff, N. L. R. (2008). Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biol.* 83, 171–193.doi: 10.1086/587826

Pâques, M. (1991). Vitrification and micropropagation: causes, remedies and prospects. *Acta Hortic.* 289, 283–290.doi: 10.17660/ActaHortic.1991.289.76

Paradela, M., De Paz, P., and Gallego, P. (2001). Variability in the organogenic response of kiwi from material undergoing micropropagation. *Melhoramiento* 37, 79–83.

Poothong, S., and Reed, B. M. (2014). Modeling the effects of mineral nutrition for improving growth and development of micropropagated red raspberries. *Sci. Hortic.* 165, 132–141.doi: 10.1016/j.scienta.2013.10.040

Poothong, S., and Reed, B. M. (2015). Increased $CaCl_2$, $MgSO_4$, and KH_2PO_4 improve the growth of micropropagated red raspberries. *In Vitro Cell Dev. Biol. Plant* 51, 648–658.doi: 10.1007/s11627-015-9720-y

Poothong, S., and Reed, B. M. (2016). Optimizing shoot culture media for *Rubus* germplasm: the effects of NH_4^+, NO_3^-, and total nitrogen. *In Vitro Cell Dev. Biol Plant* 52, 265–275.doi: 10.1007/s11627-016-9750-0

Quoirin, M., and Lepoivre, P. (1977). Improved media for *in vitro* culture of *Prunus* sp. *Acta Hortic.* 77, 437–442. doi: 10.17660/actahortic.1977.78.54

Ramage, C. M., and Williams, R. R. (2002). Mineral nutrition and plant morphogenesis. *In Vitro Cell Dev. Biol. Plant* 38, 116–124.doi: 10.1079/IVP2001269

Reed, B. M., and Hummer, K. E. (1995). "Conservation of germplasm of strawberry (*Fragaria* species)," in *Cryopreservation of Plant Germplasm I Biotechnology in Agriculture and Forestry*, ed. Y. P. S. Bajaj (Berlin: Springer), 354–370. doi: 10.1007/978-3-662-03096-7_25

Reed, B. M., Wada, S., DeNoma, J., and Niedz, R. P. (2013a). Improving *in vitro* mineral nutrition for diverse pear germplasm. *In Vitro Cell Dev. Biol. Plant* 49, 343–355.doi: 10.1007/s11627-013-9504-1

Reed, B. M., Wada, S., DeNoma, J., and Niedz, R. P. (2013b). Mineral nutrition influences physiological responses of pear *in vitro*. *In Vitro Cell Dev. Biol. Plant* 49, 699–709.doi: 10.1007/s11627-013-9556-2

Revilla, M. A., Rey, M. A., Gonzalez-Rio, F., Gonzalez, M. V., Diaz-Sala, C., and Rodriguez, R. (1992). "Micropropagation of Kiwi (*Actinidia* spp.)," in *High-Tech and Micropropagation II*, ed. Y. P. S. Bajaj (Berlin: Springer), 399–423. doi: 10.1007/978-3-642-76422-6_21

Seelye, J. F., and Butcher, S. M. (1991). *In vitro* response of *Actinidia* leaf and callus tissue to thidiazuron. *N. Z. J. Crop. Hortic. Sci.* 19, 447–450. doi: 10.1080/01140671.1991.10422891

Shao, Q., Rowe, R. C., and York, P. (2006). Comparison of neurofuzzy logic and neural networks in modelling experimental data of an immediate release tablet formulation. *Eur. J. Pharm. Sci.* 28, 394–404. doi: 10.1016/j.ejps.2006.04.007

Sonnewald, U. (2013). "Physiology of metabolism," in *Strasburger's Plant Sciences: Including Prokaryotes and Fungi*, eds A. Bresinsky, C. Körner, J. W. Kadereit, G. Neuhaus, and U. Sonnewald (Berlin: Springer), 239–409. doi: 10.1007/978-3-642-15518-5_5

Sotiropoulos, T. E., and Dimassi, K. N. (2004). Response to increasing rates of boron and NaCl on shoot proliferation and chemical composition of *in vitro* kiwifruit shoot cultures. *Plant Cell Tissue Organ Cult.* 79, 285–289.doi: 10.1007/s11240-004-4609-1

Standardi, A. (1981). Micropropagazione dell' *Actinidia chinensis* Pl. mediante coltura *in vitro* di apici meristematici. *Frutticoltura* 43, 23–27.

StatSoft-Inc. (2014). *STATISTICA, Version 12*. Tulsa, OK: StatSoft Inc.

Wada, S., Niedz, R. P., DeNoma, J., and Reed, B. M. (2013). Mesos components ($CaCl_2$, $MgSO_4$, KH_2PO_4) are critical for improving pear micropropagation. *In Vitro Cell Dev Biol Plant*. 49, 356–365. doi: 10.1007/s11627-013-9508-x

Zielińska, S., and Kêpczyńska, E. (2013). Neural modeling of plant tissue cultures: a review. *BioTechnologia* 94, 253–268. doi: 10.5114/bta.2013.46419

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hameg, Arteta, Landin, Gallego and Barreal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fpls.2020.624273

Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean

*Mohsen Yoosefzadeh-Najafabadi[1], Hugh J. Earl[1], Dan Tulpan[2], John Sulik[1] and Milad Eskandari[1]**

[1] Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada, [2] Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada

Recent substantial advances in high-throughput field phenotyping have provided plant breeders with affordable and efficient tools for evaluating a large number of genotypes for important agronomic traits at early growth stages. Nevertheless, the implementation of large datasets generated by high-throughput phenotyping tools such as hyperspectral reflectance in cultivar development programs is still challenging due to the essential need for intensive knowledge in computational and statistical analyses. In this study, the robustness of three common machine learning (ML) algorithms, multilayer perceptron (MLP), support vector machine (SVM), and random forest (RF), were evaluated for predicting soybean (*Glycine max*) seed yield using hyperspectral reflectance. For this aim, the hyperspectral reflectance data for the whole spectra ranged from 395 to 1005 nm, which were collected at the R4 and R5 growth stages on 250 soybean genotypes grown in four environments. The recursive feature elimination (RFE) approach was performed to reduce the dimensionality of the hyperspectral reflectance data and select variables with the largest importance values. The results indicated that R5 is more informative stage for measuring hyperspectral reflectance to predict seed yields. The 395 nm reflectance band was also identified as the high ranked band in predicting the soybean seed yield. By considering either full or selected variables as the input variables, the ML algorithms were evaluated individually and combined-version using the ensemble–stacking (E–S) method to predict the soybean yield. The RF algorithm had the highest performance with a value of 84% yield classification accuracy among all the individual tested algorithms. Therefore, by selecting RF as the metaClassifier for E–S method, the prediction accuracy increased to 0.93, using all variables, and 0.87, using selected variables showing the success of using E–S as one of the ensemble techniques. This study demonstrated that soybean breeders could implement E–S algorithm using either the full or selected spectra reflectance to select the high-yielding soybean genotypes, among a large number of genotypes, at early growth stages.

Keywords: artificial intelligence, data-driven model, ensemble methods, high-throughput phenotyping, random forest, recursive feature elimination

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Omar Vergara-Diaz, University of Barcelona, Spain
Michael Gomez Selvaraj, Consultative Group on International Agricultural Research (CGIAR), United States

****Correspondence:***
Milad Eskandari
meskanda@uoguelph.ca

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *31 October 2020*
Accepted: *10 December 2020*
Published: *12 January 2021*

Citation:
Yoosefzadeh-Najafabadi M, Earl HJ, Tulpan D, Sulik J and Eskandari M (2021) Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean. Front. Plant Sci. 11:624273. doi: 10.3389/fpls.2020.624273

INTRODUCTION

The world population is projected to exceed nine billion individuals by 2050, which will require significant improvements in the yield of major crops that contribute to global food security (Tilman et al., 2009; Foley et al., 2011; Alexandratos and Bruinsma, 2012; Dubey et al., 2019). Increasing the yield is the primary goal of most plant breeding programs for major crops, such as soybean (*Glycine max*), which is the world's most widely grown leguminous crop and an important source of protein and oil for food and feed (Hartman et al., 2011). In the area of plant breeding, however, measuring primary traits, such as yield, which is under influenced by a combination of quantitative and qualitative traits, in large breeding populations consisting of several thousand genotypes is time and labor-consuming (Araus and Cairns, 2014; Cai et al., 2016; Xiong et al., 2018). Breeding for yield is known as a highly complex and non-linear process due to the genetic and environmental factors (Collins et al., 2008). Therefore, breeding approaches that are established based on secondary traits (e.g., yield component traits and reflectance bands), which are strongly correlated with the primary trait, enable plant breeders to efficiently recognize promising lines at early growth stages (Ma et al., 2001; Jin et al., 2010; Montesinos-López et al., 2017).

The combination of high-throughput genotyping and phenotyping technologies have enabled plant breeders to make their early growth stage selections more accurate while it reduced the evaluation time and cost in their breeding programs (Rutkoski et al., 2016). Although there has been significant progress in high-throughput genotyping in recent years with a direct impact on current plant breeding challenges (Araus and Cairns, 2014; Tardieu et al., 2017; Araus et al., 2018), acquisition of high-throughput field phenotyping is still a bottleneck in most breeding programs (Furbank and Tester, 2011; Araus et al., 2018).

Remote sensing of spectral reflectance is considered as an efficient high-throughput phenotyping tool (Araus and Cairns, 2014; Tardieu et al., 2017), which aims to measure the spectral reflectance efficiently at several plant growth and development stages in large breeding populations (Rutkoski et al., 2016). It is well documented that the spectral properties are genotype-specific and influenced by the anatomy, morphology, and physiology of plants (Kycko et al., 2018; Schweiger et al., 2018) and, therefore, can be used for screening plant genotypes with different agronomic potentials.

Analyzing large datasets consisting of spectral reflectance data requires intensive computational and statistical analyses, which is still challenging in many plant breeding programs (Lopez-Cruz et al., 2020). Nowadays, machine learning algorithms have drawn attention from researchers to develop model-based breeding methods that can improve the efficiency of breeding processes (Hesami et al., 2020a). Recently, one of the most common artificial neural networks (ANNs), the multilayer perceptron (MLP) developed by Pal and Mitra (1992), has been broadly used for modeling and predicting complex traits, such as yield, in different breeding programs (Geetha, 2020). MLP can be considered as a non-linear computational method employed for various tasks such as classification and regression of complex systems (Chen and Wang, 2020; Hesami et al., 2020b). This algorithm is able to detect the connection and relationship between the input and output (target) variables and quantify the inherent knowledge existing in the datasets (Ghorbani et al., 2016; Hesami et al., 2020b). This algorithm includes several highly interconnected processing neurons that can be used in parallel to detect a solution for a specific problem (Ghorbani et al., 2016; Geetha, 2020). Support vector machines (SVMs), developed by Vapnik (2000), are known as one of the powerful and easy to use machine learning algorithms that can recognize patterns and behavior of non-linear relationships (Auria and Moro, 2008; Su et al., 2017). Some of the advantages of SVMs over MLP are linked to the complexity of the networks. SVMs usually use a large number of learning problem formulations leading to solving a quadratic optimization problem (Feng et al., 2020; Hesami et al., 2020b). In theory, SVM has to be better performance because of using structural risk minimization inductive principles rather than the empirical risk minimization inductive principle (Belayneh et al., 2014). In addition to MLP and SVM, random forest (RF) (Breiman, 2001) is another method for data modeling with a computational efficient training phase and very high generalization accuracy. RF has been broadly used in areas such as object recognition (Lepetit et al., 2005), skin detection (Khan et al., 2010), plant phenomics (Falk et al., 2020), and genomics (Mokry et al., 2013).

Machine learning algorithms are subject to overfitting, mainly because of limited training data and dependent on single predictive models (Ali et al., 2014; Feng et al., 2020). Ensemble techniques, in which a group of algorithms are exploited to combine all the possible predictions for the ultimate prediction used to address this shortage (Dietterich, 2000). By using ensemble models, the predictive performances were improved for yield prediction in Alfalfa (Feng et al., 2020), Nicosia wastewater treatment plant (Nourani et al., 2018), and plant lncRNAs (Simopoulos et al., 2018). Boosting, bagging, and stacking are three of the most commonly used ensemble models (Dietterich, 2000; Feng et al., 2020). The bagging method was first introduced by Breiman (1996) as a variance reduction approach for different algorithms such as decision trees or other algorithms that employed variable selection and fitting in a linear model (Galar et al., 2011). Boosting algorithms have been introduced by Schapire (1999) to serve as the alternative for the bagging method (Drucker and Cortes, 1996). Unlike bagging methods, which are parallel ensemble techniques, boosting methods are known as sequential ensemble techniques of base models by exploiting the dependencies of each algorithm (Dietterich, 2000; Feng et al., 2020). Many studies reported the successfulness of using bagging-RF and stochastic gradient boosting in predicting the yield of different crops (Pal, 2007; Gandhi et al., 2016; Aghighi et al., 2018; Zhang Z. et al., 2019). Bagging and boosting ensemble techniques commonly combine homogeneous algorithms for interpretation, while stacking methods tend to use heterogeneous algorithms and adjust the difference between them to increase precision (Dietterich, 2000; Zhou, 2012; Feng et al., 2020).

The successful use of machine learning algorithms for predicting the performance of different agronomic traits,

including yield, are reported in Alfalfa (Feng et al., 2020), *Senecio* species (Carvalho et al., 2013), grassland (Feilhauer et al., 2017; Rocha et al., 2018), and chrysanthemum (Hesami et al., 2019). However, the application of machine learning algorithms for predicting soybean yield from hyperspectral reflectance data is still unexplored and required comprehensive studies. Ensemble-based methods have successfully been applied to improve the prediction accuracies in artificial intelligence and computer vision (Ali et al., 2014; Feng et al., 2018, 2020; Ju et al., 2018) and, therefore, they may improve the accuracy of the yield prediction in this study. Thus, the main objectives of this study are: (1) to investigate the potential use of hyperspectral reflectance for predicting soybean yield, (2) to identify appropriate time-point of soybean growth stages for collecting hyperspectral reflectance to maximize yield prediction accuracy, and (3) to have a comparative study of individual and ensemble machine learning algorithms to improve the accuracy of predicting yield. The results of this study might help soybean breeders to increase the efficiency of selecting superior lines by estimating the final yield at early growth stages using spectral reflectance combined with machine learning approaches.

MATERIALS AND METHODS

Experimental Locations and Plant Materials

The research was conducted at the University of Guelph, Ridgetown Campus, in 2018 and 2019. A panel of 250 soybean genotypes was grown under field condition at two locations: Ridgetown (42°27′14.8″N 81°52′48.0″W, 200 m above sea level) and Palmyra (42°25′50.1″N 81°45′06.9″W, 195 m above sea level), in Ontario, Canada, during two consecutive growing seasons in 2018 and 2019 (**Figure 1**).

The soybean genotypes used in this study were the core germplasms of the soybean breeding program at the University of Guelph, Ridgetown, that have been collected in the past 35 years and used for genetic studies and cultivar developments. The experiments were conducted using randomized complete block designs (RCBD) with two replications in four environments (two locations × two years). Overall, there were 500 soybean plots per environment and 1000 soybean plots per year. In order to reduce the possible spatial variability in the field, each experiment was analyzed by nearest-neighbor analysis (NNA) as one of the error control strategies by using double covariate analysis (Stroup and Mulitze, 1991; Bowley, 1999; Katsileros et al., 2015). Each plot consisted of five rows, each 4.2 m long with a row spacing of 43 cm. The seeding rate was 57 seeds/m^2. At the end of the season, the three inside rows were machine harvested for estimating total seed yields (Ton ha^{-1}).

Phenotypic Evaluations

Yield Collection

Soybean seed yield (Ton ha^{-1}) was measured using three out of five harvested rows for each plot and adjusted to a 13% moisture level. The best linear unbiased prediction (BLUP) as a mixed model was used to calculate the average seed yield production for each soybean genotype across different environments (Goldberger, 1962).

Hyperspectral Reflectance Data Collection

In this study, the focus was on the spectral reflectance bands that are typically classified as the visible (VIS) and near-infrared (NIR) spectral components (Albetis et al., 2017). Canopy hyperspectral reflectance measurements were collected during the soybean growth and development stages at R4, where pods are 1.91 cm long at one of the four uppermost nodes, and R5, where seeds are 0.31 cm long in pods at one of the four uppermost nodes (Pedersen et al., 2004).

Each soybean genotype's hyperspectral reflectance properties were collected using a UniSpec-DC Spectral Analysis System (PP Systems International, Inc., 110 Haverhill Road, Suite 301 Amesbury, MA, United States). The machine covers 250 reflectance bands between 350 nm and 1,100 nm with a bandwidth of 3 nm. The field-of-view of the sensor was approximately 25° and covered a sample area of 0.25 m^2. Dark reference was used for calibrating the dual channels, and Spectralon panels were used to characterize incoming solar radiation. For each plot, three measurements were recorded, and their average, calculated by the BLUP model, was used as the reflectance band datapoint. All of the measurements were performed as close to solar noon as possible—the data for each stage were collected in 1 day, from 11:00 AM to 2:00 PM, to minimize the signal noise associated with the environment.

Data Pre-processing and Statistical Analyses

The existence of noise during hyperspectral reflectance measurement is inevitable, typically caused by sensors and electronic fluctuations (Ozaki et al., 2006). Therefore, it would be critical to have a pre-processing step for the collected data in order to increase the accuracy of the study. The hyperspectral data and yield of 250 soybean genotypes were pre-processed using the R software (version 3.6.1) to remove potential noises that randomly occur across the whole spectra resulting in misinterpretations. After checking the quality of reflectance bands and detecting outliers by using principal component analysis (PCA) for each genotype, 245 genotypes were selected for further analyses (Serneels and Verdonck, 2008). As a result of sensor-specific artifacts, reflectance bands at the two edges of the hyperspectral reflectance spectrum, 350–395 nm and 1,005–1,100 nm, were removed from the original data. The collected contiguous hyperspectral reflectance data was also reduced from 395 to 1005 nm with a 3 nm interval to a 10 nm interval leading to a total of 62 variables. Data scaling and centering were applied in order to improve reflectance properties in the pre-processing and the pre-treatment steps (Rossel, 2008). For each reflectance band variable, the Savitzky–Golay filter was applied for improving the signal-to-noise ratio (Savitzky and Golay, 1964).

As shown in **Figure 2**, the measured soybean yield was divided into four classes with equal numbers (∼) of data points in ascending order: Low (0–24.99% of total yield), medium-low (25–49.99% of total yield), medium-high (50–74.99% of total yield), and high (75–100% of total yield) yield. While

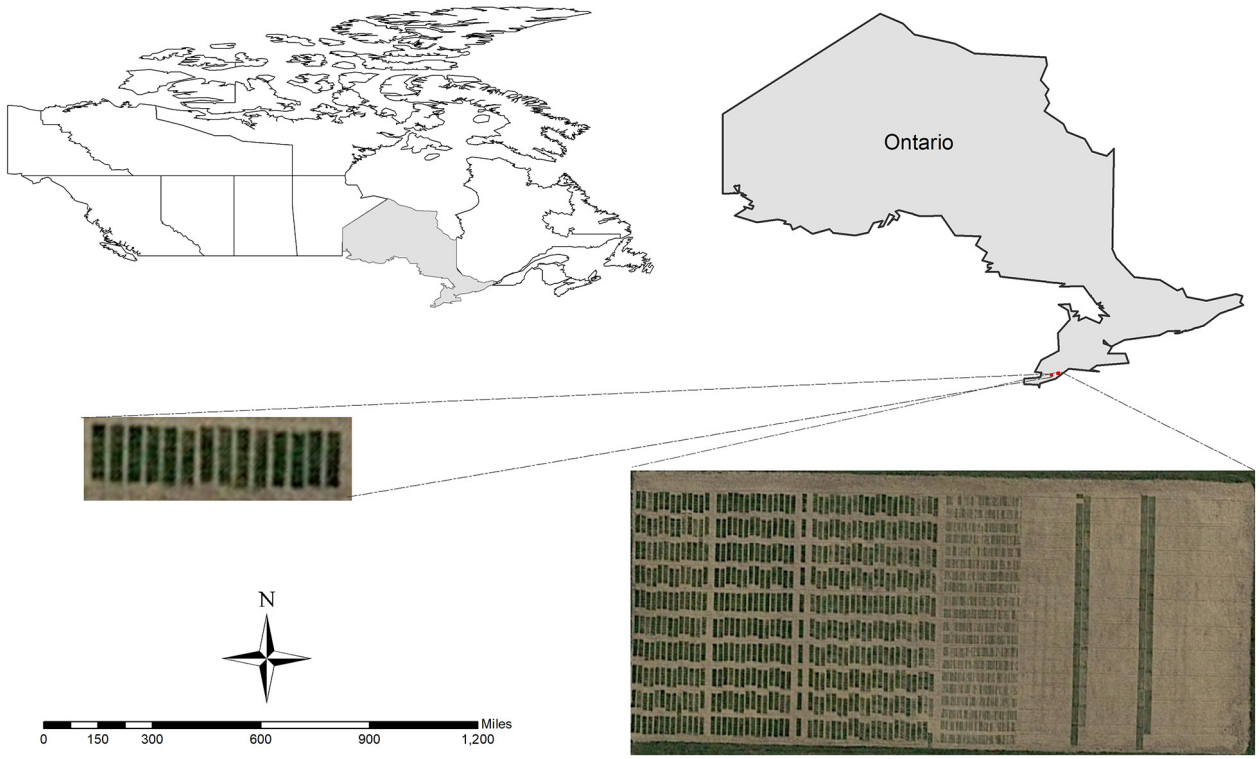

FIGURE 1 | The location of the experiments in 2018 and 2019.

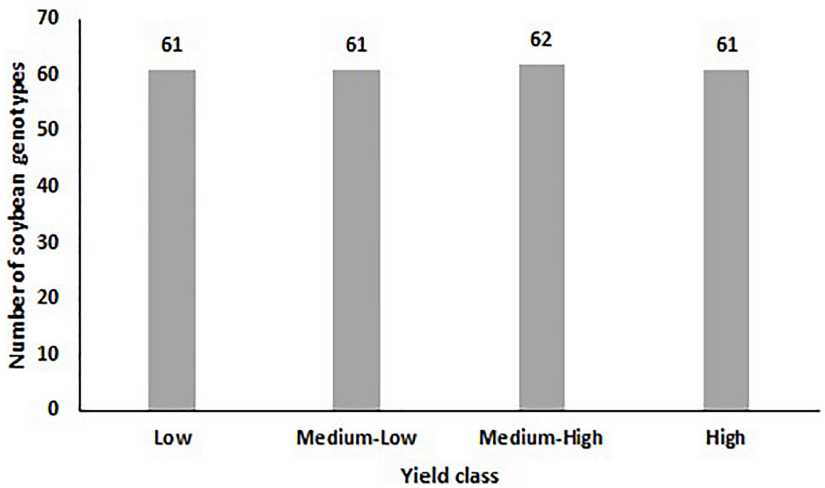

FIGURE 2 | The distribution of soybean genotypes in each yield class.

62 reflectance bands were considered as input variables, the classified yield was chosen as the output variable.

Variable Selection

Feature or variable selection is usually applied before developing the machine learning algorithms for reducing the data dimensionality, specifically in the small training datasets. One of the common approaches for variable selection is the recursive feature elimination (RFE) approach, which is easy to configure and effectively select the most relevant variables that predict the output (Chen and Jeong, 2007). Therefore, the RFE was run to indicate the initial variable importance scores and eliminate the reflectance band variables with the lowest importance score. Afterward, the process was recursively repeated until the ranking was indicated for all the reflectance bands. The package

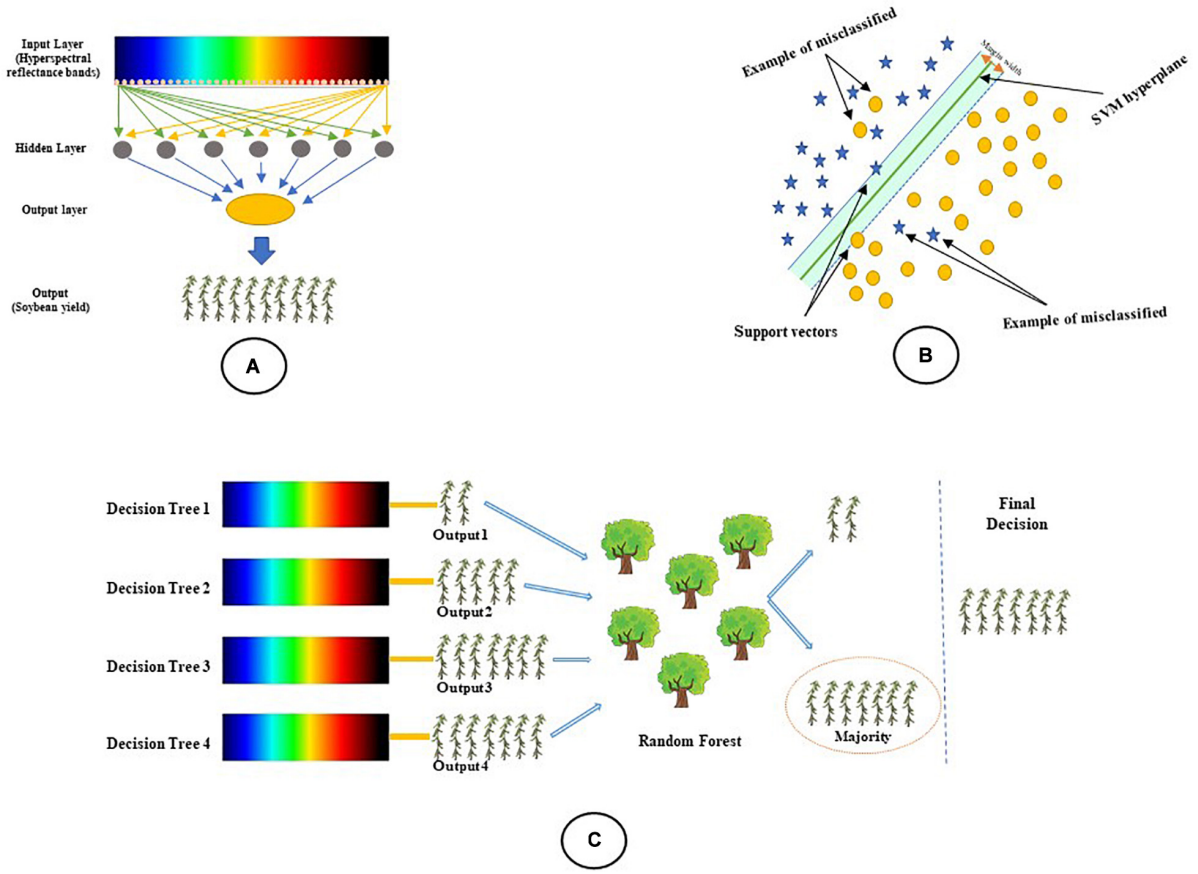

FIGURE 3 | A schematic representation of the machine learning algorithms used in this study to classify the soybean yield using reflectance bands: **(A)** Multilayer perceptron, **(B)** Support vector machine, and **(C)** Random forest.

caret (Kuhn, 2008) in R software version 3.6.1 was used for running RFE.

Data-Driven Modeling

Three of the most commonly used algorithms in the literature, multilayer perceptron (MLP), the support vector machine (SVM), and random forest (RF) (Filippi and Jensen, 2006; Chen et al., 2007; De Castro et al., 2012; Makantasis et al., 2015; Zhang N. et al., 2019; Šestak et al., 2019), were chosen and used for predicting the soybean yield. **Figure 3A** shows the MLP algorithm including an input layer, one or more hidden layers, and an output layer of completely interconnected neurons. Each neuron unit produces an output based on a sigmoid function derived from a linear combination of outputs from a previous layer (Wang et al., 2009). SVM (**Figure 3B**) is a set of related supervised learning methods that can recognize patterns used for classification analyses (Suykens and Vandewalle, 1999; Shao et al., 2012). The objective of SVM is to use hyperplanes for determining the optimal separation of yield classes. The random forest (RF) approach generates a series of trees representing a subset *n* of independent observations (**Figure 3C**). A detailed description of these machine learning algorithms can be found in Taillardat et al. (2016) and Meinshausen (2006). All of the relevant parameters in each machine learning algorithm were optimized based on the input variables.

We employed an ensemble method based on a stacking strategy (E–S) to improve the prediction performance. The results from individual algorithms were collected and combined together via the stacking procedure described in Dietterich (2000), where an algorithm with the highest accuracy performance was selected as the metaClassifier for this ensemble model. The Weka software version 3.9.4 (Hall et al., 2009) was used for running all machine learning algorithms and the ensemble method.

Quantification of Machine Learning Performance

In this study, the fivefold cross-validation strategy (Siegmann and Jarmer, 2015) with 10 repetitions was used to measure the classification quality of all the tested ML algorithms (**Figure 4**). In order to evaluate each algorithm, the values of precision (Eq. 1), recall (Eq. 2) as a measure of sensitivity, F-measure (Eq. 3), and Matthews correlation coefficient (MCC, Eq. 4) for validation

dataset were measured using the following formulas:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

where *TP* stands for the number of true positives, *TN* is the number of true negatives, *FP* stands for the number of false positives, and *FN* is the number of false negatives.

Visualizing and Analyzing

The Microsoft Excel software (2016), *ggplot2* (Wickham, 2011), and *ggvis* (Dennis, 2016) packages in the R software version 3.6.1 were used to conduct statistical analyses and visualize the results.

RESULTS

Yield Statistics and Spectral Profiles

In the current study, the average yield of 245 soybean genotypes, evaluated in four environments, ranged from 2.58 to 5.71 ton ha^{-1} with a mean and standard deviation of 4.22 and 0.57 ton ha^{-1}, respectively. The minimum, mean, and maximum values of each reflectance bands evaluated for all the genotypes across the four environments at the R4 and R5 growth stages are reported in **Figure 5**. At both growth stages, while the reflectance values showed small ranges of variation among the genotypes between 395 and 695 nm, the bands greater than 705 nm showed large variations within the population.

Variable Selection

The importance values of all 62 reflectance band variables for predicting yield were estimated using the RFE strategy for the R4 (**Table 1**) and R5 (**Table 2**) growth stages. For the R4 growth stage, the 1005 nm and the 605 nm bands had the highest and lowest importance values (%) for classifying the soybean yield, respectively. Based on RFE analysis, 56 of the reflectance bands were selected for training the algorithm, as shown in **Figure 6A**. At the R5 growth stage, the highest and lowest importance values (%) were found at 395 nm and the 725 nm bands, respectively. Out of 62 reflectance bands, 21 reflectance bands were selected to train the algorithms based on RFE strategy, which were considered selected variables (-VS) for further analyses. Among the 21 selected reflectance bands, three bands were in the violet, six in the blue, two in the green, eight in the red, and two were in the near-infrared (NIR) regions of the spectrum (**Figure 6B**). By using RFE for the R4 growth stage dataset, the top five high importance reflectance bands were located in the violet and NIR regions of the spectra. However, for R5, the violet and red regions had the top five high importance reflectance bands (**Tables 1**, **2**). The violet region, specifically the 395 nm band, had the highest importance values in both growth stages. The plotting of the soybean yield versus reflectance values at 395 nm (R5 growth stage) illustrated that the values for 395 nm in the high yielding class ranged from 0.009 to 0.016 which lower than values for the low yielding class, ranged from 0.020 to 0.029 (**Figure 7**). The difference between the reflectance values of high and low yielding classes was statistically significant at the significance level of 0.05 (data were not shown). Among all the tested bands, the 395 nm band measured at R5 was considered as the best solo reflectance band for discriminating soybeans for their yield potential.

Growth Stage Comparison

In order to investigate which of the growth stages is better for collecting reflectance data and predicting the soybean yield, the reflectance bands collected at each soybean growth stage were analyzed using the three machine learning algorithms. The average classification accuracy for validation dataset ranged between 12 and 43% using the R4 data and between 12 and 99% using the R5 data, which indicated that the R5 soybean growth stage is, in general, a better stage for collecting reflectance data if the goal is to predict the yield (**Figure 8**). Therefore, R5 was selected for further machine learning algorithm analyses. The results of individual and ensemble machine learning algorithms using R4 data are available in **Figure 8** and **Supplementary Figure 1**.

Comparative Analysis of the Developed Models

All three algorithms (RF, MLP, and SVM), as well as the E–S model, were trained using both full (62 bands) and selected (21 bands) variables at R5, and the summaries of the confusion matrices were presented in **Supplementary Table 1**. Regarding the comparative analyses of individual algorithms using all variables, RF, MLP, and SVM had the highest to lowest MCC values equal to 0.84, 0.76, and 0.66, respectively (**Figure 9A**). For the selected variables, the MCC values for RF and MLP declined to 0.80 and 0.73, respectively, while the value for SVM slightly increased to 0.73. The E–S method outperformed all the individual algorithms obtaining an MCC value of 0.93, using all variables, and 0.87, using selected variables (**Figure 9A**). In general, among all the individual tested algorithms, the RF algorithm had the highest performance with the values of 84 and 80% yield classification accuracy using all variables and selected variables, respectively.

When using full variables, the precision values for RF, MLP, and SVM were 0.91, 0.83, and 0.82, respectively. However, by using selected variables, the precision values for RF and MLP declined to 0.87 and 0.78, respectively, while the SVM performance was improved (0.87) when compared against all variables (**Figure 9B**). The E–S model had a precision of 0.96 using all variables and 0.90 using selected variables. Using all variables, the highest recall value was obtained for RF with a value of 0.84, followed by MLP and SVM with the values of 0.83 and 0.68, respectively.

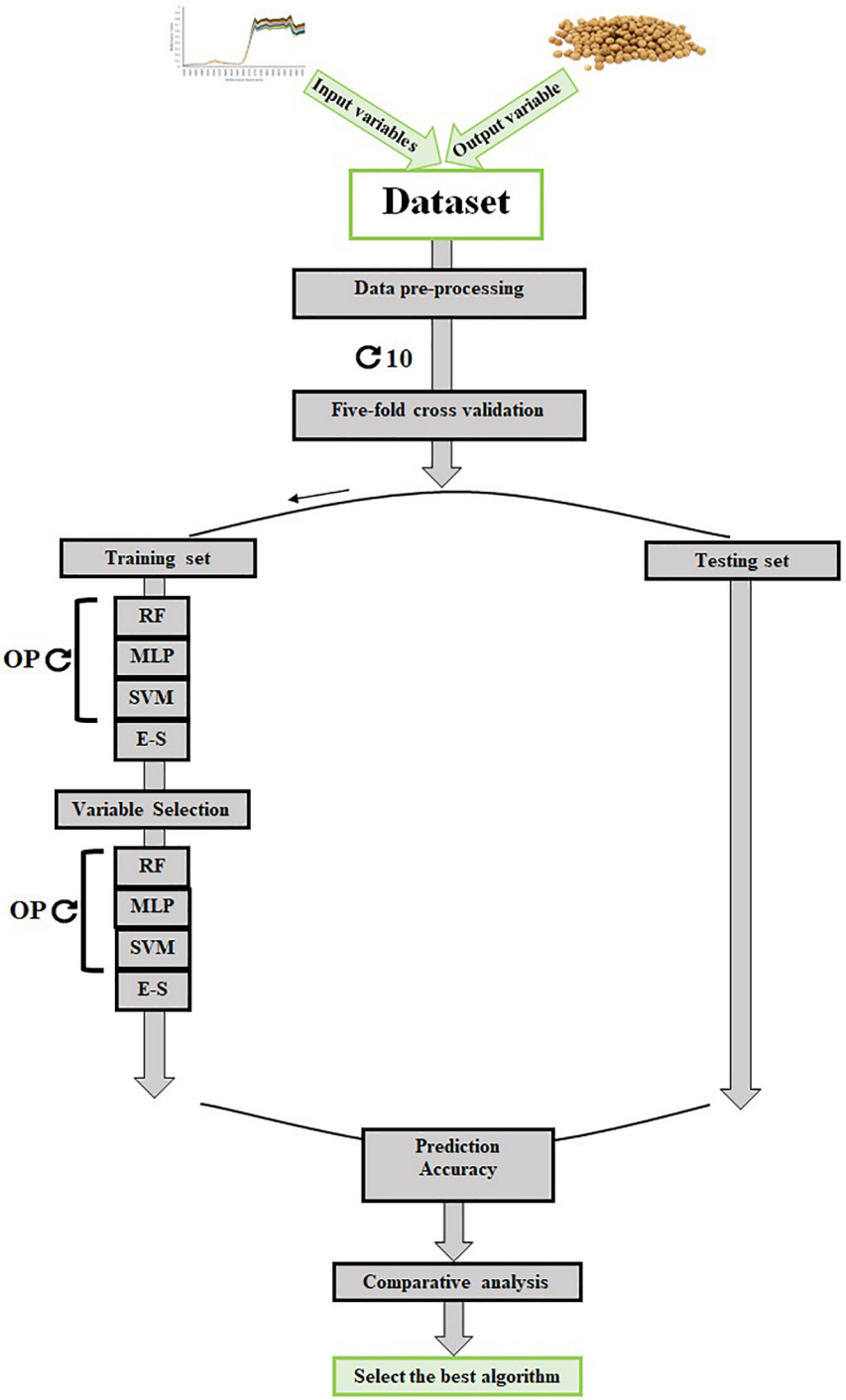

FIGURE 4 | The scheme of data collection and machine learning algorithm development and validation. OP, optimizing parameters; MLP, multilayer perceptron; SVM, support vector machine; RF, random forest; E–S, ensemble–stacking strategy.

However, the recall value of SVM increased to 0.72 using selected variables. The recall values of RF and MLP declined when selected variables were used (**Figure 9C**). E–S had the highest recall values, with 0.94 and 0.90 for full and selected variables, respectively.

To have a better interpretation of precision and recall values, the F-measure was evaluated for each and every algorithm. Using all variables, the F-measures of RF, MLP, and SVM were estimated to be 0.87, 0.81, and 0.71, respectively (**Figure 9D**). F-measure values were decreased for RF (0.84)

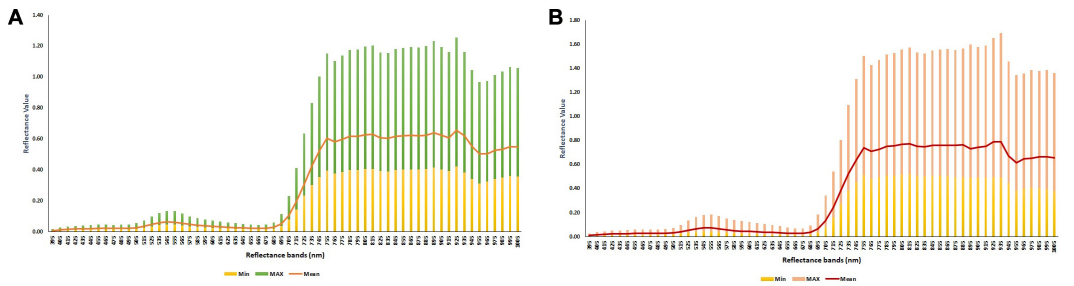

FIGURE 5 | The minimum, mean, and maximum values of each reflectance band were measured for 245 soybean genotypes evaluated at **(A)** R4 and **(B)** R5 growth stages at four different field environments.

TABLE 1 | Reflectance band ranking using the recursive feature elimination (RFE) strategy at R4 soybean growth stage.

Reflectance band (nm)	Ranking	Reflectance band (nm)	Ranking
1005	1	775	32
395	2	655	33
945	3	675	34
755	4	665	35
985	5	825	36
995	6	485	37
705	7	695	38
745	8	475	39
965	9	615	40
955	10	465	41
715	11	515	42
975	12	625	43
905	13	685	44
725	14	565	45
885	15	575	46
875	16	535	47
895	17	645	48
765	18	545	49
845	19	635	50
915	20	555	51
865	21	415	52
735	22	505	53
925	23	595	54
855	24	455	55
805	25	585	56
795	26	445	57
935	27	425	58
835	28	435	59
815	29	525	60
405	30	495	61
785	31	605	62

TABLE 2 | Reflectance band ranking using the recursive feature elimination (RFE) strategy at R5 soybean growth stage.

Reflectance band (nm)	Ranking	Reflectance band (nm)	Ranking
395	1	965	32
665	2	845	33
675	3	865	34
655	4	905	35
405	5	915	36
685	6	515	37
645	7	695	38
435	8	885	39
445	9	875	40
635	10	895	41
475	11	585	42
485	12	925	43
495	13	935	44
415	14	575	45
625	15	955	46
425	16	715	47
455	17	755	48
465	18	535	49
765	19	975	50
615	20	555	51
775	21	985	52
795	22	995	53
815	23	525	54
805	24	545	55
785	25	565	56
505	26	945	57
605	27	705	58
825	28	1005	59
855	29	745	60
595	30	735	61
835	31	725	62

and MLP (0.80) using selected variables. However, for SVM algorithm, the F-measure value was increased to 0.77 using selected variables. The E–S algorithm overperformed all the individual machine learning algorithms by having an F-measure value of 0.94, using all variables, and 0.90, using selected variables.

DISCUSSION

One of the objectives of this study was to find the best growth stage for collecting reflectance data suitable for predicting soybean yields. In this study, the hyperspectral reflectance data were collected at the reproductive stages of R4 and R5, in which

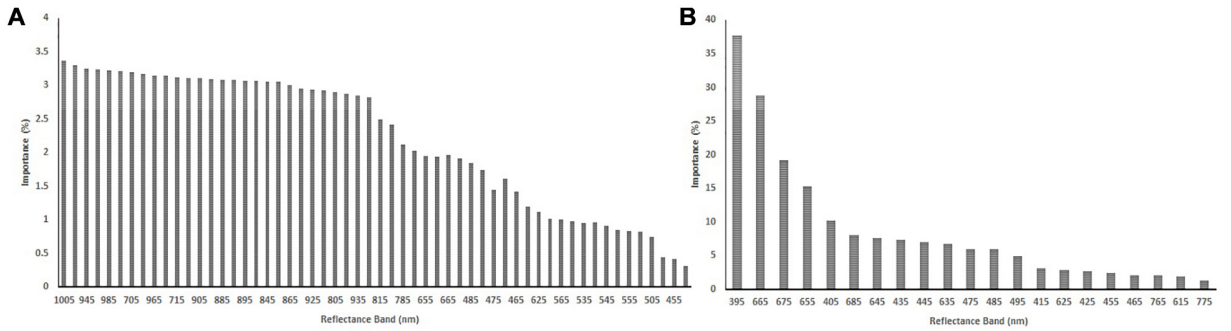

FIGURE 6 | The importance of selected variables based on the recursive feature elimination (RFE) strategy for soybean reflectance bands measured at R4 **(A)** and R5 **(B)** soybean growth stages.

pods and seeds are developed. R4 and R5 are known as critical growth and development stages in soybean since stresses can impose significant impacts on the yield at these stages and, therefore, soybean genotypes with different levels of tolerance to stresses can be discriminated from one another at these stages (Sweeney et al., 2003). For example, the results of a study by Eck (1987) showed that imposing soybeans to water deficit stress during the R1 to R3 growth stages reduce seed yields up to 9-13%. However, imposing the same soybeans to water deficit stress during R4 to R5 reduced seed yields up to 46%. Water deficit stress can less influence the total seed yield when occurring anytime beyond the R5 growth and development stage. Therefore, measuring hyperspectral reflectance at R4 and R5 would be more informative for predicting the soybean yield classes since the final yield production has been to some extent established at these two stages for all the genotypes. Our results indicated that R5 is a better stage to measure reflectance bands for predicting the yield. Ma et al. (1996) reported significant correlations between leaf photosynthetic rates and leaf greenness at R4 and R5, while this correlation was not significant at R6. In Soybean, the leaf photosynthetic rate can be changed significantly during the growth stages that, in turn, can empower different genotypes to recover the yield losses caused by temporary environmental stresses (Ferris et al., 1998; Siebers et al., 2015). Studies showed that environmental stresses at R5 can damage the soybean yield greater than that at R4 (Fehr et al., 1981) since the plants have less time to recover for yield before physiological maturity. It can be hypothesized that predicting yield of genotypes with different genetic potential by using reflectance bands that are measured at R5 is more reliable since the final yield productions have already been established, to some extent, for all the genotypes. The current study showed that the reflectance bands collected at R5 are more reliable and informative for predicting yield than the data collected at R4.

Several studies reported the strong correlation between reflectance bands and yield in different crop plants such as alfalfa (Kayad et al., 2016; Feng et al., 2020), wheat (Prey and Schmidhalter, 2019), maize (Lane et al., 2020), rice (Wan et al., 2020), and sugarcane (Verma et al., 2020). The visible reflectance bands can be splitted into three main regions, red (650–700 nm), green (495–570 nm), and violet–blue (390–495 nm) (Hennessy et al., 2020). Most studies were emphasized the importance of red spectral bands or the combined use of red and red edge bands as one solid index in predicting the total yield (Jolly et al., 2005; Filippa et al., 2018; Lykhovyd, 2020; Phan et al., 2020; Tiwari and Shukla, 2020). In this study, we identified highly ranked bands in the violet and red regions for classifying the soybean seed yield, centered at 395 nm, 665 nm, and 675 nm (**Table 2**). The violet and red spectral regions can be associated with the absorption of plant pigments such as carotenoid, anthocyanins, and chlorophyll (Merzlyak et al., 2003; Richter et al., 2016; Hennessy et al., 2020). Carotenoid plays a pivotal role in discrimination of senescent leaves (Richter et al., 2016; Hennessy et al., 2020). The importance of 395 nm band in soybean yield prediction at R5 growth stage can refer to the fact that soybean at R5 growth stage initiates the senescence and decay of chlorophyll resulting in better discrimination of the genotypes with different photosystems functioning and photoprotection capabilities. However, there is no report on the solid effect of the 395 nm reflectance band in the physiological process of soybean or any other plants.

In order to have accurate yield prediction and avoid model overfitting, machine learning algorithms may benefit from using a variable selection process to reduce the dimensionality of the data to an appropriate level (Hennessy et al., 2020). Existing variable selection methods can be categorized based on their applications, complexities, and accuracy (Zheng et al., 2020). One of the most commonly used variable selection methods is the RFE approach that provides an acceptable performance with moderate computational exertions (Guyon et al., 2002; Granitto et al., 2006). The successful use of RFE to reduce the number of input variables has been reported in many studies (Granitto et al., 2006; Chen and Jeong, 2007; Feng et al., 2020). The efficiency of using selected variables for predicting classified soybean yield over full reflectance band variables was evaluated using the RFE method. Using RFE method might decrease the value of the parameters such as precision, recall, MCC, and F-measure to avoid overfitting (Loughrey and Cunningham, 2004). This is a small price to pay, especially if the decrease in performance is not significant. Among all the tested individual machine learning algorithms, RF had the highest performance using either full or

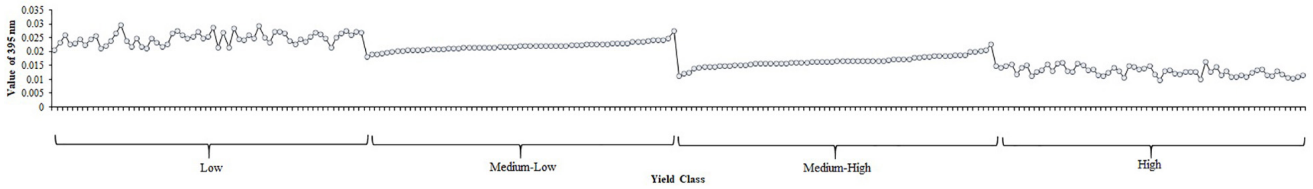

FIGURE 7 | The soybean yield classes versus the 395 nm reflectance band at R5 growth stages.

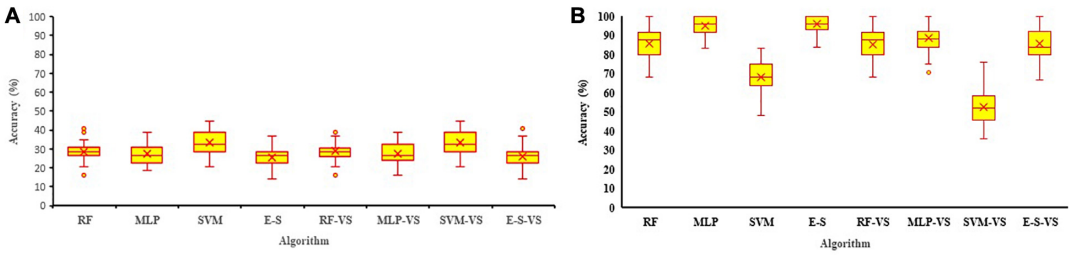

FIGURE 8 | The accuracy of RF, MLP, SVM, and E–S algorithms for predicting soybean yield using full and RFE selected variables (-VS) measured at R4 **(A)** and R5 **(B)** soybean growth stages in four environments. The mean performance was shown as × in each figure. MLP, multilayer perceptron; SVM, support vector machine; RF, random forest; E–S, ensemble–stacking strategy; RFE, recursive feature elimination.

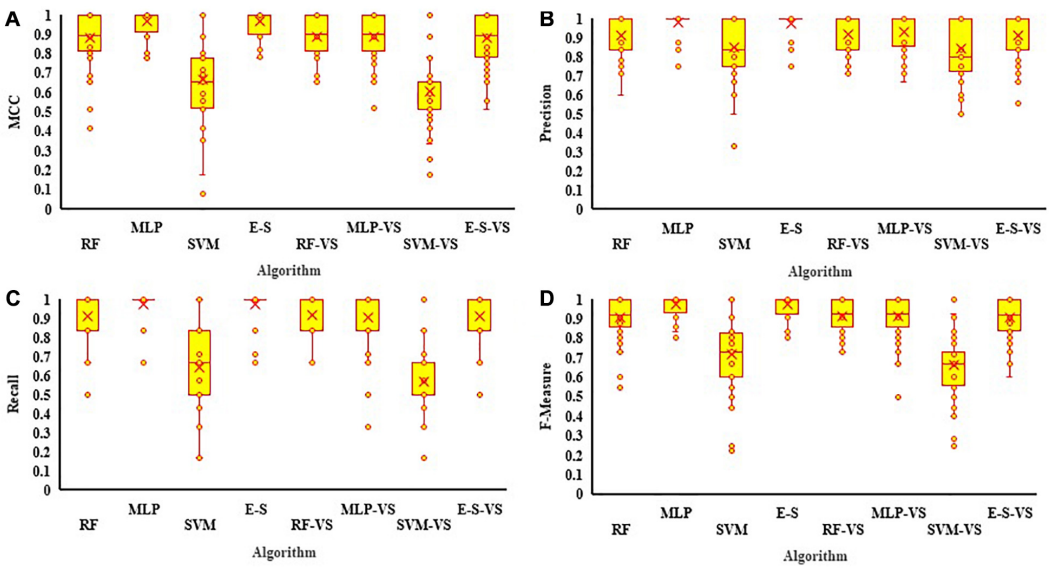

FIGURE 9 | The estimate values of **(A)** Matthews correlation coefficient (MCC), **(B)** Precision, **(C)** Recall, and **(D)** F-Measure for RF, MLP, SVM, and the E–S algorithms used for predicting soybean yield from all and selected variables collected (-VS) at the R5 growth stage. The mean performance is indicated with × in each figure. MLP, multilayer perceptron; SVM, support vector machine; RF, random forest; E–S, ensemble–stacking strategy; RFE, recursive feature elimination.

selected variables. This high performance may come from the nature of the RF algorithm, in which trees are trained using multiple random subsamples of the original dataset. This feature gives RF this ability to generate better and more stable predictions for new instances not necessarily included in the training dataset (Liaw and Wiener, 2002).

Multilayer perceptron was another machine learning algorithm that was exploited in this study. MLP was previously applied in different areas such as weed science (Tamouridou et al., 2017) or drought tolerance (Etminan et al., 2019), but not in soybean for yield prediction. This study found MLP to be the second-best machine learning algorithm for predicting the soybean yield. Previous studies reported a high likelihood of overfitting for neural network algorithms (Lawrence and Giles, 2000; Murakoshi, 2005). For MLP, common parameters such as the number of hidden layers, the

number of neurons in each layer, or training time can be used to control overfitting; however, the degree of overfitting would vary throughout the input variables (Lawrence and Giles, 2000).

Support vector machine is also one of the most common machine learning algorithms that have been broadly used in different areas such as plant tissue culture (Hesami and Jones, 2020), image classification (Lin et al., 2011), genes classification (Duan et al., 2005), and drug disambiguation (Björne et al., 2013). SVM is usually used when scientists have to deal with large numbers of features and high sparsity (Nguyen and De la Torre, 2010). Although the prediction accuracy of the SVM algorithm was lower than the values for RF and MLP in this study, its performance was slightly increased when the selected variables were used. An increase in SVM performance using selected variables was also reported in previous studies (Su and Yang, 2008; Tan et al., 2010; Alirezanejad et al., 2020). It might be due to this fact that selecting relevant variables can improve the performance of SVM through ameliorating its feature interpretability, computational efficiency, and generalization performance (Nguyen and De la Torre, 2010; Roy et al., 2015).

In order to see we can improve the prediction accuracy in this study through the combined use of the machine learning algorithms, RF, MLP, and SVM were used in constructing E–S, and RF was chosen as the metaClassifier for this ensemble algorithm. By using the E–S approach, we improved the prediction accuracy using either full or selected variables. A successful use of the E–S method has recently been reported for predicting the yield in alfalfa (Feng et al., 2020). When using the E–S approach, it is necessary to include self-sufficient, independent, and diverse ML algorithms in the analyses (Araya et al., 2017; Feng et al., 2020), which have a minimum dependency from one another and sufficient powers to predict the dependent variable, soybean yield classes in this study (Araya et al., 2017; Feng et al., 2020). The above criteria are important to be considered when individual ML algorithms are selected to combine in a given E–S analysis. In this study, RF, MLP, and SVM are selected as individual algorithms to be used in the E–S analyses because of their independent prediction methods as well as having different training approaches. By using the E–S approach, the prediction accuracy increased to 0.93, using all variables, and to 0.87, using selected variables, showing the success of using E–S as one of the ensemble techniques.

CONCLUSION

Pre-harvest soybean yield classifications and estimations are important for grain policy-making and food security across worldwide. Spectral reflectance is considered as an efficient phenotyping tool that can help breeders to make their selections at lower cost at a fast pace. The objectives of this study were to demonstrate the best soybean growth stage for measuring hyperspectral reflectance and evaluating the three most commonly used ML algorithms along with introducing the E–S method in predicting the soybean yield using reflectance variables. Soybean R5 growth stage was identified as the better stage than R4 for measuring hyperspectral reflectance. In addition to using 62 reflectance bands as the full variables, the RFE method was used to reduce the dimensionality of the data, and therefore, 21 most important bands were selected as the selected reflectance variables. Using both full and selected reflectance variables, RF overperformed all individual algorithms. Therefore, RF was selected as the metaClassifier for E–S. E–S had the highest prediction accuracy as one of the ensembles combined approaches compared to an individual ML algorithm. Therefore, E–S was recommended as a reliable and appropriate ML algorithm for predicting the soybean yield using reflectance variables. This study provides an applicable pipeline for using hyperspectral reflectance data and suitable ML algorithms for the development of high yielding soybeans, which can be used in large soybean breeding programs for selecting high-yielding soybeans at pre-harvesting stages. The developed methodology in this study can open a reliable and new window in using spectral reflectance for selecting high yielding genotypes in different crops.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MY-N performed the experiments, modeled, summed up, and wrote the manuscript. HE, DT, and JS revised the manuscript. ME designed and lead the experiments and revised the manuscript. All authors have read and approved the final manuscript.

FUNDING

This project was funded in part by the Grain Farmers of Ontario (GFO) and SeCan. The funding bodies did not play any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

ACKNOWLEDGMENTS

The authors would like to acknowledge the technical assistance of Dr. Sepideh Torabi, Mr. Bryan Stirling, Mr. John Kobler, Mr. Robert Brandt, and all the soybean breeding crew at the University of Guelph, Ridgetown Campus.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.624273/full#supplementary-material

REFERENCES

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H. S., and Radiom, S. (2018). Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 OLI. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* 11, 4563–4577. doi: 10.1109/JSTARS.2018.2823361

Albetis, J., Duthoit, S., Guttler, F., Jacquin, A., Goulard, M., Poilvé, H., et al. (2017). Detection of flavescence dorée grapevine disease using unmanned aerial vehicle (UAV) multispectral imagery. *Remote Sens.* 9:308. doi: 10.3390/rs9040308

Alexandratos, N., and Bruinsma, J. (2012). *World Agriculture Towards 2030/2050: the 2012 Revision*. Rome: Food and Agriculture Organization of the United Nations, Agricultural Development Economics Division (ESA).

Ali, I., Cawkwell, F., Green, S., and Dwyer, N. (2014). "Application of statistical and machine learning models for grassland yield estimation based on a hypertemporal satellite remote sensing time series," in *Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium*, (IEEE), 5060–5063. doi: 10.1109/IGARSS.2014.6947634

Alirezanejad, M., Enayatifar, R., Motameni, H., and Nematzadeh, H. (2020). Heuristic filter feature selection methods for medical datasets. *Genomics* 112, 1173–1181. doi: 10.1016/j.ygeno.2019.07.002

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Araya, D. B., Grolinger, K., ElYamany, H. F., Capretz, M. A., and Bitsuamlak, G. (2017). An ensemble learning framework for anomaly detection in building energy consumption. *Energy Build.* 144, 191–206. doi: 10.1016/j.enbuild.2017.02.058

Auria, L., and Moro, R. A. (2008). *Support Vector Machines (SVM) as a Technique for Solvency Analysis*. Berlin: Deutsches Institut für Wirtschaftsforschung (DIW). DIW Discussion Papers, No. 811 doi: 10.2139/ssrn.1424949

Belayneh, A., Adamowski, J., Khalil, B., and Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J. Hydrol.* 508, 418–429. doi: 10.1016/j.jhydrol.2013.10.052

Björne, J., Kaewphan, S., and Salakoski, T. (2013). "UTurku: drug named entity recognition and drug-drug interaction extraction using SVM classification and domain knowledge," in *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (* SEM),: Proceedings of the Seventh International Workshop on Semantic Evaluation*, (SemEval), 651–659.

Bowley, S. (1999). *A hitchhiker's guide to statistics in plant biology*. Guelph, Ont: Any Old Subject Books.

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Cai, G., Yang, Q., Chen, H., Yang, Q., Zhang, C., Fan, C., et al. (2016). Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. *Sci. Rep.* 6:21625. doi: 10.1038/srep21625

Carvalho, S., Macel, M., Schlerf, M., Moghaddam, F. E., Mulder, P. P., Skidmore, A. K., et al. (2013). Changes in plant defense chemistry (pyrrolizidine alkaloids) revealed through high-resolution spectroscopy. *ISPRS J. Photogrammetry Remote Sens.* 80, 51–60. doi: 10.1016/j.isprsjprs.2013.03.004

Chen, J.-C., and Wang, Y.-M. (2020). Comparing activation functions in modeling shoreline variation using multilayer perceptron neural network. *Water* 12:1281. doi: 10.3390/w12051281

Chen, L., Huang, J., Wang, F., and Tang, Y. (2007). Comparison between back propagation neural network and regression models for the estimation of pigment content in rice leaves and panicles using hyperspectral data. *Int. J. Remote Sens.* 28, 3457–3478. doi: 10.1080/01431160601024242

Chen, X.-W., and Jeong, J. C. (2007). "Enhanced recursive feature elimination," in *Proceedings of the Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, (IEEE), 429–435. doi: 10.1109/ICMLA.2007.35

Collins, N. C., Tardieu, F., and Tuberosa, R. (2008). Quantitative trait loci and crop performance under abiotic stress: where do we stand? *Plant Physiol.* 147, 469–486. doi: 10.1104/pp.108.118117

De Castro, A.-I., Jurado-Expósito, M., Gómez-Casero, M.-T., and López-Granados, F. (2012). Applying neural networks to hyperspectral and multispectral field data for discrimination of cruciferous weeds in winter crops. *Sci. World J.* 2012:630390. doi: 10.1100/2012/630390

Dennis, T. (2016). "Using R and ggvis to create interactive graphics for exploratory data analysis," in *Data Visualization: a Guide to Visual Storytelling for Libraries*, ed. L. Magnuson. (Lanham, MD: Rowman & Littlefield).

Dietterich, T. G. (2000). "Ensemble methods in machine learning," in *Proceedings of the International Workshop on Multiple Classifier Systems*, (Springer), 1–15. doi: 10.1007/3-540-45014-9_1

Drucker, H., and Cortes, C. (1996). Boosting decision trees. *Adv. Neural Inform. Process. Systems* 8, 479–485.

Duan, K.-B., Rajapakse, J. C., Wang, H., and Azuaje, F. (2005). Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans. Nanobiosci.* 4, 228–234. doi: 10.1109/TNB.2005.853657

Dubey, A., Kumar, A., Abd_Allah, E. F., Hashem, A., and Khan, M. L. (2019). Growing more with less: breeding and developing drought resilient soybean to improve food security. *Ecol. Indicators* 105, 425–437. doi: 10.1016/j.ecolind.2018.03.003

Eck, H. V. (1987). Plant water stress at various growth stages and growth and yield of soybeans. *Field Crops Res.* 17, 1–16. doi: 10.1016/0378-4290(87)90077-3

Etminan, A., Pour-Aboughadareh, A., Mohammadi, R., Shooshtari, L., Yousefiazarkhanian, M., and Moradkhani, H. (2019). Determining the best drought tolerance indices using artificial neural network (ANN): insight into application of intelligent agriculture in agronomy and plant breeding. *Cereal Res. Commun.* 47, 170–181. doi: 10.1556/0806.46.2018.057

Falk, K. G., Jubery, T. Z., Mirnezami, S. V., Parmley, K. A., Sarkar, S., Singh, A., et al. (2020). Computer vision and machine learning enabled soybean root phenotyping pipeline. *Plant Methods* 16:5. doi: 10.1186/s13007-019-0550-555

Fehr, W., Lawrence, B., and Thompson, T. (1981). Critical stages of development for defoliation of soybean 1. *Crop Sci.* 21, 259–262. doi: 10.2135/cropsci1981.0011183X002100020014x

Feilhauer, H., Somers, B., and van der Linden, S. (2017). Optical trait indicators for remote sensing of plant species composition: predictive power and seasonal variability. *Ecol. Indicators* 73, 825–833. doi: 10.1016/j.ecolind.2016.11.003

Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., et al. (2020). Alfalfa yield prediction using UAV-Based hyperspectral imagery and ensemble learning. *Remote Sens.* 12:2028. doi: 10.3390/rs12122028

Feng, P., Ma, J., Sun, C., Xu, X., and Ma, Y. (2018). A novel dynamic android malware detection system with ensemble learning. *IEEE Access* 6, 30996–31011. doi: 10.1109/ACCESS.2018.2844349

Ferris, R., Wheeler, T., Hadley, P., and Ellis, R. (1998). Recovery of photosynthesis after environmental stress in soybean grown under elevated CO2. *Crop Sci.* 38, 948–955. doi: 10.2135/cropsci1998.0011183X003800040012x

Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Sonnentag, O., Humphreys, E., et al. (2018). NDVI derived from near-infrared-enabled digital cameras: applicability across different plant functional types. *Agricult. Forest Meteorol.* 249, 275–285. doi: 10.1016/j.agrformet.2017.11.003

Filippi, A. M., and Jensen, J. R. (2006). Fuzzy learning vector quantization for hyperspectral coastal vegetation classification. *Remote Sens. Environ.* 100, 512–530. doi: 10.1016/j.rse.2005.11.007

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342. doi: 10.1038/nature10452

Furbank, R. T., and Tester, M. (2011). Phenomics–technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., and Herrera, F. (2011). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Systems Man Cybernet. Part C (Applications and Reviews)* 42, 463–484. doi: 10.1109/TSMCC.2011.2161285

Gandhi, N., Armstrong, L. J., Petkar, O., and Tripathy, A. K. (2016). "Rice crop yield prediction in India using support vector machines," in *Proceedings of the 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, (IEEE), 1–5. doi: 10.1109/JCSSE.2016.7748856

Geetha, M. (2020). Forecasting the crop yield production in trichy district using fuzzy C-Means algorithm and multilayer perceptron (MLP). *Int. J. Knowledge Systems Sci. (IJKSS)* 11, 83–98. doi: 10.4018/IJKSS.2020070105

Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., and Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environ. Earth Sci.* 75:476. doi: 10.1007/s12665-015-5096-x

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Am. Statist. Assoc.* 57, 369–375. doi: 10.1080/01621459.1962.10480665

Granitto, P. M., Furlanello, C., Biasioli, F., and Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometr. Intell. Lab. Systems* 83, 83–90. doi: 10.1016/j.chemolab.2006.01.007

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi: 10.1023/A:1012487302797

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11, 10–18. doi: 10.1145/1656274.1656278

Hartman, G. L., West, E. D., and Herman, T. K. (2011). Crops that feed the World 2. soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security* 3, 5–17. doi: 10.1007/s12571-010-0108-x

Hennessy, A., Clarke, K., and Lewis, M. (2020). Hyperspectral classification of plants: a review of waveband selection generalisability. *Remote Sens.* 12:113. doi: 10.3390/rs12010113

Hesami, M., Condori-Apfata, J. A., Valderrama Valencia, M., and Mohammadi, M. (2020a). Application of artificial neural network for modeling and studying in vitro genotype-independent shoot regeneration in wheat. *Appl. Sci.* 10:5370. doi: 10.3390/app10155370

Hesami, M., Naderi, R., Tohidfar, M., and Yoosefzadeh-Najafabadi, M. (2020b). Development of support vector machine-based model and comparative analysis with artificial neural network for modeling the plant tissue culture procedures: effect of plant growth regulators on somatic embryogenesis of chrysanthemum, as a case study. *Plant Methods* 16, 1–15. doi: 10.1186/s13007-020-00655-9

Hesami, M., and Jones, A. M. P. (2020). Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. *Appl. Microbiol. Biotechnol.* 104, 9449–9485. doi: 10.1007/s00253-020-10888-10882

Hesami, M., Naderi, R., Tohidfar, M., and Yoosefzadeh-Najafabadi, M. (2019). Application of adaptive neuro-fuzzy inference system-non-dominated sorting genetic algorithm-II (ANFIS-NSGAII) for modeling and optimizing somatic embryogenesis of chrysanthemum. *Front. Plant Sci.* 10:869. doi: 10.3389/fpls.2019.00869

Jin, J., Liu, X., Wang, G., Mi, L., Shen, Z., Chen, X., et al. (2010). Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in Northeast China. *Field Crops Res.* 115, 116–123. doi: 10.1016/j.fcr.2009.10.016

Jolly, W. M., Nemani, R., and Running, S. W. (2005). A generalized, bioclimatic index to predict foliar phenology in response to climate. *Global Change Biol.* 11, 619–632. doi: 10.1111/j.1365-2486.2005.00930.x

Ju, C., Bibaut, A., and van der Laan, M. (2018). The relative performance of ensemble methods with deep convolutional neural networks for image classification. *J. Appl. Statist.* 45, 2800–2818. doi: 10.1080/02664763.2018.1441383

Katsileros, A., Drosou, K., and Koukouvinos, C. (2015). Evaluation of nearest neighbor methods in wheat genotype experiments. *Commun. Biometry Crop Sci.* 10, 115–123.

Kayad, A. G., Al-Gaadi, K. A., Tola, E., Madugundu, R., Zeyada, A. M., and Kalaitzidis, C. (2016). Assessing the spatial variability of alfalfa yield using satellite imagery and ground-based data. *PLoS One* 11:e0157166. doi: 10.1371/journal.pone.0157166

Khan, R., Hanbury, A., and Stoettinger, J. (2010). "Skin detection: a random forest approach," in *Proceedings of the 2010 IEEE International Conference on Image Processing*, (IEEE), 4613–4616. doi: 10.1109/ICIP.2010.5651638

Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Statist. Software* 28, 1–26. doi: 10.18637/jss.v028.i05

Kycko, M., Zagajewski, B., Lavender, S., Romanowska, E., and Zwijacz-Kozica, M. (2018). The impact of tourist traffic on the condition and cell structures of alpine swards. *Remote Sens.* 10:220. doi: 10.3390/rs10020220

Lane, H. M., Murray, S. C., Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Rooney, D. K., et al. (2020). Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *Plant Phenome J.* 3:e20002. doi: 10.1002/ppj2.20002

Lawrence, S., and Giles, C. L. (2000). "Overfitting and neural networks: conjugate gradient and backpropagation," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, (IEEE), 114–119. doi: 10.1109/IJCNN.2000.857823

Lepetit, V., Lagger, P., and Fua, P. (2005). "Randomized trees for real-time keypoint recognition," in *Proceedings of the 2005 IEEE*, (IEEE), 775–781. doi: 10.1109/CVPR.2005.288

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R news* 2, 18–22.

Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., et al. (2011). "Large-scale image classification: fast feature extraction and SVM training," in *Proceedings of the CVPR*, (IEEE), 1689–1696. doi: 10.1109/CVPR.2011.5995477

Lopez-Cruz, M., Olson, E., Rovere, G., Crossa, J., Dreisigacker, S., Mondal, S., et al. (2020). Regularized selection indices for breeding value prediction using hyper-spectral image data. *Sci. Rep.* 10, 1–12. doi: 10.1038/s41598-020-65011-2

Loughrey, J., and Cunningham, P. (2004). "Overfitting in wrapper-based feature subset selection: the harder you try the worse it gets," in *Proceedings of the International Conference on Innovative Techniques and Applications of Artificial Intelligence*, (Springer), 33–43. doi: 10.1007/1-84628-102-4_3

Lykhovyd, P. (2020). Sweet corn yield simulation using normalized difference vegetation index and leaf area index. *J. Ecol. Eng.* 21, 228–236 doi: 10.12911/22998993/118274

Ma, B., Dwyer, L. M., Costa, C., Cober, E. R., and Morrison, M. J. (2001). Early prediction of soybean yield from canopy reflectance measurements. *Agronomy J.* 93, 1227–1234. doi: 10.2134/agronj2001.1227

Ma, B., Morrison, M. J., and Dwyer, L. M. (1996). Canopy light reflectance and field greenness to assess nitrogen fertilization and yield of maize. *Agronomy J.* 88, 915–920. doi: 10.2134/agronj1996.00021962003600060011x

Makantasis, K., Karantzalos, K., Doulamis, A., and Doulamis, N. (2015). "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, (IEEE), 4959–4962. doi: 10.1109/IGARSS.2015.7326945

Meinshausen, N. (2006). Quantile regression forests. *J. Mach. Learn. Res.* 7, 983–999.

Merzlyak, M. N., Solovchenko, A. E., and Gitelson, A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biol. Technol.* 27, 197–211. doi: 10.1016/S0925-5214(02)00066-2

Mokry, F. B., Higa, R. H., de Alvarenga Mudadu, M., de Lima, A. O., Meirelles, S. L. C., da Silva, M. V. G. B., et al. (2013). Genome-wide association study for backfat thickness in Canchim beef cattle using Random Forest approach. *BMC Genet.* 14:47. doi: 10.1186/1471-2156-14-47

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., los Campos, G., Alvarado, G., Suchismita, M., et al. (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13:4. doi: 10.1186/s13007-016-0154-2

Murakoshi, K. (2005). Avoiding overfitting in multilayer perceptrons with feeling-of-knowing using self-organizing maps. *BioSystems* 80, 37–40. doi: 10.1016/j.biosystems.2004.09.031

Nguyen, M. H., and De la Torre, F. (2010). Optimal feature selection for support vector machines. *Pattern Recogn.* 43, 584–591. doi: 10.1016/j.patcog.2009.09.003

Nourani, V., Elkiran, G., and Abba, S. (2018). Wastewater treatment plant performance analysis using artificial intelligence–an ensemble approach. *Water Sci. Technol.* 78, 2064–2076. doi: 10.2166/wst.2018.477

Ozaki, Y., McClure, W. F., and Christy, A. A. (2006). *Near-infrared Spectroscopy in Food Science and Technology*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/0470047704

Pal, M. (2007). Ensemble learning with decision tree for remote sensing classification. *World Academy Sci. Eng. Technol.* 36, 258–260.

Pal, S. K., and Mitra, S. (1992). Multilayer perceptron, fuzzy sets, classifiaction. *IEEE Trans. Neural Networks* 3, 683–697. doi: 10.1109/72.159058

Pedersen, P., Kumudini, S., Board, J., and Conley, S. (2004). *Soybean Growth and Development*. University Extension Ames, IA: Iowa State University.

Phan, P., Chen, N., Xu, L., and Chen, Z. (2020). Using multi-temporal MODIS NDVI data to monitor tea status and forecast yield: a case study at tanuyen, laichau, vietnam. *Remote Sens.* 12:1814. doi: 10.3390/rs12111814

Prey, L., and Schmidhalter, U. (2019). Simulation of satellite reflectance data using high-frequency ground based hyperspectral canopy measurements for in-season estimation of grain yield and grain nitrogen status in winter wheat. *ISPRS J. Photogrammetry Remote Sens.* 149, 176–187. doi: 10.1016/j.isprsjprs.2019.01.023

Richter, R., Reu, B., Wirth, C., Doktor, D., and Vohland, M. (2016). The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area. *Int. J. Appl. Earth Observation Geoinform.* 52, 464–474. doi: 10.1016/j.jag.2016.07.018

Rocha, A., Groen, T., Skidmore, A., Darvishzadeh, R., and Willemen, L. (2018). Machine learning using hyperspectral data inaccurately predicts plant traits under spatial dependency. *Remote Sens.* 10:1263. doi: 10.3390/rs10081263

Rossel, R. A. V. (2008). ParLeS: software for chemometric analysis of spectroscopic data. *Chemometrics Intell. Lab. Systems* 90, 72–83. doi: 10.1016/j.chemolab.2007.06.006

Roy, K., Kar, S., and Das, R. N. (2015). "Statistical methods in QSAR/QSPR," in *A Primer on QSAR/QSPR Modeling*, eds K. Roy, S. Kar, and R. N. Das (Berlin: Springer), 37–59. doi: 10.1007/978-3-319-17281-1_2

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Párez, L. G., Crossa, J., et al. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 (Bethesda)* 6, 2799–2808. doi: 10.1534/g3.116.032888

Savitzky, A., and Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639. doi: 10.1021/ac60214a047

Schapire, R. E. (1999). "A brief introduction to boosting," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, (Florham Park, NJ), 1401–1406.

Schweiger, A. K., Cavender-Bares, J., Townsend, P. A., Hobbie, S. E., Madritch, M. D., Wang, R., et al. (2018). Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nat. Ecol. Evol.* 2, 976–982. doi: 10.1038/s41559-018-0551-1

Serneels, S., and Verdonck, T. (2008). Principal component analysis for data containing outliers and missing elements. *Comput. Statist. Data Anal.* 52, 1712–1727. doi: 10.1016/j.csda.2007.05.024

Šestak, I., Zgorelec, Ž, Perèin, A., Mesiæ, M., and Galiæ, M. (2019). "Prediction of soybean leaf nitrogen content using proximal field spectroscopy," in *Proceedings of the 54th Croatian & 14th International Symposium on Agriculture*, (Vodice).

Shao, Y., Zhao, C., Bao, Y., and He, Y. (2012). Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy. *Food Bioprocess Technol.* 5, 100–107. doi: 10.1007/s11947-009-0267-y

Siebers, M. H., Yendrek, C. R., Drag, D., Locke, A. M., Rios Acosta, L., Leakey, A. D., et al. (2015). Heat waves imposed during early pod development in soybean (G lycine max) cause significant yield loss despite a rapid recovery from oxidative stress. *Global Change Biol.* 21, 3114–3125. doi: 10.1111/gcb.12935

Siegmann, B., and Jarmer, T. (2015). Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. *Int. J. Remote Sens.* 36, 4519–4534. doi: 10.1080/01431161.2015.1084438

Simopoulos, C. M., Weretilnyk, E. A., and Golding, G. B. (2018). Prediction of plant lncRNA by ensemble machine learning classifiers. *BMC Genom.* 19:316. doi: 10.1186/s12864-018-4665-2

Stroup, W., and Mulitze, D. (1991). Nearest neighbor adjusted best linear unbiased prediction. *Am. Statistician* 45, 194–200. doi: 10.1080/00031305.1991.10475801

Su, C.-T., and Yang, C.-H. (2008). Feature selection for the SVM: an application to hypertension diagnosis. *Exp. Systems Appl.* 34, 754–763. doi: 10.1016/j.eswa.2006.10.010

Su, Q., Lu, W., Du, D., Chen, F., Niu, B., and Chou, K.-C. (2017). Prediction of the aquatic toxicity of aromatic compounds to tetrahymena pyriformis through support vector regression. *Oncotarget* 8, 49359–49369. doi: 10.18632/oncotarget.17210

Suykens, J. A., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.* 9, 293–300. doi: 10.1023/A:1018628609742

Sweeney, D. W., Long, J. H., and Kirkham, M. (2003). A single irrigation to improve early maturing soybean yield and quality. *Soil Sci. Soc. Am. J.* 67, 235–240. doi: 10.2136/sssaj2003.2350

Taillardat, M., Mestre, O., Zamo, M., and Naveau, P. (2016). Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Monthly Weather Rev.* 144, 2375–2393. doi: 10.1175/MWR-D-15-0260.1

Tamouridou, A. A., Alexandridis, T. K., Pantazi, X. E., Lagopodi, A. L., Kashefi, J., Kasampalis, D., et al. (2017). Application of multilayer perceptron with automatic relevance determination on weed mapping using UAV multispectral imagery. *Sensors* 17:2307. doi: 10.3390/s17102307

Tan, M., Wang, L., and Tsang, I. W. (2010). "Learning sparse svm for feature selection on very high dimensional datasets," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, (Haifa).

Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27, R770–R783. doi: 10.1016/j.cub.2017.05.055

Tilman, D., Socolow, R., Foley, J. A., Hill, J., Larson, E., Lynd, L., et al. (2009). Beneficial biofuels—the food, energy, and environment trilemma. *Science* 325, 270–271. doi: 10.1126/science.1177970

Tiwari, P., and Shukla, P. (2020). "Artificial neural network-based crop yield prediction using NDVI, SPI, VCI feature vectors," in *Proceedings of the Information and Communication Technology for Sustainable Development Proceedings of ICT4SD*, (Springer), 585–594. doi: 10.1007/978-981-13-7166-0_58

Vapnik, V. (2000). *The Nature of Statistical Learning Theory*. New York, NY: SpringerVerlag. doi: 10.1007/978-1-4757-3264-1

Verma, A. K., Garg, P. K., Hari Prasad, K., and Dadhwal, V. K. (2020). Modelling of sugarcane yield using LISS-IV data based on ground LAI and yield observations. *Geocarto Int.* 35, 887–904. doi: 10.1080/10106049.2018.1544291

Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., et al. (2020). Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer–a case study of small farmlands in the South of China. *Agricult. Forest Meteorol.* 291:108096. doi: 10.1016/j.agrformet.2020.108096

Wang, Y., Wang, F., Huang, J., Wang, X., and Liu, Z. (2009). Validation of artificial neural network techniques in the estimation of nitrogen concentration in rape using canopy hyperspectral reflectance data. *Int. J. Remote Sens.* 30, 4493–4505. doi: 10.1080/01431160802577998

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Rev. Comput. Statist.* 3, 180–185. doi: 10.1002/wics.147

Xiong, Q., Tang, G., Zhong, L., He, H., and Chen, X. (2018). Response to nitrogen deficiency and compensation on physiological characteristics, yield formation, and nitrogen utilization of rice. *Front. Plant Sci.* 9:1075. doi: 10.3389/fpls.2018.01075

Zhang, N., Pan, Y., Feng, H., Zhao, X., Yang, X., Ding, C., et al. (2019). Development of Fusarium head blight classification index using hyperspectral microscopy images of winter wheat spikelets. *Biosystems Eng.* 186, 83–99. doi: 10.1016/j.biosystemseng.2019.06.008

Zhang, Z., Jin, Y., Chen, B., and Brown, P. (2019). California almond yield prediction at the orchard level with a machine learning approach. *Front. Plant Sci.* 10:809. doi: 10.3389/fpls.2019.00809

Zheng, W., Zhu, X., Wen, G., Zhu, Y., Yu, H., and Gan, J. (2020). Unsupervised feature selection by self-paced learning regularization. *Pattern Recogn. Lett.* 132, 4–11. doi: 10.1016/j.patrec.2018.06.029

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL: CRC press. doi: 10.1201/b12207

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yoosefzadeh-Najafabadi, Earl, Tulpan, Sulik and Eskandari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

METHODS
published: 21 January 2021
doi: 10.3389/fpls.2020.599886

Fusion of Deep Convolution and Shallow Features to Recognize the Severity of Wheat *Fusarium* Head Blight

Chunyan Gu[1,2†], Daoyong Wang[1,2†], Huihui Zhang[3], Jian Zhang[4], Dongyan Zhang[1,2] and Dong Liang[2]*

[1] Institute of Plant Protection and Agro-products Safety, Anhui Academy of Agricultural Sciences, Hefei, China, [2] National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei, China, [3] Water Management and Systems Research Unit, USDA Agricultural Research Service, Fort Collins, CO, United States, [4] Macro Agriculture Research Institute, College of Resource and Environment, Huazhong Agricultural University, Wuhan, China

OPEN ACCESS

Edited by:
Spyros Fountas, Agricultural University of Athens, Greece

Reviewed by:
Pengfei Chen, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, China
Jingcheng Zhang, Hangzhou Dianzi University, China

***Correspondence:**
Dongyan Zhang
zhangdy@ahu.edu.cn

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *28 August 2020*
Accepted: *30 November 2020*
Published: *21 January 2021*

Citation:
Gu C, Wang D, Zhang H, Zhang J, Zhang D and Liang D (2021) Fusion of Deep Convolution and Shallow Features to Recognize the Severity of Wheat Fusarium Head Blight. Front. Plant Sci. 11:599886. doi: 10.3389/fpls.2020.599886

A fast and nondestructive method for recognizing the severity of wheat *Fusarium* head blight (FHB) can effectively reduce fungicide use and associated costs in wheat production. This study proposed a feature fusion method based on deep convolution and shallow features derived from the high-resolution digital Red-green-blue (RGB) images of wheat FHB at different disease severity levels. To test the robustness of the proposed method, the RGB images were taken under different influence factors including light condition, camera shooting angle, image resolution, and crop growth period. All images were preprocessed to eliminate background noises to improve recognition accuracy. The AlexNet model parameters trained by the ImageNet 2012 dataset were transferred to the test dataset to extract the deep convolution feature of wheat FHB. Next, the color and texture features of wheat ears were extracted as shallow features. Then, the Relief-F algorithm was used to fuse the deep convolution feature and shallow features as the final FHB features. Finally, the random forest was used to classify and identify the features of different FHB severity levels. Results show that the recognition accuracy of the proposed fusion feature model was higher than those of models using other features in all conditions. The highest recognition accuracy of severity levels was obtained when images were taken under indoor conditions, with high resolution (12 MB pixels), at 90° shooting angle during the crop filling period. The Relief-F algorithm assigned different weights to the features under different influence factors; it made the fused feature model more robust and improved the ability to recognize wheat FHB severity levels using RGB images.

Keywords: *Fusarium* head blight, transfer learning, Relief-F, fusion feature, random forest

INTRODUCTION

Fusarium head blight (FHB) mainly caused by *Fusarium graminearum* is a devastating disease of wheat and has a serious impact on wheat production worldwide, especially in China (Huang and Mcbeath, 2010). FHB-infected wheat will produce deoxynivalenol (DON) toxin that is poisonous to humans or animals and can persist for a long time in the food chain (Palacios et al., 2017;

Peiris et al., 2017). If FHB can be detected effectively and disease severity level can be determined precisely, it can be controlled timely by applying fungicides. Particularly, the right number of doses of fungicides can be appropriately allocated according to the severity level to reduce the cost of fungicide application and protect ecological environment to a great extent (Yuan and Zhang, 2000). Recognition of wheat FHB was usually performed visually by experienced plant protectors in fields (Fernando et al., 2017). It is subjective, time-consuming, and laborious. Recently, some studies have utilized hyperspectral technology and image processing technologies for FHB recognition. Hyperspectral technology has high-level technical requirements and high costs. And also, it has a high demand for the natural environments such as light and wind and so on when collecting hyperspectral data (Bauriegel et al., 2011; Jin et al., 2018). Image processing technologies have strong generality, high efficiency, low cost, and low operating requirements in disease recognition (Mohd et al., 2019; Pantazi et al., 2019). Red-green-blue (RGB)–based images have been widely used in wheat crops. Although some valuable progress has been made (Jin et al., 2017; Aarju and Sumit, 2018), there is still a need to improve rating FHB severity level accurately by utilizing RGB images.

The first key request is to extract effective features from RGB images (Wang and Paliwal, 2003). Zahra and Davud (2015) extracted texture, color, and shape features of RGB images to recognize wheat fungal diseases with an accuracy of 98.3%. Frederic and Pierre (2007) extracted color and texture features of wheat ear RGB images to identify the wheat ear regions. Liu and Cui (2015) distinguished wheat from the background based on RGB and Lab color space and used the random forest (RF) algorithm to accurately segment the targeted winter wheat from RGB image of the canopy. The abovementioned color and texture features are widely used for crop disease identification (Zhu et al., 2017; Xiao et al., 2018), so they are fundamental references for wheat FHB identification. Among color features, the RGB channel, HSV channel, and Lab channel are widely used because they can effectively express the differences between diseased and other areas (Pydipati et al., 2006; Meunkaewjinda et al., 2008; Yao et al., 2009). Among texture features, the gray-level co-occurrence matrix (GLCM) can reflect the comprehensive information of the image about the direction, adjacent interval, and amplitude change (Chaki et al., 2015; Gavhale et al., 2015; Xie and He, 2016). Compared to color and texture features, deep convolution feature can well excavate the deep features information in images (Shervin et al., 2016; Lu et al., 2017). The LeNet-5 model proposed by Yann et al. (1998) has made the convolutional neural network achieve excellent results in the field of handwritten digit recognition for the first time and established the reputation of convolutional neural network in image recognition. In the ImageNet 2012 competition, the AlexNet deep convolutional neural network proposed by Alex et al. (2012) won the championship. Its classification results were much better than other traditional machine learning classification algorithms. The AlexNet deep convolutional neural network has attracted widespread attention since it was used in crop identification with good accuracy (Mostafa et al., 2017; Wei et al., 2018).

The accuracy of disease recognition from RGB images depends on the contribution from each extracted feature. Researchers have proposed feature selection algorithms for feature screening (Mitra et al., 2002; Marko and Igor, 2003; Valliammal and Geethalakshmi, 2012). Peng et al. (2005) proposed a minimal-redundancy-maximal-relevance criterion (mRMR) algorithm. The basic idea of mRMR was to use the theory of relevance in information theory and the size of mutual information as a measure of the correlation between features, as well as the sexual standards of features and category labels. Kira and Rendell (1992) proposed the Relief-F algorithm, which assigns different weights to all features according to the relevance of each feature and category. This algorithm is favorable by researchers because of its high efficiency, good results, and no limitation on data types (Durgabai et al., 2014; Wang et al., 2016).

Therefore, in this study we proposed the following procedure, especially a feature fusion method, to recognize the severity of wheat FHB. First, the deep convolution features of RGB images was extracted using the AlexNet convolutional neural network, and the color features and texture features of the images were extracted as shallow features. Next, to improve the recognition accuracy of wheat FHB severity levels, the Relief-F algorithm was used to fuse the extracted deep convolution feature and shallow features. Finally, the RF algorithm (Bosch et al., 2007) was used to model the features under different influence factors to explore the performance of the fusion features.

MATERIALS AND METHODS

Study Area and Image Acquisition

The experimental base in this study locates at Anhui Academy of Agricultural Sciences (117°14′E, 31°53′N), in Auhui Province, China. The field experiments of wheat FHB were conducted from April 28, 2018 (flowering period), to May 14, 2018 (ripening period). **Figure 1** shows the experimental field, which was divided into two sections: one was inoculated with FHB fungus inoculation, and the other was naturally grown. The inoculation section was gradually infected to form different levels of infection. A Nikon D3200 camera (**Table 1**) (effective pixels 6,016 × 4,000, focal length: 26 mm, aperture: f/8, exposure time: 1/250 s) was used to collect wheat ear images on sunny and cloudless days to reduce image distortion due to changing weather conditions. A total of 3,600 images of the wheat ear with FHB infection were collected. Among them, 1,200 images were randomly selected for the AlexNet learning, and the remaining 2,400 images were used for wheat FHB classification.

Images were acquired to ensure that only one wheat ear was in the lens with a black cloth as background, while the following influence factors were considered (**Table 2**):

1. Light condition

   Outdoor—under the influence of natural light and the light is uneven.
   Indoor—to minimize the influence of other illumination, a halogen lamp provides the light source in the dark room to make the wheat ears receive the light evenly.

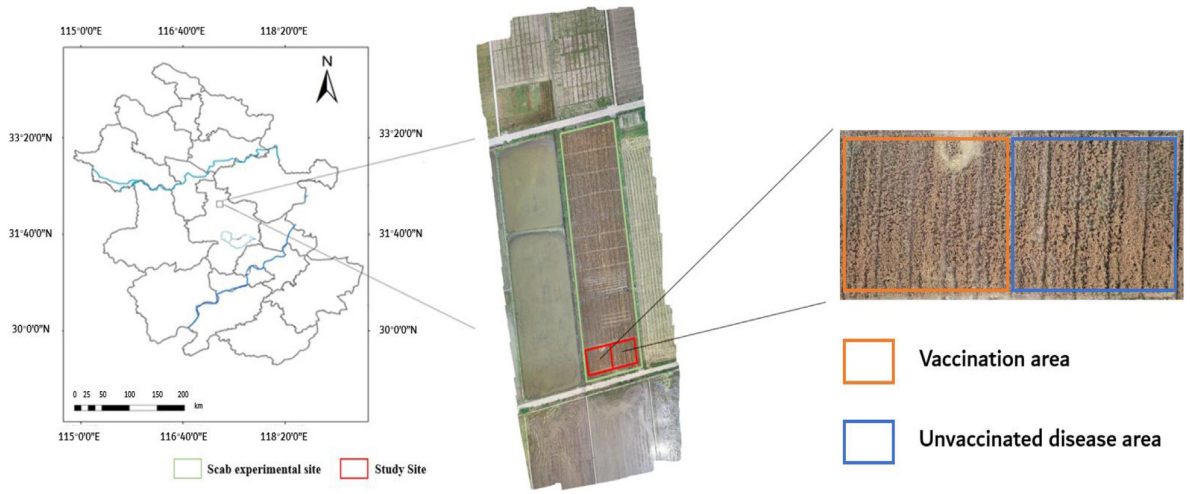

FIGURE 1 | Location of the study site and an RGB image of the experimental plot was taken by a DIJ Spirit 4Pro drone at an altitude of 50 m on May 14, 2018.

2. Shooting angle

   The angle of camera lens and wheat ear was set up at 30°, 45°, and 90° (Zhao et al., 2013).

3. Image resolution

   Nikon D3200 was adjusted to the resolution of 12 million effective pixels (4,512 × 3,000) and 6 million effective pixels (3,008 × 2,000), respectively.

4. Wheat growth period

   Flowering, filling, and ripening period. The data were collected in the middle of each growth period.

While images were taken, the actual disease level of the ear in each image was manually identified by professional personnel. The GBT 15796-2011 Rules for Monitoring and Forecast of the Wheat Head Blight was referred to determine the infected level of FHB. The disease was classified into six levels based on the ratio (R) of wheat ear lesion area to wheat ear area, as Level 0: $0 \leq R \leq 0.01$, Level 1: $0.01 < R \leq 0.1$, Level 2: $0.1 < R \leq 0.2$, Level 3: $0.2 < R \leq 0.3$, Level 4: $0.3 < R \leq 0.4$, and Level 5: $R > 0.4$ (**Figure 2**).

TABLE 1 | Nikon D3200 manufacturing parameters.

Nikon D3200 manufacturing parameters	
Boundary dimension	125 × 96 × 76.5 mm
Sensor type	CMOS
Sensor size	APS frame (23.2 × 15.4 mm)
Maximum pixels	24.72 million
Effective pixels	24.16 million
Image processor	EXPEED 3
Image resolution	16 × 4,000 (L), 4,512 × 3,000 (M), 3,008 × 2,000 (S)

Methods

The overall procedure to determine the FHB-infected level by RGB images is shown in **Figure 3**. First, the raw images were preprocessed to remove interference information. Then, the deep convolution feature of the preprocessed images was extracted based on the AlexNet transfer learning, and then the color and texture features of the preprocessed images were extracted as shallow features. Next, the deep convolution feature and shallow features were merged, and the Relief-F algorithm was used to calculate the weights of the merged features. The weight values were normalized to make the weights more numerically comparable. Then, the weight value was multiplied by its corresponding feature. To improve the accuracy of the model, the final features were normalized and used as fusion features. Finally, all fusion features were input into a RF model to recognize the FHB severity level. Details on each step are given in the following sections.

Image Preprocessing

Figure 4 shows an example of preprocessing raw images. First, the raw image was gray-scaled (Dougherty and Lotufo, 2003a), and the Otsu (Feige, 1999) threshold method was used for binarization. Next, a morphological region threshold filter (Dougherty and Lotufo, 2003b) was used to remove the noises such as the small dust on the black cloth. Then, a morphological open–close operation was used to remove the awn from the wheat ear to obtain the binary image of interest. Finally, the binary image was combined with the original image to produce a pseudocolor image.

Transfer Learning Based on the AlexNet to Extract Deep Convolution Feature

The AlexNet (Alex et al., 2012) was used as the research network to extract the deep convolution feature of FHB images. It was not enough to train an excellent network with a small sample size,

TABLE 2 | RGB images of the wheat ears were taken under different influence factors.

Influence factor			Control influence factors			
			Light	Camera shooting angle (degrees)	Image resolution (MB pixels)	Wheat growth period
Design of influence factors in the experiment	Light	Indoor		90	12	Filling
		Outdoor		90	12	Filling
	Shooting angle (degree)	30	Outdoor		12	Filling
		45	Outdoor		12	Filling
		90	Outdoor		12	Filling
	Image resolution (MB pixels)	12	Indoor	90		Filling
		6	Indoor	90		Filling
	Wheat growth period	Flowering	Outdoor	90	12	
		Filling	Outdoor	90	12	
		Ripening	Outdoor	90	12	

FIGURE 2 | The images of FHB-infected wheat ears. **(A)** Healthy wheat ear with a disease level of 0; **(B)** wheat ear with a disease severity level of 1, **(C)** 2, **(D)** 3, **(E)** 4, and **(F)** 5.

so the AlexNet model parameters trained using the ImageNet 2012 (Alex et al., 2012) dataset were transferred to FHB image sets for training. The parameters of the first five convolutional layers and corresponding pooling layers were retained, and the parameters of the three fully connected layers were trained. In addition, as the requested input of the AlexNet network was a 227 × 227 RGB image, the edges of the original FHB images were filled with 0 so that the aspect ratio of each image was 1. And then the images were resampled to 227 × 227 using bilinear interpolation (Kirkland, 2010). The network structure was shown in **Figure 5**.

A total of 1,200 images were used for the AlexNet transfer learning, where the training set had 840 images and the validation set had 360 images. The training parameters were set as follows: the learning rate was set to 0.0001, Maxepochs was set to 300, and the batch size was set to 20. The learning rate determines how quickly the parameter moves to the optimal value. Maxepochs represent the total number of trainings. The batch size indicates the number of samples used in each training batch in the training set. The final training time was 0.33 h, and the verification accuracy was 0.867. Results show that the AlexNet's transfer learning could be used to recognize FHB, but accuracy was not good enough. Therefore, this study proposed to extract the deep convolution feature of the disease images through the network obtained by transfer learning and recognize the severity of the disease based on the deep convolution feature.

Shallow Features

The shallow features were extracted from the color and texture of FHB images as follows:

1. Color features: Select the B component of the RGB color space, the *a* component of the Lab color space

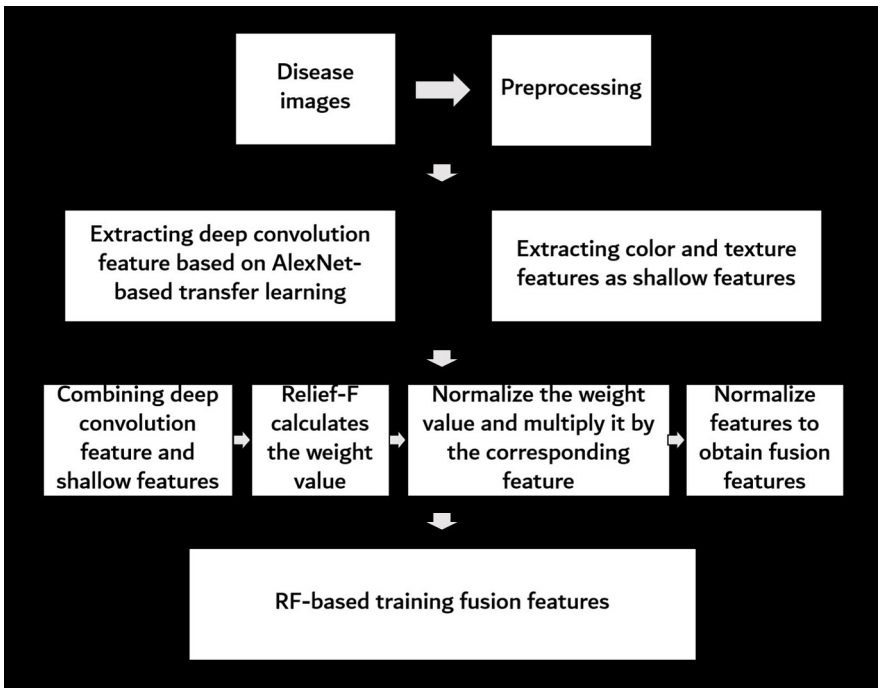

FIGURE 3 | Flowchart of image processing procedure used in the study.

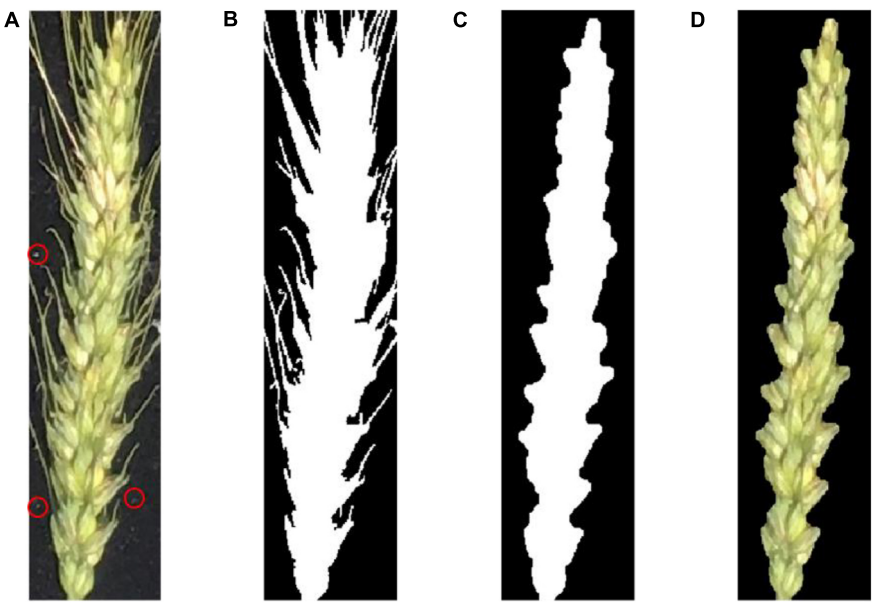

FIGURE 4 | (A) A raw image of a wheat ear. The red circle indicates some fine dust on the black cloth; **(B)** binary image after removing background noise; **(C)** binary image after removing wheat awn from the ear; **(D)** pseudocolor image of the wheat ear in the area of interest.

(Gauch and Hsia, 1992), and the S component of the HSV color space (Sural et al., 2002) of the disease image to describe the color features. Among them, the *a* color ranges from dark green (low brightness value), gray (medium brightness value), to bright pink (high brightness value). S stands for saturation. The higher the saturation, the darker the color.

2. Texture features: The mean and variance of energy, entropy, inverse different moment, correlation, and contrast in GLCM (Haralick et al., 1973) were selected to describe the texture features of FHB images. Energy reflects the uniformity and texture of the gray distribution of the image. Entropy is a measure of the amount of information in the image.

Inverse different moment reflects local changes in the texture of the image. Correlation reflects the consistency of the image texture. Contrast reflects the image sharpness and depth of texture grooves.

Feature Fusion of Deep Convolution Feature and Shallow Features

The deep convolution feature and shallow features were combined. The Relief-F algorithm was used to iterate 100 times to calculate the average weight of the combined features and normalize the weight value. The weight value was multiplied with its corresponding feature to obtain the weighted feature. Finally, the fused features were obtained by normalizing the weighted features to improve the accuracy of the model. The fusion formula was as follows:

$$Fweight_l = \frac{w_l}{\sum_{l=1}^{k} w_l} \times Fcascade_l \quad (1)$$

$$Fnormalization_{i,j} = \frac{Fweight_{i,j}}{\sum_{j=1}^{n} Fweight_{i,j}} \quad (2)$$

where $_{Fcascade_l}$ is the l feature set specially collected for cascade sample, $l = 1, 2, \ldots, k$, k is the feature dimension, $_{w_l}$ is the feature weight of the l, $_{Fweight_l}$ is the l feature set after calculating the

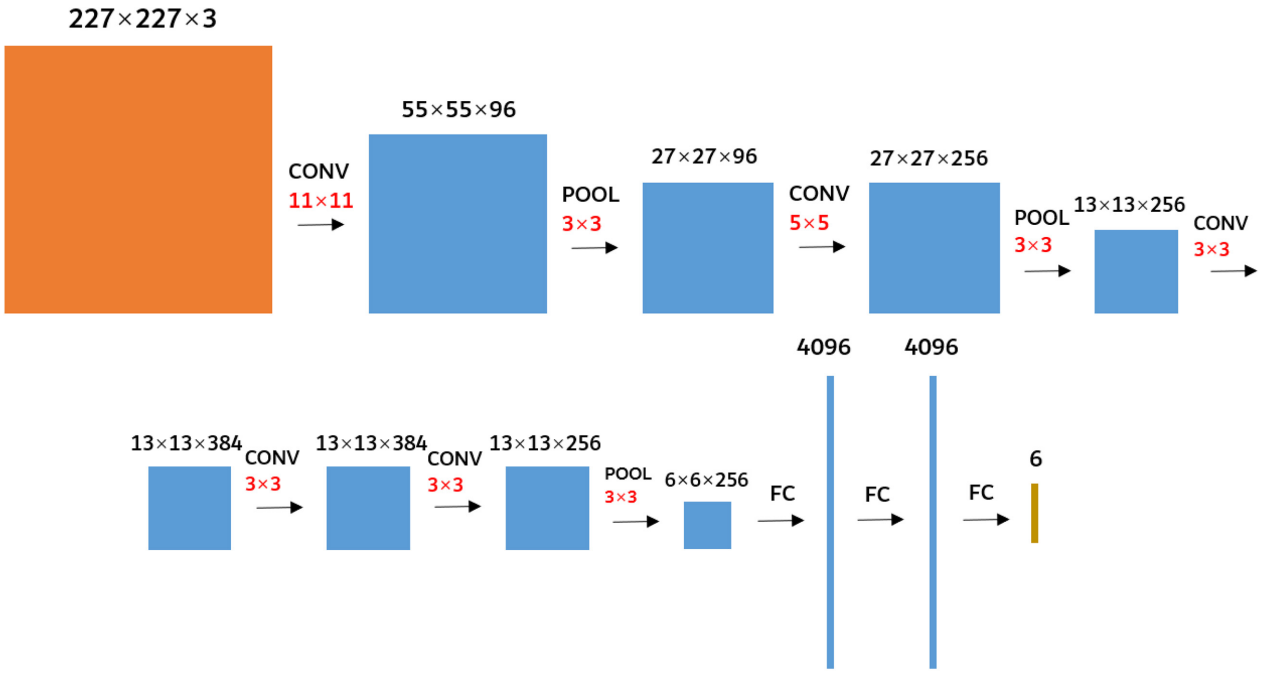

FIGURE 5 | Diagram of the AlexNet network structure. The black numbers represent the size of the feature map. The red number represents the size of the kernels. CONV represents the convolution. POOL represents the maximum pooling. FC represents the fully connected.

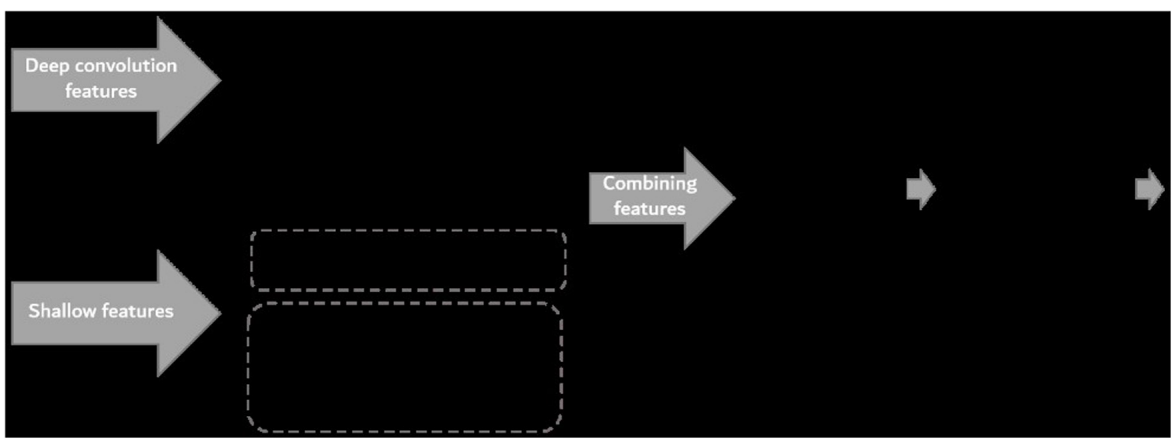

FIGURE 6 | Structure of feature fusion network.

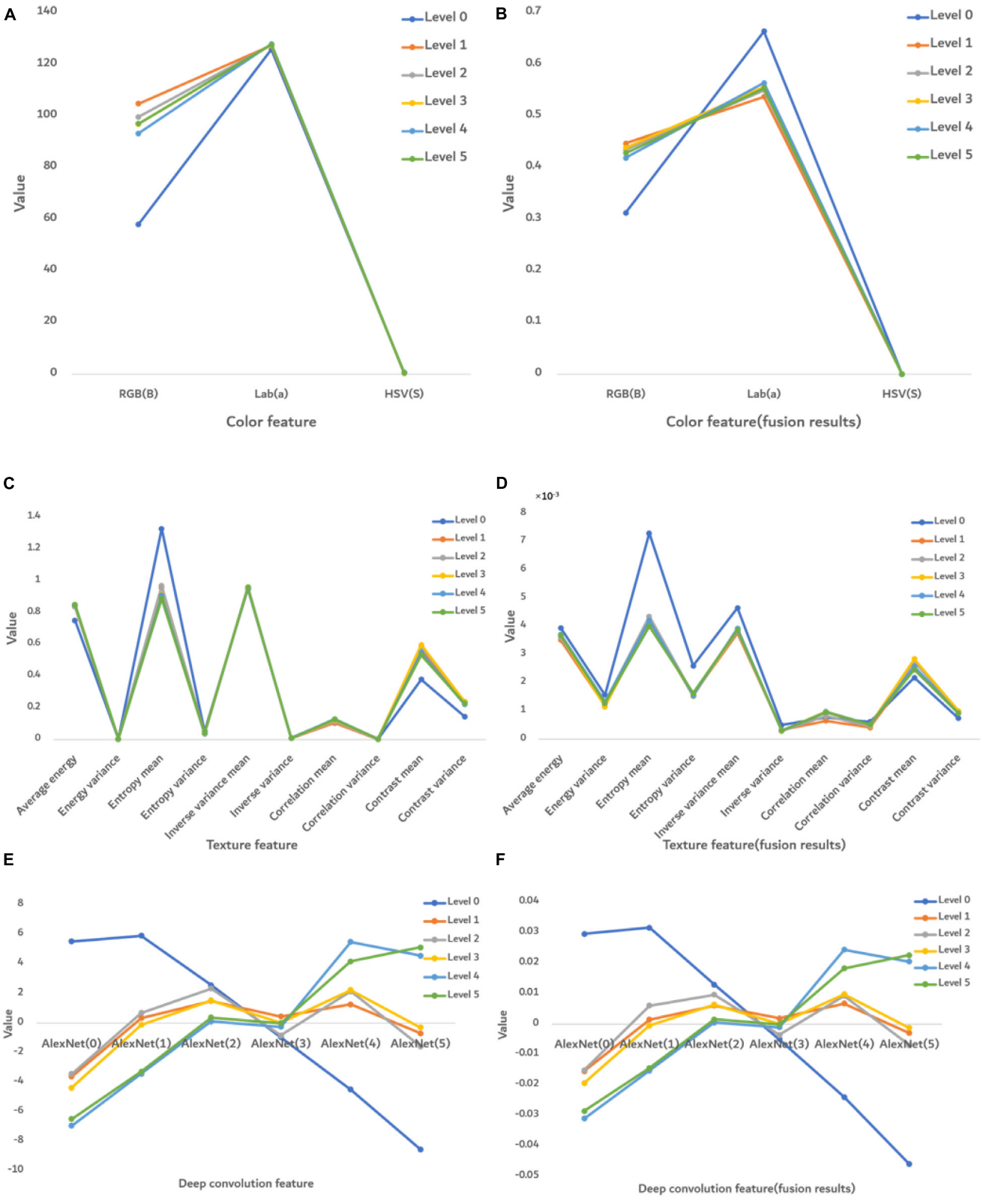

FIGURE 7 | The numerical performance of features at each disease severity level. **(A)** Color features, **(B)** color features after fusion, **(C)** texture features, **(D)** texture features after feature fusion, **(E)** deep convolution feature, and **(F)** deep convolution feature after feature fusion. The color lines are the average eigenvalue of all samples at each severity level.

TABLE 3 | Number of images in the training and test set of each disease level under different influence factors.

Influence factor	Sample set	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Light	Training set	90	307	164	157	117	215
Indoor	Test set	39	131	70	68	51	91
Outdoor	Test set	39	131	70	68	51	91
Angle	Training set	73	222	96	93	57	82
30°	Test set	32	96	42	42	27	38
45°	Test set	32	96	42	42	27	38
90°	Test set	32	96	42	42	27	38
Resolution	Training set	67	277	137	132	79	148
Low resolution	Test set	29	119	58	57	35	62
High resolution	Test set	29	119	58	57	35	62
Growth period	Training set	57	185	93	77	73	142
Flowering period	Test set	24	80	39	34	32	64
Filling period	Test set	24	80	39	34	32	64
Ripening period	Test set	24	80	39	34	32	64

weight; $_{F\text{weight}_{i,j}}$ is the eigenvalue of the row i, column j of the feature set after weight calculation, i = 1,2,. . ., m, j = 1,2,. . ., n, and m are the number of rows in the feature set, n is the number of columns in the feature set, and $_{F\text{normalization}_{i,j}}$ is the normalized feature set.

The feature fusion structure is given in **Figure 6**. The six-dimension deep convolution feature and shallow features (13-dimension) were extracted from input images. Relief-F was used to calculate the weights of the deep convolution feature (six-dimension) and shallow features (13-dimension) and the final fused features had 19 dimensions.

Random Forest

Random forest algorithm (Cutler et al., 2011) was a machine learning algorithm composed of multiple decision tree classification models. First, N training sets were extracted from the original dataset using the bootstrap (Cutler et al., 2011) sampling technique. Then, a classification regression tree was established for each training set, N decision tree models were generated, and N classification results were obtained. Finally, the prediction results of N decision trees were set to determine the category of the new sample by voting. RF gives results based on the prediction results of multiple decision trees, so even if some decision trees are misclassified, the final classification results were still correct.

RESULTS

The results were compared based on the measured severity level and the fusion feature classification.

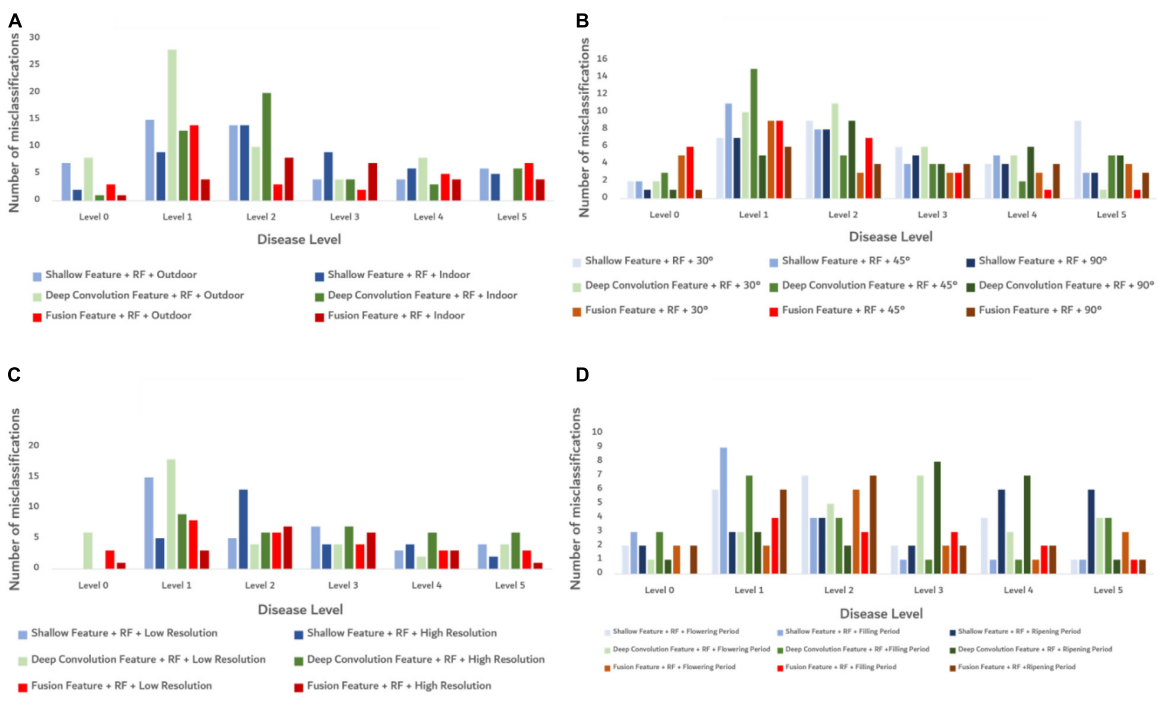

FIGURE 8 | Classification error diagram of the models at each severity level under different conditions. **(A)** The misclassification of each model under different light conditions. **(B)** The misclassification of each model under different shooting angle conditions. **(C)** The misclassification of each model under different resolution conditions. **(D)** The misclassification of each model under different growth periods.

The algorithm was developed in Pycharm2017 and completed on Windows 10 PC with 12-core Intel core i7-6800k CPU (3.40 GHz), 16 GB RAM, and dual GTX1080Ti GPU. The influence factors of light, shooting angle, image resolution, and growth period were considered.

TABLE 4 | Model performance under different influence factors.

Influence factor		Model	Model accuracy	Cross validation accuracy	Time (s)
Light	Outdoor	Shallow features + RF	0.8889	0.8880	15.70
	Outdoor	Deep convolution feature + RF	0.8711	0.8746	13.34
	Outdoor	Fusion features + RF	0.9244	0.9213	16.25
	Indoor	Shallow features + RF	0.9000	0.9000	15.54
	Indoor	Deep convolution feature + RF	0.8956	0.8933	12.91
	Indoor	Fusion features + RF	0.9378	0.9310	15.87
Angle	30°	Shallow features + RF	0.8664	0.8685	7.02
	30°	Deep convolution feature + RF	0.8736	0.8783	6.41
	30°	Fusion features + RF	0.9025	0.9087	7.08
	45°	Shallow features + RF	0.8809	0.8804	6.92
	45°	Deep convolution feature + RF	0.8773	0.8728	6.35
	45°	Fusion features + RF	0.9025	0.9098	6.99
	90°	Shallow features + RF	0.8989	0.8925	6.99
	90°	Deep convolution feature + RF	0.8917	0.8807	6.90
	90°	Fusion features + RF	0.9206	0.9170	7.23
Resolution	Low resolution	Shallow features + RF	0.9056	0.9002	9.25
	Low resolution	Deep convolution feature + RF	0.8944	0.8920	7.67
	Low resolution	Fusion features + RF	0.9250	0.9225	9.16
	High resolution	Shallow features + RF	0.9222	0.9212	9.14
	High resolution	Deep convolution feature + RF	0.9056	0.9050	8.50
	High resolution	Fusion features + RF	0.9417	0.9339	9.33
Growth period	Flowering period	Shallow features + RF	0.9194	0.9165	6.98
	Flowering period	Deep convolution feature + RF	0.9158	0.9176	6.02
	Flowering period	Fusion features + RF	0.9414	0.9484	7.13
	Filling period	Shallow features + RF	0.9304	0.9286	6.95
	Filling period	Deep convolution feature + RF	0.9267	0.9264	6.26
	Filling period	Fusion features + RF	0.9524	0.9516	7.21
	Ripening period	Shallow features + RF	0.9158	0.9143	7.05
	Ripening period	Deep convolution feature + RF	0.9194	0.9165	6.31
	Ripening period	Fusion features + RF	0.9377	0.9407	6.92

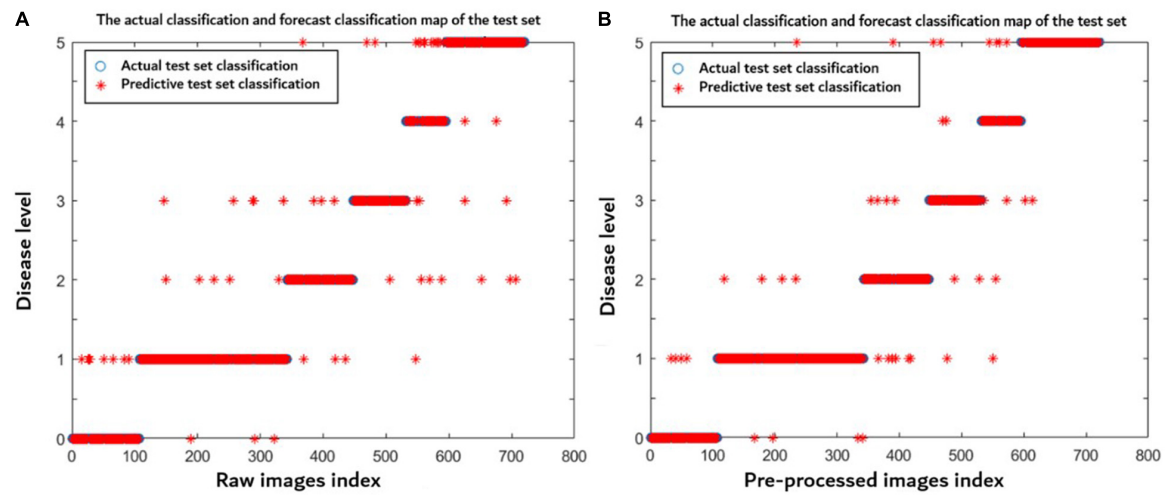

FIGURE 9 | Disease level prediction distribution of each model using **(A)** raw images and **(B)** preprocessed images. The blue symbol indicates the actual disease severity level, and the red symbol indicates the predicted disease severity level.

Feature Extraction

Figure 7 shows the numerical performance of features at each severity level. The value of each feature was obtained from averaging 2,400 images.

Before fusion (**Figure 7A**), only the RGB (B) feature was significantly different at each severity level, whereas after fusion (**Figure 7B**), the features were significantly different in RGB (B) and Lab (a). However, HSV(S) feature was the same before or after fusion, so this S feature is not able to distinguish FHB. The texture features before fusion could not effectively distinguish severity levels in **Figure 7C**, whereas after fusion they show some differences. It indicates that the ability of disease classification was improved by feature fusion. **Figures 7E,F** show that the deep convolution feature was able to distinguish disease severity levels both before and after fusion.

Model Construction

To compare the performance of each model under different influence factors more intuitively, this study constructed each model based on the collected sample set and comprehensively evaluated the model results. The image distribution at each disease level under different influence factors was shown in **Table 3**.

Random forest algorithm was used to build the models under different influence factors, and the classification errors of the models at each severity level are shown in **Figure 8**.

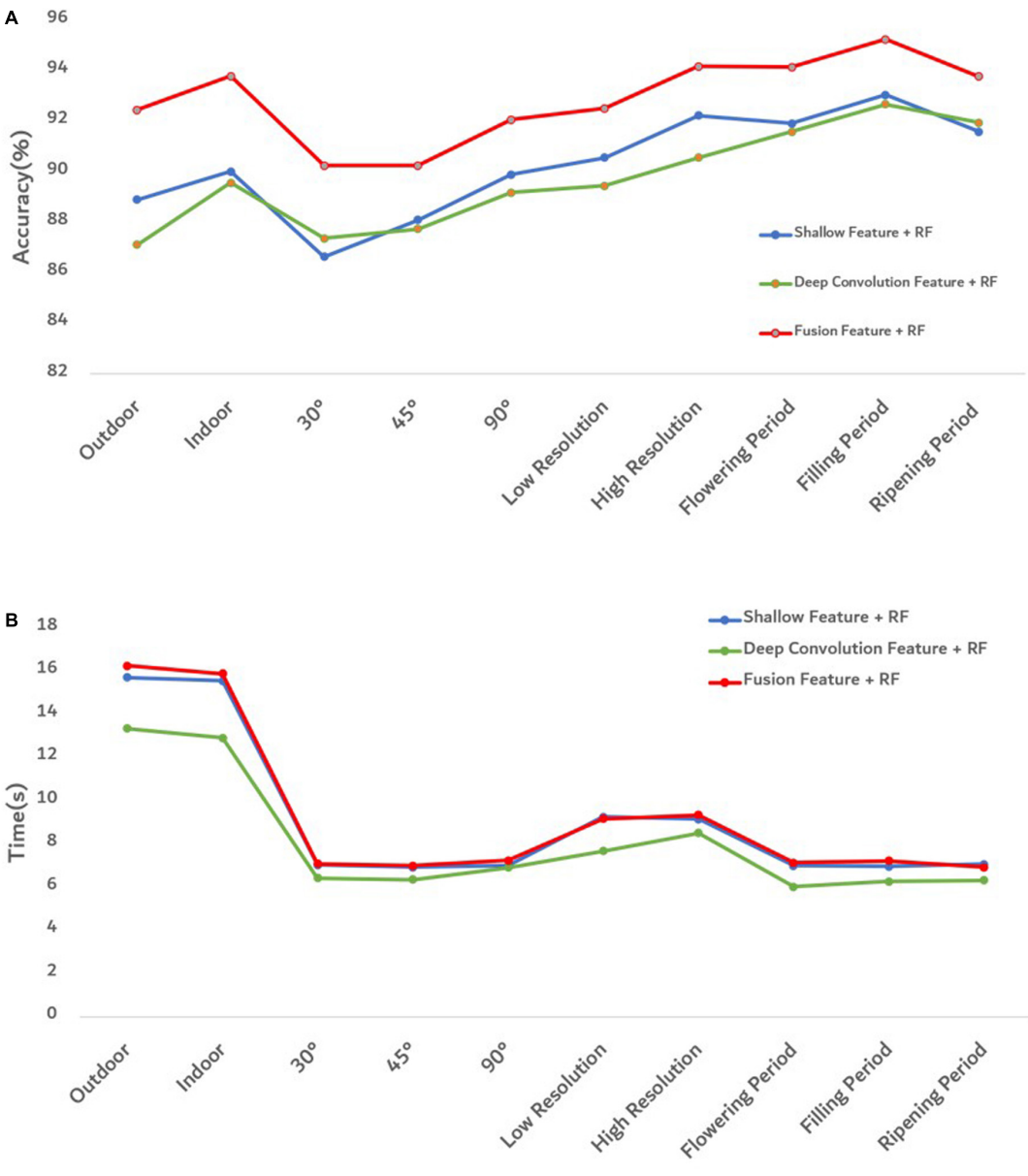

FIGURE 10 | Evaluation diagram of each model under different influence factors. **(A)** The accuracy of each model. **(B)** The training and prediction time of each model.

Under all influencing factors, the fusion features model has the least number of misclassifications. When the disease levels are 1 and 2, each model has more misclassifications than other disease levels. It shows that the fusion features performed better than other features in the models under different influence factors with a smaller number of misclassifications. But when the disease level was 1 or 2, the recognition accuracy of each model is poor. To further study the efficiency and stability of the model, the 10-fold cross-validation (Kohavi, 1995) was used to cross-verify the data, and the predicted time of each model training was calculated. The results are shown in **Table 4**. The accuracy of each model was better under indoor than outdoor lighting conditions. Under three observation angles, the 90° observation angle was the best in terms of accuracy. The images with higher resolution appropriately improved model accuracy. The identification accuracy was the highest for images taken during the crop filling period. Considering both model accuracy and training prediction time, the fusion features proposed in this study performed better than using the shallow features or deep convolution feature independently under different influence factors. This result indicates that the fusion features had high accuracy and strong robustness in the recognition of FHB severity level.

DISCUSSION

Analysis of Image Preprocessing Results

In order to better discuss the advantages of preprocessing in this article, a total of 2,400 disease images under different influence factors were made into one sample set including 1,680 images in the training set and 720 images in the test sets. The raw and preprocessed images were used to build the models through the method proposed in this article, respectively. **Figure 9** indicates that the model established using preprocessed images has a better prediction distribution of the disease level and higher model accuracy (0.943) than using the raw images (0.922). After preprocessing, the noises and wheat ears were removed, so the accuracy of the model was improved.

Analysis of Fusion Feature

To better evaluate the goodness of fusion features, we evaluated the goodness of fusion features through the accuracy of each model and the prediction time of model training (**Figure 10**).

Figure 10A shows that the proposed fusion feature had a better performance on the accuracy of the model than the model constructed by other features. Under different influence factors, the proposed fusion method had stronger robustness, and model accuracy was greater than 90%, which was 2–5% higher than the recognition accuracy of deep convolution feature or shallow features. The efficiency of an algorithm is also important in practical applications (Yu et al., 2013). The predicted runtime of the model was used to evaluate the efficiency of the method. **Figure 10B** shows that although the fused features were higher in dimensions than other features, the model training time still performed good. To better analyze the effectiveness of the feature fusion method, FHB images with different influence factors were mixed as one sample set. After extracting the features of the sample set, 100 iterations were performed using Relief-F. The weight results obtained each time are shown in **Figure 11**. As the sample set was generated by mixing all influence factors, this increases the difficulty of

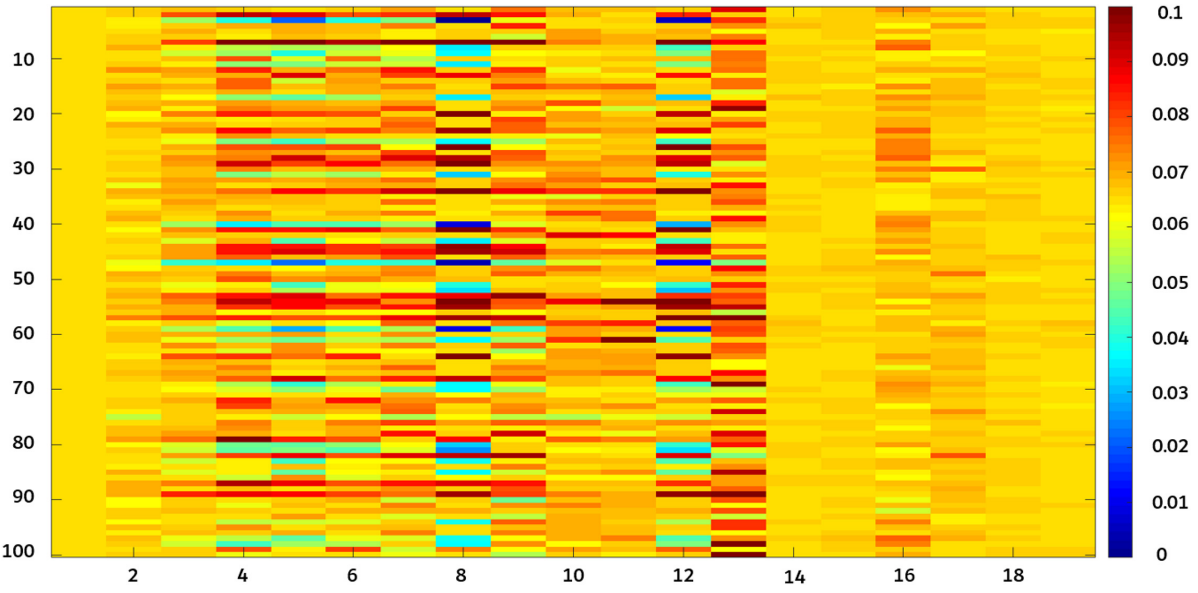

FIGURE 11 | Relief-F iterates 100 times for each weight value. X-axis 1–6 correspond to deep convolution feature (AlexNet0, AlexNet1, AlexNet2, AlexNet3, AlexNet4, and AlexNet5); 7–9 correspond to color features [RGB(B), Lab(a), and HSV(S)]; and 10–19 correspond to texture features (average energy, energy variance, entropy mean, entropy variance, inverse different moment mean, inverse different moment variance, correlation mean, correlation variance, contrast mean, and contrast variance).

FIGURE 12 | The average of final normalized weight value. The orange bar in the figure represents the weight value of different features calculated by the Relief-F algorithm.

training. In the meanwhile, the sample size of 2,400 was relatively small for deep learning. Therefore, the weight value of the deep convolution feature fluctuated greatly during the iteration process. For different growth periods and light conditions, the corresponding contribution of each color feature was different, resulting in a large fluctuation in the weight value of each color feature. For different resolution and shooting angles, the extracted texture features were different, so the weight values of texture features also fluctuated. Notice that the weight values of the deep convolution feature and shallow features fluctuated from low to high. It indicates that some features made more contributions when identifying certain influence factors than other factors. Therefore, when some features do not perform well in recognition, their weight values can be decreased so the weight values of other favorable features can be increased. Our results confirm that FHB image sets in this article can be well described by both deep convolution feature and the shallow features.

In **Figure 12**, the bar in the figure represents the weight value of different features calculated by the Relief-F algorithm. The larger the value was, the greater the contribution of the feature made. There were similar maxima in both deep convolution feature and shallow features, indicating that both deep convolution feature and shallow features made great contributions to identifying disease severity levels. The shallow features can well reflect the situation of different disease levels from the color and texture information of the images. Among the color features, the channel a of Lab color space performed better. In the texture features, except for the entropy variance and contrast variance, the other features performed well. The deep convolution feature can discover deep information well through convolutional processing.

Analysis of Different Influence Factors

To investigate the influence of different factors on the fusion feature method, FHBs under different conditions were analyzed, respectively (**Figure 13**).

In **Figure 13(1)**, the image taken in the outdoor environment was influenced by the difference of light and mirror effect (Barbedo, 2018), which has an impact on feature extraction. In an indoor environment, the image received relatively uniform light. Although there was a partial shadow in the gap between

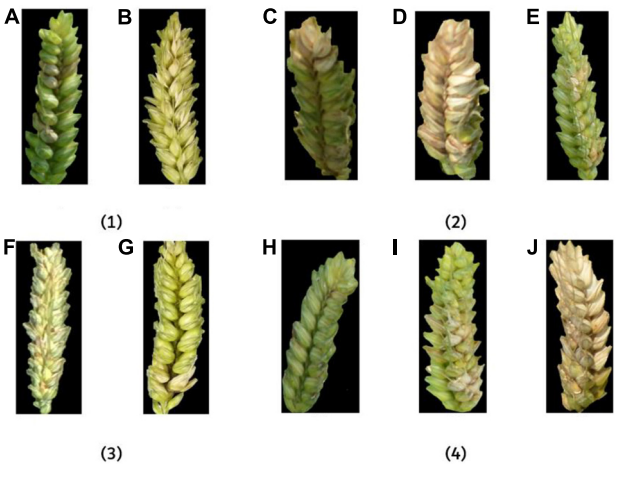

FIGURE 13 | The example images of FHB-infected wheat ears under different influence factors. In (1), **(A)** outdoor and **(B)** indoor; in (2), **(C)** 30°, **(D)** 45°, and **(E)** 90°; in (3), **(F)** low-resolution and **(G)** high-resolution; in (4), **(H)** flowering period, **(I)** filling period, and **(J)** ripening period.

the ears in some images, it had little effect on disease recognition. However, it can be seen from the results that the accuracy difference of indoor and outdoor on the fusion feature model was less than 1% (**Table 3**), which shows that the fusion method has a certain resistance to the influence of light. In **Figure 13(2)**, the larger the angle, the more information the camera can capture and the better the description of the disease. The smaller the angle, the more likely overlap will appear in the image and cause certain errors. According to the results, the recognition accuracy of 30° and 45° was similar, and the performance of 90° was the best among the three angles. In **Figure 13(3)**, high-resolution images were more informative and performed well in feature extraction. It can be seen from the results that the accuracy of high-resolution images was higher than that of low-resolution images. In **Figure 13(4)**, the color of the disease was not the same at different growth periods. FHB has just erupted during the flowering period, so the disease features were not obvious. FHB was more obvious in the filling period. The color difference between the normal wheat ear and FHB in the ripening period was small. Thus, the recognition accuracy during the filling period was the highest, and the recognition accuracy during the flowering period and the ripening period was basically the same.

To sum up, the model constructed with the fusion feature method has a certain resistance to different influence factors. More comparative experiments to explore a good collection environment should be considered in future research with valuable information provided by the study.

CONCLUSION

The study proposed a method to recognize disease severity levels of FHB-infected wheat ears using RGB images, which were taken under different influence factors, such as light condition, shooting angle, image resolution, and crop growth period. The deep convolution feature and shallow features extracted from these images were analyzed as contrast experiments for FHB identification. The feature fusion method was then proposed based on the deep convolution and shallow features under different influence factors. Results show that the recognition accuracy of the fusion features model was higher than that of using the deep convolution feature or shallow features alone. The prediction time of the feature fusion model was good, and it performed more robust under different influence factors. The highest accuracy of recognizing severity levels was obtained when images were taken indoor, with high resolution (12 MB pixels), at 90° shooting angle and during the crop filling period. The proposed feature extraction method has significant advantages in the identification of wheat FHB disease severity levels and provides important technical support for plant protection in precision fungicide application and the development of disease control methods.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CG, DW, DZ, and DL: conceptualization. CG and DW: data curation. DW: formal analysis and methodology. DW and DZ: investigation and resources. DW, HZ, and DZ: writing – original draft. CG, HZ, and JZ: writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant Nos. 41771463 and 41771469) and the Anhui Provincial Major Science and Technology Project (18030701209) and Natural Science Foundation of Anhui Province (Grant No. 1908085MC95).

REFERENCES

Aarju, D., and Sumit, N. (2018). Wheat leaf disease detection using machine learning method- a review. *Int. J. Comput. Sci. Mob. Comput.* 7, 124–129.

Alex, K., Ilya, S., and Geoffrey, E. H. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25.

Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. doi: 10.1016/j.biosystemseng.2018.05.013

Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U., and Herppich, W. B. (2011). Early detection of *Fusarium* infection in wheat using hyper-spectral imaging. *Comput. Electron. Agric.* 75, 304–312. doi: 10.1016/j.compag.2010.12.006

Bosch, A., Zisserman, A., and Muoz, X. (2007). "Image classification using random forests and ferns," in *Proceedings of the 2007 IEEE 11th International Conference on Computer Vision* (Piscataway, NJ: IEEE).

Chaki, J., Parekh, R., and Bhattacharya, S. (2015). Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognit. Lett.* 58, 61–68. doi: 10.1016/j.patrec.2015.02.010

Cutler, A., Cutler, D. R., and Stevens, J. R. (2011). Random forests. *Mach. Learn.* 45, 157–176.

Dougherty, E. R., and Lotufo, R. A. (2003a). *Gray-Scale Morphology*.

Dougherty, E. R., and Lotufo, R. A. (2003b). *Hands-on Morphological Image Processing*. Bellingham, WA: SPIE-The International Society for Optical Engineering.

Durgabai, R. P. L., Ravi, B. Y., and Vijayawada, A. P. (2014). Feature selection using reliefF algorithm. *Int. J. Adv. Res. Comput. Commun. Eng.* 3, 8215–8218.

Feige, U. (1999). A threshold of ln n for approximating set cover. *J. ACM* 45.

Fernando, A. M. S., Dennis, P., Jitendra, P., and Sherif, S. S. (2017). Assessment of *Fusarium* and deoxynivalenol using optical methods. *Food Bioproc. Tech.* 10, 34–50. doi: 10.1007/s11947-016-1788-9

Frederic, C., and Pierre, G. (2007). "Texture or color analysis in agronomic images for wheat ear counting," in *Proceedings of the 3rd International IEEE Conference on Signal-image Technologies and Internet-based System* (Washington, DC: IEEE Computer Society).

Gauch, J. M., and Hsia, C. W. (1992). Comparison of three-color image segmentation algorithms in four color spaces. *Proc. Spie Vis. Commun. Image Process.* 1818, 1168–1181.

Gavhale, K. R., Gawande, U., and Hajari, K. O. (2015). "Unhealthy region of citrus leaf detection using image processing techniques," in *Proceedings of the International Conference for Convergence of Technology-2014* (Piscataway, NJ: IEEE).

Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *Stud. Media Commun.*, SMC-3, 610–621.

Huang, M. J., and Mcbeath, J. (2010). Plant diseases, pests and food security. *Springer Netherlands* 35, 117–156. doi: 10.1007/978-1-4020-9180-3_5

Jin, X., Jie, L., Wang, S., Qi, H. J., and Li, S. W. (2018). Classifying wheat hyperspectral pixels of healthy heads and *Fusarium* head blight disease using a deep neural network in the wild field. *Remote Sens.* 10:395. doi: 10.3390/rs10030395

Jin, X., Liu, S., and Baret, F. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sens. Environ.* 198, 105–114. doi: 10.1016/j.rse.2017.06.007

Kira, K., and Rendell, L. A. (1992). "The feature selection problem: traditional methods and a new algorithm," in *Proceedings of the 10th National Conference on Artificial Intelligence*, Vol. 2, (San Jose, CA: AAAI Press), 129–134.

Kirkland, E. J. (2010). "Bilinear interpolation," in *Advanced Computing in Electron Microscopy* (Boston, MA: Springer). doi: 10.1007/978-1-4419-6533-2_12.

Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of 40th International Joint Conference on Artificial Intelligence, 1995* (Burlington, MA: Morgan Kaufmann Publishers Inc.).

Liu, Y., and Cui, R. (2015). Segmentation of winter wheat canopy image based on visual spectral and random forest algorithm. *Guang Pu Xue Yu Guang Pu Fen Xi* 35, 3480–3484.

Lu, Y., Yi, S., Zeng, N., Liu, Y., and Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. doi: 10.1016/j.neucom.2017.06.023

Marko, R., and Igor, K. (2003). Theoretical and empirical analysis of reliefF and RreliefF. *Mach. Learn.* 53, 23–69.

Meunkaewjinda, A., Kumsawat, P., Attakitmongcol, K., and Srikaew, A. (2008). "Grape leaf disease detection from color imagery using hybrid intelligent system," in *Proceedings of the 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, Krabi. doi: 10.1109/ecticon.2008.4600483

Mitra, P., Murthy, C. A., and Pal, S. (2002). Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 301–312. doi: 10.1109/34.990133

Mohd, S. K., Sabura, B. U., and Hemalatha, S. (2019). Anthracnose disease diagnosis by image processing, support vector machine and correlation with pigments. *J. Plant Pathol.* 101. doi: 10.1007/s42161-019-00268-9

Mostafa, M. G., Berrin, Y., and Erchan, A. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 235, 228–235. doi: 10.1016/j.neucom.2017.01.018

Palacios, S. A., Erazo, J. G., Ciasca, B., Lattanzio, V. M. T., Reynoso, M. M., Farnochi, M. C., et al. (2017). Occurrence of deoxynivalenol and deoxynivalenol-3-glucoside in durum wheat from Argentina. *Food Chem.* 230, 728–734. doi: 10.1016/j.foodchem.2017.03.085

Pantazi, X. Z., Moshou, D., and Tamouridou, A. A. (2019). Automated leaf disease detection in different crop species through image features analysis and one class classifiers. *Comput. Electron. Agric.* 156, 96–104. doi: 10.1016/j.compag.2018.11.005

Peiris, K. H. S., Dong, Y., Davis, M. A., Bockus, W. W., and Dowell, F. E. (2017). Estimation of the deoxynivalenol and moisture contents of bulk wheat grain samples by FT-NIR spectroscopy. *Cereal Chem. J.* 94, 677–682. doi: 10.1094/cchem-11-16-0271-r

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/tpami.2005.159

Pydipati, R., Burks, T. F., and Lee, W. S. (2006). Identification of citrus disease using color texture features and discriminant analysis. *Comput. Electron. Agric.* 52, 49–59. doi: 10.1016/j.compag.2006.01.004

Shervin, M., Amirali, A., and Yao, W. (2016). "An experimental study of deep convolutional features for iris recognition," in *Proceedings of IEEE Signal Processing in Medicine and Biology Symposium* Philadelphia, PA.

Sural, S., Qian, G., and Pramanik, S. (2002). "Segmentation and histogram generation using the HSV color space for image retrieval," in *Proceedings of the International Conference on Image Processing* (Piscataway, NJ: IEEE).

Valliammal, N., and Geethalakshmi, S. N. (2012). "Efficient feature fusion, selection and classification technique for plant leaf image retrieval system," in *Proceedings of the International Conference on Computational Science* (New York, NY: ACM).

Wang, X., and Paliwal, K. K. (2003). Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognit.* 36, 2429–2439. doi: 10.1016/s0031-3203(03)00044-x

Wang, Z., Zhang, Y., Chen, Z., Yang, H., Sun, Y., Kang, J., et al. (2016). "Application of reliefF algorithm to selecting feature sets for classification of high resolution remote sensing image," in *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (Beijing. doi: 10.1109/igarss.2016.7729190

Wei, T. J., Siow-Wee, C., and Sameem, B. A. K. (2018). "Deep learning for plant species classification using leaf vein morphometric," in *Proceedings of the IEEE/ACM Transactions on Computational Biology and Bioinformatics* (Piscataway, NJ: IEEE).

Xiao, M., Ma, Y., and Feng, Z. (2018). Rice blast recognition based on principal component analysis and neural network. *Comput. Electron. Agric.* 154, 482–490. doi: 10.1016/j.compag.2018.08.028

Xie, C., and He, Y. (2016). Spectrum and image texture features analysis for early blight disease detection on eggplant leaves. *Sensors* 16:676. doi: 10.3390/s16050676

Yann, L., Leon, B., Yoshua, B., and Patrick, H. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Yao, Q., Guan, Z., Zhou, Y., Tang, J., Hu, Y., and Yang, B. (2009). "Application of support vector machine for detecting rice diseases using shape and color texture features," in *Proceedings of the 2009 International Conference on Engineering Computation*, Hong Kong. doi: 10.1109/icec.2009.73

Yu, C., Jin, B., Lu, Y., Chen, X., Yi, Z., Kai, Z., et al. (2013). "Multi-threshold image segmentation based on firefly algorithm," in *Proceedings of the 2013 9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 16–18 October*, Beijing.

Yuan, Z., and Zhang, Y. (2000). Pesticide and environment. *Shanghai Chemcai Ind.* 17, 4–5. (in Chinese with English Abstract),

Zahra, S., and Davud, A. (2015). "Designing a classifier for automatic detection of fungal diseases in wheat plant: by pattern recognition techniques," in *Proceedings of the 23rd Iranian Conference on Electrical Engineering* (Piscataway, NJ: IEEE).

Zhao, J., Yuan, L., and Huang, L. (2013). "Characterization of powdery mildew in winter wheat using multi-angular hyperspectral measurements," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium* (Piscataway, NJ: IEEE).

Zhu, L., Wu, M., and Wan, X. (2017). Image recognition of rapeseed pests based on random forest classifier. *Int. J. Inf. Tech. Web Eng.* 12, 1–10. doi: 10.4018/ijitwe.2017070101

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Gu, Wang, Zhang, Zhang, Zhang and Liang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 January 2021
doi: 10.3389/fpls.2020.599705

Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation

Jennifer Hobbs[1*†], *Prajwal Prakash*[1,2†], *Robert Paull*[3], *Harutyun Hovhannisyan*[1], *Bernard Markowicz*[1] *and Greg Rose*[1]

[1] IntelinAir, Inc., Champaign, IL, United States, [2] Department of Electrical Engineering, Columbia University, New York, NY, United States, [3] Department of Tropical Plant and Soil Sciences, University of Hawaii at Manoa, Honolulu, HI, United States

Natural flowering affects fruit development and quality, and impacts the harvest of specialty plants like pineapple. Pineapple growers use chemicals to induce flowering so that most plants within a field produce fruit of high quality that is ready to harvest at the same time. Since pineapple is hand-harvested, the ability to harvest all of the fruit of a field in a single pass is critical to reduce field losses, costs, and waste, and to maximize efficiency. Traditionally, due to high planting densities, pineapple growers have been limited to gathering crop intelligence through manual inspection around the edges of the field, giving them only a limited view of their crop's status. Through the advances in remote sensing and computer vision, we can enable the regular inspection of the field and automated inflorescence counting enabling growers to optimize their management practices. Our work uses a deep learning-based density estimation approach to count the number of flowering pineapple plants in a field with a test MAE of 11.5 and MAPD of 6.37%. Notably, the computational complexity of this method does not depend on the number of plants present and therefore efficiently scale to easily detect over a 1.6 million flowering plants in a field. We further embed this approach in an active learning framework for continual learning and model improvement.

Keywords: deep learning-artificial neural network (DL-ANN), active learning, pineapple, computer vision, remote sensing-GIS, weakly supervised, counting, density estimation

OPEN ACCESS

Edited by:
Reza Ehsani,
University of California, Merced, United States

Reviewed by:
Luis Gonzaga Santesteban, Public University of Navarre, Spain
Yong He, Chinese Academy of Agricultural Sciences, China

****Correspondence:***
Jennifer Hobbs
jennifer+research@intelinair.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *27 August 2020*
Accepted: *09 December 2020*
Published: *28 January 2021*

Citation:
Hobbs J, Prakash P, Paull R, Hovhannisyan H, Markowicz B and Rose G (2021) Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation. Front. Plant Sci. 11:599705. doi: 10.3389/fpls.2020.599705

1. INTRODUCTION

Specialty crops, such as pineapple (*Ananas comosus* L.), present unique challenges and require sophisticated approaches to maximize productivity. Growers of large area crops such as corn or soybean have access to GPS-based yield maps and precisely apply inputs such as fertilizer and water considering field variability. Specialty crop growers lack access to these data as their crops tend to be hand-harvested. Because of this, specialty growers have been at a disadvantage, having to make decisions without this level of insight.

Growers of these high-value crops make a number of key decisions in every growing cycle. For pineapple, data supporting these decisions are generally limited to visual ground observations. But these observations are from the periphery where spatial and temporal variability, stage of growth, and development cannot be determined or quantified across the entire field. Walking through the

field is difficult as the plants grow very close together: 30,000 plants per acre (**Figure 1**). This lack of complete, real-time information about field conditions can lead to poor decisions resulting in too little or too much water, fertilization, pesticides, and growth regulators, or poor planning and scheduling of planting and harvest resources, including equipment and labor.

The natural flowering of pineapple affects fruit development and quality, and impacts harvest (Py et al., 1987; Bartholomew et al., 2002; Zhang and Kovacs, 2012; Sanewski et al., 2018). Pineapple growers use chemicals which produce ethylene (Ethephon) to induce flowering so that most plants within a field produce fruit of high quality ready to harvest at about the same time (Paull and Duarte, 2011; Bartholomew, 2013). The ideal situation would be for a grower to harvest the entire field in one pass when there is little variation in flowering, significantly increasing productivity and eliminating the cost of additional harvests.

Advances in aerial imagery collection (e.g., drones, UAVs) and remote sensing allow the grower insight into his field that was previously unattainable (Jung-Rothenhaeusler et al., 2014). UAVs have been deployed in large scale pineapple operations to reduce erosion and manage crop fertilization programs (Jung-Rothenhaeusler et al., 2014). However, their application to other aspects of managing pineapple production such as counting and identifying flowering pineapple plants from such imagery remains challenging because: 1. Pineapple inflorescence dramatically change in appearance (both size and color) as they develop and mature (Bartholomew et al., 2002; Zhang H. et al., 2016) 2. The global appearance of fields varies significantly due to lighting, shadowing, and other illumination differences. 3. A single field may have 1–2 million plants; methods where computational efficiency scales with the number of entities would be prohibitive at scale.

Our work leverages the advances of deep learning to automatically count and localize flowering pineapple plants, which may be in the millions for a single field (**Figure 2**). We use a counting-by-density-estimation approach to produce a density map of pineapple inflorescence across the field. This approach determines the density distribution of fruits across all regions of the field and identifies areas which are ready for harvest or delayed in development. Our approach produces results occasionally better than the human annotations.

Additionally, we embed this density-estimation framework in an active learning paradigm. After the density-inference is complete for a new image, we extract discrete locations of each inflorescence using a peak finding algorithm. These points are sent to human annotators for corrections and the model is retrained on the new data; this enables the model performance to improve as it sees more and more data while reducing the burden on human annotators. While active learning has been

FIGURE 1 | (A) A ground-level view of a pineapple field shows the large number and high density of plants (25,000–30,000 plants per acre) which makes inspecting the interiors of the blocks difficult. **(B)** An oblique view of a pineapple plant at the mid-flowering stage and early fruit growth stage tucked away among the leaves. **(C)** The top-down view shows the characteristic red center of a plant at an early stage of flowering called the "red-bud early cone" stage.

FIGURE 2 | Our model identifies the density of flowering across multiple blocks of a pineapple field. The flowering density is depicted on a spectrum from low (yellow) to high (red) and regions of no-flowering are shown as transparent (i.e., green from imagery shows through). A single, example field of 61.5 acres has over 1.6 million plants. This field has blocks at all stages of flowering and early fruit development. On the right hand side of the field, the plants are at the early stages of flowering with many still vegetative. The grid-like pattern observed across the image corresponds to the access roads (16 feet wide) surrounding each block with each block being 126 feet wide and of varying length; the model has accurately identified these as non-flowering areas. Several of the blocks are seen to have lower flowering density as they were just beginning to flower and still largely vegetative as confirmed in ground-inspections by horticulturalists.

previously applied to plant counting tasks in a counting-by-detection paradigm (Ghosal et al., 2019), our novel approach extracts discrete peaks that can be corrected by annotators while maintaining the computational advantages of the counting-by-density-estimation approach.

Finally, we demonstrate the usefulness and qualitative performance of our approach through field inspections. We see that the algorithm performs well across all stages of flowering, even though the appearance of the inflorescence in each stage varies. Our algorithm successfully identifies areas of stunted flowering occurring naturally or due to other circumstances (e.g., irrigation, fertilization, spraying for flower induction). The inspections also showed that there was about 1.4% plants missing in a row and about 12% of plants had fruit that were small, on short fruit stems or covered by leaves from adjoining plants and hence not easily discernible from above.

2. RELATED WORK

2.1. Counting Methods

Work in the area of dense-crowd-counting has inspired much of our current work (Loy et al., 2013; Sindagi and Patel, 2018). Within the broader domain of (entity) counting, approaches fall under one of three categories: counting by detection, counting by regression, and counting by density estimation (Sindagi and Patel, 2018).

Detection-based approaches are most applicable when the entities are large and well-separated, occlusions are limited, and the number of entities is small. These may take the form of sliding-window approaches that detect all or part of the entity in question (Li et al., 2008; Dollar et al., 2011) and sum the detections over the entire image. With the success of deep learning, many of these traditional approaches have been replaced with neural network-based detection and segmentation algorithms (Ren et al., 2015; Redmon et al., 2016; He et al., 2017), but these new methods still seek to solve the counting problem through the precise localization of all desired entities in the image. Key drawbacks to these methods are they tend to be computationally heavy, the time complexity often scales with the number of entities present, they often have an upper-limit of detectable entities before encountering memory issues, and they tend to struggle as occlusion becomes more pronounced or the entities become small. Additionally, detection methods, like Faster R-CNN (Ren et al., 2015) and YOLO (Redmon et al., 2016) require bounding box annotations and Mask R-CNN (He et al., 2017) further requires dense instance

mask annotations, all of which are extremely time consuming to acquire.

In contrast, counting by regression approaches eliminate the need to determine locations of each entity and seek only to determine the number of entities present (Chan and Vasconcelos, 2009; Ryan et al., 2009; Chen et al., 2012); these approaches also have benefited tremendously from deep learning-based architectures (Wang et al., 2015). However, these methods when used on their own provide only the total count, without any information as to how the entities are distributed across the image.

Density estimation approaches have proven very successful (Lempitsky and Zisserman, 2010; Pham et al., 2015; Xu and Qiu, 2016) especially when combined with deep architectures (Boominathan et al., 2016; Onoro-Rubio and López-Sastre, 2016; Zhang Y. et al., 2016; Sam et al., 2017, 2019) when we desire localization in addition to a final count. Many of these leverage fully convolutional neural networks (FCNs) to predict a density (Xie et al., 2018; Ma et al., 2019) across the image; this density can be integrated to provide the count over a region. Furthermore, these methods tend to outperform detection-based methods in highly occluded scenarios. They also require only simple point-annotations which can be acquired far more quickly than the bounding-box or instance-mask annotations needed by detection methods. Additionally, because the output density map is itself a single-channel image, not a collection of bounding boxes, the computational complexity is independent of the number of entities present. Our approach follows these methods as inflorescence may be occluded by other portions of the plant, and the number of inflorescence in a given image could be extremely large.

2.2. Active Learning

Deep learning approaches require a large amount of labeled data to maximize their performance and therefore a significant demand can be put on human annotators to gather such data. To offset these demands, significant work has been done in weakly, semi, and self-supervised learning approaches (Rosenberg et al., 2005; Zhu, 2005; Ratner et al., 2019; Xie et al., 2020). Most relevant to the present work are the weakly supervised approaches that incrementally train a model on a selection of data, correct any erroneous predictions using a human annotator or "oracle," and then retrain the model on the larger set of correctly annotated data (Li et al., 2013; Zhou et al., 2016). Active Learning is a subset of this domain which further explores the optimal selection of data for training (Settles et al., 2008; Settles, 2009; Huang et al., 2010). Many of these approaches rely on finding disagreement sets between different models trained for the same task (Dagan and Engelson, 1995; McCallum and Nigam, 1998) while others seek to find regions of uncertainty directly (Cohn et al., 1994) in the input space. The goal of our work is not around proposing a new or better query strategy, but to demonstrate how an active learning approach can improve results and reduce annotation cost in this domain.

2.3. Applications in Agriculture

Both traditional computer vision and deep learning-based approaches have been used for a variety of counting-based agricultural applications. The work of Guo et al. (2018) and Malambo et al. (2019) used detection-based techniques to detect sorghum heads in a field. Similarly, Gené-Mola et al. (2020) used Mask-RCNN to fully identify and segment apples on trees in an orchard. To count palm trees from UAV imagery, Li et al. (2017) used a CNN-based detection approach. Very recently, Osco et al. (2020) used an approach very similar to ours to count the number of citrus trees in a grove. Where they sought to count every tree present, in our work we seek to count only those plants who have begun to flower.

Particularly related to our work is Ghosal et al. (2019) who used a RetinaNet-based approach (Lin et al., 2017) to simultaneously regress the total count and individual bounding boxes of sorghum heads. This network was embedded into their "automated annotation protocol" (i.e. active learning system; Settles, 2009). We similarly embed our network into an active learning paradigm to enable continual learning. However, our counting approach is based on density-estimation approaches and does not rely on bounding box detections as in the above work.

3. MATERIALS AND METHODS

3.1. Data

We acquired raw imagery via a DJI Matrice 210 drone equipped with a DJI X3 three band (RGB) camera flown at a height of 200 ft above the pineapple fields (**Figure 3**). Individual images were stitched together using a third party system (Pix4Dmapper) to produce a single large-scale image for each block. During the stitching process, orthorectification is performed using the RGB image and a digital elevation model (DEM) of the field (Gao et al., 2009; Laliberte et al., 2010).

From this full dataset we randomly sampled 866 patches (512 × 512) across flights over 12 blocks from three fields for annotations. Annotators marked the center of each inflorescence with a point-label, producing 76,659 total point annotations. The data was split such that 650 patches for training and 130 patches for validation were sampled from multiple blocks belonging to an initial set of fields and 106 patches for testing were sampled from blocks belonging to an entirely different set of fields. That is, no field which appeared in the test set appeared in either the training or validation sets.

For training, we performed the following augmentation steps: the original sample (and label) was rotated by a random angle and randomly cropped to 256 × 256. For testing and validation, the original 512 × 512 patches were split into four non-overlapping 256 × 256 images.

3.2. Density Estimation

To produce the target density map, the point labels generated by annotation were blurred using a two-dimensional isotropic Gaussian filter. That is, given an image I with pixels $\boldsymbol{x}_m$ annotated with points $\boldsymbol{z}_n = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\} \, | \boldsymbol{z}_i \in \mathbf{R}^2$ where N is the total

FIGURE 3 | A photo of the drone preparing to begin its imagery collection flight over the pineapple fields, with the Pacific Ocean in the distance on the north side of Oahu, Hawaii.

number of points annotated in that image, we define the *ground truth* density map $\boldsymbol{D}$ to be a kernel density estimate given by:

$$D(\boldsymbol{x}_m) \stackrel{\text{def}}{=} \sum_{n=1}^{N} \mathcal{N}(\boldsymbol{x}_m; \boldsymbol{z}_n, \sigma^2 \mathbf{1}_{2\times 2}) \tag{1}$$

$$= \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\boldsymbol{x}_m - \boldsymbol{z}_n\|_2^2}{2\sigma^2}\right) \tag{2}$$

We explored values in $[1, 2, 6, 10, 20]$ for σ, the standard deviation of the Gaussian kernel, and found that $\sigma = 6$ provided the best results both in terms of MSE as well as steps needed for convergence.

We used the mean squared error (MSE) between the target and predicted density maps $\hat{\boldsymbol{D}}(\boldsymbol{x}_m)$ as our loss function and is given by

$$MSE = \frac{1}{M} \sum_{k=1}^{M} \|\boldsymbol{D}(\boldsymbol{x}_m) - \hat{\boldsymbol{D}}(\boldsymbol{x}_m)\|_2^2. \tag{3}$$

Adam Optimizer was used with a learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.99$, and weight decay of $1e^{-5}$. The model was trained with a batch size of 20 on a machine equipped with an NVIDIA Titan RTX for up to 1,000 epochs; the final model was halted using early stopping after 30 steps. In early work, we used a machine with a Tesla P4 with a batch size of 10. The optimal model on this hardware was not reached until 630 epochs (which is why the maximum allowed epochs was set to 1,000) and did not yield as good of results as the final model trained on the Titan RTX.

Our model used the fully-convolutional encoder-decoder structure of U-net (Ronneberger et al., 2015), taking in the 3 (RGB) input channels and producing a single-channel output corresponding to the inforescence density (**Figure 4**). Each convolutional block consisted of 3x3 convolution followed by batch normalization (Ioffe and Szegedy, 2015) and a ReLU nonlinearity. Max Pooling with a 2×2 kernel with a stride of 2 was used in the encoder after every two convolutional blocks. In the decoder, we used a 2×2 transposed convolution for upsampling. We used *same* padding throughout.

The final layer consists of a 1D convolution followed by ReLU activation: this ensures that every point in output layer is positive, which is required by our density prediction task. Note, that the output density is *not* required to be $[0, 1]$, but only positive; if many inflorescences are located closely together, their densities

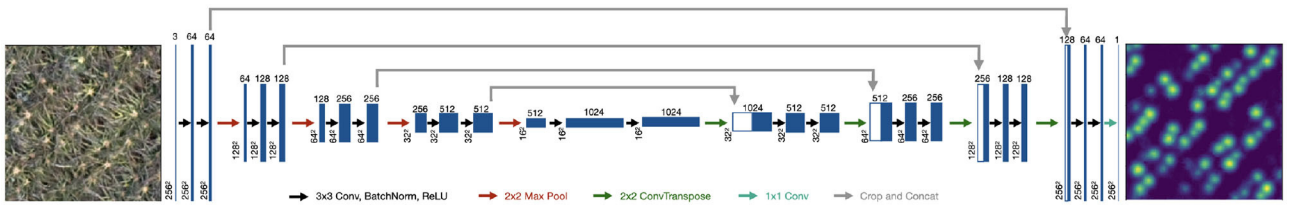

FIGURE 4 | Our architecture follows the encoder-decoder structure of U-net where the input is an RGB image and the output is a density map.

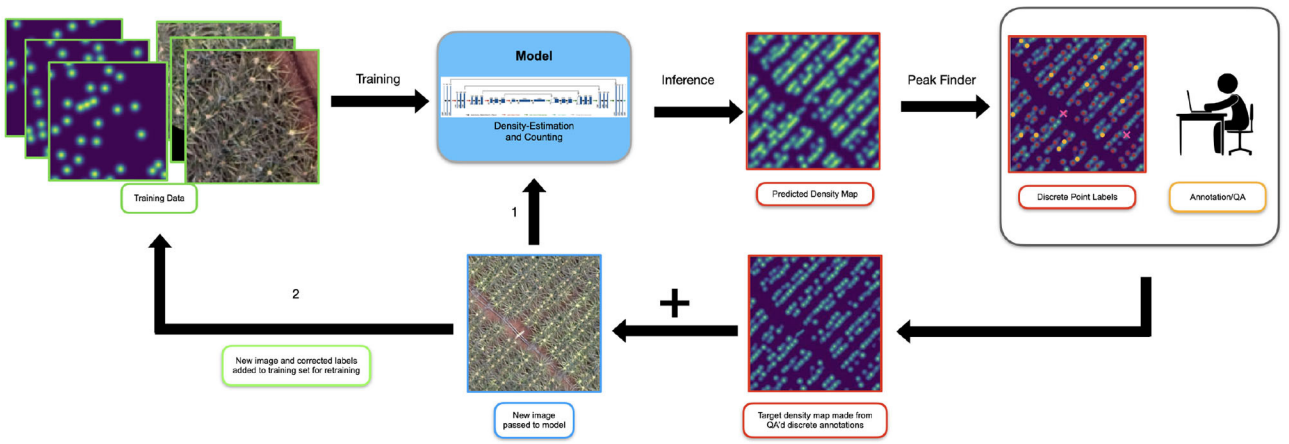

FIGURE 5 | The setup for our active learning system. Initially the model is trained on images which have been annotated from scratch. When new data is initially presented to the model (1), it is passed directly in for inference. The model generates the predicted density map and the peak finding algorithm is used to extract discrete locations of inflorescences. A human annotator or "oracle" reviews and corrects the discrete points and a target density map is created for the new sample. Note, the "predicted" points shown here has been made artificially poor to emphasize the actions of the annotator. This map and the original image are added to the training set (2) and the model is retrained.

could add to >1 in some places. In practice, we did not see this occur and therefore a final sigmoid activation could be used in place of the ReLU to enforce a range of [0, 1]. However, we found that the final ReLU activation outperforms these alternatives.

3.2.1. Total Count

The output of the U-net is a single channel density map of the flowering plants across the field. To get the total count of inflorescence $\hat{T}_c$ in a particular region, in this case the sample window, we integrated over the density map to produce the final count. That is, $\hat{T}_c = \int \boldsymbol{D}(\boldsymbol{x}_m)dx$. Note that dx corresponds to the spatial window captured by a single pixel and therefore in practice this equations to taking the sum of the prediction matrix.

3.3. Weak Supervision and Active Learning

3.3.1. Weakly Supervised Annotation Framework

We used a weakly supervised approach to continually feed more (annotated) data to the model. **Figure 5** shows an overview of this approach.

As discussed previously, the model was trained on an initial training dataset. When new (unseen) data becomes available, it is passed directly to the model for inference; the U-net produces a predicted density map for that image.

Commonly in an active learning paradigm, the model output is cleaned up directly by human annotators and fed back into the model for retraining. However, cleaning up the density map directly is a challenging annotation task because the "location" of the inflorescence is non-uniformly spread over a set of pixels. Where inflorescence distributions overlap, it is unclear how the density map should be appropriately altered; inconsistency and ambiguity from the annotators would degrade, not enhance model performance.

To overcome this, we developed a procedure to extract discrete locations of points that can be submitted for re-annotation from this final density map. We first threshold the image so regions of low density, below γ, are removed. Next, we use a 2D local-max finding algorithm common to most image processing toolkits to identify peaks requiring a minimum distance of δ between peaks. We found that $\gamma = 0.05$ and $\delta = 4$ work well in practice although these values can be dynamically changed in the annotation interface to best support the annotation process. Note that because of the filtering applied during this process, the sum over these peaks $\hat{T}_d$ will *always* be *less* than the overall predicted count obtained by integrating over the predicted density map $\hat{T}_c$. This is not problematic and in fact, we found anecdotally that annotators (aka. the "oracle") seem to be more efficient and accurate at adding missed detections as opposed to deleting false positives.

Next the set of discrete point annotations, after having been corrected by the oracle, is smoothed with the same Gaussian filter used on the initial data to create a new target (i.e., ground-truth) density map label. This new label along with the image is added to the training set for retraining. The validation set is left unchanged.

Retraining occurs whenever a "sufficient" amount of new annotated data is acquired: sufficiency is usually determined by operational constraints such as cost or compute time. After the model is retrained, if it outperforms the previous model on the validation set, it is promoted to the current version and used for subsequent inference. This process is repeated as desired.

3.3.2. Active Learning

In the passive weakly supervised approach, new samples are fed to the model randomly. However, we also seek to minimize annotator burden and maximize the efficiency of the model training process by prioritizing the most "useful" and informative samples for annotation and retraining. Therefore in the Active Learning approach, we prioritize samples in the following manner:

1. The total count $\hat{T}_c$ for a given (new) sample is computed by integrating over the predicted density map.
2. The peak-finding algorithm is applied to identify discrete locations of flowers. The number of discrete points is $\hat{T}_d$.
3. The absolute difference $CountDiff = (\hat{T}_c - \hat{T}_d)$ is computed.
4. Samples are ranked according to *CountDiff* and the samples with the greatest differences are prioritized for annotation and retraining.

3.3.3. Impact of Data Quantity and Learning Strategy

In practice, new data will be passed to the model during passive and active learning. However, to quantify the impact of more data on the model performance which the active learning system affords, we conducted an experiment in which we incrementally trained the original model on growing amounts of the original training set.

In the following experiments, the validation and test sets were identical to before. Only the subset of training data which the model was shown at each step was varied. For clarity, we denote the set of training data which was not currently being used at that step of training "the training (data) pool."

The model was initially trained on a 50 samples of training data and validated against the full validation set. Inference was run on the test set and the performance was recorded. Additional samples were selected from the training pool and added to the initial 50 samples according to the following procedure:

1. Inference was run on the test set to record performance for that amount of training data.
2. Inference was run against the training data pool.
3. *CountDiff* was computed for all samples in the training data pool.
4. Those samples with the largest value were added to the training set for the next round of training.
5. The model was retrained.
6. These steps were repeated until all data from the training data pool had been added to the training set.

We added data and retrained at levels of [50, 100, 250, 500, 650] samples. Results are shown in section 4.2.

3.4. Ground Inspections

To provide ground-level verification of the model's output and to demonstrate how this application could potentially be incorporated into one's management practices, we inferenced and conducted ground inspections of a block.

After model training, validation, and testing was complete, we ran inference on a completely unseen block; this block belonged to the same field and was under the same management as those areas used for training-validation-testing, but was not previously shown to the model. In particular, pineapples in the field were induced to flower when the plants were large enough by spraying with a chemical (Ethephon) that breaks down to release ethylene; ethylene is the natural inducer of flowering in Bromeliads of which pineapple is a member. A density map for the entire block was constructed to enable clear visualization of the distribution of inflorescences across the block and easy identification of any areas which may be exhibiting stunted development or early inflorescence.

Three horticulturalists familiar with pineapple flowering evaluated inflorescence in that block. Inflorescence number and their visibility were counted in a 50 feet bed that has two rows of pineapple plants; this evaluation was repeated four times. Qualitative evaluation was carried out by walking around the field block's perimeter and estimating the stage of flowering as red bud, early or late cone and early mid, late flowering and dry petal stage or early fruit development of each block in a field. Red Bud is the first noticeable stage of inflorescence development, with the cone stage being the later stage of inflorescence development before flowers begin to open from the base of the inflorescence cone.

4. RESULTS

4.1. Density Estimation

4.1.1. Model Performance

Results from our approach are shown in **Figure 6**. The per-pixel MSE validation loss was 0.0033 and the test loss was 0.0038. Qualitatively we see the predicted density maps closely resemble the target maps. In certain cases, particularly when the inflorescences are redder in appearance (corresponding to earlier stages of flowering), the outputs of the model occasionally appear more correct than the initial human annotations.

4.1.2. Total Count

Integration of the predicted density maps over the entire image provides us with a prediction of the total number of inflorescence. For each original image we compared the actual number of flowers to the number predicted by the model as seen in **Figure 7**. Because the U-net is a fully convolutional network, it is amenable to figures of variable sizes so long as the pooling operations result in integer dimensions. So for this analysis, we inferenced the original 512 × 512 images without any augmentation (i.e., rotation or cropping) in the training,

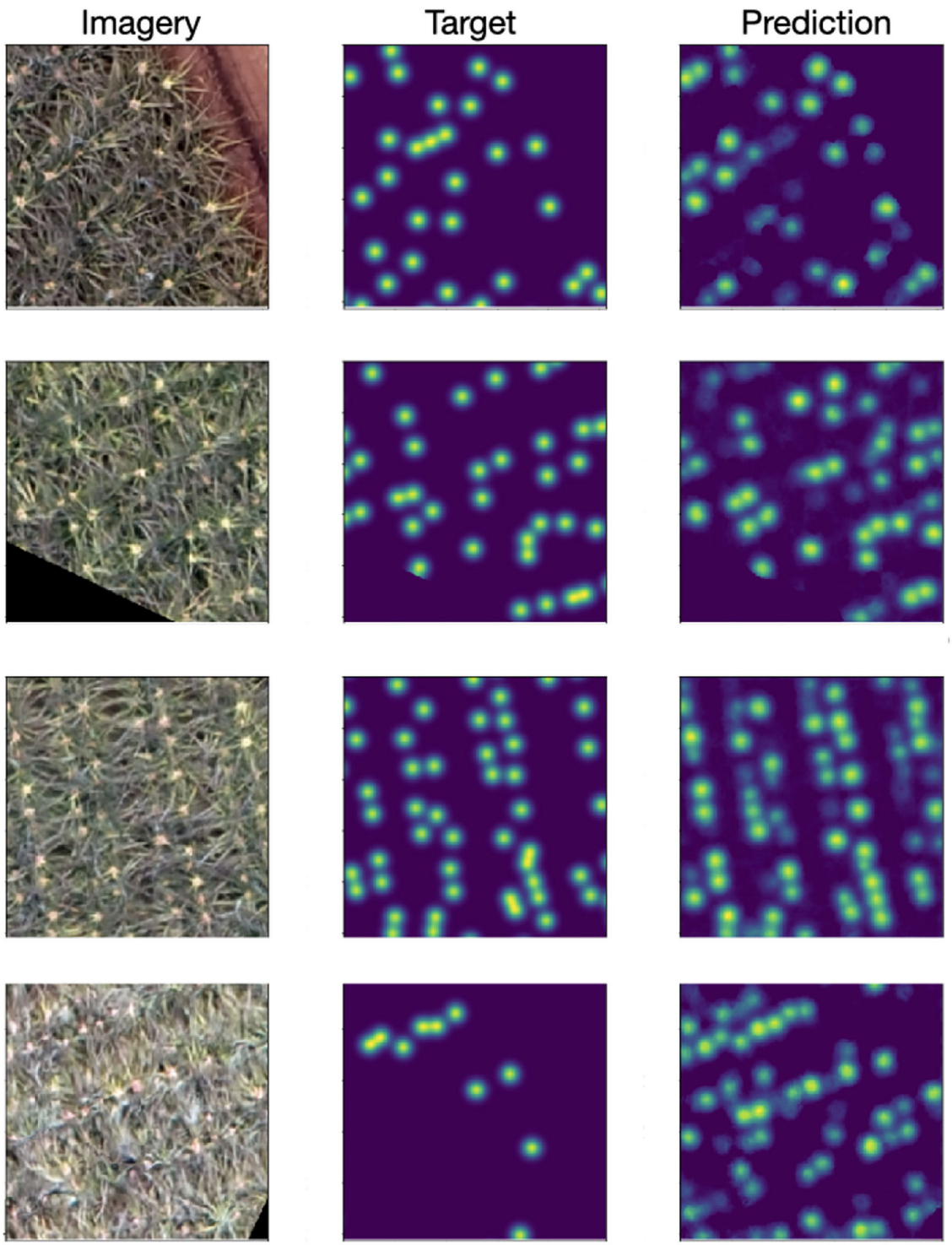

FIGURE 6 | (Left) Input RGB image. **(Middle)** Target density maps generated from the human point annotations and smoothed with a Gaussian kernel with $\sigma = 6$. **(Right)** Predicted density map. Particularly when the inflorescence are less well defined, the model can be seen to outperform the human annotations (bottom row).

validation, and test sets. We see that in all three splits, the data falls close to the x=y line with a mean absolute error (MAE) of 11.5 and mean absolute percent deviation (MAPD) of 6.37% on the test set.

4.1.3. Computational Efficiency

The computational efficiency of this approach offers key advantages. Inference speed is 0.04 sec/sample on a single P4 GPU and under 0.0039 sec/sample on a single Titan RTX.

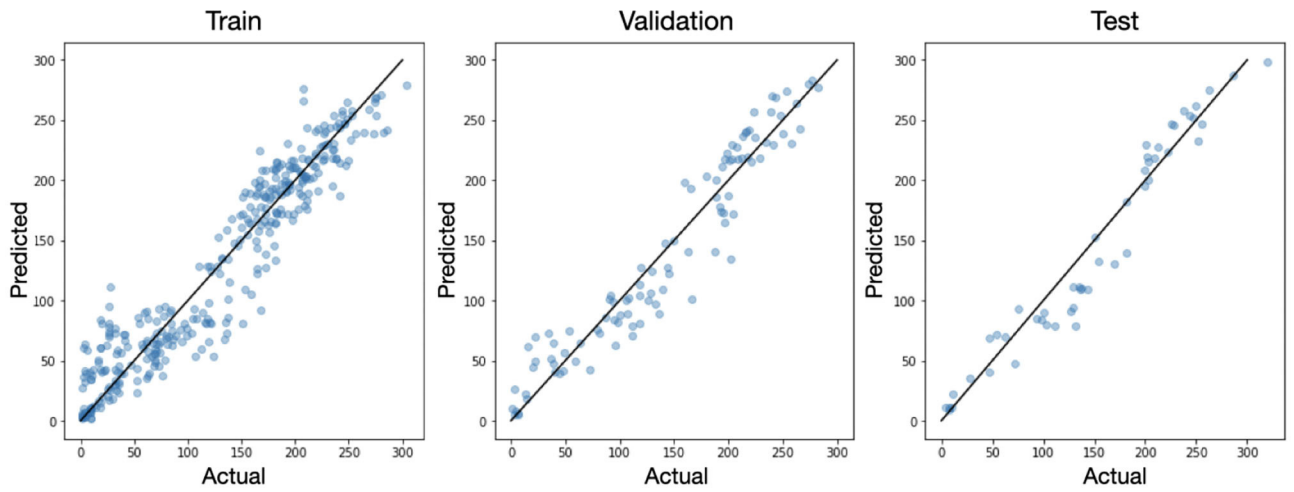

FIGURE 7 | The actual vs. predicted number of inflorescence shown for each sample in the training, validation, and test sets. The black line corresponds to x=y.

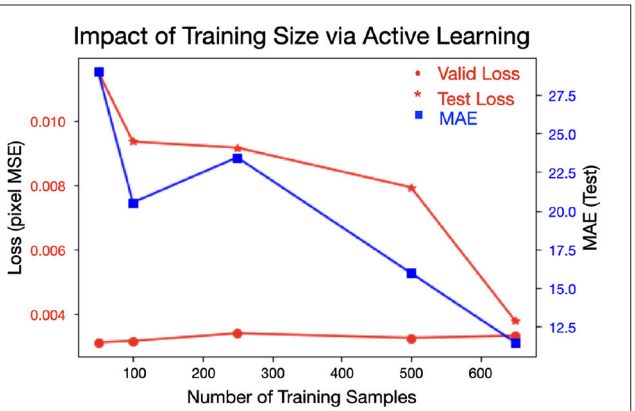

FIGURE 8 | Increasing the amount of (labeled) training data in a smart fashion decreases test loss as well as the MAE on the test set. The validation loss slightly increases as more data is added, suggesting less over-fitting is occurring as more data is added.

Especially with appropriate compilation steps which would even further increase efficiency, this speed would enable the model to be run in real-time, potentially allowing for on-the-fly decision making.

4.2. Impact of Data Quantity on Performance

Figure 8 shows the impact of enlarging the dataset via our active learning approach. Recall that the validation set was in the same domain as the training set (i.e., different samples from the same fields) while the test set was out-of-domain (i.e., samples from a completely unseen field). We see that as more data was added, the test loss (red stars) decreased, as we hoped. Additionally, the MAE on the test set (generally) improved. The validation loss slightly increased, but not significantly. This may suggest that as the quantity of data is increased, the model is less likely to (over)fit to the in-domain samples of the training and validation sets, while the generalizability (as seen in the test performance) improves.

4.3. Qualitative Analysis and Ground Inspections

Using the final supervised model, we ran inference on a new field (independent of the train, validation, or test set) to generate its density map. Horticulturists then inspected the field, particularly focusing on areas which the model deemed to be low-density.

Figure 9 shows the inspected blocks, predicted density map, and several ground-level images taken during inspection. The density map (**Figure 9A**) draws your attention to key areas on the field. Interesting features of the plot such as irrigation and drainage lines become readily apparent due to the absence of inflorescence. Other areas of low density are also visible. **Figure 9B** exhibited average inflorescence as predicted by the model and confirmed by the horticulturalists. In **Figure 9C**, failed forcing was evident in two beds in the middle of the block. We saw that while most rows had successfully flowered at almost 98% fruiting, a single bed down almost half the length of the middle of the block showed poor forcing at only 62%. This lack of flowering was possibly due to either a blocked sprayer nozzle or incomplete overlap between the sprayer arms. Automatically and immediately identifying issues caused by equipment provides tremendous value to the grower so the issue does not become present in other regions of the field.

Location (**Figure 9D**) was predicted to have a below-average flowering rate, however, inspection showed that the rate was comparable to surrounding areas. Plants in this area were shorter than average, resulting in the early fruit being more easily obscured. Very evident from the model's output was the area indicated at **Figure 9E**, which appeared to be completely lacking inflorescence. Field inspection confirms this was indeed true: a sizable portion of this block had remained vegetative and failed to flower in this triangular area. A complete absence of flowering in a pattern like this at the end of this block as likely due to the

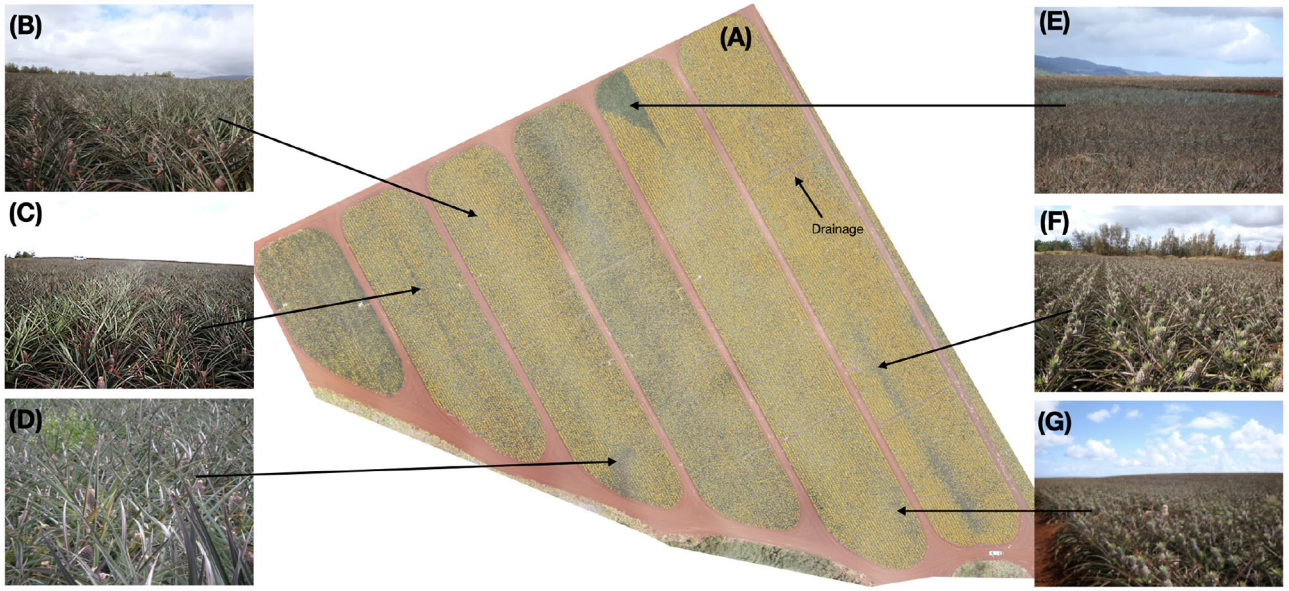

FIGURE 9 | (A) Inference was run on a previously unseen block and the density map is overlaid on the imagery (transparent: none, low: yellow, red: high). Variation across the block is clearly visible. In some cases, these come from known field-issues such as drainage or irrigation lines. Others, however, require inspection from horticulturist to determine the source. **(B)** This region of the field has mostly normal flowering. **(C)** Failed forcing is seen in the middle bed of this block and extends nearly half the length in the block where the amount of flowering (62% flowering) in that bed of two rows is lower than the surrounding beds (98% flowering). **(D)** The model predicted slightly lower-than-average density in this region, however, inspection showed the ground-truth density was normal. Plants in this area appear to be shorter than average; this can cause the young fruit to be more easily obscured from the imagery. **(E)** A vegetative region of the field is clearly visible from the density map. **(F,G)** Mosaicking issues resulting in fuzzy imagery resulted in the model predicting lower-than-average density in these areas when ground-truth inspection revealed otherwise. Best viewed electronically.

spray rig running out of chemical as it approached the end of its run.

Horticulturists inspected the areas (**Figures 9F,G**) which the model predicted as low density. Ground inspection indicated that this region is in fact flowered more than the model predicted. Examination of the original imagery shows that this region of the image was blurry, likely due to an issue during mosaicking, resulting in an artificially low prediction from the model. This will be explored further in section 5.

5. DISCUSSION

5.1. Impact on Specialty Crops

Modifying management practices with data on field conditions goes beyond reducing costs for the farmers. By identifying flowering plants at their earliest stages across entire fields, the application of chemicals can be precisely applied and limited in extent. By monitoring the progression of plant development across the field, harvest times can be optimized so that fruit are picked at their peak development, limiting waste and maximizing return. On-going work is considering the potential to predict marketable fruit and percentage of unharvested fruit because of small size; this possibility is supported by the variation in flowering densities predicted by our algorithm.

5.2. Active Learning and Uncertainty Sampling

In the present work we have embedded our model in an active learning framework to continually collect new annotations and repeatedly retrain the model for continual learning and improvement. While the capacity of neural networks is immense (Brown et al., 2020), training on an ever-growing amount of data can be computationally cumbersome and expensive. Therefore, it can be advantageous to (re)train the model only on the subset of data which is "challenging," that is, near the decision-boundary. This is the motivation behind our selection criteria for sample prioritization.

The focus of this work was not to determine the most optimal data selection process, but to identify *an* approach that could be used to reduce annotator burden and improve model performance. Here we have exploited a subtlety of the framework by noticing that "more difficult" examples tend to produce less well defined peaks that are more likely to be dropped during the peak finding step. Use of techniques such as uncertainty or adversarial sampling can be employed to identify data that should be inspected for annotation and fed back to the model for retraining (Žliobaitė et al., 2013; Mayer and Timofte, 2020). Even though we are far from having too much annotated data for the current model, the incorporation of hard example mining techniques like those mentioned above are still useful for prioritizing which samples the

annotators correct first; exploring these techniques is the focus on ongoing work.

5.3. Orthorectification and Mosaicking

All of the models in this work were trained on data from large, orthorectified, mosaicked images. Orthorectification is a central part of remote sensing analysis, particularly when involving agriculture, because it controls for the effects of image perspective and relief; agronomic indices based on ground reflectance values rely on these corrections. As such, traditional CV algorithms are largely dependent on the mosaicking and orthorectification process. However, deep learning approaches, like those used here, rely on learned, non-linear features involving shape/structure, and color. This enables them to be more robust to variations such as lighting/reflectance shifts and able to generalize to broader domains as opposed to relying on upstream algorithms to identify and/or control for these variations. Since mosaicking requires the program to identify key points for alignment, a very uniform field with high density planting, present challenges and may lead to blurriness in some assembled areas of a mosaic.

It is likely that the current model, trained on mosaicked-orthorectified images, would initially perform slightly less well on unseen, non-orthorectified imagery because that data is slightly out-of-domain. However, it is reasonable to believe that with minimal fine-tuning and retraining on such imagery, the model would perform equivalently well in the new domain; as humans, the task of identifying flowers from either sources is equivalent in difficulty and both tasks would be considered a "Type 1" process (Kahneman, 2011). Enabling inference directly on the raw imagery would cut out a time-consuming step of the processing pipeline and enable a wide range of live and on-device applications. As future work, we will examine the impact of working directly on raw images both from RGB and specific spectral bands and explore transfer-learning approaches to adapt the model to this new, but similar, domain.

Additionally, we saw in section 4.3 that the model performed less well on regions of the field which were fuzzy, potentially due to mosaicking issues. This is not surprising as degraded image quality would be expected to result in poorer performance. Nevertheless, as the inflorescence in this region are still discernible by humans from the fuzzy imagery, we believe that with additional annotation and training on degraded imagery, the model will be able to learn how to handle such sources of noise and generalize to a greater range of image quality.

5.4. Extension to Multiple Scales and Other Domains

All of the data here was flown at 200', producing images with similar statistical structure (i.e., all of the plants and inflorescence are roughly the same size). To make this algorithm broadly useful across many environments, we would like it to perform well across a variety of (reasonable) flight heights and resultant resolutions. Additionally, we would like to determine the minimum required resolution (i.e., maximum height flown) which delivers quality results; flying at a higher elevation would allow the data to be collected more rapidly.

Handling multiple scales is another place where deep learning shines over traditional computer vision algorithms. Flying at a given height allows the model to learn that inflorescence are

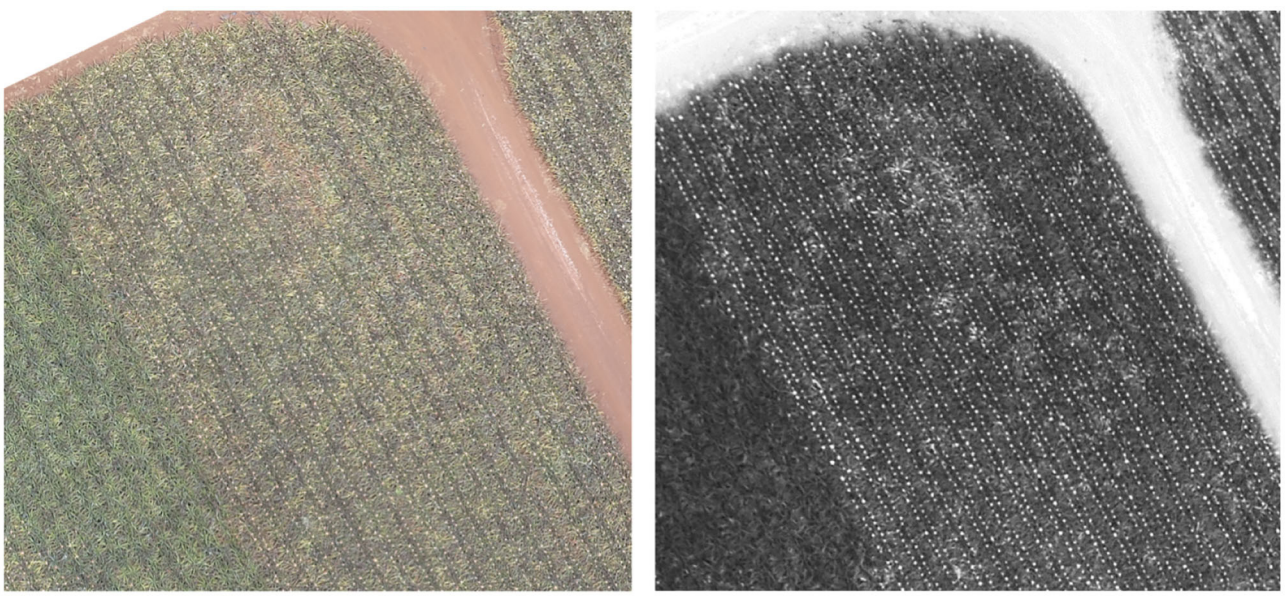

FIGURE 10 | An image of the field taken with RGB **(Left)** and Red-Edge **(Right)**. The inflorescence are readily apparent in the single-channel red-edge image, suggesting this would be a useful addition in future analysis. This block highlights the potential of our system as the leftmost portion of this block in both images is still vegetative with no flowering.

all roughly the same size; flying at multiple elevations would require the model to learn a more expansive filter-bank to identify inflorescence of widely varying sizes. Although, we anticipate this transfer task to be more challenging than the one from orthorectified to non-orthorectified, we are confident that the model could generalize to handle multiple scales because of the successes of deep learning approaches in the broader crowd counting space. Should the current model struggle to handle multiple scales, there are a number of scale and context-aware modifications we could make in the current framework which would address these challenges (Hossain et al., 2019; Liu et al., 2019). Multi-scale detection in this domain is the focus of future work.

Similarly, this analysis was conducted on blocks from a single field under the same management conditions. Deep learning approaches again provide us with the ability to more easily adapt to unseen domains such as different fields under different management. Because these approaches do not rely on handcrafted rules and features but instead learn the relevant features directly from the data, knowing these management practices or appearance differences *a prior* is not necessary. Given the success of other deep learning models to generalize with increasing data, we believe the current model will generalize over a wide range of appearances, seasons, and management practices, particularly as we continue to supply it with new data efficiently obtained under the active learning paradigm.

5.5. Real-Time Edge Deployment and Alerting

A key advantage of this approach is speed of inference and lightness of the model architecture; not only is the model fast, but its performance is constant and does not degrade as the number of detected entities increases. Because a single image can be processed in <0.01 s on a GPU, this opens the possibility for real-time deployment. While the current model is trained and inferenced on a GPU, it has not yet been compiled for target hardware through an optimized runtime like TensorRT[1], further increasing the inference speed. This would enable edge deployment: one could envision running the model live while a drone is collecting the imagery and providing alerts when encountering low-density flowering areas.

The alerting component that this model enables, either real-time on the edge or after batch-processing, also has key value to growers. While the aerial imagery itself provides the growers with novel information not accessible from manual ground inspection (see **Figure 10**, left), most growers are not interested in or compelled by the raw imagery alone. Instead, most prefer to have an intelligence layer that sits on top of the imagery and alerts them to regions under their management requiring attention. Once the density map is determined by this current application, it can be handed off to a second application which identifies regions of the field which are anomalous or problematic and automatically alerts the growers accordingly; this is the focus of ongoing work.

[1]https://developer.nvidia.com/tensorrt

5.6. Beyond RGB

Although not discussed in detail here, determining the right camera and flight height/resolution was an important step in the data acquisition process. The present analysis focuses only on RGB data as we were able to obtain very good results from the three-channel images. However, other channels may further improve model performance, stability, and generalization. **Figure 10** shows a region of the field [corresponding to the area in **Figure 9E**] taken in RGB (left) and with a Red-Edge (right). The inflorescence visually "pop" in the red-edge image and are easily identifiable. Therefore, incorporating collecting additional red-edge imagery and training the model on a four-channel input could be very beneficial. Future work will explore incorporating additional channels like the red-edge seen here.

6. CONCLUSION

We have developed a density-estimation deep learning model based on a U-net backbone that accurately detects flowering pineapple plants in a field. Because of the architectural decisions made, the model is fast, lightweight, and its computational efficiency is independent of the number of inflorescence detected, allowing us to rapidly detected over 1.6 million flowering plants in a field. Our model highlights areas on the field which are vegetative or demonstrate failed forcing; growers can be alerted to these areas which would otherwise go undetected. Finally, the model will continue to improve as more corrected annotations are fed back into the model for retraining through our active learning system.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because, data may not be used for commercial purposes. Requests to access the datasets should be directed to Jennifer Hobbs, jennifer+research@intelinair.com.

AUTHOR CONTRIBUTIONS

PP, HH, and JH were involved in the model development. RP, BM, GR, and JH were involved in the data acquisition, annotation, and ground-truthing. JH took the lead in preparing the manuscript. All authors provided critical feedback and analysis.

FUNDING

This material was based on work supported by the USDA SBIR Grant Award 2019-33610-29755. RP was also supported in part by the US Department of Agriculture, National Institute of Food and Agriculture, under an agreement 58-2040-5-010 through the Agriculture Research Service and Hatch Project H862.

ACKNOWLEDGMENTS

The authors greatly appreciate the collaboration with Dole Food Company Hawaii where this research was carried out. The technical help provided by Ms. Gail Uruu is also recognized. The expert drone piloting skills and professionalism of Mike Elliott at Drone Services Hawaii was critical for the success of this project.

REFERENCES

Bartholomew, D. (2013). "History and perspectives on the role of ethylene in pineapple flowering," in *XII International Symposium on Plant Bioregulators in Fruit Production 1042*, 269–284. doi: 10.17660/ActaHortic.2014.1042.33

Bartholomew, D., Rohrbach, K., and Evans, D. (2002). *Pineapple Cultivation in Hawaii*. University of Hawaii at Manoa; College of Tropical Agriculture and Human Resources; Cooperative Extension Service; Fruits and Nut Series F&N-7.

Boominathan, L., Kruthiventi, S. S., and Babu, R. V. (2016). "Crowdnet: a deep convolutional network for dense crowd counting," in *Proceedings of the 24th ACM International Conference on Multimedia*, 640–644. doi: 10.1145/2964284.2967300

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *arXiv [Preprint] arXiv:*2005.14165.

Chan, A. B., and Vasconcelos, N. (2009). "Bayesian Poisson regression for crowd counting," in *2009 IEEE 12th International Conference on Computer Vision*, 545–551. doi: 10.1109/ICCV.2009.5459191

Chen, K., Loy, C. C., Gong, S., and Xiang, T. (2012). Feature mining for localised crowd counting. *BMVC* 1:3. doi: 10.5244/C.26.21

Cohn, D., Atlas, L., and Ladner, R. (1994). Improving generalization with active learning. *Mach. Learn.* 15, 201–221. doi: 10.1007/BF00993277

Dagan, I., and Engelson, S. P. (1995). "Committee-based sampling for training probabilistic classifiers," in *Machine Learning Proceedings 1995* (Elsevier), 150–157. doi: 10.1016/B978-1-55860-377-6.50027-X

Dollar, P., Wojek, C., Schiele, B., and Perona, P. (2011). Pedestrian detection: an evaluation of the state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 743–761. doi: 10.1109/TPAMI.2011.155

Gao, F., Masek, J. G., and Wolfe, R. E. (2009). Automated registration and orthorectification package for landsat and landsat-like data processing. *J. Appl. Remote Sens.* 3:033515. doi: 10.1117/1.3104620

Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J. R., Morros, J.-R., Ruiz-Hidalgo, J., Vilaplana, V., et al. (2020). Fruit detection and 3d location using instance segmentation neural networks and structure-from-motion photogrammetry. *Comput. Electron. Agric.* 169:105165. doi: 10.1016/j.compag.2019.105165

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., et al. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenom.* 2019:1525874. doi: 10.34133/2019/1525874

Guo, W., Zheng, B., Potgieter, A. B., Diot, J., Watanabe, K., Noshita, K., et al. (2018). Aerial imagery analysis-quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Front. Plant Sci.* 9:1544. doi: 10.3389/fpls.2018.01544

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969. doi: 10.1109/ICCV.2017.322

Hossain, M., Hosseinzadeh, M., Chanda, O., and Wang, Y. (2019). "Crowd counting using scale-aware attention networks," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1280–1288. IEEE. doi: 10.1109/WACV.2019.00141

Huang, S.-J., Jin, R., and Zhou, Z.-H. (2010). "Active learning by querying informative and representative examples," in *Advances in Neural Information Processing Systems*, 892–900.

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv [Preprint] arXiv:*1502.03167.

Jung-Rothenhaeusler, F., Traut, K., Gauggel, C., Brooke Smith, A., and Umali, R. (2014). "Pineapple plantation information system-application of geo-informatics and unmanned aerial sensing technology for efficient land use management," in *XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014)*: *IV 1111*, 255–266. doi: 10.17660/ActaHortic.2016.1111.37

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Laliberte, A. S., Herrick, J. E., Rango, A., and Winters, C. (2010). Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring. *Photogr. Eng. Remote Sens.* 76, 661–672. doi: 10.14358/PERS.76.6.661

Lempitsky, V., and Zisserman, A. (2010). "Learning to count objects in images," in *Advances in Neural Information Processing Systems 23*, eds J. D. Lafferty, C. K. L. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (Curran Associates, Inc.), 1324–1332.

Li, M., Zhang, Z., Huang, K., and Tan, T. (2008). "Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection," in *2008 19th International Conference on Pattern Recognition* (IEEE), 1–4. doi: 10.1109/ICPR.2008.4761705

Li, W., Fu, H., Yu, L., and Cracknell, A. (2017). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* 9:22. doi: 10.3390/rs9010022

Li, X., Snoek, C. M., Worring, M., Koelma, D., and Smeulders, A. W. (2013). Bootstrapping visual categorization with relevant negatives. *IEEE Trans. Multimed.* 15, 933–945. doi: 10.1109/TMM.2013.2238523

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988. doi: 10.1109/ICCV.2017.324

Liu, W., Salzmann, M., and Fua, P. (2019). "Context-aware crowd counting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5099–5108. doi: 10.1109/CVPR.2019.00524

Loy, C. C., Chen, K., Gong, S., and Xiang, T. (2013). "Crowd counting and profiling: methodology and evaluation," in *Modeling, Simulation and Visual Analysis of Crowds* (Springer), 347–382. doi: 10.1007/978-1-4614-8483-7_14

Ma, Z., Wei, X., Hong, X., and Gong, Y. (2019). "Bayesian loss for crowd count estimation with point supervision," in *Proceedings of the IEEE International Conference on Computer Vision*, 6142–6151. doi: 10.1109/ICCV.2019.00624

Malambo, L., Popescu, S., Ku, N.-W., Rooney, W., Zhou, T., and Moore, S. (2019). A deep learning semantic segmentation-based approach for field-level sorghum panicle counting. *Remote Sens.* 11:2939. doi: 10.3390/rs11242939

Mayer, C., and Timofte, R. (2020). "Adversarial sampling for active learning," in *The IEEE Winter Conference on Applications of Computer Vision*, 3071–3079. doi: 10.1109/WACV45572.2020.9093556

McCallum, A. K., and Nigam, K. (1998). "Employing EM in pool-based active learning for text classification," in *Proceedings of 15th International Conference on Machine Learning, Madison*, 350–358.

Onoro-Rubio, D., and López-Sastre, R. J. (2016). "Towards perspective-free object counting with deep learning," in *European Conference on Computer Vision* (Springer), 615–629. doi: 10.1007/978-3-319-46478-7_38

Osco, L. P., de Arruda, M. d. S., Junior, J. M., da Silva, N. B., Ramos, A. P. M., Moryia, É. A. S., et al. (2020). A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS J. Photogr. Remote Sens.* 160, 97–106. doi: 10.1016/j.isprsjprs.2019.12.010

Paull, R. E., and Duarte, O. (2011). *Tropical Fruits, Vol. 1*. CABI. doi: 10.1079/9781845936723.0000

Pham, V.-Q., Kozakaya, T., Yamaguchi, O., and Okada, R. (2015). "Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 3253–3261. doi: 10.1109/ICCV.2015.372

Py, C., Lacoeuilhe, J. J., and Teisson, C. (1987). *The Pineapple. Cultivation and Uses*. Paris: G.-P. Maisonneuve et Larose.

Ratner, A. J., Hancock, B., and Ré, C. (2019). "The role of massively multi-task and weak supervision in software 2.0," in *CIDR*.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788. doi: 10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 91–99.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Rosenberg, C., Hebert, M., and Schneiderman, H. (2005). "Semi-supervised self-training of object detection models," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)*, Vol. 1 (Breckenridge, CO), 29–36. doi: 10.1109/ACVMOT.2005.107

Ryan, D., Denman, S., Fookes, C., and Sridharan, S. (2009). "Crowd counting using multiple local features," in *2009 Digital Image Computing: Techniques and Applications*, 81–88. doi: 10.1109/DICTA.2009.22

Sam, D. B., Sajjan, N. N., Maurya, H., and Babu, R. V. (2019). "Almost unsupervised learning for dense crowd counting," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33*, 8868–8875. doi: 10.1609/aaai.v33i01.33018868

Sam, D. B., Surya, S., and Babu, R. V. (2017). "Switching convolutional neural network for crowd counting," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4031–4039. doi: 10.1109/CVPR.2017.429

Sanewski, G. M., Bartholomew, D. P., and Paull, R. E. (2018). *The Pineapple: Botany, Production and Uses*. CABI. doi: 10.1079/9781786393302.0000

Settles, B. (2009). *Active Learning Literature Survey*. Technical report, University of Wisconsin-Madison Department of Computer Sciences.

Settles, B., Craven, M., and Ray, S. (2008). "Multiple-instance active learning," in *Advances in Neural Information Processing Systems*, 1289–1296.

Sindagi, V. A. and Patel, V. M. (2018). A survey of recent advances in CNN-based single image crowd counting and density estimation. *Pattern Recogn. Lett*. 107, 3–16. doi: 10.1016/j.patrec.2017.07.007

Wang, C., Zhang, H., Yang, L., Liu, S., and Cao, X. (2015). "Deep people counting in extremely dense crowds," in *Proceedings of the 23rd ACM International Conference on Multimedia*, 1299–1302. doi: 10.1145/2733373.2806337

Xie, Q., Luong, M.-T., Hovy, E., and Le, Q. V. (2020). "Self-training with noisy student improves imagenet classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10687–10698. doi: 10.1109/CVPR42600.2020.01070

Xie, W., Noble, J. A., and Zisserman, A. (2018). Microscopy cell counting and detection with fully convolutional regression networks. *Comput. Methods Biomech. Biomed. Eng*. 6, 283–292. doi: 10.1080/21681163.2016.1149104

Xu, B., and Qiu, G. (2016). "Crowd density estimation based on rich features and random projection forest," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–8. doi: 10.1109/WACV.2016.7477682

Zhang, C., and Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precis. Agric*. 13, 693–712. doi: 10.1007/s11119-012-9274-5

Zhang, H., Sun, W., Sun, G., Liu, S., Li, Y., Wu, Q., et al. (2016). Phenological growth stages of pineapple (*Ananas comosus*) according to the extended biologische bundesantalt, bundessortenamt and chemische industrie scale. *Ann. Appl. Biol*. 169, 311–318. doi: 10.1111/aab.12292

Zhang, Y., Zhou, D., Chen, S., Gao, S., and Ma, Y. (2016). "Single-image crowd counting via multi-column convolutional neural network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 589–597. doi: 10.1109/CVPR.2016.70

Zhou, P., Cheng, G., Liu, Z., Bu, S., and Hu, X. (2016). Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. *Multidimen. Syst. Signal Process*. 27, 925–944. doi: 10.1007/s11045-015-0370-3

Zhu, X. J. (2005). *Semi-Supervised Learning Literature Survey*. Technical report, University of Wisconsin-Madison Department of Computer Sciences.

Žliobaitė, I., Bifet, A., Pfahringer, B., and Holmes, G. (2013). Active learning with drifting streaming data. *IEEE Trans. Neural Netw. Learn. Syst*. 25, 27–39. doi: 10.1109/TNNLS.2012.2236570

Conflict of Interest: JH, PP, HH, BM, and GR were employed by IntelinAir, Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hobbs, Prakash, Paull, Hovhannisyan, Markowicz and Rose. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

CORRECTION
published: 17 May 2021
doi: 10.3389/fpls.2021.695397

Corrigendum: Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation

Jennifer Hobbs[1*†], *Prajwal Prakash*[1,2†], *Robert Paull*[3], *Harutyun Hovhannisyan*[1], *Bernard Markowicz*[1] *and Greg Rose*[1]

[1] IntelinAir, Inc., Champaign, IL, United States, [2] Department of Electrical Engineering, Columbia University, New York, NY, United States, [3] Department of Tropical Plant and Soil Sciences, University of Hawaii at Manoa, Honolulu, HI, United States

Keywords: deep learning-artificial neural network (DL-ANN), active learning, pineapple, computer vision, remote sensing-GIS, weakly supervised, counting, density estimation

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Jennifer Hobbs
jennifer+research@intelinair.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *14 April 2021*
Accepted: *15 April 2021*
Published: *17 May 2021*

Citation:
Hobbs J, Prakash P, Paull R, Hovhannisyan H, Markowicz B and Rose G (2021) Corrigendum: Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation. Front. Plant Sci. 12:695397. doi: 10.3389/fpls.2021.695397

A Corrigendum on

Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation

by Hobbs, J., Prakash, P., Paull, R., Hovhannisyan, H., Markowicz, B., and Rose, G. (2021) Front. Plant Sci. 11:599705. doi: 10.3389/fpls.2020.599705

In the original article, there was a mistake in **Figure 8** as published. The labels for the Validation and Test losses were switched in the legend. The Validation loss should be indicated as a red line with circular markers and the test loss should be indicated as a red line with starred markers. The text and caption described the figure correctly and therefore remain unchanged. The corrected **Figure 8** appears below.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Copyright © 2021 Hobbs, Prakash, Paull, Hovhannisyan, Markowicz and Rose. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

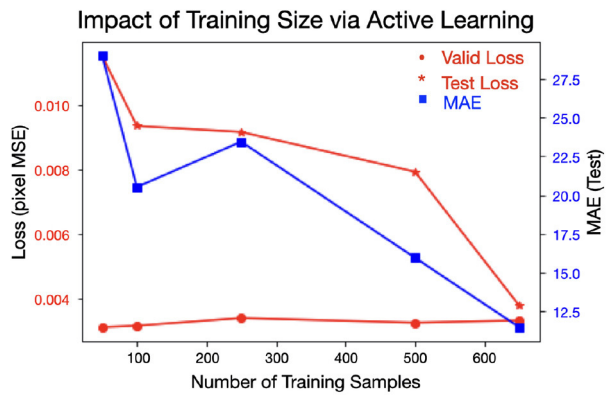

FIGURE 8 | Increasing the amount of (labeled) training data in a smart fashion decreases test loss as well as the MAE on the test set. The validation loss slightly increases as more data is added, suggesting less over-fitting is occurring as more data is added.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 February 2021
doi: 10.3389/fpls.2021.604510

Combining Multi-Dimensional Convolutional Neural Network (CNN) With Visualization Method for Detection of *Aphis gossypii* Glover Infection in Cotton Leaves Using Hyperspectral Imaging

Tianying Yan[1,2†], Wei Xu[3,4†], Jiao Lin[3], Long Duan[1,2], Pan Gao[1,2], Chu Zhang[5,6,7*] and Xin Lv[2,3*]*

[1] College of Information Science and Technology, Shihezi University, Shihezi, China, [2] Key Laboratory of Oasis Ecology Agriculture, Shihezi University, Shihezi, China, [3] College of Agriculture, Shihezi University, Shihezi, China, [4] Xinjiang Production and Construction Corps Key Laboratory of Special Fruits and Vegetables Cultivation Physiology and Germplasm Resources Utilization, Shihezi, China, [5] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [6] Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou, China, [7] School of Information Engineering, Huzhou University, Huzhou, China

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Jakub Nalepa, Silesian University of Technology, Poland
Lucas Costa, University of Florida, United States

***Correspondence:**
Pan Gao
gp_inf@shzu.edu.cn
Chu Zhang
chuzh@zjhu.edu.cn
Xin Lv
lxshz@126.com

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *09 September 2020*
Accepted: *11 January 2021*
Published: *15 February 2021*

Citation:
Yan T, Xu W, Lin J, Duan L, Gao P, Zhang C and Lv X (2021) Combining Multi-Dimensional Convolutional Neural Network (CNN) With Visualization Method for Detection of Aphis gossypii Glover Infection in Cotton Leaves Using Hyperspectral Imaging. Front. Plant Sci. 12:604510. doi: 10.3389/fpls.2021.604510

Cotton is a significant economic crop. It is vulnerable to aphids (*Aphis gossypii* Glovers) during the growth period. Rapid and early detection has become an important means to deal with aphids in cotton. In this study, the visible/near-infrared (Vis/NIR) hyperspectral imaging system (376–1044 nm) and machine learning methods were used to identify aphid infection in cotton leaves. Both tall and short cotton plants (Lumianyan 24) were inoculated with aphids, and the corresponding plants without aphids were used as control. The hyperspectral images (HSIs) were acquired five times at an interval of 5 days. The healthy and infected leaves were used to establish the datasets, with each leaf as a sample. The spectra and RGB images of each cotton leaf were extracted from the hyperspectral images for one-dimensional (1D) and two-dimensional (2D) analysis. The hyperspectral images of each leaf were used for three-dimensional (3D) analysis. Convolutional Neural Networks (CNNs) were used for identification and compared with conventional machine learning methods. For the extracted spectra, 1D CNN had a fine classification performance, and the classification accuracy could reach 98%. For RGB images, 2D CNN had a better classification performance. For HSIs, 3D CNN performed moderately and performed better than 2D CNN. On the whole, CNN performed relatively better than conventional machine learning methods. In the process of 1D, 2D, and 3D CNN visualization, the important wavelength ranges were analyzed in 1D and 3D CNN visualization, and the importance of wavelength ranges and spatial regions were analyzed in 2D and 3D CNN visualization. The overall results in this study illustrated the feasibility of using hyperspectral imaging combined with multi-dimensional CNN to detect aphid infection in cotton leaves, providing a new alternative for pest infection detection in plants.

Keywords: *Aphis gossypii* Glover, machine learning, aphid infection, hyperspectral imaging, convolutional neural network (CNN), visualization

INTRODUCTION

Cotton is rich in cellulose and is the largest source of natural textiles (Ma et al., 2016). It has important applications in the medical field and an important position in the global economy (Rather et al., 2017). However, cotton plants are vulnerable to pests during the 6-month growth period (Wilson et al., 2018). Aphids (*Aphis gossypii* Glovers) are one of the most invasive pests in cotton plants (Wang et al., 2018). Aphids can reproduce rapidly within a few days, hiding in the lower surface of leaves and the core of young leaves. The small size and fast reproduction of aphids are the main obstacles in the control process of cotton pests. Besides, the back color of the juvenile aphids is similar to the plant color, which is not easy to be distinguished. The mature aphids have migratory and strong mobility. Once small-scale aphid pests occur in the cultivation area, the scale of the pests is likely to spread rapidly in a short time, and the cotton yield and quality will be reduced accordingly (Chen et al., 2018a).

Aphids are not only common in cotton crops, but also in traditional crops (Chen et al., 2019; Szczepaniec, 2018; Thorpe et al., 2016). Armstrong et al. (2017) studied sugarcane aphids and found suitable resistance genes in sugarcane. Hough et al. (2017) quantified the effect of temperature on the growth of soybean aphid populations. Kafeshani et al. (2018) used Taylor's power law and Iwao's patchiness to evaluate the spatial distribution of aphids on two citrus species (Satsuma mandarin and Thomson navel). Related scholars studied aphids from the perspectives of physiology and biochemistry, aiming to reduce the impact of aphids on crops. Combined with the experience of the Australian cotton industry, the excessive use of pesticides in the early stages would lead to high resistance in the offspring of aphids (Herron and Wilson, 2017). Therefore, quickly identifying and obtaining information on aphids, formulating efficient management strategies, and reducing the frequency of pesticide use are vital steps to increase crop yields and reduce aphids' resistance to pesticides.

At present, the classification of pests and diseases based on imaging technology has been widely used. This technology is mainly based on pest morphology, plant texture, and morphological changes. Wang et al. (2017b) used the convolution neural network (CNN) to classify RGB images of crop pests containing 82 common pest types in a complex farmland background. Sunoj et al. (2017) used three cameras (digital single-lens reflex camera, consumer-grade digital camera, and smartphone) to take the RGB images of the front side of the leaves under infection by soybean aphids, and classified aphids according to the shape parameters. Deng et al. (2018) extracted features from RGB images of pests in complex environments, and input the features into Support Vector Machine (SVM) for recognition.

Spectroscopic technologies have been used as effective alternatives for pest and pest infection detection. Some scholars have used spectroscopy to study the degree of damage of insect pests to plants, and identify the pest infection. Moscetti et al. (2015) used near-infrared (NIR) spectroscopy to detect and remove olive fruits damaged by fruit flies. Canário et al. (2017) used spectral technology to identify tomato plants in the early stage of infection by whitefly. Basati et al. (2018) used visible/near-infrared (Vis/NIR) spectroscopy to detect wheat samples infected by pests.

Hyperspectral imaging technology combines imaging and spectroscopy techniques to detect the two-dimensional (2D) geometric space and one-dimensional (1D) spectral information of the target, and can quickly and non-destructively analyze the research object (Gao et al., 2019). Hyperspectral imaging has been widely used in plant science. It can be used to evaluate important parameters of plant health, such as nutrients, plant biomass, biological stress, and abiotic stress (Thomas et al., 2018). Plant diseases and pest detection is a significant research field of hyperspectral imaging. Previous studies have proved that hyperspectral imaging can identify the outbreak and dynamics of plant diseases and pests (Moghadam et al., 2017; Huang et al., 2018). In the application of hyperspectral imaging in pest detection as a branch of plant science research direction, the current main research includes pest identification (Liu et al., 2016) and pest infection degree classification (Lu and Ariana, 2013). Identification and segmentation of insect infected areas are one of the vital steps in pest detection. Related scholars have already identified and segmented infected areas (Tian et al., 2015). Hyperspectral images (HSIs) can provide a huge number of features, including spectral features and spatial features. As high-dimensional data, HSIs have a large amount of data information. How to mine valuable information has become a difficult problem.

Deep learning (DL) is currently a more concerning data processing method, and it has a wide range of applications in hyperspectral image processing (Signoroni et al., 2019). DL methods combined with spectral features of HSIs have been widely used in plant science, such as plant disease classification (Han and Gao, 2019). In addition, it is a new trend to select key wavelength images from HSIs and extract spatial features for disease segmentation (Feng et al., 2020). Although the performance of DL is better than conventional machine learning methods, DL methods based on spectral features or spatial features use less valuable information, ignoring the spatial features, or spectral features of HSIs. There is still room for improvement in the performance of DL. Previous studies have proved that the performance of DL methods based on spectral-spatial features is fine. A conventional method is to fuse separately extracted spectral features and spatial features (Zhao and Du, 2016; Wang et al., 2017a). Another method is to use three-dimensional (3D) CNN, whose 3D convolution kernel directly combines local spectral-spatial features (Wang et al., 2019a). At present, there are various DL architectures that combine the spectral-spatial features of HSIs, such as Resnet and DenseNet (Paoletti et al., 2019; Zhong et al., 2018).

Although DL can handle high-dimensional data, redundant features in HSIs are a huge challenge. HSIs contain much information irrelevant to the research target, which increases the computational burden, reduces the analysis efficiency, and interferes with the analysis results. Thus, dimensionality reduction (band selection and feature extraction) is a critical measure for the application of HSIs in various fields. Dimensionality reduction can be divided into linear and

non-linear methods. Conventional linear methods include principal component analysis and factor analysis, which can extract features and select bands through correlation coefficients. The main non-linear dimensionality reduction methods are Isomap and Auto-Encoder, which can extract features (Kozal et al., 2013). DL, as a non-linear dimensionality reduction method, its convolutional layer can transform HSIs into low-dimensional features (Zhang et al., 2020a), and the constructed DL architectures (e.g., attention-based CNN) can select the optimal band subset (Cai et al., 2020; Lorenzo et al., 2020). Although DL can achieve better results in many cases, it is meaningful to identify the part of the input data that has a greater contribution to the research target due to a large amount of input data, so as to reduce the input of useless information (or information with low contribution) in the future research. DL visualization, which can interpret research results, is an effective way to find important features of research goals (Yosinski et al., 2015; Zhou et al., 2016). The saliency map is a DL visualization method, and its principle is to reflect the main contribution area of the input data through the gradient of the backpropagation (Simonyan et al., 2013). 2D CNN based on DL and DL visualization methods have made rapid progress in plant phenotypic stress, involving plant diseases and pests (Singh et al., 2018). Hyperspectral imaging, which can perform 1D analysis, 2D analysis, and 3D analysis, can provide spectral (1D) information, spatial (2D) information, and spectral-spatial (3D) information. At present, 1D CNN combined with hyperspectral imaging can be used for plant disease detection (Han and Gao, 2019). 3D CNN has been partially researched in plant disease detection (Nagasubramanian et al., 2019). However, there are few studies using 1D, 2D, and 3D CNN for the same plant using hyperspectral imaging, especially pest detection. Meanwhile, the use of DL visualization to find high contribution regions of input data is usually ignored.

This study aimed to explore the application of multi-dimensional CNN in aphid infection identification. In this study, hyperspectral imaging and DL were used to diagnose leaves in cotton plants infected by aphids (*A. gossypii* Glovers). Using HSIs containing a single leaf, the spectra, and RGB images were extracted. CNN models using 1D analysis in extracted spectra, 2D analysis in extracted RGB images, and 3D analysis in hyperspectral images were established, and compared with conventional machine learning methods (Logistic Regression, LR; Support Vector Machine, SVM; Nearest Neighbors, NN; Decision Tree, DT). At the same time, the results of 1D, 2D, and 3D CNN were visualized. Important wavelength ranges were discovered through the 1D and 3D CNN visualization, and important infection regions were discovered through the 2D and 3D CNN visualization.

MATERIALS AND METHODS

Sample Preparation

Cotton plants were cultivated in greenhouses, which in the North Second District of Shihezi University, Shihezi City, Xinjiang Uygur Autonomous Region, China (86°3′34″ E, 44°18′58″ N). The cotton plants (Lumianyan 24) were planted on October 15, 2019. During the cultivation process until the seedling stage of the cotton plant, cotton plants were cultivated under two different light intensities and the same medium. It resulted in 20 high and 20 short cotton plants. Half of the cotton plants of each morphology type were inoculated with aphids, and the other half of the cotton plants without aphids were used as the control group. In the plant cultivation process, the leaves in cotton plants would not curl up and deform under the infection of aphids, which was suitable for shooting. At the same time, the blade size of leaves was similar, which was conducive to modeling. During the plant cultivation, the highest temperature of the culture environment was 28.5°C, the lowest temperature was 15.5°C, and the average temperature was 20.85°C. The highest relative humidity was 56%, the lowest relative humidity was 32%, and the average relative humidity was 44.475%.

As shown in **Figure 1**, RGB images taken from the overhead view of infected cotton plants and healthy cotton plants are shown. Some areas of cotton leaves infected by aphids are white. The reason is that the area is covered with aphid secretions, which is a carbohydrate that reflects most of the light.

Hyperspectral Image Acquisition and Preprocessing

The research object of this study was healthy leaves and infected leaves and the HSIs were taken from December 5, 2019 to December 25, 2019, with an interval of 5 days. For each sampling and shooting time, cotton plants were destructively sampled, and the front of the leaves was shooted. To reduce the interference of biochemical factors such as the open state of leaves, the shooting time was fixed at around 14:00 (UTC/GMT + 08:00).

In this study, the Vis/NIR hyperspectral imaging systems were composed of four modules, including an imaging module, an illumination module, a lifting module, and a software module. The imaging module was SOC 710VP camera (Surface Optics Corporation, San Diego, CA, United States). The camera had a push broom and dual CCD detectors. When the sample was taken at a fixed position, the SOC 710VP hyperspectral camera used the internal translation push-broom mechanism to scan samples. The HSI size was 128 wavebands × 520 pixels × 696 pixels, each pixel contained the full spectrum in the range of 376–1044 nm with the spectral resolution of 5 nm. The lighting module was composed of two halogen lamps with a power source of 75 W. The lifting platform module placed the shooting object, and the imaging module could fully capture the shooting object by lifting. The software module was used to control HSI acquisition. The shooting integration time of the hyperspectral camera was 25 ms, the aperture value was F1.4, and the shooting height was 86 cm. During the HSI acquisition process, the imaging conditions and system parameters were kept unchanged. After HSI acquisition, the original HSIs were calibrated to reflectance images according to Equation 1.

$$I_c = \frac{I_r}{2I_g} \quad (1)$$

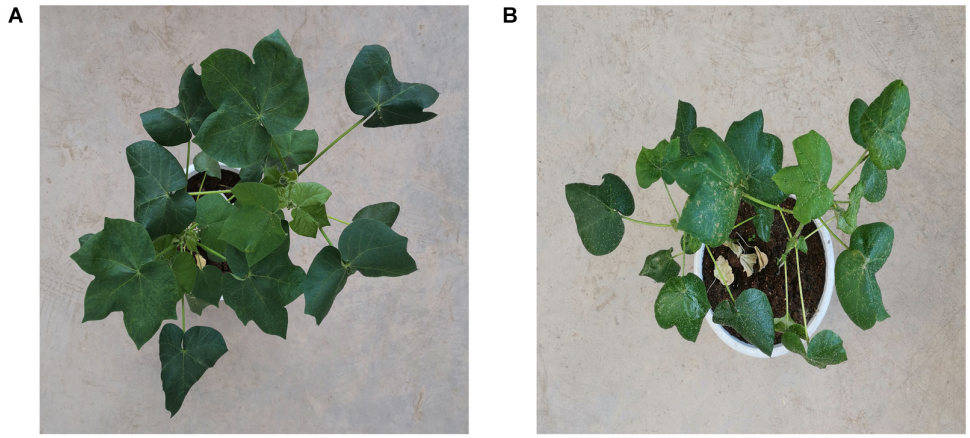

FIGURE 1 | (A) An image taken from the overhead view of healthy cotton plants. **(B)** An image taken from the overhead view of cotton plants infected by *Aphis gossypii* Glovers.

Where I_c is the reflectance image, I_r is the original image, and I_g is the gray (combined with 50% black and 50% white) reference image.

The Savitzky–Golay smoothing filter (the kernel size was $5 \times 5 \times 5$, the polynomial order was 3, the filter calculated the filtered value at the central node of the kernel) was used to reduce the random noise on the reflectance HSIs. The 3D data cube of a cotton leaf HSI is shown in **Figure 2**.

Data Set Construction

In this study, a total of 256 HSIs of infected leaves and healthy leaves were collected. Considering that wavelengths under 450 nm and over 1000 nm have more noise, as a limitation of the sensor (Malenovsky et al., 2006). Only the wavelengths in the range of 461–988 nm were studied.

The software provided by SOC was used to synthesize RGB images at wavelengths of 461, 548, and 698 nm in HSIs.

In HSIs, the area containing a single leaf was regarded as a region of interest (ROI). The pixel-wise spectra (461–988 nm) in the ROI were extracted and averaged to represent the sample. Due to the average spectra contains redundant and collinear information, and the use of the first derivative can suppress background information (Jin and Wang, 2016). The first derivative spectra of the average spectra were used.

Data Analysis Methods

Conventional Machine Learning Methods

Logistic regression (LR) is a generalized linear regression analysis model (Stoltzfus, 2011). LR uses the Logistic Sigmoid function to convert the output into a probability value to predict the label. The basic LR model deals with binary classification problems. For the LR model, the regularization parameter C is used to solve the model fitting problem. In this study, the optimization range of C was in $[10^{-5}–10^{5}]$.

Support Vector Machine (SVM) is a common classification algorithm used for supervised learning (Jian et al., 2016). The principle of SVM is to find the hyperplane with the largest interval in the feature space. SVMs are currently divided into linear SVM, polynomial SVM, radial basis function (RBF) SVM, and sigmoid SVM. The SVM using the "linear" kernel function is essentially a linear classifier, similar to LR. To compare with LR, the kernel optimization range was in ("polynomial," "sigmoid," "RBF") in SVM. For the SVM model, the regularization parameter C and the kernel coefficient γ are used to solve the model fitting problem. In this study, the optimization range of C and γ were all in $[10^{-5}–10^{5}]$.

Nearest Neighbors (NN) is a widely used pattern recognition method (Zhang et al., 2018). The principle of NN is to find training samples that meet the first K shortest distances of distance test samples and predict the label based on these training samples. In this study, the optimization range of K was in [1, 30].

Decision Tree (DT) is a supervised learning method for classification (Wang et al., 2019b). Its purpose is to create a model that can learn simple decision rules from data features. DT uses the rules to predict the label of a test sample. DT learns data through if-then-else decision rules and estimates the label of the predicted sample. The deeper the decision tree, the more complex the decision rules and the better the fit to the training samples. For the DT model, the parameter max_depth is used to limit the depth of the tree. In this study, the optimization range of max_depth was in [1, 30].

Convolutional Neural Network

Convolutional Neural Network (CNN) is a neural network based on convolutional layers. The CNN model usually consists of five parts as input, convolution, pooling, dense connection, and output. There are differences in these five parts of the current mainstream CNN models (Khan et al., 2019). Since Resnet solves the problem of network degradation in DL, it becomes the backbone network for subsequent research (Sanchez-Matilla et al., 2020; Veit and Belongie, 2020; Zhang et al., 2020b). In this study, Resnet-18 was used as the backbone network to construct CNNs (He et al., 2016).

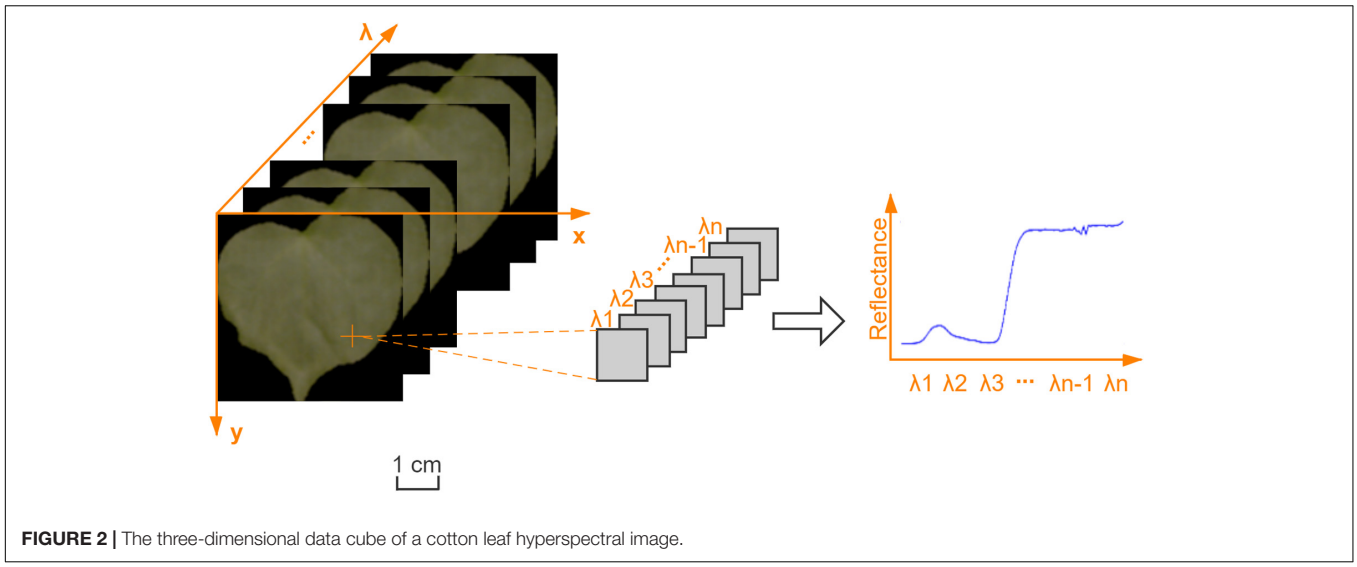

FIGURE 2 | The three-dimensional data cube of a cotton leaf hyperspectral image.

For spectra, the network layers of Resnet-18 were adjusted. The process of sliding windows of each network layer for 2D analysis was adjusted to the process of sliding windows for 1D analysis. For RGB images, Resnet-18 was directly used. For HSIs, the classification model suitable for 3D analysis was designed. The residual blocks were used in the model. The residual block allowed the training of the deep network to proceed smoothly. The main reason was that the stack of the residual block could effectively return the gradient, and the skip connection was added based on the stack. Since HSIs had three dimensions (depth, height, and width), and the amount of data was large, the network structure based on 3D CNN should not be complicated. Otherwise, there would be insufficient computing power.

In this study, 3D CNN consisted of two convolutional layers, two batch normalization layers, two max-pooling layers, two residual blocks, a global average pooling layer, and a dense layer, followed by a Softmax layer. Since the size of a larger convolution kernel will improve the performance of the network, smaller size of the convolution kernel will increase the convergence speed of the network (Cai et al., 2018; Tan and Le, 2019). Considering the convolution kernel size as a compromise, the convolution kernel with the size of 9 × 3 × 3 was used as the first convolutional layer, the convolution kernel with the size of 3 × 1 × 1 was used as the second convolutional layer, and the number of channels was 5 and 3 in turn. The Rectified Linear Unit (ReLU) was used as the activation function of the convolution output. The size of max-pooling layers was 3. The 1 × 1 × 1 convolutional layer was used in the first residual block and not used in the second residual block. The number of channels in two residual blocks was 3 and 5 in turn. 3D CNN architecture is shown in **Figure 3**.

The CNN models could be used for data dimensionality reduction (Song et al., 2019). The CNN models non-linearly map high-dimensional data to low-dimensional space. In this study, the RGB images and hyperspectral images reduced by the CNN models were used for modeling based on conventional machine learning methods. The global pooling layer features in the CNN models were used in the modeling process of conventional machine learning methods.

Saliency Map of Convolutional Neural Network

The saliency map is a CNN visualization method that can reflect the impact of each data element on the classification results. In this study, the saliency map visualization method proposed by Simonyan et al. (2013) was used. When the sample label was correctly predicted, each element in the data would have a corresponding contribution value, and the magnitude of the contribution value reflected the importance of the elements. The contribution value was visualized by saliency maps, which could effectively observe the important regions of the sample identified by CNN.

Given a data D_0 of category c in the test set, after being classified by the CNN model, the score value S_c will be obtained. If the predicted category is consistent with the true category, the weight can be calculated. The approximate calculation process is carried out according to Equation 2.

$$w = \text{abs}\left(\frac{\partial S_c}{\partial D}|D_0\right) \tag{2}$$

Where w is the absolute value of the derivative of score S_c concerning data D_0, and w is valid only when the predicted category is consistent with the true category.

In the case of a HSI, the wavelength variable v of the pixel (i, j) of the image I corresponds to the w element whose index is $h(i, j, v)$. To obtain the contribution value of a single category for each pixel (i, j), the maximum value M of w on all wavelength variables is used, as shown in Equation 3.

$$M = \max_v\left(w_{h(i,j,v)}\right) \tag{3}$$

In this study, another interpretation method was visualized for the saliency map. As shown in Equation 4, the cumulative contribution C of wavelengths in the test set is calculated, and

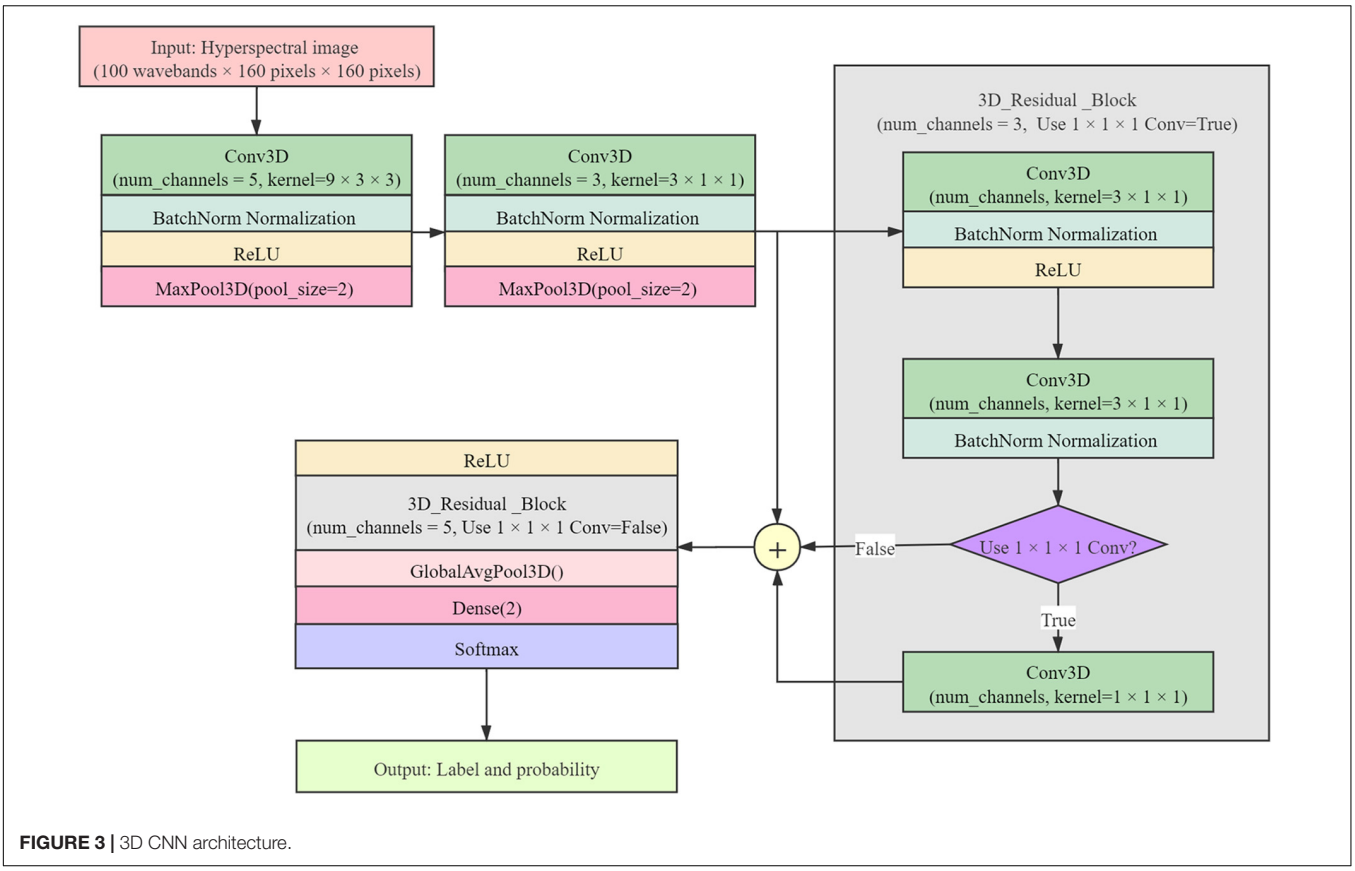

FIGURE 3 | 3D CNN architecture.

the L1-norm is used to normalize by column.

$$C = \frac{\sum_j \in (1, 2, \ldots S)\, M_j}{|| \sum_j \in (1, 2, \ldots S) M_j ||_1} \quad (4)$$

Among them, M_j is the saliency map of the jth sample, and the number of samples is S.

To compare with the spectral wavelengths and observe the wavelengths in the hyperspectral images with the high contribution rate, the L1-norm visualization method proposed by Nagasubramanian et al. (2019) was used in this study.

Model Evaluation

To better evaluate the performance and stability of the models, the training set, validation set, and test set of HSIs in each shooting period were divided according to the ratio of 3:1:1. The division of spectral data set, RGB image data set, and HSI data set is shown in **Table 1**. The first derivative spectra, RGB images, and HSIs of all samples had a one-to-one correspondence in each set. In each set, the number of samples of each class was almost equal, with slight differences due to the number of leaves in different cotton plants. Due to the small amount of data in different periods, it was impossible to explain the best period to detect infection. However, the number of healthy samples and infected samples in general could be used to model and explore important wavelengths and regions.

TABLE 1 | The division of data set.

Data Set Type	Training	Validation	Test
First derivative spectra	75/78[a]	26/25	24/28
RGB images	75/78	26/25	24/28
Hyperspectral images	75/78	26/25	24/28

[a] *The number of category samples is expressed as: the number of healthy samples/number of infected samples.*

Batch normalization could speed up model convergence and shorten the model building time (Ioffe and Szegedy, 2015). In this study, batch normalization was used for all sample sets before training.

Bayesian optimization algorithm (BOA) was used in the parameter optimization process of conventional machine learning methods (Pelikan and Goldberg, 1999). The classification accuracy was used to evaluate the performances of each model, which was calculated as the ratio of the number of correctly classified samples to the total number of samples. The training set was learned by the models, and the model parameters were optimized by BOA 200 times. The models with the highest prediction accuracy in the validation set were saved and evaluated in the test set.

In this study, for CNN models, the batch size was 10 and the epoch was 200 while training the models, the optimization algorithm was set to SGD, the initial learning rate was set to 0.1,

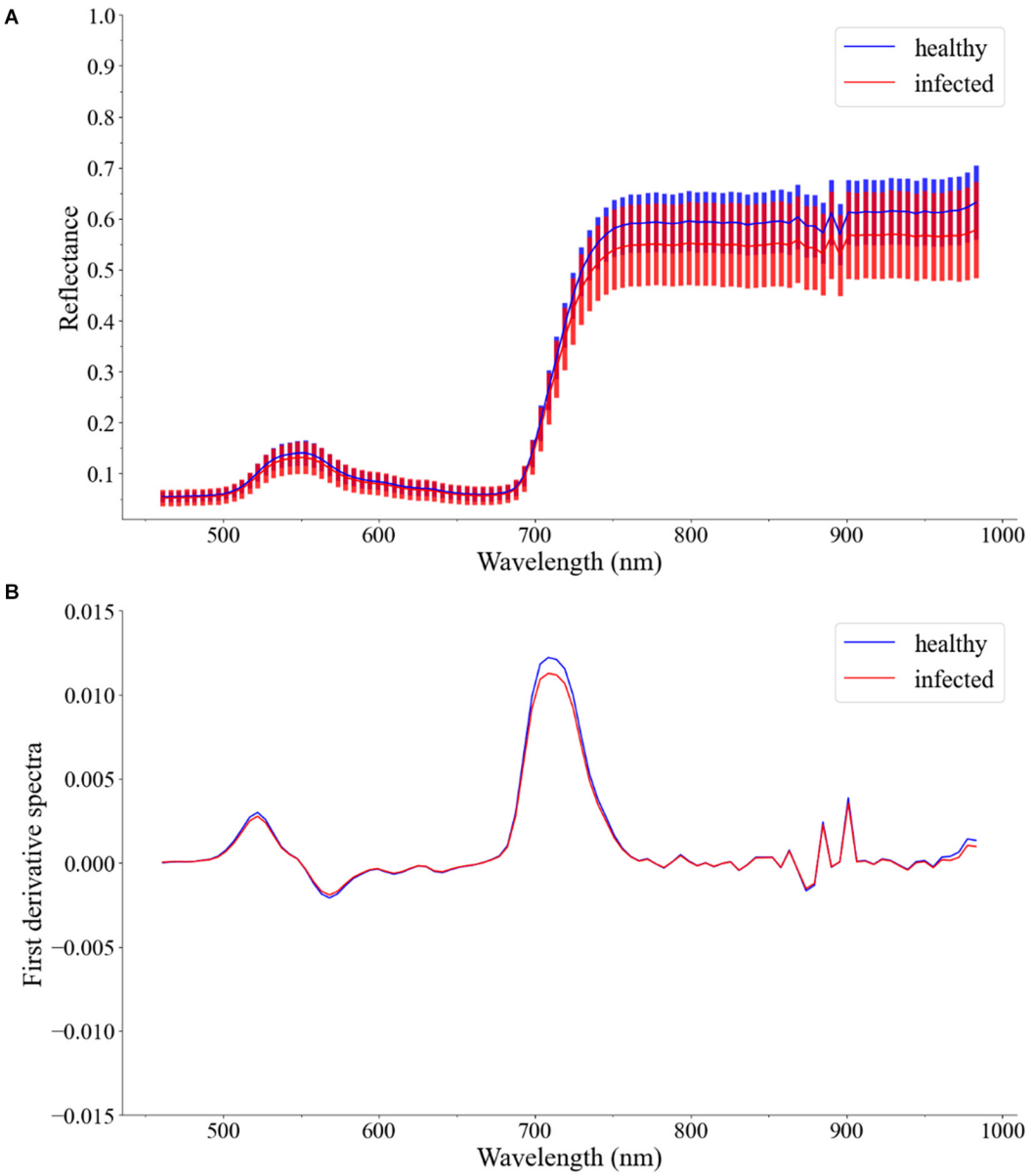

FIGURE 4 | (A) Vis/NIR average spectra (461–988 nm) and standard deviation for healthy leaves and infected leaves. **(B)** First derivative spectra of average spectra.

and the learning rate was gradually adjusted to 0.01 during the training process. The CNN models were trained from scratch and initialized using the Xavier method. During the training process, the CNN models with the best fit in the validation set, and the lowest loss value were selected, and the test set was used to evaluate the models.

Software and Hardware

In this study, Python scripting language (version 3.7.6, 64 bit) was used for numerical calculations. Conventional machine learning methods were implemented on the python library package scikit-learn (version 0.23.1). The CNN models were built on the MXNet (version 1.5.0) framework (Amazon, Seattle, WA, United States). All data analysis procedures were implemented on a computer with 16 GB of RAM, the NVIDIA GEFORCE GTX 1080Ti GPU, and the Intel Core i7-9700K CPU.

RESULTS

Spectral Profiles

In this study, the first derivative spectra were used to build the models. **Figure 4A** shows the Vis/NIR average spectra (461–988 nm) and standard deviation for each class of leaves. **Figure 4B** shows the first derivative spectra of average spectra.

For the Vis/NIR average spectra, the reflectance of healthy leaves was slightly higher. There was a large overlap in the standard deviation areas of the reflectance of healthy leaves and infected leaves. For the first derivative spectra of average spectra, there was no difference between healthy leaves and infected leaves, except around 720 nm. In general, the spectra of different types of samples cannot provide clear enough discrimination. Therefore, other classification methods should be considered.

Classification Models

For conventional machine learning models, the features of the global pooling layer reduced by CNN models were used for modeling and evaluation. The CNN models used for dimensionality reduction were trained from scratch and could completely predict the categories of all samples. For CNN models, the images were resized to reduce the influence of sample shape on DL. The resized RGB image size was 160 pixels × 160 pixels and the resized HSI size was 100 wavebands (461–988 nm) × 160 pixels × 160 pixels. The classification results of each model are shown in **Table 2**. The classification results of some other conventional classification methods are reflected in **Supplementary Table 1**.

For the first derivative spectra, the CNN model performed best, with an accuracy rate of 98.08% in the test set. For RGB images, LR and CNN model performed best, with an accuracy rate of 84.62% in the test set. For HSIs, the CNN model performs best, with an accuracy rate of 88.46% in the test set.

For the first derivative spectra, RGB images, and HSIs, CNN performed best, and LR, SVM, and DT model perform worse in turn, and the worst was the NN model.

For the performance of the same model in different datasets, most models performed best in the first derivative spectra, followed by HSIs, and the worst performance in RGB images.

Visualization of Convolutional Neural Network

In this study, the test set used for 1D CNN was visualized by Equations 2 and 3. The samples of healthy leaves and infected leaves were normalized so that the sum of the wavelength contribution value of each sample was 1. The normalized results were displayed visually. As shown in **Figure 5**, the row coordinates represent the wavelength, and the ordinate represents the sample.

For all the samples in the test set, the wavelength with the largest contribution was concentrated in the NIR wavelength range of 750–950 nm, followed by the Vis wavelength range of 460–660 nm.

The saliency maps of healthy leaves and infected leaves were visually explained in Equation 4. The visualization results are shown in **Figure 6**.

In **Figure 6**, the cumulative contribution value is very low in the red-side spectral range of 660–750 nm. The most contributory wavelength ranges were concentrated in 750–950 nm, followed by 460–660 nm, which also explains the important wavelength ranges of the saliency maps. The 460–600 nm wavelengths and 750–950 nm wavelengths of the spectral data set were respectively intercepted, and 1D CNN was used for re-modeling. As a result, in the performance of the test set, the accuracy of 460–600 nm wavelengths was 71.15%, and the accuracy of 750–950 nm wavelengths was 90.38%. The results confirmed the effectiveness of the visualization method.

In this study, the RGB images and the HSIs were selected from the test set of healthy leaves and infected leaves (one-to-one correspondence between RGB image and HSI samples). As shown in **Figure 7**, the saliency maps are visualized on the test set, which includes the samples of two classes in five periods, and the labels of all samples are correctly predicted.

Cotton leaves infected by aphids will attach aphid secretions. Aphid secretion is a carbohydrate product secreted by *A. gossypii* Glovers, which has a high reflection effect. The secretion of aphids causes the leaves to turn white. At the same time, the texture of real leaves under the infection of aphids is changed. It could be seen from the RGB images that part of the infected leaves was white.

In the saliency map of the 2D CNN, it was found that the pixel area with the largest contribution value was concentrated on the aphids' secretions (leaf whiteness) and leaf textures in infected leaves. 2D CNN looked for the presence of these areas that affect classification in healthy leaves. And there was no whitening in some areas of healthy leaves caused by aphid secretions. Therefore, in the saliency map of 2D CNN, there were obviously few areas with larger contribution values, and the leaf textures were mainly displayed in healthy leaves.

In 3D CNN visualization, the pixel area with the largest contribution values was always concentrated near the leaf veins and leaf edges.

By using the method proposed by Nagasubramanian et al. (2019), the visualization result of the wavelength contribution rate of HSIs is shown in **Figure 8**.

The results showed that the 3D CNN model was not sensitive to the 750–950 nm wavelengths in HSIs, and had excellent sensitivity to the 460–600 nm wavelengths. In HSIs, the pixel area with the largest contribution rate depends on the 460–600 nm wavelengths. The 460–600 nm wavelengths and 750–950 nm wavelengths of the HSI set were respectively intercepted, and 3D CNN was used for re-modeling. As a result, in the performance of the test set, the accuracy of 460–600 nm wavelengths was 88.46%. And the accuracy of 750–950 nm wavelengths was 53.85%, which was similar to blind guessing. The results confirmed the effectiveness of the visualization method. It could also indicate that the 3D CNN model paid attention to the area around the leaf veins and the leaf edges in the 460–600 nm wavelengths of the HSIs.

DISCUSSION

Aphids not only appear on cotton crops but also often appear on other traditional crops. Aphids affect the growth and development of crops, which in turn affect the quality and yield of agricultural products grown on crops. From a physiological point of view, the selection of aphid-resistant crops is one of

TABLE 2 | Classification accuracy of the conventional machine learning methods and convolutional neural network (CNN).

Data set type	Methods	Category values	Training			Validation			Test		
			0	1	Accuracy	0	1	Accuracy	0	1	Accuracy (%)
First derivative spectra	LR	0[a]	71	4		22	4		21	3	
		1	7	71		4	21		4	24	
		Total			92.81			84.31			86.54
	SVM	0	74	1		22	4		15	9	
		1	2	76		2	23		5	23	
		Total			98.04			88.24			73.08
	NN	0	75	0		17	9		14	10	
		1	0	78		7	18		6	22	
		Total			100.00			68.63			69.23
	DT	0	67	8		14	12		20	4	
		1	0	78		8	17		5	23	
		Total			94.77			60.78			82.69
	CNN	0	75	0		26	0		24	0	
		1	0	78		0	25		1	27	
		Total			100.00			100.00			98.08
RGB images	LR	0	75	0		26	0		24	0	
		1	0	78		0	25		8	20	
		Total			100.00			100.00			84.62
	SVM	0	75	0		26	0		23	1	
		1	0	78		0	25		9	19	
		Total			100.00			100.00			80.77
	NN	0	75	0		26	0		24	0	
		1	3	75		1	24		13	15	
		Total			98.04			98.04			75.00
	DT	0	75	0		24	2		21	3	
		1	0	78		0	25		9	19	
		Total			100.00			96.08			76.92
	CNN	0	75	0		26	0		24	0	
		1	0	78		0	25		8	20	
		Total			100.00			100.00			84.62
Hyperspectral images	LR	0	75	0		23	3		21	3	
		1	0	78		4	21		6	22	
		Total			100.00			86.27			82.69
	SVM	0	75	0		26	0		21	3	
		1	0	78		2	23		5	23	
		Total			100.00			96.08			84.62
	NN	0	75	0		26	0		24	0	
		1	15	63		8	17		11	17	
		Total			90.20			84.31			78.85
	DT	0	75	0		22	4		18	6	
		1	0	78		4	21		4	24	
		Total			100.00			84.31			80.77
	CNN	0	70	5		25	1		22	2	
		1	6	72		3	22		4	24	
		Total			92.81			92.16			88.46

[a]*0 means the label of the healthy leaves, 1 means the label of the infected leaves.*

the effective measures to reduce the impact of aphids, but the process of crop selection and breeding takes a long time (Cabral et al., 2018; Zhan et al., 2020). From a chemical and physical point of view, the infection caused by aphids can be captured by the imaging system (Chen et al., 2018b). Currently, there are few studies on aphid detection based on near-ground object hyperspectral imaging. In this study, the Vis/NIR hyperspectral imaging system was used to detect aphid infection and direct sensory monitoring of aphid infection. It can provide a reference for the direct control of aphids on crops.

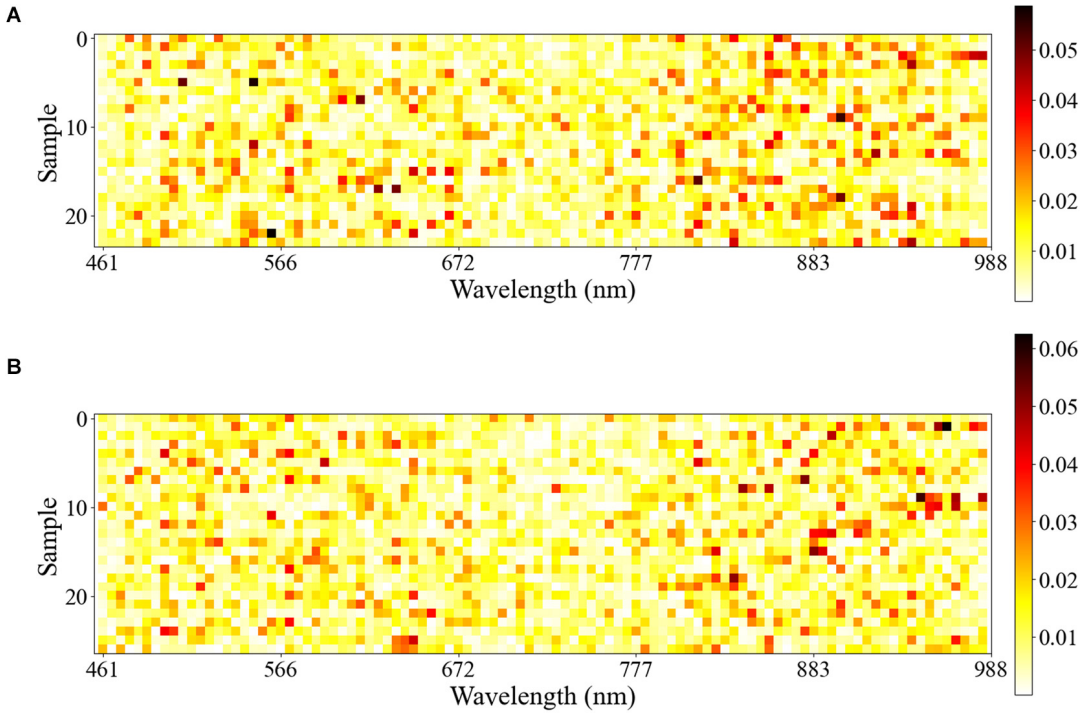

FIGURE 5 | The row coordinate represents the wavelength, and the ordinate represents the sample correctly predicted by 1D CNN. The depth of each row of color represents the importance of the wavelength of the corresponding sample in the process of identifying *A. gossypii* Glovers infection. **(A)** Visualization of the first derivative spectra test set of healthy leaves. **(B)** Visualization of the first derivative spectra test set of infected leaves.

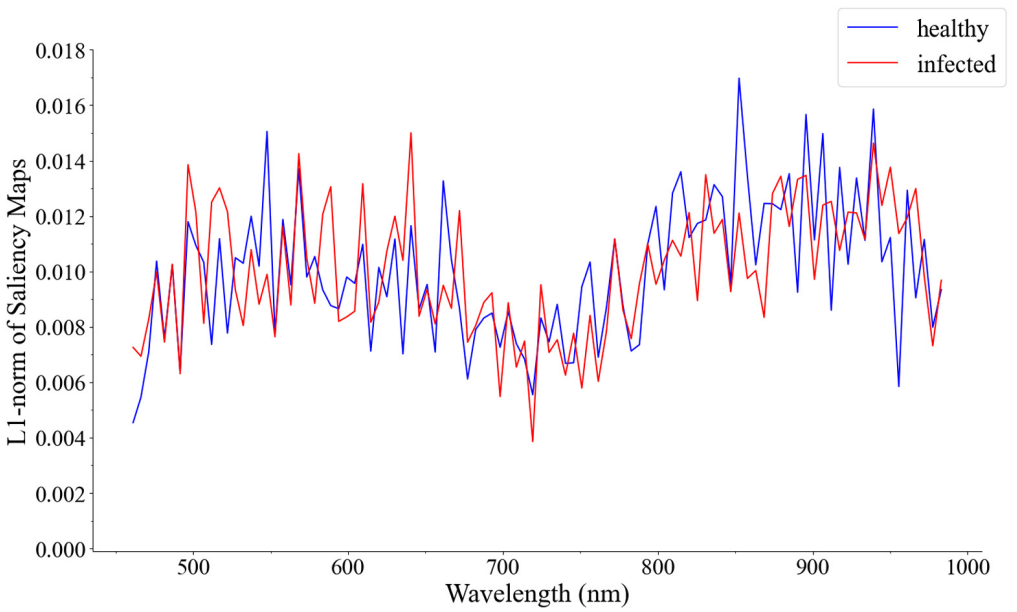

FIGURE 6 | Visualization of the cumulative contribution of wavelengths of samples correctly predicted in the 1D CNN in the first derivative spectra test set.

Due to hyperspectral imaging can obtain spatial and spectral information of the research object, both spatial and spectral information in hyperspectral imaging can be used to detect pest infection and disease infection (Ahmad et al., 2018). Therefore, the spectral information and spatial information based on hyperspectral imaging have been introduced in previous research. Previous studies have determined that using spectral information in hyperspectral imaging is highly effective for detecting pest

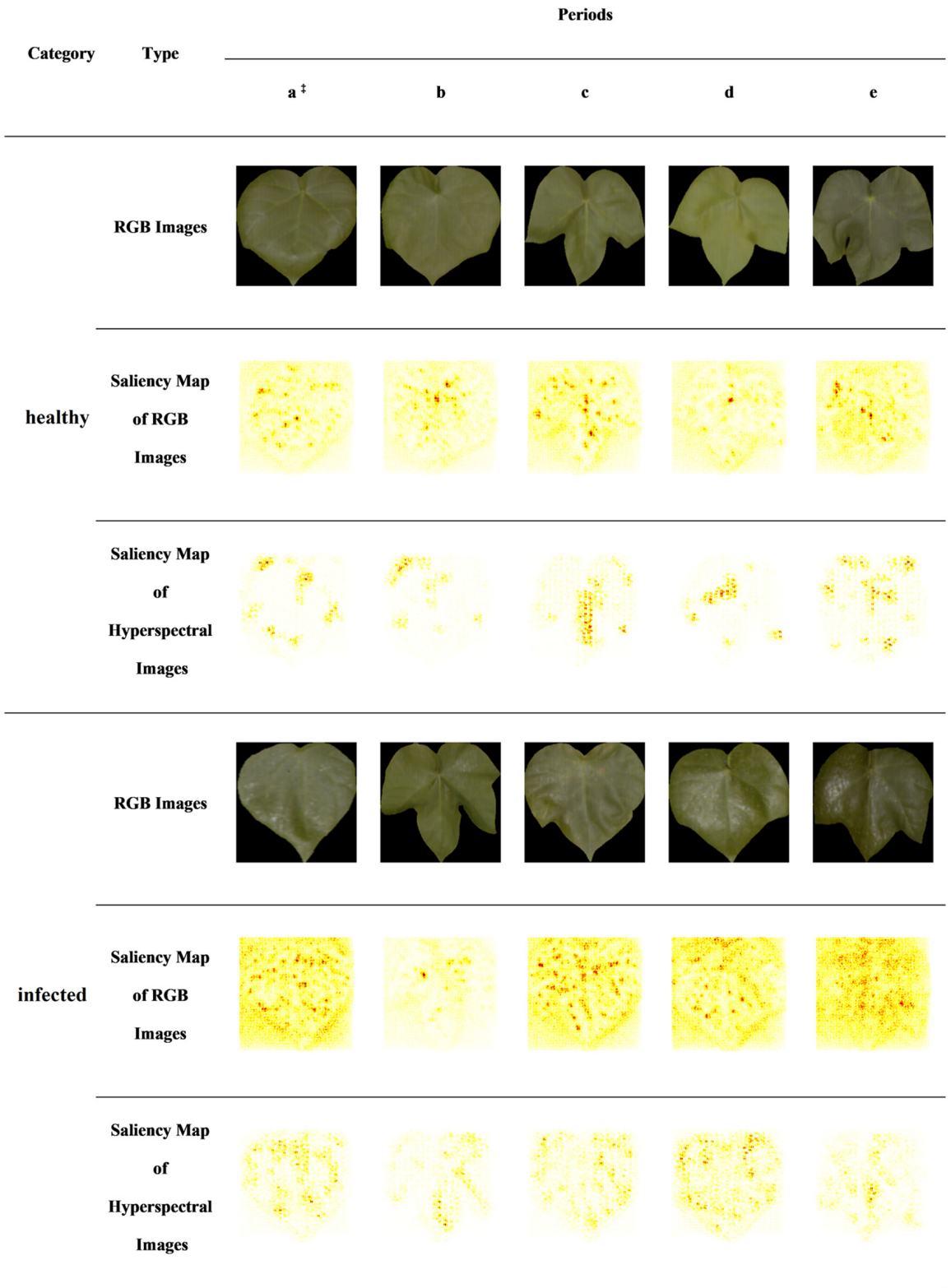

FIGURE 7 | CNN-based saliency map of RGB images and hyperspectral images. The saliency maps of RGB images are based on 2D CNN, and the saliency maps of hyperspectral images are based on 3D CNN. In each saliency map, the features of the darker regions have a greater impact on the identification results. ‡a–e indicates the five periods of dataset collection in turn.

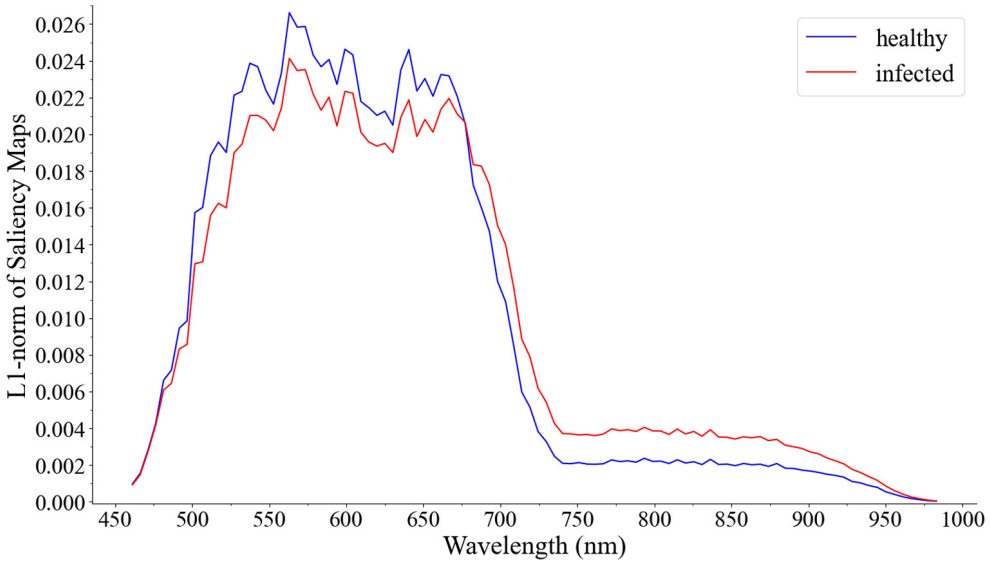

FIGURE 8 | Visualization of the wavelength contribution of the hyperspectral image test set.

infection and disease infection (Rady et al., 2017). It is feasible to use the average spectra of hyperspectral imaging or the pixel-wise spectra for 1D analysis and detection of infection. Zhou et al. (2019) used average spectra to detect infection. Qiu et al. (2019) used the spatial information of spectral imaging and used the pixel-wise spectra to detect the infected area. However, the 1D analysis does not make full use of the spatial information of hyperspectral imaging. Due to a large amount of hyperspectral imaging data, the selection of key wavelength images is important for the use of spatial information (Li et al., 2019). According to the 2D analysis of the key wavelength information, the plant infection area can be marked with higher precision. However, 1D analysis and 2D analysis are not sufficient for the mining and utilization of spatial information or spectral information. The 3D analysis makes effective use of spatial information and spectral information, and it has a few applications in plant disease monitoring. At present, there are few studies on simultaneous 1D analysis, 2D analysis, and 3D analysis for the same infection. In this study, 1D analysis, 2D analysis, and 3D analysis were used to detect aphid infection in cotton leaves, and the multi-dimensional detection results were all good. On the whole, 1D analysis is worthy of consideration for rapid detection of infection, and 2D analysis and 3D analysis can be used to detect the infected area.

Currently, conventional machine learning methods are used to monitor plant diseases and insect pests. Meanwhile, the DL method has been widely used in the monitoring of plant diseases and insect pests. 1D DL and 2D DL are widely used in plant diseases and insect pest detection. 3D DL has been partially applied in the monitoring of plant diseases and insect pests. However, the DL method has not been reasonably explained in the detection of plant diseases and insect pests. In this study, the saliency map was used to visualize the DL model. Through visualization, important wavelengths and spatial regions were discovered. The important wavelengths and spatial regions were consistent with actual conditions. Overall, the visualization of DL provides new ideas for the interpretation of the application of plant pests and disease detection in DL in the future.

Overall, the classification results of the 1D analysis, 2D analysis, and 3D analysis were fine. In the field of spectroscopy, DT and NN models were prone to overfitting problems, which may be the reason for the poor results (Liu et al., 2008; Mandrell et al., 2020). RGB images lost a lot of spectral information and only contain color information, which may be the reason for their poor performance. Due to the computing power of the computer in this study, HSIs contain a lot of spectral information and spatial information, but the 3D CNN model may not make full use of information. For the first derivative spectra, while the spatial information was lost, the spectra were simple but contain enough information, and most models were easy to learn data features under the existing computer computing power, which may be the reason for the outstanding effect. Considering the impact of the training set size on the overall performance of the investigated classification methods, the size of the validation and test sets were kept constant, and the training set size was sequentially expanded from 25/27, which was the same size as the validation and test sets, to 75/78 (the number of healthy samples/number of infected samples). The results are shown in **Supplementary Table 2**. The overall performance was poor when the training set size was 25/27, and when the training set size was 50/52, the overall performance approximated the overall performance of the training set size of 75/78. With the increase of training set size, the overall performance grew slowly. The overall performance was likely to be the best when the training set size was 75/78.

Aphids pierce and suck plant tissues on tender leaves, tender stems, buds, and floral organs, and other young parts of the plant, and suck the juice, which will make the veins and leaves

green, yellow, white, or thin (Dubey et al., 2013). These will cause changes in the color, texture, and spectral reflectance of the leaves. In the process of visualization, the saliency maps showed that the 1D CNN model was interested in the 750–950 nm wavelengths, followed by the spectral range of 460–660 nm. The 2D CNN could capture the color and texture characteristics of the leaves, and the model itself did not notice the aphids. The main interest spectral range of the 3D CNN model was 460–660 nm, and its interest area was around the leaf veins and the leaf edges. Combined with the L1-norm visualization of HSIs, it was found that 3D CNN was not interested in the spectral range of 750–950 nm. Besides, for the 1D and 3D CNN models, the datasets in the range of 460–660 nm and 750–950 nm were re-modeling, respectively, and the test results were consistent with the visualization results, indicating the effectiveness of the visualization methods. It can be used for wavelength selection. However, 1D CNN and 3D CNN models had differences in the regions of interest of the corresponding data sets. The reason may be that the 1D CNN model captured the information of the overall structural change of the blade and the information of the spectral reflectance change in the NIR spectral range. For the 3D CNN model, it could capture the changes in the spectral reflectance of the leaves in the Vis spectral range of the HSIs, but it was difficult to capture the NIR information about the changes in the internal chemical composition of the leaves.

In this study, only the infected leaves and the healthy leaves in cotton plants were studied. Since HSIs were obtained in a greenhouse, the interference factors affecting the HSIs were controllable. In the controllable environment, typical infected samples and healthy samples were obtained, and the difference between the two categories of HSIs was large. However, there are many uncontrollable factors in field conditions, which cause uncertainty and variations in samples. The differences between healthy and infected samples might not be so large, and a large number of field experiment samples should be studied in future research. Our study provided an initial assessment of pest detection in cotton, and provide a potential method for rapid and non-invasive pest detection. In future researches, the developed method will be validated and updated based on the in-field experiments.

CONCLUSION

In this study, the Vis/NIR hyperspectral imaging system (376–1044 nm) and machine learning methods were used to identify aphid infection in cotton leaves. Spectra, RGB images, and hyperspectral images containing a single leaf were used to build classification models. Spectra did not contain spatial information. However, the spectral information in the spectra was simple but rich, which was conducive to the learning of existing computing power and models, and had achieved excellent results. The RGB images and the hyperspectral images contained spatial information, and the characteristics of the spatial region that affect the classification results could be found. Compared with the RGB images, the hyperspectral images contained a lot of spectral information. The classification results of the 3D CNN used to identify aphid infection were better than 2D CNN and worse than 1D CNN. It was recommended that 1D CNN could be used to quickly and accurately identify aphid infection. In the visualization of 1D CNN, it was found that the important spectral regions of the spectra were concentrated in the Vis (460–660 nm) and NIR (750–950 nm) range. In the visualization of 2D CNN and 3D CNN, the spatial regions of cotton leaves changed after aphid infection were found. 2D CNN could be used to find aphid infection areas. 3D CNN combined features of 1D CNN and 2D CNN, it could be used to discover the aphid infection area while discovering important spectral regions. Based on the exploration of CNNs in multiple dimensions on aphid infection, the effects of CNNs in various dimensions were compared, which provided data reference for related scholars and provided new ideas for future research on pest infection.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.6084/m9.figshare.13668314.

AUTHOR CONTRIBUTIONS

TY and JL prepared materials and used the hyperspectral imaging system to obtain hyperspectral images. TY and WX wrote the manuscript. TY completed the development and implementation of methods. LD tested the results. WX, PG, and CZ designed the experiment. CZ made guidance for the writing of the manuscript. PG and CZ provided funding for this work. XL reviewed the initial design of the experiments and made comments. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China, grant numbers 61965014 and 61705195 and the Scientific and Technological Research Projects in Key Areas of Xinjiang Production and Construction Corps (XPCC), grant number 2020AB005.

ACKNOWLEDGMENTS

The authors want to thank Jingcheng Xu, a master student at the Agricultural College of Shihezi University in China, for providing guidance on the use of hyperspectral imaging systems.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.604510/full#supplementary-material

REFERENCES

Ahmad, M. N., Shariff, A. R. M., and Moslim, R. (2018). Monitoring insect pest infestation via different spectroscopic techniques. *Appl. Spectrosc. Rev.* 53, 836–853. doi: 10.1080/05704928.2018.1445094

Armstrong, J. S., Mbulwe, L., Sekula-Ortiz, D., Villanueva, R. T., and Rooney, W. L. (2017). Resistance to melanaphis sacchari (Hemiptera: Aphididae) in forage and grain sorghums. *J. Econ. Entomol.* 110, 259–265. doi: 10.1093/jee/tow261

Basati, Z., Jamshidi, B., Rasekh, M., and Abbaspour-Gilandeh, Y. (2018). Detection of sunn pest-damaged wheat samples using visible/near-infrared spectroscopy based on pattern recognition. *Spectrochim. Acta* 203, 308–314. doi: 10.1016/j.saa.2018.05.123

Cabral, C., Wollenweber, B., Antonio, C., Rodrigues, A. M., and Ravnskov, S. (2018). Aphid infestation in the phyllosphere affects primary metabolic profiles in the arbuscular mycorrhizal hyphosphere. *Sci. Rep.* 8:1. doi: 10.1038/s41598-018-32670-1

Cai, H., Zhu, L., and Han, S. (2018). ProxylessNAS: Direct neural architecture search on target task and hardware. *arXiv*. Preprint.

Cai, Y., Liu, X., and Cai, Z. (2020). BS-Nets: An End-to-End framework for band selection of hyperspectral image. *IEEE Trans. Geosci. Remote* 58, 1969–1984. doi: 10.1109/TGRS.2019.2951433

Canário, D. V. P., Figueiredo, E., Franco, J. C., and Guerra, R. (2017). Detecting early mealybug infection stages on tomato plants using optical spectroscopy. *Eur. J. Hortic. Sci.* 82, 341–348. doi: 10.17660/eJHS.2017/82.3.4

Chen, J., Fan, Y., Wang, T., Zhang, C., Qiu, Z., et al. (2018a). Automatic segmentation and counting of aphid nymphs on leaves using convolutional neural networks. *Agron.Basel* 8:8080129. doi: 10.3390/agronomy8080129

Chen, T., Zeng, R., Guo, W., Hou, X., Lan, Y., et al. (2018b). Detection of Infection in Cotton (Gossypium hirsutum L.) Caused by Aphids Using Leaf Level Hyperspectral Measurements. *Sens. Basel* 18:18092798. doi: 10.3390/s18092798

Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z., et al. (2019). Genome sequence of the corn leaf aphid (Rhopalosiphum maidis Fitch). *Gigascience* 8:giz033. doi: 10.1093/gigascience/giz033

Deng, L., Wang, Y., Han, Z., and Yu, R. (2018). Research on insect pest image detection and recognition based on bio-inspired methods. *Biosyst. Eng.* 169, 139–148. doi: 10.1016/j.biosystemseng.2018.02.008

Dubey, N. K., Goel, R., Ranjan, A., Idris, A., Singh, S. K., et al. (2013). Comparative transcriptome analysis of Gossypium hirsutum L. In response to sap sucking insects: Aphid and whitefly. *BMC Genomics* 14:241. doi: 10.1186/1471-2164-14-241

Feng, X., Zhan, Y., Wang, Q., Yang, X., Yu, C., et al. (2020). Hyperspectral imaging combined with machine learning as a tool to obtain high-throughput plant salt-stress phenotyping. *Plant J.* 101, 1448–1461. doi: 10.1111/tpj.14597

Gao, P., Xu, W., Yan, T., Zhang, C., Lv, X., et al. (2019). Application of Near-Infrared hyperspectral imaging with machine learning methods to identify geographical origins of dry Narrow-Leaved oleaster (Elaeagnus angustifolia) fruits. *Foods* 8:8120620. doi: 10.3390/foods8120620

Han, Z., and Gao, J. (2019). Pixel-level aflatoxin detecting based on deep learning and hyperspectral imaging. *Comput. Electron. Agr.* 164:104888. doi: 10.1016/j.compag.2019.104888

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas: CVPR).

Herron, G. A., and Wilson, L. J. (2017). Can resistance management strategies recover insecticide susceptibility in pests? A case study with cotton aphid Aphis gossypii (Aphididae: Hemiptera) in Australian cotton. *Austral. Entomol.* 56, 1–13. doi: 10.1111/aen.12236

Hough, A. R., Nechols, J. R., McCornack, B. P., Margolies, D. C., Sandercock, B. K., et al. (2017). The Effect of Temperature and Host Plant Resistance on Population Growth of the Soybean Aphid Biotype 1 (Hemiptera: Aphididae). *Environ. Entomol.* 46, 58–67. doi: 10.1093/ee/nvw160

Huang, L., Yang, L., Meng, L., Wang, J., Li, S., et al. (2018). Potential of Visible and Near-Infrared Hyperspectral Imaging for Detection of Diaphania pyloalis Larvae and Damage on Mulberry Leaves. *Sens. Basel* 18:18072077. doi: 10.3390/s18072077

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*. Preprint.

Jian, C., Gao, J., and Ao, Y. (2016). A new sampling method for classifying imbalanced data based on support vector machine ensemble. *Neurocomputing* 193, 115–122. doi: 10.1016/j.neucom.2016.02.006

Jin, J., and Wang, Q. (2016). Hyperspectral indices based on first derivative spectra closely trace canopy transpiration in a desert plant. *Ecol. Inform.* 35, 1–8. doi: 10.1016/j.ecoinf.2016.06.004

Kafeshani, F. A., Rajabpour, A., Aghajanzadeh, S., Gholamian, E., and Farkhari, M. (2018). Spatial distribution and sampling plans with fixed level of precision for citrus aphids (Hom., Aphididae) on two orange species. *J. Econ. Entomol.* 111, 931–941. doi: 10.1093/jee/tox380

Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S. (2019). A survey of the recent architectures of deep convolutional neural networks. *arXiv*. Preprint. doi: 10.1007/s10462-020-09825-6

Kozal, A. O., Teke, M., and Ilgin, H. A. (2013). "Comparative analysis of hyperspectral dimension reduction methods," in *Signal Processing and Communications Applications Conference*, (Turkey: IEEE).

Li, J., Luo, W., Wang, Z., and Fan, S. (2019). Early detection of decay on apples using hyperspectral reflectance imaging combining both principal component analysis and improved watershed segmentation method. *Postharvest Biol. Tec.* 149, 235–246. doi: 10.1016/j.postharvbio.2018.12.007

Liu, F. T., Ting, K. M., Yu, Y., and Zhou, Z. (2008). Spectrum of variable-random trees. *J. Artif. Intell. Res.* 32, 355–384. doi: 10.1613/jair.2470

Liu, G., He, J., Wang, S., Luo, Y., Wang, W., et al. (2016). Application of Near-Infrared hyperspectral imaging for detection of external insect infestations on jujube fruit. *Int. J. Food Prop.* 19, 41–52. doi: 10.1080/10942912.2014.923439

Lorenzo, P. R., Tulczyjew, L., Marcinkiewicz, M., and Nalepa, J. (2020). Hyperspectral band selection using Attention-Based convolutional neural networks. *IEEE Access* 8, 42384–42403. doi: 10.1109/ACCESS.2020.2977454

Lu, R., and Ariana, D. P. (2013). Detection of fruit fly infestation in pickling cucumbers using a hyperspectral reflectance/transmittance imaging system. *Postharvest Biol. Tec.* 81, 44–50. doi: 10.1016/j.postharvbio.2013.02.003

Ma, D., Hu, Y., Yang, C., Liu, B., Fang, L., et al. (2016). Genetic basis for glandular trichome formation in cotton. *Nat. Commun.* 7:10456. doi: 10.1038/ncomms10456

Malenovsky, Z., Albrechtova, J., Lhotakova, Z., Zurita-Milla, R., Clevers, J. G. P. W., et al. (2006). Applicability of the PROSPECT model for Norway spruce needles. *Int. J. Remote Sens.* 27, 5315–5340. doi: 10.1080/01431160600762990

Mandrell, C. T., Holland, T. E., Wheeler, J. E., Esmaeili, S. M. A., Amar, K., et al. (2020). Machine learning approach to raman spectrum analysis of MIA PaCa-2 pancreatic cancer tumor repopulating cells for classification and feature analysis. *Life Basel* 10:10090181. doi: 10.3390/life10090181

Moghadam, P., Ward, D., Goan, E., Jayawardena, S., and Hernandez, E. (2017). "Plant disease detection using hyperspectral imaging," in *2017 International Conference on Digital Image Computing: Techniques and Applications*, (Sydney: DICTA), doi: 10.1109/DICTA.2017.8227476

Moscetti, R., Haff, R. P., Stella, E., Contini, M., Monarca, D., et al. (2015). Feasibility of NIR spectroscopy to detect olive fruit infected by Bactrocera oleae. *Postharvest Biol. Tec.* 99, 58–62. doi: 10.1016/j.postharvbio.2014.07.015

Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., et al. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* 15:8. doi: 10.1186/s13007-019-0479-8

Paoletti, M. E., Haut, J. M., Plaza, J., and Plaza, A. (2019). Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm.* 158, 279–317. doi: 10.1016/j.isprsjprs.2019.09.006

Pelikan, M., and Goldberg, D. (1999). *BOA: The bayesian optimization algorithm*. Burlington: Morgan Kaufmann, doi: 10.1007/978-3-540-32373-0_3

Qiu, R., Yang, C., Moghimi, A., Zhang, M., Steffenson, B. J., et al. (2019). Detection of fusarium head blight in wheat using a deep neural network and color imaging. *Remote Sens. Basel* 11:11222658. doi: 10.3390/rs11222658

Rady, A., Ekramirad, N., Adedeji, A. A., Li, M., and Alimardani, R. (2017). Hyperspectral imaging for detection of codling moth infestation in GoldRush apples. *Postharvest Biol. Tec.* 129, 37–44. doi: 10.1016/j.postharvbio.2017.03.007

Rather, A. M., Mahato, S., Maji, K., Gogoi, N., and Manna, U. (2017). 'Reactive' nano-complex coated medical cotton: A facile avenue for tailored release of small molecules. *Nanoscale* 9, 16154–16165. doi: 10.1039/c7nr03990e

Sanchez-Matilla, R., Li, C. Y., Shamsabadi, A. S., Mazzon, R., and Cavallaro, A. (2020). Exploiting vulnerabilities of deep neural networks for privacy protection. *IEEE Trans. Multimed.* 22, 1862–1873. doi: 10.1109/TMM.2020.2987694

Signoroni, A., Savardi, M., Baronio, A., and Benini, S. (2019). Deep learning meets hyperspectral image analysis: A multidisciplinary review. *J. Imaging* 5:52. doi: 10.3390/jimaging5050052

Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv*. Preprint.

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Song, X., Hu, G., Luo, J., Feng, Z., Yu, D., et al. (2019). Fast SRC using quadratic optimisation in downsized coefficient solution subspace. *Signal Process.* 161, 101–110. doi: 10.1016/j.sigpro.2019.03.007

Stoltzfus, J. (2011). Logistic regression: A brief primer. *Acad. Emerg. Med.* 18, 1099–1104. doi: 10.1111/j.1553-2712.2011.01185.x

Sunoj, S., Sivarajan, S., Maharlooei, M., Bajwa, S. G., Harmon, J. P., et al. (2017). Identification and counting of soybean aphids from digital images using shape classification. *Trans. ASABE* 60, 1467–1477. doi: 10.13031/trans.12105

Szczepaniec, A. (2018). Interactive effects of crop variety, insecticide seed treatment, and planting date on population dynamics of sugarcane aphid (Melanaphis sacchari) and their predators in late-colonized sorghum. *Crop Prot.* 109, 72–79. doi: 10.1016/j.cropro.2018.03.002

Tan, M., and Le, Q. V. (2019). MixConv: Mixed depthwise convolutional kernels. *arXiv*. Preprint.

Thomas, S., Kuska, M. T., Bohnenkamp, D., Brugger, A., Alisaac, E., et al. (2018). Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective. *J. Plant Dis. Protect.* 125, 5–20. doi: 10.1007/s41348-017-0124-6

Thorpe, P., Cock, P. J. A., and Bos, J. (2016). Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *BMC Genomics* 17:6. doi: 10.1186/s12864-016-2496-6

Tian, Y., Cheng, Y., Wang, X., and Liu, S. (2015). Recognition method of insect damage and stem/calyx on apple based on hyperspectral imaging. *Trans. Chin. Soc. Agricult. Engine.* 31, 325–331. doi: 10.3969/j.issn.1002-6819.2015.04.046

Veit, A., and Belongie, S. (2020). Convolutional networks with adaptive inference graphs. *Int. J. Comput. Vis.* 128, 730–741. doi: 10.1007/s11263-019-01190-4

Wang, C., Ma, N., Ming, Y., Wang, Q., and Xia, J. (2019a). Classification of hyperspectral imagery with a 3D convolutional neural network and J-M distance. *Adv. Space Res.* 64, 886–899. doi: 10.1016/j.asr.2019.05.005

Wang, D., Xie, N., Yi, S., Liu, C., Jiang, H., et al. (2018). Bioassay-guided isolation of potent aphicidal Erythrina alkaloids against Aphis gossypii from the seed of Erythrina crista-galli L. *Pest Manag. Sci.* 74, 210–218. doi: 10.1002/ps.4698

Wang, L., Zhang, J., Liu, P., Choo, K. R., and Huang, F. (2017a). Spectral-spatial multi-feature-based deep learning for hyperspectral remote sensing image classification. *Soft. Comput.* 21, 213–221. doi: 10.1007/s00500-016-2246-3

Wang, P., Tu, Y., and Tseng, Y. J. (2019b). PgpRules: A decision tree based prediction server for P-glycoprotein substrates and inhibitors (vol 35, pg 4535, 2019). *Bioinformatics* 35:4535. doi: 10.1093/bioinformatics/btz546

Wang, R., Zhang, J., Dong, W., Yu, J., Xie, C. J., et al. (2017b). A crop pests image classification algorithm based on deep convolutional neural network. *Telkomnika* 15, 1239–1246. doi: 10.12928/TELKOMNIKA.v15i3.5382

Wilson, L. J., Whitehouse, M. E. A., and Herron, G. A. (2018). The management of insect pests in australian cotton: An evolving story. *Annu. Rev. Entomol.* 63, 215–237. doi: 10.1146/annurev-ento-020117-043432

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv*. Preprint.

Zhan, Y., Tian, H., Ji, X., and Liu, Y. (2020). Myzus persicae (Hemiptera: Aphididae) infestation increases the risk of bacterial contamination and alters nutritional content in storage Chinese cabbage. *J. Sci. Food Agr.* 100, 3007–3012. doi: 10.1002/jsfa.10331

Zhang, B., Zhao, L., and Zhang, X. (2020a). Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sens. Environ.* 247:111938. doi: 10.1016/j.rse.2020.111938

Zhang, J., Lu, C., Wang, J., Yue, X., Lim, S., et al. (2020b). Training convolutional neural networks with Multi-Size images and triplet loss for remote sensing scene classification. *Sens. Basel* 20:20041188. doi: 10.3390/s20041188

Zhang, S., Li, X., Zong, M., Zhu, X., and Wang, R. (2018). Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans. Neur. Net. Lear.* 29, 1774–1785. doi: 10.1109/TNNLS.2017.2673241

Zhao, W., and Du, S. (2016). Spectral-Spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Trans. Geosci. Remote* 54, 4544–4554. doi: 10.1109/TGRS.2016.2543748

Zhong, Z., Li, J., Luo, Z., and Chapman, M. (2018). Spectral-Spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE T. Geosci. Remote* 56, 847–858. doi: 10.1109/TGRS.2017.2755542

Zhou, B., Khosla, A., Lapedriza, À, Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. *arXiv*. Preprint.

Zhou, R., Jin, J., Li, Q., Su, Z., Yu, X., et al. (2019). Early detection of magnaporthe oryzae-Infected barley leaves and lesion visualization based on hyperspectral imaging. *Front. Plant Sci.* 9:1962. doi: 10.3389/fpls.2018.01962

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yan, Xu, Lin, Duan, Gao, Zhang and Lv. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 22 February 2021
doi: 10.3389/fpls.2021.591333

DAM: Hierarchical Adaptive Feature Selection Using Convolution Encoder Decoder Network for Strawberry Segmentation

*Talha Ilyas[1], Muhammad Umraiz[1], Abbas Khan[1] and Hyongsuk Kim[1,2]**

[1] Division of Electronic Engineering, Intelligent Robots Research Center, Jeonbuk National University, Jeonju, South Korea,
[2] Division of Electronic and Information Engineering, Jeonbuk National University, Jeonju, South Korea

Autonomous harvesters can be used for the timely cultivation of high-value crops such as strawberries, where the robots have the capability to identify ripe and unripe crops. However, the real-time segmentation of strawberries in an unbridled farming environment is a challenging task due to fruit occlusion by multiple trusses, stems, and leaves. In this work, we propose a possible solution by constructing a dynamic feature selection mechanism for convolutional neural networks (CNN). The proposed building block namely a dense attention module (DAM) controls the flow of information between the convolutional encoder and decoder. DAM enables hierarchical adaptive feature fusion by exploiting both inter-channel and intra-channel relationships and can be easily integrated into any existing CNN to obtain category-specific feature maps. We validate our attention module through extensive ablation experiments. In addition, a dataset is collected from different strawberry farms and divided into four classes corresponding to different maturity levels of fruits and one is devoted to background. Quantitative analysis of the proposed method showed a 4.1% and 2.32% increase in mean intersection over union, over existing state-of-the-art semantic segmentation models and other attention modules respectively, while simultaneously retaining a processing speed of 53 frames per second.

Keywords: semantic segmentation, convolutional neural network, encoder-decoder architecture, fruit segmentation, channel attention, spatial attention, segmentation grad-cam, autonomous harvesting

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Thiago Teixeira Santos, Brazilian Agricultural Research Corporation (EMBRAPA), Brazil
Lucas Costa, University of Florida, United States

***Correspondence:**
Hyongsuk Kim
hskim@jbnu.ac.kr

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *04 August 2020*
Accepted: *25 January 2021*
Published: *22 February 2021*

Citation:
Ilyas T, Umraiz M, Khan A and Kim H (2021) DAM: Hierarchical Adaptive Feature Selection Using Convolution Encoder Decoder Network for Strawberry Segmentation. Front. Plant Sci. 12:591333. doi: 10.3389/fpls.2021.591333

INTRODUCTION

Since the evolution of deep convolutional neural networks (DCNNs) from neural networks (Krizhevsky et al., 2012), machine learning has shown unprecedented performance on a number of machine vision and pattern recognition tasks such as image classification (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; He et al., 2016; Szegedy et al., 2016; ur Rehman et al., 2018), object detection and localization (Ren et al., 2015; Redmon et al., 2016; He et al., 2017; Nizami et al., 2020), and semantic and instance segmentation (Long et al., 2015; Ronneberger et al., 2015; Badrinarayanan et al., 2017; Bolya et al., 2019). Recently, unsupervised algorithms are also gaining popularity (Epifanio and Soille, 2007; Zhao and Kit, 2011; Xia and Kulis, 2017; Ilyas et al., 2020) due to their certain advantages over supervised ones (Huang et al., 2017; Lin et al., 2017c;

Ilyas et al., 2020). Moreover, deep learning has also demonstrated unparalleled performance in the field of bioinformatics and computational biology (Wahab et al., 2019, 2020; Park et al., 2020).

Where DCNNs have found several intuitive applications in various fields in our everyday lives, they are also being used in agriculture for autonomous harvesting and seeding. A lot of work has been done in literature in this regard, like crop and weed classification (Dyrmann et al., 2016, 2017; Grinblat et al., 2016; Kussul et al., 2017), plant detection (Mohanty et al., 2016; Khan et al., 2020), land cover classification (Ienco et al., 2017; Kussul et al., 2017), and crop disease identification (Fuentes et al., 2018). Just like any other machine vision task, the implementation of DCNNs in agriculture comes with its own set of problems. By the same token, the real time segmentation and detection of strawberries in an unconstrained farm environment is a challenging task, as strawberries usually grow in clusters and are occluded by leaves, branches, and other fruits. Due to different light intensities sometimes backgrounds and fruits have the same texture and color. These commonly occurring phenomena in farms makes the task more difficult and reduces the accuracy of DCNNs (Sa et al., 2016; Xiong et al., 2019).

Strawberries are some of the most highly valued crops as they give the best yield under sheltered environments, and thus have a very high production cost (Sa et al., 2016). The most crucial time for strawberry crop is harvesting time because the fruit becomes overripe quickly and if picking gets behind it effects the whole crop. Moreover, hiring skilled laborers in horticulture accounts for most of the cultivation cost. This crop also needs intensive post-harvest care (Guerrero et al., 2017). Because of all these expenses, horticulture industries in general are bound to have small profit margins. In some regions, labor cost makes up more than half of the total production cost, e.g., 60% in Norway (Xiong et al., 2019). Furthermore, there is a decline in interest of joining the agriculture industry among the new generation of workers (Adhikari et al., 2019). Under all these challenges the food industry must keep up with the demands of the ever-growing population.

To overcome such problems, one potential solution is autonomous harvesting as it can reduce labor cost to a minimum and increase the crop yield quality by timely harvesting. Due to outstanding performances of DCNNs in computer vision tasks, robotics and unmanned systems are now faster and more reliable than ever. Which in turn has allowed their adoption into many real-life applications like the detection of crop rows, weeds, and seeding beds in fields of maize and rice (Guerrero et al., 2017; Adhikari et al., 2019; Ma et al., 2019).

In this work, we proposed a DCNN named Straw-Net, to precisely segment and classify the fruits into specified classes in real time. In the case of strawberries, this is difficult to achieve because they usually grow in clusters and within the same cluster, and tend to have different sizes, shapes, and colors. In some cases, severe occlusion may also occur which renders the fruit almost invisible. By taking all these shortcomings into account, we designed an adaptive self-contained attention mechanism (i.e., dense attention module, DAM) for our network, which is capable of learning both channel and spatial interdependencies and can learn 'what' is important and 'where' to put more focus. We verify the efficacy of the proposed attention module quantitatively via benchmark metrics and qualitatively via modified Grad-CAM (Selvaraju et al., 2017). Grad-CAM is usually used for classification models, however, this paper extends its applicability to segmentation models. The dataset used in this paper is collected from different strawberry farms across the Republic of Korea under different lighting and weather conditions for better generalization of real-life scenarios. Our main contributions are listed below:

- We propose a single attention module (DAM) for both channel and spatial attention, and a parallel dilated convolution module (PDC) for aggregating multi-scale context.
- We validate the effectiveness of DAM and PDC by ample ablation experiments.
- We propose an optimal location for integrating our attention module in any existing network and compare results with other existing attention mechanisms.
- We propose a technique for visual interpretation of segmentation networks by modifying Grad-CAM.
- A new dataset for the semantic segmentation of strawberries is introduced, consisting of four classes depending on the ripeness level of fruit as shown in **Figures 1**, **2** (see section "Materials and Methods" for details).

RELATED WORK

Autonomous Harvesting

A great deal of work has been done in the field of autonomous harvesting of fruits using various classical image analysis and pattern recognition techniques (Yamamoto et al., 2010; Feng et al., 2012; Ouyang et al., 2012; Qingchun et al., 2012; Zhang et al., 2012; Hayashi et al., 2013, 2014). One autonomous strawberry harvester (ASH) in particular AGROBOT SW6010 (Agrobot, 2020) has gained a lot of popularity. It uses morphological, color, and shape analysis for the identification and selection of strawberries to harvest, and uses 24 robotic arms to perform harvesting. FFRobotics (Kahani, 2017) have introduced the FFBot for harvesting apples, which needs a human supervisor to control and monitor the harvesting process. More recently, Hravest CROO (Harvest Croo, 2021) has introduced an ASH which divides the fruit picking into three steps, (a) grab the leaves, (b) 3D inspection of plant, and finally (c) pick the fruit. This simple robotic framework allows them to increase harvesting speed. Autonomous harvesters are also gaining popularity in other areas of precision agriculture like weeding. NAIO Technologies (Barthes, 2010) has introduced multiple autonomous weeding robots like OZ, TED, and DINO for easier vegetable weeding on large-scale farms.

The performance and accuracy of any autonomous harvester relies heavily on the perception system used and how the visual information obtained is being processed. An earlier era of vision-based harvesters used monocular devices to obtain 2D visual

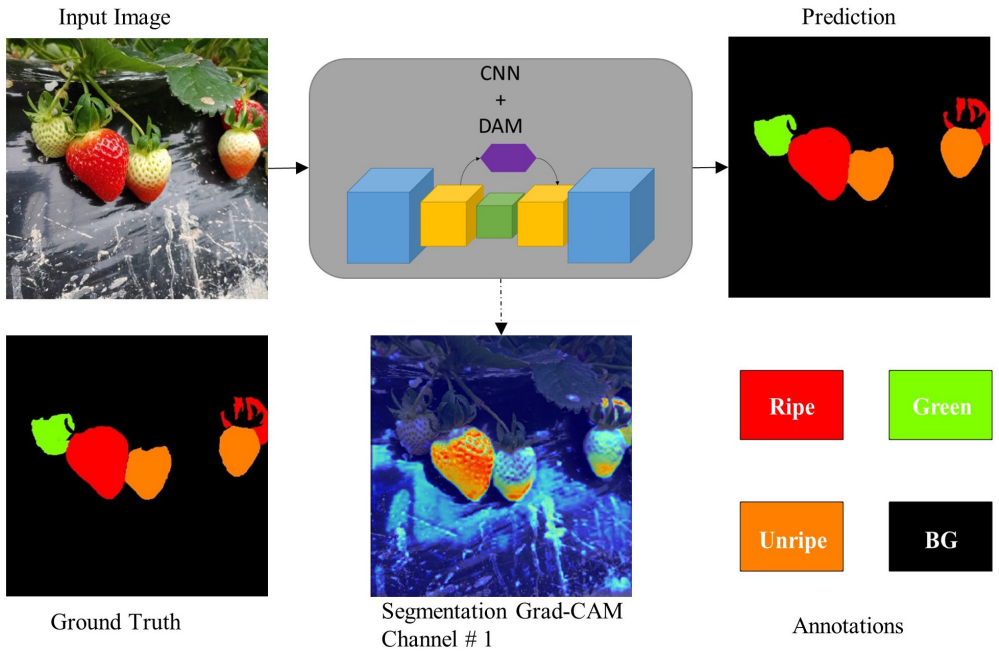

FIGURE 1 | Overview of the proposed DCNN and the dataset with corresponding annotations. Seg-Grad-CAM represents the attention maps for ripe class (i.e., channel # 1).

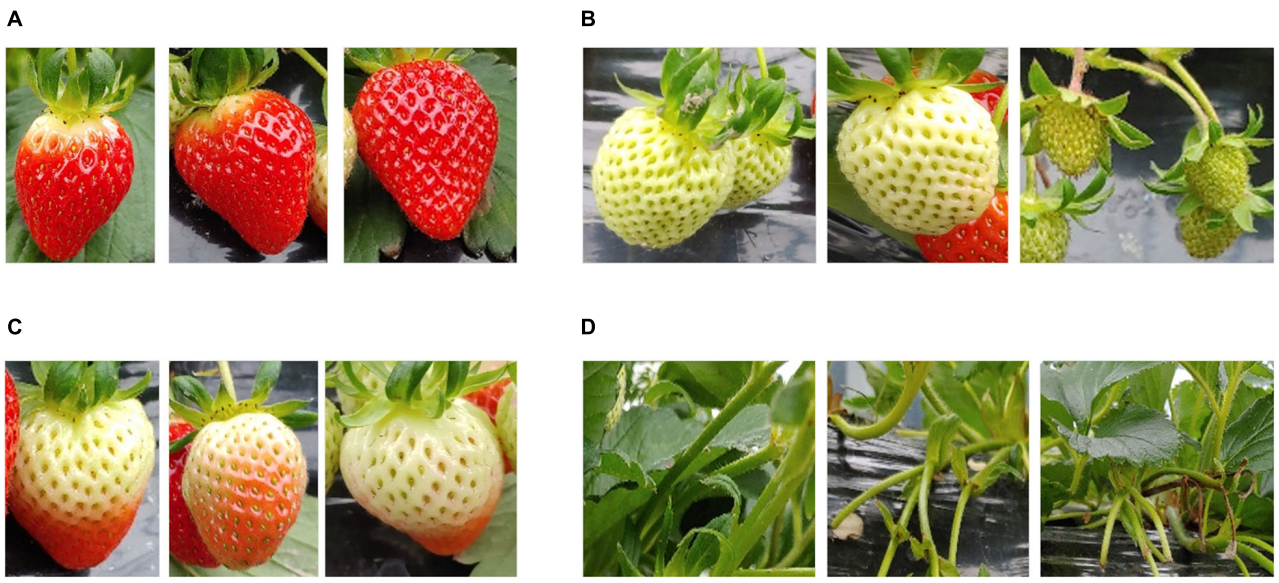

FIGURE 2 | A few representative instances belonging to each class (stage of ripeness) in the strawberry segmentation (SS1K) dataset. **(A)** Ripe; edible quality, **(B)** green; not ready for harvest, **(C)** unripe; export quality, **(D)** background.

data, e.g., Grand d'Esnon et al. (1987) classified fruits based on their texture and geometry and Edan et al. (2000) used two monochromatic cameras to produce a stereo-like effect for better localization of melons. Similarly, Yamamoto et al. (2014) introduced a stationary strawberry picking mechanism which used three different monochromatic light sources for the coloration measurement and spectral analysis of fruits and leaves to precisely localize the fruit for picking. With this mechanism they achieved an effective yield rate of 67%. But these monochromatic vision systems were highly susceptible to light intensity changes. Later, some works used stereovision to obtain 3D map of fruits via triangulation (Sun et al., 2011). A predecessor of AGROBOT SW6010 used stereo RGB-D images for tomato harvesting (Buemi, 1995; Buemi et al., 1996). Using

RGB-D images obtained via a Binocular-stereo vision camera, Ge et al. (2019) constructed a 3D point cloud to localize the pickable fruit. They used a Mask-RCNN as a backbone of their computer vision-based control to classify the strawberry into two classes, i.e., ripe and unripe. Following this pipeline they were able to improve the picking accuracy to 74%. Similarly, for recognizing clustered tomatoes and classifying them into overlapping and adhering regions, Xiang et al. (2014) used stereovision to obtain a depth-map, the reported accuracy for clustered tomato detection was 87.9%. Sensor calibration plays a vital role in the performance of stereovision systems. Recently, laser-based distance measuring systems (LiDAR) and spectral imaging are also doing wonders in precision agriculture. Zhang et al. (2015) combined computer vision with near-infrared structured lighting, and using a single multispectral camera was able to reconstruct the 3D surface of the apple for calyx and stem recognition. The results showed a 97.5% average accuracy.

These aforementioned ASHs rely heavily on classical mathematical algorithms; Ouyang et al. (2012) introduced a pipeline consisting of a series of image preprocessing and denoising techniques and then used the optimal thresholding (Otsu) algorithm for strawberry segmentation. Zhao et al. (2016) used a grayscale co-occurrence matrix to extract features of fruits from various color spaces. Whereas, different color spaces have different properties and their own application domain. Wei et al. (2014) used an OHTA color space, a modified version of the Otsu algorithm, to achieve an impressive detection accuracy of more than 95%. But it had a major drawback of not being able to detect green-colored fruits. Similarly, Qingchun et al. (2012) used a HSV color space to extract features for strawberry harvesting, results showed a 86% successful harvest rate. All the above-mentioned methods can detect fruits under controlled environments, but the detection accuracy drops when illumination changes. Moreover, depending on the variation in orientation, size, and shape of fruit, these methods require a lot of parameter tuning (Durand-Petiteville et al., 2017). In short, although several of the ASH prototypes have been developed to segment and classify fruits accurately in real time, their performance remains susceptible to unconstrained environments. This is where machine learning comes in.

Agricultural farms are unconstrained natural environments or semi-constrained at very best. Machine learning has found intuitive applications in many fields, because of its adaptive learning ability, like in healthcare (Ronneberger et al., 2015; Işın et al., 2016; Kauanova et al., 2017), autonomous driving (Fujiyoshi et al., 2019; Hofmarcher et al., 2019; Imai, 2019), and weed and crop detection (Grinblat et al., 2016; Mohanty et al., 2016; Dyrmann et al., 2017; Kussul et al., 2017; Fuentes et al., 2018). But very little work has been done in detecting fruits and classifying them according to their ripeness level. Lamb and Chuah (2018) used a single-stage detector SSD (Liu et al., 2016) to detect strawberries and attained a maximum average precision of 87.7%, but were not able to achieve real-time performance (see section "Real-Time Performance Barrier") even after using various network compression techniques. Bargoti and Underwood (2017) proposed an image processing framework using a simple CNN and a multi-scale multi-layer perceptron (ms-MLP) to detect and count apples, with an F1-score of 85.8%. Their algorithm was a multi-stage setup which used watershed and circular Hough transform to detect the individual fruits. Hence falling short of real-time performance. Chen et al. (2019) used a faster region-based convolutional neural network (F-RCNN) (Ren et al., 2015) for predicting strawberry production rate using aerial farm images. Sa et al. (2016) presented an approach for fruit detection in field farms using an F-RCNN and showed its generalization to many different farm fields. Moreover, Yu et al. (2019) combined a Mask-RCNN with a feature pyramid network (FPN) for better feature extraction, to detect mature strawberries (one class) with a precision rate of 95%, but were not able to break the real-time performance barrier (see section "Real-Time Performance Barrier"). Whereas our proposed encoder-decoder based CNN is able to explicitly detect and classify fruits according to specified ripeness levels while still maintaining a processing speed of 53 fps on standard resolution images.

Semantic Segmentation

Since the dawn of fully convolutional networks (FCNs) (Long et al., 2015) semantic segmentation has gained a lot of popularity. Following the main idea of embedding low contextual information in a progressive manner to preserve spatial and temporal information, a lot of encoder-decoder architecture has been introduced in literature. Deconv-Nets (Noh et al., 2015) introduced transposed convolution called deconvolution, for learning the upsampling process. SegNets (Badrinarayanan et al., 2017) introduced unpooling (i.e., inverse of pooling) to upsample the score maps in a gradual way. To remedy the loss of localization information by the subsequent downsampling of feature maps, U-net (Ronneberger et al., 2015) proposed skip-connections between the encoder and decoder to preserve spatial information. Further, the intermediate layers were exploited by RefineNet (Lin et al., 2017a) with skip-connections, which uses multipath refinement via different convolutional modules to get final predictions. Global Convolutional Network (Peng et al., 2017) tried to increase the receptive field by factorizing large kernels into smaller ones to get global contextual embeddings. PSP-Net (Zhao et al., 2017) used spatial pyramid pooling at different scales, and Deeplab (Chen et al., 2017) used atrous convolutions with different dilation rates for exploiting multi-scale information. Contrary to previous works that exploited intermediate layers by modifying identity skip-connections (Lin et al., 2017a; Peng et al., 2017) and those that use contextual multi-scale embedding for context gathering, our proposed network integrates the representational power of both of these types of networks to achieve better segmentation results.

Dilated Separable Convolution

More recently, networks like Dilated ResNet (DRN) (Yu et al., 2017) used dilated convolutions (Yu and Koltun, 2015) to increase the valid receptive field size while still maintaining the same computational cost (i.e., number of parameters and FLOPs). Furthermore, Deeplab-v3+ (Chen et al., 2018) combined dilated convolution with depth-wise separable convolution (Chollet, 2017). By doing so they achieved a significant performance

boost while keeping the model complexity to a minimum. These convolutions have been adopted by many recent algorithms (Jin et al., 2014; Wang et al., 2016; Howard et al., 2017; Zhang et al., 2018). In our network, we have also used the dilated separable convolution for better performance.

Attention Mechanism

Attention plays a vital role in human perception (Rensink, 2000; Corbetta and Shulman, 2002). As a matter of fact, neurons present in the primary visual cortex of cats (Hubel and Wiesel, 1962) have inspired the construction of DCNNs (LeCun et al., 1989). Neurons in the human visual system do not process the whole semantic scene at once. Instead the neurons try to process the scenery in a sequence and they adaptively focus on only the salient features of the scenery in front of them (Woo et al., 2018).

Recent algorithms have also tried to equip DCNNs with such attention mechanisms to improve their performance (Lin et al., 2017b; Shen et al., 2018). More recently, Fu et al. (2019) proposed a self-attention mechanism for integrating local and global semantic features. Their mechanism consisted of two modules, one for position attention (PAM) and one for channel attention (CAM). Because of heavy matrix multiplications, both modules were far too computationally expensive. Whereas, squeeze and excite (SE) networks (Hu et al., 2018) recalibrated the feature maps depending upon their importance, while keeping the computational overhead to a minimum. Although in Hu et al. (2018), the authors implicitly refer to the SE module as an attention mechanism, this can be explicitly considered as one, as shown by Park et al. (2018) and Woo et al. (2018). Recently, a block attention module (BAM) (Park et al., 2018) and convolution block attention module (CBAM) (Woo et al., 2018) achieved a significant performance boost in an ImageNet-1K classification challenge by adding spatial attention to SE modules. These aforementioned modules also consisted of two separate blocks for generating channel and spatial attention. In contrast to these works, we extend the use of attention mechanisms to the segmentation task. Moreover, different from existing works, instead of using two separate blocks for channel and spatial attention, we propose one block for both tasks, to avoid computational overhead and reduce inference time. We propose a gating mechanism to control the flow of multi-scale information from different stages of the backbone network (encoder) to suitable upsampling stages of the decoder. By doing so, we are able to achieve better category specific attention masks. Our adaptive self-contained attention mechanism can learn both channel and spatial interdependencies and can dynamically emphasize or suppress features according to their importance. Detailed ablation experiments verify the effectiveness of our module (see section "Results and Discussion").

MATERIALS AND METHODS

Image Acquisition

Strawberry images were collected from several strawberry farms across Jeonju-si District, Jeollabuk-do, Republic of Korea during the growing season (2019). All the strawberry farms adopted a hedgerow planting system (Strawberry, 2020) as shown in **Figure 3**. The data acquisition was carried out at a distance of 40 cm, using a 24.1 MPx Canon EOS-200D-based platform with a CMOS sensor. We chose this distance so that the device could capture sufficiently large scenery for processing, and at this distance ASH would be able to perform suitable target searching and harvesting. During different time periods and under varying weather and lighting conditions, we acquired 1500 images. Images were stored in the JPEG format and all had a resolution of 6288 × 4056 pixels. We stored the data in high resolution to avoid being limited in available resolution at later processing stages.

Dataset Construction and Annotation

We started with primary data filtering and removed the images which were blurred or contained no strawberry fruit at all. After the primary filtering step, we ended up with 1000 unlabeled images. Out of the total 1000 unlabeled images we randomly selected 750 images for training, 100 for validation, and 150 for testing.

Then with the help of experts in the strawberry harvesting field, we divided the strawberry fruits into four classes depending on the ripeness level. We labeled them as follows: (a) ripe; (edible quality) ready for harvesting, (b) green; not ready for harvesting, (c) unripe; (export quality) that can be harvested if the farm had to export the strawberries to far away destinations, and (d) background. Some representative instances belonging to each class (stage of ripeness) are shown in **Figure 2**. According to field experts and the Food and Agriculture Organization (FAO, 2020), a strawberry which is less than 70% matured should be considered as export quality. Because any more than that and there is a chance that the fruit may rot over long journeys. So, one might say that labeling the unripe class is somewhat intuitive. After deciding the ripeness level, we labeled the images as shown in **Figure 1**. There also exists a data imbalance between the classes, such that per batch there are a greater number of ripe and green strawberries than unripe ones, as shown in **Figure 4**. We will discuss this problem of data

FIGURE 3 | Data acquisition process.

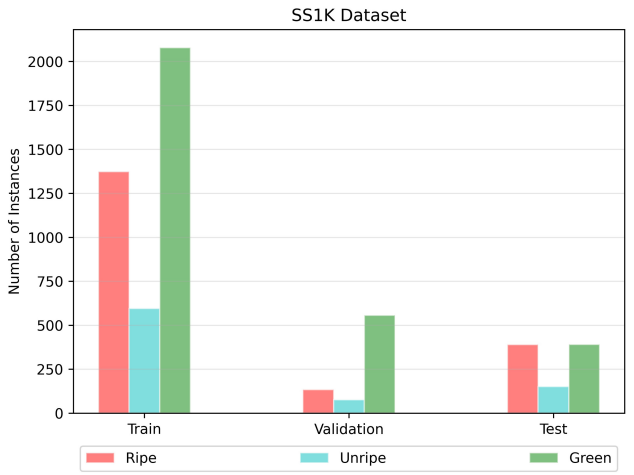

FIGURE 4 | The number of instances of each class in the strawberry segmentation (SS1K) dataset.

imbalance in the performance analysis (see section "Results and Discussion"). From this point onward for ease of notation we will call this strawberry segmentation dataset SS1K (1K for the total number of samples).

NETWORK ARCHITECTURE

First, we will describe the backbone of our architecture, i.e., the encoder part and the blocks used within, and then we will discuss the design considerations for our attention module. Finally, we will describe the decoder design choice and how to integrate the attention module in any existing network.

Encoder Design Considerations

A modified FCN for real-time segmentation of strawberry fruit, named Straw-Net is shown in **Figure 5A**. The encoder consists of SE-ResNet (Hu et al., 2018)-like blocks, with a few modifications. The SE-ResNet block consists of two parts, one being the ResNet bottleneck and the other being the SE-module as shown in **Figure 5B**. In the ResNet bottleneck, instead of using simple convolution, we decided to use the dilated separable convolution, which is a combination of dilated and depth-wise separable convolution (Chen et al., 2018). It allows the network designer to freely control the feature map's size and filter's effective receptive field (ERF), while significantly reducing the network computational cost. Depth-wise separable convolution disentangles the normal convolution into a depth-wise (or channel-wise) convolution followed by a point-wise convolution. This decomposition allows the DCNN to achieve better performance with much fewer parameters. In dilated convolutions, ERF can be easily changed by changing the dilation rate 'd_i' (Yu et al., 2017), where normal convolution is a special case of dilated convolution with $d = 1$. Increasing the ERF at each stage of the network helps the convolutional filters to aggregate multi-scale contextual information more efficiently.

After processing the features by bottleneck layers, next these feature-maps are passed through the SE-module (Hu et al., 2018), shown in **Figure 5B**. SE-modules recalibrate the feature maps by obtaining their channel-wise statistics via global average pooling (GAP). The GAP outputs a vector of size n, where n is same as the number of filter channels. Then this vector is passed through a multi-layer perceptron (MLP) to obtain a weighing vector of size n. This vector is then used to adaptively emphasize or suppress the feature maps according to their importance. For more details about SE-modules, we refer interested readers to Hu et al. (2018). Moreover, skip-connection allows for uninterrupted gradient flow to the earlier layers for better training.

Data in raw images are mostly redundant so a large kernel size with high stride can be used to process the raw image and make it ready for deeper layers to process. Using a high stride also reduces the dimensions which will in turn reduce the computational overhead (Hasanpour et al., 2016; He et al., 2016). Keeping that in mind, firstly the image is passed through a normal convolution layer with 16 filters of size 7×7 and stride 2. Now this processed input is passed through the successive SE-ResNet blocks as shown in **Figure 5A**. Each convolutional layer in SE-ResNet is followed by a batch normalization (BN) and ReLu activation, unless explicitly stated.

The network backbone consists of six stages. All stages consist of two SE-ResNet blocks. Among those, the first three stages $S_{i\in\{1,2,3\}}$ are followed by subsequent pooling operations for reducing feature map size. In the next stages $S_{i\in\{4,5,6\}}$ we do not perform a pooling operation. Because, after using the stride = 2 in the first layer and the three subsequent pooling operations in the first three stages, the extracted feature map size is 16 times smaller than the input at the end of the encoder. Reducing it further will result in the loss of a lot of useful localization information, making the decoding process more difficult. In the first two stages, the dilation rate is set to $d = 1$, and in the next three stages, the dilation rate is doubled for every next stage, i.e., $d_{i\in\{3,4,5\}} = \{2,4,8\}$ for stages $S_{i\in\{3,4,5\}}$. The final stage S_6 again has a dilation rate of $d_6 = 1$ to avoid the gridding artifact (Yu and Koltun, 2015).

Parallel Dilated Convolution Module (PDC)

We go deeper into the DCNNs, even though the deeper layers have a large theoretical receptive field (TRF) but their effective receptive field (ERF) is much smaller than the theoretical one as shown by Zhou et al. (2014). Information regarding global context plays a vital role in scene segmentation (Peng et al., 2017; Zhao et al., 2017). So, at the end of the encoder we probe the feature maps of the last stage (i.e., S_6) for aggregating global and sub-region context by incorporating the PDC module shown in **Figure 6B**. PDC acts as a hierarchical global module prior to using dilated convolution at different dilation rates to extract global contextual information from S_6's feature maps at multiple scales. We perform detailed ablation experiments to show the effectiveness of PDC and compare it with other multi-scale feature aggregation modules of Zhao et al. (2017) and Chen et al. (2018) in Section "Ablation Study for (PDC)."

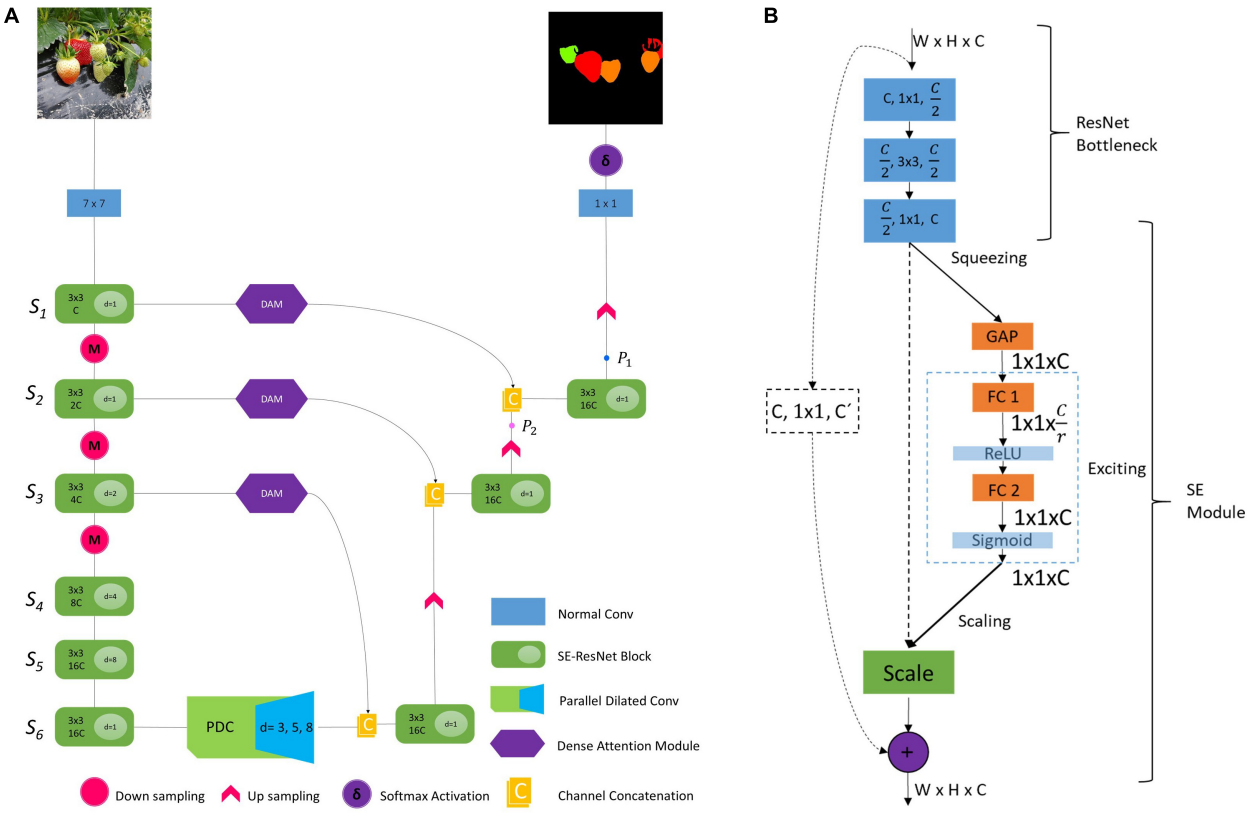

FIGURE 5 | (A) Straw-Net, complete architecture. Here S_i represents different stages of network $i \in \{1,2,\ldots 6\}$. c represents the number of channels inside each SE-ResNet block (we set $c = 16$), and d is the dilation rate. For details on points P_1 and P_2 see Section "Ablation Study for DAM." **(B)** SE-ResNet module; in the first three convolutional blocks, the first value represents the number of input feature maps, the second value represents the kernel size, and the third value represents the number of output channels. The dashed arrows represent the identity mapping. GAP, FC, and r represent global average pooling, densely connected layers, and reduction ratio, respectively.

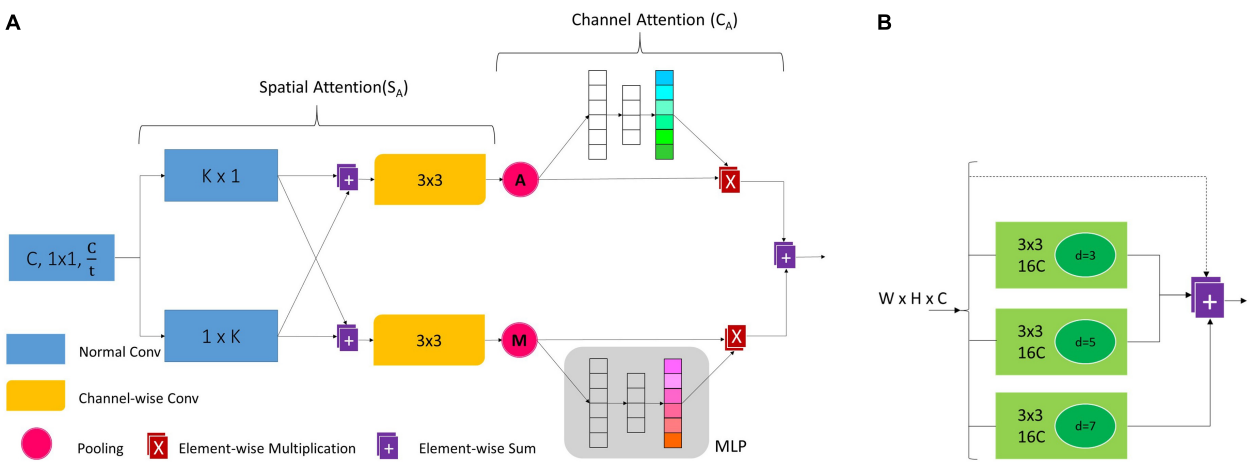

FIGURE 6 | (A) Dense attention module, here K is kernel size and t is the reduction factor. **(B)** Parallel dilated convolution module, here C represent the no. of feature maps and d is the dilation rate.

Dense Attention Module (DAM)

To control the flow of information from encoder to decoder via skip-connections we incorporate the dense attention modules on skip-connections. We found that this is the best location to make the most out of these attention modules. These modules perform *'feature surgery'* on the feature maps coming from the encoder

which are rich in localization information. They help in efficient feature fusion between encoder feature maps (which focus on 'where' the target object is) and decoder feature maps (which focus on 'what' the target object is). The whole operation can be summarized as follows;

$$F_{s_i}^{''} = DAM\left(F_{s_i}\right) \tag{1}$$

Given an input feature map $F_{s_i} \in \mathbb{R}^{W_{S_i} \times H_{S_i} \times C_{s_i}}$ from stage S_i of the encoder, the DAM computes the refined feature map $F_{s_i}^{''} \in \mathbb{R}^{W^{'} \times H^{'} \times C^{'}}$ to be concatenated with decoder feature maps. Usually, the low-level feature maps have a large number of channels (e.g., 128 or 256). So, DAM first reduces the number of channels of the corresponding low-level feature maps by a factor of t such that $F_{s_i} \in \mathbb{R}^{W_{s_i} \times H_{s_i} \times C^{'}}$ where $C^{'} = \frac{C_{s_i}}{t}$. To avoid the suppression of information in rich decoder feature maps by the low-level encoder feature maps, we set $t = 4$ in our experiments. The contextual information is aggregated using a large kernel size $f_n^{K \times K}$. To reduce the number of computations and inference time we decompose the one $f_n^{K \times K}$ filter into two parallel $f_n^{1 \times K}$ and $f_n^{K \times 1}$filters. Here 'n' represents the normal convolutional filter and we set $K = 7$. Then the results of both these convolutions are added in their respective parallel branches as shown in **Figure 6A**. Next, we pass these feature maps through a depth (channel)-wise convolutional layer (in their respective branches) of filter size 3×3, i.e., $f_c^{3 \times 3}$ where 'c' represents the depth-wise convolutional filter. In the depth-wise convolution, one filter convolves spatially on only one feature map making the output feature maps spatially enhanced as shown by Gao et al. (2018). So, the channel specific spatial attention for both branches is computed as,

$$F_{s_i}^{S_A} = f_c^{3 \times 3}\left(f_n^{1 \times K}\left(F_{S_i}\right) + f_n^{K \times 1}\left(F_{S_i}\right)\right) \tag{2}$$

Here, the superscript S_A refers to spatial attention in the top (S_t) and bottom (S_m) branch. Next, this channel-specific spatial attention is recapitulated using both average and max pooling operations generating different feature descriptors.

$$F_{S_i}^{avg} = AvgPool\left(F_{s_i}^{s_t}\right) \tag{3}$$

$$F_{S_i}^{max} = MaxPool\left(F_{s_i}^{s_m}\right) \tag{4}$$

Here, $F_{s_i}^{pool} \in {}'\mathbb{R}^{W^{'} \times H^{'} \times C^{'}}$ where W' and H' represent the pooled (average and max) width and height of the feature maps. Unlike previous work (Hu et al., 2018), we argue that instead of using only average pooling, exploiting both pooling operations to gather distinct global characteristics helps the module to infer distinct channel-wise attention in both branches independently. Exploiting both average and max pooling features greatly improves the network's representational power (see section "Ablation Study for DAM"). After pooling, these 3D feature descriptors are passed through an MLP to obtain a 1D descriptor vector $F_{s_i}^{C_A} \in \mathbb{R}^{1 \times 1 \times C^{'}}$, for obtaining channel attention C_A for both the top (C_t) and bottom (C_m) branch. MLP consists of one GAP layer for obtaining channel-wise statistical data and two neuron layers. These 1D vectors can now be used to scale their respective 3D feature maps according to their importance. In short, the channel attention is obtained as follows;

$$F_{s_i}^{C_t} = \ MLP\left(F_{S_i}^{avg}\right) \tag{5}$$

$$F_{s_i}^{C_m} = \ MLP\left(F_{S_i}^{max}\right) \tag{6}$$

$$F_{s_i}^{C_t} = \ \sigma\left[W_2\left(W_1\left(GAP\left(F_{S_i}^{avg}\right)\right) + b_1\right) + b_2\right] \tag{7}$$

$$F_{s_i}^{C_m} = \ \sigma\left[W_2\left(W_1\left(GAP\left(F_{S_i}^{max}\right)\right) + b_1\right) + b_2\right] \tag{8}$$

Here $W_1 \in \mathbb{R}^{C^{'} \times C^{'}/r}$ and $b_1 \in \mathbb{R}^{C^{'}/r}$ are the weights and biases of the hidden neuron layer while $W_2 \in \mathbb{R}^{C^{'}/r \times C^{'}}$ and $b_2 \in \mathbb{R}^{C^{'}}$ belong to the output neuron layer. Finally, the output of the module is now calculated as;

$$F_{s_i}^{''} = \ \left(F_{s_i}^{C_t} \otimes \ F_{S_i}^{avg}\right) + \left(F_{s_i}^{C_m} \otimes \ F_{S_i}^{max}\right) \tag{9}$$

Where, $\bigotimes$ denotes element-wise multiplication.

Decoder Design Choices

We propose a simple yet effective decoder for our network as shown in **Figure 5A**. Our decoder bilinearly upsamples the feature map by a factor of 16 in subsequent steps. In the first step, the output of PDC is concatenated with the refined feature maps (i.e., output of DAM) of the third stage of the encoder then processed through a SE-ResNet block and finally upsampled by a factor of 2. The second and third steps also upsample the feature maps after concatenating and processing the feature maps in the same way. The only difference is that the second step upsamples by a factor of 2 while the third step upsamples by a factor of 4. This subsequent upsampling of feature maps after obtaining attention from DAM helps the network to further refine the segmentation results after each step. Lastly, the network's output is obtained by performing a 1×1 convolution followed by Softmax activation.

Implementation Details

Firstly, we resized all the images and segmentation masks to a 512×512 resolution without preserving the aspect ratio, to reduce training time and computational requirements. We also carried out extensive data augmentation during training to increase dataset size and to avoid overfitting. As for augmentation techniques used, we only selected those transformations which were suitable for segmentation problems and increased the network's robustness. To be precise, we used random crop-and-resize, random mirroring along the vertical axis, random rotation, and lastly, random brightness and saturation distortion.

In the encoder, for the number of channels in each stage, we set $C = 16$. At each stage, the number of channels (C) and the dilation rate (d) were successively increased as shown in **Figure 5A**. For the SE-ResNet block (**Figure 5B**), following Hu et al. (2018), we set the reduction ratio to $r = 8$. In the PDC module for global context aggregation, we set dilation rate

to d = {3,5,7}, respectively, for the three parallel branches as shown in **Figure 6B**. Regarding DAM, implementation details are provided in Section "Dense Attention Module (DAM)." For training, following Chen et al. (2018) and Fu et al. (2019), we employed an Adam optimizer along with poly learning rate policy where,

$$l_{r_new} = l_r * \left(1 - \frac{iter}{total_iter}\right)^{power} \quad (10)$$

Here, we set power = 0.9, $l_r = 0.005$ and we used weighted cross entropy as a loss function. We adopted dropout of 0.25 and set the mini batch_size = 4. The network is trained for 9K iterations.

RESULTS AND DISCUSSION

Ablation Study for DAM

To evaluate the effectiveness of DAM we performed several experiments and the results are reported in **Table 1**. Our baseline consisted of a simple encoder and decoder as described in Section "Network Architecture" along with simple U-Net (Ronneberger et al., 2015)-like skip-connections. Baseline did not include DAM and PDC modules. It can be seen clearly from **Table 1** that DAM significantly increases the mean IoU from 79.57% to 88.79%, with a slight increase in computational cost. Furthermore, the experiments also show that if we use different pooling operations AMP (i.e., average and max) in different branches then the network performs better as compared to the attention module with only one pooling (i.e., AP or MP). Ablation studies also show the effect of inclusion and exclusion of channel (C_A) and spatial attention (S_A) from DAM [see section "Dense Attention Module (DAM)"].

Ablation Study for PDC

There are a number of modules available for multi-scale context aggregation for obtaining better feature representation from encoder feature maps, like the pyramid pooling module (PPM) of PSP-Net (Zhao et al., 2017). Furthermore, our PDC module is closer to the atrous spatial pyramid pooling modules namely ASPP (v2 and v3) introduced by Deeplab_v2 and Deeplab_v3 (Chen et al., 2017, 2018), respectively. We used a PDC module because it has smaller memory requirement, less floating-point operations (FLOPs), and number of parameters with almost identical performance. The results are summarized in **Table 2**.

DAM Visualization With Segmentation Grad-CAM

For the qualitative analysis, we apply the Grad-CAM (Selvaraju et al., 2017), to show the effects of DAM. Grad-CAM is a gradient-based visualization method, which tries to explain the reasoning behind the decisions made by the DCNNs. It was mainly proposed for classification networks. We propose a modified version of Grad-CAM to evaluate the results of the semantic segmentation model making it into Segmentation Grad-CAM (SGC). If $\left\{A^k\right\}_{k=1}^{K}$ represents the feature map of a selected layer with K feature maps then Grad-CAM calculates the heatmaps by taking the gradient of y^c (logit for a given class) w.r.t to all N pixels (indexed by u, v), in all feature maps of $\left\{A^k\right\}_{k=1}^{K}$. But in the case of segmentation models, instead of y^c (a single value), for each class we have y_{ij}^c (a whole feature map). In this case, the gradients are computed by taking the mean of all M pixels (indexed by i, j) in the feature map of class 'c.' Finally, the weighing vector α_k^c is calculated as;

$$\alpha_k^c = \frac{1}{N}\sum_{u,\,v}\left(\frac{\delta\frac{1}{M}\sum_{(i,j)} y_{ij}^c}{\delta A_{u,v}^k}\right) \quad (11)$$

The heatmaps are then generated by;

$$L_{SGC}^c = ReLu\left(\sum_k\left(\alpha_k^c A^k\right)\right) \quad (12)$$

Thus, SGC can produce heatmaps which explain the reasoning behind the grouping of individual pixels of the input image in one segmented region in the output. We display the activated attention maps of our network at two points in the decoder as shown in **Figure 5A**: firstly, after obtaining attention from DAM of stage S_1 (i.e., point P_1) and secondly after obtaining attention

TABLE 1 | Ablation studies on DAM.

Baseline	AP	MP	AMP	PDC	CA	SA	mIoU (%)	FLOPs (G)	Memory usage (GB)	Param. (M)
✓				✓			79.57	**8.11**	1.82	**0.7**
✓	✓			✓	✓	✓	87.13	14.6	1.69	1.26
✓		✓		✓		✓	87.58	14.6	1.69	1.26
✓			✓	✓	✓		87.06	9.49	**1.56**	1.09
✓			✓		✓	✓	87.17	11.01	1.67	1.05
✓			✓	✓	✓	✓	**88.79**	14.6	1.8	1.26

Here AP and MP represent only using average or max pooling in both branches of DAM. AMP represents average and max pooling both being used in respective parallel branches of DAM. PDC is parallel dilated convolution module. C_A and S_A are channel and spatial attention sub-parts of DAM. Bold is used to highlight the best results.

TABLE 2 | Ablation studies for PDC.

Metric		PDC	PPM	ASPP_v2	ASPP_v3
		Dil_rate = {3,5,7} + Skip-connect	**Bin-size = {1,2,3,6}**	**Dil-rate = {6,12,18,24}**	**Dil-rate = {1,6,12,18} + img_pooling**
mIoU	(%)	88.79	86.49	84.67	**88.92**
FLOPs	(G)	14.6	**11.21**	15.86	15.31
Memory	(GB)	**1.8**	1.8	1.83	1.82
Param.	(M)	1.26	**1.22**	3.43	1.64

PPM represents pyramid pooling module of Zhao et al. (2017) and ASPP_v2 and v3 represents the atrous spatial pyramid pooling modules of Chen et al. (2017) and Chen et al. (2018), respectively. Bold is used to highlight the best results.

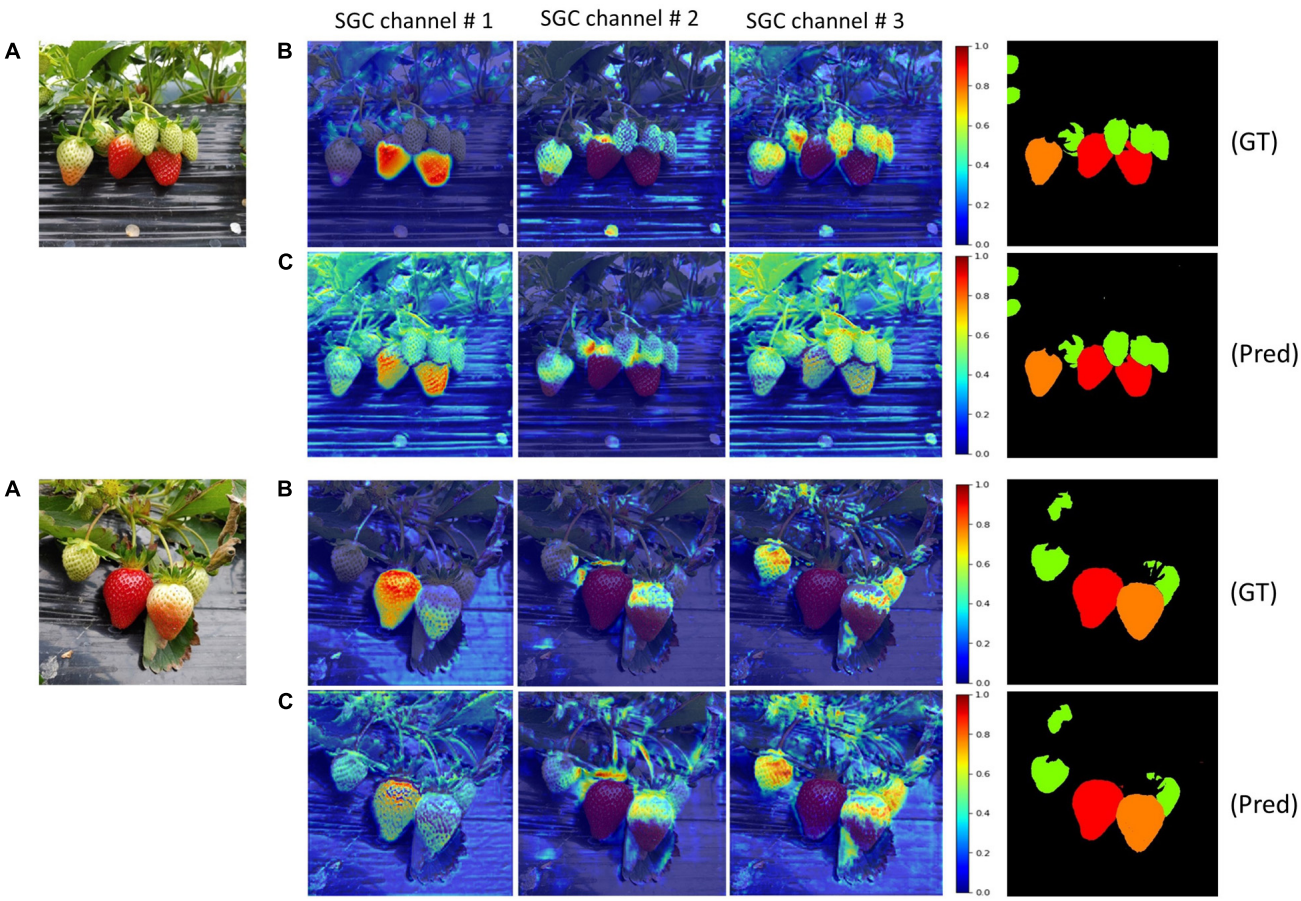

FIGURE 7 | (A) Input images; **(B)** heatmaps at point P1; **(C)** heatmaps at point P2. GT represents ground truth and Pred. is the network's prediction (best viewed in color).

TABLE 3 | Comparison of results on SS1K dataset.

Method	IOR	FLOPS (G)	Param. (M)	Memory usage (GB)	Test mIoU (%)	Precision (%) thresh = 0.75			Recall (%) thresh = 0.75			FPS (sec)
						Ripe	Unripe	Green	Ripe	Unripe	Green	
FCN-8s	1:1	20.82	2.67	5.23	76.1	92.25	74.89	90.04	75.08	78.93	87.02	39
Seg-Net	1:1	41.33	5.46	–	80.7	94.48	**87.85**	84.88	92.49	80.72	90.06	7
U-Net	1:1	70.52	4.65	8.24	83.68	94.27	77.72	89.75	84.52	82.4	89.11	32
DAN	8:1	34.4	**0.65**	5.2	77.12	95.35	65.47	87.58	56.78	65.86	77.12	10
BAM	1:1	20.6	1.1	5.28	87.3	95.63	79.46	**94.31**	91.13	84.86	86.39	36
CBAM	1:1	19.8	1.1	4.8	86.47	95.74	76.84	89.68	86.68	**86.62**	**91.21**	40
PSP-Net	8:1	34.5	0.71	5.4	85.65	94.91	86.42	91.68	84.84	64.64	86.60	18
GCN	1:1	**8.6**	0.85	2.79	80.51	90.98	72.84	87.85	92.64	60.01	83.46	51
Deeplab_v2	1:1	28.6	3.09	2.94	87.07	**95.45**	81.47	90.4	87.13	81.1	87.44	49
Deeplab_v3+	1:1	14.2	1.2	2.93	84.69	93.51	80.21	85.05	88.59	68.44	87.97	**53**
Straw-Net	1:1	14.6	1.26	**1.8**	**88.79**	94.35	86.51	94.16	**94.45**	78.19	89.64	**53**

IOR, input to output ratio of network; FPS, frames per second. For methods; bold is used to highlight the proposed method. For results; bold is used to highlight best results.

from DAM of stage S_2 (i.e., point P_2). The channel #s {1,2,3} correspond to the ripe, unripe, and green class of strawberry, respectively. It can be seen from **Figure 7B** that the heatmaps of all the classes at point P_1 gets further refined and have clearer semantic meaning than those at point P_2. Which shows the effectiveness of incorporating the DAM on skip-connections.

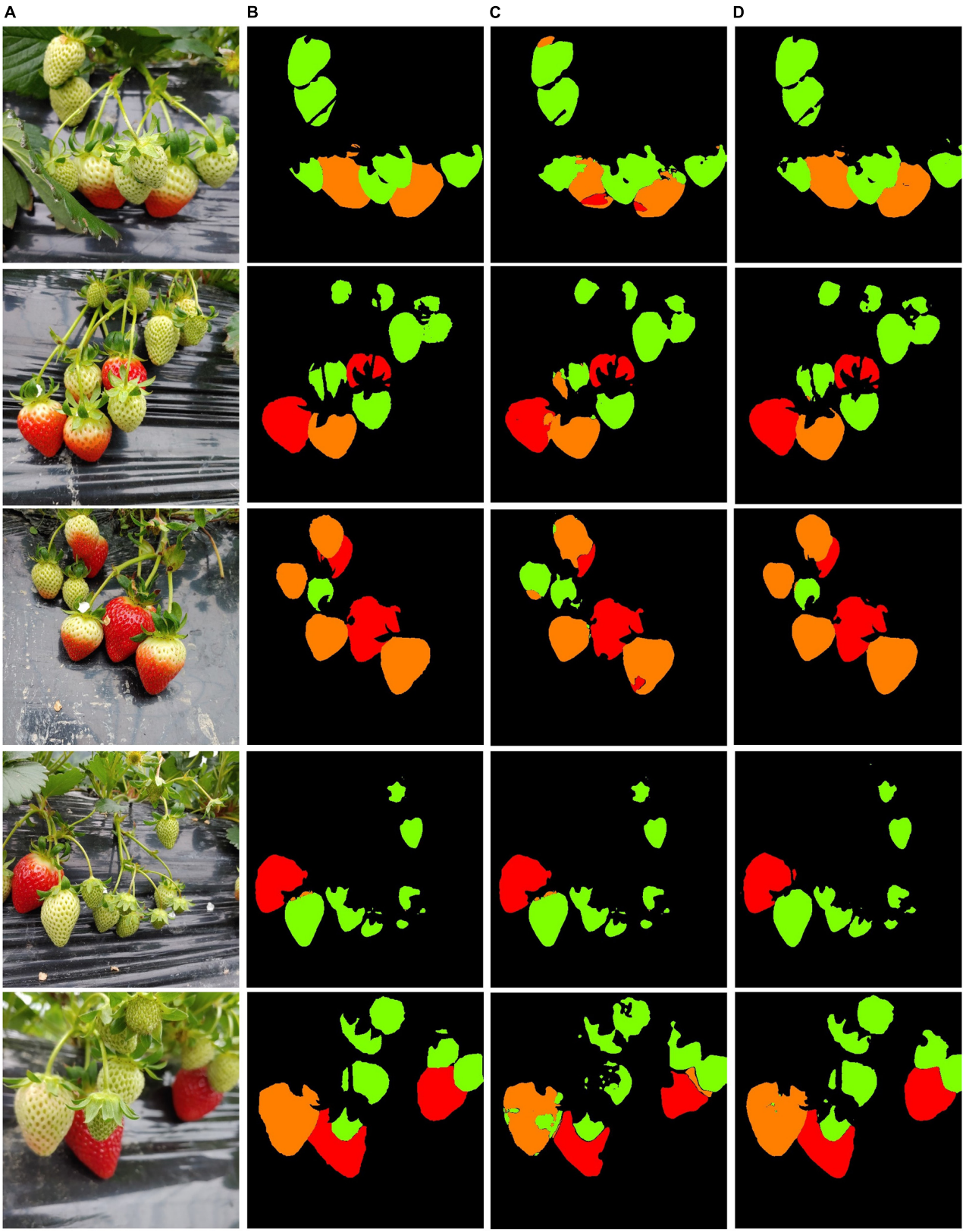

FIGURE 8 | Visualization of some semantic segmentation results on the SS1K dataset. **(A)** Raw images. **(B)** Ground truth. **(C)** Deeplab v3+. **(D)** Straw-Net (best viewed in color).

For better visualization, all the heatmaps in **Figure 7** have been rescaled to the same size.

Comparison With State-of-the-Art Networks

In this sub-section, we compare the results of our network on the SS1K dataset with other existing state-of-the-art models in semantic segmentation. We evaluate all the models on different benchmark metrics and report the results in **Table 3** and **Figure 8** shows some visual semantic segmentation results. All the values reported in **Table 3** are an average of 10 runs by default. Here IOR represents the input image to segmented output ratio. All the networks have a 1:1 ratio which means they output feature maps of the same size as the input, except DAN (Fu et al., 2019) and PSP-net (Zhao et al., 2017), their segmented output is eight times smaller than the input. Intersection over union value is averaged over all four classes. For precision and recall, the values are reported for each class separately and are calculated at a threshold of 0.75. It can be seen from **Table 3** that our Straw-Net outperforms all other existing networks overall for real-time semantic segmentation of strawberry fruits. All the metrics including frames per second (fps) are calculated for 512 × 512 resolution images, on a single Nvidia Titan RTX-2080 GPU. From **Table 3**, it can be seen that our proposed network, even though incorporating an attention mechanism is much faster, requires less memory (GB) and less floating-point operations (FLOPs) as compared to other attention networks like DAN (Fu et al., 2019), BAM (Park et al., 2018), and CBAM (Woo et al., 2018). On the other hand, compared to other existing state-of-the-art segmentation models like Deeplab_v2 and Deeplab_v3 (Chen et al., 2017, 2018), our proposed approach is able to achieve a highest mean intersection over union (mIoU) value and comparable precision recall scores. The detailed architectures of all the networks used for comparison are provided as **Supplementary Material**.

Further Analysis

To analyze the results further and to see which classes confuse the network resulting in lower performance, we plot a precision-recall (PR) curve and confusion matrix of the final segmentation results, as shown in **Figures 9**, **10**. From the results we can analyze the networks performance visually and see which classes or features are highlighted by neurons. Moreover, it will also help us to take precautionary measures to avoid inter-class confusions. For instance, in **Figure 9** the confusion matrix shows that the network is more confused between ripe and unripe strawberries rather than between unripe and green strawberries.

Another reason for this instability in confusion matrix is data imbalance, as shown in **Figure 4**. Because, there are fewer samples of unripe strawberries per batch as compared to the other two classes. We can analyze the effect of this imbalance from PR-Curves. We plot a PR-Curve for each class in **Figure 10**. PR-Curves represent a trade-off between precision and recall at different thresholds. The area under the precision-recall curve is usually denoted as AUC (i.e., area under the curve). A high value of AUC means high precision and recall. Whether we want high precision or high recall depends on the application domain. In **Figure 10**, ISO-F1 curves represents the lines in precision-recall space which have the same F1-values. We can see from **Figure 10** that the AUC for unripe strawberries is much less than the ripe and green strawberries which in turn means that precision and recall values are also low for unripe strawberries. Moreover, the micro-average curve represents the mean of PR-Curves of all classes considering data imbalance. Whereas, the macro-average curve represents the mean PR-Curve without considering data imbalance.

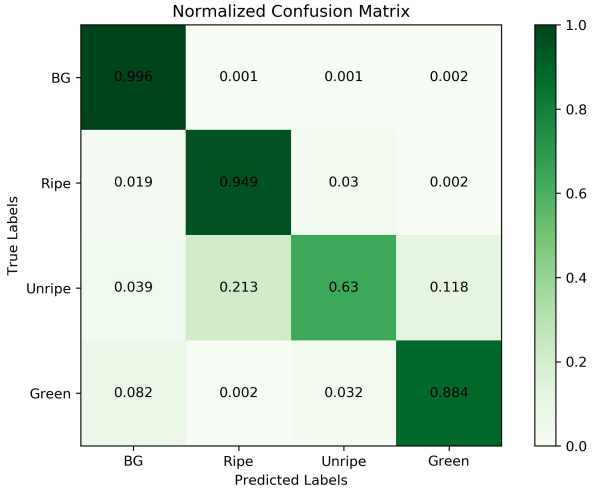

FIGURE 9 | Confusion matrix for Straw-Net architecture for semantic segmentation of the SS1K dataset.

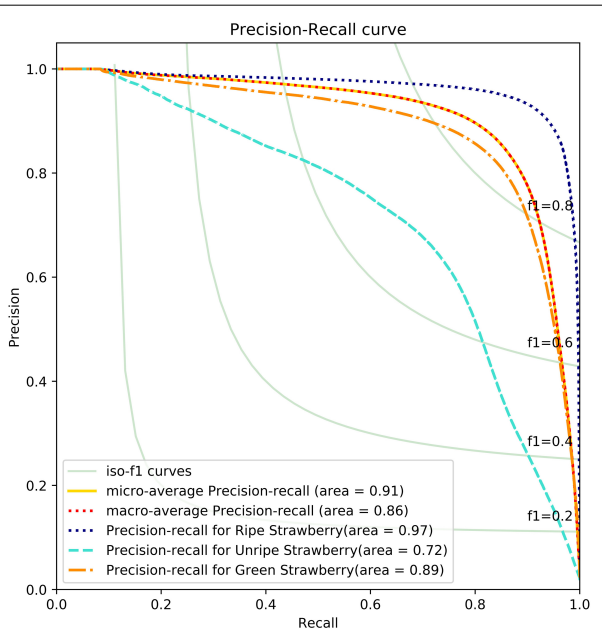

FIGURE 10 | Precision recall curve obtained for Straw-Net architecture on the SS1K dataset (Best viewed in color).

TABLE 4 | Comparison of different system configurations on network's (Straw-Net) inference speed.

System	OS	CPU	Clock speed (GHz)	GPU (Nvidia)	FLOPS (Tera)	Power consumption (Watts)	FPS (sec)
Server	Linux 18.04	Core i9-9940X	3.3	RTX-2080	14.2	–	53
Desktop PC	Linux 16.04	Core i7-9700	3.0	GTX-1650	5.5	180 ∼ 300	28.9
Desktop PC	Linux 16.04	Core i7-9700	3.0	None	–	180 ∼ 300	13.8
Portable devices							
Laptop0	Windows 10	Core i7-10750	2.6	RTX-2070	6.6	170	40
Laptop1	Windows 10	Core i7-9750	2.59	None	–	70	6.32
Laptop1	Windows 10	Core i7-9750	2.59	GTX-1650	3.2	120	21.3
Laptop2	Windows 10	Core i5-8265	1.8	None	–	48	3.38
Embedded systems							
Nvidia Jetson TX2	Linux 16.04	ARM-Cortex A57	2.0	Pascal GPU	1.3	35	15.3

Real-Time Performance Barrier

Neurons in the human visual system can interpret 10 to 12 fps and perceive them individually (Read and Meyer, 2000), whereas higher frame rates are perceived as motion. To reduce eye strain, the standard frame rate was set to be anywhere between 16 and 25 fps (Brown, 2014). Nowadays, all available video cameras have the minimum frame rate of 24 fps (Brunner, 2021).

For example, let us assume that a camera is generating 24 fps and sending those frames as an input to the proposed architecture, then the proposed algorithm should be able to process all those frames within a second to produce an output that is perceivable to the human eye. Therefore, if an algorithm can achieve a speed above this threshold ($\geq$16 fps) it is said to have crossed the real-time barrier, where this limitation is mainly generated by the human perception system. In the case of ASH, if an algorithm has a processing speed of $\geq$24 fps it means that it will generate outputs (i.e., strawberry segments) after processing all the input frames. The processing speed of 53 fps was the maximum frame rate that was achieved during the experiments with the highest system configuration, i.e., RTX-2080 GPU and Core i9-9940X CPU as shown in **Table 4**. In contrast, for most sluggish situations, let us consider that a system can only process 3 fps (Laptop2 Core i5-8265 no GPU). In this case we might have to quantize our frames so that the network can process them before the next batch arrives. Thus, the statistical value of the output generated by a 53 fps system would be higher than the output generated by a 3 fps system.

Our model is adaptive, easily scalable, has a small computational footprint of 14.6 GFLOPS (**Table 3**), and an even smaller memory footprint of 1.8 GB (**Table 3**). Therefore, it can be easily implemented on machines with low computational power like laptops with (40 and 21.3 FPS) or without (3.38 FPS) GPU or even on embedded systems like an NVIDIA Jetson TX2 board (15.3 FPS) without any loss in precision and accuracy. Therefore, any system configuration in ASH operating at the speed of $\geq$16 fps would overcome the real-time barrier and will be suitable for autonomous harvesting.

Effect of Input Resolution

To demonstrate the effect of change in resolution on the inference speed and precision of the network, we consider two more resolutions in addition to 512 × 512, i.e., 256 × 256 (low resolution) and 1024 × 1024 (high resolution). The results are reported in **Table 5** (all experiments were performed under the same conditions).

From the results we can see that if we reduce the resolution to 256 × 256 the computational complexity of the network is reduced considerably, and the speed is increased. Moreover, there is no significant decline in mIOU, but if we look at the AP, it is decreased by 14.8%. In contrast, if we analyze the case of high resolution (1024 × 1024), we can see that there is a little increase of about 0.58% in the network's performance, but the computational complexity has exploded, and inference speed is now considerably slower than the 512 × 512 version. Therefore, we recommend using the 512 × 512 resolution.

TABLE 5 | Performance comparison for input images of varying resolution.

Metric		Image resolution		
		256 × 256	***512 × 512***	***1024 × 1024***
mIoU	(%)	85.31	**88.79**	88.21
mAP	(%)	76.87	91.67	**92.06**
FLOPs	(G)	**3.6**	14.6	58
Memory	(GB)	**0.46**	1.8	7.19
Param.	(M)	1.26	1.26	1.26
FPS	(sec)	**69**	53	20

*GPU used for comparison is RTX-2080 (**Table 4**). Bold is used to highlight the best results.*

CONCLUSION

In this paper, a new dataset (i.e., SS1K) is introduced for the segmentation of strawberries into four classes depending upon the ripeness of the fruit (including a background class). The proposed segmentation network named Straw-Net improves the performance of ASHs in unconstrained and natural farming environments. Also, a real-time attention mechanism (DAM) is developed for integrating local and global semantic features efficiently. DAM controls the flow of information between the network's encoder and decoder, enabling efficient feature

fusion. Integrating adaptive feature fusion on skip-connections results in improved segmentation and classification ability of the network as shown by Segmentation Grad-CAM. The proposed attention mechanism can be integrated with any existing DCNN without any modification. By incorporating DAM in our baseline model, we achieved a significant performance boost while keeping the computational complexity to a minimum. Moreover, the effectiveness of DAM is verified by performing extensive ablation experiments. To verify the overall efficacy of the proposed approach, we compared the results with other attention mechanisms as well as with existing state-of-the-art segmentation models. Results demonstrated enhanced performance, i.e., improved mIoU, recall, and precision score with the proposed method on the strawberry segmentation problem. Our future work involves incorporating the proposed approach with ASH for deployment in strawberry farms.

DATA AVAILABILITY STATEMENT

The datasets generated for this study is available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

TI designed the study, collected the data, performed the experiments, analyzed the data, and wrote the manuscript. MU and AK collected the data and performed the experiments. HK supervised and administered the overall project and reviewed and edited the writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the Korean government (NRF-2019R1A2C1011297 and NRF-2019R1A6A1A09031717) and the US Air Force Office of Scientific Research under grant number FA9550-18-1-0016.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.591333/full#supplementary-material

REFERENCES

Adhikari, S. P., Yang, H., and Kim, H. (2019). Learning semantic graphics using convolutional encoder-decoder network for autonomous weeding in paddy field. *Front. Plant Sci.* 10:1404. doi: 10.3389/fpls.2019.01404

Agrobot (2020). *Agrobot - Agricultural Robots*. Available online at: https://www.agrobot.com/ [Accessed July 6, 2020]

Badrinarayanan, V., Kendall, A., and Cipolla, R. J. (2017). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi: 10.1109/tpami.2016.2644615

Bargoti, S., and Underwood, J. (2017). Image segmentation for fruit detection and yield estimation in apple orchards. *J. Field Robot.* 34, 1039–1060. doi: 10.1002/rob.21699

Barthes, A. (2010). *NAIO Technologies, Automated Robots and Farming Tools*. Available online at: https://www.naio-technologies.com/en/agricultural-equipment/ [Accessed October 10, 2020]

Bolya, D., Zhou, C., Xiao, F., and Lee, Y. J. (2019). "Yolact: real-time instance segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, (New Jersey, NJ: IEEE), 9157–9166.

Brown, J. (2014). "Audio-visual palimpsests: resynchronizing silent films with "Special" music," in *The Oxford Handbook of Film Music Studies*, ed. D. Neumeyer. (New York, NY: Oxford University Press), 588–610.

Brunner, D. (2021). *Frame Rate: A Beginner's Guide*. Available online at: https://www.techsmith.com/blog/frame-rate-beginners-guide/#:~{}:text=24fps%20%E2%80%93This%20is%20the%20standard,and%20viewed%20at%2024%20fps. [Accessed January 1, 2021]

Buemi, F. (1995). "Agrobot: a robotic system for greenhouse operations", in *Proceedings of the 4th Workshop on Robotics in Agriculture & the Food Industry*, (Toulouse: IARP)

Buemi, F., Massa, M., Sandini, G., and Costi, G. (1996). The agrobot project. *Adv. Space Res.* 18, 185–189. doi: 10.1016/0273-1177(95)00807-q

Chen, L. -C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach Intell.* 40, 834–848. doi: 10.1109/tpami.2017.2699184

Chen, L. -C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision ECCV*, (Munich: ECCV), 801–818.

Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., et al. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sens.* 11:1584. doi: 10.3390/rs11131584

Chollet, F. (2017). "Xception: deep learning with depthwise separable convolutions", in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE), 1251–1258.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Durand-Petiteville, A., Vougioukas, S., and Slaughter, D. C. (2017). Real-time segmentation of strawberry flesh and calyx from images of singulated strawberries during postharvest processing. *Comput. Electron. Agric.* 142, 298–313. doi: 10.1016/j.compag.2017.09.011

Dyrmann, M., Jørgensen, R. N., and Midtiby, H. S. (2017). RoboWeedSupport-detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Adv. Anim. Biosci.* 8, 842–847. doi: 10.1017/s2040470017000206

Dyrmann, M., Karstoft, H., and Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. doi: 10.1016/j.biosystemseng.2016.08.024

Edan, Y., Rogozin, D., Flash, T., and Miles, G. E. (2000). Robotic melon harvesting. *IEEE Trans. Rob. Autom.* 16, 831–835. doi: 10.1109/70.897793

Epifanio, I., and Soille, P. J. (2007). "Morphological texture features for unsupervised and supervised segmentations of natural landscapes," in *Proceedings of the IEEE Transactions on Geoscience and Remote Sensing*, (New Jersey, NJ: IEEE), 45, 1074–1083. doi: 10.1109/tgrs.2006.890581

FAO (2020). *Chapter 1: Requirements for Producing and Exporting Organic Products to Major Markets*. Rome: FAO.

Feng, Q., Wang, X., Zheng, W., Qiu, Q., and Jiang, K. (2012). New strawberry harvesting robot for elevated-trough culture. *Int. J. Agric. Biol. Eng.* 5:1.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., et al. (2019). "Dual attention network for scene segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 3146–3154.

Fuentes, A. F., Yoon, S., Lee, J., and Park, D. S. (2018). High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9:1162. doi: 10.3389/fpls.2018.01162

Fujiyoshi, H., Hirakawa, T., and Yamashita, T. (2019). Deep learning-based image recognition for autonomous driving. *IATSS Res.* 43, 244–252. doi: 10.1016/j.iatssr.2019.11.008

Gao, H., Wang, Z., and Ji, S. (2018). "Channelnets: compact and efficient convolutional neural networks via channel-wise convolutions. *Adv. Neural Inf. Process. Syst.* 5197–5205.

Ge, Y., Xiong, Y., Tenorio, G. L., and From, P. J. (2019). Fruit localization and environment perception for strawberry harvesting robots. *IEEE Access* 7, 147642–147652. doi: 10.1109/access.2019.2946369

Grand d'Esnon, A., Rabatel, G., Pellenc, R., and Journeau, A. (1987). *Magali—A Self-Propelled Robot to Pick Apples*. ASAE Paper No. 87-1037. St. Joseph, Mich: ASAE.

Grinblat, G. L., Uzal, L. C., Larese, M. G., and Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 127, 418–424. doi: 10.1016/j.compag.2016.07.003

Guerrero, J. M., Ruz, J. J., and Pajares, G. (2017). Crop rows and weeds detection in maize fields applying a computer vision system based on geometry. *Comput. Electron. Agric.* 142, 461–472. doi: 10.1016/j.compag.2017.09.028

Harvest Croo (2021). *Agricultural Robotics*. Available online at: https://harvestcroo.com/ [Accessed January 7, 2021]

Hasanpour, S. H., Rouhani, M., Fayyaz, M., and Sabokrou, M. (2016). Lets keep it simple, using simple architectures to outperform deeper and more complex architectures. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1608.06037 (accessed January 3, 2020).

Hayashi, S., Yamamoto, S., Saito, S., Ochiai, Y., Kamata, J., Kurita, M., et al. (2014). Field operation of a movable strawberry-harvesting robot using a travel platform. *Jpn. Agric. Res. Q.* 48, 307–316. doi: 10.6090/jarq.48.307

Hayashi, S., Yamamoto, S., Saito, S., Ochiai, Y., Nagasaki, Y., and Kohno, Y. (2013). Structural environment suited to the operation of a strawberry-harvesting robot mounted on a travelling platform. *Eng. Agric. Environ. Food.* 6, 34–40. doi: 10.1016/s1881-8366(13)80015-8

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, (New Jersey, NJ: IEEE) 2961–2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE), 770–778.

Hofmarcher, M., Unterthiner, T., Arjona-Medina, J., Klambauer, G., Hochreiter, S., and Nessler, B. (2019). "Visual scene understanding for autonomous driving using semantic segmentation," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, eds W. Samek, G. Montavon, A. Vedaldi, L. Hansen, and K. R. Müller (Berlin: Springer), 285–296. doi: 10.1007/978-3-030-28954-6_15

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1704.04861 (accessed December 14, 2019).

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks,", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 7132–7141.

Huang, Z., Pan, Z., and Lei, B. (2017). Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. *Remote Sens.* 9:907. doi: 10.3390/rs9090907

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154.2.

Ienco, D., Gaetano, R., Dupaquier, C., and Maurel, P. (2017). Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geosci. Remote Sens. Lett.* 14, 1685–1689. doi: 10.1109/lgrs.2017.2728698

Ilyas, T., Khan, A., Umraiz, M., and Kim, H. (2020). SEEK: a framework of superpixel learning with CNN Features for unsupervised segmentation. *Electronics* 9:383. doi: 10.3390/electronics9030383

Imai, T. (2019). Legal regulation of autonomous driving technology: current conditions and issues in Japan. *IATSS Res.* 43, 263–267. doi: 10.1016/j.iatssr.2019.11.009

Işın, A., Direkoğlu, C., and Şah, M. (2016). Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Comput. Sci.* 102, 317–324. doi: 10.1016/j.procs.2016.09.407

Jin, J., Dundar, A., and Culurciello, E. (2014). Flattened convolutional neural networks for feedforward acceleration. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1412.5474 (accessed January 30, 2020).

Kahani, A. (2017). *The Future of Fresh Fruit Harvest*. Available online at: https://www.ffrobotics.com/ [Accessed January 8, 2020]

Kauanova, S., Vorobjev, I., and James, A. P. (2017). "Automated image segmentation for detecting cell spreading for metastasizing assessments of cancer development," in *Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (New Jersey, NJ: IEEE), 2382–2387

Khan, A., Ulhaq, A., Robinson, R., and Rehman, M. U. (2020). "Detection of vegetation in environmental repeat photography: a new algorithmic approach in data science," in *Statistics for Data Science and Policy Analysis*. (Berlin: Springer), 145–157. doi: 10.1007/978-981-15-1735-8_11

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* 14, 778–782. doi: 10.1109/lgrs.2017.2681128

Lamb, N., and Chuah, M. C. (2018). "A strawberry detection system using convolutional neural networks," in *Proceedings of the IEEE International Conference on Big Data (Big Data)*, (New Jersey, NJ: IEEE), 2515–2520.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1, 541–551. doi: 10.1162/neco.1989.1.4.541

Lin, G., Milan, A., Shen, C., and Reid, I. (2017a). "Refinenet: multi-path refinement networks for high-resolution semantic segmentation", in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE), 1925–1934.

Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., et al. (2017b). A structured self-attentive sentence embedding. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1703.03130 (accessed February 13, 2020).

Lin, Z., Ji, K., Kang, M., Leng, X., and Zou, H. (2017c). Deep convolutional highway unit network for SAR target classification with limited labeled training data. *IEEE Geosci. Remote Sens. Lett.* 14, 1091–1095. doi: 10.1109/lgrs.2017.2698213

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, (Berlin: Springer), 21–37.

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 3431–3440.

Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y., et al. (2019). Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PLoS One* 14:e0215676. 10.1371/journal.pone.0215676 doi: 10.1371/journal.pone.0215676

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls.2016.01419

Nizami, I. F., Ur Rehman, M., Majid, M., and Anwar, S. (2020). *Natural Scene Statistics Model Independent No-Reference Image Quality Assessment using Patch Based Discrete Cosine Transform*. Berlin: Springer.

Noh, H., Hong, S., and Han, B. (2015). "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, (New Jersey, NJ: IEEE) 1520–1528.

Ouyang, C., Li, D., Wang, J., Wang, S., and Han, Y. (2012). "The research of the strawberry disease identification based on image processing and pattern recognition," in *Proceedings of the International Conference on Computer and Computing Technologies in Agriculture* (Berlin: Springer), 69–77. doi: 10.1007/978-3-642-36124-1_9

Park, J., Woo, S., Lee, J. -Y., and Kweon, I. (2018). Bam: bottleneck attention module. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1807.06514 (accessed March 19, 2020).

Park, S., Wahab, A., Nazari, I., Ryu, J. H., and Chong, K. T. (2020). i6mA-DNC: prediction of DNA N6-Methyladenosine sites in rice genome based on dinucleotide representation using deep learning. *Chemometr. Intell. Lab. Syst.* 204:104102. doi: 10.1016/j.chemolab.2020.104102

Peng, C., Zhang, X., Yu, G., Luo, G., and Sun, J. (2017). "Large kernel matters–improve semantic segmentation by global convolutional network", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 4353–4361.

Qingchun, F., Wengang, Z., Quan, Q., Kai, J., and Rui, G. (2012). "Study on strawberry robotic harvesting system," in *Proceeding of the 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)* (New Jersey, NJ: IEEE), 320–324.

Read, P., and Meyer, M.-P. (2000). *Restoration of Motion Picture Film*. Amsterdam: Elsevier.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster r-cnn: towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 1, 91–99.

Rensink, R. A. (2000). The dynamic representation of scenes. *Vis. Cogn.* 7, 17–42. doi: 10.1080/135062800394667

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer), 234–241.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and Mccool, C. J. (2016). Deepfruits: a fruit detection system using deep neural networks. *Sensors* 6:1222. doi: 10.3390/s16081222

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, (New Jersey, NJ: IEEE) 618–626.

Shen, T., Zhou, T., Long, G., Jiang, J., Pan, S., and Zhang, C. (2018). "Disan: Directional self-attention network for rnn/cnn-free language understanding," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. (Menlo Park, CA: AAAI)

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1409.1556 (accessed October 19, 2019).

Strawberry (2020). *Growing Strawberries*. Available online at: https://strawberryplants.org/growing-strawberries/ [Accessed July 27, 2020].

Sun, J., Lu, B., and Mao, H. (2011). Fruits recognition in complex background using binocular stereovision. *J. Jiangsu U. Nat. Sci. Ed.* 32, 423–427.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 2818–2826.

ur Rehman, M., Khan, S. H., Rizvi, S. D., Abbas, Z., and Zafar, A. (2018). "Classification of skin lesion by interference of segmentation and convolotion neural network," in *Proceedings of the 2018 2nd International Conference on Engineering Innovation (ICEI)* (New Jersey, NJ: IEEE), 81–85.

Wahab, A., Ali, S. D., Tayara, H., and Chong, K. T. (2019). iIM-CNN: intelligent identifier of 6mA sites on different species by using convolution neural network. *IEEE Access* 7, 178577–178583. doi: 10.1109/access.2019.2958618

Wahab, A., Mahmoudi, O., Kim, J., and Chong, K. T. (2020). DNC4mC-deep: identification and analysis of DNA N4-methylcytosine sites based on different encoding schemes by using deep learning. *Cell* 9:1756. doi: 10.3390/cells9081756

Wang, M., Liu, B., and Foroosh, H. (2016). Design of efficient convolutional layers using single intra-channel convolution, topological subdivisioning and spatial" bottleneck" structure. *arXiv [Preprint]* Avilable at: https://arxiv.org/abs/1608.04337 (accessed October 26, 2019).

Wei, X., Jia, K., Lan, J., Li, Y., Zeng, Y., and Wang, C. J. O. (2014). Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot. *Optik* 125, 5684–5689. doi: 10.1016/j.ijleo.2014.07.001

Woo, S., Park, J., Lee, J.-Y., and So Kweon, I. (2018). "Cbam: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision* (Munich: ECCV), 3–19. doi: 10.1007/978-3-030-01234-2_1

Xia, X., and Kulis, B. (2017). W-net: a deep model for fully unsupervised image segmentation. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1711.08506 (accessed December 7, 2020).

Xiang, R., Jiang, H., and Ying, Y. (2014). Recognition of clustered tomatoes based on binocular stereo vision. *Comput. Electron. Agric.* 106, 75–90. doi: 10.1016/j.compag.2014.05.006

Xiong, Y., Peng, C., Grimstad, L., From, P. J., and Isler, V. (2019). Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper. *Comput. Electron. Agric.* 157, 392–402. doi: 10.1016/j.compag.2019.01.009

Yamamoto, S., Hayashi, S., Saito, S., Ochiai, Y., Yamashita, T., and Sugano, S. J. (2010). Development of robotic strawberry harvester to approach target fruit from hanging bench side. *IFAC Proc. Vol.* 43, 95–100. doi: 10.3182/20101206-3-jp-3009.00016

Yamamoto, S., Hayashi, S., Yoshida, H., and Kobayashi, K. (2014). Development of a stationary robotic strawberry harvester with a picking mechanism that approaches the target fruit from below. *Jpn Agric. Res. Q.* 8, 261–269. doi: 10.6090/jarq.48.261

Yu, F., and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1511.07122 (accessed March 21, 2020).

Yu, F., Koltun, V., and Funkhouser, T. (2017). "Dilated residual networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 472–480.

Yu, Y., Zhang, K., Yang, L., and Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* 163:104846. doi: 10.1016/j.compag.2019.06.001

Zhang, B., Huang, W., Wang, C., Gong, L., Zhao, C., Liu, C., et al. (2015). Computer vision recognition of stem and calyx in apples using near-infrared linear-array structured light and 3D reconstruction. *Biosyst. Eng.* 139, 25–34. doi: 10.1016/j.biosystemseng.2015.07.011

Zhang, K., Yang, L., Wang, L., Zhang, L., and Zhang, T. (2012). Design and experiment of elevated substrate culture strawberry picking robot. *Trans. Chin. Soc. Agric. Machinery* 43, 165–172.

Zhang, X., Zhou, X., Lin, M., and Sun, J. (2018). "Shufflenet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 6848–6856.

Zhao, H., and Kit, C. (2011). Integrating unsupervised and supervised word segmentation: the role of goodness measures. *Inf. Sci.* 181, 163–183. doi: 10.1016/j.ins.2010.09.008

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE) 2881–2890.

Zhao, Y., Gong, L., Huang, Y., and Liu, C. (2016). A review of key techniques of vision-based control for harvesting robot. *Comput. Electron. Agric.* 127, 311–323. doi: 10.1016/j.compag.2016.06.022

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2014). Object detectors emerge in deep scene cnns. *arXiv [Preprint]* Available online at: https://arxiv.org/abs/1412.6856 (accessed May 6, 2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ilyas, Umraiz, Khan and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 March 2021
doi: 10.3389/fpls.2021.608732

Deep Segmentation of Point Clouds of Wheat

Morteza Ghahremani[1,2]*, *Kevin Williams*[1], *Fiona M. K. Corke*[1], *Bernard Tiddeman*[2], *Yonghuai Liu*[3] and *John H. Doonan*[1]

[1] National Plant Phenomics Centre, Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, United Kingdom, [2] Department of Computer Science, Aberystwyth University, Aberystwyth, United Kingdom, [3] Department of Computer Science, Edge Hill University, Ormskirk, United Kingdom

The 3D analysis of plants has become increasingly effective in modeling the relative structure of organs and other traits of interest. In this paper, we introduce a novel pattern-based deep neural network, Pattern-Net, for segmentation of point clouds of wheat. This study is the first to segment the point clouds of wheat into defined organs and to analyse their traits directly in 3D space. Point clouds have no regular grid and thus their segmentation is challenging. Pattern-Net creates a dynamic link among neighbors to seek stable patterns from a 3D point set across several levels of abstraction using the K-nearest neighbor algorithm. To this end, different layers are connected to each other to create complex patterns from the simple ones, strengthen dynamic link propagation, alleviate the vanishing-gradient problem, encourage link reuse and substantially reduce the number of parameters. The proposed deep network is capable of analysing and decomposing unstructured complex point clouds into semantically meaningful parts. Experiments on a wheat dataset verify the effectiveness of our approach for segmentation of wheat in 3D space.

Keywords: 3D analysis, segmentation, convolutional neural network, deep learning, pattern, point cloud, wheat

OPEN ACCESS

Edited by:
Reza Ehsani,
University of California, Merced, United States

Reviewed by:
Shawn Carlisle Kefauver,
University of Barcelona, Spain
Tao Duan,
Institute of Microelectronics (CAS), China

***Correspondence:**
Morteza Ghahremani
mog9@aber.ac.uk

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *21 September 2020*
Accepted: *24 February 2021*
Published: *24 March 2021*

Citation:
Ghahremani M, Williams K, Corke FMK, Tiddeman B, Liu Y and Doonan JH (2021) Deep Segmentation of Point Clouds of Wheat. Front. Plant Sci. 12:608732. doi: 10.3389/fpls.2021.608732

1. INTRODUCTION

Three- and four-dimensional phenotyping has the potential to provide reliable, comprehensive information on morphological and developmental traits in plants. With recent improvements in image acquisition and 3D reconstruction, future studies would benefit from rapidly assessing 3D models (Chaudhury et al., 2018; Bernotas et al., 2019; Chaudhury and Godin, 2020; Artzet et al., under review). Accurate 3D models enable quantitative analyses of various traits, and a high-throughput spatial and temporal 3D analysis tool could monitor impacts of different treatments in experiments and, ultimately, management decisions in production conditions. 3D or higher-order data, however, requires complex processes for both acquisition and computation while quality can vary due to numerous factors such as imaging noise, occlusion, spikes, holes, lack of homogeneity, and interference from cluttered backgrounds. Despite the obvious attractions, few segmentation techniques have been reported for 3D point clouds of plants and they tend to require specific conditions that cannot easily be generalized.

Wheat is globally important with more than 700 million tonnes of grain produced annually (FAO report 2020)[1]. The grain-filling period of wheat is a key growth period that directly influences yield. There is widespread interest in estimating the number of ears per unit area

[1]http://www.fao.org/worldfoodsituation/csdb/en/

(Ferrante et al., 2017) and other traits crucial for determining yield from images. Ear segmentation is therefore critical in estimating yield in wheat (Bi et al., 2010; Kun et al., 2011; Chen et al., 2016; Alharbi et al., 2018; Tan et al., 2020). Manual data collection, involving visual inspection of the standing crop, is labor intensive and time-consuming. Image processing and computer vision techniques facilitate high-throughput counting of ears. Such techniques can rapidly estimate yield, potentially accurately and with minimal human intervention.

Deep learning has been invaluable for the development of high-throughput pipelines that undertake 2D image analysis of wheat and many other plants (Qiongyan et al., 2017; Hasan et al., 2018; Wang X. et al., 2019; Hamidinekoo et al., 2020). Learning methods capable of extracting high-level features from raw input data with minimal human intervention would be useful for high-throughput pipelines. Lack of depth information is a major drawback of current 2D imaging, limiting the accurate quantitative evaluation of many traits. In this study, we demonstrate that deep learning techniques can also be used to directly segment 3D geometric wheat data, acquired using standard 3D structure from motion techniques (Furukawa and Ponce, 2010; Jay et al., 2015; Schönberger and Frahm, 2016; Schönberger et al., 2016). In this report, we propose a novel network that efficiently handles highly complex 3D point clouds. Unlike most segmentation techniques that heavily rely on data and its distribution, our proposed network extracts stable patterns from point clouds across different levels of features obtained through the K-nearest neighbor algorithm. Our network is thus more robust to variation in the density of point cloud data, typical imaging distortions, and noise. To the best of our knowledge, this paper is the first study to segment and analyse ears directly within the point cloud domain via deep learning. The proposed framework has been validated using 690 wheat point clouds, captured at different times during the growth cycle. The results indicate that our deep learning method is robust and can accommodate irregular point clouds that are noisy and contain irrelevant outliers.

In section 2, we review previously reported segmentation techniques in plant science. The proposed pattern-based deep neural network (Pattern-Net) is detailed in section 3. Section 4 reports and discusses the experimental results of Pattern-Net on the wheat dataset. Section 5 relates our findings to previous studies and, finally, conclusions and future work are provided in section 6.

2. BACKGROUND

Segmentation of ears is challenging due to their highly complicated and varied shapes and numbers and unpredictable interaction with their background. Most studies to date have been carried out in the 2D domain using standard images (Chopin et al., 2016; Zhou et al., 2018; Misra et al., 2020). A hybrid approach (Chopin et al., 2016) uses a-priori information about the shape of leaves and local image orientations to fit active contour models to features that are missed during the initial segmentation. Mohanty et al. (2016) applied a deep learning method for plant disease detection. Madec et al. (2019) employed a CNN to identify ears from low-spatial-resolution RGB images. Ubbens and Stavness (2017) implemented deep convolutional neural networks (CNNs), successfully estimating leaf number from an image database of Arabidopsis rosettes. Sadeghi-Tehran et al. (2019) developed a deep CNN-based classification technique to automatically identify and count the number of ears in images taken under natural field conditions. Recently, a 2D CNN model (Xu et al., 2020) extracted the contour features of ears using a K-means clustering algorithm and then classified the segmented images using a five-layered CNN. These examples clearly demonstrate the potential of these approaches to extract useful biologically relevant information from images and the feasibility of scaling to accommodate very large datasets.

Previous methods for segmenting point clouds considered constraints and used learning-based optimization techniques such as clustering, support vector machine (SVM) etc. (Paulus et al., 2013; Li et al., 2018). Gélard et al. (2017) segmented leaves using a geometrical constraint and Euclidean cluster extraction method. Liu et al. (2018) exploited a revised version of Euclidean distance and spectral clustering to segment individual leaves from a variety of plants including wheat. Multi-view vision segmentation techniques (Guo and Xu, 2017; Shi et al., 2019) have been applied to stereo multi-view 2D images. The performance of three learning methods including SVM, boosting, and K-means clustering in the segmentation of soybean plants were compared in Zhou et al. (2019), where K-means clustering outperformed the other methods in terms of processing efficiency and segmentation accuracy. We previously used a semi-automatic method for segmentation of leaf and petiole in Grapevine to quantify drought responses from images (Briglia et al., 2020). Jin et al. (2018) proposed an indirect method for 3D object detection and segmentation, whereby a region-based CNN (RCNN) is used to detect objects in 2D images projected from 3D points.

Since traditional point cloud-based segmentation methods consider some constraints that depend on traits of interest, the generalization of such methods is not straightforward. The efficiency of previous methods is also questionable in highly complex noisy 3D models. To address these drawbacks, a tensor-based technique has been developed that represents highly-complex models by their first- and second-order tensors without requiring pre-defined shape assumptions and constraints (Elnashef et al., 2019). Most recently, Li et al. (2019) employed a 3D joint filtering operator for leaf segmentation. Here, we introduce a new procedure for segmentation of 3D point cloud data from plants using deep neural networks.

A deep learning-based point cloud segmentation named PointNet (Qi et al., 2017a) has been recently proposed that is capable of extracting high-level features from raw input data via learning on sufficient 3D CAD models[2] of various objects. The mean accuracy of PointNet is an impressive 84% and has been further improved by Qi et al. (2017b), Shen et al. (2018), Guerrero et al. (2018), Landrieu and Simonovsky (2018), and Wang Y. et al.

[2] 3D CAD models are online available at http://modelnet.cs.princeton.edu

(2019). Despite poor existing segmentation methods, 3D point cloud deep learning segmentation methods can effectively handle complex models across a wide array of species. Applying these techniques to typical 3D models of plants ($>10^4$ points) is almost impossible since current GPU devices are unable to process such large models. These issues motivated us to further develop a light deep network for point cloud segmentation (Ghahremani et al., 2020) that is highly effective for architectural models. However, direct application of this method to plant point cloud data did not yield satisfactory results since plants tend to occupy volumetric space in a very different manner from buildings for example—with complex structures, configurations, occlusion, and often cluttered background. Here we expand our recent segmentation method (Ghahremani et al., 2020) to wheat point clouds. To the best of our knowledge, the proposed network provides the first practical segmentation of plant parts directly within the point cloud domain. We provide thorough empirical and theoretical analysis on the stability and efficiency of the proposed Pattern-Net method using more than 690 wheat point clouds and demonstrate its ability to extract biologically meaningful data in terms of accurate ear counts and ear-length estimates.

3. PROPOSED 3D POINT CLOUD SEGMENTATION NETWORK

The goal is to establish and train a deep neural network that converts an input point set $P = \{p_1, \ldots, p_M\}$ into a set of segmentation labels. Here, M denotes the total number of 3D points and they are represented as a set of 3D coordinates. The ground-truth label is a vector of length M, $\Gamma = \{\gamma_1, \ldots, \gamma_M\}$, where γ_i is the label of i-th point. Since there are N segmentation labels, thus $\gamma_i \leq N$. The output of the network is a vector of predicted labels, i.e., $\hat{\Gamma} = \{\hat{\gamma}_1, \ldots, \hat{\gamma}_M\}$. The principles of the proposed Pattern-Net are explained in the following sections. Ghahremani et al. (2020) provided more details about implementation.

3.1. Network Properties

A segmentation network for a point cloud set must meet the following four requirements about invariance (Qi et al., 2017a,b; Ghahremani et al., 2020):

Property I (permutation invariance): This property states that the segmentation labels must be invariant to changes in the order of 3D points. If γ_i and γ_j are the segmentation labels of 3D points p_i and p_j, respectively, then

$$[\gamma_i, \gamma_j] = [\gamma_j, \gamma_i], \quad \forall i, j \in \{1, \ldots, M\}, \tag{1}$$

where [.] indicates the order. Unlike pixels in images or voxels in volumetric grids, a 3D point cloud set has no order and due to its irregular format, the segmentation network must be invariant to the order of the points.

Property II (transformation invariance): The segmentation results must not be varied by changes in affine transformation, i.e.,:

$$\Gamma_{r[p_1,\ldots,p_M]+t} = \Gamma_{[p_1,\ldots,p_M]}. \tag{2}$$

3D models may be captured or described under different viewpoints (rotation) and translations (position) at different growth time (scaling). These factors must not influence the segmented labels when a network segments a point cloud of interest.

Property III (3D points relations): In point cloud domain the relationship between 3D points, denoted by R, is determined by their distance from each other:

$$R\{p_i, p_j\} = D(p_i, p_j). \tag{3}$$

The distance metrics could be Euclidean distance, Manhattan distance, cosine distance, etc. Points in the point cloud domain are not isolated and their neighbors represent meaningful parts/organs that execute particular functions and produce particular behaviors.

Property IV (resolution-invariance): The density of 3D points (or equivalently the number of points) must not influence the performance of the segmented regions. The density of the point cloud influences the relationship parameter defined in Equation (3), but the overall segmentation results must remain unchanged.

These four properties provide the foundation for the design of our network.

3.2. Network Architecture

The basic steps of the proposed segmentation network are depicted in **Figure 1**. The framework has five main layers: points downsampler (PD), search pattern (SP), learn pattern (LP), linkage patterns (LPs), and fully connected (FC) layers.

The input 3D point set is first decomposed into "L" levels by the PD layer. Inside each scale level, the relationship between each query point and its neighbors is sought by the KNN algorithm embedded in the SP layer and then is learned as a pattern by the LP layer. There are several interactions between the SP and the LP layers for extracting the deep patterns from the relationships of 3D points. The linkage features (LPs) layer links all learnt patterns across all levels and finally an FC layer predicts the segmentation labels. In the following, we detail these layers.

Points Downsampling (PD) Layer: Image acquisition is undertaken at different zoom levels and growth times that directly affect quality, density, and quantity of the point clouds. The function of this layer is to make the deep network independent of the quantity and distribution of points (*Property IV*). To this end, we decompose the input 3D point cloud, P, into L sets via a random downsampling operator, in such a way that all the 3D point subsets, $P\{l\}$, $l \in \{1, \ldots, L\}$, are completely different while their overall schemes/abstracts are similar to each other:

$$P\{l\} \cap P\{k\} = \emptyset \quad \forall l, k \in \{1, \ldots, L\}, \tag{4}$$

$$\bigcup_{l=1}^{L} P\{l\} = P, \tag{5}$$

$$R\{l\} \simeq R\{k\} \quad \forall l, k \in \{1, \ldots, L\}. \tag{6}$$

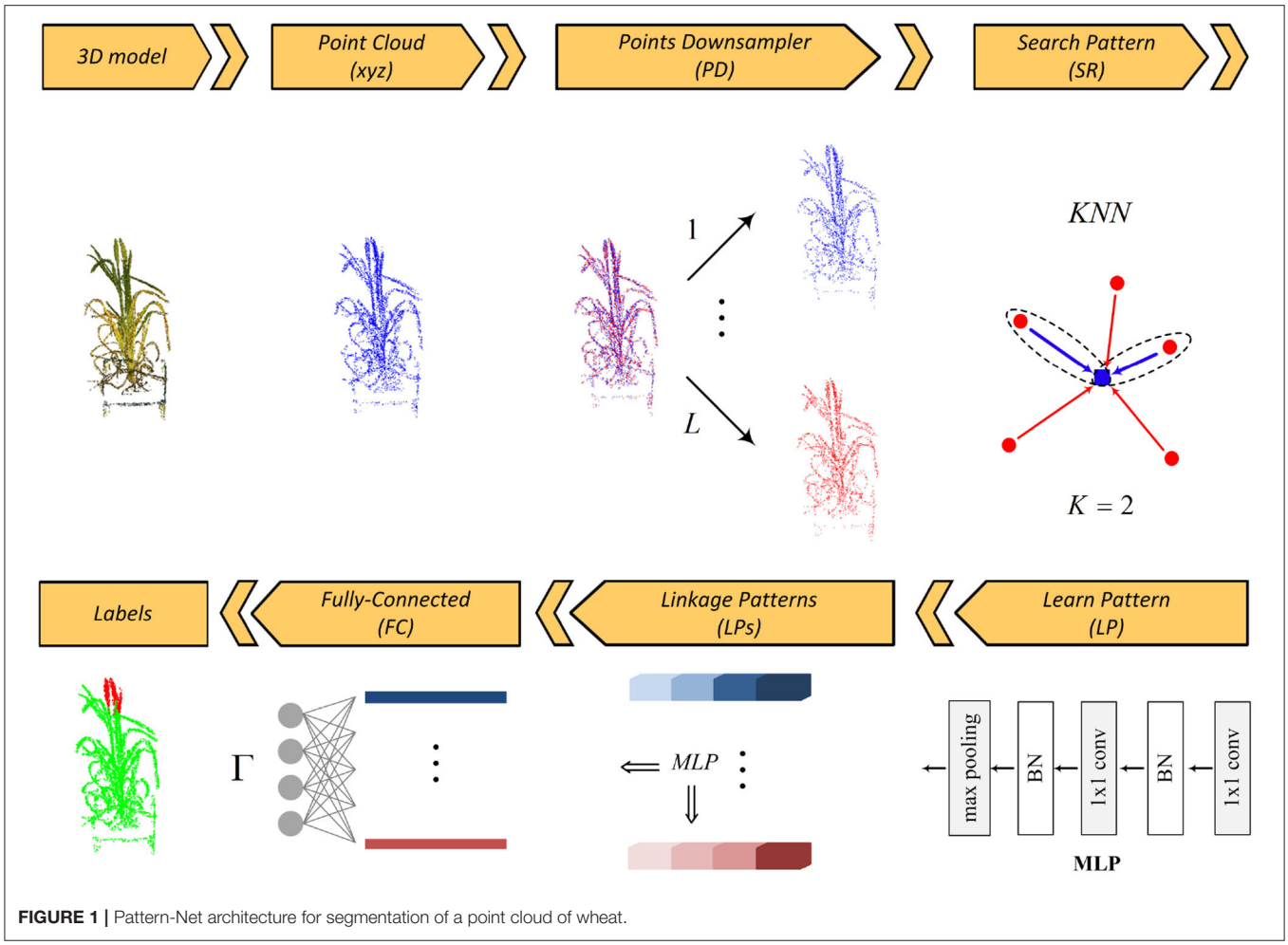

FIGURE 1 | Pattern-Net architecture for segmentation of a point cloud of wheat.

As illustrated in **Figure 1**, the segmented regions of a plant remain unchanged across different sampling levels while none of the decomposed point sets shares identical points. The main idea is to enforce the network keeping patterns of a 3D point cloud throughout all sampling levels. Similar organs/parts across different sampling levels share similar global features and this will assign considerable weights to such organs in the LPs layer, while dissimilar organs have smaller weights that are removed by a dropout operator. As will be discussed in "search pattern," this strategy also effectively helps the network not to be saturated with its K nearest neighbors while keeping the radius of the neighborhood reasonable. In short, the main advantages of using multi-level sampling analysis are:

- Detection of hidden general patterns by decomposing a complex point cloud into simpler ones;
- Making a balance between the searching area and K responses; and
- Efficiently reducing the computational complexity of the KNN algorithm.

Search Pattern (SP) Layer: The task of this layer is to search all possible relationships between the query point/feature ($\mathbf{f}_q = (f_{q,x}, f_{q,y}, f_{q,z})$) and its neighbors via the KNN algorithm (*Property III*). For each of K nearest neighbor responses ($\mathbf{f}_i = (f_{i,x}, f_{i,y}, f_{i,z}), i = \{1, 2, .., K\}, i \neq q$), we compute all three possible edges emanating from the query point along three axes (i.e., $\mathbf{f}_i - \mathbf{f}_q$), and stack it with the query point coordinates/feature $\mathbf{f}_q$. Thus, there is a feature space of size $K \times 6$ for each query point. Adding edges to the feature space is important as KNN sorts K nearest responses and how far KNN responses are from the query point should be taken into account.

Learn Pattern (LP) Layer: The function of this layer is to find and to learn a meaningful relationship/pattern between all input 3D points via a two consecutive 2D convolution kernel followed by a batch normalization operator. A max-pooling operator is then applied to the output weights to get the features of the query point. The max-pooling is a symmetric function that guarantees that the extracted features are permutation-invariant (*Property I*). The combination of 2D convolution kernels, batch normalization and max-pooling operators is often called multi-layer perceptron (MLP) (Qi et al., 2017a). Inside each decomposed set, relationships between each query point and its neighbors are sought by the SP layer and then learned by the LP layer. This is done by applying and concatenating four

MLPs {32, 32, 32, 32}, yielding from low-level features to high-level ones. Hence, there is a feature vector of length 128 for each 3D point inside each decomposed set.

Linkage Patterns (LPs) Layer: This layer contains several MLP layers and it aims to link the patterns that are similar across all the decomposed levels. As can be seen in the figure, the LPs layer is fed by all the low-level and high-level features. By applying a max-pooling operator to the features of the points inside a sampling set, a description vector of length 128 is obtained. We arrange all the local description vectors $\psi_l, l \in \{1, \ldots, L\}$ in a matrix Ψ. We then apply an MLP to the whole cube features of the points to yield a global description vector ϕ. As discussed in section 3.3, the global description vector is used as a guideline for extracting stable patterns in the feature space.

Fully-Connected (FC) Layer: This layer functions as a decoder and maps the patterns extracted in the preceding layer into Γ labels. The output of the LPs layer is decoded by three consecutive MLPs {256, 256, Γ}. The drop-rate of all the decoding MLPs except the last one is fixed at $\frac{2}{3}$.

3.3. Network Loss Function

The goal of the LPs layer is to make the local vectors $\psi_l, l \in \{1, \ldots, L\}$ as close to the global one ϕ as possible for the detection of the stable patterns inside the given point cloud. Assume that there is a linear relationship between the cloning and global description vectors, i.e., $\phi = \Psi\omega$, the estimated coefficients ω can be computed by the Moore–Penrose inverse (Penrose, 1955), i.e.,

$$\omega = \Psi^{\dagger}\phi = (\Psi^T\Psi)^{-1}\Psi^T\phi. \tag{7}$$

The Moore–Penrose pseudo-inverse could be simply implemented by singular value decomposition (SVD) (Brake et al., 2019). The coefficient vector ω measures the contribution of each local set in the resulting global one. The variance $\sigma(\omega)$ of elements of ω approaches zero if all the local description vectors are close to the global one. We add this term into the loss function as follows:

$$L(\theta) = \underbrace{-\frac{1}{n}\sum_{i=1}^{M}\sum_{k=1}^{\Gamma}\Omega_k y_{ik}\log p_{ik}}_{\textit{segmentation loss}} + \lambda \underbrace{\sigma(\omega)}_{\textit{linear mapping loss}} \tag{8}$$

In the above equation, the first term is the cross-entropy function for computing the loss of the predicted labels and the second term forces the network to yield zero standard deviation for the coefficients obtained by the linear mapping. y_{ik} is one-hot encoded labels and y_{ik} is scaled softmax logits. λ is a predetermined hyperparameter. In the segmentation of plants, some organs are of more interest than others; for example, the segmentation of ears is more important than those of the other organs. To deal with imbalanced distributions of organ-specific point clouds, we have added a dynamic coefficient vector, Ω, into Equation (8), which is defined as

$$\Omega_k = |C_k - \frac{\sum_{i=1}^{M}(\gamma_i == k)}{M}|, \tag{9}$$

where C_k is a probability constant that determines the significance of the k-th segmented organ.

4. EXPERIMENTAL RESULTS

4.1. Data Acquisition and Preparation

Spring wheat (variety Paragon) was used to acquire the images for modeling. These plants comprised part of Experiment W048 being undertaken to benchmark wheat growth under LED lighting. Briefly, they were grown as single plants in 1 L capacity pots containing Levington F2 peat-based compost. After germination, plants were grown on a conveyor based automated watering and imaging system (Lemnatec, Germany) at National Plant Phenomics Centre (NPPC)[3] and grown under white LED Sunblaster (Kroptek, Sussex UK) luminaries at light level of 400 μM m^{-2} s^{-1}. Pots were watered daily to a target weight equivalent to either 75% (well-watered) or 35% (droughted) of field capacity and grown to maturity. The image acquisition system employed a pair of freestanding DSLR cameras in carefully calibrated locations that have been piggybacked onto the propriety LemnaTec platform, which acts as a delivery and lighting system for routine image collection. An in-line turntable was used to rotate subjects through 360 degrees and camera triggering was controlled and synchronized by prototype software, and image collection was based on commands from "gphoto2"[4]. Each image acquisition event provides 74 high-resolution multi-view images (6,000×4,000 px.) per plant. For the purposes of this analysis, we used images from 10 individuals grown under well-watered conditions and 10 individuals grown under drought, and a total of 690 point clouds were reconstructed and selected for segmentation.

The 3D models were reconstructed from the multi-view images by COLMAP (Schönberger and Frahm, 2016; Schönberger et al., 2016). COLMAP includes two phases: structure-from-motion (SfM) for sparse reconstruction and multi-view stereo (PMVS) for dense reconstruction. SfM extracts the calibration parameters including intrinsic and extrinsic parameters/matrices from the multi-view images. To this end, we detected keypoints from images by FFD (Ghahremani et al., 2021) and then extracted features from the keypoints by InterTex feature descriptor (Ghahremani et al., 2021). Exhaustive matching (Codreanu et al., 2013) was applied to the features to find corresponding keypoints in the multi-view images. The matched keypoints were then verified by geometric verification and finally, the structure and motion reconstruction were extracted (Schönberger and Frahm, 2016). PMVS (Schönberger et al., 2016) projected the 2D images into 3D space using the transformation matrices obtained by SfM and forms point clouds as outputs.

We annotated the point clouds using MeshLab software (Ranzuglia et al., 2013). Regions of interest were extracted and labeled into one of two semantic categories—ear and non-ear. Thus, the number N of labels is equal to 2 and examples are shown in **Figure 2**. The segmentation task was

[3]https://www.plant-phenomics.ac.uk/
[4]https://www.gphoto.org/

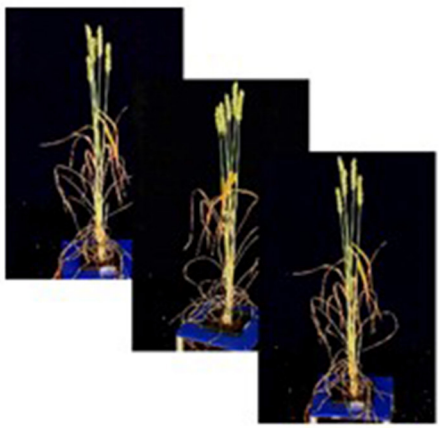

FIGURE 2 | Samples of the captured multi-view images, their reconstructed point clouds and annotated ones. Ears in annotated point clouds are shown in red and non-ears in green.

repeated under a different number of input points ranging from 512 to 16,384. Final harvest measurements including plant height, ear number, and ear length were used for independent verification of the segmentation results.

4.2. Evaluation Metrics

The segmented point clouds were assessed by the mean intersection-over-union (*m*IoU) and mean accuracy (*m*A). These metrics are widely used for assessing segmentation results.

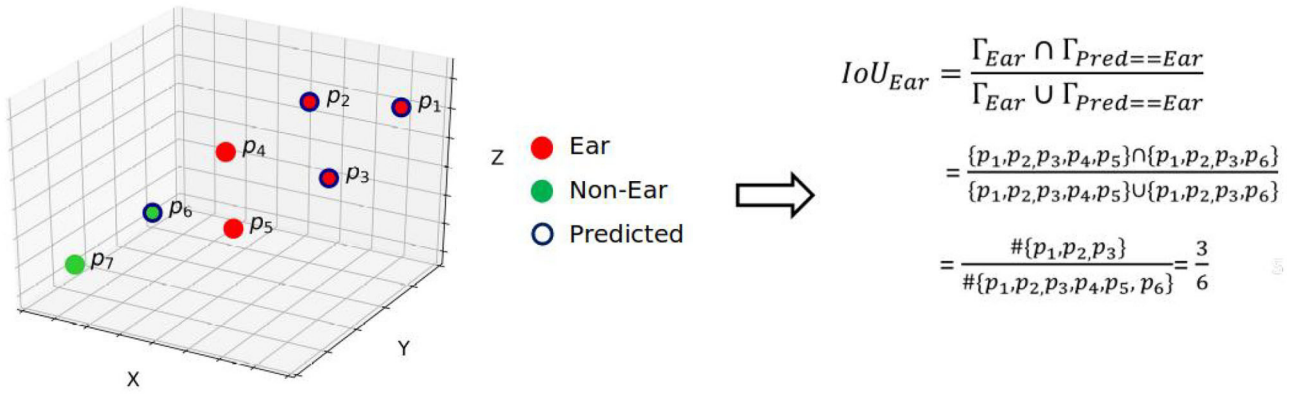

FIGURE 3 | Computation of IoU for the predicted ears in the point cloud domain.

According to commonly accepted definition, accuracy is the ratio of true predicted labels to the whole points and IoU is the number of points common between the labels (Γ) and predicted ones ($\hat{\Gamma}$) divided by the total number of points present across both the labels and predicted ones, i.e.,

$$IoU_k = \frac{\Gamma_k \cap \hat{\Gamma}_k}{\Gamma_k \cup \hat{\Gamma}_k},\ m \in \{1, \ldots, N\}. \tag{10}$$

The procedure for computing IoU of ears is illustrated in **Figure 3**. The average of all the organs' IoUs, i.e., $\frac{\sum_{k=1}^{N} IoU_k}{N}$, yields the *m*IoU. We also assessed the segmentation results using Pearson correlation coefficient (R^2) and root relative mean square (RRMSE):

$$RRMSE = \sqrt{\frac{1}{C}\sum_{i=1}^{C}\left(\frac{B_i - \hat{B}_i}{B_i}\right)^2}, \tag{11}$$

where B_i is the ground-truth counted ears and $\hat{B}_i$ is the predicted ones. C is the total number of point clouds processed and it equals 690 in this study.

Given paired data $\left\{(B_1, \hat{B}_1), \ldots, (B_i, \hat{B}_i)\right\}$ consisting of C pairs, Pearson correlation coefficient R^2 is defined as:

$$R^2 = \frac{\sum_{i=1}^{C}(B_i - \bar{B})(\hat{B}_i - \bar{\hat{B}})}{\sqrt{\sum_{i=1}^{C}(B_i - \bar{B})^2}\sqrt{\sum_{i=1}^{C}(\hat{B}_i - \bar{\hat{B}})^2}}, \tag{12}$$

where

$$\bar{B} = \frac{1}{C}\sum_{i=1}^{C} B_i,\ \bar{\hat{B}} = \frac{1}{C}\sum_{i=1}^{C} \hat{B}_i. \tag{13}$$

4.3. Data Preparation for Training and Testing

The wheat dataset was randomly split into 580 training, 30 validation and 80 test samples. The code was implemented in TensorFlow 1.12 (Abadi et al., 2015) on a 64-bit computer with Intel(R) Xeon(R) Gold 6130 CPU @ 2.10 GHz processors, 48 GB RAM, and two Tesla P100-PCIE-16GB GPU devices. The entire model was trained by minimizing the loss function stated in Equation (8). We used the Adam optimization algorithm with a constant learning rate of 0.001, and we reduced the learning rate until 0.0001 using the exponential decay function. Since there exists a direct relationship between the complexity and the required GPU resources, we have also carried out the training procedure on a light version of Pattern-Net, called light Pattern-Net, where the size of MLPs is half of the Pattern-Net, i.e., 16. The batch size, hyperparameter λ and parameter L were set to 10, 10,000, and 8, respectively. Because of the agronomic importance of the ear, C_{ear} in Equation (9) was set to 1 and the other category, i.e., $C_{non-ear}$, was set to 0.95. During the training step, the point clouds were augmented by randomly rotating, scaling and translating, in order to ensure that the network was transformation invariant, required by *Property II*.

TABLE 1 | Segmentation results of the proposed method on the wheat dataset.

Network	#Input points (xyz)	Mean accuracy (*m*A)%	Mean IoU (*m*IoU) %
Pattern-Net	512	91.17	80.19
	1,024	91.76	80.85
	2,048	92.07	81.26
	4,096	92.20	81.34
	8,192	92.27	81.74
Light Pattern-Net	10,240	87.32	76.47
	12,288	87.61	76.77
	14,336	87.97	77.12
	16,384	88.13	77.25

4.4. Results

The results are summarized in **Table 1**. The light Pattern-Net version works quite well but the most promising results are

obtained by the Pattern-Net. Accuracy of above 91% indicates that when we increase the number of 3D points from 512 to 8,192, both the mean accuracy and the mean IoU results of the network are improved, as expected. Samples of results (**Figure 4**) show that the difference between the predicted labels and the reference mainly occurred in the border between the ear and the non-ear regions. This aspect of Pattern-Net is more favorable when we measure the dimension of attributes of interest. As seen in the table, the mean IoU of dimensions of segmented organs is above 80%. Deep learning-based networks can be improved by increasing the number of input samples. So, if one needs higher precision in the test experiments, then the network must be trained with additional relevant samples. We also carried out experiments for inputs with more than 8,192 points. To this end, we had to decrease the size of MLPs to half of the original because of a limitation in RAM available in our GPU. As shown in **Table 1**, the light Pattern-Net still works well with mean accuracy around 87% and achieving 88.13% mean accuracy for input point clouds of size 16,384 points. Typically, 16,384 points is considered to represent a dense model for plants with dimensions of <50 cm (height) × 50 cm (width) × 50 cm (length).

The impact of the coefficient vector, Ω, defined in Equation (9) is shown in **Figure 5**. The network works well when the weights are in the interval of [0.9,1] and achieves its best performance for $C_{ear} = 1.$ and $C_{non-ear} = 0.95$. The dynamic coefficients balance between the loss of the majority non-ears points and that of the minority ear ones during training. Since vector Ω is a predetermined hyperparameter, we need to tune this parameter just once during training and the test step does not require the vector. The R^2 and RRMSE results of the counted wheat samples with different ear numbers for training, validation, and test sets are reported in **Figure 6**. The ear number varies in the range of $\{0, 1, 2, \ldots, 8\}$. The R^2 results of the counted wheat between the automatic segmentation and the manually annotated ones in MeshLab are all higher than 0.91 and RRMSE all <0.3. The R^2 result of the validation step is less than that of the test one due to the lower number of wheat samples, which is 30. The R^2 of the counted wheat samples by Pattern-Net for the test dataset is more than 0.92, indicating the reliability of the proposed network in segmentation of the unseen test wheat samples.

4.5. The Manually Collected Post-data

Final post-harvest measurements of the two treatments for all 20 plants were collected manually. The difference between the predicted counted ears from Pattern-Net and the ground-truth data from physical post-harvest counting of ears was computed and the detailed distribution of errors is shown in **Figure 7**, where the ears of the most samples were counted correctly and the mean absolute difference of count errors is as low as 0.3. An important aspect of our method is that the length of ears was also predicted by the segmentation and their average results are shown in **Figure 8**. We collected the ground-truth values for ear length and plant height in MeshLab as well as by direct physical measurement of the plant material. For facilitating the comparison, the length of ears was normalized by the height of plants providing relative ear length. The R^2 of average relative ear length between the segmentation and the actual ground-truth is 0.67 for the plants grown under drought conditions, which is on par with 0.695 of the ground-truth values annotated in MeshLab. The difference is as small as 0.025 and this figure for the plants grown under well water conditions is about 0.06. To determine the basis for differences between the MeshLab ground truth and the segmented results from Pattern-Net, we carefully compared the two and found that the classification of the border region between ear and non-ear regions could influence the predicted length of ears (**Figure 4**). Accurate classification of the border region remains a challenging task that needs further investigation.

5. DISCUSSION

Geometrically accurate models of individuals that can be computationally interrogated would be of great value in quantifying and understanding phenotypic variation, both in fundamental biological studies as well in commercial production scenarios. Typical plants have a complex and variable body shape as well as a plastic developmental programme that can continue to alter their morphology across their entire life cycle. Their complex and variable shape present numerous challenges to building and analysing models at a speed and cost appropriate to their use, while progressive developmental change may necessitate repeated modeling of the same individual. The potential benefits of rapid cost-effective 3D modeling extend well beyond basic morphology, as many physiological processes also vary across the plant body, both spatially and temporally, so that emerging non-contact physiological assessment methods (Dieleman et al., 2019) often require complex correction for shape.

A number of different technologies have been developed, including LASER, Time of Flight, and LIDAR to capture information from living plants for modeling (Paulus, 2019). Medical imaging approaches, such as μCT scanning, have also been applied to plants, particularly for ears of wheat (Hughes et al., 2019) and analogous structures from other crops such as sorghum inflorescences (Li et al., 2020) but the trade-offs involved in image acquisition generally mean that the approach is applicable to either low numbers of complete plants or somewhat larger numbers of parts of plants. The capital investment in the scanning equipment is also substantial, putting this out of reach of most researchers. The image acquisition method we used is highly convenient in that it utilizes consumer-grade cameras and can be easily transferred to other labs and situations. The SfM method is widely used and the models produced are composed of 3D point clouds. These are a common format and there is much freely available software, such as MeshLab, for converting them into virtual objects with solid surfaces that then can be imported into CAD packages (for engineering, generally) or other analysis pipelines where features can be extracted, identified, and/or estimated. This approach works quite well for geometrically simple objects that generate clean simple models with relatively few outliers in the point cloud. However, plants are

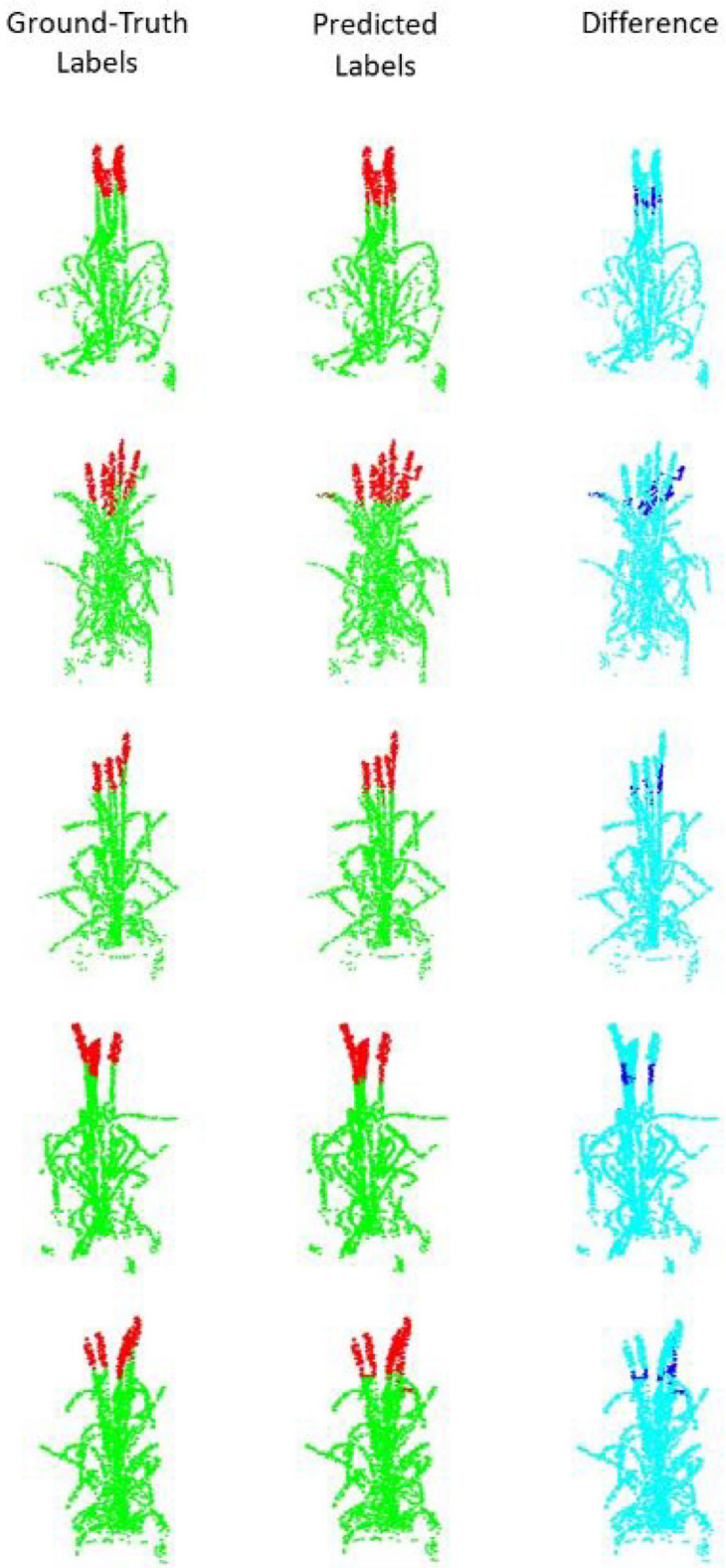

FIGURE 4 | Segmentation results for point clouds containing 2,048 points. **(Left)** The ground truth samples (ears are shown in red and non-ears in green) annotated in MeshLab; **(middle)** The predicted/segmented labels; **(Right)** The difference between the ground truth and the segmented results is shown in dark blue.

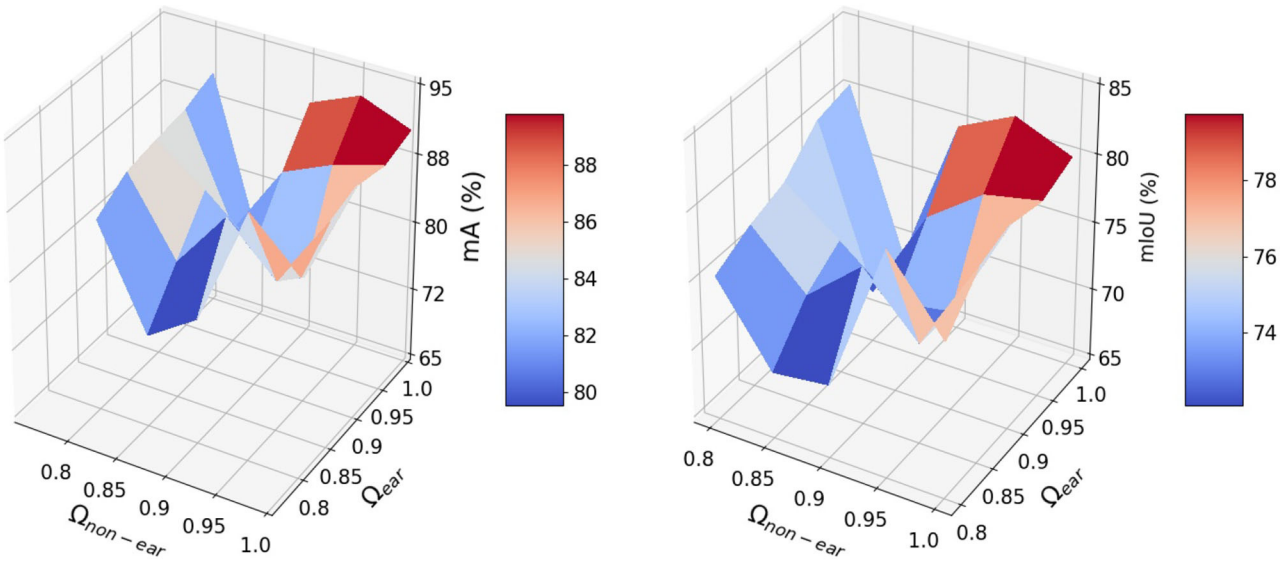

FIGURE 5 | The influence of the dynamic coefficient Ω on the segmentation results.

complex topologically and extensive occlusion tends to yield sub-optimal models that do not lend themselves to being converted to accurate surface-based models—on one hand, the outlying points tend to create spurious surfaces while on the other hand, occlusion and other imaging issues can lead to artifacts such as "holes" where there should be "tissue." To solve these issues, various modifications to surface-based approaches have been developed with some success: we previously used a projection method to assess leaf angle during the imposition of drought stress in grapevines (Briglia et al., 2020). Pound et al. (2016) used an elegant patch and boundary-refinement method to reconstruct accurate models of wheat and rice leaves that they could extend to whole canopies.

However, Pattern-Net bypasses many of these issues by undertaking much of the analysis directly in the point cloud domain. Our results indicate that Pattern-Net can detect, classify, and measure features directly in the 3D point clouds with sufficient accuracy to compare with manual phenotyping. Also, and notwithstanding the current limitations on GPU resources, Pattern-Net can already be scaled to accommodate the analyses of many 100's of individual models. With access to more powerful facilities, we envisage that Pattern-Net would be capable of supporting longitudinal phenotyping of large genetically defined populations, such as MAGIC and diversity mapping populations (Camargo et al., 2016).

We and others have previously reported methods to produce models based on 3D point clouds and to identify biologically relevant features, including from wheat (Liu et al., 2018) and from diverse other species (Lou et al., 2014; Briglia et al., 2020). Different published ear detection methods compared with manual counting indicate Pattern-Net has a high level of correct feature identification ($R^2 > 0.9$). Fernandez-Gallego et al. (2018) achieved correlations of up to $R^2 = 0.75$ between their computer vision method using 2D images of field grown wheat and manual counting. Sadeghi-Tehran et al. (2019) used superpixels and CNN pretrained by a VGG16 model[5] to achieve R^2 of 0.94 on 126 test images. TasselNetV2+ (Lu and Cao, 2020) achieved $R^2 = 0.91$ on the WEDD[6] dataset (Madec et al., 2019). We tested TasselNetV2+ on our multi-view wheat samples. We used the pre-trained model released by the authors[7] and the images were resized to 1,280 × 720 px. Since each sample consists of 74 multi-view images which are highly occluded, we ran TasselNetV2+ over all 74 images for each individual plant and took the maximum values as the predicted number of ears. The performance of TasselNetV2+ is shown in **Figure 7**. We also developed an image-based CNN using Faster RCNN ResNet101[8]. Faster RCNN was trained on the WEDD dataset. In both cases, the image-based techniques show lower accuracy compared to our 3D-based pipeline (**Figure 7**). The presence of occlusion in 2D images is inevitable, and the 3D-based pipeline can better deal with this problem. 3D models provide realistic depth that allows one to explore more accurately and enrich our understanding of the plant structures. The high cost of computing memory, however, is still a big challenge for processing in 3D space. Pattern-Net and its light version need $1.1M$ and $514K$ parameters, respectively. Our network gets to 92.3% test accuracy in 300 epochs of training, where the running-time for input 8,192 points is 253 seconds per epoch. The training time for the light Pattern-Net is 406 seconds per epoch for the input of 16,384 points.

It should be noted here that we used only a single variety of wheat, Paragon, whereas some of the 2D performance is

[5]https://keras.io/api/applications/vgg/
[6]https://github.com/simonMadec/Wheat-Ears-Detection-Dataset
[7]https://github.com/poppinace/tasselnetv2plus
[8]https://docs.openvinotoolkit.org/2020.2/_models_intel_faster_rcnn_resnet101_coco_sparse_60_0001_description_faster_rcnn_resnet101_coco_sparse_60_0001.html

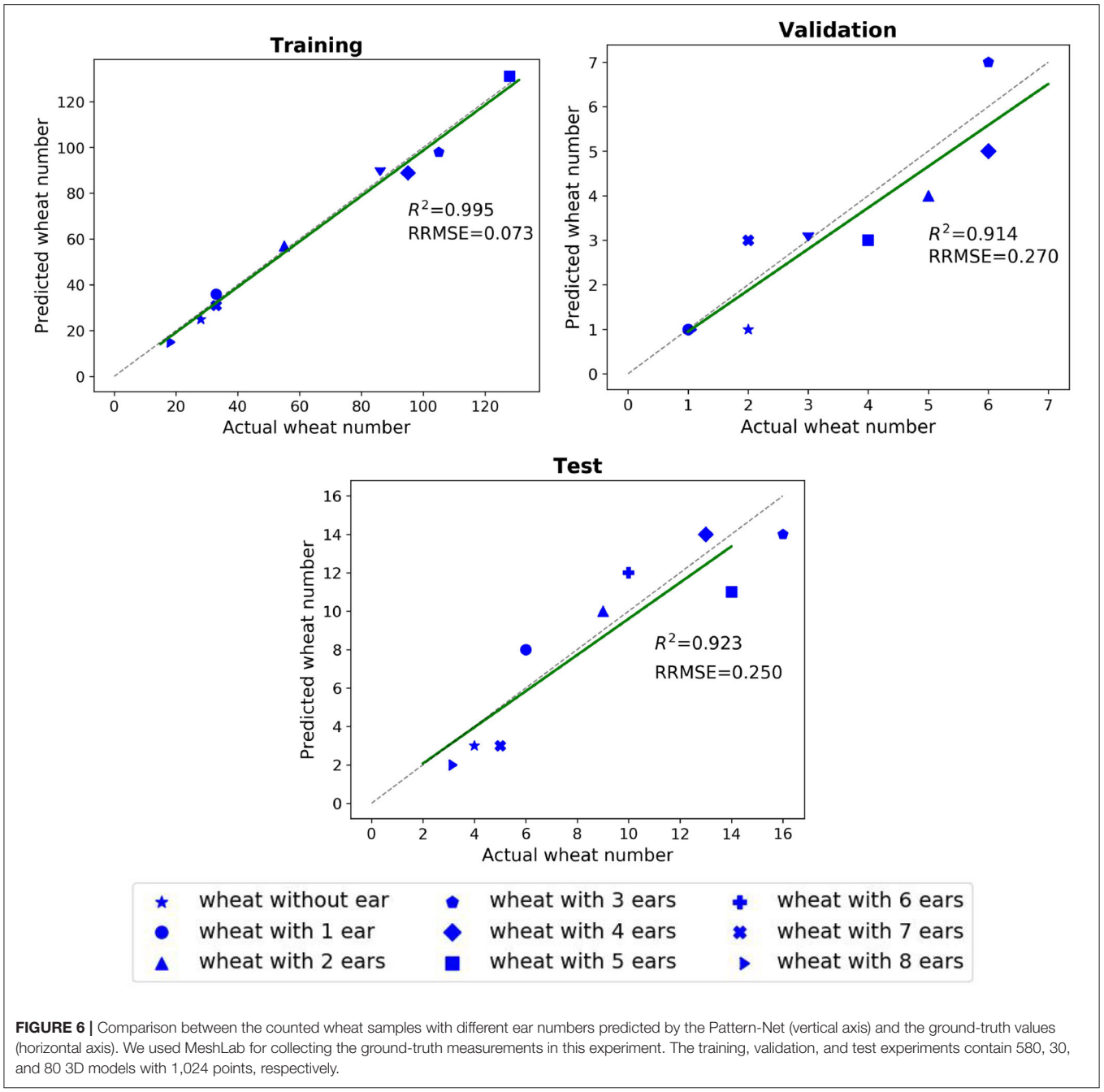

FIGURE 6 | Comparison between the counted wheat samples with different ear numbers predicted by the Pattern-Net (vertical axis) and the ground-truth values (horizontal axis). We used MeshLab for collecting the ground-truth measurements in this experiment. The training, validation, and test experiments contain 580, 30, and 80 3D models with 1,024 points, respectively.

given over many varieties and under less constrained imaging conditions (outdoors). Therefore, it is likely that Pattern-Net would require additional training before applying to other wheat cultivars or related cereals. Also, the definition of the boundary zone between ear and non-ear could be improved. This issue has arisen previously in the 2D analysis of rice panicles (the equivalent grain bearing structure to ears in wheat) and been solved by dual imaging with higher and lower resolution cameras followed by co-registration and a bespoke analysis pipeline (Huang et al., 2013). While many computer vision methods, in both 2D and 3D domains, can provide accurate feature recognition and counting, measurement of those features remains a challenge for plant phenotyping. We previously used an indirect RCNN to detect leaves in the 2D images projected from 3D point cloud models of grapevines subjected to drought and successfully quantified leaf angle to estimate a plant's response to stress (Briglia et al., 2020). Pattern-Net is capable of not only recognizing and counting ears accurately but also estimating their length, all within the 3D domain. Notwithstanding the issues associated with accurate recognition of the ear-non ear boundary in the point cloud, the output from Pattern-Net was well correlated ($R^2 > 0.6$) with manual

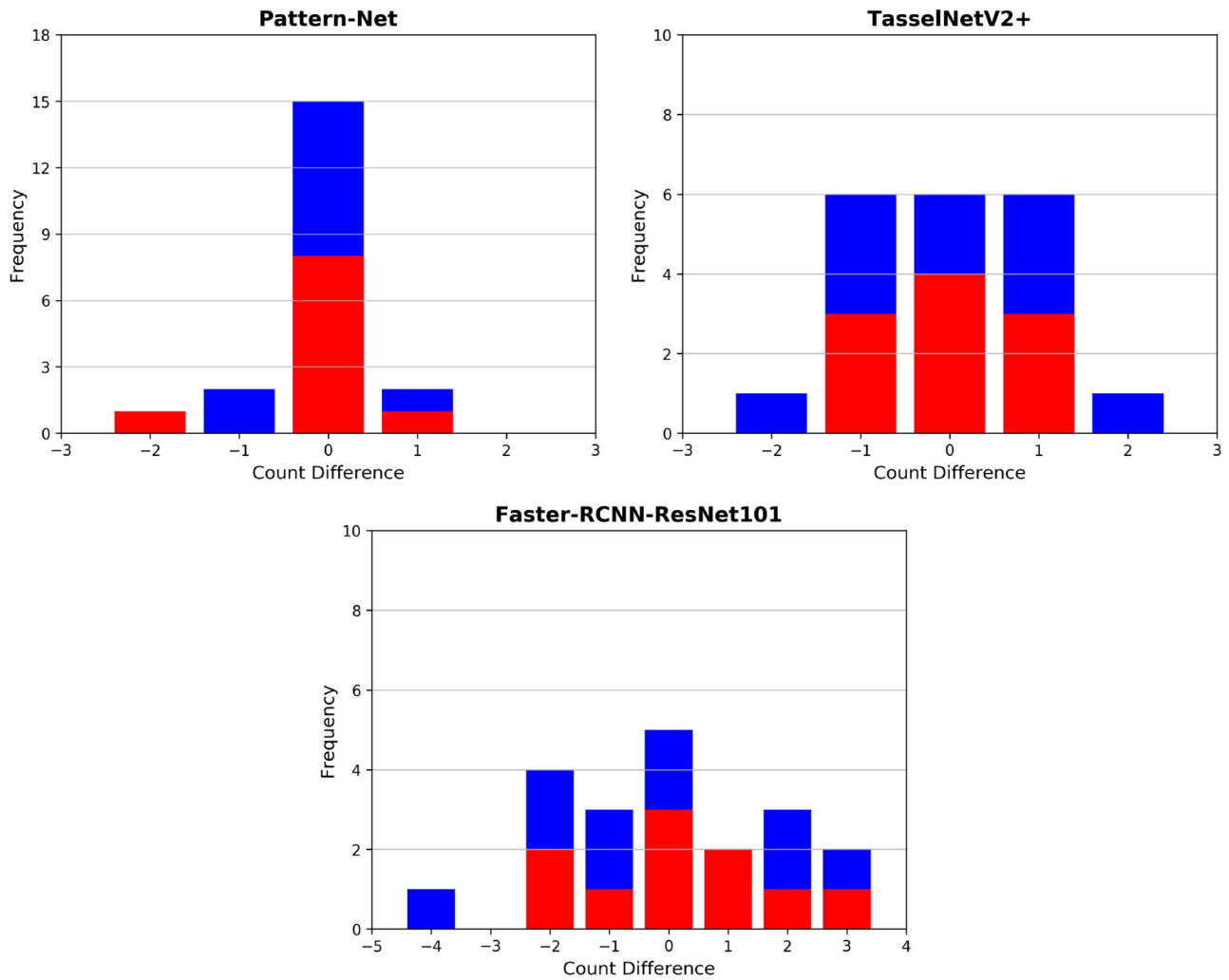

FIGURE 7 | Histogram of count errors between the ears predicted by our 3D-based pipeline and 2D image-based approaches and the corresponding physical ground-truth measurements collected post-harvest; the image-based techniques include TasselNetV2+ (Lu and Cao, 2020) and Faster RCNN. The results of the individuals grown under well-watered and drought conditions are shown in red and in blue, respectively. 3D models with 2,048 points were used here. 2D images with 1,280 × 720 px were used for the image-based techniques.

measurements for both well-watered and droughted plants. An innovation that may have helped modeling was additional viewpoints provided by the cameras. An interesting emerging approach is active imaging (Gibbs et al., 2018) where the camera(s), on a robotic arm, is relocated as required to overcome occlusion and to optimize the 3D model in a re-iterative manner. Such a system could be integrated into the conveyor system, in a similar manner to the dual-camera system used in this study. However, there are likely to be additional costs either in terms of image acquisition time, or computing power to ensure rapid real-time modeling and analysis.

To justify the additional costs, the 3D domain must add additional value and Pattern-Net begins to achieve this objective by providing quantification of a key morphological feature, ear length. This varies between cultivars and Siddique and Whan (1993) proposed that the ear to stem ratio might be a better indicator of yield potential than harvest index (HI) because the ratio is largely unperturbed by post-anthesis drought. They conceded that ear to stem ratio could only be used in early generations due to its labor-intensive data acquisition. Image-based approaches have the potential to reduce that labor burden, and Pattern-Net provides this metric as one of its outputs. As expected, the value of the ear: total plant height, manually measured or computationally inferred, increases slightly in the drought treatment and therefore Pattern-Net may be able to contribute to emerging Speed-Breeding (Watson et al., 2019).

6. CONCLUSION AND FUTURE WORK

In this study, we have developed a CNN method for direct segmentation of 3D point clouds that is less susceptible to

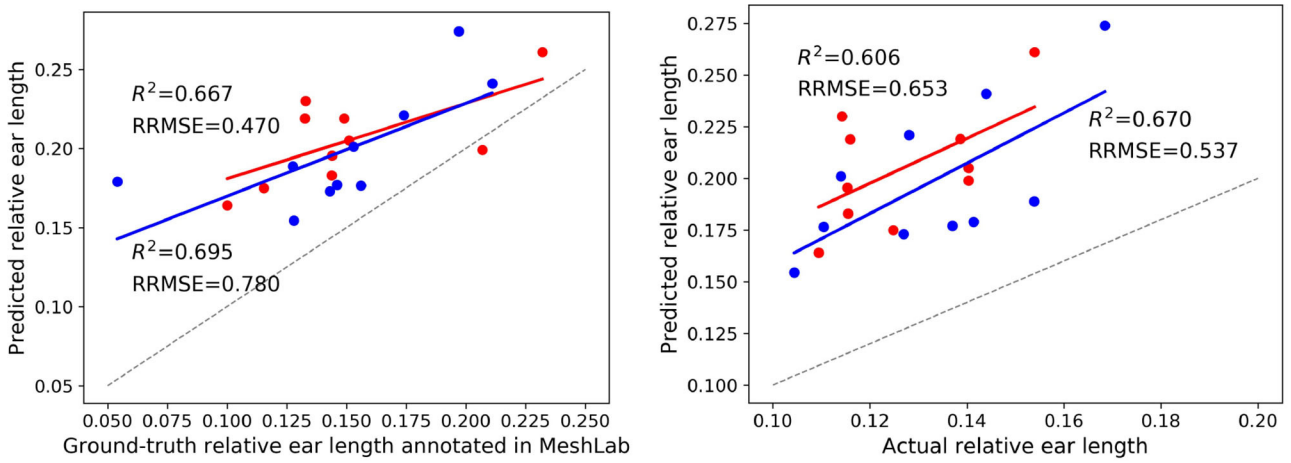

FIGURE 8 | Results on the ear length of two treatments including 10 individuals grown under well water conditions (in red) and 10 individuals grown under drought conditions (in blue). **(Left)** Comparison between the average relative ear length identified by the Pattern-Net and the corresponding ground-truth values measured in MeshLab. **(Right)** Comparison between the relative ear length identified by the Pattern-Net and the corresponding physical ground-truth measurements collected post-harvest. 3D models with 2,048 points were used here.

outliers. It is also invariant to changes in translation, rotation and scale. The key idea is to decompose the wheat point clouds into multiple subsets with similar structural information and then to force the network to learn and identify stable patterns. The network could successfully cope with the large-scale input point clouds ranging from 10,240 to 16,384 points and the results indicate that it is less prone to overfitting. This methodology provides a promising direction for robust analysis and understanding of plant point clouds although accurate estimation of ear length needs further improvement. While we have applied Pattern-Net to the relatively constrained datasets obtained from pot-grown wheat, this or similar approaches could be applied to field crops and canopies. The rapid and accurate assessment of the reproductive parts of many crops can be facilitated by image-based methods. For example, a dual-camera system has been developed for measuring harvested rice panicles (Huang et al., 2013). We expect that the principles developed within Pattern-Net can be applied to many other cereal crops, but in the context of intact plants.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MG, BT, YL, and JD designed the experiment. KW and FC undertook data acquisition, experimental design, and plant treatments. MG developed the deep learning and data analysis, testing and evaluation tasks, and drafted the manuscript. All authors have read and revised the manuscript.

FUNDING

This project was supported by BBSRC under grant numbers BB/S016538/1, BB/R02118X/1, BB/CCG1730/1, and BB/M0666/1. MG acknowledges DCDS and President scholarships from Aberystwyth University.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the HPC resources provided by Supercomputing Wales (SCW) and Aberystwyth University. We also thank Jason Brook and Karen Askew for technical assistance and Katie L. Awty-Carroll for allowing us to use the multi-view camera system prior to publication. We thank the Editor and two reviewers for their constructive comments that have improved the quality of the paper.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.

Alharbi, N., Zhou, J., and Wang, W. (2018). "Automatic counting of wheat spikes from wheat growth images," in *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM* (Funchal: SCITEPRESS-Science and Technology Publications), 346–355. doi: 10.5220/0006580403460355

Bernotas, G., Scorza, L. C., Hansen, M. F., Hales, I. J., Halliday, K. J., Smith, L. N., et al. (2019). A photometric stereo-based 3d imaging system using computer vision and deep learning for tracking plant growth. *GigaScience* 8:giz056. doi: 10.1093/gigascience/giz056

Bi, K., Jiang, P., Li, L., Shi, B., and Wang, C. (2010). Non-destructive measurement of wheat spike characteristics based on morphological image processing. *Trans. Chin. Soc. Agric. Eng.* 26, 212–216. doi: 10.3969/j.issn.1002-6819.2010.12.036

Brake, D. A., Hauenstein, J. D., Schreyer, F.-O., Sommese, A. J., and Stillman, M. E. (2019). Singular value decomposition of complexes. *SIAM J. Appl. Algeb. Geomet*. 3, 507–522. doi: 10.1137/18M1189270

Briglia, N., Williams, K., Wu, D., Li, Y., Tao, S., Corke, F., et al. (2020). Image-based assessment of drought response in grapevines. *Front. Plant Sci*. 11:595. doi: 10.3389/fpls.2020.00595

Camargo, A. V., Mott, R., Gardner, K. A., Mackay, I. J., Corke, F., Doonan, J. H., et al. (2016). Determining phenological patterns associated with the onset of senescence in a wheat magic mapping population. *Front. Plant Sci*. 7:1540. doi: 10.3389/fpls.2016.01540

Chaudhury, A., and Godin, C. (2020). Skeletonization of plant point cloud data using stochastic optimization framework. *Front. Plant Sci*. 11:773. doi: 10.3389/fpls.2020.00773

Chaudhury, A., Ward, C., Talasaz, A., Ivanov, A. G., Brophy, M., Grodzinski, B., et al. (2018). Machine vision system for 3d plant phenotyping. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 16, 2009–2022. doi: 10.1109/TCBB.2018.2824814

Chen, Z.-Z., Feng, Q., Shen, C., Wang, J., and Wang, L. (2016). Algorithms for pedigree comparison. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 15, 422–431. doi: 10.1109/TCBB.2016.2550434

Chopin, J., Laga, H., and Miklavcic, S. J. (2016). A hybrid approach for improving image segmentation: application to phenotyping of wheat leaves. *PLoS ONE* 11:e0168496. doi: 10.1371/journal.pone.0168496

Codreanu, V., Dong, F., Liu, B., Roerdink, J. B., Williams, D., Yang, P., et al. (2013). "GPU-asift: a fast fully affine-invariant feature extraction algorithm," in *2013 International Conference on High Performance Computing & Simulation (HPCS)* (Helsinki), 474–481. doi: 10.1109/HPCSim.2013.6641456

Dieleman, J. A., De Visser, P. H. B., Meinen, E., Grit, J. G., and Dueck, T. A. (2019). Integrating morphological and physiological responses of tomato plants to light quality to the crop level by 3d modeling. *Front. Plant Sci*. 10:839. doi: 10.3389/fpls.2019.00839

Elnashef, B., Filin, S., and Lati, R. N. (2019). Tensor-based classification and segmentation of three-dimensional point clouds for organ-level plant phenotyping and growth analysis. *Comput. Electron. Agric*. 156, 51–61. doi: 10.1016/j.compag.2018.10.036

Fernandez-Gallego, J. A., Kefauver, S. C., Gutiérrez, N. A., Nieto-Taladriz, M. T., and Araus, J. L. (2018). Wheat ear counting in-field conditions: high throughput and low-cost approach using RGB images. *Plant Methods* 14, 1–12. doi: 10.1186/s13007-018-0289-4

Ferrante, A., Cartelle, J., Savin, R., and Slafer, G. A. (2017). Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. *Field Crops Res*. 203, 114–127. doi: 10.1016/j.fcr.2016.12.028

Furukawa, Y., and Ponce, J. (2010). Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell*. 32, 1362–1376. doi: 10.1109/TPAMI.2009.161

Gélard, W., Devy, M., Herbulot, A., and Burger, P. (2017). "Model-based segmentation of 3d point clouds for phenotyping sunflower plants," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4 VISAPP* (Porto), 459–467. doi: 10.5220/0006126404590467

Ghahremani, M., Liu, Y., and Tiddeman, B. (2021). Ffd: Fast feature detector. *IEEE Trans. Image Process*. 30, 1153–1168. doi: 10.1109/TIP.2020.3042057

Ghahremani, M., Tiddeman, B., Liu, Y., and Behera, A. (2020). "Orderly disorder in point cloud domain," in *European Conference on Computer Vision* (Glasgow: Springer), 494–509. doi: 10.1007/978-3-030-58604-1_30

Ghahremani, M., Zhao, Y., Tiddeman, B., and Liu, Y. (2021). Interwoven texture-based description of interest points in images. *Pattern Recogn*. 2021:107821. doi: 10.1016/j.patcog.2021.107821

Gibbs, J. A., Pound, M., French, A. P., Wells, D. M., Murchie, E., and Pridmore, T. (2018). Plant phenotyping: an active vision cell for three-dimensional plant shoot reconstruction. *Plant Physiol*. 178, 524–534. doi: 10.1104/pp.18.00664

Guerrero, P., Kleiman, Y., Ovsjanikov, M., and Mitra, N. J. (2018). PCPNet learning local shape properties from raw point clouds. *Geometry Learning*. 37, 75–85. doi: 10.1111/cgf.13343

Guo, J., and Xu, L. (2017). Automatic segmentation for plant leaves via multiview stereo reconstruction. *Math. Probl. Eng*. 2017:9845815. doi: 10.1155/2017/9845815

Hamidinekoo, A., Garzón-Martínez, G. A., Ghahremani, M., Corke, F. M., Zwiggelaar, R., Doonan, J. H., et al. (2020). Deeppod: a convolutional neural network based quantification of fruit number in arabidopsis. *GigaScience* 9:giaa012. doi: 10.1093/gigascience/giaa012

Hasan, M. M., Chopin, J. P., Laga, H., and Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* 14:100. doi: 10.1186/s13007-018-0366-8

Huang, C., Yang, W., Duan, L., Jiang, N., Chen, G., Xiong, L., et al. (2013). Rice panicle length measuring system based on dual-camera imaging. *Comput. Electron. Agric*. 98, 158–165. doi: 10.1016/j.compag.2013.08.006

Hughes, N., Oliveira, H. R., Fradgley, N., Corke, F. M., Cockram, J., Doonan, J. H., et al. (2019). μ ct trait analysis reveals morphometric differences between domesticated temperate small grain cereals and their wild relatives. *Plant J*. 99, 98–111. doi: 10.1111/tpj.14312

Jay, S., Rabatel, G., Hadoux, X., Moura, D., and Gorretta, N. (2015). In-field crop row phenotyping from 3d modeling performed using structure from motion. *Comput. Electron. Agric*. 110, 70–77. doi: 10.1016/j.compag.2014.09.021

Jin, S., Su, Y., Gao, S., Wu, F., Hu, T., Liu, J., et al. (2018). Deep learning: individual maize segmentation from terrestrial lidar data using faster r-CNN and regional growth algorithms. *Front. Plant Sci*. 9:866. doi: 10.3389/fpls.2018.00866

Kun, B., Pan, J., Chongwei, T., Feifei, H., and Cheng, W. (2011). The design of wheat variety bp classifier based on wheat ear feature. *Chinese Agric. Sci. Bull*. 6, 465–467. Available online at: https://en.cnki.com.cn/Article_en/CJFDTotal-ZNTB201106094.htm

Landrieu, L., and Simonovsky, M. (2018). "Large-scale point cloud semantic segmentation with superpoint graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 4558–4567. doi: 10.1109/CVPR.2018.00479

Li, D., Cao, Y., Shi, G., Cai, X., Chen, Y., Wang, S., et al. (2019). An overlapping-free leaf segmentation method for plant point clouds. *IEEE Access* 7, 129054–129070. doi: 10.1109/ACCESS.2019.2940385

Li, D., Cao, Y., Tang, X.-S., Yan, S., and Cai, X. (2018). Leaf segmentation on dense plant point clouds with facet region growing. *Sensors* 18:3625. doi: 10.3390/s18113625

Li, M., Shao, M.-R., Zeng, D., Ju, T., Kellogg, E. A., and Topp, C. N. (2020). Comprehensive 3d phenotyping reveals continuous morphological variation across genetically diverse sorghum inflorescences. *New Phytol*. 226, 1873–1885. doi: 10.1111/nph.16533

Liu, J., Liu, Y., and Doonan, J. (2018). "Point cloud based iterative segmentation technique for 3d plant phenotyping," in *2018 IEEE International Conference on Information and Automation (ICIA)* (Fujian), 1072–1077. doi: 10.1109/ICInfA.2018.8812589

Lou, L., Liu, Y., Han, J., and Doonan, J. H. (2014). "Accurate multi-view stereo 3d reconstruction for cost-effective plant phenotyping," in *International Conference Image Analysis and Recognition* (Vilamoura: Springer), 349–356. doi: 10.1007/978-3-319-11755-3_39

Lu, H., and Cao, Z. (2020). Tasselnetv2+: A fast implementation for high-throughput plant counting from high-resolution RGB imagery. *Front. Plant Sci*. 11:1929. doi: 10.3389/fpls.2020.541960

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol*. 264, 225–234. doi: 10.1016/j.agrformet.2018.10.013

Misra, T., Arora, A., Marwaha, S., Chinnusamy, V., Rao, A. R., Jain, R., et al. (2020). Spikesegnet-a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* 16, 1–20. doi: 10.1186/s13007-020-00582-9

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci*. 7:1419. doi: 10.3389/fpls.2016.01419

Paulus, S. (2019). Measuring crops in 3d: using geometry for plant phenotyping. *Plant Methods* 15, 1–13. doi: 10.1186/s13007-019-0490-0

Paulus, S., Dupuis, J., Mahlein, A.-K., and Kuhlmann, H. (2013). Surface feature based classification of plant organs from 3d laserscanned point clouds for plant phenotyping. *BMC Bioinformatics* 14:238. doi: 10.1186/1471-2105-14-238

Penrose, R. (1955). A generalized inverse for matrices. *Math. Proc. Cambrid. Philos. Soc*. 51, 406–413. doi: 10.1017/S0305004100030401

Pound, M. P., French, A. P., Fozard, J. A., Murchie, E. H., and Pridmore, T. P. (2016). A patch-based approach to 3d plant shoot phenotyping. *Mach. Vis. Appl.* 27, 767–779. doi: 10.1007/s00138-016-0756-8

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017a). "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 652–660.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017b). "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems* (California), 5099–5108.

Qiongyan, L., Cai, J., Berger, B., Okamoto, M., and Miklavcic, S. J. (2017). Detecting spikes of wheat plants using neural networks with laws texture energy. *Plant Methods* 13:83. doi: 10.1186/s13007-017-0231-1

Ranzuglia, G., Callieri, M., Dellepiane, M., Cignoni, P., and Scopigno, R. (2013). "Meshlab as a complete tool for the integration of photos and color with high resolution 3d geometry data," in *CAA 2012 Conference Proceedings* (Southampton: Pallas Publications; Amsterdam University Press), 406–416.

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., and Hawkesford, M. J. (2019). Deepcount: In-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks. *Front. Plant Sci.* 10:1176. doi: 10.3389/fpls.2019.01176

Schönberger, J. L., and Frahm, J.-M. (2016). "Structure-from-motion revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV), 4104–4113. doi: 10.1109/CVPR.2016.445

Schönberger, J. L., Zheng, E., Pollefeys, M., and Frahm, J.-M. (2016). "Pixelwise view selection for unstructured multi-view stereo," in *European Conference on Computer Vision (ECCV)* (Amsterdam), 501–518. doi: 10.1007/978-3-319-46487-9_31

Shen, Y., Feng, C., Yang, Y., and Tian, D. (2018). "Mining point cloud local structures by kernel correlation and graph pooling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 4548–4557. doi: 10.1109/CVPR.2018.00478

Shi, W., van de Zedde, R., Jiang, H., and Kootstra, G. (2019). Plant-part segmentation using deep learning and multi-view vision. *Biosyst. Eng.* 187, 81–95. doi: 10.1016/j.biosystemseng.2019.08.014

Siddique, K., and Whan, B. (1993). Ear: stem ratios in breeding populations of wheat: significance for yield improvement. *Euphytica* 73, 241–254. doi: 10.1007/BF00036703

Tan, C., Zhang, P., Zhang, Y., Zhou, X., Wang, Z., Du, Y., et al. (2020). Rapid recognition of field-grown wheat spikes based on a superpixel segmentation algorithm using digital images. *Front. Plant Sci.* 11:259. doi: 10.3389/fpls.2020.00259

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* 8:1190. doi: 10.3389/fpls.2017.01190

Wang, X., Xuan, H., Evers, B., Shrestha, S., Pless, R., and Poland, J. (2019). High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat. *GigaScience* 8:giz120. doi: 10.1101/527911

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.* 38:146. doi: 10.1145/3326362

Watson, A., Hickey, L. T., Christopher, J., Rutkoski, J., Poland, J., and Hayes, B. J. (2019). Multivariate genomic selection and potential of rapid indirect selection with speed breeding in spring wheat. *Crop Sci.* 59, 1945–1959. doi: 10.2135/cropsci2018.12.0757

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., et al. (2020). Wheat ear counting using k-means clustering segmentation and convolutional neural network. *Plant Methods* 16, 1–13. doi: 10.1186/s13007-020-00648-8

Zhou, C., Liang, D., Yang, X., Xu, B., and Yang, G. (2018). Recognition of wheat spike from field based phenotype platform using multi-sensor fusion and improved maximum entropy segmentation algorithms. *Remote Sens.* 10:246. doi: 10.3390/rs10020246

Zhou, J., Fu, X., Zhou, S., Zhou, J., Ye, H., and Nguyen, H. T. (2019). Automated segmentation of soybean plants from 3d point cloud using machine learning. *Comput. Electron. Agric.* 162, 143–153. doi: 10.1016/j.compag.2019.04.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ghahremani, Williams, Corke, Tiddeman, Liu and Doonan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fpls.2021.634103

Tomato Anomalies Detection in Greenhouse Scenarios Based on YOLO-Dense

*Xuewei Wang and Jun Liu**

Shandong Provincial University Laboratory for Protected Horticulture, Blockchain Laboratory of Agricultural Vegetables, Weifang University of Science and Technology, Weifang, China

Greenhouse cultivation can improve crop yield and quality, and it not only solves people's daily needs but also brings considerable gains to the agricultural staff. One of the most widely cultivated greenhouse crops is tomato, mainly because of its high nutritional value and its good taste. However, there are a number of anomalies for the tomato crop that pose a threat for its greenhouse cultivation. Detection of tomato anomalies in the complex natural environment is an important research direction in the field of plant science. Automated identification of tomato anomalies is still a challenging task because of its small size and complex background. To solve the problem of tomato anomaly detection in the complex natural environment, a novel YOLO-Dense was proposed based on a one-stage deep detection YOLO framework. By adding a dense connection module in the network architecture, the network inference speed of the proposed model can be effectively improved. By using the K-means algorithm to cluster the anchor box, nine different sizes of anchor boxes with potential objects to be identified were obtained. The multiscale training strategy was adopted to improve the recognition accuracy of objects at different scales. The experimental results show that the mAP and detection time of a single image of the YOLO-Dense network is 96.41% and 20.28 ms, respectively. Compared with SSD, Faster R-CNN, and the original YOLOv3 network, the YOLO-Dense model achieved the best performance in tomato anomaly detection under a complex natural environment.

Keywords: deep learning, plant diseases recognition, DenseNet, real-field scenarios, object detection

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Jaafar Abdulridha, University of Florida, United States
Athanasios Balafoutis, Centre for Research and Technology Hellas (CERTH), Greece

***Correspondence:**
Jun Liu
liu_jun860116@wfust.edu.cn

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *27 November 2020*
Accepted: *09 March 2021*
Published: *09 April 2021*

Citation:
Wang X and Liu J (2021) Tomato Anomalies Detection in Greenhouse Scenarios Based on YOLO-Dense. Front. Plant Sci. 12:634103. doi: 10.3389/fpls.2021.634103

INTRODUCTION

With the development of the economy, agriculture is transforming from traditional to modern, and greenhouses, as an important support for modern agriculture, are widely used. A greenhouse can be kept untouched by the external environment and at the same time not be restricted by the geographical and seasonal factors of each crop cultivation, thus showing the capabilities of controlled environment agricultural production (Xie et al., 2017). Tomato, an important vegetable variety grown in greenhouses, is one of the most prevalent fruits and vegetables cultivated worldwide. Tomato is highly favored by global consumers because it is rich in antioxidant lycopene, multivitamins, and minerals and has the advantages of low heat and considerable cultivation benefits (Heuvelink, 2005). However, a research field with huge potential about tomato cultivation

is dealing with the detection of possible threats because crop diseases are one of the main factors affecting the yield and quality of agricultural products.

Diseased crops usually show discoloration, necrosis, deformity, decay, and wilting after infection (Alemu, 2015). Most of the disease phenomena will be reflected in the leaves of crops (Pethybridge and Nelson, 2015), and the judgment of diseases through the leaves of crops has become one of the common means in the field of agriculture (Barbedo, 2016). Traditional identification of crop diseases is accomplished manually by farmers in the field, and incorrect diagnosis and unnecessary pesticide application are common (Juncheng et al., 2018). Not only that, the traditional manual judgment method has the disadvantages of being time-consuming and labor-intensive, and human-subjective factors play a major role (Ghazi et al., 2017).

China is one of the largest tomato-producing and -consuming countries in the world (Xu et al., 2000), and tomato production is one of the important ways for farmers to increase their income (Liu, 2018). Early detection of tomato anomalies in the complex natural environment is of great significance to reduce the cost of manual identification and improve tomato quality and yield. The primary task and design difficulty are the real-time and accurate identification and spatial detection of tomato anomalies. The growth stages of tomato plants present high diversification, and tomato anomaly images in natural environments can be easily influenced by light, occlusion, and background noise, which cause some difficulties in the detection of tomato anomalies. Therefore, the rapid and accurate identification of tomato anomalies in the complex environment is a key problem in achieving an automated inspection of tomato anomalies.

When using object detection algorithms, usually feature extractors such as histogram of gradient (HOG), scale-invariant feature transform (SIFT), and Haar-like feature are manually designed to extract object features, which are then given as input to the support vector machine (SVM), iterator AdaBoost, random forest (RF), and other classifiers for classification and recognition. The most common object detection methods for tomato anomalies are basically based on color and shape features for feature extraction and analysis (Xu et al., 2011; Martinelli et al., 2015). The generalization of the most common method is poor, and it is difficult to extract reasonable features. Also, the calculation complexity is high, and therefore sometimes the requirements of accuracy and speed for real-time detection are not fulfilled. Most of them do not take into account the influencing factors in the complex environment of a greenhouse and have insufficient robustness against the changes of various features, so it is difficult to meet the actual requirements. Deep convolutional neural networks (DCNNs), which have emerged in recent years, provide a new idea for tomato anomaly object detection.

Deep learning-based detection algorithms can be divided into region-based and regression-based. The region-based method generates candidate regions by selective search algorithm and then classifies them using convolutional neural networks. A few of these methods are RCNN (Girshick et al., 2014), Fast R-CNN (Shahid et al., 2015), Faster R-CNN (Ren et al., 2017), and SPP-Net (He et al., 2014). This kind of region-based, two-step method can achieve high detection accuracy, but it has the disadvantages of complex network and slow detection speed. Regression-based methods such as SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016) frame take the object detection problem as a regression one, so the object class probability and position coordinates can be directly regressed. The YOLO series algorithm based on regression have fast processing speed and high accuracy, so they have been widely used in actual scenarios, such as fruit detection (Xu et al., 2020), mask-wearing detection (Ren and Liu, 2020), and traffic sign detection (Zhang et al., 2021). YOLOv2 (Redmon and Farhadi, 2017) and YOLOv3 (Li Y. et al., 2019) were improved on the basis of the YOLO algorithm, which further enhances the detection effect. However, the network structure of the fast regression-based detection algorithm remains large, and therefore the speed of deploying to embedded services is slow while the deployment cost is high.

Artificial intelligence has been widely used in agriculture in recent years (Tang et al., 2020). Farmers have gradually begun to use smartphones to detect crop anomalies (Prasad et al., 2014). In view of the problem of crop anomaly detection, Li W. et al. (2019) proposed a pipeline based on deep learning for locating and counting agricultural pests in image by a self-learning saliency feature map, and the average accuracy (mAP) achieved was 0.885. Xing et al. (2019) developed a simple but effective CNN model based on a self-built dataset, which increased the complexity of cross-channel operation and made the frequency of feature reuse adapt to the network depth, but this algorithm cannot be widely used in systems of general performance as it requires a high amount of computation. For the detection of tomato anomalies, Fuentes et al. (2018) proposed a detection algorithm in a real, natural environment based on Faster R-CNN, and the average accuracy it achieved was up to 96% for 10 common tomato anomalies including leaf mold, gray mold, canker, plague, miner, low temperature, powdery mildew, whitefly, nutritional excess, and yellow leaf curl. Zhang Y. et al. (2020) also proposed a detection method based on the improved Faster R-CNN algorithm, but this method has a slow detection speed and is not suitable for the real-time detection of tomato anomalies with large datasets.

YOLOv3 is an end-to-end object detection algorithm based on Darknet-53, and multiscale feature fusion is done via FPN (Feature Pyramid Networks) (Lin et al., 2017). YOLOv3 has the characteristics of fast detection speed and strong comprehensive performance, but when it is directly applied to certain specific detection objects, due to the influence of scene complexity and feature diversification, the detection effect cannot meet the requirements, so it needs to be improved. Zhang X. et al. (2020) improved YOLOv3 using an expanded spatial pyramid, achieving a speed of 56 frames/second and an average accuracy of 82.2%. Chen et al. (2020) designed an anchor box (a set of initial prediction boxes with fixed width and height, and the number and width of the anchor boxes clustered by different datasets are different) that was more suitable for face detection, and a new loss function was used to replace the square sum loss function, which reduced the model error and improved the convergence speed. Fu et al. (2020) automatically detected kiwifruit in orchards by

improving the YOLOv3-tiny model; two convolution cores of 3 × 3 and 1 × 1 were added to the fifth and sixth convolution layers of the YOLOv3-tiny model to develop the deep YOLOv3-tiny (DY3TNet) model. The experimental results showed that the improved DY3TNet model had small volume and reduced computational complexity, thus realizing real-time detection. Xu et al. (2020) improved YOLOv3 by using soft-NMS (non-maximum suppression) instead of NMS to reduce the loss of the prediction bounding box due to green mango overlap, which can meet the requirements of real-time detection for robotic picking.

The objective of this study is to introduce the idea of dense connection in DenseNet (Dense Convolutional Network) (Huang et al., 2016) into a YOLOv3 basic network, and a YOLO-Dense object detection algorithm is proposed. DenseNet is a densely connected network structure. All layers in the network are directly connected. The input of each layer in the network is the intersection of all the output layers in front, and the feature map learned in each layer of the network will also be directly transmitted to the latter layer as the output, so the multiplexing of features can be enhanced. The improved method begins by designing the YOLO Dense network based on the structure of Darknet-53 in YOLOv3. Then, the improved K-means clustering algorithm was used for calculating the anchor bounding boxes, in order to reduce the impact of the initial points on the clustering results. Finally, the last step of the method involves the multiscale training of the network.

METHOD FOR IMPROVING THE YOLOv3 MODEL

YOLOv3 has excellent detection effects in the field of object detection, but for the detection task of tomato anomalies in the complex natural environment, the network needs to be improved. In this study, an improved YOLO-Dense network model was proposed based on the characteristics of the self-made tomato anomaly dataset. The improvement scheme is shown in **Figure 1**.

In the YOLOv3 network structure, the Darknet-53 network was used to extract features. Due to the increase in the number of layers in the network, it is possible that an overfitting or gradient (diffusion, explosion) problem will occur. Therefore, the idea of residual network is used in the Darknet-53 network: the original output of a layer is directly connected to the back layer, and the residual layer is constructed between layers of the same dimension. The problem of gradient disappearance in deep neural network is solved by means of Jump-Layer connection.

This study draws on a DenseNet network idea: the input of each layer in the network is the sum of the output of all the layers in front, and the output of this layer will also propagate backward, becoming part of the input of the latter layer. The new YOLO-Dense network, which uses dense connection, can realize feature fusion through dimension connection on the channel, which is helpful for feature extraction of tomato anomalies. Meanwhile, it reduces parameters and calculation costs and improves network efficiency. Therefore, it is necessary to change part of the residual layer in the underlying network Darknet-53 of YOLOv3 into a densely connected network and refer to the DenseNet network for naming. The modified structure can make more effective use of the features extracted in the prediction layer, and thus its detection speed is faster than the original YOLOv3.

Figure 2 shows a descriptive graphical depiction of the YOLO-dense on how the system works.

The improved network structure model is shown in **Figure 3** below:

The specific ways of improvement are as follows: the dimensions of the 13th layer to the 36th layer of Darknet-53 are the same, and there is no need to be transformed, so the residual layer in Darknet-53 is changed to the dense connection layer, and the shortcut layers in the 15th, 18th, 21st, 24th, 27th, 30th, 33rd, and 36th layers of Darknet-53 are changed to route layers. The original residual layer is changed to dense connection layer to achieve the dense connection of the features of the same dimension, and the design was named

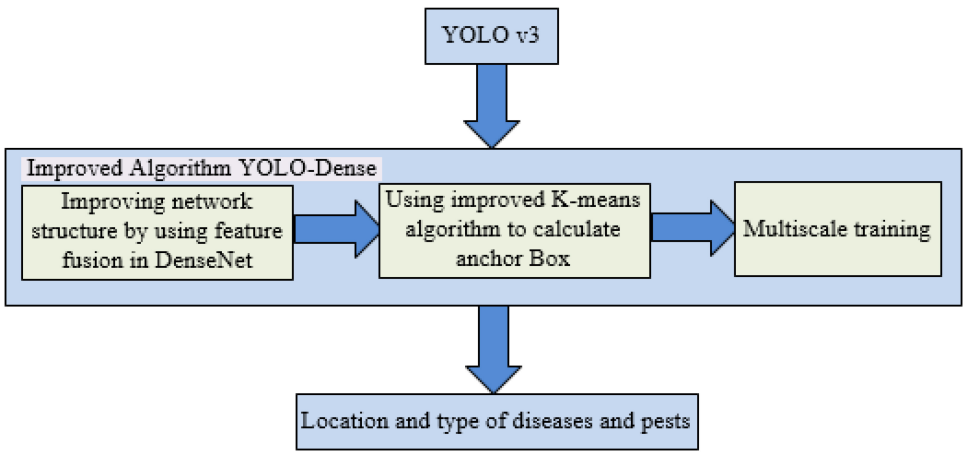

FIGURE 1 | Schematic diagram for improving the YOLOv3 model 2.1 YOLO-Dense network design.

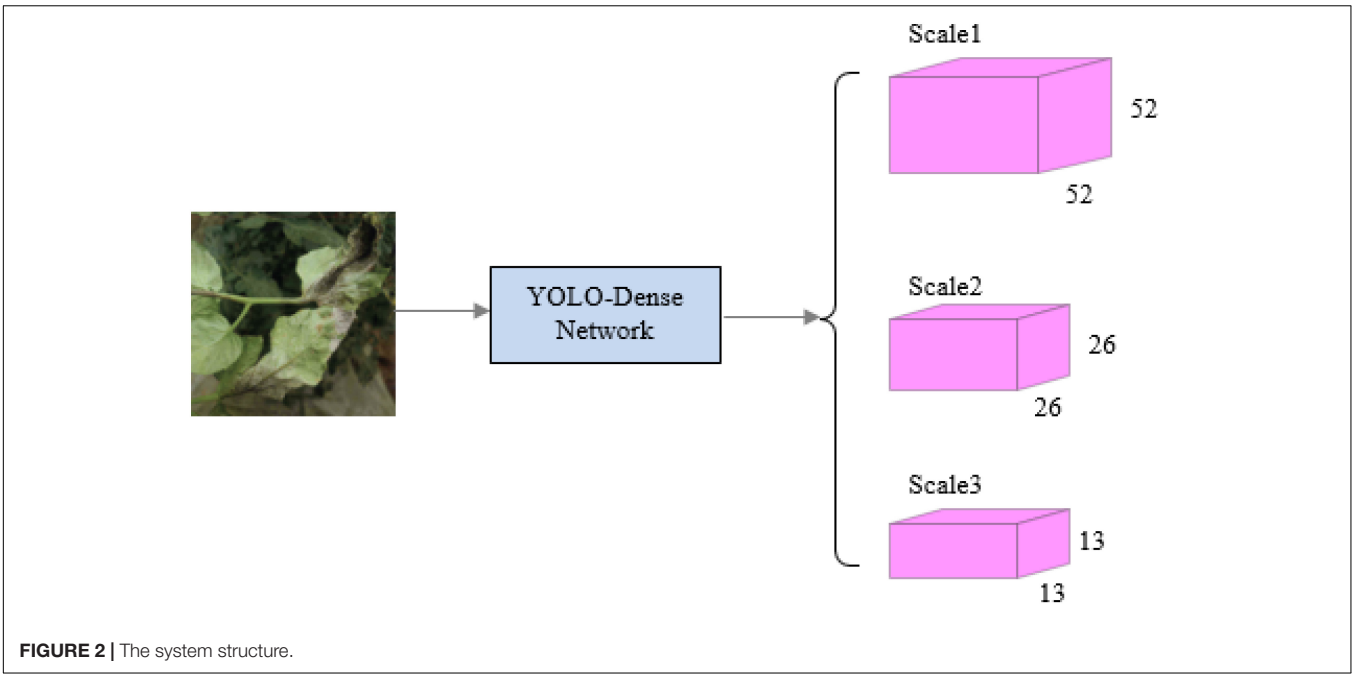

FIGURE 2 | The system structure.

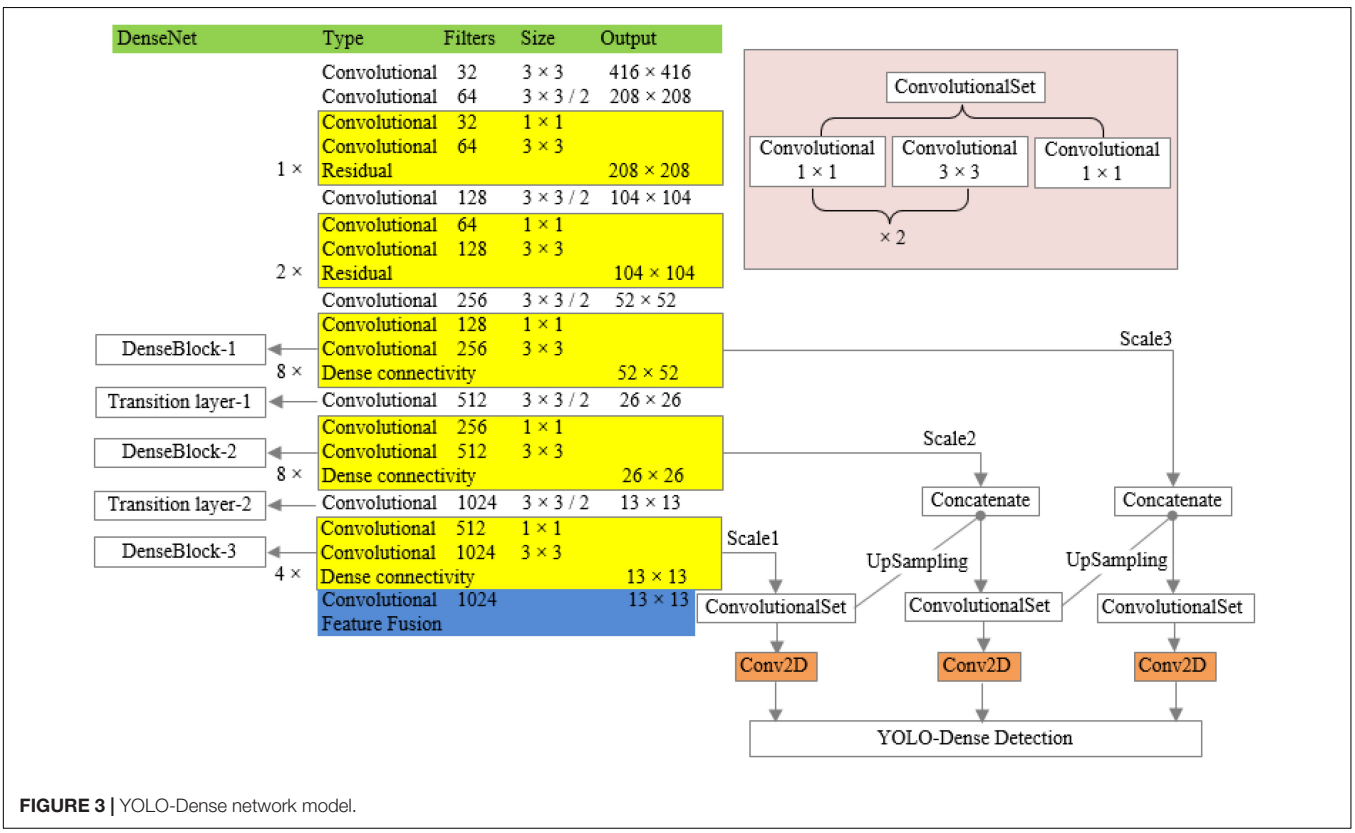

FIGURE 3 | YOLO-Dense network model.

DenseBlock-1. The shortcut layers of Layers 40, 43, 46, 49, 52, 55, 58, and 61 were changed to route layers, the original residual layer was changed to dense connection layer, and the design was named DenseBlock-2. Similarly, the shortcut layers of Layers 65, 68, 71, and 74 were changed to route layers, and the original residual layer was changed to dense connection layer, and the design is named DenseBlock-3. The 37-layer convolution layer in Darknet-53 is similar to the transition layer of the DenseNet network in function, both of which reduce the dimensionality of the output feature map, so the 37-layer

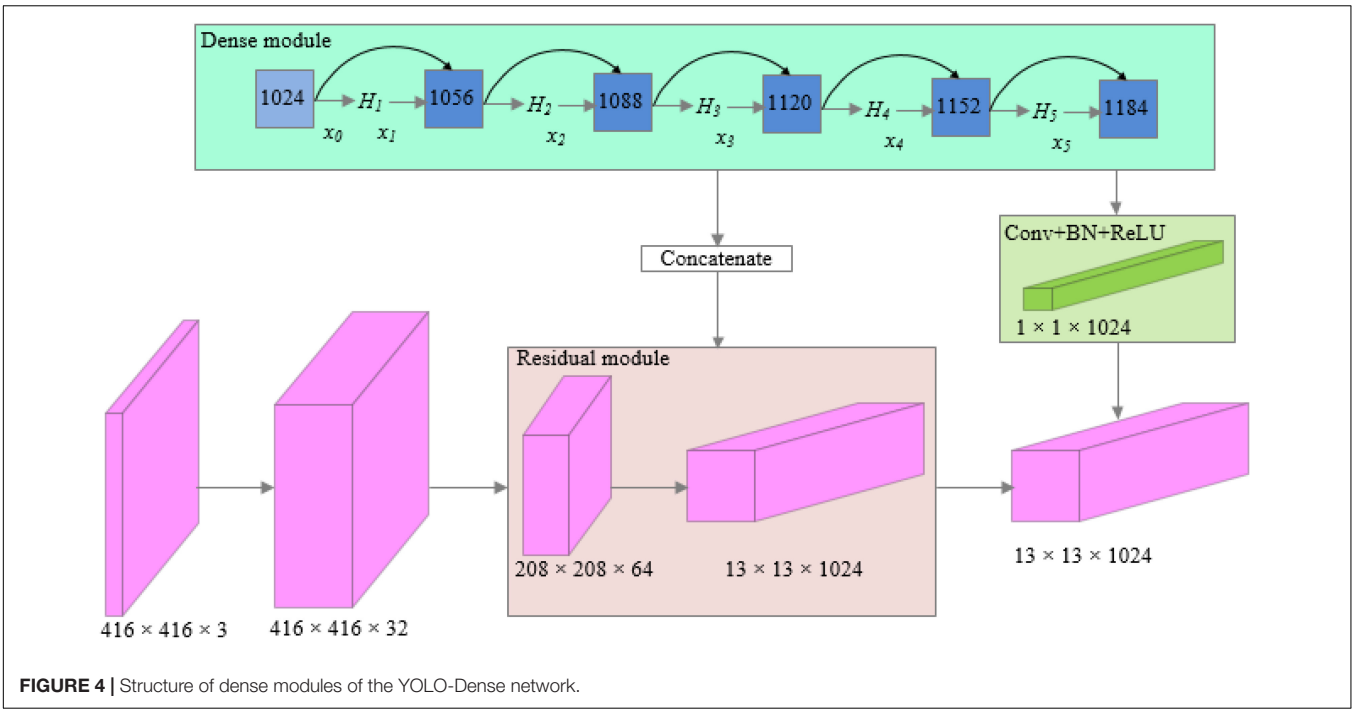

FIGURE 4 | Structure of dense modules of the YOLO-Dense network.

layer is renamed Transitionlayer-1; similarly, the 62-layer layer is renamed Transitionlayer-2.

The structure of dense modules of the improved network is shown in **Figure 4**, which can realize multilayer feature multiplexing and fusion and avoid the computational complexity caused by the new structure.

In the dense module:

$$x_l \; H_l\left(\left[x_0, x_1, \cdots, x_{l-1}\right]\right), \; l \; 1, 2, 3, 4, 5 \qquad (1)$$

In the abovementioned formula, and x_0 is the input feature map of the module, x_1 is the output of the first layer. $\left[x_0, x_1, \cdots, x_{l-1}\right]$ is the concatenation of $x_0, x_1, \cdots, x_{l-1}$. $H_l\left(\left[x_0, x_1, \cdots, x_{l-1}\right]\right)$ is the combination function of BN (batch normalization), ReLU (rectified linear units), and convolution, to realize the l layer nonlinear transformation.

The YOLO-Dense network uses the YOLO detection layer for class output and uses the three different prediction scales to detect objects of different sizes, with different prediction scales of 13 × 13, 26 × 26, and 52 × 52, respectively. The predictive scale output feature map has two sets of dimensions: the dimension of extracted features, such as 13 × 13, and the second dimension is calculated by using the following formula:

$$dimensionality \; + \; B \times (5 + C) \qquad (2)$$

In the abovementioned formula, B indicates the number of bounding boxes per prediction and C is the number of classes of the bounding boxes. So, another dimension is 3 × (5+6) = 33. In the output layer of the network, the Softmax classifier used in the original YOLO network cannot identify and locate two kinds of anomalies in the same grid correctly. Therefore, this study uses a separate logistic classifier for each different disease category to predict the confidence score that each anchor box belongs to a specific category and replaces the original Softmax classifier with it.

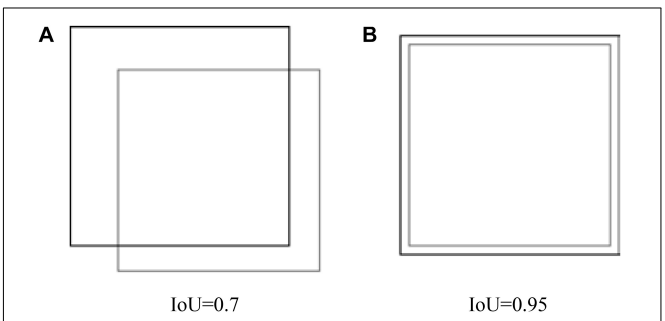

FIGURE 5 | IoU schematic. **(A)** IoU=0.7. **(B)** IoU=0.95.

Anchor Box Calculation Using the Improved K-Means Clustering Algorithm

YOLOv3 borrows the idea of using the anchor box in Faster R-CNN. The anchor box is used as *a priori* box to assist in predicting the object bounding box, which is designed according to different datasets. For the self-built dataset in this study, the anchor box needs to be recalculated.

The K-means algorithm usually uses Euclidean distance, Chebyshev distance, Manhattan distance, and other methods as distance measures to calculate the distance between two points. YOLO v3 has used the K-means clustering algorithm and got nine prediction boxes. The K-means clustering algorithm of YOLO v3 is based on features from the PASCAL-VOC dataset, which in

turn produces a prediction box, and since images in the PASCAL-VOC dataset have large gaps from tomato anomaly features, it has poor detection accuracy for tomato anomalies and is not suitable for tomato anomaly detection tasks. In addition, the K-means clustering algorithm randomly chooses K points as the initial clustering center (i.e., there are k classifications), and this random way increases the randomness of the clustering and affects the clustering effect of the algorithm. In this study, if the clustering algorithm uses these commonly used distances, the larger the candidate box generated, the greater the error, so it will not produce good detection results. The main focus of this study is small object detection; thus, the original anchor is not applicable. It is necessary to find a more suitable anchor box by the clustering algorithm, which can help improve the average accuracy and speed of small-object detection. Considering that the K-means algorithm is sensitive to the initial value setting, and it is easy to converge to the local optimum when the dataset is large, this study uses the K-means + + (Arthur and Vassilvitskii, 2007) algorithm to obtain the initial value before clustering. Therefore, the similarity between bounding boxes can be calculated by a custom distance formula, which is as follows:

$$d\left(box,\ centroid\right)\ =\ 1-IoU\left(box,\ centroid\right) \tag{3}$$

In the abovementioned formula, the centroid is the bounding box selected as centers in clustering; the box is the bounding box labeled in samples; and IoU(box,centroid) represents the merging ratio of sample annotation boxes and cluster center boxes (intersection over union, IoU), that is, the merging of the intersection ratio of the detection result and the ground truth. As shown in Formula (4) and **Figure 5**.

$$IoU\ =\frac{DetectionResult\cap GroundTruth}{DetectionResult\cup GroundTruth} \tag{4}$$

(a) IoU = 0.7 (b) IoU = 0.95

When the IOU value is maximum, the annotation box and anchor box match best, d(box,centroid) is minimum, and the annotation box is assigned to the cluster center. Compared with the K-means clustering algorithm, it uses the strategy that the initial center points are as far away from each other as possible as to measure the average overlap degree of the object cluster, so that the clustering results are not affected by the random selection of the initial cluster center point distance, and the clustered prior box is closer to the object box of the dataset.

Let K = 1, 2,...12. Cluster analysis was performed on the dataset samples, and the relationship between the IoU and K was obtained as shown in **Figure 6**.

It is observed in **Figure 6** that when the number of anchor boxes is 9, the average IoU reaches 94.6%, and there is no important improvement thereafter. To balance the IoU and network complexity, the clustering result of K = 9 is taken as the anchor box size in the network, i.e., (52, 20), (65, 29), (73, 32), (84, 36), (89, 40), (97, 46), (109, 58), (122, 63), and (136, 71).

Multiscale Training

Compared with the YOLO model, the model proposed in this study does not contain a fully connected layer, so it is possible to try different sizes of input images for multiscale training when training convolutional neural networks. By training input images at different scales, the model can achieve the task of adapting to object detection at multiple scales.

Since the improved network contains four residual modules and one dense connection block, the minimum size image is 1/32 of the input image, so when changing the image size, it is necessary to ensure that it is a multiple of 32. Therefore, the image size of the training set is divided into a variety of image scales, which are {320, 352,..., 608}. During training, a scale training is selected randomly 10 times per iteration. The schematic diagram of the multiscale training process is shown in **Figure 7**.

The multiscale training benefits the robustness of the model by being able to accept any size of image as input and therefore improves the performance indicators of the model network.

EXPERIMENT DESIGN

Experimental Platform

The experimental environment configuration is as follows: Intel i7-9750H processor, Nvidia GeForce RTX 2060 graphics card, 16-GB memory, and operating system Ubuntu 16.04. The experiment adopts Python programming language, and the deep learning framework is Keras TensorFlow.

Dataset Construction

Due to the lack of datasets of tomato anomalies in the complex natural environment and because the quality of images is not high, within the context of this study, a tomato anomaly dataset was created by taking photos from complex natural environments and from the Internet as well. The dataset contains 15,000 images of tomato anomalies in various scenarios. Dataset images include occlusion, shading, and other situations. The dataset was converted into VOC2007 dataset format according to the experimental requirements, and the data were annotated with LableImg annotation software.

Model Training

In the original data set, 70%, 20%, and 10% images of each category were selected to form the training set, validation set, and test set, respectively. In the YOLO-Dense model proposed in this study, the training process uses the trained weight file of the original YOLOv3 as the initialization parameter. Because different network structures need to be trained in the comparative experiment, and the number of iterations to achieve the optimal detection performance is also different, this study monitors dynamically during training and saves the weight file of the network every 1,000 iterations in order to select the best weight file to prevent overfitting. In the iteration process, after 100 and 150 rounds, the learning rate decay factor is set to 0.1, that is, the current learning rate is 0.1 times the previous learning rate. Some experimental training parameter settings that improve the optimal network detection effect are shown in **Table 1**. At the same time, a multiscale training strategy is used, i.e., random selection of an input image size every 10 rounds to achieve the effect that the model can adapt to image features of different size.

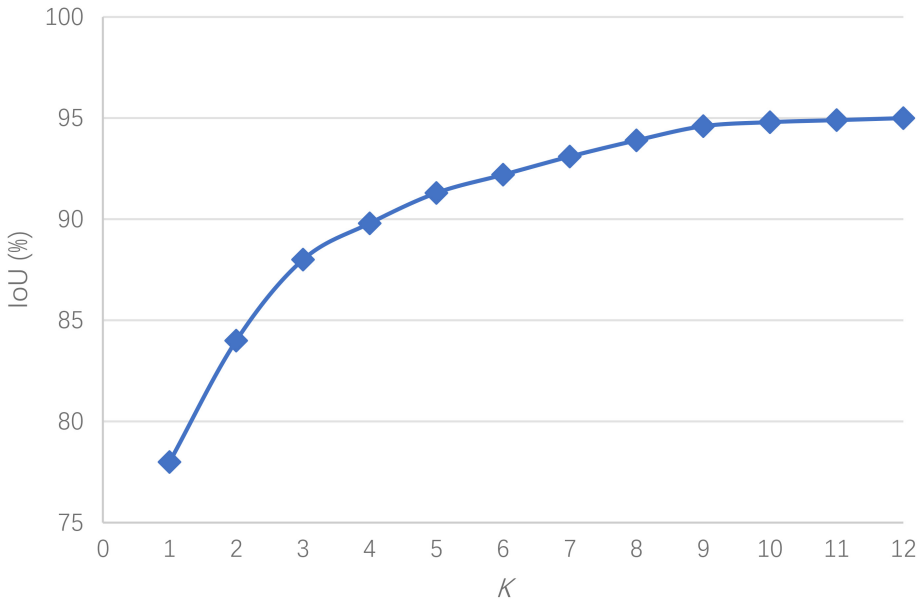

FIGURE 6 | The relationship between the IoU and K.

FIGURE 7 | Schematic diagram of the multiscale training process.

Evaluating Indicator

The detection accuracy of each category in the detection of tomato anomalies is very important. False-positive and false-negative detection may cause the risk of further spread of the disease. Therefore, average precision (AP) and mean average precision (mAP) were selected as the evaluation indexes of object detection algorithm in this study. These two evaluation indicators take into account the accuracy rate (Precision, P) and recall rate (Recall, R):

$$P\left(classes\right) = \frac{TP}{TP + FP} \tag{5}$$

$$R\left(classes\right) = \frac{TP}{TP + FN} \tag{6}$$

TABLE 1 | Selection of key parameters.

Parameter name	Parameter value
Batch size	64
Learning rate	0.0026
Decay	0.0054
Momentum	0.9
Factor	0.1

$$AP = \int_0^1 P(R)dR \tag{7}$$

$$mAP = \frac{\sum_{i=1}^{N} APi}{N} \tag{8}$$

Taking the gray mold category in the detection object of this research as an example, TP in the above formula indicates that the detection model detects the correct gray mold object as the number of gray mold diseases, FP indicates the number of false detection of other categories of object as gray mold diseases, and FN indicates the number of false detection of the correct gray mold object as objects of other categories. The values of recall rate and accuracy rate are taken as abscissa and ordinate, respectively, and a P–R curve is drawn, and the area under the curve is AP. For all categories (the number of categories is denoted as N), the average precision is obtained by calculating AP and taking the mean value. mAP is an important index for evaluating the performance of the model, which can reflect the overall performance of the network model and avoid the problem of extreme performance of some categories and weakening the performance of other categories in the evaluation process.

EXPERIMENTAL RESULTS AND ANALYSIS

Comparison of Algorithm Performance Under Different Resolution Images

The multiscale training method makes the model robust to different resolution images. The corresponding models of this study were trained by changing the resolution of the input image to 320 × 320, 416 × 416, 544 × 544, and 608 × 608, respectively. Then, based on the obtained detection model, the test set was tested separately by adjusting the threshold of the comprehensive score of tomato anomaly detection, and the corresponding Precision–Recall curve of each model was obtained. **Figure 8** shows the Precision–Recall curve of the model proposed in this study at four different image resolutions, and **Table 2** gives the results table of its specific detection evaluation index.

From the above table, we can see that the algorithm performance under different resolution images is satisfying, the mAP value of the model can reach more than 90%, and the detection time of a single image can be controlled within 30 ms. As the resolution of the input image increases, the size of the output feature map, the number of output grids, and the mAP value of the model increase as well; however, the detection speed decreases. When the input resolution is 608 × 608, the detection time of a single image of the model increases to 29.98 ms, affecting the real-time performance of the system. Therefore, it is necessary to select the appropriate resolution for tomato anomaly detection after considering the detection accuracy and detection speed.

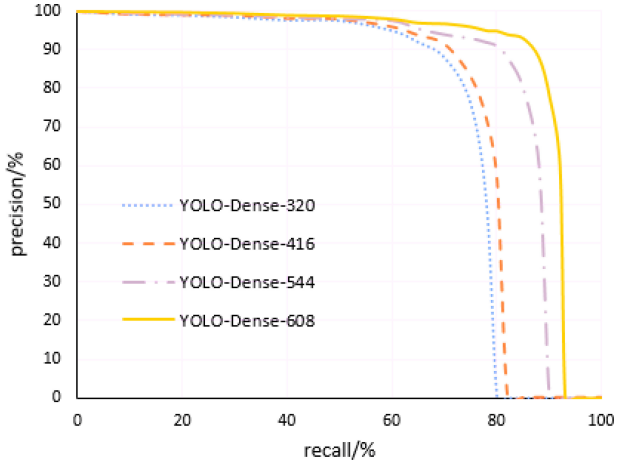

FIGURE 8 | Precision–Recall curves of the proposed model at four different image resolutions.

TABLE 2 | Algorithm performance under different resolutions.

Resolution	mAP (%)	Detection time/(ms)
320 × 320	90.26	17.68
416 × 416	92.32	18.99
544 × 544	96.40	20.27
608 × 608	96.98	29.98

Comparison of Detection Accuracy

The test set is used for testing, and the experimental results are compared with the results of other commonly used object detection algorithms such as the original YOLOv3, Faster R-CNN, and SSD. The results are shown in **Table 3**.

Compared with Faster R-CNN, SSD, and other algorithms, the YOLO-Dense algorithm has higher detection accuracy in terms of accuracy (AP). Compared with the original YOLOv3, the proposed algorithm in this study achieves an improvement in the average detection accuracy of 12 categories of detection objects. The main reason is that the original YOLOv3 network directly divides an image into 7 × 7 grid, and each grid is predicted with 2 bounding boxes, and each grid predicts only one object, which easily leads to inaccurate positioning. In contrast, as for the YOLO-Dense algorithm, it was mentioned before that during training various image sizes could be used in order for the model to be robust and accurate. Moreover, because the original YOLOv3 network only detects one class for each prediction box, and there are different objects of anomalies in this dataset,

TABLE 3 | Comparison of detection accuracy (AP) (%).

Disease/pest	SSD	Faster R-CNN	The original YOLOv3	YOLO-Dense
Early blight	85.46	91.35	90.84	96.71
Late blight	85.67	91.22	90.67	96.68
Yellow leaf curl virus	84.78	91.17	89.98	96.28
Brown spot	85.19	91.08	89.69	93.98
Coal pollution	85.03	91.26	90.52	96.86
Gray mold	84.99	91.41	90.17	96.78
Leaf mold	85.11	90.96	89.65	96.56
Navel rot	85.79	90.89	89.37	96.55
Leaf curl disease	84.98	90.71	89.44	96.48
Mosaic	84.43	90.29	89.32	96.51
Leaf miner	79.12	86.31	85.38	94.95
Greenhouse whitefly	78.01	80.58	79.69	94.01

the detection result on the closely related Leaf miner and the greenhouse whitefly is not ideal.

Faster R-CNN has high accuracy in large-object recognition, but the accuracy of small-object recognition is very low. The reason is that Faster R-CNN introduces the anchor idea to predict 9 anchor boxes with equal ratio of length to width and area ratio in each position of the feature map, which greatly improves the position prediction ability of the model. However, because Faster R-CNN does not cluster the anchor box, the anchor box for the PASCAL VOC dataset is not applicable to the characteristics of this dataset, resulting in low accuracy of small object detection. On the contrary, the large object is closer to the characteristics of the PASCAL VOC dataset, and the large size diseases are clearly distinguishable, so the detection accuracy can reach more than 90%.

In the actual object detection process of tomato anomalies, the original YOLOv3 algorithm did not perform satisfactorily due to the complex scenes and small differences among different disease classes, especially in the detection of occluded objects and small objects. This study attempts to improve on the basis of the YOLOv3 algorithm and proposes a detection algorithm for tomato anomaly detection task in greenhouse scenarios. Applying DenseNet to the detection of tomato anomalies can improve the expression ability of the network model and thus improve the detection accuracy.

Comprehensive Performance Comparison

In the case of the low recognition rate of the original YOLO algorithm for tomato anomaly detection, the proposed newly improved network structure YOLO-Dense obtained its optimal performance at 8000 iterations. A comprehensive performance comparison is shown in **Table 4**.

For the YOLO-Dense network proposed in this study, although the network appears redundant when densely connected, it cannot increase parameters and computation too much, thus not having too much impact on the detection speed. Especially, the concatenation method of dense network connection is adopted, which enables each layer to obtain

TABLE 4 | Comprehensive performance comparison.

Algorithm name	mAP (%)	Time/(ms)	False detection rate (%)	Missed detection rate (%)
SSD	84.32	25.69	1.38	1.29
Faster R-CNN	90.67	2868.94	1.87	1.97
YOLO v3	88.31	21.18	1.05	1.14
YOLO-Dense	**96.41**	**20.28**	**0.61**	**0.96**

Bold values highlight the performance of the proposed algorithm.

gradient and input signal directly from loss function, so as to train a deeper network, further improving the detection accuracy of the network, reduce the detection speed, and improve the overall performance of the network. This study also draws on the idea of anchor, which predicts 3 anchor boxes (i.e., 3×3 anchor boxes per grid) for multiple scales of YOLO-Dense, so that it is equal to the number of anchor boxes of Faster R-CNN and sets different scales for different sizes of objects to detect, which greatly improves the detection accuracy. The output layer uses the Logistic classifier instead of the original Softmax classifier, which improves the final detection accuracy. It is also seen from **Table 4** that the network proposed in this study has the highest average detection accuracy (mAP), the lowest false detection rate, and a missed detection rate, but the detection time is 141 times faster than that of Faster R-CNN. Compared with other algorithms, the speed of Faster R-CNN is the slowest. The biggest difference of Faster R-CNN is that regression and classification are separated. Thus, the detection speed of Faster R-CNN is far behind the other three network frameworks.

By using dense connections, feature fusion and reuse are achieved. By improving the K-means algorithm, object bounding box dimensions are clustered and anchor boxes are calculated for self-made tomato anomaly image datasets; multiscale training is used to enhance the robustness of the model against different sizes. The experimental results show that the YOLO-Dense algorithm improves the detection rate of small objects and occluded objects. Compared with the commonly used algorithms such as SSD, Faster R-CNN, and original YOLOv3, the detection effect is better and the robustness is stronger.

CONCLUSION

This study proposes a tomato anomaly detection method based on the deep-learning YOLO framework. Integrating the dense connection idea of DenseNet into the feature extraction part of the original YOLO network realizes the high fusion and multiplexing of feature information. Meanwhile, the improved K-means clustering algorithm is used for anchor box calculation to improve the matching degree between prior anchors and the feature map, so as to adapt to the detection task of tomato anomalies and improve the detection accuracy. The experiment shows that the optimized model has high detection accuracy and fast speed. The model has strong robustness to the detection of tomato anomalies in the complex natural environment, with an

average accuracy of 96.41%. Also, the detection time of a single image is only 20.28 ms. The experiment verifies that this method can be used for the detection of tomato anomalies in the complex natural environment. Among its potential applications are in handheld devices, edge computing terminals, and other systems.

In conclusion, compared with the other three algorithms, the YOLO-Dense algorithm has certain advantages in performance. The model makes full use of low-level feature information to improve the detection rate of small objects; dense connection reduces the interference of useless features to the model, realizes feature enhancement, improves the detection effect of occluded objects, and improves the model performance. The experimental results prove the effectiveness of the algorithm.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JL and XW conducted the experiments and data analysis, and wrote the manuscript. XW revised the manuscript. Both authors read and approved the manuscript.

FUNDING

This study was supported by the Facility Horticulture Laboratory of Universities in Shandong with project numbers 2019YY003, 2018YY016, 2018YY043, and 2018YY044; School-level High-level Talents Project 2018RC002; Youth Fund Project of Philosophy and Social Sciences of Weifang College of Science and Technology with project numbers 2018WKRQZ008 and 2018WKRQZ008-3; Key Research and Development Plan of Shandong Province with project number 2020RKA07036, 2019RKA07012, and 2019GNC106034; Research and Development Plan of Applied Technology in Shouguang with project number 2018JH12; 2018 innovation fund of the Science and Technology Development Centre of the China Ministry of Education with project number 2018A02013; 2019 basic capacity construction project of private colleges and universities in Shandong Province; Weifang Science and Technology Development Program with project numbers 2019GX071, 2019GX081, and 2019GX082; and special project of ideological and political education of Weifang University of Science and Technology (W19SZ70Z01).

ACKNOWLEDGMENTS

Appreciations are given to the editors and reviewers of the Journal.

REFERENCES

Alemu, K. (2015). Detection of diseases, identification and diversity of viruses: a review. *J. Biol. Agric. Healthc.* 5, 204–213.

Arthur, D., and Vassilvitskii, S. (2007). "K-Means++: the advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007*, (New Orleans, LA: ACM).

Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60. doi: 10.1016/j.biosystemseng.2016.01.017

Chen, W., Huang, H., Peng, S., Zhou, C., and Zhang, C. (2020). Yolo-face: a real-time face detector. *Vis. Comput.* 9. doi: 10.1007/s00371-020-01831-7 [Epub ahead of print].

Fu, L., Feng, Y., Wu, J., Liu, Z., Gao, F., Majeed, Y., et al. (2020). Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precis. Agric.* 1–23. doi: 10.1007/s11119-020-09754-y

Fuentes, A. F., Sook, Y., Jaesu, L., and Sun, P. D. (2018). High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9:1162. doi: 10.3389/fpls.2018.01162

Ghazi, M. M., Yanikoglu, B., and Aptoula, E. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 235, 228–235. doi: 10.1016/j.neucom.2017.01.018

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition CVPR*, (Columbus, OH: IEEE).

He, K., Zhang, X., Ren, S., and Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904–1916. doi: 10.1109/tpami.2015.2389824

Heuvelink, E. (2005). Greenhouse tomato production. *Crop Prod. Sci. Hortic.* 12, 394–399.

Huang, G., Liu, Z., Laurens, V. D. M., and Weinberger, K. Q. (2016). "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Honolulu, HI: IEEE).

Juncheng, M., Keming, D., Feixiang, Z., Lingxian, Z., Zhihong, G., and Zhongfu, S. (2018). A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Comput. Electron. Agric.* 154, 18–24. doi: 10.1016/j.compag.2018.08.048

Li, W., Chen, P., Wang, B., and Xie, C. (2019). Automatic localization and count of agricultural crop pests based on an improved deep learning pipeline. *Sci. Rep.* 9, 1–11.

Li, Y., Han, Z., Xu, H., Liu, L., and Zhang, K. (2019). Yolov3-lite: a lightweight crack detection network for aircraft structure based on depthwise separable convolutions. *Appl. Sci.* 9:3781. doi: 10.3390/app9183781

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. San Juan, PR: IEEE, 2117–2125.

Liu, J. (2018). Tomato yield estimation based on object detection. *J. Adv. Comput. Intell. Intell. Inform.* 22, 1120–1125. doi: 10.20965/jaciii.2018.p1120

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., and Berg, A. C. (2016). "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, (Cham: Springer International Publishing).

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., et al. (2015). Advanced methods of plant disease detection. a review. *Agron. Sustain. Dev.* 35, 1–25.

Pethybridge, S. J., and Nelson, S. C. (2015). Leaf doctor: a new portable application for quantifying plant disease severity. *Plant Dis.* 99, 1504200709 24008.

Prasad, S., Peddoju, S. K., and Ghosh, D. (2014). "Energy efficient mobile vision system for plant leaf disease identification," in *Proceedings of the Wireless Communications & Networking Conference*, (Istanbul: IEEE).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV), 779–788. doi: 10.1109/CVPR.2016.91

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, Honolulu, HI, 6517–6525.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/tpami.2016.2577031

Ren, X., and Liu, X. (2020). Mask wearing detection based on yolov3. *J. Phys. Conf. Ser.* 1678:012089. doi: 10.1088/1742-6596/1678/1/012089

Shahid, N., Kalofolias, V., Bresson, X., Bronstein, M., and Vandergheynst, P. (2015). "Robust principal component analysis on graphs," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, (Santiago: IEEE), 2812–2820.

Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., and Zou, X. (2020). Recognition and localization methods for vision-based fruit picking robots: a review. *Front. Plant Sci.* 11:510. doi: 10.3389/fpls.2020.00510

Xie, J., Yu, J., Chen, B., Feng, Z., and Siddique, K. H. M. (2017). Facility cultivation systems: a chinese model for the planet. *Adv. Agron.* 145, 1–42. doi: 10.1016/bs.agron.2017.05.005

Xing, S., Lee, M., and Lee, K. K. (2019). Citrus pests and diseases recognition model using weakly dense connected convolution network. *Sensors* 19:3195. doi: 10.3390/s19143195

Xu, G., Zhang, F., Shah, S. G., Ye, Y., and Mao, H. (2011). Use of leaf color images to identify nitrogen and potassium deficient tomatoes. *Pattern Recognit. Lett.* 32, 1584–1590. doi: 10.1016/j.patrec.2011.04.020

Xu, Z., Shou, W., Huang, K., Zhou, S., Li, G., Tang, G., et al. (2000). The current situation and trend of tomato cultivation in china. *Acta Physiol. Plant.* 22, 379–382. doi: 10.1007/s11738-000-0061-y

Xu, Z. F., Jia, R. S., Sun, H. M., Liu, Q. M., and Cui, Z. (2020). Light-yolov3: fast method for detecting green mangoes in complex scenes using picking robots. *Appl. Intell.* 50, 1–18.

Zhang, B., Wang, G., Wang, H., Xu, C., Li, Y., and Xu, L. (2021). Detecting small chinese traffic signs via improved yolov3 method. *Math. Probl. Eng.* 2021, 1–10. doi: 10.1155/2021/8826593

Zhang, X., Gao, Y., Wang, H., and Wang, Q. (2020). Improve yolov3 using dilated spatial pyramid module for multi-scale object detection. *Int. J. Adv. Robot. Syst.* 17:172988142093606. doi: 10.1177/1729881420936062

Zhang, Y., Song, C., and Zhang, D. (2020). Deep learning-based object detection improvement for tomato disease. *IEEE Access* 8:56607–56614. doi: 10.1109/access.2020.2982456

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 April 2021
doi: 10.3389/fpls.2021.621168

Biotic Yield Losses in the Southern Amazon, Brazil: Making Use of Smartphone-Assisted Plant Disease Diagnosis Data

Anna C. Hampf[1,2], Claas Nendel[1,3], Simone Strey[4] and Robert Strey[4]*

[1] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany, [2] Albrecht Daniel Thaer Institute of Agricultural and Horticultural Sciences, Humboldt-Universität zu Berlin, Berlin, Germany, [3] Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany, [4] Progressive Environmental and Agricultural Technologies (PEAT) GmbH, Hannover, Germany

Pathogens and animal pests (P&A) are a major threat to global food security as they directly affect the quantity and quality of food. The Southern Amazon, Brazil's largest domestic region for soybean, maize and cotton production, is particularly vulnerable to the outbreak of P&A due to its (sub)tropical climate and intensive farming systems. However, little is known about the spatial distribution of P&A and the related yield losses. Machine learning approaches for the automated recognition of plant diseases can help to overcome this research gap. The main objectives of this study are to (1) evaluate the performance of Convolutional Neural Networks (ConvNets) in classifying P&A, (2) map the spatial distribution of P&A in the Southern Amazon, and (3) quantify perceived yield and economic losses for the main soybean and maize P&A. The objectives were addressed by making use of data collected with the smartphone application *Plantix*. The core of the app's functioning is the automated recognition of plant diseases via ConvNets. Data on expected yield losses were gathered through a short survey included in an "expert" version of the application, which was distributed among agronomists. Between 2016 and 2020, *Plantix* users collected approximately 78,000 georeferenced P&A images in the Southern Amazon. The study results indicate a high performance of the trained ConvNets in classifying 420 different crop-disease combinations. Spatial distribution maps and expert-based yield loss estimates indicate that maize rust, bacterial stalk rot and the fall armyworm are among the most severe maize P&A, whereas soybean is mainly affected by P&A like anthracnose, downy mildew, frogeye leaf spot, stink bugs and brown spot. Perceived soybean and maize yield losses amount to 12 and 16%, respectively, resulting in annual yield losses of approximately 3.75 million tonnes for each crop and economic losses of US$2 billion for both crops together. The high level of accuracy of the trained ConvNets, when paired with widespread use from following a citizen-science approach, results in a data source that will shed new light on yield loss estimates, e.g., for the analysis of yield gaps and the development of measures to minimise them.

Keywords: plant pathology, animal pests, pathogens, machine learning, digital image processing, disease diagnosis, crowdsourcing, crop losses

OPEN ACCESS

Edited by:
Reza Ehsani, University of California, Merced, United States

Reviewed by:
Shaun M. Sharpe, Agriculture and Agri-Food Canada, Canada
Simone Ugo Maria Bregaglio, Council for Agricultural and Economics Research (CREA), Italy

****Correspondence:***
Anna C. Hampf
anna.hampf@zalf.de

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *25 October 2020*
Accepted: *12 March 2021*
Published: *15 April 2021*

Citation:
Hampf AC, Nendel C, Strey S and Strey R (2021) Biotic Yield Losses in the Southern Amazon, Brazil: Making Use of Smartphone-Assisted Plant Disease Diagnosis Data. Front. Plant Sci. 12:621168. doi: 10.3389/fpls.2021.621168

HIGHLIGHTS

- ConvNets were trained to identify 420 crop disease classes under diverse conditions.
- Crowdsourcing can significantly improve the data basis for algorithm training.
- Expected yield losses to pests and diseases in the Southern Amazon are below global estimates.
- Annual soybean and maize yield losses to pests and diseases each amount to 3.75 million tonnes.
- Citizen science data can help to identify yield gaps and advance the field of crop loss research.

INTRODUCTION

Pathogens and animal pests (P&A) are major challenges to global food security, directly affecting the quantity (reduced productivity) and quality (e.g., reduced content of valuable nutrients, poorer market quality, and inferior storage characteristics) of food (Oerke, 2006)[1]. They can cause devastating yield losses, leading to malnutrition and starvation, as several examples in history have shown (e.g., the Irish Potato Famine (1845–49), caused by potato leaf blight; and witches' broom disease, *Moniliophthora perniciosa*, which destroyed Brazil's leading position in world cocoa production). Globally, direct yield losses to P&A were estimated to range between 20 and 30% for major food and cash crops (Oerke and Dehne, 2004; Oerke, 2006; Savary et al., 2019). Besides these direct effects on food provision, P&A also have indirect effects on the environment (e.g., pesticide use, soil contamination), public health (e.g., mycotoxin contamination) and the economic performance of rural communities (Savary et al., 2012).

The Southern Amazon (specifically, the states of Mato Grosso and Pará) is Brazil's largest domestic producing region of cotton (64% of national output), maize (34%) and soybeans (28%) (CONAB, 2019). High annual rainfalls and relatively long wet seasons with reliable onset dates allow for the cultivation of two crops in one season (Arvor et al., 2014). Early maturing soybean cultivars are grown at the onset of the rainy season and are either followed by maize or cotton. The high production intensity as well as the warm and humid climate, however, make the region susceptible to the outbreak and spread of P&A. Soybean, maize and cotton production are expected to decrease by 30–40% if farmers do not make use of pesticides to control major P&A (CEPEA, 2019). One of the largest threats to crop production in the Southern Amazon is the fungus *Phakopsora pachyrhizi*, commonly known as Asian soybean rust, causing yield losses of up to 90% (Godoy et al., 2016). Since its first occurrence in Brazil in the early 2000s, the fungus has caused annual yield losses in the range of 360,000–4.6 million tonnes, and economic losses (grain loss +pest control costs) of approximately US$0.18–2.38 billion per year (Godoy et al., 2016).

However, although P&A can cause immense crop damage and economic losses, there are very few systematic research and monitoring programmes on the impact of P&A on crop performance and their spatial distribution. Yield loss data is often based on a limited number of site-specific tests or a particular pathogen over one season. As a result, there has been a persistent and chronic lack of knowledge on the frequency and extent of crop losses caused by plant diseases (Esker, 2012; Nelson, 2017). Moreover, biotic yield losses are largely ignored in yield gap analysis. Yield gaps are an essential concept in crop loss research, defined as the difference between potential yields and actual yields (van Ittersum et al., 2013). While yield losses due to nutrient- and water deficiency were extensively explored using crop modelling, such studies for P&A or weeds are still missing. One major challenge to quantifying P&A-related yield losses is the extremely large diversity of plant diseases, the diversity of life cycles of these organisms and the enormous number of interactions that may exist between P&A and their host crops (Donatelli et al., 2017; Savary et al., 2018).

Various methodological approaches have been used to identify P&A and to quantify associated yield losses, including field experiments (Savary et al., 2016), expert surveys (Savary et al., 2019), simulation modelling (Bregaglio and Donatelli, 2015; Donatelli et al., 2017), remote sensing (Mahlein, 2016), image recognition techniques (Barbedo, 2013; Barbedo et al., 2016), and deep learning models (Boulent et al., 2019). Deep learning model and in particular convolutional neural networks (ConvNets) have recently achieved impressive identification performances in various visual classification tasks, such as the automatic identification of plants and animals (LeCun et al., 2015; ImageCLEF, 2018). Due to their capacity to generalise, they can overcome many of the challenges (e.g., diseases with similar symptoms, multiple simultaneous disorders in a single plant) faced by traditional classification methods (e.g., thresholding, fuzzy classifier, feature-based rules), which appear to be either too specific (identifying just a small number of pathogens) or too sensitive (functioning only under strict operation conditions) (Barbedo, 2013; Boulent et al., 2019).

Several studies demonstrated that ConvNets can be trained to identify a large number of different plant-disease combinations with an accuracy of 85–99% (Mohanty et al., 2016; Ferentinos, 2018; Boulent et al., 2019). The accuracy of these models, however, drastically fell to 25–30% when they were tested on images taken under conditions other than the training dataset (Mohanty et al., 2016; Ferentinos, 2018). The acquisition of a large, verified database with P&A images from different geographic locations as well as the maximisation of real-condition images in the training dataset are two of the major challenges to further improving ConvNets' performance (Mohanty et al., 2016; Barbedo, 2018b; Ferentinos, 2018). Integrated into mobile devices such as smartphones, ConvNets can be turned into valuable decision support tools for farmers, allowing for plant disease diagnosis on a massive—indeed global—scale (Hughes and Salathe, 2015; Mohanty et al., 2016; Ferentinos, 2018). For instance, Picon et al. (2019) implemented their trained model into various mobile devices and obtained balanced accuracies of 86 and 98% for two different wheat

[1]The term *crop losses* refers to both quantitative and qualitative losses, whereas the term *yield losses* covers quantitative losses only. The focus of this study is on yield losses.

diseases. In another study, Ramcharan et al. (2019) deployed a ConvNet in a mobile app to identify three different cassava diseases in Tanzania but reported a 32% drop in the classification performance when shifting from the test dataset to real-world images.

To sum up, the most important research gaps are a lack of data on the spatial distribution of plant diseases and associated yield losses in the Southern Amazon, a lack of a large verified database for the training and further improvements of ConvNets and a lack of implementation of deep learning technologies for the automated recognition of plant diseases into a practical tool for farmers and/or extension workers. The study seeks to address these research gaps by targeting the following objectives:

(i) Evaluating the performance of ConvNets in classifying P&A
(ii) Mapping the spatial distribution of P&A in the Southern Amazon, Brazil
(iii) Quantifying perceived yield losses for main soybean and maize diseases
(iv) Discussing the potential benefits and limitations of an automated plant disease classification and possible implications for the field of crop loss research.

The objectives were addressed in a joint effort by the Leibniz Centre for Agricultural Landscape Research (ZALF) and PEAT GmbH (Progressive Environmental and Agricultural Technologies). In 2016, PEAT launched *Plantix*, a mobile decision support application for farmers, extension workers and gardeners that uses image recognition and deep learning to diagnose P&A. As part of this study, the *Plantix* library was expanded to include P&A common to the (sub)tropical environment of the Southern Amazon and a 3 months field test was carried out in 2016 to capture field-condition images and to test and promote the app *in situ*. Since then, *Plantix* users have captured more than a million images of P&A in Brazil.

MATERIALS AND METHODS

Study Area

The states of Mato Grosso (MT) and Pará (PA)—located in the Southern Amazon of Brazil (**Figure 1A**)—are dominated by highly industrialised agricultural systems, mainly consisting of soybean-maize and soybean-cotton rotations. In 2019, farmers in MT produced approximately 32, 31, and 4.5 million tonnes of soybean, maize and cotton (seeds and lint) on 9.6, 4.9, and 1 million ha (Mha) of cropland, respectively (**Figure 1B**; CONAB, 2020). The climate in the study area is sub(tropical), with pronounced dry and wet seasons and annual precipitation rates ranging from approximately1,000 mm in South MT to over 3,000 mm in Northern PA (INMET, 2019). However, the warm and humid climate, as well as changes in the production system (e.g., expansion of the agricultural frontier northward, extended sowing periods, lack of rotation) have led to a high incidence and spread of P&A in the study area (Godoy et al., 2016; Fundação, 2019). The main diseases affecting agricultural production in the study area are Asian soybean rust (in soybean); common and tropical rust (in maize); and anthracnose and *Ramularia* blight (in cotton) (ABRAPA, 2011; Fundação, 2019). The most damaging pests are those that feed on multiple crops (polyphagous pests) and disperse across fields and over extended periods, such as the lesser cornstalk borer (*Elasmopalpuslignosellus*), the cotton bollworm (*Helicoverpa zea*, also known ascorn earworm), and the fall armyworm (*Spodoptera frugiperda*) (Fundação, 2019). A detailed overview of the main P&A affecting agricultural production in the study area is given in **Supplementary Tables 1,2**. To reduce the impact of P&A on crop production, genetically modified (GM) crops have been increasingly grown in Brazil. In the 2017–18 cropping season, insect-resistant seeds, herbicide-tolerant seeds or a combination of both were planted on 97, 91, and 84% of soybean, maize (second season) and cotton fields, respectively (Céleres, 2018). Likewise, the use of pesticides in Brazil increased from approximately50 million tonnes in 1990 to 378 million tonnes in 2017 (FAOSTAT, 2019), with MT reporting the largest amount of pesticide use (Pignati et al., 2017). The costs of pesticides (fungicides, herbicides and insecticides) have been estimated to account for 16, 9, and 27% of the total production costs for soybean, maize and cotton, respectively (CEPEA, 2019).

The *Plantix* Application and Its Workflow

Georeferenced images of P&A were collected by users of the *Plantix* smartphone application. The application was released by PEAT in 2016 and is freely available in different languages for any smartphone using the Android operating system. The core function of the app is the automated classification of P&A using ConvNets and involves four steps: (1) taking a picture of the infected plant; (2) classifying the image using several ConvNets; (3) confirming or rejecting the diagnosis by the user; and (4) receiving further information on causes, preventive measures and control options (**Figure 2**). When taking a picture of a diseased plant in the field, the user can upload the image either directly to a remote server, or the image can be stored on the smartphone and uploaded as soon as a functioning internet connection is available. This enhances the app's usability in rural and remote areas with low mobile internet connectivity. Once uploaded, the image is classified using multiple ConvNets (one network to determine if the image contains a relevant crop or no plant at all (e.g., an object); one network to classify the crop type; and one to classify the disease). Then, the most similar crop disease combinations (further referred to as "classes") are displayed to the user and ranked according to their softmax probabilities (see section "Convolutional Neural Networks and Softmax Probability"). Based on this probability ranking as well as a symptom description and reference images for comparison, the user can either confirm or reject the diagnosis. Once the diagnosis is confirmed, the user receives further information on causes, preventive measures, and biological or chemical treatment options. The app can also be used as an offline library, which currently (as of June 2020) contains a description of 592 P&D (267 fungal diseases, 191 insects, 51 bacteria, 51 viruses, 21 mites, and 11 deficiencies). Currently, the ConvNets can automatically detect 231 plant diseases and deficiencies on 49 different species, resulting in a total of 420 classes. Although

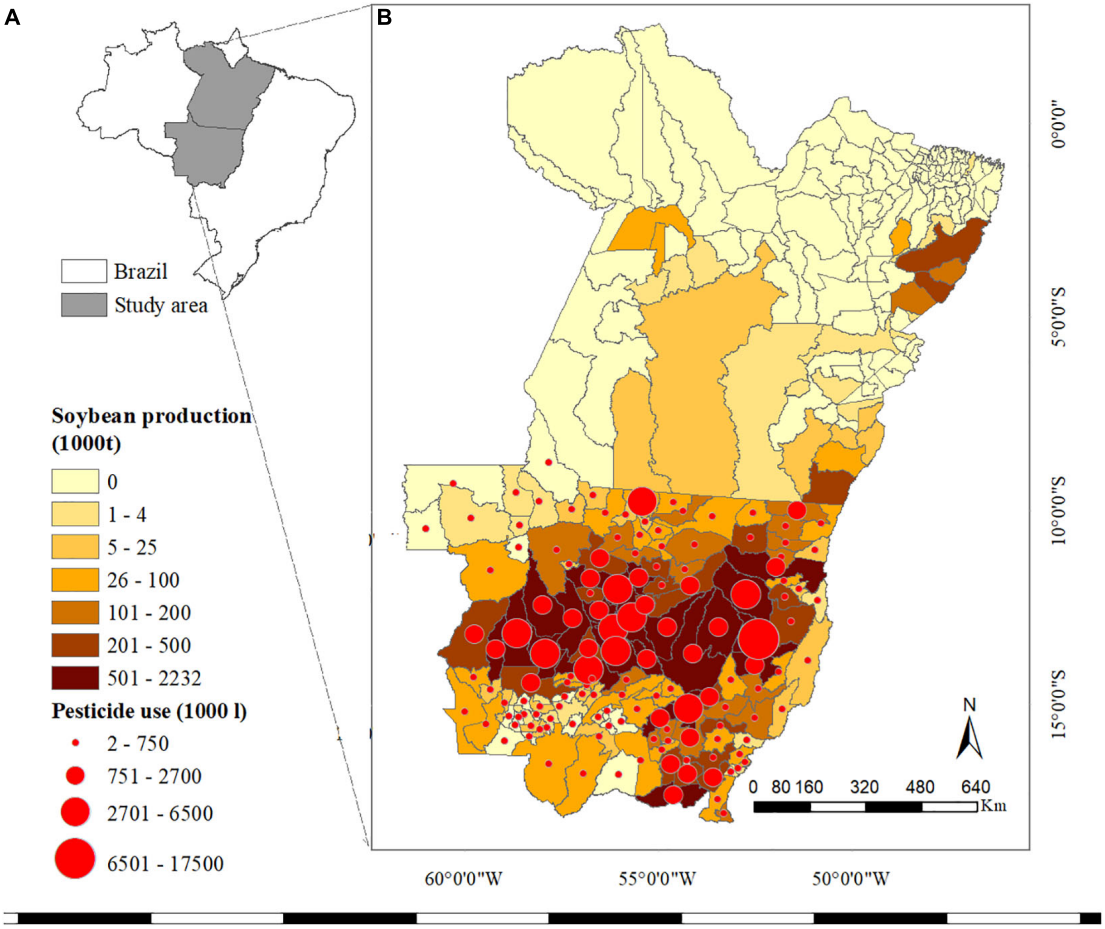

FIGURE 1 | (A) Study of area within Brazil. **(B)** Soybean production by municipality in Mato Grosso and Pará in thousands of tonnes in 2018, as well as pesticide use in litres in MT. IBGE (2019b), INDEA (2020). Maps created using ArcMap 10.6.1.

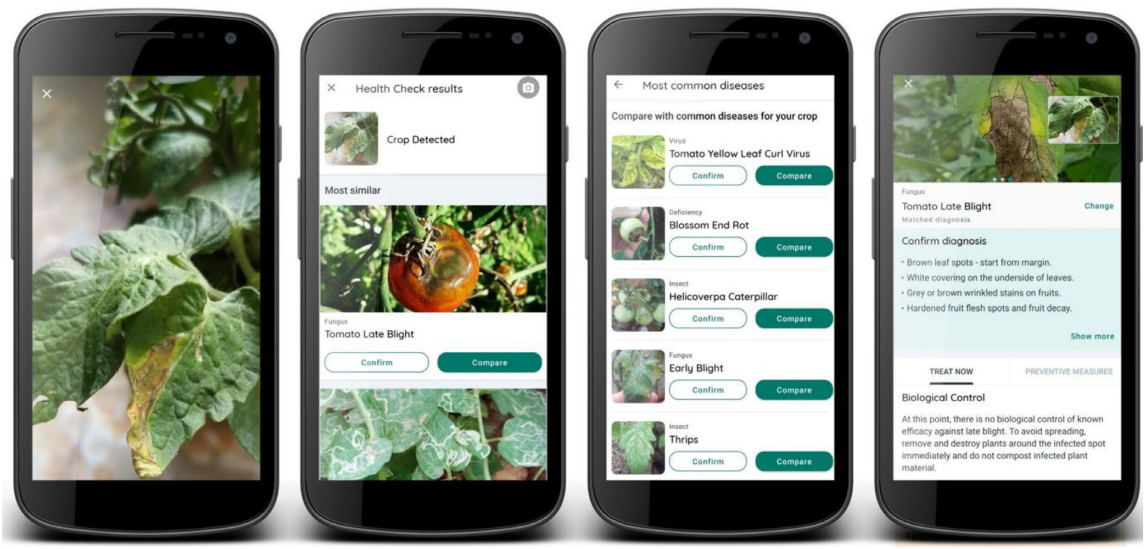

FIGURE 2 | Function flow of the smartphone application Plantix: (1) The user takes a picture of the diseased plant organ. (2) ConvNets classify the image. (3) The user confirms or rejects the diagnosis. (4) The app displays additional information on symptoms, preventive measures and P&A control mechanisms.

deficiencies can also be detected, the focus of the application and this study is on P&A.

Convolutional Neural Networks and Softmax Probability

Image classification in *Plantix* is done via ConvNets, a type of deep neural network, which processes data that comes in the form of an array—for example, a colour image composed of three 2D arrays containing pixel intensities in the three colour channels (LeCun et al., 2015). The architecture of a typical ConvNet mainly consists of convolutional layers and pooling layers. The role of the convolutional layers is to detect local conjunctions of features from the previous layer, whereas the role of the pooling layers is to merge semantically similar features into a single one (LeCun et al., 2015). Multiple ConvNets, such as GoogleNet, AlexNet, and ResNet, have been trained and tested for the classification of plant diseases (Boulent et al., 2019). Other networks, such as EfficientNet (source code/weights) were especially designed for the use on mobile devices. EfficientNet achieves state-of-the-art accuracy with fewer parameters and fewer number of floating-point operations (FLOPs) than other current ConvNets (Tan and Le, 2019; Tan et al., 2019). Due to their high accuracy, this study trained and tested ConvNets using the EfficientNet architecture on a large crowdsourced image database held by PEAT.

To facilitate the interpretation of the network's output, the convolutional and pooling layers are followed by a fully connected layer. The logits contained in this last layer are converted into probabilities using an activation function, most commonly softmax. Softmax normalises the input array into a scale between 0 and 1, with the sum of the softmax output resulting in 1 (Sharma et al., 2020). In multi-class classification, which is the case of plant disease recognition in *Plantix*, the output of the softmax activation function is given as a vector with probabilities for each class, e.g., [0.2, 0.6, 0.1, . . .]. The class with the highest probability among all the distributed probabilities is the top-1 prediction. *Plantix* displays the top-1 prediction to the user as the most likely disease, but other predicted classes with lower probabilities (top-2, top-3) can be shown on lower-ranking positions, thus serving as a decision support tool. Although the prediction probability from a softmax distribution has a poor direct correspondence to confidence, correctly classified examples tend to have a greater maximum softmax probability than erroneously classified or out-of-distribution examples (Hendrycks and Gimpel, 2017). Therefore, in this study, only images with a top-1 softmax probability above 0.5 for both the predicted crop and disease were retained in the final dataset.

Model Training and Testing

The ConvNets implemented in *Plantix* were trained and tested on a large crowdsourced image database, collected either directly by *Plantix* users or by agronomists. While agronomists helped to gather images of less frequent diseases, a high share of images directly collected by *Plantix* users increases the diversity in the image dataset in terms of e.g., image quality, geographic location and smartphone devices. The larger the diversity in the image dataset and the better it reflects the reality of the operational environment, the greater the robustness of the trained model (Barbedo, 2018b; Boulent et al., 2019). Images used for model training and testing were not only collected in Brazil (see section "Fieldwork in Mato Grosso") but also in other world regions (e.g., Germany, India). All images were either directly annotated by agronomists or annotated and validated afterward by plant experts. The final image dataset was split into a training (2/3) and testing (1/3) subset. Using transfer learning, one ConvNet was trained to classify species (crop ConvNet) and another one to classify P&D and deficiencies (disease ConvNet). The ConvNets are trained with a cosine annealing learning rate over 20 epochs. The total time for training on a machine using two Nvidia GeForce RTX 2080 Ti GPUs is 28 h.

The performance of the ConvNets in identifying plant diseases was assessed by comparing the predicted label (ConvNet classification) to the actual label (expert classification) for each element of the test dataset and calculating the three following evaluation metrics: (1) precision, (2) recall, and (3) F_1 score. Also, the proportion of images where the correct class was among the top-3 predicted classes was calculated. "Precision" designates the number of images correctly labelled as belonging to the positive class (true positives) divided by the total number of images labelled as belonging to the positive class (sum of true positives and false positives). "Recall" is defined as the number of true positives divided by the total number of images that actually belong to the positive class (sum of true positives and false negatives). The F_1 score is the harmonic mean of the "precision" and the "recall" figures (Powers, 2011). All metrics are based on the binary confusion matrix (**Table 1**). **Table 2** provides an overview of how to calculate each metric.

Fieldwork in Mato Grosso

Three months of fieldwork (from September to December 2016) was carried out in MT to collect training images and to promote the app among farmers and their organisations. Before the fieldwork, an online survey was conducted among local agronomists and plant experts to identify the main soybean, maize and cotton plant diseases common to the Southern Amazon. The online survey asked agronomists to rank a literature-based pre-selection of P&A according to their importance and/or to name additional P&A. The survey was

TABLE 1 | Confusion matrix for a classification task.

		Prediction (ConvNet classification)	
		Positive	**Negative**
Actual (expert classification)	Positive	True positive (Tp)	False positive (Fp)
	Negative	False negative (Fn)	True negative (Tn)

TABLE 2 | Formula of evaluation metrics used to assess ConvNet performance.

N°	*Evaluation indices*	*Formula*
1	Precision	$\frac{Tp}{Tp+Tp}$
2	Recall (sensitivity)	$\frac{Tp}{Tp+Fn}$
3	F_1 score	$2 * \frac{precision * recall}{precision + recall}$

sent to agronomists of the Brazilian Agricultural Research Corporation (Embrapa), as well as to universities and research institutes specialising in agronomy. The P&A they identified were incorporated into the *Plantix* library with a description of their symptoms as well as preventive measures and control options. Besides, the library and menu of *Plantix* were translated into Portuguese.

During fieldwork, more than 50 farms in Southeast and Central Mato Grosso were visited to test *Plantix* in the field and collect images that could be used for model training. One crucial element of the fieldwork was to advertise the app among farmers, research institutes, students and the general public to ensure a large engagement in the crowdsourcing project. Advertising materials were distributed at universities and research institutes and sent to public and private farmers' organisations, such as Aprosoja and the Mato Grosso Research, Assistance and Rural Extension Company (EMPAER), as well as to more than 200 local offices of the Rural Workers' Union (FETAGRI) and the Rural Union (FAMATO). One example of these advertising materials is shown in **Supplementary Figure 1**.

Survey on Perceived Yield Losses and Yield Loss Estimates at the State Level

Besides the above mentioned online survey, a second survey was conducted during the cropping season 2016–17 to gather information on perceived yield losses to P&A. This survey was directly included in an "expert" version of *Plantix*, which was distributed exclusively among agronomists and other plant experts. When taking a picture of a diseased plant, the agronomists were asked to roughly estimate expected (future) yield losses. Six different answer ranges were possible: 0–5, 5–10, 10–20, 20–50, 50–70, and more than 70%. The reason to limit its distribution to plant experts was to ensure the highest possible data quality. However, this also limited the spatial coverage of the survey and most yield losses estimates were provided for Central MT, causing a potential location bias.

To get an approximation of possible yield and economic losses at the municipality and state level, the study assumed that yield loss estimates provided for Central MT would be representative of other production sites in the Southern Amazon. Hence, the expert-based average yield loss estimates of each disease were merged with the kernel density map of the respective disease (see sections "Description of Cleaned-Up Dataset", "Spatial Distribution" and "Expected Yield Losses"), resulting in spatial yield loss maps of the most important soybean and maize P&A. The mean of these spatial yield loss maps was taken for both soybean and maize and the expected percentage yield losses per municipality were estimated. Next, data on absolute crop production between 2016 and 2018 at the municipality level (IBGE, 2019b) and expected percentage yield losses per municipality were used to calculate absolute yield losses at the municipality and state level. Finally, economic yield losses were estimated assuming average prices of US $355 and $159 per metric tonne for soybean and maize, respectively, for the 2016–2018 cropping seasons in accordance with data from the International Monetary Fund (IMF, 2020).

Kernel Density Estimation

To visualize the spatial distribution of the predicted P&A, kernel density maps were generated using the *tmaptools* (Tennekes, 2019) package of the open-source software program R (R Core Team, 2020). Kernel density estimation produces a risk map that is interpolated from incident locations in a defined study area. It generalizes or "smooths" discrete data points in a way that a continuous surface area is produced (Hart and Zandbergen, 2014). Here, a 2D kernel density estimator was applied with a bandwidth set to 1/50th of the shorter side of the study area and the resolution of the output raster was set to 1 km^2. Kernel densities below 0.0001 were set to NA. The output raster were plotted using the R packages *raster* (Hijmans, 2019), *rasterVis* (Lamigueiro and Hijmans, 2019), and *RColorBrewer* (Neuwirth, 2014).

RESULTS

Evaluation of the ConvNets' Performance

Table 3 summarises the evaluation metrics for the crop and disease ConvNet trained on the *Plantix* image dataset. The evaluation metrics indicate a high performance of the crop and disease ConvNets in identifying 420 classes with a precision of 91.11%, recall of 90.61%, an F_1 score of 90.86%, and top-3 accuracy of 98.81% (weighted summary; see **Table 3**). **Table 3** also gives the weighted mean of the metrics for 18 maize diseases and 19 soybean diseases, indicating a lower precision for soybean disease detection than for maize.

Plantix Dataset and Data Cleaning

Between November 2016 and May 2020, *Plantix* users captured about 1.05 million images in Brazil, of which approximately 980,000 showed a plant, whereas the remaining images contain objects. All images containing objects were discarded from the dataset. Most of the images were taken in South Brazil, in the states of Sao Paulo (∼190,000), Santa Catarina (∼172,000) and Minas Gerais (∼2,000). *Plantix* users in the Southern Amazon captured 77,611 P&A images, of which 80% came from MT. From this dataset, all images showing ornamental plants were removed, further reducing the dataset to 70,266 images. Since users often took multiple images at the same location and at the same time, only one image per disease per camera session was allowed. This reduced the dataset to 44,926 images. Finally, the images

TABLE 3 | Summary of evaluation metrics for the crop and disease ConvNet trained using the *Plantix* image dataset.

Crop	Precision	Recall	F_1 score	Top3	N° diseases
Weighted mean	90.86	91.11	90.61	98.81	420
Mean	88.08	89.63	86.58	97.98	420
Maize	88.40	91.51	89.82	98.99	18
Soybean	78.72	76.64	76.86	96.82	19

For maize and soybean diseases, weighted metrics are shown.

were filtered according to their softmax probability (see section "Convolutional Neural Networks and Softmax Probability"). All images with a top-1 softmax probability below 0.5 for either the predicted crop or disease type were removed from the dataset, reducing it to 15,921 images. This corresponds to about 20% of the original dataset. The results presented in this study are based on this cleaned-up dataset.

Description of Cleaned-Up Dataset

The cleaned-up dataset contains 15,921 images of P&A and deficiencies that were taken by *Plantix* users in the Southern Amazon between 2016 and 2020. According to the predictions provided by the ConvNets, the dataset holds images of 395 different classes; some, however, are only represented by a few images. A complete list of all classes for which more than 50 pictures were taken can be found in the **Supplementary Table 3**. The bulk of the images were collected in the main production areas of MT, specifically the central north (Sinop, Sorriso, Lucas do Rio Verde), the southwest (Campo Novo do Parecis, Tangará da Serrá) and the southeast (Primavera do Leste and Campo Verde). Multiple images were also collected along Highway BR-163, which connects Cuiabá (MT) with Santarém (PA) and serves as a soybean export corridor, as well as along the Trans-Amazonian Highway BR-230. **Figure 3** shows the spatial distribution of images collected by *Plantix* users in MT and PA between November 2016 and May 2020, as well as the main land-use types in the study area.

Spatial Distribution

Spatial Distribution of Predicted Crop Types

Plantix predicted most of the images as showing diseased maize plants (1,973), followed by citrus (1,921), including orange, lemon and tangerine, soybean (1,583), pepper (1,459), tomato (1,390), mango (954), banana (824), cotton (453), rice (442), onion (384), eggplant (369), cucumber (356), papaya (347), and lettuce (313). **Figure 4** shows the kernel density of images collected by *Plantix* users in MT and PA between November 2016 and May 2020 according to predicted crop types.

Spatial Distribution of Predicted Pest Types

The ConvNet that processed images according to disease type predicted most images as showing insects (4,692) or fungal diseases (4,402). Fewer pictures were predicted to show bacteria (899), viruses (577) and mites (366). There were also many images labelled as deficiencies (1,534), e.g., nitrogen, magnesium or iron deficiency. The pathogen class of "others" (219) groups abiotic damage, such as pesticide burn, herbicide damage or sunburn. The "disease" ConvNet also predicted numerous images as containing healthy plants (3,461). **Figure 5** shows the kernel density of images collected by *Plantix* users in MT and PA between November 2016 and May 2020, according to predicted pathogen types.

Spatial Distribution of Predicted Soybean Pathogens and Animal Pests

The "disease" ConvNet interpreted 1,454 out of 1,583 images as showing diseased soybean plants, and the remaining images as healthy soybean plants. These images were classified as stink bugs on soybean (227), brown spot of soybean (159), tobacco caterpillar (138), potassium deficiency (125), anthracnose of soybean (121), downy mildew of soybean (106), frogeye leaf spot (105), Asian soybean rust (94), target spot of soybean (92), and sudden death syndrome (86). Fewer images were predicted to show the fall armyworm (41), the *helicoverpa* caterpillar (35), soybean looper (32), stem rot (25), leaf miner flies (19), boron deficiency (16), spider mites (15), powdery mildew of soybean (9), and castor semi-looper (8). Of the 10 most frequently predicted soybean P&A, all except tobacco caterpillar are mentioned by the Mato Grosso Foundation among the most common soybean diseases found in the Southern Amazon (**Supplementary Tables 1**, **2**). Most images were collected between mid-October and mid-February, which corresponds to the main soybean cropping season, whereas images predicted as containing maize plants were mainly collected either during sowing in March or before harvest in July. The timing of the image data collection is shown in more detail in **Supplementary Figure 2**. **Figure 6** shows the kernel density of the main soybean pests and diseases based on images collected by *Plantix* users in MT and PA between November 2016 and May 2020.

Spatial Distribution of Predicted Maize Pathogens and Animal Pests

Between 2016 and 2020, *Plantix* users captured 1,973 images of maize plants in MT and PA, Brazil, of which 1,854 were interpreted by the "disease" ConvNet to show diseased maize plants and 109 healthy maize plants. According to the system's predictions, most images were likely to show the fall armyworm (466), bacterial stack rot (233), maize rust (233), grey leaf spot of maize (172), magnesium deficiency (164), northern leaf blight (116), phosphorus deficiency (107), boron deficiency (90), and aphids (75). Fewer images were interpreted to show potassium deficiency (64), nitrogen deficiency (45), stemborer damage (36), fusarium ear rot (23), maize smut (16), and goss wilt (14). Of the 10 most frequently predicted maize P&A, all except northern leaf blight, aphids, maize smut and goss wilt are mentioned by the Mato Grosso Foundation among the most common maize diseases found in the Southern Amazon (**Supplementary Tables 1,2**). **Figure 7** shows the kernel density of the main P&A affecting maize production in the Southern Amazon.

Expected Yield Losses

Expected Yield Losses According to Pathogens and Animal Pests

Agronomists reported expected yield losses for 2,419 images. For soybean and maize, respectively, yield losses were reported for 19 and 13 different classes, including two and five deficiencies, based on 409 and 250 corresponding images. The survey reveals that expected soybean and maize yield losses due to P&A were on average 12.16 and 16%, respectively. However, there were large differences in expected yield losses according to different P&A. For soybean, expected yield losses were highest for the sudden death syndrome (23%), followed by castor semi-looper (21.67%), and fall armyworm (16.6%). Expected maize yield losses were highest for maize rust (22.11%), bacterial stalk rot (20.27%),

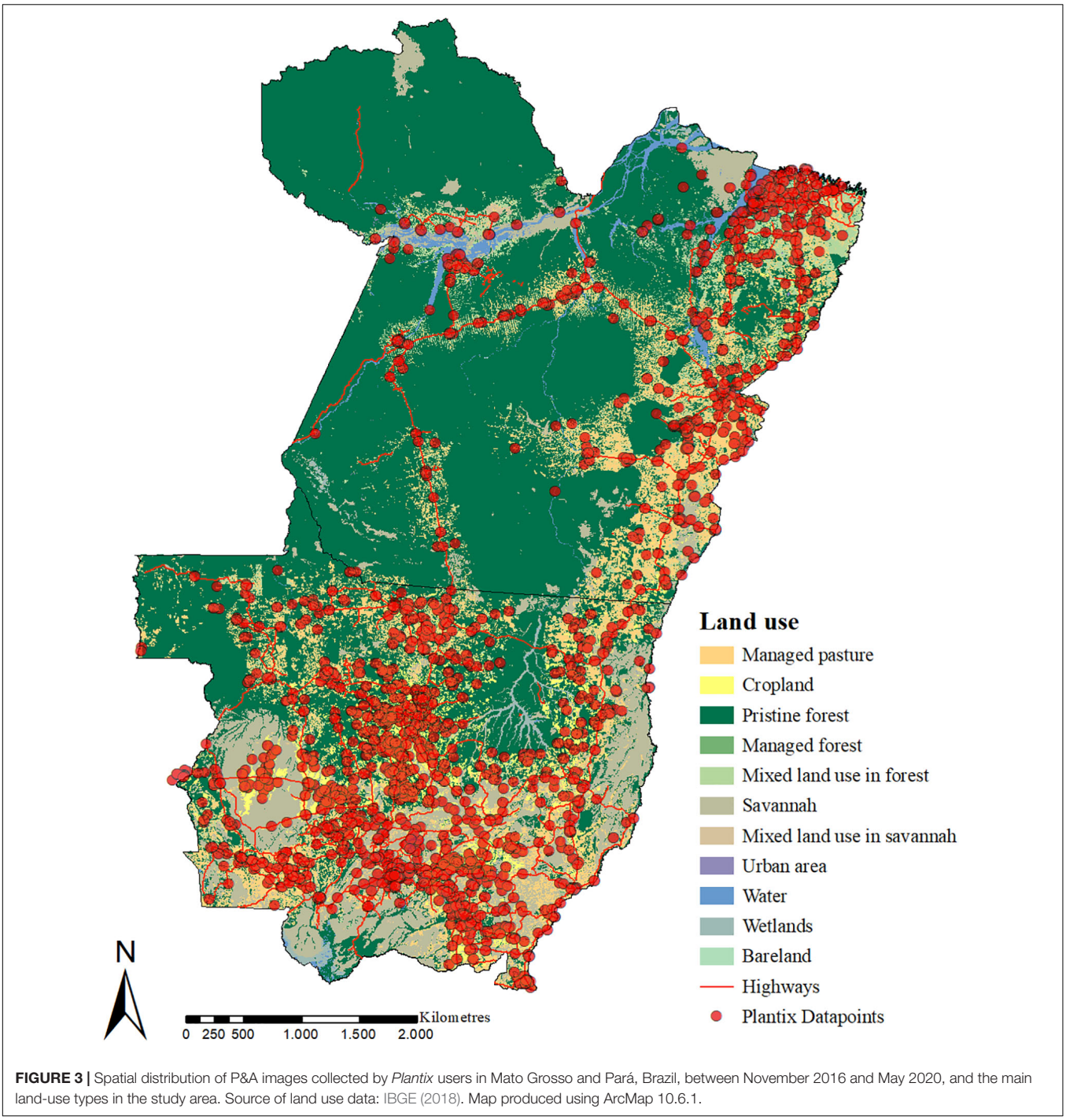

FIGURE 3 | Spatial distribution of P&A images collected by *Plantix* users in Mato Grosso and Pará, Brazil, between November 2016 and May 2020, and the main land-use types in the study area. Source of land use data: IBGE (2018). Map produced using ArcMap 10.6.1.

and stemborer damage (16.6%). Differences in expected maize yield losses according to pest types were rather low, with 18.8% for bacteria, 15.79% for fungi and 15.43% for insects. Likewise, expected soybean yield losses varied little among different pest types, with 13.05% for mites, 12.63% for fungi and 11.67% for insects. **Figure 8** shows the expected soybean and maize yield losses according to different P&A.

The study results indicate that the biggest threat to maize production are maize rust, bacterial stalk rot and the fall armyworm, as these three P&A cause high yield losses and are also among the most widespread diseases according to the ConvNet predictions and information provided by the Mato Grosso Foundation (Fundação, 2019). For soybean, the picture is less clear: the Sudden Death Syndrome was reported to cause the highest yield losses, but it was relatively seldom predicted by the ConvNets. P&A like anthracnose, downy mildew, frogeye leaf spot, stink bugs and brown spot seem to pose a greater threat as they are widespread and cause average yield losses of 10–15%.

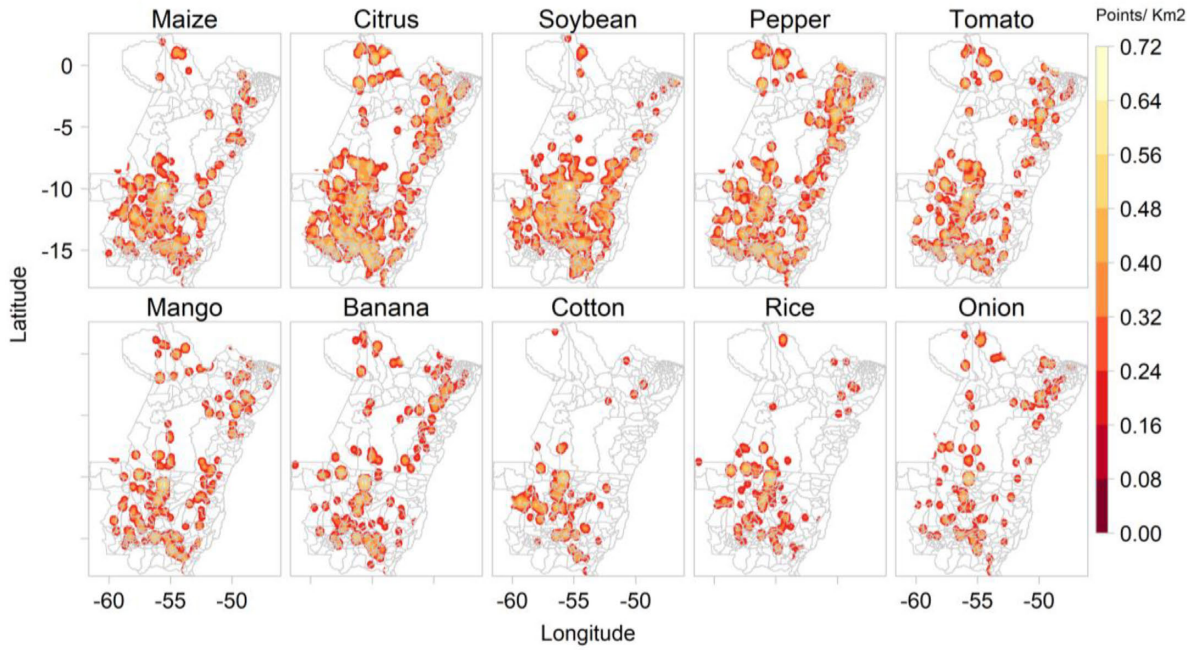

FIGURE 4 | Kernel density of images collected by *Plantix* users in Mato Grosso and Pará, Brazil, between November 2016 and May 2020, according to predicted crop types.

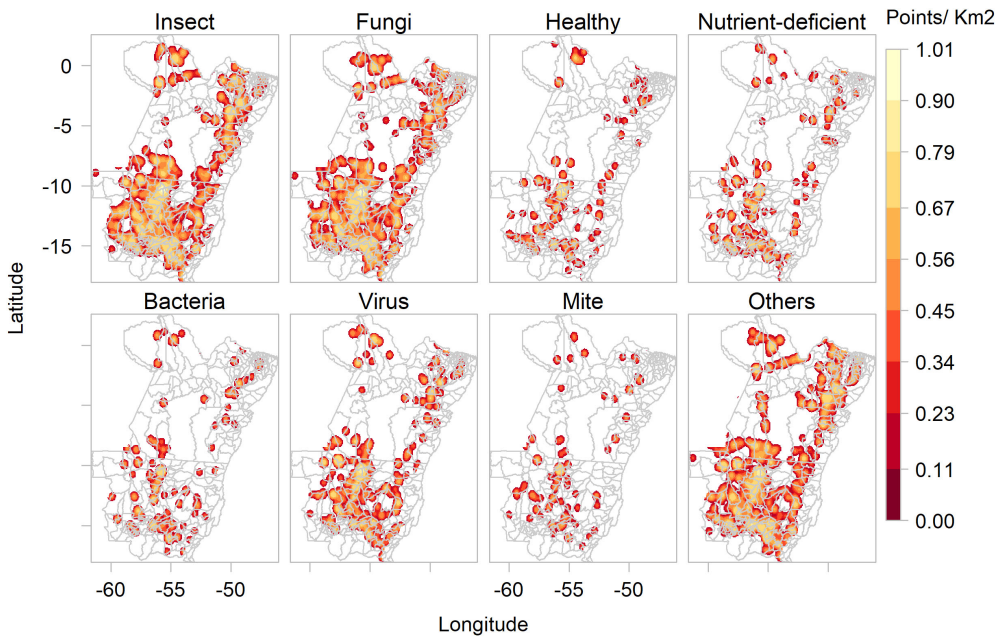

FIGURE 5 | Kernel density of images collected by *Plantix* users and agronomists in Mato Grosso and Pará, Brazil, between November 2016 and May 2020, according to predicted pest types and healthy plants.

The Asian soybean rust, which was a long time the most severe disease in the study area, seems to be relatively well controlled with average yield losses of 10.5%.

Figure 9 shows examples of P&A images for which agronomists in Mato Grosso provided an estimate of expected yield losses. Most of the images showed diseased plants with mild symptoms and expected yield losses below 20%. The examples demonstrate that the angle, distance and quality of the recorded images may vary considerably and that the background may also be noisy, containing other plant material or soil. Moreover,

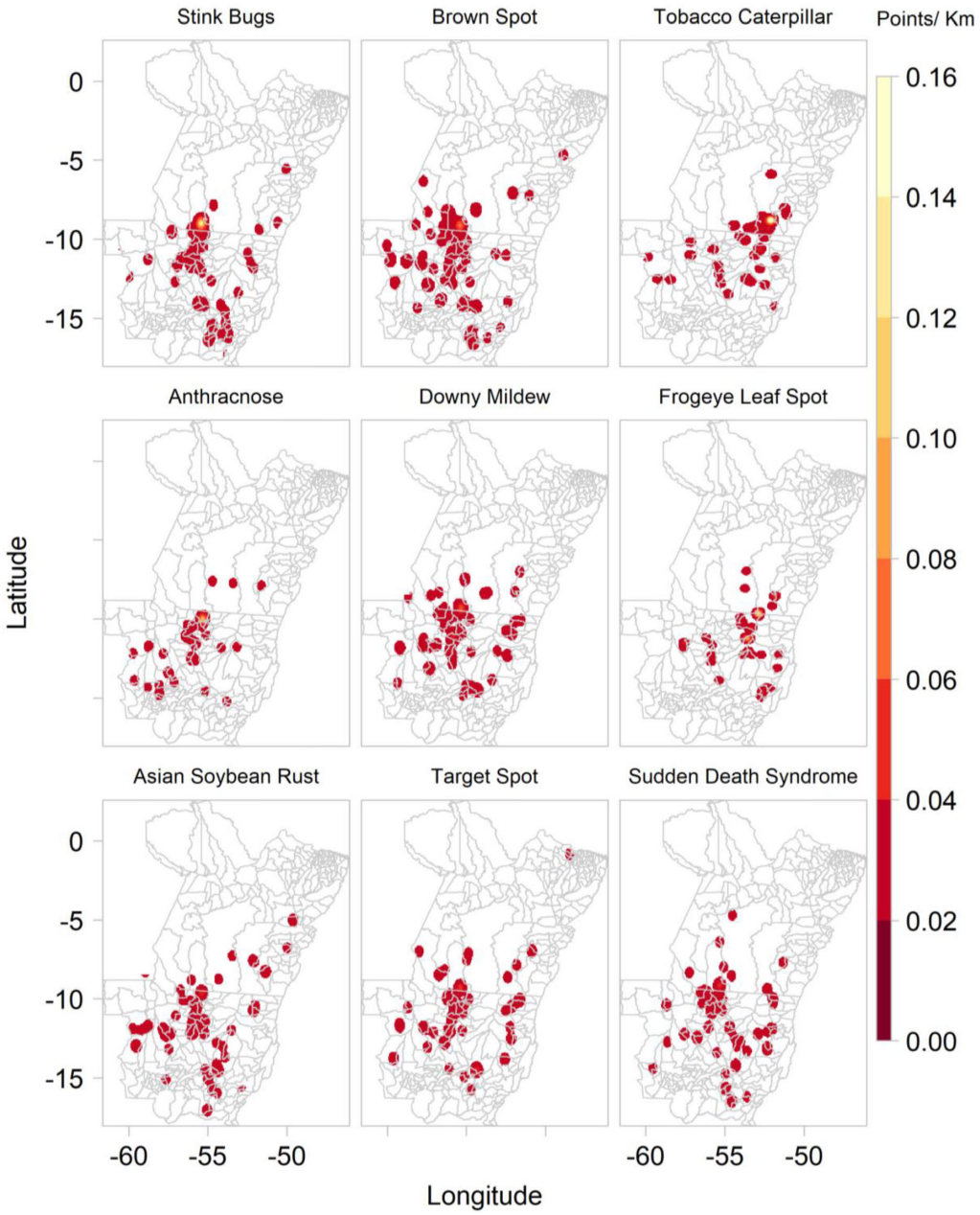

FIGURE 6 | Kernel density of main soybean pathogens and animal pests based on ConvNet predictions provided for images collected by *Plantix* users and agronomists in Mato Grosso and Pará, Brazil, between November 2016 and May 2020.

mild symptoms may be hard to recognise from an image alone, especially non-foliar diseases (e.g., Anthracnose of soybean). These images also exemplify the difficulty plant experts may be confronted with when annotating P&A images.

Expected Yield and Economic Losses per Municipality and at the State Level

Between 2016 and 2018, farmers in MT and PA produced an annual average of approximately 24.6 million tonnes of maize (first and second season) and approximately 31 million tonnes of soybean (IBGE, 2019b). Most of the production originated from Central MT (the municipalities of Sorriso and Nova Mutum, **Figure 10A**), where also most of the images with estimates on expected yield losses were collected. The expected yield loss estimates at the municipality level reveal that the percentage yield losses range between 7.5 and 23% for soybean and between 11.6 and 22.12% for maize (**Figure 10B**). Overall, yield loss estimates were available for 78 and 60% of all soybean and maize-producing municipalities, which represent 95 and 97% of the region's total respective soybean and maize production. The

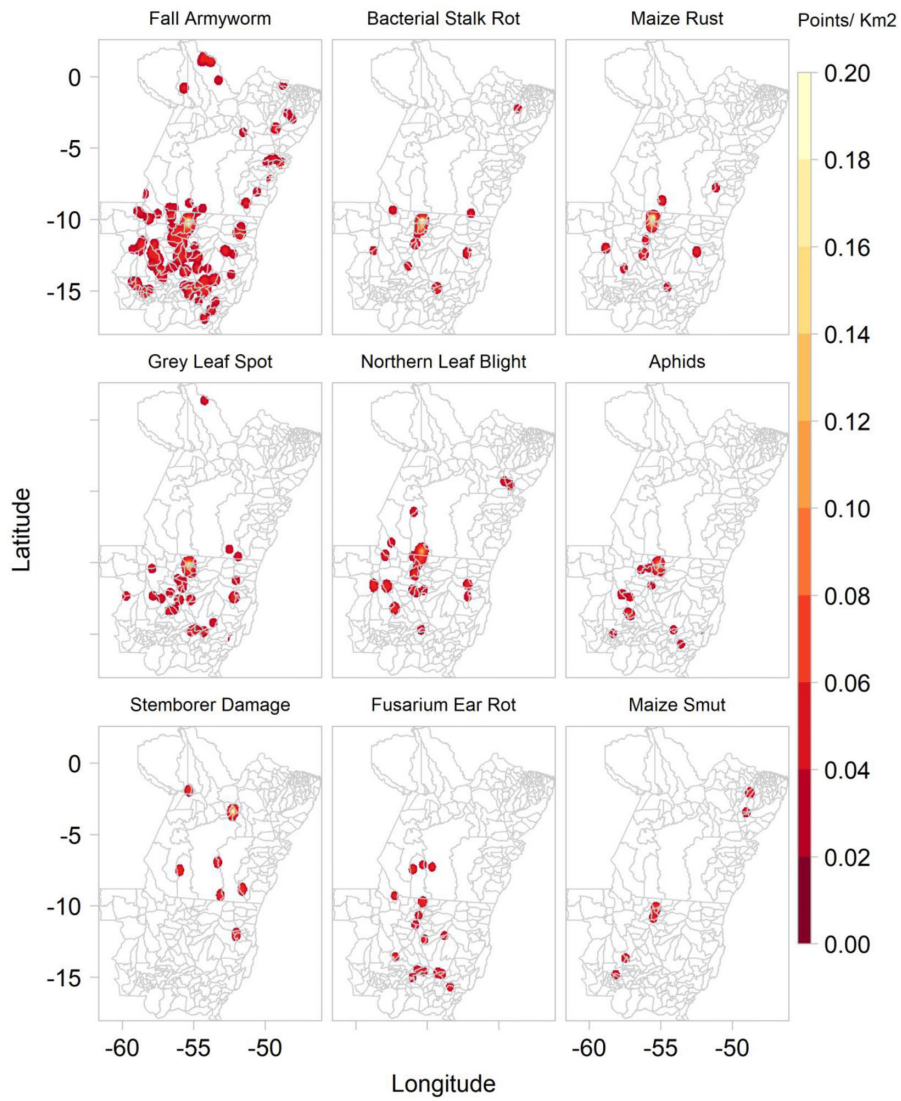

FIGURE 7 | Kernel density of the main maize pathogens and animal pests based on ConvNet predictions provided for images collected by *Plantix* users in Mato Grosso and Pará, Brazil, between November 2016 and May 2020.

estimation of yield losses at the state level reveals that, on average, 3.74 and 3.75 million tonnes of soybean and maize, respectively, were lost in MT and PA between 2016 and 2018 (**Figure 10C**). This translates into economic losses of around US $2 billion per cropping season ($1.33 and $0.6 billion for soybean and maize, respectively).

DISCUSSION

The launch of the *Plantix* mobile application demonstrated that ConvNets trained for the automatic classification of plant diseases are not only ready to be put into operational use, but that such a decision support tool can achieve great attention and widespread use among farmers, generating data that can, in turn, be used by science. Based on the lessons learned from *Plantix* in Brazil, the following section outlines the potential benefits and limitations of an automated plant disease classification via a mobile application, as well as possible benefits for the scientific community—especially in the field of crop loss research.

Crowdsourcing the Collection of P&A Images

ConvNets—like other supervised machine learning algorithms—require large amounts of human-annotated data to be trained successfully. However, the development of such a human-annotated image database for plant diseases has been one of the major challenges in further improving the performance of ConvNets and making them fully operational in the field (Barbedo, 2018a,b; Boulent et al., 2019). One option to generate such a database is to crowdsource it. PEAT has relied on

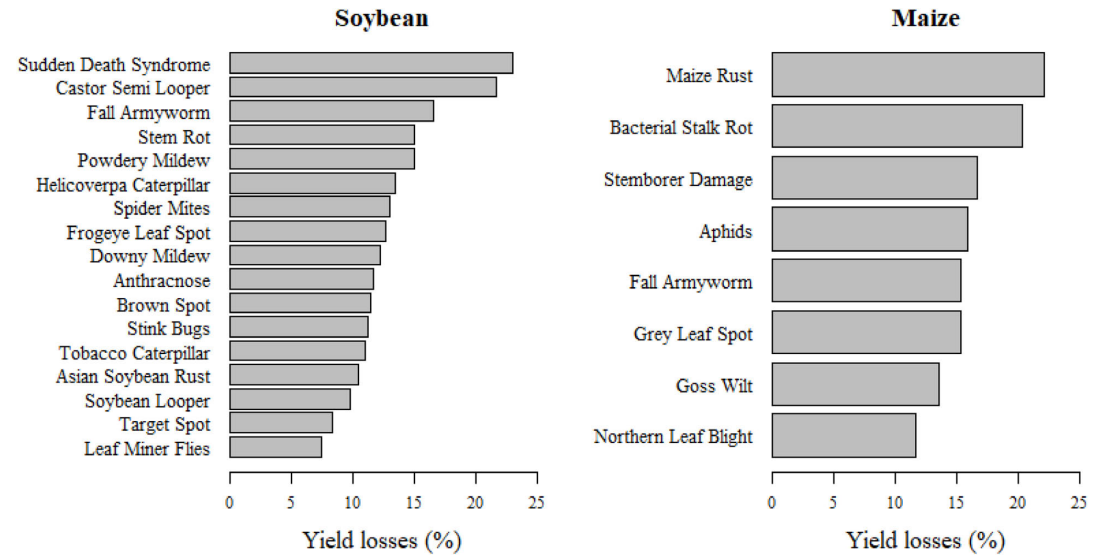

FIGURE 8 | Expected soybean and maize yield losses (%) in Mato Grosso and Pará, Brazil, according to individual pathogens and animal pests.

just such a crowdsourcing process, as the images taken by *Plantix* users constantly add to a database that can—at least in part—be used for model training. The advantage of this crowdsourced image data collection process is that images are acquired at different locations, at different hours of the day, under different meteorological conditions and with different smartphone devices, thus accurately reflecting the reality of the operational environment. The growing field of citizen science could help scientists to crowdsource the collection of P&A images. Citizen science means the involvement of the general public in scientific work, often in collaboration with or under the direction of professional scientists (European Commission, 2013). Citizen science projects can be found in fields ranging from astronomy to medicine and computer science to earth observations, including from the field of plant pathology (Walther and Kampen, 2017; Luigi Nimis et al., 2018; D'Agostino et al., 2020).

Potential Benefits and Limitations of ConvNet Systems as a Decision Support Tool for Farmers

The implementation of ConvNets trained for the automatic classification of plant diseases into a mobile application proved to be a useful decision support tool for farmers and gardeners. In particular, the high top-3 score (98%) indicates that a diagnosis given by the application can help farmers to identify the correct diseases among a pre-selected list. A simple, easy-to-use and free tool is particularly attractive for small-scale farmers, who often do not have access to agricultural extension services or who lack the financial means for such services. One major advantage of such a tool is that it can be used by anyone (regardless of education level or scouting experience) with an internet-enabled smartphone. Furthermore, the services are immediately available at any time, but leave the user with the final decision of whether and how to protect or treat the plants. The app can also help farmers to reduce the number of pesticide applications by promoting the adoption of non-chemical methods, such as pheromones, biopesticides or the removal and burning of affected plant parts (integrated pest management). Moreover, the functionality of the app can be expanded to include an early warning system by sending push-notifications in case a disease has been spotted in a nearby plot or by recommending fertiliser use and the timing of sowing, among other benefits.

One disadvantage of ConvNets in diagnosing plant diseases is the timing of the detection: Since this form of detection relies on visual symptoms, the earliest possible detection is when symptoms are visible to the human eye and can be recorded by a camera. However, plants may be affected by a pathogen much earlier and consequently react to its presence with, e.g., a reduction in the photosynthesis rate, which induces an increase in fluorescence and heat emission (Martinelli et al., 2014). Therefore, the earliest management of a pest or disease can only happen if there has already been a visual change in the plant material. Moreover, the success and usability of such a tool not only depends on its classification performance but also its availability in different languages and the inclusion of locally relevant diseases. A major hurdle for any system's use is also the affordability of smartphones and internet connectivity. Although the smartphone penetration rate is increasing rapidly (45% of the Brazilian population actively used a smartphone in 2019; Newzoo, 2019), it is still limited in rural areas and among farm households. According to the latest agricultural census, in MT and PA, only 26% of farmers in MT and only 13% in PA have access to the internet. Even fewer farmers use the mobile internet: 14 and 10% in MT and PA, respectively (IBGE, 2019a). Attempts to distribute such a smartphone app in other countries, e.g., Nigeria, basically failed because too few people own smartphones in rural areas.

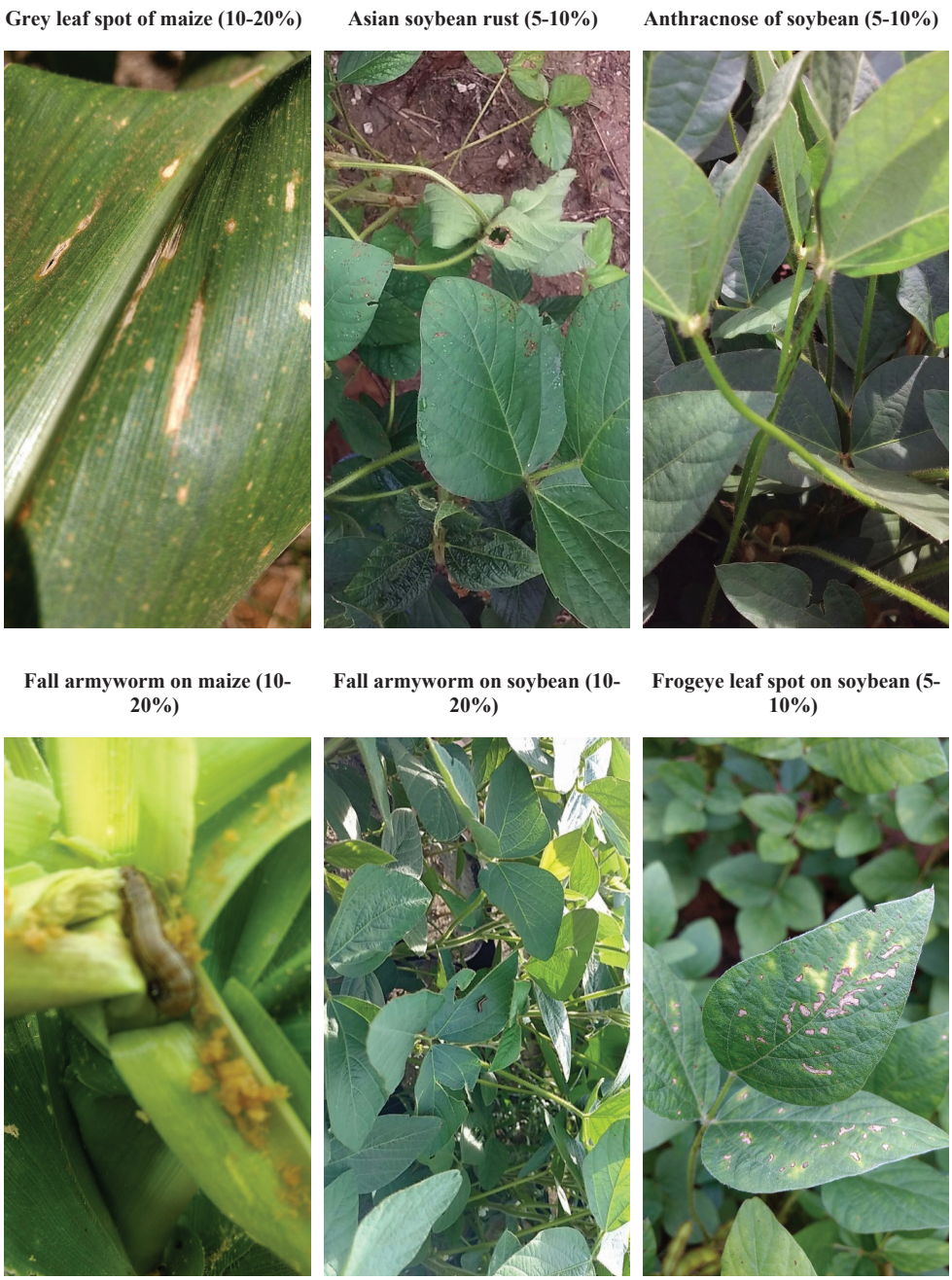

FIGURE 9 | Examples of images of diseased soybean and maize plants and associated expected yield losses provided by agronomists in the state of MT, Brazil, in the 2016–2017 cropping season.

Reducing Yield Losses to Pests and Diseases

The study results indicate that soybean and maize yield losses amount, on average, to 12 and 16%, respectively, and are hence slightly lower than reported yield loss estimates at the global scale of 19 and 21% for these two crops (Oerke, 2006; Savary et al., 2019). This is likely due to the massive and prophylactic application of pesticides in the study area (Pignati et al., 2017; INDEA, 2020). Another explanation for the rather low yield loss estimates is that much of the data was collected in Central MT, where the most productive and experienced farmers are located, and where the cultivation of soybean and maize is an established practice. The study results might, therefore, underestimate yield losses in other municipalities and hence also at the state level. Estimated annual soybean yield losses of 3.74 million tonnes, as well as corresponding economic losses of US $1.33 billion, are within the range of estimates provided by Godoy et al. (2016). The

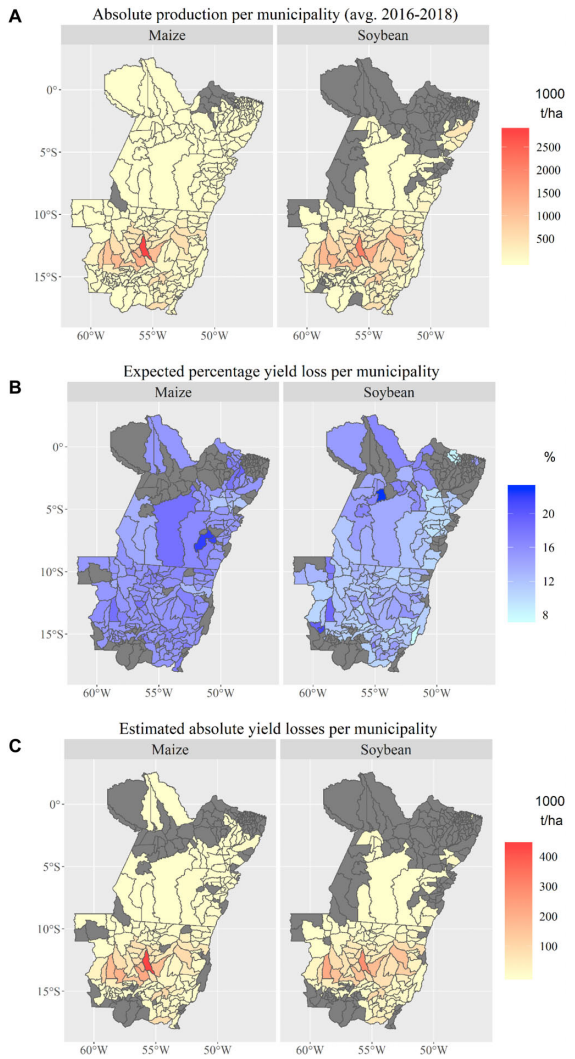

FIGURE 10 | (A) Absolute maize and soybean production in municipalities of Mato Grosso and Pará, Brazil; **(B)** expected maize and soybean yield losses (%) due to the most common P&A in the study area; and **(C)** absolute maize and soybean production losses per municipality in Mato Grosso and Pará, Brazil.

results also indicate that there is a large variation in crop losses due to specific P&A, which is in agreement with the findings of Savary et al. (2019), but a rather low variation in yield losses according to pathogen types (e.g., fungi, bacteria).

P&A are mainly controlled through the intensive use of pesticides in combination with the cultivation of GM crops, which may cause negative effects for human health, such as acute and chronic intoxication (Pignati et al., 2017) and biodiversity loss, as well as the development of pest resistance to pesticides (Karlsson Green et al., 2020). One alternative, holistic approach to combating pests is integrated pest management (IPM), which combines preventive and curative methods, and only applies chemical pesticides when there is an urgent need (Karlsson Green et al., 2020). A field experiment jointly established by the Embrapa and Aprosoja in MT to test the efficacy of IPM demonstrated that areas managed using IPM measures produced the same yield as areas with conventional management, but used approximately 50% less insecticide (Bueno et al., 2020). Despite its large potential to decrease pesticide use as well as production costs, the adoption of IPM in Brazil sharply declined in the 2000s due to the introduction of double-cropping and no-tillage systems (Panizzi, 2013). Reviving the adoption of IPM among farmers through targeted public policies and governmental funding agencies, as well as the adaptation of IPM to new circumstances and production systems, can help to minimise biotic yield losses while maintaining environmental quality.

Possible Implications for Crop Loss Research and Limitations of This Study

This study contributes to the field of crop loss research by providing probability distribution maps and yield loss estimates for the main soybean and maize P&A of the Southern Amazon, one of Brazil's—and the world's—most important agricultural regions. These yield loss estimates fill a major data gap and comprise one of the few spatially explicit available datasets for different P&A in Brazil. The analysis provided here can easily be extended to other crops or world regions as more data becomes available, which in turn will enable future researchers to train the underlying ConvNets for more crop disease combinations. Besides, the georeferenced images can be combined with other spatial data (e.g., climate, soil data) to identify factors influencing the outbreak and spread of diseases (Wieland et al., 2017) and to model and predict their spatio-temporal distribution. The georeferenced images collected by *Plantix* users can also be used for other purposes, such as ground-truth labels for the classification of crops and diseases via satellite images. One example of such an application can be found in Wang et al. (2020), who used the *Plantix* image database and deep learning to map crop types in southeast India. Nonetheless, the probability distribution maps and reported yield loss estimates provided in the present study must be interpreted with caution: despite the data cleaning steps applied, which reduced the original dataset by 80%, the probability distribution maps in this study might be biased, as the collection of data points depended on the number of active users in an area. Although images were collected in almost all crop-producing areas and a great deal of effort was devoted to advertising the application throughout the study area, some diseases might be underrepresented or might not have been captured at all.

CONCLUSION

The overall objective of this study was to map the spatial distribution of the main soybean and maize diseases in the Southern Amazon and to quantify the associated yield losses by making use of data collected using the *Plantix* smartphone application. Soybean and maize yield losses to P&A in the Southern Amazon were found to be lower than biotic yield losses reported for these crops in other

world regions. A likely explanation is the massive and prophylactic application of large amounts of pesticides in the study area. Integrated pest management can be a sustainable alternative to the intensive use of pesticides, helping to minimise negative outcomes for human health, biodiversity and the environment. ConvNets can aid farmers in the early detection and non-chemical control of P&A, while crowdsourcing may aid researchers in gathering training data that accurately reflects the target operational environment. The high level of accuracy of the trained ConvNets, paired with widespread use through a citizen science approach, provides a unique source of data that allows scientists to get a new angle on yield loss estimates, e.g., for the analysis of yield gaps and the development of measures to minimise them.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AH, RS, and CN: conceptualisation. AH and RS: data preparation. AH: methodology and writing—original draft preparation. AH, SS, RS, and CN: writing—review and editing. CN: supervision. All authors have read and agreed on the published version of the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge the scholarship and travel funds awarded to the corresponding author by the Heinrich Böll Foundation. We would like to thank all *Plantix* users who collected images of diseased plants within the study area, as well as all the agronomists who provided yield loss estimates. We were grateful to Ralf Wieland from ZALF and Thomas Friedel from PEAT for sharing their expertise in the field of machine learning. We were thankful for the helpful comments and suggestions made by the reviewers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.621168/full#supplementary-material

REFERENCES

ABRAPA (2011). .*Algodão no Cerrado do Brasil*, 2nd Edn, ed. E. C. Freire (Goiânia: Mundial Gráfica).

Arvor, D., Dubreuil, V., Ronchail, J., Simões, M., and Funatsu, B. M. (2014). Spatial patterns of rainfall regimes related to levels of double cropping agriculture systems in Mato Grosso (Brazil). *Int. J. Climatol.* 34, 2622–2633. doi: 10.1002/joc.3863

Barbedo, J. G. A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. *SpringerPlus* 2, 660.

Barbedo, J. G. A. (2018a). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. doi: 10.1016/j.biosystemseng.2018.05.013

Barbedo, J. G. A. (2018b). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53. doi: 10.1016/j.compag.2018.08.013

Barbedo, J. G. A., Koenigkan, L. V., and Santos, T. T. (2016). Identifying multiple plant diseases using digital image processing. *Biosyst. Eng.* 147, 104–116. doi: 10.1016/j.biosystemseng.2016.03.012

Boulent, J., Foucher, S., Theau, J., and St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10:941. doi: 10.3389/fpls.2019.00941

Bregaglio, S., and Donatelli, M. (2015). A set of software components for the simulation of plant airborne diseases. *Environ. Modell. Softw.* 72, 426–444. doi: 10.1016/j.envsoft.2015.05.011

Bueno, A. F., Panizzi, A. R., Hunt, T. E., Dourado, P. M., Pitta, R. M., and Goncalves, J. (2020). Challenges for adoption of integrated pest management (IPM): the soybean example. *Neotrop. Entomol.* 50, 5–20. doi: 10.1007/s13744-020-00792-9

Céleres (2018). *20 Anos da Adoção da Biotecnologia Agrícola no Brasil: Lições Aprendidas e Novos Desafios*. Uberlandia: Celeres.

CEPEA (2019). *Mensuração Econômica da Incidência de Pragas e Doenças no Brasil: Uma Aplicação Para as Culturas de Soja, Milho e Algodão*. Piracicaba: CEPEA.

CONAB (2019). *Séries Históricas de Área Plantada, Produtividade e Produção, Relativas às Safras 1976/77 a 2018/19 de Grãos*. Rio Branco: CONAB.

CONAB (2020). *Séries Históricas de Área Plantada, Produtividade e Produção, Relativas às Safras 1976/77 a 2018/19 de Grãos*. Available online at: https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras?start=20 (accessed Augest 15, 2020).

D'Agostino, D., Law-Green, D., Watson, M., Novara, G., Tiengo, A., Sandrelli, S., et al. (2020). A citizen science exploration of the X-ray transient sky using the EXTraS science gateway. *Future Gen. Comput. Syst.* 111, 806–818. doi: 10.1016/j.future.2019.10.030

Donatelli, M., Magarey, R. D., Bregaglio, S., Willocquet, L., Whish, J. P. M., and Savary, S. (2017). Modelling the impacts of pests and diseases on agricultural systems. *Agric. Syst.* 155, 213–224. doi: 10.1016/j.agsy.2017.01.019

Esker, P. (2012). Crop loss analysis and global food supply: focusing now on required harvests. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 7, 1–14. doi: 10.1079/pavsnnr20127052

European Commission (2013). *Green Paper on Citizen Science. Citizen Science for Europe. Towards a Better Society of Empowered Citizens and Enhanced Research*. Brussels: European Commission.

FAOSTAT (2019). *Pesticides use Dataset*. Available online at: http://www.fao.org/faostat/en/#data/RP (accessed December 01, 2019).

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009

Fundação, M. T. (2019). *Boletim de Pesquisa N°18, 2017/2018*. Rondonópolis: Fundação, MT.

Godoy, C. V., Seixas, C. D. S., Soares, R. M., Marcelino-Guimarães, F. C., Meyer, M. C., and Costamilan, L. M. (2016). Asian soybean rust in Brazil: past, present, and future. *Pesq. Agropec. Bras.* 51, 407–421. doi: 10.1590/s0100-204x2016000500002

Hart, T., and Zandbergen, P. (2014). Kernel density estimation and hotspot mapping. *Policing Int. J. Police Strategies Manag.* 37, 305–323. doi: 10.1108/pijpsm-04-2013-0039

Hendrycks, D., and Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks. *Paper Presented at the 5th International Conference on Learning Representations*, Toulon.

Hijmans, R. J. (2019). *Raster: Geographic Data Analysis and Modeling. R Package Version 3.0-7*. Available online at: https://CRAN.R-project.org/package=raster (accessed June, 2019).

Hughes, D. P., and Salathe, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* [Preprint]. 1511.08060.

IBGE (2018). *Monitoramento da Cobertura e uso da Terra do Brasil 2014-2016*. Rio de Janeiro: IBGE.

IBGE (2019a). *Censo Agropecuário: Resultados definitivos 2017*. Available online at: https://censos.ibge.gov.br/agro/2017 (accessed October 25, 2019).

IBGE (2019b). *Pesquisa Agrícola Municipal. Tabela 5457: Área plantada ou Destinada à colheita, Área Colhida, Quantidade Produzida, Rendimento Médio e Valor da Produção das Lavouras Temporárias e Permanentes*. Available online at: https://sidra.ibge.gov.br/pesquisa/pam/tabelas (accessed November 13, 2019).

ImageCLEF (2018). *ImageCLEF- Cross-Language Image Retrieval Evaluations: ExperLifeCLEF 2018*. Available online at: https://www.imageclef.org/node/231 (accessed October 10, 2018).

IMF (2020). *Primary Commodity Price System. Market Prices for Non-Fuel and Fuel Commodities*. Available online at: https://data.imf.org/?sk=471DDDF8-D8A7-499A-81BA-5B332C01F8B9 (accessed August 18, 2020).

INDEA (2020). *Relatório de Comércio de Agrotóxicos Consolidado*. Available online at: http://www.indea.mt.gov.br/-/6099478-agrotoxicos?ciclo= (accessed April 6, 2020).

INMET (2019). *Estações e Dados. Banco de Dados Meteorológicos para Ensino e Pesquisa*. Available online at: http://www.inmet.gov.br (accessed May 17, 2019).

Karlsson Green, K., Stenberg, J. A., and Lankinen, A. (2020). Making sense of Integrated Pest Management (IPM) in the light of evolution. *Evol. Appl.* 13, 1791–1805. doi: 10.1111/eva.13067

Lamigueiro, O. P., and Hijmans, R. J. (2019). *RasterVis. R Package Version 0.47*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Luigi Nimis, P., Pittao, E., Altobelli, A., De Pascalis, F., Laganis, J., and Martellos, S. (2018). Mapping invasive plants with citizen science. A case study from Trieste (NE Italy). *Plant Biosyst.* 153, 700–709. doi: 10.1080/11263504.2018.1536085

Mahlein, A. K. (2016). Plant disease detection by imaging sensors – Parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251. doi: 10.1094/PDIS-03-15-0340-FE

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., et al. (2014). Advanced methods of plant disease detection. A review. *Agron. Sustain. Dev.* 35, 1–25. doi: 10.1007/s13593-014-0246-1

Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls.2016.01419

Nelson, A. (2017). Crop pests: crop-health survey aims to fill data gaps. *Nature* 541:464. doi: 10.1038/541464a

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes. R Package Version 1.1-2*.

Newzoo (2019). *Global Mobile Market Report*. Available online at: https://resources.newzoo.com/hubfs/Reports/2019_Free_Global_Mobile_Market_Report.pdf (accessed August 12, 2020).

Oerke, E. C. (2006). Crop losses to pests. *J. Agric. Sci.* 144:31. doi: 10.1017/s0021859605005708

Oerke, E. C., and Dehne, H. W. (2004). Safeguarding production—losses in major crops and the role of crop protection. *Crop Protect.* 23, 275–285. doi: 10.1016/j.cropro.2003.10.001

Panizzi, A. R. (2013). History and contemporary perspectives of the integrated pest management of soybean in Brazil. *Neotrop. Entomol.* 42, 119–127. doi: 10.1007/s13744-013-0111-y

Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., and Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* 161, 280–290. doi: 10.1016/j.compag.2018.04.002

Pignati, W. A., Lima, F., Lara, S. S., Correa, M. L. M., Barbosa, J. R., Leao, L., et al. (2017). Spatial distribution of pesticide use in Brazil: a strategy for Health Surveillance. *Cien. Saude Colet.* 22, 3281–3293. doi: 10.1590/1413-812320172210.17742017

Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, merkedness & correlation. *J. Mach. Learn. Technol.* 2, 37–63.

R Core Team (2020). *R: A language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramcharan, A., McCloskey, P., Baranowski, K., Mbilinyi, N., Mrisho, L., Ndalahwa, M., et al. (2019). A Mobile-Based Deep Learning Model for Cassava Disease Diagnosis. *Front Plant Sci* 10:272. doi: 10.3389/fpls.2019.00272

Savary, S., Ficke, A., Aubertot, J.-N., and Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food Secur.* 4, 519–537. doi: 10.1007/s12571-012-0200-5

Savary, S., Jouanin, C., Félix, I., Gourdain, E., Piraux, F., Willocquet, L., et al. (2016). Assessing plant health in a network of experiments on hardy winter wheat varieties in France: multivariate and risk factor analyses. *Eur. J. Plant Pathol.* 146, 757–778. doi: 10.1007/s10658-016-0955-1

Savary, S., Nelson, A. D., Djurle, A., Esker, P. D., Sparks, A., Amorim, L., et al. (2018). Concepts, approaches, and avenues for modelling crop health and crop losses. *Eur. J. Agron.* 100, 4–18. doi: 10.1016/j.eja.2018.04.003

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3, 430–439. doi: 10.1038/s41559-018-0793-y

Sharma, S., Sharma, S., and Althaiya, A. (2020). Activation functions in neural networks. *Int. J. Eng. Appl. Sci. Technol.* 4, 310–316.

Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., et al. (2019). MnasNet: platform-aware neural architecture search for mobile. *Paper Presented at the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA.

Tan, M., and Le, Q. V. (2019). EfficientNet: rethinking model scaling for convolutional neural networks. *Paper Presented at the 36th International Conference on Machine Learning*, Long Beach, CA.

Tennekes, M. (2019). *Tmaptools: Thematic Map Tools. R Package Version 2.0-2*. Available online at: https://github.com/mtennekes/tmaptools (accessed September, 2019).

van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., and Hochman, Z. (2013). Yield gap analysis with local to global relevance—A review. *Field Crops Res.* 143, 4–17. doi: 10.1016/j.fcr.2012.09.009

Walther, D., and Kampen, H. (2017). The citizen science project 'Mueckenatla' helps monitor the distribution and spread of invasive mosquito species in Germany. *J. Med. Entomol.* 54, 1790–1794. doi: 10.1093/jme/tjx166

Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., et al. (2020). Mapping crop types in southeast India with smartphone crowdsourcing and deep learning. *Remote Sens.* 12:2957. doi: 10.3390/rs12182957

Wieland, R., Kerkow, A., Früh, L., Kampen, H., and Walther, D. (2017). Automated feature selection for a machine learning approach toward modeling a mosquito distribution. *Ecol. Modell.* 352, 108–112. doi: 10.1016/j.ecolmodel.2017.02.029

Conflict of Interest: SS and RS are the founders of the company PEAT GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hampf, Nendel, Strey and Strey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 May 2021
doi: 10.3389/fpls.2021.620273

Multiscale Parallel Algorithm for Early Detection of Tomato Gray Mold in a Complex Natural Environment

*Xuewei Wang and Jun Liu**

Shandong Provincial University Laboratory for Protected Horticulture, Blockchain Laboratory of Agricultural Vegetables, Weifang University of Science and Technology, Weifang, China

Plant disease detection technology is an important part of the intelligent agricultural Internet of Things monitoring system. The real natural environment requires the plant disease detection system to have extremely high real time detection and accuracy. The lightweight network MobileNetv2-YOLOv3 model can meet the real-time detection, but the accuracy is not enough to meet the actual needs. This study proposed a multiscale parallel algorithm MP-YOLOv3 based on the MobileNetv2-YOLOv3 model. The proposed method put forward a multiscale feature fusion method, and an efficient channel attention mechanism was introduced into the detection layer of the network to achieve feature enhancement. The parallel detection algorithm was used to effectively improve the detection performance of multiscale tomato gray mold lesions while ensuring the real-time performance of the algorithm. The experimental results show that the proposed algorithm can accurately and real-time detect multiscale tomato gray mold lesions in a real natural environment. The F1 score and the average precision reached 95.6 and 93.4% on the self-built tomato gray mold detection dataset. The model size was only 16.9 MB, and the detection time of each image was 0.022 s.

Keywords: multiscale, convolutional neural network, tomato gray mold, object detection, intelligent agriculture, deep learning, plant diseases

OPEN ACCESS

Edited by:
Spyros Fountas, Agricultural University of Athens, Greece

Reviewed by:
Evangelos Anastasiou, Agricultural University of Athens, Greece
Yunchao Tang, Zhongkai University of Agriculture and Engineering, China

***Correspondence:**
Jun Liu
liu_jun860116@wfust.edu.cn

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *22 October 2020*
Accepted: *30 March 2021*
Published: *11 May 2021*

Citation:
Wang X and Liu J (2021) Multiscale Parallel Algorithm for Early Detection of Tomato Gray Mold in a Complex Natural Environment. Front. Plant Sci. 12:620273. doi: 10.3389/fpls.2021.620273

INTRODUCTION

Plant diseases are the main cause of food loss in the world's economy. Food loss from crop infections caused by pathogens such as bacteria, viruses, and fungi is a persistent problem. This situation is further complicated by the fact that disease is more likely to metastasize globally now than ever before. In order to minimize the damage caused by diseases during crop growth, crop prevention is imperative. Traditionally, crop inspections and plant diseases are determined by farmers or experts with some training or experience. This manual method is expensive because it requires continuous monitoring and is not feasible for larger areas.

Tomato is one of the largest vegetable crops planted in China because of its variety, abundant nutrition, and high yield. However, in recent years, gray mold, leaf mold, early blight, late blight, and other common diseases of greenhouse tomato frequently occur, which brings serious threats to the yield and quality of greenhouse tomato. Tomato gray mold is a worldwide infectious disease caused by *Botrytis cinerea*, and it is harmful to the growth of greenhouse tomato. The disease has the characteristics of easy occurrence, rapid spread, strong fungicide resistance, and great economic loss. It has become one of the key factors to inhibit the safe production of greenhouse tomato (Elad and Shtienberg, 1995). In addition to harming tomato, the pathogen can also harm more than 20

crops such as eggplant, pepper, and cucumber. The incidence of tomato gray mold mostly starts from the tip of the leaf and expands inward along the veins in a "V" shape, initially watery, and then yellowish brown, with deep and shallow striate lines on the edges. The boundary between the disease lesion and healthy tissue is clear. Gray mold layer can emerge on the surface of the lesion when the humidity is high, and the region above the pathogenic part can be affected leading to death in severe cases (O'Neill et al., 1997). According to a preliminary field investigation, the tomato yield loss caused by this disease is generally 15–20% and even reaches about 30% in severe cases. The yield loss of tomato gray mold in greenhouse has caused serious losses to national economy and people's life and has become the main reason that restricts the high efficiency and safe production of vegetables. Therefore, effective control of disease occurrence becomes the key to achieve sustainable development of the tomato industry. At present, chemical agents are mainly used to control tomato gray mold in production. The amount of fungicides used to control gray mold is more than 60% of the amount of fungicides used throughout the planting season. Due to frequent fungicide use, fungicide resistance is prominent and the prevention effect is unsatisfactory (Borges et al., 2015). In the concept of digital agriculture, precise and rational application of fungicides is the direction of agricultural development, and achieving rational application of fungicides is one of the important measures to ensure high yield and safety of vegetables and fruits. It can not only effectively control the occurrence of diseases but also effectively reduce environmental pollution. One of the preconditions for the rational application of fungicides is that information of vegetable growth status must be accurately obtained, among which the key basic technology is to quickly and accurately obtain the types of vegetable diseases and their degree of disease damage. Therefore, the rapid and accurate diagnosis of diseases is an important measure to ensure the high yield and safety of vegetables, which has great practical significance to improve the green, safe, and sustainable production capacity of vegetables.

Existing studies mainly focus on exploring the occurrence of diseases, and there are only a few studies dealing with the early detection of diseases (Nigam and Jain, 2020). Traditional diagnostic methods for plant diseases usually obtain diagnostic results after comprehensive analysis of plant diseases by plant protection workers based on experience and pathological analysis, but these traditional detection methods are inefficient, involve a heavy workload, have poor real-time performance, and are unable to achieve early and rapid diagnosis, often delaying the optimal treatment period of diseases, increasing the dosage of fungicides, and increasing costs and environmental pollution (Martinelli et al., 2015). Because of the complexity and variability of diseases of a large number of plants, even experienced phytopathologists cannot accurately diagnose a specific disease. It is also worth noting that many agricultural regions are difficult to be properly monitored throughout the process (Barbedo, 2013). Early detection of pathogens is essential to reduce disease transmission and promote effective management practices (Sankaran et al., 2010). It is very important to seek a rapid and accurate early detection method.

With the rapid development of facility agriculture in Shouguang City, Shandong Province, China, tomato gray mold has increasingly become a limiting factor affecting tomato production and development, especially in early spring and late autumn tomato cultivation. Using deep neural network to extract features is better than traditional feature extraction methods, and this study will continue to use deep learning methods to detect tomato gray mold lesion objects with different sizes. In this study, images of healthy and infected tomato gray mold leaves in a real natural environment were collected. A multiscale parallel network structure from dense to sparse was proposed based on the universal object detection method YOLO (you only look once). The aims were to establish a rapid and accurate early detection model for tomato gray mold and provide scientific basis for the early diagnosis of tomato gray mold.

RELATED WORK

Intelligent Agriculture

Intelligent agriculture, relying on modern information technology, has achieved precision management and visual diagnosis of agricultural production through intelligent perception of the agricultural production environment and data analysis. It is the highest form of agricultural development. Machine vision and its associated emerging technologies hold great potential in intelligent agricultural applications (Tang et al., 2020). Li J. et al. (2020) presented a reliable algorithm based on field RGB-D camera, which can detect and locate fruiting branches of multiple litchi groups accurately and automatically in a large environment. Chen et al. (2020) established a measurement framework of orchard harvesting operation based on multivision technology, and the experimental results show that the proposed adaptive stereo matching strategy has high matching accuracy and stable performance for different sampling depths. Lin et al. (2020) presented a fruit detection method in a natural environment using partial shape matching and probabilistic Hough transform, and experiments on datasets of citrus, tomato, pumpkin, melon, luffa, and mango show that this method is competitive for the detection of most types of fruits in a natural environment, such as green, orange, circular, and noncircular. These studies show that intelligent agriculture based on machine vision and image processing technology has become a key research field in the new agricultural information technology.

Plant Disease Detection

As an important part of intelligent agriculture, plant disease detection provides a theoretical basis for the scientific formulation of disease control measures and scientific application of drugs. With the application of convolutional neural network in the field of computer vision, the research of plant disease detection has developed rapidly. Convolutional neural network (CNN) is known as a general function simulator, and its ability to fit features is much stronger than that designed based on experience. Initially, researchers simply applied CNN to plant disease detection task, applied network structure for classification

to detection task, and only used CNN to extract features from samples. After that, researchers proposed end-to-end plant disease detection models, which could achieve better detection results under a relatively ideal environment (Mohanty et al., 2016; Amara et al., 2017; Durmus et al., 2017; Wang et al., 2017; Aravind et al., 2018; Brahimi et al., 2018; Ferentinos, 2018).

In recent years, the anchor frame-based plant disease detection method has achieved remarkable success, and the most representative one is the Faster region-based CNN (R-CNN-based plant disease detection method). Fuentes et al. (2017) first used Faster R-CNN to locate tomato diseases and pests directly, combined with deep feature extractors such as VGG-Net and ResNet, and the mean average precision (mAP) value reached 85.98% in a dataset containing 5,000 tomato diseases and pests of nine categories. Fuentes et al. (2018) and Fuentes et al. (2019) improved Faster R-CNN on the backbone structure and ROI pooling according to the characteristics of plant diseases and pest detection and the mAP reached 92.5%. Ozguven and Adem (2019) proposed a Faster R-CNN structure for automatic detection of beet leaf spot disease by changing the parameters of the CNN model; 155 images were trained and tested. The results show that the overall correct classification rate of this method is 95.48%. Zhou et al. (2019) presented a fast rice disease detection method based on the fusion of FCM-KM and Faster R-CNN. The application results of 3,010 images showed that the detection accuracy and time of rice blast, bacterial blight, and sheath blight were 96.71%/0.65 s, 97.53%/0.82 s, and 98.26%/0.53 s, respectively. The breakthroughs achieved in the existing studies are amazing. However, the faster R-CNN-based method is a two-stage detection method with a large amount of calculation and is time-consuming.

At present, object detection methods based on deep learning emerge endlessly, and many researchers have improved new methods on plant disease detection to predict the location and class of the lesions. Jiang et al. (2019) proposed the INAR-SSD model, and the test on a self-built apple leaf disease dataset achieved a performance of 78.80% mAP with a high detection speed of 23.13 FPS. Sun et al. (2020) presented an instance detection method improved on the basis of single-shot detector (SSD) to detect maize leaf blight under a complex background. The proposed method combined data preprocessing, feature fusion, feature sharing, disease detection, and other steps. The mAP of the new model is higher (from 71.80 to 91.83%) than that of the original SSD model. The FPS of the new model has also improved (from 24 to 28.4), reaching the standard of real-time detection. Bhatt et al. (2019) presented a method to detect pests and diseases on images captured under uncontrolled conditions in tea gardens. YOLOv3 was used to detect pests and diseases. While ensuring real-time availability of the system, about 86% mAP was achieved with 50% IoU. Singh et al. (2020) concluded that although plant disease detection technology has developed rapidly, the methods can be only effectively used for a restricted number of plants.

Object Detection Network Structures

The task of plant disease detection is similar to the method of general object detection, and the task of plant disease detection can be seen as the specific application of a general object. As for the classical two-stage universal object detection method, the first stage is responsible for extracting candidate windows, which are the input of the second stage, and the second stage is responsible for the accurate detection task (Ren et al., 2016). The two-stage structures have many parameters and are time-consuming. For example, when performing ROI pooling operation, the candidate box obtained in the first stage needs to be cut out from a high-dimensional feature map. This operation needs a large amount of calculation and requires the aid of parallel computing devices such as GPU.

The emergence of one-stage universal object detection methods has solved this problem. Famous one-stage detection methods include SSD (Liu et al., 2016) and YOLO (Redmon and Farhadi, 2016, 2018; Redmon et al., 2016). Compared with the traditional convolutional neural network, the SSD selects VGG16 as the trunk of the network and adds a feature pyramid network to obtain features from different layers and make predictions. YOLO considers the detection task as a regression problem and uses global information to directly predict the bounding box and category of the object to achieve end-to-end detection of a single CNN network. YOLO can achieve global optimization and greatly improve the detection speed while satisfying higher accuracy. These methods are equivalent to the previous stage of the two stages, eliminating the time-consuming ROI pooling operation and, thus, have an innate speed advantage. The one-stage object detection algorithm directly adds the detection head to the backbone network for classification and regression, uses the whole image as the input of the network, and directly returns the position of the bounding box and the category to which it belongs at the output layer.

In summary, the main difference between the two networks is that the two-stage network needs to first generate a candidate box (proposal) that may contain the lesions, and then further execute the object detection process. In contrast, the one-stage network directly uses the features extracted in the network to predict the location and class of the lesions. In the field of plant diseases and pest detection which emphasizes detection accuracy at this stage, more models based on the two-stage network are used.

Multiscale Object Detection of Plant Diseases

In actual plant disease detection, multiscale plant disease objects are common in a real natural environment. This study considers how to detect small-scale objects and large-scale objects in the same frame image, such as in an agricultural Internet of Things monitoring scenario, where the proximal plant leaves and the distant plant leaves may have very different scales. In the case of front shooting of the surveillance camera, because some plant disease objects are far away from the camera, the object size is small and it makes the plant disease objects occupy very small pixels in the image. Also, the corresponding area contains less information, which is prone to missed detection, affecting the detection accuracy of the algorithm. Therefore, it is difficult to identify and locate small-scale plant disease targets in the field of target detection.

Because small objects consist of very few pixels and generally only occupy less than 5% of the whole image, it is difficult to extract enough features for CNN. In order to improve the detection performance of small objects, it is usually necessary to combine image super-resolution, large-scale feature map prediction, deep and shallow feature fusion, and other feature enhancement methods (Li D. et al., 2020; Liu and Wang, 2020; Zhao et al., 2020). However, these methods will bring additional parameters and calculation while improving the performance of small object detection, resulting in the reduction of the real-time performance of the algorithm, and it is difficult to deploy to the terminal with a small amount of computation. Therefore, how to improve the performance of small object detection of the algorithm to meet the needs of a real natural environment without introducing too much additional calculation cost to ensure real-time performance is an urgent problem to be solved.

In order to improve the detection ability of multiscale tomato gray mold objects, reduce the missed detection rate, and improve the detection efficiency, MobileNetv2-YOLOv3, which is known for its speed, was selected in this study (Sandler et al., 2018). As a basic detection network, a multiscale parallel tomato gray mold detection algorithm (MP-YOLOv3) with high real-time performance was constructed by combining multiscale pixel feature fusion and efficient channel attention mechanism to enhance the quality of small object features. The contributions of this study are summarized as follows:

(1) A multiscale feature fusion strategy is proposed to fuse feature maps of the skeleton network MobileNetv2 in different scales from high to low, which enhances the small object information carried by feature maps and provides rich semantic information for the prediction layer of the network, thereby effectively improving the small object detection ability of the algorithm.
(2) In order to highlight the useful feature channels and suppress the feature channels with small contribution, we introduce an efficient channel attention module before the detection layer to assign weights to the feature channels according to their importance, which effectively improves the detection performance.
(3) In order to reduce the impact of network complexity on detection speed, a tomato gray mold detection model MP-YOLOv3 with high real time and robustness is proposed, which adopts a parallel processing mode on the architecture and uses a buffer queue between the various functions of object detection to reduce the waiting time in detection.
(4) The model was validated on the self-built tomato gray mold dataset. Compared with existing algorithms, it achieved good results in small-scale tomato gray mold object detection, significantly improved the accuracy of tomato gray mold object detection, reduced the occurrence of missed detection, and could meet the practical application needs.

MATERIALS

Dataset Collection

Since there is no published image database of tomato gray mold disease in a real natural environment, 1,000 images of tomato gray mold pathogen in a natural environment were collected from the Internet of Things monitoring video of tomato greenhouse in Shouguang City, Shandong Province, China. Meanwhile, in order to expand the sample dataset, 263 images of tomato gray mold were obtained by the network crawler method. A total of 1,263 images were collected. The images include conditions on cloudy and sunny days, objects such as branches and leaves forming shadows or shelters on the surface of tomato leaves, etc.

Data Annotation

In image labeling, the minimum outer rectangle of each lesion is labeled with the LabelImg tool[1] to ensure that there is only one tomato gray mold lesion object in each rectangular labeling frame and as few background pixels as possible. After image annotation, 263 images under different weather and light conditions were selected as the test set, and the remaining 1,000 images were used for network training. Details of 263 images selected are shown in **Table 1**.

Data Enhancement

When training deep learning model, the more and comprehensive the training data, the stronger the recognition ability (Theodoridis, 2015). Therefore, to enrich the image training dataset, better extract image features, and avoid overfitting, this study uses a variety of methods to enhance the dataset. Due to uncertain factors such as illumination direction and weather, the illumination conditions during image acquisition are very complex. In order to improve the generalization ability of the training model, the original image is processed by eight methods: brightness enhancement and attenuation, color enhancement and attenuation, contrast enhancement and attenuation, and sharpness enhancement and attenuation (Bloice et al., 2017). After image amplification, the original annotation is still valid. Nine thousand images after image enhancement were used for training and parameter optimization validation of subsequent improved network. One thousand images were randomly selected from 9,000 images as

[1]https://github.com/tzutalin/labelImg

TABLE 1 | Detailed information of samples in test images.

Conditions	Independent leaves	Indistinct leaves	Shaded leaves	Occluded leaves
Number of images	68	62	74	59
Number of annotated lesions	469	457	516	413

the validation set, and the remaining 8,000 images were used as the training set. There was no overlap between the training set and the test set.

METHODS

Considering the real-time requirement of tomato gray mold detection in a real scene, we chose MobileNetv2-YOLOv3 as the basic detection network. However, due to the insufficient accuracy caused by its focus on efficient convolution operation, we combined multiscale feature fusion and efficient channel attention to improve it. This section will give the detailed design methods of the multiscale feature fusion module and the high-efficiency channel attention module and propose the flow of parallel structure.

The overall framework of the multiscale parallel tomato gray mold early detection algorithm is shown in **Figure 1**. The algorithm is divided into two parts. The first part extracts the object features through MobileNetv2, and the second part detects tomato gray mold objects through the object prediction part of YOLOv3. First, the image resolution is adjusted to 416 × 416 and it is inputted into the MobileNetv2 network to extract features, and the fused feature map is obtained by multiscale feature fusion. Second, these feature maps are enhanced by the efficient channel attention module, and the weight of feature channels is assigned according to their importance. Third, through MobileNetv2, a 13 × 13 × 1,024-dimensional tensor is obtained, and through a 1 × 1 convolutional kernel for convolution operation, a S × S × 18-dimensional tensor is obtained. Finally, this tensor is used to predict the location of the tomato gray mold.

Multiscale Feature Fusion Module

In tomato gray mold object detection, due to the distance between the tomato leaf and the camera, the size of the tomato gray mold object presented on the image is also different. The size of the last layer of the feature layer is only 13 × 13, which is 1/32 of the original input image, which makes the feature layer lose some feature information of smaller objects. In deep neural networks, the higher the layers, the smaller the size of the feature map, and the richer the semantic information contained. The lower feature layer has greater resolution and retains more details in the original image, which is conducive to determining the location of the object. To simultaneously utilize the detailed information in the shallow feature map extracted from the MobileNetv2 skeleton network and the semantic information in the deep feature map, a multiscale feature fusion module is proposed in this study. The specific approach is to improve the network's ability to detect small-size tomato gray mold objects by fusing high-level features with low-level features and predicting them on multiple-scale feature maps. The feature map after up-sampling is combined with the feature map with the size of 26 × 26 in the convolution process as the basis for the second prediction. Then the feature map with the size of 52 × 52 and 104 × 104 is obtained in this way for the third and fourth prediction, respectively, as shown in **Figure 1**.

Efficient Channel Attention Module

When generating fusion feature maps, our algorithm uses the method of channel splicing between the feature maps obtained from the up-sampling and the feature maps extracted from the skeleton network, resulting in large differences in the information carried by each channel fusion feature map. The importance of different channels of fusion features is different for the scale detection, so it is necessary to introduce efficient channel attention module after multiscale merging, so that the model can learn the importance of different channel features. In order to selectively highlight the effective feature channels and suppress the feature channels with small contribution, we detect the feature channels with different scales by assigning weight to the feature channels according to the importance degree through the efficient channel attention module, which effectively improves the detection performance. The principle is shown in **Figure 2**.

As the spatial characteristics of different channels should have a certain correlation, if a channel has a high correlation with its adjacent channels, it means that the feature contains more subject features. We adopt an efficient channel attention module that can learn the feature correlation between channels. In the efficient channel attention module, in order to highlight the feature correlation between channels, the feature channel dimension is compressed. The original feature channel $H \times W \times C$ is transformed to $1 \times 1 \times C$ by global pooling, and global features in channel dimension are obtained. One-dimensional convolution with convolution kernel size k is used to extract and integrate information between each channel and its k neighborhood channels to obtain correlation parameters L_i between channels.

$$L_i = \sum_{j=1}^{k} \alpha^j C_i^j, \; C_i^j \in \varphi_i^k \tag{1}$$

In the abovementioned formula, α^j represents one-dimensional convolution kernel parameters, and φ_i^k represents k neighborhood channels of feature channel C_i. The larger the L_i, the stronger the correlation between feature channels C_i and φ_i^k, that is, the more useful information C_i contains. To make the network focus on the feature channels with more useful information, ω_i is denoted as the weight of the feature channel, that is, the activation value of each channel obtained by L_i passing through sigmoid function.

$$\omega_i = \sigma(L_i) \tag{2}$$

In the abovementioned formula, σ represents sigmoid activation function. The weighted output feature channel is obtained by multiplying the weight with the original channel feature value. The weighted output feature channel is conducive to highlighting the key features of the object and weakening the nonimportant features.

As shown in **Figure 1**, the efficient channel attention module was added to the proposed model. Feature maps

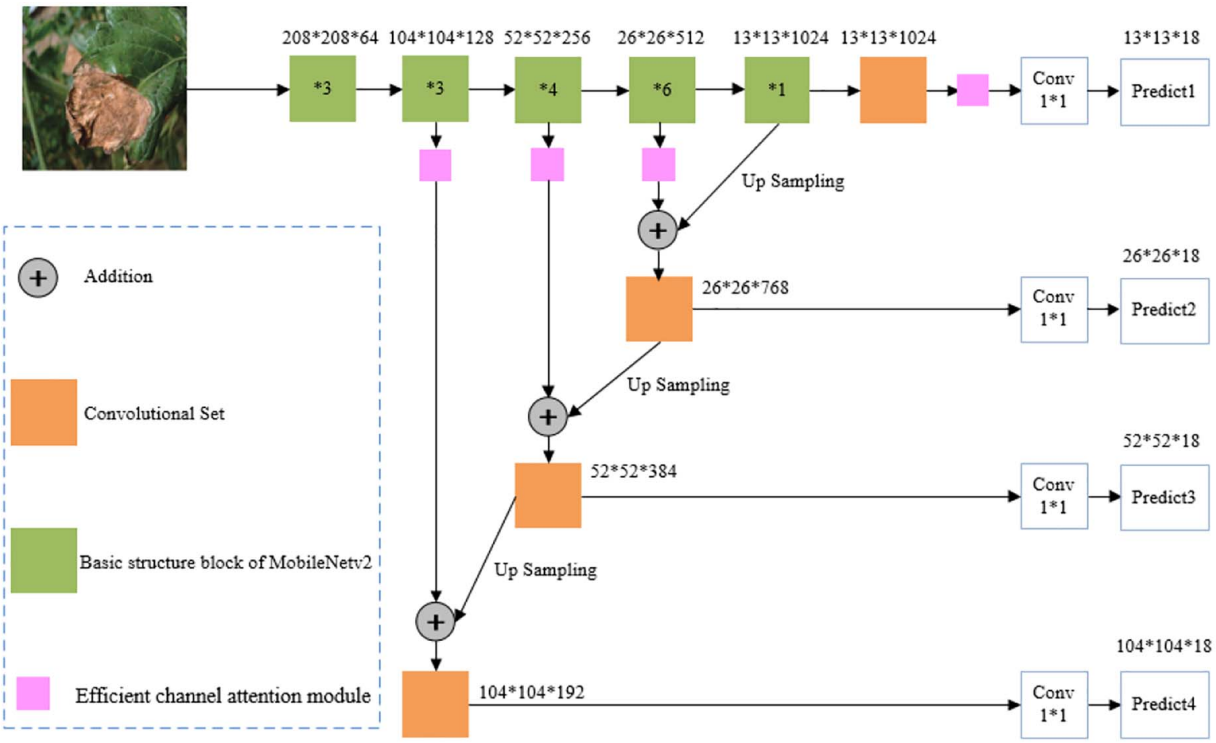

FIGURE 1 | Overall framework of a multiscale parallel algorithm for the early detection of tomato gray mold in a complex natural environment.

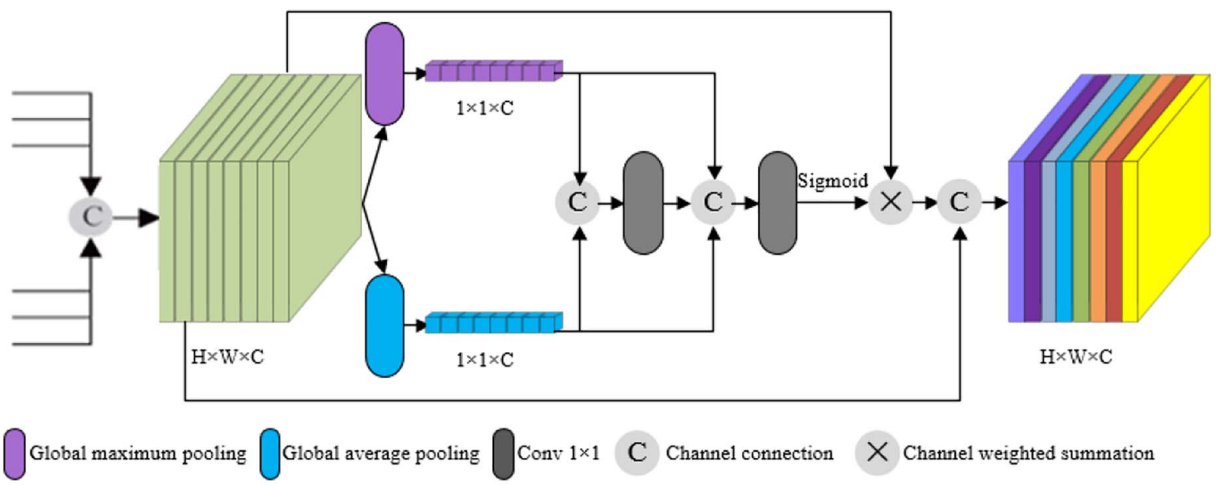

FIGURE 2 | Efficient channel attention module.

close to the proposed model's prediction layers have a higher predictive impact, taking into account that more and ineffective computation would be added if efficient channel attention modules were used for all convolutional layers. Therefore, the convolution and addition layers preceding the prediction layers in the model served as inputs to the efficient channel attention module. Using the efficient channel attention module, high weights are assigned to tomato gray mold disease features in the convolutional feature map and low weights to the natural background. The final output prediction layer assigns more weight to the image information of interest, and the information in each channel will contain more accurate and more information on tomato gray mold disease characteristics. Thus, the detection rate is effectively improved, and the small object lesions which are confused by local occlusion and natural background can be easily missed.

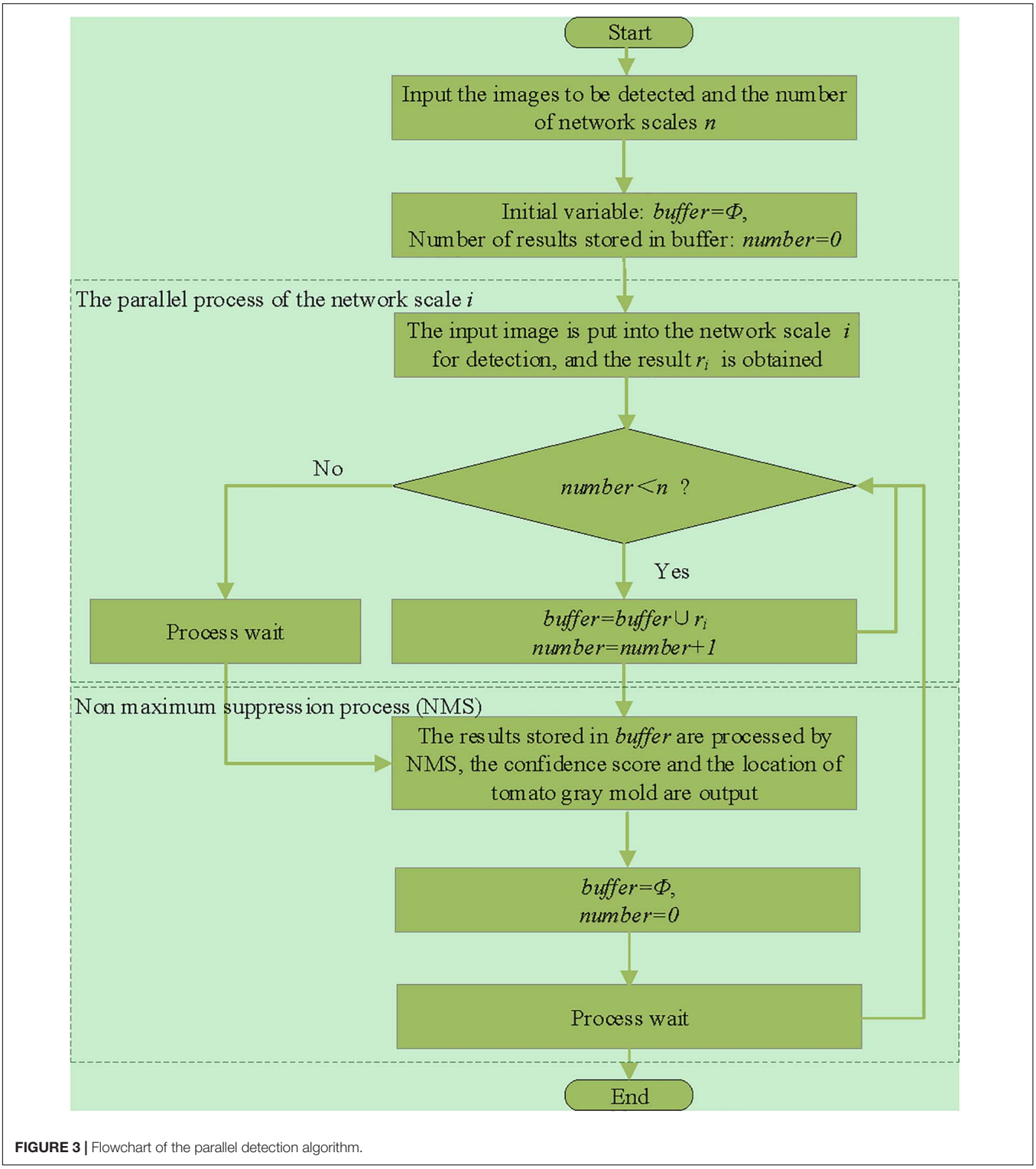

FIGURE 3 | Flowchart of the parallel detection algorithm.

Parallel Network Structure

A serial multiscale network will increase the time overhead. To ensure the real-time performance of detection, the improved model uses parallel detection algorithm. The flowchart of the algorithm is shown in **Figure 3**. The multiscale network changes from serial to parallel operation, and the time overhead is reduced from the time of detection of multiple-scale networks to the time of detection of a single-scale network. It ensures that the time overhead of the model will not increase substantially while improving the detection performance.

Superiority of the Proposed Algorithm MP-YOLOv3

To present the superiority of the proposed algorithm MP-YOLOv3, the algorithm is compared with Tiny-YOLOv3, MobileNetv2-YOLOv3, MobileNetv2-SSD, and Faster R-CNN regarding their advantages and disadvantages of their specific features. The comparison is shown in **Table 2**.

It can be seen from **Table 2** that the proposed algorithms are improved by combining multiscale feature fusion and efficient channel attention on the basis of MobileNetv2-YOLOv3. With the addition of very few parameters and little impact on the speed, a high-precision real-time tomato gray mold detection algorithm is constructed.

EXPERIMENTAL DESIGN

Experimental Operation Environment

In this experiment, under Ubuntu 16.04 operating system, Caffe deep learning framework was built on i7-7700HQCPU (16 GB memory) and NVIDIA GTX 1070 GPU (8 GB memory) hardware platform, and Python language programming was used to realize the training and testing of the tomato gray mold object detection network model.

Model Training

In this study, a stochastic gradient descent method was used to train the network in an end-to-end joint manner. In order to improve the training efficiency, during network training, the network parameters were initially initialized with the pretraining model on ImageNet (Deng et al., 2009), so that good initial values of the model were achieved to reduce the training time cost and speed up model convergence. The initial learning rate was set to 0.001, the weight attenuation rate was set to 0.0005, the momentum factor was set to 0.9, and the verification period was set to 5,000, that is, the network tests the accuracy of the training model on the verification set every 5,000 iterations, and the training was stopped when the model accuracy rate reaches convergence. The maximum number of iterations was set to 50,000. Training ended when the loss dropped to around 1. The training model was saved after the training, and the model was verified with the test set. The final output of the network was the identified object and its probability of being the object of tomato gray mold, and the result only retained the region with a probability value greater than 0.8.

TABLE 2 | Comparison of the characteristics of different algorithms.

Algorithms	Network structure	Backbone network	Feature extraction capability
The proposed method	One-stage	MobileNetv2	Multiscale feature extraction
Tiny-YOLOv3	One-stage	7-layer conv + Max	The ability of deep feature extraction is poor.
MobileNetv2-YOLOv3	One-stage	MobileNetv2	The ability of deep feature extraction is poor.
MobileNetv2-SSD	One-stage	MobileNetv2	The ability of deep feature extraction is poor.
Faster R-CNN	Two-stage	ResNet	The accuracy outperforms some of the one-stage detectors, but the speed is low.

Evaluating Indicator

In this study, both recall rate and accuracy rate should be considered in the process of object detection of tomato gray mold, so F1 value and AP value were used to evaluate the recognition results.

$$F1 = \frac{2PR}{P+R} \tag{3}$$

$$P = \frac{TP}{TP+FP} \tag{4}$$

$$R = \frac{TP}{TP+FN} \tag{5}$$

$$AP = \frac{\sum P}{N} \tag{6}$$

In the abovementioned formula, P represents precision, R represents recall rate, TP (true positive) represents the number of tomato gray mold objects that the algorithm can accurately detect, FP (false positive) indicates the number of background misidentified as tomato gray mold objects, FN (false negative) indicates the number of unrecognized tomato gray mold objects, and N is the total number of images.

EXPERIMENTAL RESULTS AND ANALYSIS

Detection Results of Tomato Gray Mold

In order to verify the performance of MP-YOLOv3 proposed in this study, the identification results of the network on 263 test sets were further analyzed. There were 1,855 tomato gray mold objects in 263 test sets. The number of objects identified by this method was 1,689, of which 1,578 were tomato gray mold objects. The recall rate, accuracy, and misidentification rate of this method were 85.07, 93.43, and 6.57%, respectively. Examples of the identification results and specific identification results of this research method are shown in **Figure 4** and **Table 3**, respectively.

From **Table 3** and **Figure 4**, it can be seen that the detection effect is the best in the sparse independent leaf scenario with a recall rate of 96.59%. The blurred indistinct tomato gray mold object can be accurately identified. This method can correctly identify 86.65% of the indistinct objects in the image. This method can effectively identify the tomato gray mold object

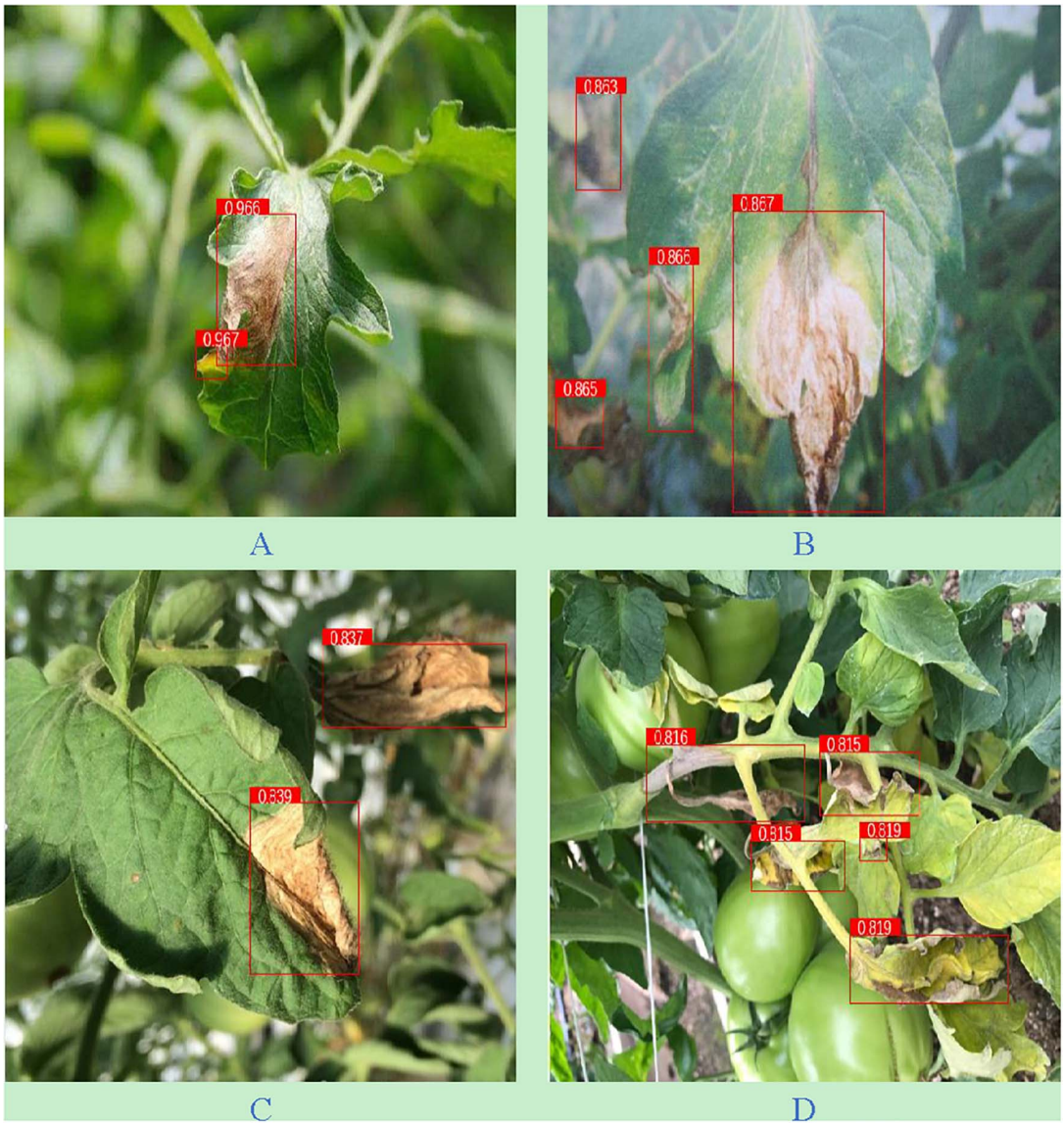

FIGURE 4 | The effect diagram of the detection method in this paper. **(A)** Independent leaves; **(B)** indistinct leaves; **(C)** shaded leaves; **(D)** occluded leaves.

TABLE 3 | Detailed detection results of tomato gray mold.

Conditions	Independent leaves	Indistinct leaves	Shaded leaves	Occluded leaves	Total
Number of objects correctly identified	453	396	432	337	1,578
Number of annotated lesions	469	457	516	413	1,855
Recall rate	96.59%	86.65%	83.72%	81.60%	85.07%

with a shaded surface, and the recognition recall rate is 83.72%. In addition, this method is also applicable when the leaves are occluded, and its recognition recall rate is 81.60%. According to **Figure 4**, MP-YOLOv3 also has a high detection accuracy for small-size lesions, which proves that multiscale detection has a good detection effect for objects with different sizes. Through the above analysis, it can be concluded that despite the phenomenon of misidentification and missed identification, the method in this study can accurately detect the tomato gray mold object in the image.

Analysis of Ablation Experiments

To examine the detection effects of the two improvement modules in MP-YOLOv3, benchmarked on the previous network, the multiscale feature fusion module and the efficient channel attention module were considered as two experimental variables

using the dataset built from this study to conduct a series of ablation experiments. Among them, "√" means joining the module and " × " means not joining it. The results are shown in **Table 4**.

As can be seen from **Table 4**, the two improved modules proposed in this study could both enhance the detection efficacy of tomato gray mold, and the combination of the two modules had the best detection efficacy, which verified the rationality of the model designed in this study. The improvement of the multiscale feature fusion module enriches the feature information of small objects in feature maps. The improvement of the efficient channel attention module enables the feature maps of network output to more efficiently characterize objects.

TABLE 4 | Detection effects of ablation experiments.

Strategies	Multiscale feature fusion module	Efficient channel attention module	F1 score/%	Average precision/%
1	×	×	87.9	85.3
2	√	×	91.7	89.6
3	×	√	93.2	91.1
4	√	√	95.6	93.4

Comparison of Different Detection Methods

To verify the detection performance of MP-YOLOv3, the algorithm is compared with Tiny-YOLOv3, MobileNetv2-YOLOv3, MobileNetv2-SSD, Faster R-CNN, and other algorithms. The experimental results are shown in **Table 5**.

It can be seen from **Table 5** that the detection effect of the algorithm in this study is the best among the compared advanced algorithms, and the detection average precision reaches 93.4%, which is 8.1% higher than the AP of MobileNetv2-YOLOv3. By improving the feature extraction network of MobileNetv2-YOLOv3, the number of layers of the network is deepened and the extracted features are more detailed, which ensures that the model improves the detection accuracy at the same time. The number of model parameters did not increase, the size of the model was only 16.9 MB, and the detection time of each image was 0.022 s, which achieved a good detection effect.

TABLE 5 | Comparison of detection results using different algorithms.

Algorithms	F1 score/%	Average precision/%	Single image detection time/second (s)
The proposed method	95.6	93.4	0.022
Tiny-YOLOv3	86.8	84.1	0.023
MobileNetv2-YOLOv3	87.9	85.3	0.022
MobileNetv2-SSD	88.5	86.6	0.035
Faster R-CNN	89.9	87.8	0.126

CONCLUSION AND FUTURE DIRECTIONS

Conclusions

(1) The method proposed in this study can identify the tomato gray mold object from images with complex background, and it is expected to be applied in tomato growth information monitoring and tomato disease automated inspection. Compared with the traditional method of disease detection, this method is more challenging. On the one hand, the object of tomato gray mold at the early stage of growth is smaller; on the other hand, the object of tomato gray mold at this time is very similar to the background color, resulting in the use of traditional methods which cannot effectively and accurately identify the lesions in the image. Deep learning theory makes early detection of tomato gray mold possible. It can automatically extract image features and is an effective detection method.

(2) The proposed method uses the multiscale feature fusion module and the efficient channel attention module to fuse the features of different scales and effectively solve the problem of insufficient semantic information of low-level features and improve the detection effect of the model on multiscale tomato gray mold objects. The experimental results show that the proposed algorithm has certain advantages over other existing algorithms and solves the problems of multiscale change, occlusion, and poor detection of small-size objects, which can improve the accuracy of object detection while ensuring a small amount of calculation.

(3) The model has excellent performance in practical application and can adapt to a complex natural environment. It lays a research foundation for subsequent disease object positioning and spraying pesticides on demand, reduces the use of chemical pesticides, and has important significance for protecting farmland ecology.

Future Directions

Crop disease detection is one of the key problems to solve automation in agricultural fields, and object detection is also one of the most difficult tasks of computer vision for a long time. In this study, the task of tomato gray mold object detection was studied, innovative algorithms were proposed, and some progress was made. However, there are still some problems that deserve further study.

(1) In this study, an early detection model of tomato gray mold disease was proposed, and future work needs to further solve the problem of missed detection of under extreme shooting angle to achieve accurate early diagnosis of tomato gray mold disease at different parts under different shooting conditions.

(2) The current work should be transplanted to mobile terminals, such as smartphones, tablet PC, etc., to improve practicability and increase the modern atmosphere. Later in the practical application test, a large number of data will be used to continuously improve the practicability and accuracy of tomato gray mold detection.
(3) Various environmental parameters of greenhouse tomato crops will be collected in real time by the Internet of Things technology, and a tomato gray mold early-warning model will be constructed. Tomato gray mold will be early-warned by analyzing the real-time collected data.
(4) Although this study can achieve excellent detection and recognition results, data-driven deep learning technology requires a large number of samples to support, and it is difficult to obtain sufficient sample size in the field of plant disease monitoring with a wide variety of species. In the future, we will solve this problem from the aspect of small sample disease detection.
(5) Additional care would be necessary with poorly supervised learning when applied to automatic pest detection in our area due to the high cost to deal with labeling work.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JL and XW conducted the experiments and data analysis and wrote the manuscript. XW revised the manuscript. Both authors read and approved the manuscript.

FUNDING

This study was supported by the Facility Horticulture Laboratory of Universities in Shandong with project numbers 2019YY003, 2018YY016, 2018YY043, and 2018YY044, School Level High-Level Talents Project 2018RC002, Youth Fund Project of Philosophy and Social Sciences of Weifang College of Science and Technology with project numbers 2018WKRQZ008 and 2018WKRQZ008-3, Key Research and Development Plan of Shandong Province with project numbers 2020RKA07036, 2019RKA07012, and 2019GNC106034, Research and Development Plan of Applied Technology in Shouguang with project number 2018JH12, 2018 Innovation Fund of Science and Technology Development Centre of the China Ministry of Education with project number 2018A02013, 2019 Basic Capacity Construction Project of Private Colleges and Universities in Shandong Province, Weifang Science and Technology Development Programme with project numbers 2019GX071, 2019GX081, and 2019GX082, Special Project of Ideological and Political Education of Weifang University of Science and Technology (W19SZ70Z01), and Shandong social science planning project (21CPYJ20).

REFERENCES

Amara, J., Bouaziz, B., and Algergawy, A. (2017). "A deep learning based approach for banana leaf diseases classification," in *Lecture Notes in Informatics (LNI) Datenbanksysteme für Business, Technologie und Web (BTW 2017) – Workshopband*, eds B. Mitschang, D. Nicklas, F. Leymann, H. Schöning, M. Herschel, and J. Teubner (Bonn: Gesellschaft Für Informatik), 79–88.

Aravind, K. R., Raja, P., and Aniirudh, R. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Comput. Sci.* 133, 1040–1047. doi: 10.1016/j.procs.2018.07.070

Barbedo, J. G. A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. *Springerplus* 2:660.

Bhatt, P. V., Sarangi, S., and Pappula, S. (2019). "Detection of diseases and pests on images captured in uncontrolled conditions from tea plantations," in *Proceedings of the SPIE, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV, 1100808*, Vol. 11008, (Bellingham, WAS: SPIE). doi: 10.1117/12.2518868

Bloice, M. D., Stocker, C., and Holzinger, A. (2017). Augmentor: an image augmentation library for machine learning. *J. Open Source Softw.* 2:19.

Borges, ÁV., Saraiva, R. M., and Maffia, L. A. (2015). Biocontrol of gray mold in tomato plants by clonostachys rosea. *Trop. Plant Pathol.* 40, 71–76. doi: 10.1007/s40858-015-0010-3

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., and Moussaoui, A. (2018). "Deep learning for plant diseases: detection and saliency map visualisation," in *Human and Machine Learning*, eds J. Zhou and F. Chen (Cham: Springer International Publishing), 93–117. doi: 10.1007/978-3-319-90403-0_6

Chen, M., Tang, Y., Zou, X., Huang, K., Huang, Z., Zhou, H., et al. (2020). Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology. *Comput. Electron. Agric.* 174:105508. doi: 10.1016/j.compag.2020.105508

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). "Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Computer Vision and Pattern Recognition*, (Miami, FL), 248–255.

Durmus, H., Gunes, E. O., and Kirci, M. (2017). "Disease detection on the leaves of the tomato plants by using deep learning," in *Proceedings of the 6th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2017*, (Fairfax, VA).

Elad, Y., and Shtienberg, D. (1995). Botrytis cinerea in greenhouse vegetables: chemical, cultural, physiological and biological controls and their integration. *Integr. Pest Manag. Rev.* 1, 15–29. doi: 10.1007/bf00140331

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009

Fuentes, A. F., Yoon, S., and Park, D. S. (2019). Deep learning-based phenotyping system with glocal description of plant anomalies and symptoms. *Front. Plant Sci.* 10:1321. doi: 10.3389/fpls.2019.01321

Fuentes, A. F., Yoon, S., Lee, J., and Park, D. S. (2018). High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9:1162. doi: 10.3389/fpls.2018.01162

Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17:2022. doi: 10.3390/s17092022

Jiang, P., Chen, Y., Liu, B., He, D., and Liang, C. (2019). Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access* 7, 59069–59080. doi: 10.1109/ACCESS.2019.2914929

Li, D., Wang, R., Xie, C., Liu, L., and Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors* 20:578. doi: 10.3390/s20030578

Li, J., Tang, Y., Zou, X., Lin, G., and Wang, H. (2020). Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots. *IEEE Access* 8, 117746–117758. doi: 10.1109/ACCESS.2020.3005386

Lin, G., Tang, Y., Zou, X., Cheng, J., and Xiong, J. (2020). Fruit detection in natural environment using partial shape matching and probabilistic Hough transform. *Precis. Agric.* 21, 160–177. doi: 10.1007/s11119-019-09662-w

Liu, J., and Wang, X. (2020). Early recognition of tomato gray leaf spot disease based on mobilenetv2-yolov3 model. *Plant Methods* 16:83.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016). "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision—ECCV*, (Amsterdam), 21–37.

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., et al. (2015). Advanced methods of plant disease detection. A review. *Agron. Sustain. Dev.* 35, 1–25.

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls.2016.01419

Nigam, S., and Jain, R. (2020). Plant disease identification using deep learning: A review. *Indian J. Agric. Sci.* 90, 249–257.

O'Neill, T. M., Shtienberg, D., and Elad, Y. (1997). Effect of some host and microclimate factors on infection of tomato stems by botrytis cinerea. *Plant Dis.* 81, 36–40. doi: 10.1094/pdis.1997.81.1.36

Ozguven, M. M., and Adem, K. (2019). Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms. *Phys. A Statal Mech. Appl.* 535:122537. doi: 10.1016/j.physa.2019.122537

Redmon, J., and Farhadi, A. (2016). YOLO9000: better, faster, stronger. *arXiv* [Preprint] arXiv:1612.08242,

Redmon, J., and Farhadi, A. (2018). YOLOv3: an incremental improvement. *arXiv* [Preprint] 1–6. arXiv:1804.02767,

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, NV), 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/tpami.2016.2577031

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Salt Lake City, UT: IEEE), 4510–4520.

Sankaran, S., Mishra, A., Ehsani, R., and Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric.* 72, 1–13. doi: 10.1016/j.compag.2010.02.007

Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., and Nagasubramanian, K. (2020). Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* 26, 53–69. doi: 10.1016/j.tplants.2020.07.010

Sun, J., Yang, Y., He, X., and Wu, X. (2020). Northern maize leaf blight detection under complex field environment based on deep learning. *IEEE Access* 8, 33679–33688. doi: 10.1109/ACCESS.2020.2973658

Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., and Zou, X. (2020). Recognition and localization methods for vision-based fruit picking robots: a review. *Front. Plant Sci.* 11:510. doi: 10.3389/fpls.2020.00510

Theodoridis, S. (2015). "Neural networks and deep learning," in *Machine Learning*, ed. S. Theodoridis (Athens: National and Kapodistrian University of Athens), 875–936.

Wang, G., Sun, Y., and Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Comput. Intell. Neurosci.* 2017, 1–8. doi: 10.1155/2017/2917536

Zhao, Y., Liu, L., Xie, C., Wang, R., Wang, F., Bu, Y., et al. (2020). An effective automatic system deployed in agricultural internet of things using multi-context fusion network towards crop disease recognition in the wild. *Appl. Soft Comput.* 89:106128. doi: 10.1016/j.asoc.2020.106128

Zhou, G., Zhang, W., Chen, A., He, M., and Ma, X. (2019). Rapid detection of rice disease based on FCM-KM and faster R-CNN fusion. *IEEE Access* 7, 143190–143206. doi: 10.1109/ACCESS.2019.2943454

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

METHODS
published: 10 June 2021
doi: 10.3389/fpls.2021.575751

Leaf Counting: Fusing Network Components for Improved Accuracy

Guy Farjon, Yotam Itzhaky, Faina Khoroshevsky and Aharon Bar-Hillel*

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er Sheva, Israel

Leaf counting in potted plants is an important building block for estimating their health status and growth rate and has obtained increasing attention from the visual phenotyping community in recent years. Two novel deep learning approaches for visual leaf counting tasks are proposed, evaluated, and compared in this study. The first method performs counting *via* direct regression but using multiple image representation resolutions to attend leaves of multiple scales. The leaf count from multiple resolutions is fused using a novel technique to get the final count. The second method is detection with a regression model that counts the leaves after locating leaf center points and aggregating them. The algorithms are evaluated on the Leaf Counting Challenge (LCC) dataset of the Computer Vision Problems in Plant Phenotyping (CVPPP) conference 2017, and a new larger dataset of banana leaves. Experimental results show that both methods outperform previous CVPPP LCC challenge winners, based on the challenge evaluation metrics, and place this study as the state of the art in leaf counting. The detection with regression method is found to be preferable for larger datasets when the center-dot annotation is available, and it also enables leaf center localization with a 0.94 average precision. When such annotations are not available, the multiple scale regression model is a good option.

Keywords: image-based plant phenotyping, leaf counting, counting with convolutional neural networks, fusing network components for counting, growth rate estimation

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Lucas Costa, University of Florida, United States
Massimo Minervini, IMT School for Advanced Studies Lucca, Italy

****Correspondence:***
Guy Farjon
guyfar@post.bgu.ac.il

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *24 June 2020*
Accepted: *10 May 2021*
Published: *10 June 2021*

Citation:
Farjon G, Itzhaky Y, Khoroshevsky F and Bar-Hillel A (2021) Leaf Counting: Fusing Network Components for Improved Accuracy. Front. Plant Sci. 12:575751. doi: 10.3389/fpls.2021.575751

1. INTRODUCTION

Object counting is important for a variety of tasks in the agriculture and phenotyping domains. Estimating the number of fruits, flowers, and trees helps farmers make better decisions on cultivation practices, plant disease prevention, and the size of the harvest labor force (Rahnemoonfar and Sheppard, 2017). For example, estimating the number of flowers in a field can be used to predict harvest dates from peak flowering, track the response of flowering to environmental conditions, and evaluate the effects of cultural practices (Adamsen et al., 2000). The chemical thinning process of apple trees starts by counting the number of flowers during the blossom period, for flowering level and peak day estimation (Farjon et al., 2019). Determination of wheat grains quality is done by counting wheat seedlings in the field and estimating their density (Liu et al., 2016).

In this study, we focus on the task of leaf counting. The number of leaves a plant has is one of the visual key traits (phenotype) describing its development and growth (Dobrescu et al., 2017). It enables growth rate estimation and is related to the health status of the plant and its yield potential (Telfer et al., 1997; Walter and Schurr, 1999). Manually measuring the traits of the visual plant is a slow, tedious, and expensive process (Jiang et al., 2019), and it usually requires the presence of specialized investigators (Giuffrida et al., 2016). Hence, these traits are measured on

a small random sample of plants, which might lead to a measurement bias (Aich and Stavness, 2017). An accurate automated leaf counting system will enable faster measurements that will decrease costs, and improve the overall accuracy of per-plant estimations (Aich and Stavness, 2017). Due to the importance of this task, a Leaf Counting Challenge (LCC) was initiated in recent years in the Computer Vision Problems in Plant Phenotyping (CVPPP) community and public benchmarks (Scharr et al., 2014; Bell and Dee, 2016; Minervini et al., 2016) are available.

This study presents two network architectures for leaf counting, one based on direct regression and the other on combining detection with regression. For direct regression, we suggest improving upon existing art by explicitly employing fusion over multiple-scale analysis. The number of leaves is regressed from multiple image resolutions, created by a Feature Pyramid Network (FPN) (Xie et al., 2015), thus accounting for both small and large leaves. For each scale, a network is trained to provide both an estimation of the leaves count and the variance of the estimation. The multiple estimates obtained from the different image resolutions are then fused based on their estimated variance. This method is termed Multiple-Scale Regression (MSR). We examine several techniques for fusing the estimators, including min-variance selection and a Maximum Likelihood Estimation (MLE) solution.

The second proposed method treats counting as a two-stage process, where leaf detection is done prior to count regression, yet both are done in a single end-to-end network. Following the density estimation literature (Xie et al., 2015, 2018), we develop an algorithm employing "leaf center" point annotations during training. At several stages, the detection network is regressing a "heat map," a map with Gaussians of pre-defined parameters placed upon the annotated leaf centers (Lempitsky and Zisserman, 2010). The heat map is gradually refined to obtain more exact leaf positions, thus allowing better separation between small adjacent leaves. An initial count estimate is obtained by applying Non-Maxima Suppression (NMS) and global summing layers. This detection-based estimate is fused with additional features to provide the final count estimate *via* direct regression. This method is termed Detection with Regression Network (DRN). The training of the network is driven by two losses accounting for the detection and count regression accuracy.

The suggested counting methods were tested on the publicly available LCC datasets containing tobacco and *Arabidopsis* plants (Bell and Dee, 2016; Minervini et al., 2016), as well as on a larger dataset of images of the banana plants. Examples of images from the datasets can be seen in **Figure 1**. Both methods obtained improvement over the current state-of-the-art on the LCC datasets, evident by the results obtained on a hidden test set held by the dataset owners. While for small datasets, the methods are comparable in performance, and DRN is preferable when a sufficient sample size is available. This method, which also provides detection of the individual location of the leaves, is hence currently the leading leaf counting method to the best of our knowledge.

The main contributions of this study are:

- It shows that merging ideas from detection and direct regression methods in a single network provides improved accuracy (compared to any other current method) in the leaf counting task. Specifically, a detection refinement mechanism is proposed based on repeated heat map regression with decreasing Gaussian kernel size.
- It shows that for direct regression, using estimators from multiple image scales is beneficial, but that it requires careful fusion of the estimators based on additional variance estimation.
- The methods suggested provide a new state-of-the-art in the leaf counting from RGB images task, as measured on the LCC challenge.

The entire code is freely accessible at https://github.com/farjon/Leaf-Counting. This work is an extension of a previously presented conference paper (Itzhaky et al., 2018), which was supplemented in several important ways. First, the DRN was improved by employing successive refinement of the heat maps produced with Gaussian kernels of decreasing size. Second and more important, careful experiments were conducted to understand the role of data size on the LCC datasets and using a novel larger dataset with 1,016 images of banana plants. These experiments reveal that the DRN method is superior to the MSR for a large enough dataset, a fact which was not clear enough in the conference paper version. Finally, a comprehensive comparison to existing art is provided, showing that the suggested algorithms compare favorably with the best current methods.

2. BACKGROUND

With the growing need for systematic plant phenotyping (Großkinsky et al., 2015) and the development of recent Convolutional Neural Network (CNN)-based techniques (Ren et al., 2015; He et al., 2016), visual leaf counting has attracted considerable attention (Giuffrida et al., 2016; Dobrescu et al., 2017; Lu et al., 2017; Teimouri et al., 2018; Jiang et al., 2019; Kuznichov et al., 2019). A basic but appealing idea is to perform counting by using some standard detection or segmentation network architecture (Ren et al., 2015; Redmon et al., 2016; He et al., 2017; Lin et al., 2017b) to detect the leaves. In this way, one also obtains information additional to counting, the exact image locations of the leaves. This information maybe valuable for applications requiring further processing, like detecting diseases or examining the morphological structure of the leaf. In Romera-Paredes and Torr (2016) a segmentation-based approach was developed and tested on the LCC dataset. First, the image goes through a Fully Convolutional Network (FCN) providing a high-dimensional feature map representation. This map is then used as input for a Recurrent Neural Network (RNN) model in which each layer application segments a different leaf. The leaf count is the number of segmented objects. The study of Ren and Zemel (2017) uses the same techniques as in Romera-Paredes

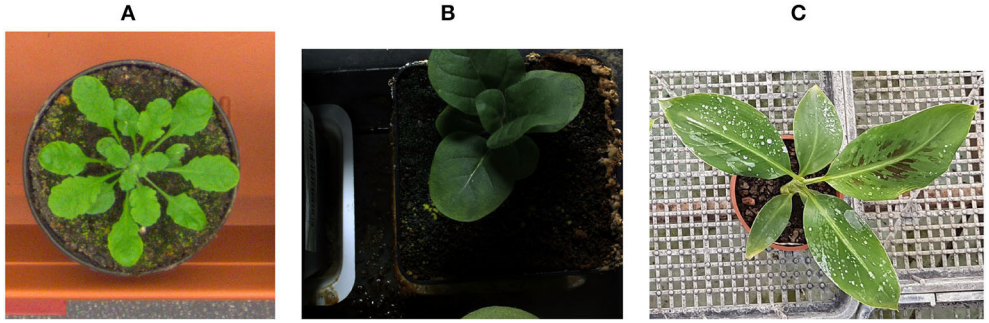

FIGURE 1 | Data example of three different plants used in the experiments. **(A)** Rosette plant, **(B)** Tobacco plant, **(C)** Banana plant.

and Torr (2016), but in addition, offers the ability of instance segmentation. A recent segmentation and counting network was presented in Kuznichov et al. (2019), where new algorithms were suggested for the generation of synthetic plant images. These images are then used as data augmentation for network training. The resulting network provides the state-of-the-art results in the Leaf Segmentation Challenge (LSC),[1] which is a segmentation plus counting benchmark on the same datasets (Scharr et al., 2014; Bell and Dee, 2016; Minervini et al., 2016) as the LCC. However, such segmentation-based methods have a significant drawback for counting since only successfully segmented leaves are counted, and their results in the counting evaluation metric are less competitive.

More direct approaches to counting use annotations requiring less effort, like dot annotations, marking only the object center (Arteta et al., 2016), or even just image-level counts (Dobrescu et al., 2017). When dot annotations are available, counting is often approached *via* a density estimation approach (Xie et al., 2015, 2018). These methods are successful for approximate counting of large object quantities (hundreds or thousands, like crowds in stadiums, cars in large parking lots, etc.), but are less suited for accurate counting of small quantities as required in the leaf counting task. If only image-level counts are available for training, counting tasks can be addressed by a direct regression approach, where a regressor is learning a direct function from image to count estimation (Dobrescu et al., 2017; Lu et al., 2017). The natural regressor choice is an adapted deep CNN. The advantages of this approach are the simplicity of the algorithm, hence the ability to train it from a smaller sample size, and the alleviation of the annotation burden. The currently leading results reported on the leaf counting CVPPP challenge datasets were obtained with a direct regression approach (Dobrescu et al., 2017; Teimouri et al., 2018).

Considering such direct approaches, Giuffrida et al. (2016) proposed a flat (non-deep) learning-based approach for leaf counting in rosette plants. They used a supervised regression model, applied to image-based descriptors learned in an unsupervised manner. Teimouri et al. (2018) treat leaf counting as a classification task, where each possible count results in a limited range is a different label. They use an ensemble of 20 similar models based on Inception-v3 architecture (Szegedy et al., 2016), pre-trained on ImageNet (Deng et al., 2009), and fine-tuned on the leaf counting datasets. They report good results but their evaluation is partial, with 168 images randomly sampled from the five LCC datasets (A1–A5). Therefore, the results cannot be fully compared to LCC bench-marked results. In Aich and Stavness (2017) the authors treated the counting task as a two-staged task. First, they used SegNet (Badrinarayanan et al., 2017) to isolate the plant from the background. Then, they used the segmented mask and the RGB image as the input to a VGG-16 based regressor. Though Aich and Stavness (2017) reports better results than Giuffrida et al. (2016), they are inferior to those of Dobrescu et al. (2017) in which a regression model based on the ResNet-50 architecture (He et al., 2016) was suggested. Trained on multiple leaf datasets and using data augmentation, they obtained the winning results in the LCC event of 2017, and the best LCC benchmark results prior to this study. The model uses only the leaf count annotations while training. Giuffrida et al. (2018) shows identical results on the LCC 2017 dataset as in Dobrescu et al. (2017), but by using multi-modal data (including near infrared channel and a fluorescence channel, in addition to RGB images), the authors showed state-of-the-art results on Cruz et al. (2016) dataset. However, when multi-modal data is not available, the methods proposed in this study are with superior accuracy.

3. METHODS

3.1. Image Datasets

Two datasets were used in the experiments:

- Leaf Counting Challenge data: The challenge includes four different datasets termed A1–A4. Datasets A1, A2, and A4 contain images of the *Arabidopsis* plant and the A3 dataset contains images of young tobacco plants. Datasets A1, A2, and A3 are relatively small containing 128, 31, and 27 images, respectively, while A4 is larger with 624 images. Images were taken from a top viewpoint, using a camera positioned ∼1 m above the plants, under controlled illumination conditions.

[1]https://competitions.codalab.org/competitions/18405.

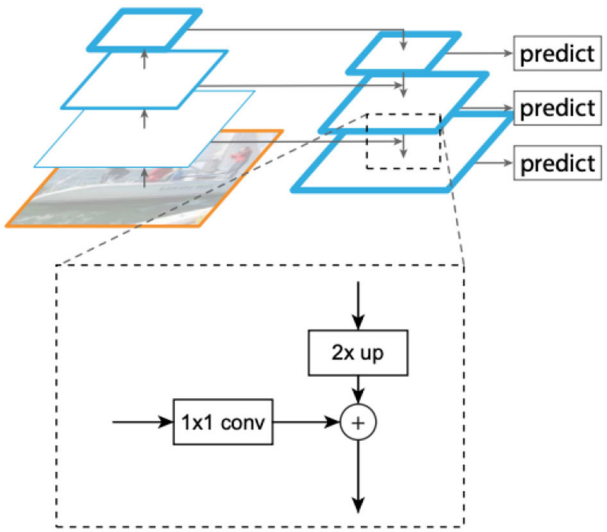

FIGURE 2 | Creating representations at multiple resolutions from a standard Convolutional Neural Network (CNN) (Lin et al., 2017a). Each such tensor (termed $P_2 - P_7$) involves a $X2$ up-sampling of the representation at the higher level (to get the rich semantics), and adding information from a corresponding lower CNN level through 1×1 convolutions (to add better spatial resolution). The produced multiple scale representations are close to being "equivalent" in the sense that a single classifier or regressor can be trained and applied to all of them.

Each image contains a single plant, which was monitored across several weeks (up to 7 weeks for some plants). Hence, plants age varies between several days and several weeks. For testing, the LCC organizers also provided the A5 dataset, which is sampled from A1 to A4 test sets. This dataset was created to test if models generalize well across datasets and are not dataset-specific. In addition, a combined dataset was created, which includes all images from A1 to A4, and was termed Ac (A5 is a subset of Ac). More information is shown in Scharr et al. (2014), Bell and Dee (2016), and Minervini et al. (2016).

- Banana leaves (BL) data: This dataset is larger than the LCC datasets, containing 1,016 images from a top viewpoint collected at plantations in the north of Israel. This dataset is different from the LCC dataset mainly since the leaves are relatively large and elliptical. The banana leaves were annotated using the *via* annotation tool (Dutta and Zisserman, 2019).

Each image was labeled in two annotation levels. The first is the number of leaves, and the second is a leaf center dot annotation. The size of the images varies between 441×441 and $2,448 \times 2,048$, and they were resized to $800 \times W$ where W was chosen to keep the original aspect ratio. To enrich the datasets, random transformations were applied including rotation, scaling, vertical and horizontal flips.

3.2. Direct Regression Architecture and Design

Direct regression algorithms are widely used for various counting tasks (Aich and Stavness, 2017; Dobrescu et al., 2017; Rahnemoonfar and Sheppard, 2017; Ubbens and Stavness, 2017; Aich et al., 2018). For example (Dobrescu et al., 2017), the winner of LCC 2017, uses the ResNet-50 architecture as a regressor. The top classification layers of the network were removed and replaced with a regression "head," containing two fully connected layers followed by a single leaf estimate output neuron. We aim to improve upon Dobrescu et al. (2017) by considering estimation at multiple scales. The regressor used is based on an FPN architecture (Lin et al., 2017a), extending a backbone network to produce multiple resolutions of representation. As in Lin et al. (2017a), a backbone of ResNet-50 (He et al., 2016) is used, trained on ImageNet data (Deng et al., 2009).

3.2.1. An FPN for Multiple Scale Representation

The input to the FPN is an image of arbitrary size, and the output is a feature pyramid of proportionally sized tensor representations at multiple levels. The FPN architecture aims to provide a rich representation with semantic features typically available at higher layers of a CNN, yet at multiple spatial resolutions. To obtain this, tensor representations of increasing spatial resolutions are created iteratively starting from the highest level representation. A schematic view of this process is shown in **Figure 2**. The pyramid used by the method includes representation tensors denoted by $P_3 - P_7$ in Lin et al. (2017a), where pyramid level P_j has a resolution of 2^j lower than the input. All pyramid levels have $C = 256$ channels. The details of the pyramid generally follow Lin et al. (2017a) with a few modest differences, based on Lin et al. (2017b). Specifically, we include the higher and more semantic pyramid levels P_6 and P_7 instead of P_2, which provides better localization (less important for counting). Like Lin et al. (2017b), P_6 is obtained *via* a 3×3 stride-2 convolution on ResNet-50 layer C_5, and P_7 is computed by applying Rectified Linear Unit (ReLU) followed by a 3×3 stride-2 convolution on P_6. A regression sub-model is then applied at each scale.

3.2.2. Count Regression Sub-model

The regression sub-model includes two 3×3 ReLU convolution layers with 256 output maps, followed by a global average pooling operation, flattening the maps to a compact 256×1 representation. This vector is fed into two fully connected layers with decreasing sizes of 128 and 64, respectively, followed by the output layer. This layer contains two output neurons which are estimates of the mean and variance of the expected number of leaves (as shown in details below). The architecture is summarized in **Figure 3**. As in the RetinaNet (Lin et al., 2017b) detection architecture, all the regressor sub-models share their weights, so effectively the same regressor is applied to each pyramid level. We hence use five different image pyramid levels for gaining five different leaf count estimators, with their fusion described in the section "fusing the predictors of the sub-models."

3.2.3. Modeling the Uncertainty of Each Sub-model

Following Kendall and Gal (2017), the regression modules have two outputs, $\hat{y}$ and $\hat{\sigma}^2$, corresponding to the expected leaf count and its variance. The training data is a labeled sample of images

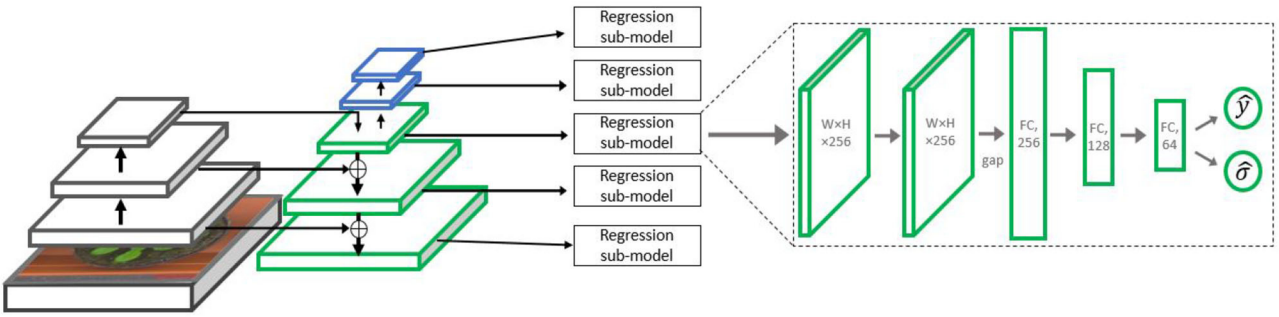

FIGURE 3 | The Multiple-Scale Regression (MSR) architecture. The left side contains the backbone ResNet-50 network, followed by the Feature Pyramid Network (FPN). On the right, the regression sub-model is displayed. Such a sub-model is attached to each P_j level of the FPN,with $j = 3, \ldots, 7$.

$\{I_i, y_i\}_{i=1}^N$ with y_i being the leaf count label. We train the network to minimize the following loss:

$$L_r = \sum_{i=1}^{N} \frac{1}{2} \exp(-s_i) \cdot ||\hat{y}_i - y_i||^2 + \frac{1}{2} s_i \tag{1}$$

where $s_i = \log \hat{\sigma}_i^2$, and $\hat{y}_i = \hat{y}(I_i)$, $\hat{\sigma}_i = \hat{\sigma}(I_i)$ are the outputs of the regressor over image *i*. This loss is applied independently to the five scales and summed to get the total loss minimized. Note that the loss is the log of a Gaussian density (up to an additive constant), but with the mean and variance being input-dependent estimates. This means that in image space regions where the regression is not accurate (i.e., where large deviations occur between $\hat{y}$ and the true y, a larger $\hat{\sigma}^2$ is inferred to reduce the loss. The opposite happens in areas of accurate prediction.

3.2.4. Fusing the Predictions of the Sub-models

The output of the count regression model is the prediction of the mean and the variance for each of the scales. Given count and variance estimates for the scales of the five FPN, new fusion options beyond plain averaging become possible. Keeping an input image index *i* fixed (and hence omitted), denote the multiple estimators for this image by $(\hat{y}^j, \hat{\sigma}^j)_{j=1}^J$. One intuitive way would be to choose the model with the lowest degree of uncertainty, i.e., the one with the lowest predicted variance for the specific input image:

$$\hat{y} = \hat{y^k} \;\; with \;\; k = argmin(\hat{\sigma^j}) \tag{2}$$

While this choice is intuitive, we know that the MLE for the consolidation of Gaussian distributions is different. Specifically, the MLE has the closed-form solution:

$$\hat{y} = \frac{\sum_{j=1}^{J} \frac{1}{(\hat{\sigma^j})^2} \hat{y^j}}{\sum_{j=1}^{J} \frac{1}{(\hat{\sigma^j})^2}} \tag{3}$$

It should be noted that the second method assumes observation independence between the results of the sub-models. This assumption is not fulfilled, since the sub-models receive as input similar representation at different resolutions. However, in practice, the fusion method given by Equation (3) was performed better and was used as default in the experiments.

3.3. Detection-Based Counting

In this approach, the network is taught to first detect leaf centers, using dot annotation provided at training, and only then count them. Toward this task, a two-dimensional heat map was created for every training image to serve as an intermediate output goal. In the heat map, a two-dimensional Gaussian is placed around each leaf center (as shown in detection sub-model section regarding Gaussian size). Each Gaussian was normalized such that its peak pixel (i.e., the leaf center location) equals to 1. The network was trained to estimate this heat map, and then use it to regress the number of leaves. The heat map, in turn, is estimated using an iterative process in which a coarse initial heat map is gradually refined through the network stages to reach a final heat map of fine spatial resolution.

Like in the MSR model, the initial representation used is based on the FPN architecture with a ResNet-50 backbone. The model does not employ multiple scale analysis, but the FPN is used nevertheless due to the combination of semantic features from the higher layers and the high spatial resolution obtained at its lower layers. A detection sub-model regressing the heat map is placed on top of the high-resolution pyramid scale (P_3). Then, a count regression sub-network accepts the heat map and tensors leading to it as input and predicts the final count output. The architecture is summarized in **Figure 4**.

3.3.1. Detection Sub-model

The detection sub-model is a small FCN, containing four 3×3 ReLU convolutional layers with 256 filters each. The final layer in the sub-model contains a single 2*D* map approximating the ground truth leaf center heat map, and it is produced by an additional 1×1 ReLU convolutional layer with a single filter. $\hat{p}$ and p denote the value of the network estimation and the target heat map, respectively, for a single map location (dropping location and layer indices for convenience). The minimized loss is a weighted smooth-L1 loss

$$L_d(\hat{p}, p) = \begin{cases} (1 - w) \cdot Huber(\hat{p}, p) & \text{, where p > 0} \\ w \cdot Huber(\hat{p}, p) & \text{, otherwise} \end{cases} \tag{4}$$

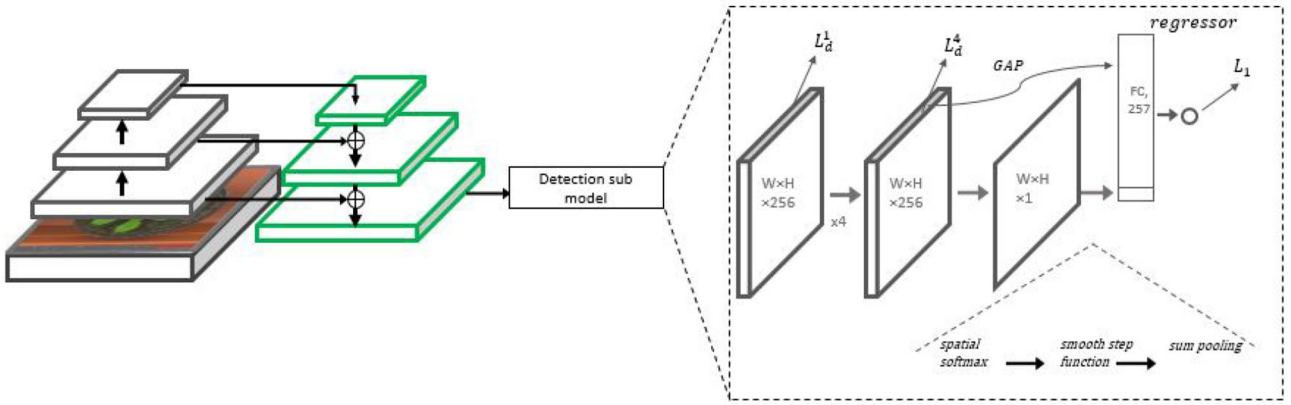

FIGURE 4 | The Detection with Regression Network (DRN) architecture. On the left side, the backbone ResNet-50 with the FPN is displayed, from which P_3 representation is taken for further processing. On the right, the detection sub-model is presented, including four 3 × 3 ReLU convolutional layers repeatedly regressing the leaf heat map. A counting head summarizes the final heat map estimate into an initial count estimate. The final count estimate is regressed from a combination of the initial estimate and a feature vector extracted from the final heat map.

With

$$Huber(\hat{p},p) = \begin{cases} |\hat{p}-p| - \frac{1}{2} & \text{, where } |\hat{p}-p| \geq 1 \\ \frac{1}{2}(\hat{p}-p)^2 & \text{, otherwise} \end{cases} \quad (5)$$

The weight parameter w ($w = 0.1$ in the experiments) keeps the total weight of positive pixels in the heat-map high. This directs the optimization toward accurate regression of these values, rather than the pixels with the value zero, which outnumber them considerably. Although the final heat map is predicted following all four convolutional layers, each of those layers predicts a heat map independently using an additional loss term L_d^k for $k = 1, \ldots, 4$. The ground truth heat maps guiding the estimates produced by early convolutional layers were created using larger Gaussian kernels. Hence the first convolutional layers estimate a coarse heat-map (created with a 7×7 Gaussian kernel), in which sometimes Gaussians may overlap, while the final layer predicts a finer heat-map (created with a 3×3 Gaussian kernel). Each of the layers in the sub-model are guided using the same loss function as the final loss. Denoting by $p_{ij}^k, \hat{p}_{ij}^k$ the true and predicted values at location (i, j) of layer k, the minimized detection loss is the sum of losses overall pixels and layers:

$$L_d = \sum_{k=1}^{4} L_d^k = \sum_{k=1}^{4} \sum_{\{(i,j)\}} L_d(\hat{p}_{ij}^k, p_{ij}^k) \quad (6)$$

3.3.2. Counting Sub-network

Given an estimated heat-map, an initial count estimate can be based on finding the number of Gaussian centers. Ideally, this number is the required leaf count. However, assuming imperfect detection, it is preferable not to use it directly as the final count, but as a strong feature provided to the final regressor. To properly count the Gaussian centers, we suggest to incorporate a smooth NMS procedure, shrinking each Gaussian to a delta function, followed by a global sum operation. Hence, the following NMS function is applied to the estimated heat map $\hat{P} \in R^{N \times M}$.

$$\begin{aligned} Q &= MaxPool(\hat{P}, (K, K)) \\ \hat{P}_{NMS} &= \hat{P} \cdot exp(-\beta \cdot |\hat{P} - Q|) \end{aligned} \quad (7)$$

Where $MaxPool(X, F)$ is the operation of max-pooling with stride 1 and an F filter size. Following this operation, a pixel in the output map $\hat{P}_{NMS}$ keeps a value close to its original value in $\hat{P}$ if it was the highest value of $\hat{P}$ in a $K \times K$ neighborhood (as shown in example **Figure 5**). The experiments indicate that large β values were effective, entailing a sharp NMS keeping only the winners. K is chosen to be the same value used in the heat map creation, so a single pixel is expected to remain active from a full Gaussian.

The NMS operation keeps mostly the pixels that are local maxima, but it does not remove noise, so $\hat{P}_{NMS}$ still contains small values in non-center pixels. We hence apply a smooth step function (sigmoid) layer of the form $y(x) = \frac{1}{1+e^{-\rho(x-t)}}$ to keep only values greater than some threshold ($t = 0.8$, $\rho = 15$ were used). The result is a detection map $\hat{D}$ containing the estimated leaf centers as its active pixels. A global sum pooling layer applied to $\hat{D}$ then gives a single number $C_{\hat{D}}$, the detection-based estimate for the leaf count.

Beyond the detection-based estimate $C_{\hat{D}}$, additional features are used for the final count regression. A Global Average Pooling (GAP) operation is applied to the fourth convolutional layer (the layer predicting the final heat map), obtaining a 256-dimensional feature vector V_4. The final count estimator $\hat{C}$ is obtained by applying linear regression in the final layer, i.e., $\hat{C} = W \cdot [V_4, C_{\hat{D}}]$ where $[V_4, C_{\hat{D}}]$ is a concatenated 257 features vector. We train the final counter to minimize an L_1 loss estimated based on the true count, which is added to the detection loss (Equation 6).

4. RESULTS AND DISCUSSION

We describe next the training procedure and evaluation metrics. Then, the results of the two suggested methods are presented and compared to current art methods. Finally, performance analysis

A

0.04	0.01	0.8	0.08	0.16	0.9	0.1
0.1	0.6	0.04	0.15	0.4	0.29	0.22
0.15	0.23	0.9	0.17	0.15	0.1	0.4
0.16	0.15	0.2	0.1	0.95	0.15	0.2
0.1	0.3	0.3	0.17	0.3	0.09	0.25
0.9	0.2	0.98	0.05	0.5	0.1	0.05

B

0.6	0.8	0.8	0.4	0.9	0.9	0.22
0.6	0.9	0.9	0.17	0.4	0.9	0.4
0.23	0.9	0.9	0.95	0.95	0.4	0.4
0.3	0.3	0.3	0.95	0.95	0.25	0.25
0.9	0.98	0.98	0.5	0.5	0.25	0.25
0.9	0.2	0.98	0.05	0.5	0.25	0.25

C

0	0	0.8	0	0	0.9	0
0	0	0	0	0.4	0	0
0	0	0.9	0	0	0	0.4
0	0	0	0	0.95	0	0
0	0	0	0	0	0	0.25
0.9	0	0.98	0	0.5	0	0

FIGURE 5 | Smooth Non-Maxima Suppression (NMS) procedure example: **(A)** an estimated heat map $\hat{P}$. **(B)** The resulting map Q after applying the *MaxPool* operation with $K = 2$. Local maxima are highlighted in red. **(C)** the output map $\hat{P}_{NMS}$.

of the suggested method is done using visualization, ablation studies, and experiments with varying sample sizes.

4.1. Training Procedure

Following Dobrescu et al. (2017), validation results were obtained using a 4-fold cross-validation procedure, and the average results over the folds were reported. Test results were obtained by evaluating the model on a hidden held-out set kept by the challenge organizers (Scharr et al., 2014; Bell and Dee, 2016; Minervini et al., 2016) (the model predictions were sent to the LCC organizers for evaluation). The tested model was trained over the accumulated dataset, with 75% of the data used for model fitting and 25% as a validation set for the early stopping criterion. We did not use the provided foreground plant mask in either of the models. In all experiments, ADAM (Kingma and Ba, 2014) optimizer was used with a learning rate of 10^{-5}.

4.2. Evaluation Metrics

For count regression evaluation, we used the metrics provided by the LCC workshop organizers (as shown in Dobrescu et al., 2017). C and $\hat{C}$ denote the true and estimated count, respectively, and these include

- The average distance $E[\hat{C} - C]$, where $E[\cdot]$ denotes the average over test images. This is not used as an accuracy score, but it measures the bias of the predictor (its tendency to over or underestimate the count) — DiC.
- The average L_1 distance $E[|\hat{C} - C|] - |DiC|$.
- The fraction of accurate agreement $E[1_{\hat{C}==C}]$ — agreement [%].
- The Mean-Squared Error $E[(\hat{C} - C)^2]$ — MSE.

For estimation of the detection performance, a recall-precision curve was used with the Average Precision (AP) metric—the area under the curve. To determine if a leaf detection (an active point in $\hat{D}$) is a hit or a miss we use the criterion introduced in the Percentage of Correct Keypoints (PCK) computation (Yang and Ramanan, 2013). With PCK, a leaf detection point is considered a hit if the distance between it and a ground truth leaf center is lower than $\alpha \cdot max(w, h)$ where (w, h) are the width and height of the bounding boxes surrounding the entire object (plant). Like in Yang and Ramanan (2013) we used $\alpha = 0.1$ to test the detection performance.

4.3. Results and Comparison to Previous Work

Cross-validation results on the datasets of the LCC challenges, as well as the new banana leaf dataset, are reported in **Table 1**. The results are more stable for datasets $A1$ and $A4$, which are larger. There is no clear winner in this examination, but the proposed MSR and DRN have a slight advantage over the LCC winner (Dobrescu et al., 2017). Also, it can be seen that when the training set is small, the MSR method has an advantage, while for larger datasets ($A4$ and BL) the DRN method usually performs better. In **Table 2**, the results of the test held-out set of the LCC are presented and compared to current best methods. The results here are more decisive. First, the MSR and DRN methods have an advantage over previous art for most datasets and metrics. Second, DRN has a clear leading position overall methods (including MSR) in all the three performance indices ($|DiC|$, agreement, and MSE). This is specifically seen in the large sample and heterogeneous conditions "all" and $A5$.

4.4. Detection and Regression Network Analysis

Some examples of the performance of the detectors are shown in **Figure 6**. As can be seen from these examples, the estimated Gaussian heat map successfully finds the vast majority of the leaves. Misses occur mostly for close leaves, high leaf occlusion cases, and very small leaves. For example, row *c* in **Figure 6** shows a typical leaf miss detection. Notice that the leaf is extremely small and that its Gaussian is too close to other Gaussians.

Table 3 shows the metric scores for several ablated model versions on the largest data set of the LCC, $A4$. The results indicate that each of the examined elements contributes to the performance of the detector. Specifically, combining the detection with a final regression and the usage of multiple intermediate losses for detection is of high importance. This model obtained 0.94 AP when trained on 80% of the Ac dataset and validated on 20% of it, indicating the achievement of both high recall and high precision. For example, for a recall value of 0.7, the precision is 0.987.

4.5. Multiple Scale Regression Analysis

Table 4 presents the results of intermediate algorithm versions between the baseline of Dobrescu et al. (2017) and the

TABLE 1 | Cross-validation results on separate datasets.

Dataset		DiC			|DiC|			(%)			MSE		
Train	**Test**	**MSR**	**DRN**	**Dobrescu et al. (2017)**	**MSR**	**DRN**	**Dobrescu et al. (2017)**	**MSR**	**DRN**	**Dobrescu et al. (2017)**	**MSR**	**DRN**	**Dobrescu et al. (2017)**
A1	A1	−0.06	**0.03**	−0.81	**0.53**	0.56	0.94	**58**	51	25	0.78	**0.7**	1.38
A2	A2	**0.61**	0.73	−2.38	**1.26**	2.05	2.38	**43**	22	38	**3.71**	7.82	12.9
A3	A3	**0.14**	−0.45	−0.57	**1.07**	1.6	1.43	**30**	21	14	**1.89**	3.83	2.57
A4	A4	**0.03**	−0.21	0.1	0.71	**0.6**	0.91	45	**49**	35	1.15	**0.85**	1.54
Ac	A1	**0.02**	0.21	0.28	0.66	0.6	**0.53**	44	45	**56**	0.86	**0.7**	0.72
Ac	A2	0.63	0.45	**−0.38**	1.09	**0.86**	0.88	35	**52**	38	3.33	2.44	**1.38**
Ac	A3	0.14	**0.02**	0.71	**0.71**	0.75	**0.71**	41	29	**57**	0.94	**0.83**	1.57
Ac	A4	**−0.01**	0.06	0.12	0.65	**0.63**	0.69	47	**51**	46	**0.97**	1.0	1.01
BL	BL	**0.05**	0.09	0.07	0.41	**0.37**	0.46	61	**69**	56	0.43	**0.42**	0.51

Results are presented for the algorithms: Multiple Scale Regression (MSR), Detection and Regression Network (DRN), and direct regression of Dobrescu et al. (2017), the LCC 2017 challenge winner. The best results for each metric is presented in bold.

TABLE 2 | Results of the MSR and the DRN on the held-out test set, with comparison to the winner of the 2017 LCC (Dobrescu et al., 2017) and to Giuffrida et al. (2018) which shows identical results.

		A1	A2	A3	A4	A5	All
DiC							
	MSR	−0.27 (1.21)	**−0.22 (0.67)**	**−0.04 (1.61)**	**0.15 (1.08)**	**0.10 (1.22)**	**0.07 (1.22)**
	DRN	**−0.24 (0.94)**	−0.67 (1.00)	−0.25 (1.01)	0.26 (1.09)	0.12 (1.08)	0.09 (1.08)
	Dobrescu et al. (2017) and Giuffrida et al. (2018)	−0.39 (1.17)	−0.78 (1.64)	0.13 (1.55)	0.29 (1.10)	0.25 (1.21)	0.19 (1.24)
	Aich and Stavness (2017)	−0.33 (1.38)	−0.22 (1.86)	2.71 (4.58)	0.23 (1.44)	0.80 (2.77)	0.73 (2.72)
	Giuffrida et al. (2016)	−0.79 (1.54)	−2.44 (2.88)	−0.04 (1.93)	–	–	–
|DiC|							
	MSR	0.70 (1.02)	**0.44 (0.53)**	1.14 (1.12)	0.76 (0.78)	0.84 (0.90)	0.83 (0.90)
	DRN	**0.48 (0.83)**	0.67 (1.00)	**0.75 (0.72)**	**0.75 (0.83)**	**0.73 (0.80)**	**0.72 (0.81)**
	Dobrescu et al. (2017) and Giuffrida et al. (2018)	0.88 (0.86)	1.44 (1.01)	1.09 (1.10)	0.84 (0.76)	0.90 (0.85)	0.91 (0.86)
	Aich and Stavness (2017)	1.00 (1.00)	1.56 (0.88)	3.46 (4.04)	1.08 (0.97)	1.66 (2.36)	1.63 (2.30)
	Giuffrida et al. (2016)	1.27 (1.15)	2.44 (2.88)	1.36 (1.37)	–	–	–
%							
	MSR	57.6	**55.6**	32.1	42.9	41.7	42.3
	DRN	**63.6**	**55.6**	**35.7**	**43.5**	**42.6**	**43.7**
	Dobrescu et al. (2017) and Giuffrida et al. (2018)	33.3	11.1	30.4	34.5	33.2	32.9
	Aich and Stavness (2017)	30.3	11.1	7.1	29.2	23.8	24.0
	Giuffrida et al. (2016)	27.3	44.4	19.6	–	–	–
MSE							
	MSR	1.48	**0.44**	2.54	**1.19**	1.50	1.49
	DRN	**0.91**	1.33	**1.07**	1.25	**1.17**	**1.17**
	Dobrescu et al. (2017) and Giuffrida et al. (2018)	1.48	3.00	2.38	1.28	1.53	1.56
	Aich and Stavness (2017)	1.97	3.11	28.0	2.11	8.28	7.90
	Giuffrida et al. (2016)	2.91	13.33	3.68	–	–	–

A5 dataset is a sample from A1 to A4 datasets. "All" dataset is a combination of all the datasets (A1–A4). The reported results are the mean and standard values of each metric. It is noticeable that the DRN algorithm gets a better score in almost each of the indices. The best results for each metric is presented in bold.

suggested multiple-scale regression. The results are of cross-validation experiments on the A1 dataset. The baseline model following Dobrescu et al. (2017) includes a ResNet-50 backbone, global average pooling on top of the C_5 convolution layer, two fully connected layers in decreasing sizes, and a final single neuron predictor. As can be seen, two main steps improve the

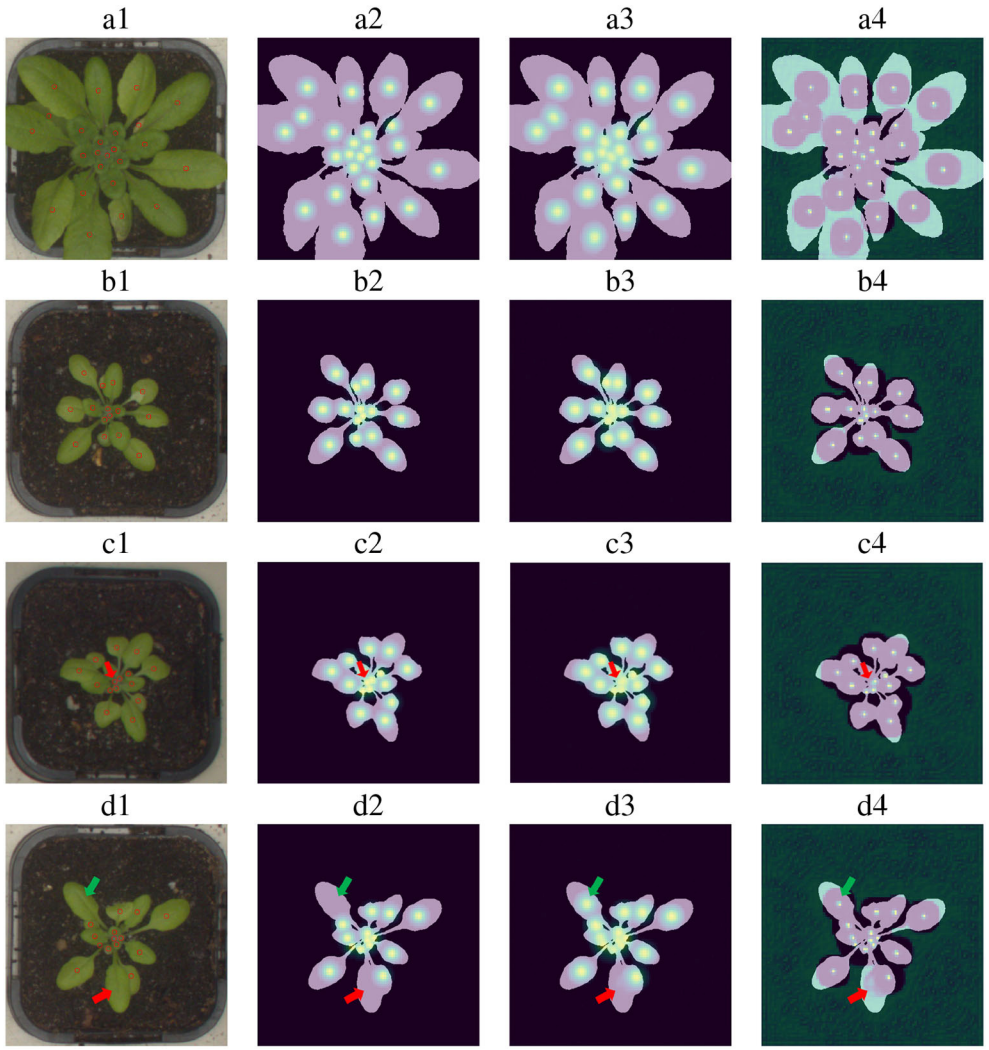

FIGURE 6 | Leaf center detection examples for four images. **Column 1:** Input images with their ground truth point annotations. **Column 2:** Ground truth Gaussian heat maps generated using human annotation. **Column 3:** Gaussian heat maps inferred by the model. **Column 4:** Leaf center predictions after the non-maxima suppression operation. Examples "a" and "b" present successful detection cases. Examples "c" and "d" include miss detection examples and wrong annotation found by the model (a miss-detection is pointed with a red arrow, wrong annotation with a green arrow).

TABLE 3 | Ablation table for the DRN.

Condition	DiC	$\|DiC\|$	(%)	MSE
1. Full model	−0.02	0.6	49	0.85
2. Fixed GT heat maps	0.02	0.63	48	0.93
3. No data Aug.	0.1	0.71	42	1.09
4. No multiple losses	0.6	1.14	41	4.91
5. Detection only	0.78	1.36	40	7.7

The table includes cross-validation results on the A4 dataset. 1. The full model. 2. Using fixed size kernels for generating the ground truth Gaussians of the heat maps. 3. Training without data augmentation. 4. Using a single detection loss at the end of the detection sub-model. 5. Using count estimator $C_{\hat{D}}$ as the single feature for count regression.

performance over this baseline: moving to pyramid features and using the maximum likelihood technique for the fusion of the estimators.

The larger dataset BL enables measurement of the relation between sample size and counting accuracy in a larger domain than before. **Figure 7** shows $|DiC|$, MSE, and counting agreement as a function of training set size. In the experiment, 150 images were held out for accuracy estimation, and models were trained using subsets of 50, 100, 250, 450, 700, and 866 images from the BL dataset. The presented results are the average of five experiments with different train-test splits. It can be seen that the DRN model significantly benefits from additional examples in the explored range.

4.6. Inference Speed

We implemented the model in Python, using Keras and TensorFlow framework. We used a single 1080-Ti GTX Nvidia GPU for training and testing. While training the models takes up to 24 h (on the largest dataset), the inference is fast. MSR

TABLE 4 | Incremental improvements over a baseline in the direct regression model, cross-validation results on A1 dataset.

Index	Input	Loss	Description	DiC	$\|DiC\|$	Agreement (%)	MSE
1	C_5	L_2 *loss*	Baseline	0.15	0.62	48	0.9
2	P_3	L_2 *loss*	pyramid 1 level	−0.07	0.55	53	0.76
3	$P_3 - P_7$	L_2 *loss*	Avg fuse	0.18	0.66	44	0.91
4	$P_3 - P_7$	*Gauss*	*min* $\hat{\sigma}$ fuse	−0.08	0.62	45	0.82
5	$P_3 - P_7$	*Gauss*	MLE fuse	−0.06	0.53	58	0.78

The table rows present the following: 1. A baseline is similar to Dobrescu et al. (2017) (the implementation of it). 2. Replacing the input of the regressor with the middle pyramid level P5. 3. Using five pyramid levels, averaging the results. 4. Using five pyramid levels with the min-sigma fusion of the predictors (Equation 2). 5. Five pyramid levels with MLE fusion (Equation 3).

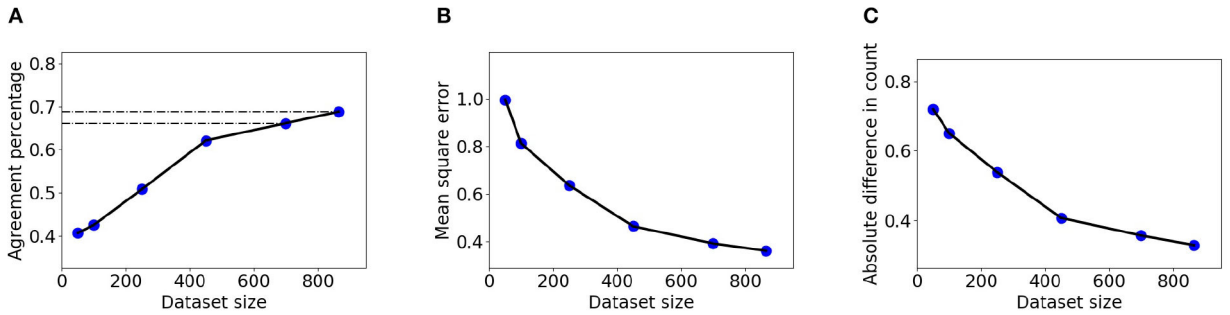

FIGURE 7 | Detection and Regression Network network accuracy as a function of dataset size for BL data, measured using the LCC evaluation criteria. **(A)** Agreement index. **(B)** MSE index. **(C)** $|DiC|$ index. the method earns, though with diminishing returns, from samples including many hundreds of data points.

reaches 20 FPS (0.05 s per image) and DRN reaches 7 FPS (0.14 s per image) for images of average size 450 × 460, typical for all LCC datasets except for A3. As expected, MSR is significantly faster than DRN since it uses a direct regression approach. For A3, containing larger images (2, 448 × 2, 048), inference took on average 0.34 s for MSR and 0.43 for DRN.

4.7. Discussion

The methods developed here have several limitations, related to difficulty factors like leaf occlusion, the number of plants in the image, illumination conditions, and leaf shape diversity. While occlusion is present in the LCC and BL datasets, it is not severe, as can be seen from the high rates of the count agreement and low $|DiC|$. Specifically, the $|DiC|$ is lower than one on all the datasets, indicating that on average at most one leaf is missed. Due to the occlusion problem, however, the current algorithmic framework is limited to handle 1–2 dozens of leaves at most. Beyond this, occlusion becomes dominant, as no single viewpoint exists from which all the leaves can be seen. For handling mature plants with many dozens of leaves, a single top viewpoint is not enough. Instead, multiple views including side views should be used, and registration of the cameras is required to avoid double counting of leaves across the viewpoints.

The models were trained using datasets containing a single plant in each image. They do not include mechanisms for handling the detection of multiple plants and the assignment of leaves to the correct plant. However, when there are several plants in a single image, a simple solution may incorporate an object detector as the first stage to detect each of the plants, and then count leaves in each detected bounding box independently. Such a two-stage method was applied successfully in Khoroshevsky et al. (2020) for counting wheat spikelets in a spike and bananas in a bunch.

The models suggested here were trained on the LCC and BL datasets, which are limited in viewpoint, illumination conditions, and leaf shape type. If one of these conditions is changed, for example having images taken outdoors rather than in a plantation illumination, re-training of the models would be required to cope with it. While re-training is required, it is likely that the architecture details will not require significant changes. To enable easier re-training, and in some cases even avoid it completely, domain adaptation methods as suggested by Giuffrida et al. (2019) can be utilized.

5. CONCLUSIONS

The two proposed counting methods improve upon the state of the art results on the test set of the LCC. The better strategy depends on the dataset size. For smaller datasets, multiple-scale regression has a slight advantage, but for a larger dataset with several hundred images the detection-based method is preferable. From an annotation effort point of view, the detection-based approach is more demanding due to its reliance on leaf dot annotation. On the other hand, it provides leaf detection capabilities in addition to counting.

For the detection-based pipe, the most important contributing element is the fusion of the detection results with a regression

component. Without it, this approach is inferior in all indices. For the multiple-scale regression approach, an FPN-based representation is helpful, but fusion must be done carefully using confidential information. An interesting and non-trivial direction may be to try and merge the advantages of both methods in a single architecture, providing detection with regression at multiple scales.

Interesting possible extensions of this study may be adapting it to specific agricultural tasks, like in-field leaf counting of weed, or growth-rate estimation in a plantation. In addition, while the suggested methods were developed for leaf counting they can be extended to other counting tasks, like counting flowers or fruits on a tree.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because our banana leaf dataset is the property of a commercial company ("Rahan") and cannot be made public. The LCC datasets are publicly available. Requests to access the datasets should be directed to guyfar@post.bgu.ac.il.

AUTHOR CONTRIBUTIONS

GF and YI developed the main components of the detection-based and regression-based algorithms, respectively, with the guidance of AB-H. FK contributed to the algorithmic developments and experiments. GF prepared the manuscript with AB-H. All authors read and approved the final manuscript.

FUNDING

This publication was funded by the Israel Innovation Authority through the Phenomics MAGNET Consortium and by the Israel Ministry of Science and Technology.

ACKNOWLEDGMENTS

We thank Dr. Oshry Markovitch for his help in planning the banana leaf experiments and data collection, and to Ortal Bakhshian who collected the banana leaf data and annotated it. This work was an extension of a previously presented conference paper (Itzhaky et al., 2018).

REFERENCES

Adamsen, F., Coffelt, T., Nelson, J. M., Barnes, E. M., and Rice, R. C. (2000). Method for using images from a color digital camera to estimate flower number. *Crop Sci*. 40, 704–709. doi: 10.2135/cropsci2000.403704x

Aich, S., Ahmed, I., Obsyannikov, I., Stavness, I., Josuttes, A., Strueby, K., et al. (2018). "Deepwheat: estimating phenotypic traits from images of crops using deep learning," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Lake Tahoe, NV), Vol. 3. doi: 10.1109/WACV.2018.00042

Aich, S., and Stavness, I. (2017). "Leaf counting with deep convolutional and deconvolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 2080–2089. doi: 10.1109/ICCVW.2017.244

Arteta, C., Lempitsky, V., and Zisserman, A. (2016). "Counting in the wild," in *European Conference on Computer Vision* (Amsterdam: Springer), 483–498. doi: 10.1007/978-3-319-46478-7_30

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi: 10.1109/TPAMI.2016.2644615

Bell, J., and Dee, H. M. (2016). *Aberystwyth Leaf Evaluation Dataset*.

Cruz, J. A., Yin, X., Liu, X., Imran, S. M., Morris, D. D., Kramer, D. M., et al. (2016). Multi-modality imagery database for plant phenotyping. *Mach. Vision Appl.* 27, 735–749. doi: 10.1007/s00138-015-0734-6

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). "Imagenet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

Dobrescu, A., Giuffrida, M. V., and Tsaftaris, S. A. (2017). "Leveraging multiple datasets for deep leaf counting," in *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)* (IEEE), 2072–2079. doi: 10.1109/ICCVW.2017.243

Dutta, A., and Zisserman, A. (2019). The VGG image annotator (VIA). *arXiv* 1904.10699.

Farjon, G., Krikeb, O., Bar-Hillel, A., and Alchanatis, V. (2019). Detection and counting of flowers on apple trees for better chemical thinning decisions. *Precis. Agric*. 21, 503–521. doi: 10.1007/s11119-019-09679-1

Giuffrida, M. V., Dobrescu, A., Doerner, P., and Tsaftaris, S. A. (2019). "Leaf counting without annotations using adversarial unsupervised domain adaptation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Long Beach, CA: IEEE), 2590–2599. doi: 10.1109/CVPRW.2019.00315

Giuffrida, M. V., Doerner, P., and Tsaftaris, S. A. (2018). Pheno-deep counter: a unified and versatile deep learning architecture for leaf counting. *Plant J.* 96, 880–890. doi: 10.1111/tpj.14064

Giuffrida, M. V., Minervini, M., and Tsaftaris, S. A. (2016). *Learning to Count Leaves in Rosette Plants*. York, NY: BMVA Press. doi: 10.5244/C.29.CVPPP.1

Großkinsky, D. K., Svensgaard, J., Christensen, S., and Roitsch, T. (2015). Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J. Exp. Bot*. 66, 5429–5440. doi: 10.1093/jxb/erv345

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)* (Venice: IEEE), 2980–2988. doi: 10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 770–778. doi: 10.1109/CVPR.2016.90

Itzhaky, Y., Farjon, G., Khoroshevsky, F., Shpigler, A., and Bar-Hillel, A. (2018). "Leaf counting: Multiple scale regression and detection using deep CNNs," in *BMVC* (Newcastle).

Jiang, B., Wang, P., Zhuang, S., Li, M., Li, Z., and Gong, Z. (2019). Leaf counting with multi-scale convolutional neural network features and fisher vector coding. *Symmetry* 11:516. doi: 10.3390/sym11040516

Kendall, A., and Gal, Y. (2017). "What uncertainties do we need in bayesian deep learning for computer vision?" in *Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA*, 5574–5584.

Khoroshevsky, F., Khoroshevsky, S., Markovich, O., Granitz, O., and Bar-Hillel, A. (2020). "Phenotyping problems of parts-per-object count," in *European Conference on Computer Vision* (Glasgow: Springer), 261–278. doi: 10.1007/978-3-030-65414-6_19

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv* 1412.6980.

Kuznichov, D., Zvirin, A., Honen, Y., and Kimmel, R. (2019). "Data augmentation for leaf segmentation and counting tasks in Rosette plants," in *CVPR CVPPP Workshop* (Long Beach, CA). doi: 10.1109/CVPRW.2019.00314

Lempitsky, V. S., and Zisserman, A. (2010). "Learning to count objects in images," in *NIPS* (Vancouver).

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017a). "Feature pyramid networks for object detection," in *CVPR* (Honolulu, HI), Vol. 1, 4. doi: 10.1109/CVPR.2017.106

Lin, T. Y., Goyal, P., Girshick, R. B., He, K., and Dollár, P. (2017b). "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)* (Venice), 2999–3007. doi: 10.1109/ICCV.2017.324

Liu, T., Wu, W., Chen, W., Sun, C., Zhu, X., and Guo, W. (2016). Automated image-processing for counting seedlings in a wheat field. *Precis. Agric*. 17, 392–406. doi: 10.1007/s11119-015-9425-6

Lu, H., Cao, Z., Xiao, Y., Zhuang, B., and Shen, C. (2017). Tasselnet: counting maize tassels in the wild via local counts regression network. *Plant Methods* 13:79. doi: 10.1186/s13007-017-0224-0

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recogn. Lett*. 81, 80–89. doi: 10.1016/j.patrec.2015.10.013

Rahnemoonfar, M., and Sheppard, C. (2017). Deep count: fruit counting based on deep simulated learning. *Sensors* 17:905. doi: 10.3390/s17040905

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 779–788. doi: 10.1109/CVPR.2016.91

Ren, M., and Zemel, R. S. (2017). "End-to-end instance segmentation with recurrent attention," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 6656–6664. doi: 10.1109/CVPR.2017.39

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 91–99.

Romera-Paredes, B., and Torr, P. H. S. (2016). "Recurrent instance segmentation," in *European Conference on Computer Vision* (Amsterdam: Springer), 312–329. doi: 10.1007/978-3-319-46466-4_19

Scharr, H., Minervini, M., Fischbach, A., and Tsaftaris, S. A. (2014). *Annotated Image Datasets of Rosette Plants*. Technical Report FZJ-2014-03837.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 2818–2826. doi: 10.1109/CVPR.2016.308

Teimouri, N., Dyrmann, M., Nielsen, P. R., Mathiassen, S. K., Somerville, G. J., and Jørgensen, R. N. (2018). Weed growth stage estimator using deep convolutional neural networks. *Sensors* 18:1580. doi: 10.3390/s18051580

Telfer, A., Bollman, K., and Poethig, R. (1997). Phase change and the regulation of trichome distribution in arabidopsis thaliana. *Development* 124, 645–654. doi: 10.1242/dev.124.3.645

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci*. 8:1190. doi: 10.3389/fpls.2017.01190

Walter, A., and Schurr, U. (1999). The modular character of growth in nicotiana tabacum plants under steady-state nutrition. *J. Exp. Bot*. 50, 1169–1177. doi: 10.1093/jxb/50.336.1169

Xie, W., Noble, J. A., and Zisserman, A. (2018). Microscopy cell counting and detection with fully convolutional regression networks. *Comput. Methods Biomech. Biomed. Eng*. 6, 283–292. doi: 10.1080/21681163.2016.1149104

Xie, Y., Xing, F., Kong, X., Su, H., and Yang, L. (2015). "Beyond classification: structured regression for robust cell detection using convolutional neural network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 358–365. doi: 10.1007/978-3-319-24574-4_43

Yang, Y., and Ramanan, D. (2013). Articulated human detection with flexible mixtures of parts. *IEEE Trans. Pattern Anal. Mach. Intell*. 35, 2878–2890. doi: 10.1109/TPAMI.2012.261

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Farjon, Itzhaky, Khoroshevsky and Bar-Hillel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fpls.2021.645899

Occlusion Robust Wheat Ear Counting Algorithm Based on Deep Learning

*Yiding Wang†, Yuxin Qin† and Jiali Cui**

School of Information Science and Technology, North China University of Technology, Beijing, China

Counting the number of wheat ears in images under natural light is an important way to evaluate the crop yield, thus, it is of great significance to modern intelligent agriculture. However, the distribution of wheat ears is dense, so the occlusion and overlap problem appears in almost every wheat image. It is difficult for traditional image processing methods to solve occlusion problem due to the deficiency of high-level semantic features, while existing deep learning based counting methods did not solve the occlusion efficiently. This article proposes an improved EfficientDet-D0 object detection model for wheat ear counting, and focuses on solving occlusion. First, the transfer learning method is employed in the pre-training of the model backbone network to extract the high-level semantic features of wheat ears. Secondly, an image augmentation method Random-Cutout is proposed, in which some rectangles are selected and erased according to the number and size of the wheat ears in the images to simulate occlusion in real wheat images. Finally, convolutional block attention module (CBAM) is adopted into the EfficientDet-D0 model after the backbone, which makes the model refine the features, pay more attention to the wheat ears and suppress other useless background information. Extensive experiments are done by feeding the features to detection layer, showing that the counting accuracy of the improved EfficientDet-D0 model reaches 94%, which is about 2% higher than the original model, and false detection rate is 5.8%, which is the lowest among comparative methods.

Keywords: wheat ear counting, transfer learning, image augmentation, attention module, deep learning

OPEN ACCESS

Edited by:
Panos M. Pardalos,
University of Florida, United States

Reviewed by:
Pouria Sadeghi-Tehran,
Rothamsted Research,
United Kingdom
Carlos Camino,
European Commission, Joint Research Centre (JRC), Italy

***Correspondence:**
Jiali Cui
jialicui@ncut.edu.cn

† These authors contributed equally to the work and share first co-authorship

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *24 December 2020*
Accepted: *19 May 2021*
Published: *11 June 2021*

Citation:
Wang Y, Qin Y and Cui J (2021) Occlusion Robust Wheat Ear Counting Algorithm Based on Deep Learning. Front. Plant Sci. 12:645899. doi: 10.3389/fpls.2021.645899

INTRODUCTION

The number of wheat ears is used as the essential information to study wheat yield (Prystupa et al., 2004; Peltonen-Sainio et al., 2007; Ferrante et al., 2017). Accurate monitoring of the number of wheat ears is necessary for growers to predict wheat harvest and growth trends. The counting of wheat ears is usually done manually, which is an extremely time-consuming work (Liu et al., 2016). In large-scale planting scenarios, the accuracy of manual counting will increase with the increase of the number of wheats. Therefore, it is indispensable to develop an efficient and automatic wheat ear counting method.

Traditionally, automatic counting methods based on image processing have been successfully used in practical applications, such as plant leaf counting and fruit counting (Giuffrida et al., 2015; Mussadiq et al., 2015; Maldonado and Barbosa, 2016; Stein et al., 2016; Aich and Stavness, 2017; Barré et al., 2017; Dobrescu et al., 2017). These methods fall into two categories. In the first class of

conventional methods, the color of the target objects is extracted and set as positive samples. The background color is set as negative samples, and then traditional machine learning classification methods, such as Support Vector Machine (SVM) are used to separate the target and background in the images. But in the actual wheat ear counting task, the varieties and maturity of wheat will be different (**Figure 1**), which lies in the fact that the preset positive sample color cannot represent wheat ears under all conditions. Methods in the second category used threshold segmentation algorithms, such as Watershed Algorithm (Bleau and Leon, 2000). Although this type of method reduces the dependence on color information, the segmentation threshold is determined by experience, which makes the algorithm have no generalization ability and low robustness.

The previous wheat ear counting methods were mainly realized by manual counting and traditional image processing methods, which has great room for improvement in precision and generalization ability. In contrast, for counting complex background and dense object distribution, deep learning has inherent advantages that can overcome some of the shortcomings of traditional methods. There are two ways to implement deep learning based wheat ear counting algorithm: semantic segmentation and object detection. The process of counting using the semantic segmentation method is reproduced below. Above all, the ears of wheat are labeled pixel by pixel in the original images, and the regions containing the ears are positive samples and other regions are negative samples. After the image is annotated, the fully convolution network such as Unet (Ronneberger et al., 2015), FCN (Long et al., 2015), etc. is usually trained in way of encoder-decoder (Grbovic et al., 2019; Sadeghi-Tehran et al., 2019; Misra et al., 2020; Xu X. et al., 2020). The trained full convolutional network can segment each wheat ear in the input images and output it in the form of a mask. There are two difficulties with this approach. First, training the fully convolution network requires pixel-level annotation. The time cost of this annotation method is almost the same as that of manually counting the number of ears in the image. Second, the mask output by fully convolutional network is not directly related to the number of wheat ears. Solving this problem usually involves designing multifaceted post-processing steps. By using object detection implementation counting, these problems can be avoided effectively. In this way, people roughly mark the positions of the upper left and lower right corners of the ears, and the detection results can be directly converted to the number of ears. Hasan et al. (2018) adopted R-CNN (Girshick et al., 2014) and Madec et al. (2019) adopted the Faster-RCNN (Ren et al., 2017) method to calculate the number of wheat ears. Later, more researchers utilized object detection methods to model wheat ear counting tasks (Mohanty et al., 2016; Xiong et al., 2019; Lu and Cao, 2020). Therefore, wheat ear counting based on deep learning was realized by object detection methods, which makes the algorithm easy to be applied in practice.

With the rapid development of deep learning theory, object detection methods based on deep learning have become a new paradigm in machine learning in recent years. Compared with traditional image processing technologies, Convolutional Neural Networks (CNN) is invariant to geometric transformation, illumination, and background differences. This feature overcomes the deficiencies of many traditional technologies. Since the advent of the R-CNN network in 2014, deep learning has made rapid progress in object detection. Then YOLO (Redmon et al., 2016), SSD (Liu et al., 2016), R-FCN (Dai et al., 2016), etc. continuously refresh the object detection accuracy level. In 2019, Google launched the EfficientDet family of models and feature fusion module called BiFPN (Tan et al., 2020). EfficientDet achieves state-of-the-art accuracy with fewer parameters compared to the previous object detection and semantic segmentation model. It contains a total of eight versions from D0 to D7. The best results can always be achieved under the constraints of the computing resources of different devices. At the same time, BiFPN also shows the best efficiency in multi-scale feature fusion. At present, the deep learning model based on EfficientDet and BiFPN is being applied to a variety of research fields, such as forest fire prevention (Xu et al., 2021), estimation of fashion landmarks (Kim et al., 2021), detection of garbage scattering areas (You et al., 2020), etc.

However, deep learning technology is not a universal method, and there will be problems in wheat ear detection and counting tasks. The species of wheat, for example, differ from other plants in that individual wheat plants have multiple ears. Therefore, there will be dozens of wheat ears in an image, which will cause serious occlusion problems (**Figure 1**). Occlusion and overlap will cause acute deviations in the detection and counting results of the model. In the study of Hasan et al. (2018) and Madec et al. (2019), counting accurately reached 86 and 91%, respectively. However, it seems that the occlusion and overlap of wheat cannot be effectively solved.

In this study, wheat ear counting adopts object detection method. So, the main objective is aimed at improving the EfficientDet-D0 model. In detecting and counting wheat ears, it focuses on addressing the problems of occlusion and overlap in the wheat ear images.

MATERIALS AND METHODS

In this study, the pipeline of the wheat ear counting algorithm based on the EfficientDet-D0 model is shown in **Figure 2**. The pipeline comprises four important parts: transfer learning, Random-Cutout image augmentation, attention module, and feature fusion module. First, the backbone network of the Effcientdet-D0 is separately trained utilizing transfer learning. Then Random-Cutout is used to augment the input images. After that, the attention module will refine the feature map output by the backbone network. Finally, feature fusion module fuses feature maps with different resolution and semantic information, followed by detection layer and Non-Maximum Suppression (NMS) to obtain the final detection results.

Dataset and Platform

The data used in this study are from the public data set called Global Wheat (David et al., 2020). Eight institutions lead the data set in seven countries: University of Tokyo, Arvalis, INRAE, University of Saskatchewan, ETH Zürich, University of

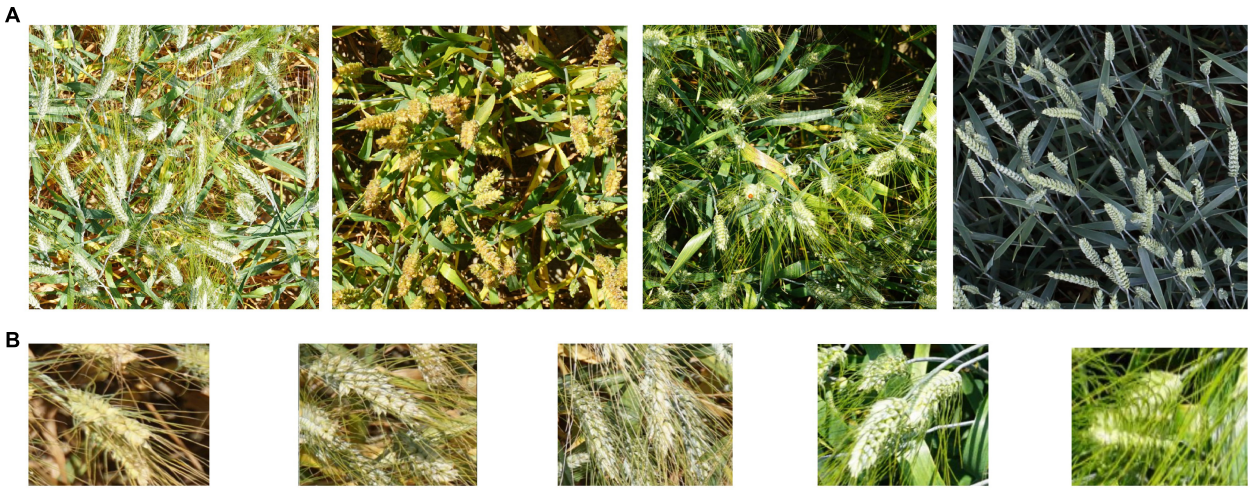

FIGURE 1 | Wheat images in the Global Wheat dataset. **(A)** Different varieties, light, maturity **(B)** Examples for occlusion and overlap.

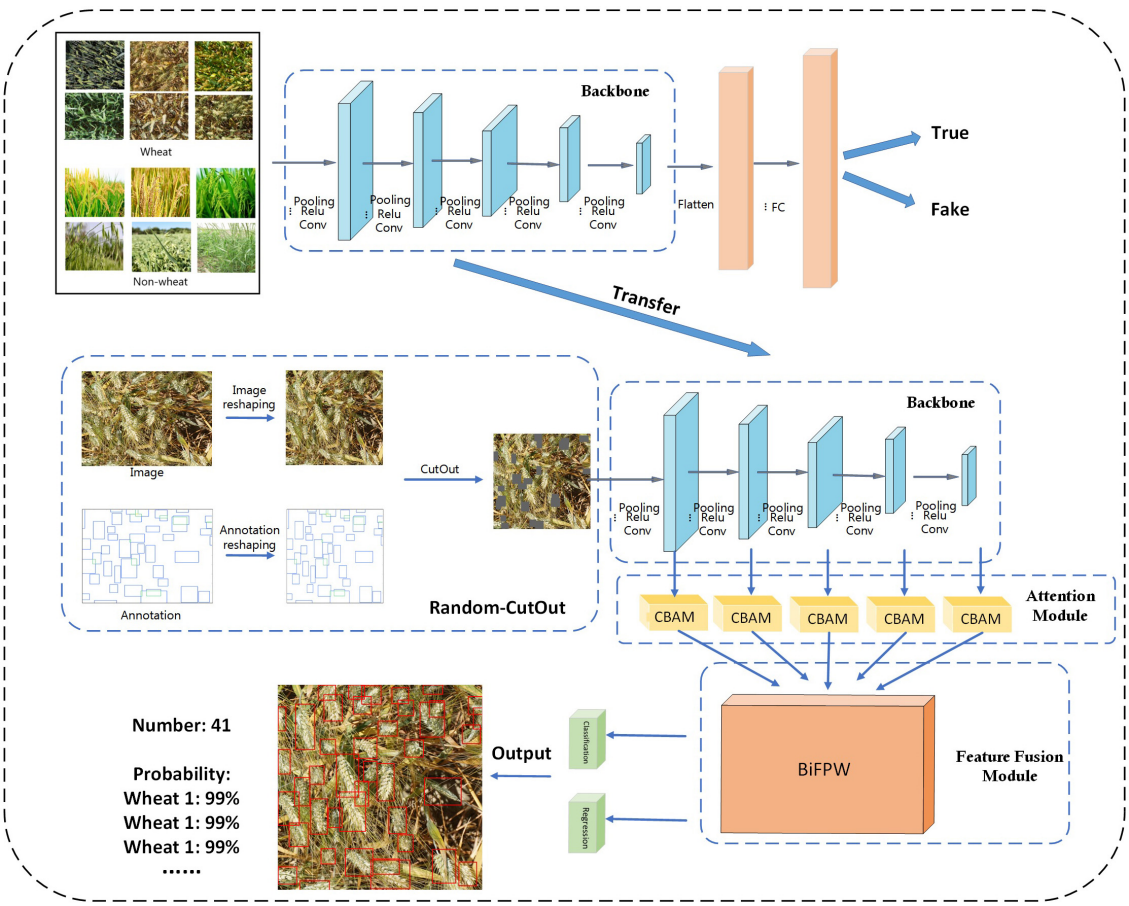

FIGURE 2 | The proposed pipeline for robust counting and detection of wheat eras.

Queensland, Nanjing Agricultural University, and Rothamsted Research. To better gauge the performance for unseen genotypes, environments, and observational conditions, this dataset covers multiple regions, including Europe (France, United Kingdom, Germany), North America (Canada), Asia (China, Japan), and Australia. All the 3,365 images were randomly split into training

set, validation set and test set without overlap. 2,693 (∼80%) images were selected as the training set, 336 images (∼10%) were used as the validation set, and the remaining 336 images (∼10%) were used as the test set. The performance of the final model is all obtained on the test set, and the data in the test set will never participate in training.

In this work, all models are trained and tested on the same device, which consists of an Intel E5-2603 V4 CPU, 1TB hard disk, and two Titan X graphics cards. The operating environment is Ubuntu16.0.4, tensorflow2.3.0 and Python3.7.

EfficientDet-D0 and BiFPN

The research in this article is based on DfficientDet-D0 object detection model. Its performance can surpass classic one-stage networks such as YOLOV3 and SSD, but its floating-point operations per second (FLPOS) is about 1/28 of homogeneous one-stage networks. Lightweight parameters enable DfficientDet-D0 to be easily deployed to hardware in practical applications, and the single inference time can satisfy the real-time counting work.

EfficientDet-D0 consists of two principal parts: the backbone network and the feature fusion module. Backbone is a model downstream module that is stacked by multiple MBConv for image feature extraction. Among them, the structure of MBConv is similar to the residual block, and effective features are extracted from the input through three steps. In the first step, MBConv uses 1×1 convolution to increase the dimension of the input. The second step is to extract the deep semantic features of the feature map with increased dimension by using depthwise separable convolution (Chollet, 2017). The third step is to integrate the input of MBConv with the deep semantic features generated in the second step as the final output.

A weighted feature fusion module BiFPN is proposed in the EfficientDet series model, shown in **Figure 3**. Compared with other superficial feature fusion layers such as FPN (Lin et al., 2017) and PANet (Wang et al., 2019), the weighted connection method is adopted inside BiFPN. All previous methods treat all input features equally, but different input features at different resolutions usually contribute unequally to the output features. Through 3×3 convolution and 1×1 convolution to achieve weighting of feature maps, the network model can learn the importance of different feature layers. This method makes multi-scale feature fusion more efficient. In CNN, low-level features contain more location and detailed information, but because less convolution layers are passed, they have less semantic information and more noise. The high-level features are full of semantic information, but the perception of details is poor. BiFPN combines the two features, making the feature map have the advantages of high-level feature maps and low-level feature maps. In the authentic wheat ear detection task, BiFPN enables the model to extract features at different scales. This significantly improves the model's multi-scale detection capabilities and detection capabilities in complex backgrounds.

After BiFPN has processed the feature map of wheat ears, each pixel of the feature map will be placed anchors. In EfficientDet-D0, the number of is usually set to 9, and these have different scales and aspect ratios. Then the classification layer model judges whether each anchor point contains background or wheat ears and returns the confidence. If the confidence is higher than 0.5, the regression layer will fine-tune the upper-left and lower-right coordinates of the anchor to make it closer to the real bounding box. The result at this time cannot be directly applied to the wheat ear counting. Since the detection process is based on an anchoring mechanism, the position of the same wheat ear usually corresponds to multiple overlapping candidate boxes. The NMS algorithm is to delete those duplicate candidate boxes. If the high-confidence candidate box is overlapped by some of the low-confidence, the low-confidence candidate boxes will be deleted.

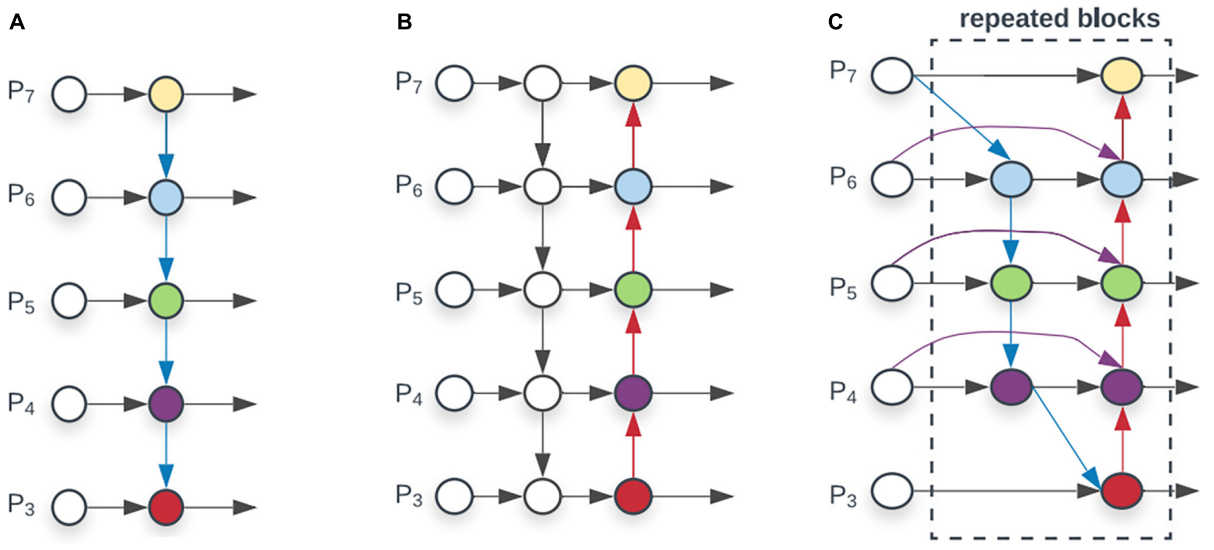

FIGURE 3 | Comparison of different feature fusion layer **(A)** FPN introduces a top-down pathway to fuse multi-scale features from level 3 to7 (P3-P7). **(B)** PANet adds bottom-up pathway on top of FPN. **(C)** BiFPN with better and efficiency trade-offs.

After NMS, the wheat ears will be independently labeled, and the number of detection results can be counted by the computer to complete the end-to-end wheat ear counting.

Backbone Training Using Transfer Learning

The predictive ability of the CNN model largely depends on the size of the data set. The more abundant the data, the better the CNN model's ability to extract image features. However, not every computer vision problem can obtain sufficient data. In this case, it is extremely difficult to train a model from scratch. Transfer learning provides a simpler and faster method. Before starting to train, the backbone of the CNN model is pre-trained on a huge data set. ImageNet (Shorten and Khoshgoftaar, 2019) is a commonly used transfer learning data set. It includes more than 14 million common images, which can provide sufficient materials for CNN training. The pre-trained backbone is sensitive to the features of the image. The trained backbone is then transferred back to the model and all parts of the model are fine-tuned using experimental data. In this way, an excellent CNN model is trained with a small amount of data.

However, there are domain gaps in the marginal distribution of ImageNet datasets and wheat datasets, and the task similarity is weak. Due to these differences, the backbone network pre-trained on the ImageNet dataset does not have a strong perception of the wheat ear features. Such a direct transfer learning method cannot get the best backbone in wheat ear detection. A serious domain shift cannot exist between learning data and training data, so a dataset was specially constructed for the pre-training backbone in this research. For a better description, this data set is defined as D1, and the wheat ear data is defined as D0. Data set D1 consists of two parts, D0 and non-wheat data. Non-wheat data includes 2,256 rice images, 561 oat images, and 274 drilgrass images. The appearance of these three crops is very close to wheat. The D1 dataset is used to train the classification task of the EffcientDet-D0 backbone with fully connected (FC) layer and classification layer (**Figure 2**). The goal of classification is to distinguish whether the image is wheat. It is not easy to accurately classify these crops, not only does the backbone need to be sensitive to simple features, but it also needs to have a strong perception of high-level semantic features of wheat ears. **Figure 4** shows the output of the middle layer of the backbone and the classification results.

Wheat Ear Counting Under Occlusion Condition

Occlusion and overlap are the primary problems faced in wheat ear detection and counting. To improve the detection accuracy, these problems must be considered in the algorithm design. This article proposes an effective solution to solve the occlusion and overlap in wheat ear detection. First, in the image preprocessing stage, Random-Cutout is used to augment the image so that the model can fully learn these tricky occlusion areas. Secondly, in the model, the adoption of the CBAM attention module can refine the features of occluded wheat ears; therefore, it makes the model detect the wheat ears from the cluttered background, while reducing the interference of background and occlusion areas.

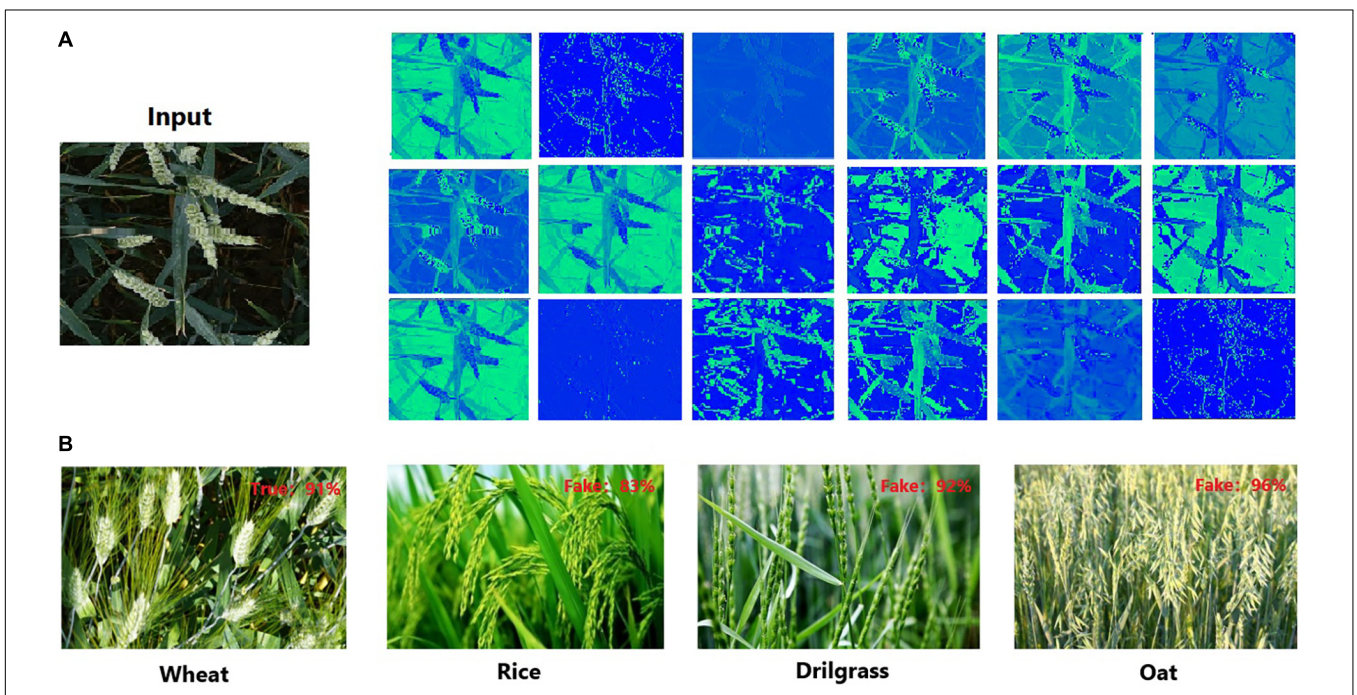

FIGURE 4 | Backbone pre-training results. **(A)** The first 24 features of the fifth convolution layers for backbone. **(B)** The classification results for backbone with full connection layer and classification layer.

Random-Cutout for Occlusion Image Augmentation

To broaden the diversity of samples and increase the model's priori knowledge of the occlusion problem, an image augmentation method is proposed for dense object detection. In an image of wheat ears, the occurrence of occlusion and overlap is often related to the distribution of wheat ears. In order to simulate the occlusion under real conditions better, some rectangles are randomly erased. In the existing approaches, such as Cutout (Devries and Taylor, 2017) and Random Erasing (Zhong et al., 2020), the completely random positions of a fixed size of the image were occluded. If these methods are applied to wheat ear counting, a few wheat ears in the image may be totally occluded and these areas will be processed as noise data (**Figure 5**).

Random-Cutout generates occlusion area randomly to meet the wheat ear growth distribution and avoid the negative effects of excessive and insufficient occlusion on model training. Depending on the distribution of real occlusion in the images, the proposed Random-Cutout algorithm combines position and size information to generate the simulated occlusion area. In terms of the location, the probability of occlusion in dense areas of wheat ears is much higher than that in sparse areas. However, it is important to emphasize that this does not mean that occlusion does not occur in sparse areas. Occlusion and overlap are also commonly associated with wheat leaves and stems. In terms of size of random occlusion, the core is to occlude wheat ears effectively without completely losing the context information. In the wheat ear dataset, the wheat ear scales in images with different field of vision are greatly different, which means that the occlusion size generated by the algorithm cannot be set to a fixed value. When the occlusion size of a large wheat ear is applied to the images of small scales wheat ears, a lot of valid context information in the image will be erased directly. Therefore, the occlusion size generated by the Random-Cutout should be adjusted adaptively according to the size of the wheat ears in the current image.

The flowchart of the Random-Cutout is shown in **Figure 6**. First, Probability Map is generated according to the distribution of wheat ears in the images to determine the approximate location of the simulated occlusion. The value of each pixel is defined as a probability value I, in which the value of the cold color area is low, and the value of the warm color is high. Next, Center Point Proposal is generated according to the Probability Map. At this time, there may be hundreds or thousands of candidate center points, and the total number of them needs to be adjusted to a suitable value N. It is necessary to randomly select N center points from all Center Point Proposal according to the number of objects in the images. Finally, a rectangle of random length H and width W is initialized from these center points and superimposed to the original images. H and W are closely linked to the size of wheat ears in the image. We conducted a lot of experiments to determine the settings of the above parameters, which are shown in **Table 1**.

CBAM for Refining Features of Partially Occluded Wheat Ears

Using the visual attention mechanism in multi-object detection model is an effective way to overcome occlusion and overlap problems. The attention module concentrates "Resources" on salience areas of the image and extracts global information from these fine-grained features. Therefore, the model can quickly filter out unwanted information and focus on the region of interest (Laskar and Kannala, 2017).

Convolutional block attention module (Woo et al., 2018) is one of the most effective attention modules. CBAM refines the feature map by calculating the weight of the features in space domain and channel domain (**Figure 7**). For feature map $F \in \mathbb{R}^{W \times H \times C}$, each channel can be regarded as a feature in the images extracted by CNN. By aggregating the relations between channels in the feature map, channel attention module can obtain the "what" features that should be paid attention to in the images. Channel attention module first uses global average pooling and global max pooling operations to generate two different channel context descriptors: F^c_{avg} and F^c_{max}, which represent average-pooled features and max-pooled features. Then these two features are input into a weight sharing module to generate a channel attention vector $M_C \in \mathbb{R}^{C \times 1}$. The weight sharing module is a multilayer perceptron (MLP) with hidden layer. The hidden layer size is set to $\mathbb{R}^{C/r \times 1}$, where, r is the scaling factor. After applying the shared network to each descriptor, the attention feature is generated by element-wise summation. Equation 1 shows how channel-wise attention is generated (Woo et al., 2018):

$$\begin{aligned} M_c(F) &= \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)\right) \end{aligned} \quad (1)$$

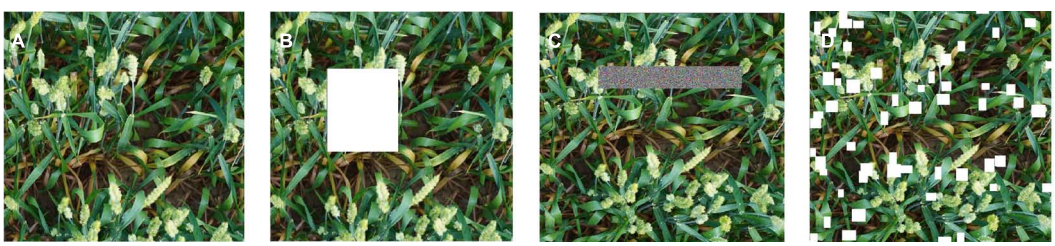

FIGURE 5 | Erasing illustration for different methods: **(A)** input image, **(B)** cutout (erased rectangles marked in white), **(C)** random Erasing (erased rectangles marked with random noise), and **(D)** Random-Cutout (erased rectangles marked in white).

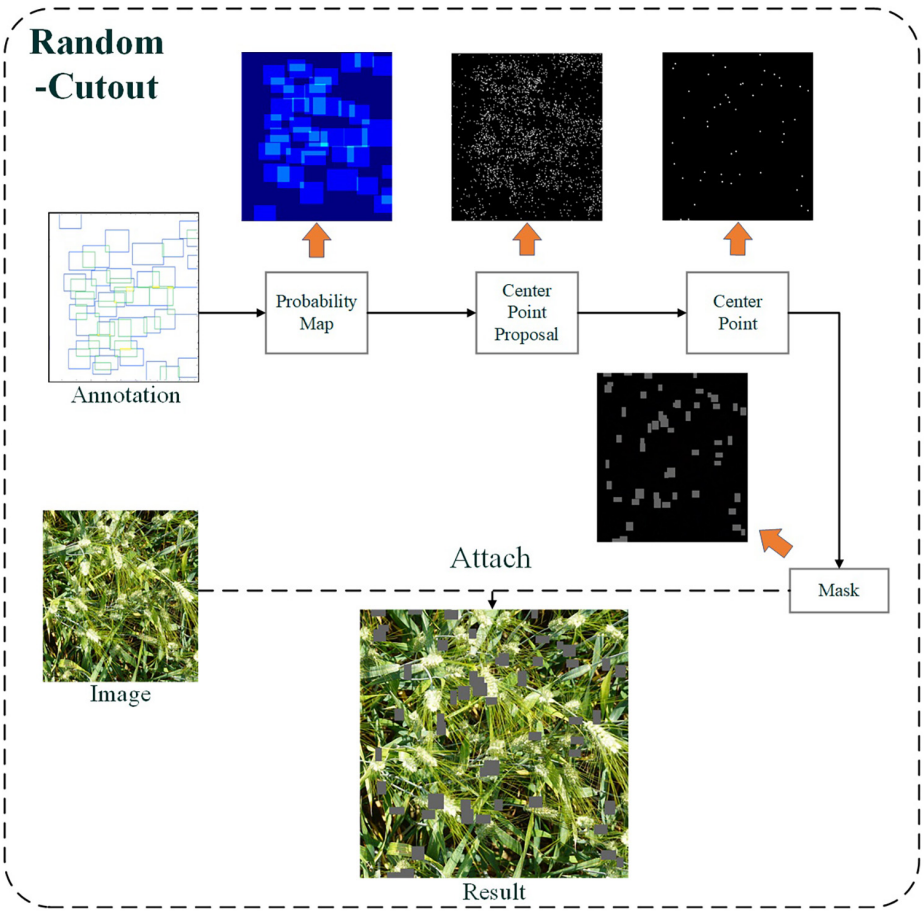

FIGURE 6 | The schematic layout of the Random-Cutout.

TABLE 1 | The setting of hyper-parameters in Random-Cutout in this research.

	Definition of parameters	Mathematical definition
I	The initial probability value of each pixel point is 0.001. When wheat ears exist at this pixel point, the probability value of the pixel point is the initial value plus the number of wheat ears multiplied by 0.003	$I_i = 0.0010.003 \times n_i$
N	1/4 of the number of wheat ears in the image	$N = N_{total}/4$
H	Random number between one quarter of the minimum length and one quarter of the maximum length of the wheat ear in the image	$\forall H \in \bigcup\left(\frac{1}{4} \times H_{min}, \frac{1}{4} \times H_{max}\right)$
W	Random number between one quarter of the minimum width and one quarter of the maximum width of the wheat ears in the image	$\forall W \in \bigcup\left(\frac{1}{4} \times W_{min}, \frac{1}{4} \times W_{max}\right)$

where, σ represents the sigmoid function, $W_0 \in \mathbb{R}^{C/r \times C}, W_1 \in \mathbb{R}^{C/r \times C/r}$ indicates the weight of MLP, W_0, W_1 share two inputs and ReLU activate function.

After generating the attention on the channel, the spatial attention can be generated through the pooling operation. Compared with channel-wise attention, spatial-wise attention is constructed more explicitly. The purpose of the spatial attention module is to obtain the prominent region in the image, that is, "where" the image needs to be paid attention to. The spatial attention module first uses max pooling and average pooling along the direction of the feature map channel to obtain two spatial descriptors: F^s_{avg} and F^s_{max}. In order to have a larger spatial receptive field for the two descriptors, a larger pool filter is usually used in this step, e.g., 7×7, 15×15. After that, spatial attention module concatenates two spatial descriptors and uses a convolution layer to generate spatial-wise attention, Equation 2 shows how spatial-wise attention is generated (Woo et al., 2018):

$$\begin{aligned} \mathrm{M}_s(F) &= \sigma\left(f^{7\times7}\left(\left[AvgPool(F); MaxPool(F)\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \end{aligned} \tag{2}$$

where, σ represents the sigmoid function. $f^{7\times7}$ represents a convolution with a convolution kernel size of 7×7.

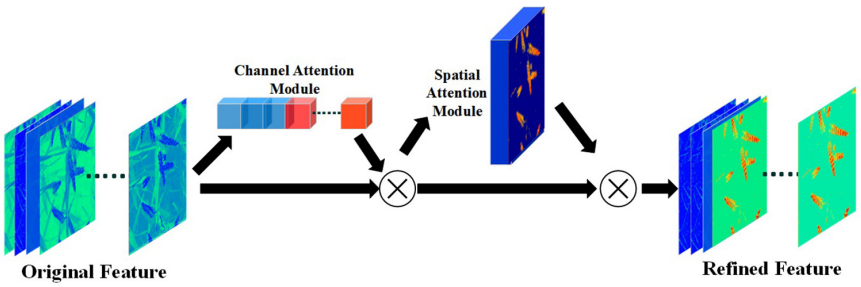

FIGURE 7 | The process of CBAM module generating channel attention and spatial attention.

After generating channel-wise attention and spatial-wise attention, the feature map can be refined twice by element-wise multiplication, this process can be described as Equation 3:

$$F' = M_c(F) \otimes F$$
$$F'' = M_s(F') \otimes F' \quad (3)$$

where, F' and F'' represent the first and second refinement results of the feature map, respectively, $\otimes$ represent element-wise multiplication.

The wheat ears are distributed in a messy background, therefore, CBAM play an extraordinary role. In this study, five CBAM are added between the EffcientDet-D0 backbone and the feature fusion layer BiFPN (**Figure 2**). Five feature maps of different scales outputted by the backbone network will be used as the training input in the attention modules, so that the model can effectively get the features of different spatial information and semantic information.

Criteria for Performance Evaluation

Evaluation indicators are objective evaluation criteria for the results of the algorithm. In different tasks, the evaluation indicators are different. In this study, counting accuracy rate (P), false detection rate (O), and frames per second (FPS) are used as performance indicators. Counting accuracy rate is the ratio between the correct number of wheat ears and the actual number of wheat ears, while false detection rate is the ratio of the number of wheat ears detected incorrectly to the total number detected. Equation 4 gives the definition of these two evaluation criteria.

$$P = \frac{N_{cor}}{N_{real}}, O = \frac{N_{err}}{N_{num}} \quad (4)$$

where, N_{cor} is the number of wheat ears that the model detects correctly, and N_{err} is the number of errors detected by model. N_{real} represents the actual number of wheat ears in the test image. N_{num} represents the total number detected by the model.

frames per second is an index to evaluate the inference speed of the model, which indicates how many images the model can process per second. Usually only when the FPS reaches 24 or more, this model is possible to achieve real-time detection. FPS

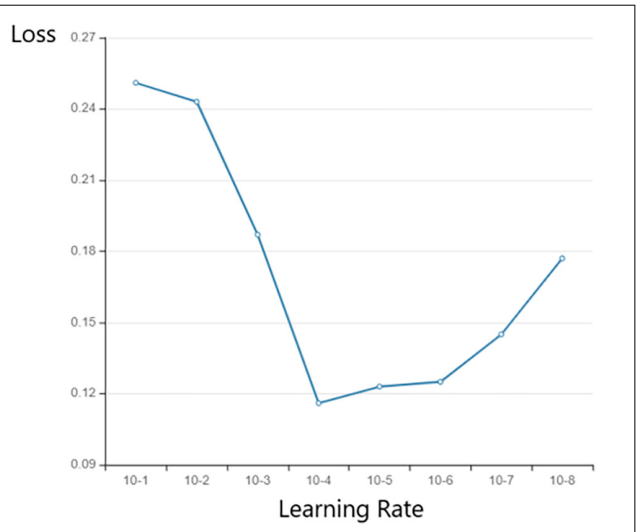

FIGURE 8 | The effect of learning rate on loss.

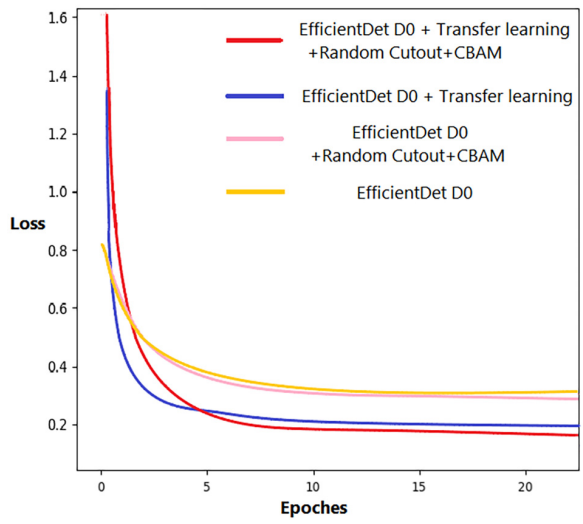

FIGURE 9 | Loss function curve.

is defined as shown in Equation 5:

$$FPS = \frac{1}{T} \tag{5}$$

where, T denotes the time used by the model to infer the image.

Hyper-Parameter Configuration and Learning Rate Optimization

In order to ensure reasonableness, the same hyper-parameters are set in the comparison experiments. The stochastic gradient descent (SGD) method is used to optimize the training of the loss function. Batch size and epoch are set to 12 and 300, respectively. The learning efficiency will be reduced by 50% every 30 iteration. At the same time, to prevent over-fitting, an early stopping strategy is set. When the loss of the verification dataset does not reduce or rise in 5 iterations, then the training will stop early.

Learning rate controls the speed of gradient descent during CNN training (Equation 6). If the learning rate is set too small, the convergence process of the model will be slow. If the learning rate is set too large, the gradient will oscillate repeatedly near the

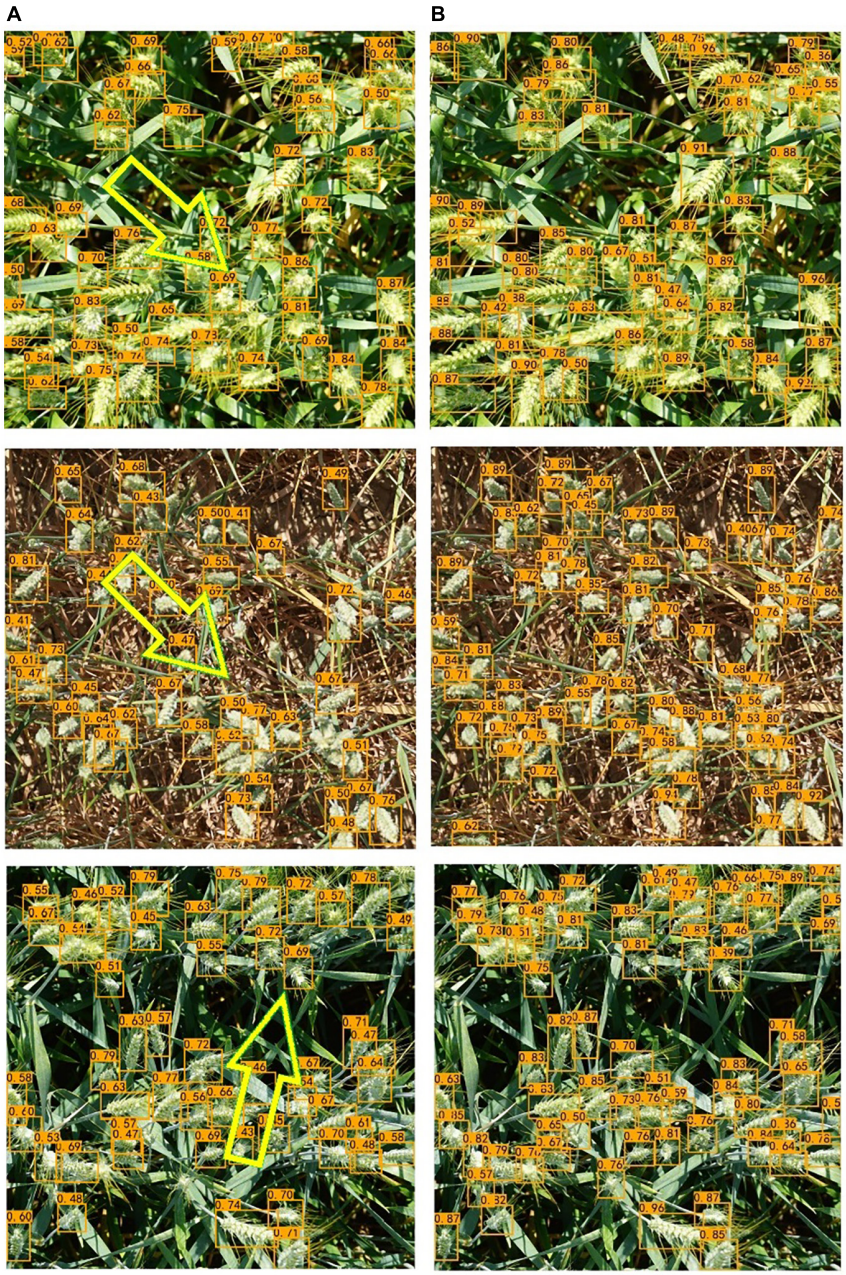

FIGURE 10 | Detection results of the two networks. **(A)** EfficientDet-D0 (Some obvious missed detections are highlighted by yellow arrows). **(B)** The improved EfficientDet-D0.

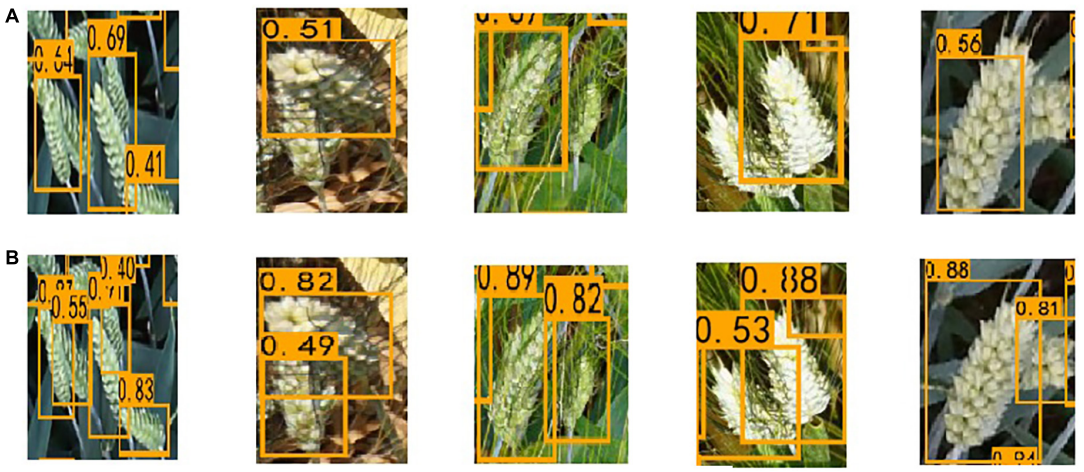

FIGURE 11 | The results of the two networks on the overlap. **(A)** EfficientDet-D0. **(B)** The improved EfficientDet-D0.

minimum or even fail to converge.

$$\theta_{i+1} = \theta_i - \alpha \frac{\partial}{\partial \theta_i} L(\theta_i) \quad (6)$$

where, θ_i represents the parameters that need to be updated during the i-th iteration, α represents the learning rate, and L represents the loss function.

In this study, we compared the influence of different levels of learning rate on the final loss value of the model (**Figure 8**) and found that the learning rate is optimal under the order of 10e-4.

It should be noted that the loss function (L) consists of two parts, classification loss function (L_{class}) and regression loss function (L_{reg}), as shown in Equation 7. The purpose of optimizing L_{class} is to allow the network to distinguish wheat ears and background, and the purpose of optimizing L_{reg} is to enable the network to locate these wheat ears accurately.

$$L = L_{class} + L_{reg} \quad (7)$$

RESULTS AND ANALYSIS

In this section, comparative experiments are first done to show that the modifications, such as transfer learning, image augmentation and CBAM, works in performance promotion. Then comparative experiments are done to show the superiority of the proposed algorithm.

Performance Comparison With EfficientDet-D0

In the comparison experiments, we first compared the improved EfficientDet-D0 with the original one. **Figure 9** shows the loss function curve of the model in four cases during the training process. To make the difference obvious, the curves in the figure are smoothed. Regardless of whether the improved model is under transfer learning conditions, the loss value is greatly reduced. It can be seen that the transfer learning method also played a role in the experiment. The loss of the model with and without transfer learning is reduced by 0.101 and 0.122, respectively. In terms of detection ability, our improved model has been significantly improved. The detection results of EfficientDet-D0 show that there are many omissions in the intensive area, but the number of missed wheat ears with our model is significantly reduced (**Figure 10**).

At the same time, it is found that the improved EfficientDet-D0 model has dramatically reduced the impact of occlusion on the detection results. Before the improvement, the model distinguished multiple adjacent wheat ears into one, which was most serious in the dense area of wheat ears. The proposed method greatly overcomes this drawback. We selected several severely occluded images in the data set and tested them on two models, respectively; the results are shown in **Figure 11**. The results show whether occlusion between the wheat leaves and ears or overlap between wheat ears, the proposed network has been dramatically improved (**Table 2**).

To visualize the difference of the improved EfficientDet-D0 and the original one, the Class Activation Mapping (CAM) (Zhou et al., 2016) is used to show the difference in network feature extraction (**Figure 12**). The thermodynamic features of different colors reveal the "attractiveness" of the regional network. Among them, the red area represents the most significant influence on the network. As the color changes from red to yellow, and finally to blue, it means that the influence gradually decreases.

TABLE 2 | The missed detection rate of the two models under types of occlusion.

Types of occlusions	EfficientDet-D0 (%)	Proposed (%)
Overlap between wheat ears and wheat ears (265 images)	13.8	8.7
Leaves cover wheat ears (112 images)	5.2	3.3
Wheat ears were not fully photographed (81 images)	2.1	0.9

TABLE 3 | Peformance comparison of different CNN methods.

Method	Transfer learning	Backbone	O (Average ± STD) (%)	P (Average ± STD) (%)	FPS
YOLOv3	×	Darknet-53	7.3 ± 0.57	90.3 ± 0.46	23
SSD	×	VGG-16	8.6 ± 0.86	88.1 ± 0.14	22
Faster-RCNN	×	Resnet-50	6.3 ± 0.21	91.1 ± 0.53	16
EfficientDet-D1	×	EfficientNet-B1	6.5 ± 0.55	91.6 ± 0.45	27
EfficientDet-D0	×	EfficientNet-B0	6.7 ± 0.46	90.8 ± 0.87	**35**
Proposed	×	EfficientNet-B0	6.3 ± 0.33	92.9 ± 0.07	30
EfficientDet-D1	√	EfficientNet-B1	6.1 ± 0.14	93.1 ± 0.35	27
EfficientDet-D0	√	EfficientNet-B0	6.4 ± 0.77	92.5 ± 0.28	**35**
Proposed	√	EfficientNet-B0	**5.8 ± 0.12**	**94.2 ± 0.19**	30

The bold font represents the best result in the experiment.

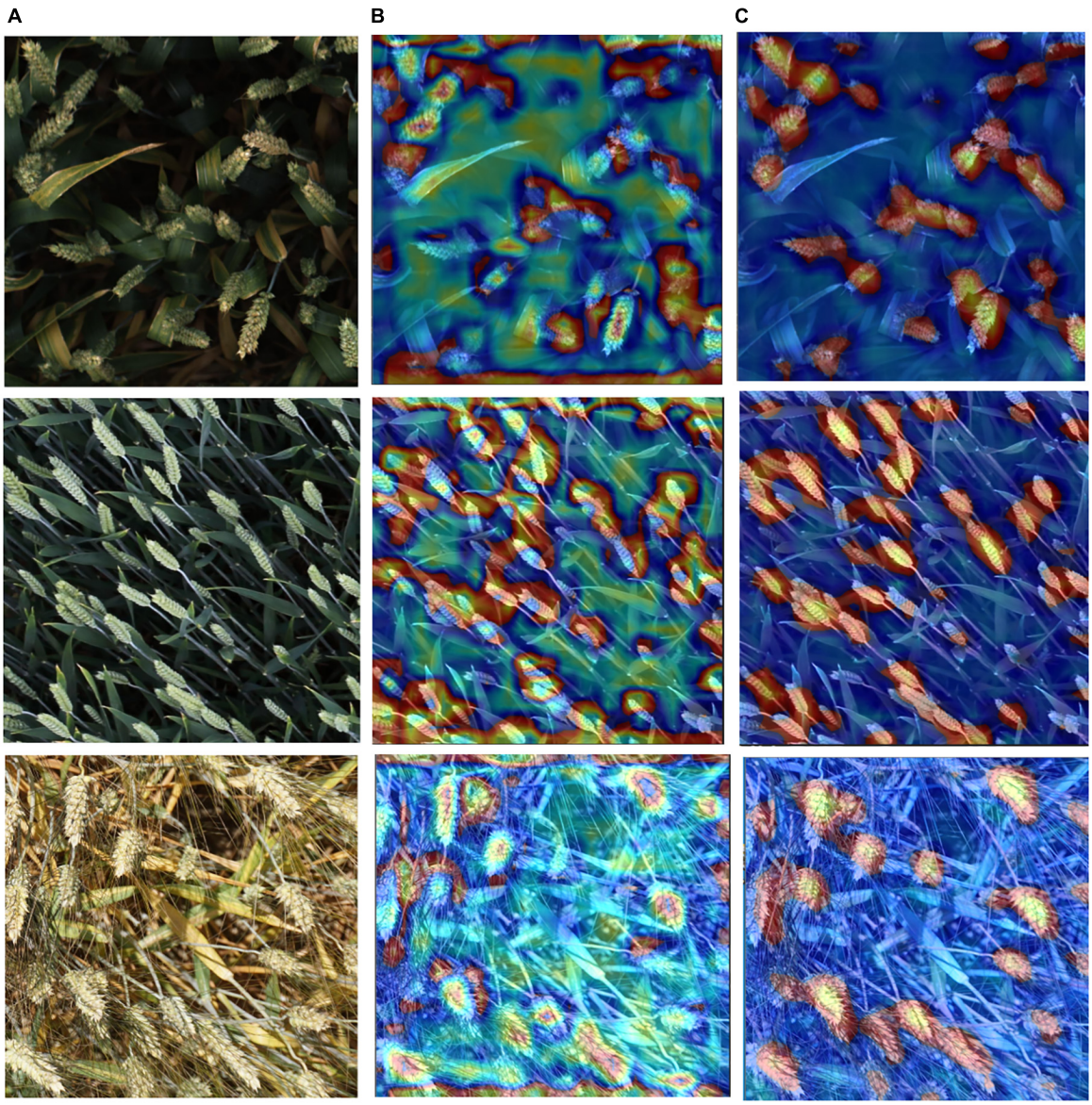

FIGURE 12 | The CAM of Test images using two networks. **(A)** Test images. **(B)** The CAM of test images with EfficientDet-D0. **(C)** The CAM of test images with the improved EfficientDet-D0.

Performance Comparison of Different CNN Methods

By displaying some results in **Figure 10**, it can be observed that both original EfficientDet-D0 and the improved one has high ability to detect wheat ears under different lighting, background, and scales, which shows the advantages of CNN in such problems. Therefore, in order to evaluate the improved model more comprehensively, we compared it with other CNNs. In previous counting studies, models such as Faster-RCNN, YOLOV3, SSD are often used (Xu C. et al., 2020). We have compared the proposed method with these models and the results are shown in **Table 3**. According to the results, although the YOLOV3 and SSD models can achieve real-time detection in forwarding inference, it has a high false detection rate and a little effect on dense multi-object detection tasks. It cannot complete the task of detecting and counting wheat ears well. Faster-RCNN is a classic two-stage neural network. The counting accuracy rate of Faster-RCNN is 0.3% higher than that of Efficientdt-D0, and the false rate is 0.4% lower. But its accuracy is still about 1.3% lower than our model and its inference time is the longest.

We also did some experiments to compare the higher version of EfficientDet-D0 (i.e., EfficientDet-D1). The accuracy of the EfficientDet-D1 increased by 0.7% compared with the EfficientDate-D0 model and the improved EfficientDet-D0 increased by 1.6% (**Table 3**).Since EfficientDet-D1 is a general-purpose object detection model that improves accuracy by expanding the size of the backbone and feature fusion modules to extract better feature expressions, this results in a decrease in the effective inference speed of EfficientDet-D1 by 22%. In contrast, the improved EfficentDet-D0 model was designed specifically for wheat ear detection to improve accuracy by reducing occlusion interference. CBAM reduced the inference speed by about 15%, but this was the tradeoff with the improvement in accuracy. In terms of false detection rate, the improved EfficientDet-D0 is 0.3% lower than EfficientDet-D1 and 0.6% lower than EfficientDet-D0. Although the accuracy increases, the error rate of the improved model does not decrease significantly. The reason is to ensure that as many ears as possible are detected in the post-processing process, the confidence threshold is usually set to a small value, which will cause some proposed regions that do not contain ears to be leaked.

From the results in section "Performance Comparison *With* EfficientDet-D0" and section "Performance Comparison of Different CNN Methods," it can be seen that transfer learning is an effective strategy in wheat ear detection. In transfer learning, the data do not need to be finely labeled, and only the categories they belong to are roughly labeled. After using transfer learning, the false detection rate of EfficientDet-D1, EfficientDet-D0 and the improved EfficientDet-D0 was reduced by 0.4%, 0.3% and 0.5%, while the counting accuracy rate was increased by 1.5, 1.7, and 1.3%, respectively.

CONCLUSION

In this article, we proposed a novel wheat ear counting algorithm. Importantly, we focus on the occlusion and overlap problems that exist under the actual growth conditions of wheat ears. Farmers and breeders take images of wheat under a certain area in the wheat field and our proposed algorithm can automatically calculate the number of wheat ears in that area, which is helpful to evaluate and predict the level of wheat yield.

The main contributions come from the three key procedures of the proposed method. First, the transfer learning method is employed to extract the high-level semantic features of wheat ears. Secondly, an image augmentation method Random-Cutout is proposed to simulate occlusion in real wheat images. Finally, convolutional block attention module (CBAM) is adopted into the EfficientDet-D0 model to refine the features and pay more attention to the wheat ears.

Extensive experiments show that the counting accuracy of the proposed algorithm reaches 94% and false detection rate is 5.8%. The performance evaluation shows that the proposed method is invariant to illumination and scale changes. Simultaneously, the proposed method had high accuracy and strong robustness for occlusion and overlap problem. We firmly believe that human beings will benefit from automatic wheat ear counting by machines, thereby reducing manual counting errors. Moreover, it greatly reduces the labor cost. The proposed model can be used as a post-processing method to plan the wheat harvesting and storage.

The methods used in this research can achieve accurate counting of wheat ears, but the research will never stop here. In the future, we will envisage using this method in more crop counting work such as apple counting, etc. Moreover, we will apply the Random-Cutout image augmentation method to more fields, not limited to agriculture, to prove its robustness to solve the occlusion problem.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YW performed the experiments and analyzed the results. YQ designed and performed the experiments, and wrote the manuscript. JC wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partly supported by the National Natural Science Foundation of China (NSFC No. 61673021), National Natural Science Foundation of China (NSFC No. 62071006), National Key R&D Program of China (NKRDPC No. 2017YFB0802300), and Hangzhou Innovation Institute, Beihang University (No. 2020-Y3-A-014).

REFERENCES

Aich, S., and Stavness, I. (2017). "Leaf Counting with Deep Convolutional and Deconvolutional Networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (Venice: Institute of Electrical and Electronics Engineers), 2080–2089.

Barré, P., Stöver, B. C., Müller, K. F., and Steinhage, V. (2017). LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics* 40, 50–56. doi: 10.1016/j.ecoinf.2017.05.005

Bleau, A., and Leon, L. J. (2000). Watershed-based segmentation and region merging. *Comput. Vis. Image Underst.* 77, 317–370. doi: 10.1006/cviu.1999.0822

Chollet, F. (2017). "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: Institute of Electrical and Electronics Engineers), 1800–1807.

Dai, J., Li, Y., He, K., and Sun, J. (2016). "R-FCN: Object Detection via Region-based Fully Convolutional Networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Vol. Vol. 29 (Barcelona), 379–387.

David, E., Madec, S., Sadeghi-Tehran, P., Aasen, H., Zheng, B., Liu, S., et al. (2020). Global Wheat Head Detection (GWHD) Dataset: A Large and Diverse Dataset of High-Resolution RGB-Labelled Images to Develop and Benchmark Wheat Head Detection Methods. *Plant Phenomics (Washington, D.C.)* 2020, 3521852.

Devries, T., and Taylor, G. W. (2017). Improved Regularization of Convolutional Neural Networks with Cutout. *ArXiv [Preprint]* ArXiv:1708.04552,

Dobrescu, A., Giuffrida, M. V., and Tsaftaris, S. A. (2017). "Leveraging Multiple Datasets for Deep Leaf Counting," in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (Venice: Institute of Electrical and Electronics Engineers), 2072–2079.

Ferrante, A., Cartelle, J., Savin, R., and Slafer, G. A. (2017). Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. *Field Crops Research* 203, 114–127. doi: 10.1016/j.fcr.2016.12.028

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 14* (Columbus, OH: Institute of Electrical and Electronics Engineers), 580–587.

Giuffrida, M. V., Minervini, M., and Tsaftaris, S. A. (2015). "Learning to Count Leaves in Rosette Plants," in *Proceedings of the Computer Vision Problems in Plant Phenotyping Workshop 2015* (London).

Grbovic, Z., Panic, M., Marko, O., Brdar, S., and Crnojevic, V. S. (2019). "Wheat Ear Detection in RGB?: and Thermal Images Using Deep Neural Networks," in *Proceedings of the 15th International Conference on MachineLearning and Data Mining, MLDM 2019*, Vol. 2 (New York, NY), 875–889.

Hasan, M., Chopin, J. P., Laga, H., and Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods* 14, 1–13.

Kim, H. J., Lee, D. H., Niaz, A., Kim, C. Y., Memon, A. A., and Choi, K. N. (2021). Multiple-Clothing Detection and Fashion Landmark Estimation Using a Single-Stage Detector. *IEEE Access* 9, 11694–11704. doi: 10.1109/access.2021.3051424

Laskar, Z., and Kannala, J. (2017). "Context aware query image representation for particular object retrieval," in *Scandinavian Conference on Image Analysis* (Cham: Springer).

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature Pyramid Networks for Object Detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: Institute of Electrical and Electronics Engineers), 936–944.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. E., Fu, C.-Y., et al. (2016). "SSD: Single Shot MultiBox Detector," in *Proceedings of the 14th European Conference on Computer Vision, ECCV 2016* (Amsterdam), 21–37.

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA: Institute of Electrical and Electronics Engineers), 3431–3440.

Lu, H., and Cao, Z. (2020). TasselNetV2+: A Fast Implementation for High-Throughput Plant Counting From High-Resolution RGB Imagery. *Frontiers in Plant Science* 11:541960.

Madec, S., Jin, X., Lu, H., Solan, B. D., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agricultural and Forest Meteorology* 264, 225–234. doi: 10.1016/j.agrformet.2018.10.013

Maldonado, W., and Barbosa, J. C. (2016). Automatic green fruit counting in orange trees using digital images. *Computers and Electronics in Agriculture* 127, 572–581. doi: 10.1016/j.compag.2016.07.023

Misra, T., Arora, A., Marwaha, S., Chinnusamy, V., Rao, A. R., Jain, R., et al. (2020). SpikeSegNet-a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* 16, 40.

Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science* 7:1419.

Mussadiq, Z., Laszlo, B., Helyes, L., and Gyuricza, C. (2015). Evaluation and comparison of open source program solutions for automatic seed counting on digital images. *Computers and Electronics in Agriculture* 117, 194–199. doi: 10.1016/j.compag.2015.08.010

Peltonen-Sainio, P., Kangas, A., Salo, Y., and Jauhiainen, L. (2007). Grain number dominates grain weight in temperate cereal yield determination: Evidence based on 30 years of multi-location trials. *Field Crops Research* 100, 179–188. doi: 10.1016/j.fcr.2006.07.002

Prystupa, P., Savin, R., and Slafer, G. A. (2004). Grain number and its relationship with dry matter, N and P in the spikes at heading in response to N×P fertilization in barley. *Field Crops Research* 90, 245–254. doi: 10.1016/j.fcr.2004.03.001

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: Institute of Electrical and Electronics Engineers), 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 1137–1149. doi: 10.1109/tpami.2016.2577031

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer International Publishing), 234–241.

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., and Hawkesford, M. J. (2019). DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Frontiers in Plant Science* 10:1176.

Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* 6, 1–48.

Stein, M., Bargoti, S., and Underwood, J. P. (2016). Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. *Sensors* 16, 1915. doi: 10.3390/s16111915

Tan, M., Pang, R., and Le, Q. V. (2020). "EfficientDet: Scalable and Efficient Object Detection," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA: Institute of Electrical and Electronics Engineers), 10781–10790.

Wang, K., Liew, J. H., Zou, Y., Zhou, D., and Feng, J. (2019). "PANet: Few-Shot Image Semantic Segmentation With Prototype Alignment," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul: Institute of Electrical and Electronics Engineers), 9197–9206.

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "CBAM: convolutional block attention module[C]," in *European Conference on Computer Vision* (Cham: Springer).

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., and Shen, C. (2019). TasselNetv2: in-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* 15, 1–14.

Xu, C., Jiang, H., Yuen, P. W. T., Ahmad, K. Z., and Chen, Y. (2020). MHW-PD: A robust rice panicles counting algorithm based on deep learning and multi-scale hybrid window. *Computers and Electronics in Agriculture* 173, 105375. doi: 10.1016/j.compag.2020.105375

Xu, R., Lin, H., Lu, K., Cao, L., and Liu, Y. (2021). A Forest Fire Detection System Based on Ensemble Learning. *Forests* 12, 217. doi: 10.3390/f12020217

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., et al. (2020). Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant Methods* 16, 106–106.

You, T., Chen, W., Wang, H., Yang, Y., and Liu, X. (2020). Automatic Garbage Scattered Area Detection with Data Augmentation and Transfer Learning in SUAV Low-Altitude Remote Sensing Images. *Mathematical Problems in Engineering* 2020, 1–13. doi: 10.1155/2020/7307629

Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. (2020). "Random Erasing Data Augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. Vol. 34, 13001–13008. doi: 10.1609/aaai.v34i07.7000

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning Deep Features for Discriminative Localization," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: Institute of Electrical and Electronics Engineers), 2921–2929.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Qin and Cui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fpls.2021.683078

Investigating a Selection of Methods for the Prediction of Total Soluble Solids Among Wine Grape Quality Characteristics Using Normalized Difference Vegetation Index Data From Proximal and Remote Sensing

Aikaterini Kasimati, Borja Espejo-Garcia, Eleanna Vali, Ioannis Malounas and Spyros Fountas*

Natural Resources Management and Agricultural Engineering, Agricultural University of Athens, Athens, Greece

OPEN ACCESS

Edited by:
Maria Balota, Virginia Tech, United States

Reviewed by:
Francesco Marinello, University of Padua, Italy
Alexandr Muterko, Institute of Cytology and Genetics, Russian Academy of Sciences (RAS), Russia

***Correspondence:**
Aikaterini Kasimati
akasimati@aua.gr

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *19 March 2021*
Accepted: *17 May 2021*
Published: *11 June 2021*

Citation:
Kasimati A, Espejo-Garcia B, Vali E, Malounas I and Fountas S (2021) Investigating a Selection of Methods for the Prediction of Total Soluble Solids Among Wine Grape Quality Characteristics Using Normalized Difference Vegetation Index Data From Proximal and Remote Sensing. Front. Plant Sci. 12:683078. doi: 10.3389/fpls.2021.683078

The most common method for determining wine grape quality characteristics is to perform sample-based laboratory analysis, which can be time-consuming and expensive. In this article, we investigate an alternative approach to predict wine grape quality characteristics by combining machine learning techniques and normalized difference vegetation index (NDVI) data collected at different growth stages with non-destructive methods, such as proximal and remote sensing, that are currently used in precision viticulture (PV). The study involved several sets of high-resolution multispectral data derived from four sources, including two vehicle-mounted crop reflectance sensors, unmanned aerial vehicle (UAV)-acquired data, and Sentinel-2 (S2) archived imagery to estimate grapevine canopy properties at different growth stages. Several data pre-processing techniques were employed, including data quality assessment, data interpolation onto a 100-cell grid (10 × 20 m), and data normalization. By calculating Pearson's correlation matrix between all variables, initial descriptive statistical analysis was carried out to investigate the relationships between NDVI data from all proximal and remote sensors and the grape quality characteristics in all growth stages. The transformed dataset was then ready and applied to statistical and machine learning algorithms, firstly trained on the data distribution available and then validated and tested, using linear and nonlinear regression models, including ordinary least square (OLS), Theil–Sen, and the Huber regression models and Ensemble Methods based on Decision Trees. Proximal sensors performed better in wine grapes quality parameters prediction in the early season, while remote sensors during later growth stages. The strongest correlations with the sugar content were observed for NDVI data collected with the UAV, Spectrosense+GPS (SS), and the CropCircle (CC), during Berries pea-sized and the Veraison stage, mid-late season with full canopy growth, for both years. UAV and SS data proved to be more accurate in predicting the sugars out of all wine grape quality characteristics, especially during a mid-late season with full canopy

growth, in Berries pea-sized and the Veraison growth stages. The best-fitted regressions presented a maximum coefficient of determination (R^2) of 0.61.

Keywords: normalized difference vegetation index, linear regression, ensemble methods, correlation, quality prediction, wine grape quality, remote sensing, precision viticulture

INTRODUCTION

Precision viticulture (PV) is a strategy to manage vineyard variability by utilizing spatiotemporal data and observations, to enhance the oenological potential of a vineyard. In addition, new technologies introduced in support of vineyard management allow for the efficiency and quality of production to be improved, and in parallel, minimizing impacts on the environment (Balafoutis et al., 2017). This is particularly relevant in regions, where high wine production quality standards warrant adopting site-specific management practices to increase grape quality and yield.

Grape quality is a complex concept that refers to achieving optimal grape composition characteristics (Dai et al., 2011). Among these, sugar and titratable acidity are commonly used to describe the quality of wine grapes at harvest. The sugar content relates to the wine concentration of alcohol after fermentation, whereas the acid content determines the taste and stability of wine (Herrera et al., 2003). The most common method used in determining wine grape quality characteristics is to perform sample-based laboratory analysis by obtaining the chemical compounds of the grapes, which can be a time-consuming, complex, and expensive process (Cortez et al., 2009).

In recent years, remote sensing is widely applicable in agriculture, specifically crop growth monitoring and crop quality and yield estimation. For example, the normalized difference vegetation index (NDVI) is a vegetation index (VI) used for spatial decision-making in vineyards (Acevedo-Opazo et al., 2008). Canopy response and NDVI can be obtained in a direct, precise, and non-destructive way from various sensors and sensor configurations to acquire different bands, using proximal, aerial, and satellite platforms, based on the distance to the assessed crop (Hall et al., 2011; Baluja et al., 2012). Presently, advanced sensing techniques have had many applications beyond their original scope, especially as computing power has drastically increased in recent years, which have allowed for more complex machine learning techniques to be used to find patterns and correlations between NDVI and specific crop quality and yield characteristics (Pantazi et al., 2016; Kamilaris and Prenafeta-Boldú, 2018).

Previous research has been conducted to estimate crop quality and yield with the assessment of VIs derived from various sensors. A common approach is to perform statistical and regression analysis, including descriptive statistics, Pearson's correlation, and regression models. The Pearson correlation coefficient has been quantified in various studies to identify the spatial correlation between NDVI and crop quality and yield (Sun et al., 2017; He et al., 2018), research dedicated to selecting key variables to predict the product quality and yield with satisfactory performance directly. Linear and multivariate regression models have been constructed and fitted to various VIs to determine the field-wide production for multiple crops, such as wheat, corn, soybean, sorghum, rice, and grapes (Magney et al., 2016; Sun et al., 2017; Petersen, 2018; Prasetyo et al., 2018). An application on table grapes by Anastasiou et al. (2018) estimated yield and quality with the assessment of VIs derived from satellite and proximal sensing at different growth stages, from veraison to harvest. The VIs exhibited different degrees of correlations with different measurement dates and sensing methods. This study showed that both satellite-based and proximal-based NDVI at both stages (veraison and harvest) presented good correlations to crop quality characteristics, with proximal sensing proving to be the most accurate in estimating table grape yield and quality characteristics. In addition to linear regression models, more advanced approaches for yield estimation have been evaluated using ensemble methods. Boosted Regression Trees, Decision Trees, and Random Forests-based machine learning approaches were used to train the models to estimate crop yield from a short time series of remotely sensed NDVI (Heremans et al., 2015; Bhatnagar and Gohain, 2020).

While previous research has studied various correlation and regression models between VIs and crop production, machine learning techniques for estimating grape quality and yield have not been thoroughly investigated yet. In this article, we propose an alternative approach to predict wine grape quality characteristics by combining machine learning techniques and NDVI data collected at different growth stages with non-destructive methods, such as proximal and remote sensing, that are currently used in PV. For this reason, extensively used regression methods have been compared against more complex methods that deal better with outliers. Finally, to evaluate and ensure the robustness of the machine learning models used in this study, a 5-fold cross-validation procedure was followed across 20 experiments as a validation technique.

MATERIALS AND METHODS

Study Area

The field site where the study was conducted was a commercial wine grape vineyard block on the Palivos Estate located in Nemea, Greece (37.8032°, 22.69412°, WGS84). The vineyard, planted with *Vitis vinifera* L. cv. "Agiorgitiko" for winemaking is located on a steep slope, and the experimental area selected for data collection was approximately 2 ha. Wine grapes were trained to a vertical shoot positioned, cane pruned double Guyot training/trellis system, with northeast-southwest row orientation and row distance of 2.2 m.

Canopy Reflectance Data Collection

Canopy reflectance was measured four times per growing season, during 2019 and 2020, starting in late May until the harvest in early September, to record the NDVI at different phenological

growth stages of the grapevines. Crop vigor was assessed at these four berry growth stages, namely, (i) Flowering, (ii) Setting, (iii) Berries pea-sized, and (iv) Veraison, with two vehicle-mounted on-the-go proximal sensors, while an unmanned aerial vehicle (UAV) and Sentinel-2 (S2) satellite imagery were used to assess the crop vigor through remote sensing (**Figure 1**; **Table 1**). A CropCircle (CC), an active proximal canopy sensor (ACS-470, Holland Scientific Inc., Lincoln, NE, United States), and a Spectrosense+GPS (SS) passive sensor (Skye Instruments Ltd., Llandrindod Wells, United Kingdom) were mounted on a tractor, located at a height of approximately 1.5 m from the soil surface and according to each growth stage, and 0.5 m horizontally from the vines, to record proximal reflectance measurements from the side and the top of the canopy, respectively, at a rate of 1 reading per second and moving at a constant speed of 8–10 km/h. Tractor steering and the relative position of the sensors remained consistent throughout the data collection since the row distance is 2.2 m. All recorded data were georeferenced with a Garmin GPS16X HVS (Garmin, Olathe, Kansas United States) and SS's built-in GPS.

Aerial imagery data were acquired on the same dates as the proximal measurements, with a Phantom 4 Pro UAV (Dà-Jiāng Innovations, Shenzhen, Guangdong, China) equipped with a multispectral Parrot Sequoia+ camera (Parrot SA, Paris, France) and its GPS, enabling to geotag all obtained images. The UAV data acquisition was performed on the same days as the proximal measurements, close to solar noon, with nadir flights at 30 m above ground height. The flight duration was approximately 10 min, and the capture interval of the multispectral camera was set at 2 s. The UAV flight plan overlap and sidelap were 80 and 70%, respectively, with a ground sampling distance (GSD) of the imagery ortho-mosaics ~3 cm. Specifically, atmospherically corrected S2 satellite images, 2A products with a 10 m pixel spatial resolution, were downloaded from the official Copernicus Open Access Hub[1] for the closest dates available to the dates of the proximal and UAV surveys.

[1]https://scihub.copernicus.eu/

FIGURE 1 | Two vehicle-mounted crop reflectance sensors, UAV-acquired data, and Sentinel-2 archived imagery were used to estimate grapevine canopy properties. UAV, unmanned aerial vehicle.

TABLE 1 | Grapevine seasonal EL growth stages of proximal and remote sensing data acquisition.

Stage	EL No	Description	Date range	Data acquisition date
Flowering	23	16–20 leaves separated; 50% caps off	June 1–June 20	CC; SS; UAV: 040619, 050620 S2: 130619, 190620
Setting	27	Young berries enlarging, bunch at right angles to stem	June 20–July 20	CC; SS; UAV: 040719, 140720 S2: 050719, 120720
Berries pea-sized	31	About 7 mm in diameter	July 20–Aug 15	CC; SS; UAV: 010819, 110820 S2: 020819, 110820
Veraison	35	Berries begin to color and enlarge	Aug 15–Sept 10	CC; SS; UAV: 280819, 260820 S2: 280819, 280820

CC, CropCircle; SS, Spectrosense+GPS; UAV, unmanned aerial vehicle; and S2, Sentinel-2.

Data Preparation

All proximal canopy reflectance data was transformed to projected coordinates (UTM Zone 34N), cleaned by removing the data points located outside the field boundaries, and interpolated (Taylor et al., 2007). The interpolated data was upscaled to 10 × 20 m cells using ArcMap v10.3 (ESRI, Redlands, CA, United States). This resulted in 100 plots across the study area and generated NDVI maps' time-series with 10 × 20 m spatial resolution, oriented parallel to the trellis lines (**Figure 2**). Similarly, the UAV-acquired imagery was combined using Pix4D software (Pix4D S.A., Prilly, Switzerland), and the generated NDVI ortho-mosaic was fitted to the vineyard's boundaries. Radiometric calibration was applied to the generated ortho-mosaic using the reference images of a radiometric calibration target (Airinov Aircalib), captured after each flight. The software automatically recognized the albedo values for each band. The data was then upscaled using a mean aggregation approach to the same 100 plots. The Sentinel-2 NDVI was calculated using bands 4 and 8, red and near-infrared, respectively, obtaining imagery with 10 m pixel spatial resolution. To place the raster dataset to the spatially correct geographic location, a spatial correction "shift," based on ground control points from the UAV detailed map was carried out, following the boundaries of the experimental field before the satellite imagery was upscaled to the 10 × 20 m plots by averaging the NDVI of any pixel centroids within the management plots. The last step of satellite image processing was to clip the NDVI according to the border of the experimental field.

Qualitative Characteristics Analysis

Wine grapes were hand-harvested at the end of each growing season, in mid-September. A regular 100-cell grid (10 × 20 m), covering the total area, was configured to facilitate field sampling to assess crop yield and wine grape quality. The total yield was determined by counting the total number of crates filled with grapes per cell, multiplying it with the average crate weight of the harvested wine grapes. Wine grape quality characteristics were assessed by randomly picking 50 berries from each vineyard cell. The qualitative analysis of substances in berries, musts, and wines, namely, the total soluble solids (°Brix), the total titratable acidity, and the pH were determined.

Statistical Analysis and Regression Methods

An initial descriptive statistical analysis was executed to assess proximal and remote sensing performance on the prediction of wine grape quality. The exploratory correlation analysis included calculating Pearson's correlation matrix to investigate the relationships between NDVI data from all four proximal and remote sensors and the wine grape quality characteristics.

Regression model analysis was also performed. Due to various possible distributions found in the input data, several algorithms were evaluated only for those data that presented Pearson's correlation, with absolute values higher than 0.5 ($|r| > 0.50$). The regression algorithms used were both linear and nonlinear, depending on the output model generated. The linear models used included Ordinary Least Square (OLS), Theil–Sen, and Huber regression models.

- Ordinary least square: It is the most common estimation method for computing linear regression models, which can be found in related works (Prasetyo et al., 2018). The OLS regression is a powerful analysis that can analyze multiple variables simultaneously to answer complex research questions. However, like many statistical analyses, it has several underlying assumptions describing properties of the

FIGURE 2 | Satellite images of the wine grapes commercial vineyard and the 100-cell grid developed parallel to the trellis lines (Google Earth Pro, 2021).

error term. Moreover, this method can be used only if the data is normally distributed since outliers tend to pull the fitted model far from the accurate result.
- Theil–Sen estimator method: Linear models are sensitive to outliers, and few outliers can skew our predictions heavily (Sen, 1968). Compared with the OLS estimator, the Theil–Sen estimator is robust against outliers. Contrary to OLS, this algorithm uses a generalization of the median instead of the mean. Moreover, it is the most popular non-parametric technique for estimating a linear trend and does not assume the underlying distribution of the input data.
- Huber regression: It is also considered a robust linear regression modeling method, less sensitive to outliers in data (Huber, 1973). Huber regression is aware of the possibility of outliers in a dataset and assigns them less weight than other examples in the dataset, contrary to Theil–Sen that ignores their presence.

Moreover, to improve our model's predictive power, nonlinear methods, Decision Trees, and different Ensemble methods based on Decision Trees, including AdaBoosting, Random Forests, and Extra Trees were evaluated in the context of this research, combining the predictions from multiple machine learning algorithms together to make more accurate predictions than the individual models. These ensemble methods take one or more decision trees and then reduce their variance and bias by applying them to boost or bootstrap aggregation (bagging).

- Decision trees: Although it can also be used for classification, the algorithm is suitable for regression problems. Decision Tree models are the foundation of all tree-based models, visually representing the "decisions" used to generate predictions. This method uses a non-parametric learning approach. Its main advantage is its straightforward interpretation. If the model is not too complex, it can be visualized to understand better why the classifier made a specific decision. Its major disadvantage is that singular decision tree models are prone to overfitting, resulting in weak, unstable predictions that could have negative consequences if the input data contain noise.
- AdaBoost: The AdaBoost (adaptive boosting) algorithm uses an ensemble-learning approach known as boosting (Freund and Schapire, 1995). First, a decision tree is retrained several times, increasing the emphasis on those data samples where the regression was inaccurate. Then, it combines the predictions from multiple "weak learners," simple decision tree models, which are added sequentially to the ensemble, correcting the predictions made by the model before it in the sequence.
- Random forest: A supervised learning algorithm that uses ensemble learning method for regression, aggregating many decision tree regressors into one model, which have been trained on different data samples drawn from the input feature (the NDVI in this study), with the bootstrap sampling technique (Breiman, 2004). As a result, the trees in random forests run in parallel, and each tree draws a random sample from the original dataset, adding some randomness that prevents overfitting.
- Extremely randomized trees: Extra Trees are similar to the Random Forest, combining the predictions from many decision trees (Geurts et al., 2006). However, it does not use bootstrap sampling but the entire original input sample. It creates many unpruned decision or regression trees from the training dataset, and predictions are made by averaging the prediction of the decision trees. It uses a random split for node creation to grow the trees, leading to a reduction in overfitting.

Although tree-based methods provide an approach for overcoming the constraints of parametric models, their limitation is that they are computationally more expensive than the traditional OLS. However, if the differences in the performances are high enough, they should be a good approach for addressing the regression modeling problem.

Fine-Tuning

Optimizing machine learning models relies on an empirical approach, and specifically, in ensemble methods, the number of estimators is not predefined. Moreover, some tree-based hyperparameters, such as the ideal maximum depth of the trees, are unknown until several values are evaluated. Thus, being able to test several model hyperparameters quickly is imperative in maximizing performance. For that reason, to start the training process, several experiments were performed by grid-search to find the best hyperparameters for this study. **Table 2** summarizes the primary hyperparameters that governed the ensemble methods. In the results section, the ones that obtained the best performances will be stated.

Evaluation Methodology

The prediction accuracy was assessed using the coefficient of determination (R^2), and root mean square error (RMSE) metrics. Additionally, to test the generalization ability of every regression model and ensure their robustness, a 5-fold cross-validation procedure was followed for each of them. Additionally, to compute the final performances more accurately, the experiments were repeated 20 times with different data splits.

TABLE 2 | Hyperparameters evaluated for optimizing the ensemble learning models.

Hyperparameter	Description	Values
Number of trees	The number of trees used in the ensemble. High values could reduce overfitting.	10, 50, 100, 200
Max. depth	The maximum depth of the trees. Too high could lead to overfitting.	No maximum, 1, 5, 7
Split criteria	The function to measure the quality of a node split. The calculation of the Gini Index is computationally faster, but it could lead to more poor results.	Information gain, Gini impurity

Software and Hardware

One main software package was used in this work: Scikit-Learn machine learning library (version 0.23.2). In addition, all the experiments were run on Ubuntu 18.04 as the OS (**Figure 3**).

RESULTS

Exploratory Correlation Analysis

The exploratory correlation analysis included calculating Pearson's correlation matrix to investigate the relationships between NDVI data from all four proximal and remote sensors and the wine grape quality characteristics. The 2019 and 2020 correlation matrices generally indicated good absolute correlations between NDVI data from all four proximal and remote sensors and total soluble solids, the sugar content measured in °Brix, ($|r| > 0.50$). However, this was not the case for the other two main wine grape quality characteristics, the total titratable acidity, and the pH, that presented no correlation with the NDVI data at any crop stage.

In the top two rows of **Table 3**, the best intra-sensor correlations between NDVI data and total soluble solids are presented. The signal stabilizes mid-late in the growing season. The maximum correlation ($|r| = 0.74$) for 2019 was observed for Spectrosense+GPS data during Berries pea-sized and the Veraison stage (i.e., mid-late season with full canopy growth). For 2020, correlations were strongest for UAV data ($|r| = 0.79$), at the same growth stages. All Sentinel-2 NDVI variables demonstrated relatively weak correlations ($0.29 < |r| < 0.57$) when correlated with the total soluble solids. For the given stages of the growing season, it was noticed that during Veraison, the NDVI data from UAV, Spectrosense+GPS, and the CC sensors correlated the best with the total soluble solids for both years.

The Pearson's correlation coefficient evolution for both 2019 and 2020 is shown in **Figure 4**. The highest correlations between the proximal- and remote-based spectral vegetation indices and the wine grape total soluble solids at different crop stages are recorded for the UAV, with the Spectrosense+GPS, the CropCircle, and the Sentinel-2 imagery following. Even though no pattern was noted in the correlation coefficient evolution, it is clear that mid-late season the NDVI correlates the best with wine grape quality characteristics.

Regression Analysis

Regression model analysis was performed only for those data that presented Pearson's correlation for NDVI data from all four proximal and remote sensors and total soluble solids, with absolute values higher than 0.5 ($|r| > 0.50$) for the different crop stages. The regression algorithms used, both linear and nonlinear regression analysis, were performed using those highly correlated NDVI data to evaluate their performance in assessing the wine grapes' quality characteristics. The regression models between NDVI data from all four proximal and remote sensors and total soluble solids presented different degrees of accuracy, depending on the model fitted, the sensor used, and the growth stage assessed. The best fitted, linear, and nonlinear regressions were observed for UAV and Spectrosense+GPS data during the mid-late season with full canopy growth in Berries pea-sized and the Veraison growth stages.

When using OLS, Theil-Sen and Huber linear regression models the best fit for the models was for the estimation of total soluble solids, during Veraison, with a coefficient of determination R^2 ranging from ($0.38 < R^2 < 0.61$) for both 2019 and 2020. The maximum coefficient of determination for the linear regression models ($R^2 = 0.61$)

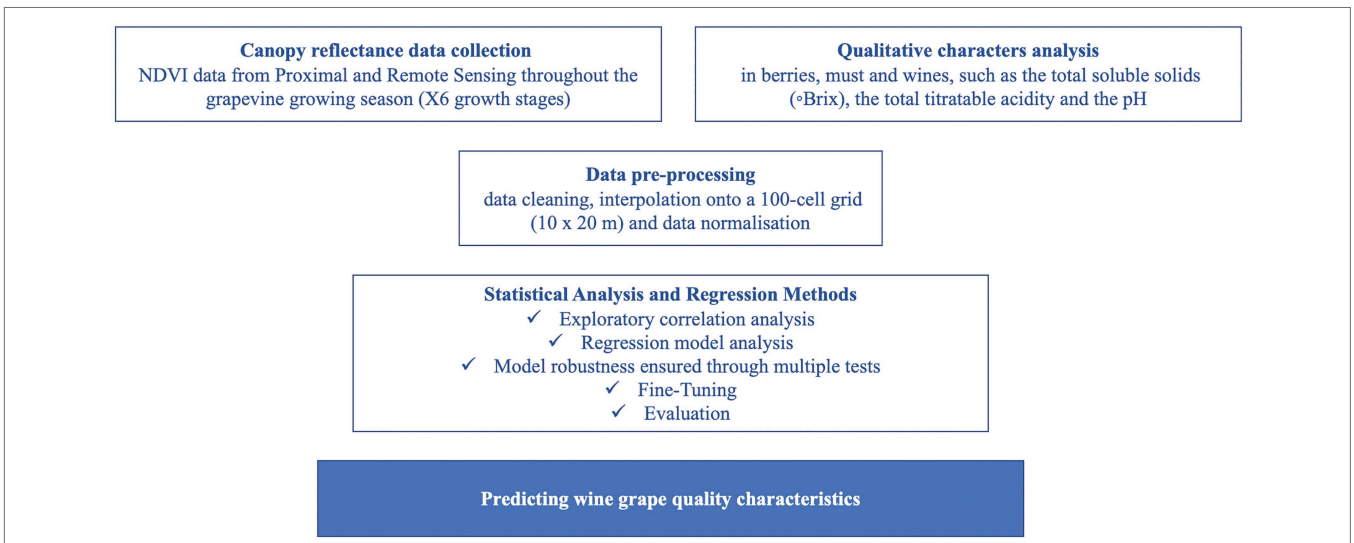

FIGURE 3 | Workflow for investigating a selection of methods for predicting wine grape quality characteristics using normalized difference vegetation index (NDVI) data from proximal and remote sensing.

TABLE 3 | Selected best performed Pearson's correlation coefficients comparisons between NDVI data from all four proximal and remote sensors and total soluble solids (i) for a given sensor (rows 1 and 2) and (ii) for given growth stages of the season (rows 3 and 4; CC, CropCircle; SS, Spectrosense+GPS; UAV; and S2, Sentinel-2).

Per sensor	**CropCircle**	**Spectrosense+GPS**	**UAV**	**Sentinel-2**
2019	0.69 (Veraison)	0.74 (Veraison)	0.63 (Veraison)	0.57 (Berries pea-sized)
	0.52 (Berries pea-sized)	0.69 (Berries pea-sized)	0.62 (Flowering)	0.54 (Veraison)
2020	0.54 (Setting)	0.70 (Setting)	0.79 (Veraison)	0.33 (Veraison)
	0.50 (Berries pea-sized)	0.67 (Veraison)	0.77 (Berries pea-sized)	0.29 (Setting)
Per growth stage	**Flowering**	**Setting**	**Berries pea-sized**	**Veraison**
2019	0.62 (UAV)	0.58 (UAV)	0.69 (SS)	0.74 (SS)
	0.36 (S2)	0.48 (S2)	0.61 (UAV)	0.69 (CC)
2020	0.76 (UAV)	0.70 (SS)	0.77 (UAV)	0.79 (UAV)
	0.63 (SS)	0.54 (CC)	0.50 (CC, SS)	0.67 (SS)

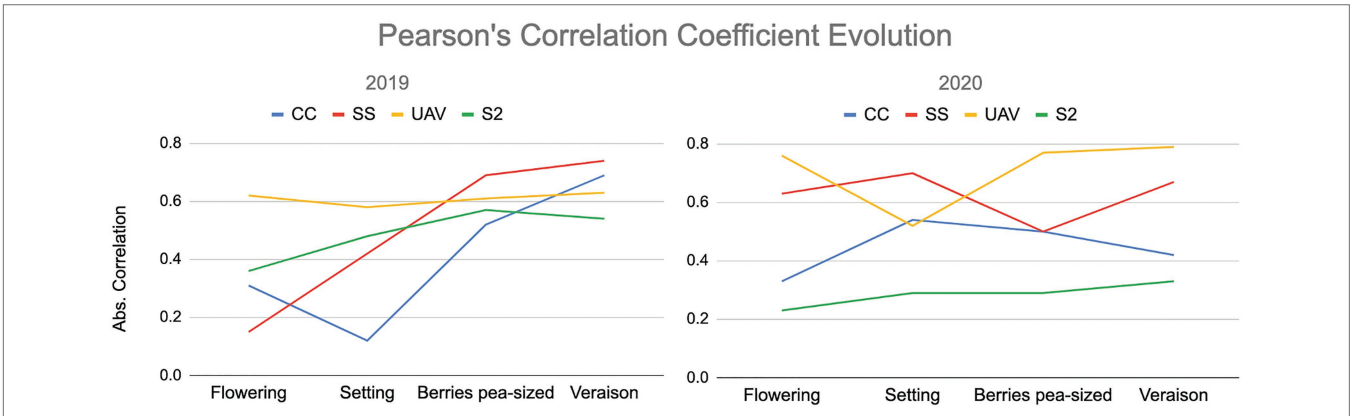

FIGURE 4 | Pearson's correlation coefficients evolution throughout the growing seasons 2019 and 2020 (legend as for **Table 3**).

TABLE 4 | Selected best performed linear regression models performed using the highly correlated NDVI data from all four proximal and remote sensors to evaluate their performance in assessing the wine grapes quality characteristics (legend as for **Table 3**).

	Sensor_growth stage	**Method**	**R^2 (avg)**	**RMSE**
2019	SS_Veraison	OLS	0.51 ± 0.09	1.45 ± 0.19
	CC_Veraison	OLS	0.42 ± 0.10	1.67 ± 0.35
	SS_Berries pea-sized	Huber	0.41 ± 0.11	1.71 ± 0.24
	UAV_Veraison	Theil-Sen	0.38 ± 0.10	1.95 ± 0.55
2020	UAV_Veraison	Theil-Sen	0.61 ± 0.03	1.37 ± 0.19
	UAV_Berries pea-sized	Huber	0.57 ± 0.04	1.55 ± 0.32
	UAV_Flowering	Theil-Sen	0.56 ± 0.06	1.75 ± 0.19
	SS_Setting	OLS	0.44 ± 0.07	1.73 ± 0.23
	UAV_Setting	OLS	0.44 ± 0.04	2.09 ± 0.45

was observed for 2020 retrieved with UAV data and fitting the Theil–Sen regression model. For 2019, the OLS seems to perform better using canopy reflectance data coming from the CropCircle and the Spectrosense+GPS proximal sensors. The selected best-performed results of the linear regression analysis are presented in **Table 4**.

Among the nonlinear methods, and to improve our models' predictive power, different Ensemble methods based on Decision Trees, including AdaBoosting, Random Forests, and Extra Trees, were evaluated, aggregating the predictions from multiple machine learning algorithms together to make more accurate predictions than the individual models. The best fit for the nonlinear model was for estimating total soluble solids, during Veraison, with the coefficient of determination R^2 ranging from ($0.42 < R^2 < 0.59$) for both 2019 and 2020. The maximum coefficient of determination for the nonlinear regression models ($R^2 = 0.59$) was observed for 2020 retrieved with UAV data and using the AdaBoost algorithm. For 2019, the Extra Trees performs better using canopy reflectance data from the CropCircle and the Spectrosense+GPS proximal sensors. In the case of the Adaboost, the best hyperparameters were 50 decision trees with a maximum depth of 1. In the case of the Extra Trees, the best hyperparameters were also 50 decision trees, but with a maximum depth of 7 and the split criteria based on Information Gain. This same configuration was the one that led to the highest performance in Random Forest (SS_Berries pea_sized and UAV_Flowering). It is important to note that although the Decision Tree classifier, as a standalone classifier, was also evaluated, its performance was always lower than the ensemble methods. The selected best-performed results of the nonlinear regression analysis are presented in **Table 5**.

TABLE 5 | Selected best performed nonlinear regression models performed using the highly correlated NDVI data from all four proximal and remote sensors to evaluate their performance in assessing the wine grapes quality characteristics (legend as for **Table 3**).

	Sensor_growth stage	Method	R^2 (avg)	RMSE
2019	SS_Veraison	Extra Trees	0.46 ± 0.03	1.64 ± 0.14
	CC_Veraison	Random Forest	0.43 ± 0.5	1.68 ± 0.32
	SS_Berries pea-sized	Random Forest	0.43 ± 0.06	1.82 ± 0.29
	UAV_Veraison	Extra Trees	0.42 ± 0.05	2.05 ± 0.51
	UAV_Flowering	Random Forest	0.39 ± 0.07	2.11 ± 0.33
2020	UAV_Veraison	AdaBoost	0.59 ± 0.05	1.41 ± 0.25
	UAV_Berries pea-sized	Extra Trees	0.53 ± 0.03	1.65 ± 0.22
	UAV_Flowering	Extra Trees	0.56 ± 0.05	1.72 ± 0.29
	SS_Setting	Extra Trees	0.43 ± 0.02	1.83 ± 0.33
	UAV_Setting	Extra Trees	0.41 ± 0.06	1.95 ± 0.35

DISCUSSION

In this article, an alternative approach to predict wine grape quality characteristics by combining machine learning techniques and NDVI data collected at different growth stages with non-destructive methods, such as proximal and remote sensing, currently used in precision viticulture, is used in precision viticulture proposed. While previous research has studied various correlation and regression models between VIs and crop production, the use of machine learning techniques for the estimation of grape quality and yield has not been thoroughly investigated yet (Sun et al., 2017; He et al., 2018; Liakos et al., 2018). This study proved that advanced sensing techniques may have many applications, especially with the help of the increasing computing power, allowing for more complex machine learning techniques to find patterns and correlations between canopy reflectance data and specific crop quality characteristics. Furthermore, extensively used regression methods have been compared against more complex methods that deal better with outliers. In addition, to evaluate and ensure the robustness of the machine learning models used in this study, a 5-fold cross-validation procedure was followed across 20 experiments. The validation technique avoids overoptimistic (or random) results based on hold-out evaluations.

Wine grapes quality refers to the achievement of optimal levels of all grape composition characteristics, with sugar content being a basic one, related to the wine concentration of alcohol after fermentation. The exploratory correlation analysis presented that the recorded canopy reflectance data from all four sensors, i.e., the pure vine NDVI extracted from two proximal sensors, a CropCirle and a Spectrosense+GPS and the "mixed pixel" UAV and Sentinel-2 imagery, showed an increasing correlation to the total soluble solids as the season progressed. Similar results have also been found by other researchers (Tagarakis et al., 2013; Fountas et al., 2014), with this study being aligned with Anastasiou et al. (2018), who estimated yield and quality with the assessment of vegetation indices derived from satellite and proximal sensing at different growth stages and their study showed that NDVI at late developmental stages of the vine growing season presented good correlations to crop quality characteristics. Also, García-Estévez et al. (2017) found that the highest correlation of NDVI derived from proximal sensing with yield parameters of wine grapes was at veraison. Sun et al. (2017), the best crop stage for estimating wine grape yield characteristics from satellite-derived data is before harvest. The strongest correlations with the sugar content were observed for NDVI data collected with the UAV, Spectrosense+GPS, and the CropCircle, during Berries pea-sized and the Veraison stage, mid-late season with full canopy growth, for both years. The weaker correlation coefficients recorded with Sentinel-2 and assessed with an overhead "mixed pixel" approach indicated less reliability for wine grapes quality characteristics predictions, which is a sensible result, as Khaliq et al. (2019) also discussed that satellite imagery resolution could not be directly used to describe vineyard variability reliably. However, this was not the case for the other two main wine grape quality characteristics, the total titratable acidity, and the pH, that presented no correlation with the NDVI data at any crop stage.

The performance of each sensor was different and affected by data acquisition parameters, such as proximity to the vines and the specific technical characteristics of the equipment used. The CropCircle and the Spectrosense+GPS proximal sensors were mounted on a tractor, recording reflectance measurements from the side and the top of the canopy, respectively, while the UAV and Sentinel-2 satellite imagery assessed the crop vigor through remote sensing from the top. The highest correlations between the proximal- and remote-based spectral vegetation indices and the wine grape total soluble solids at different crop stages are recorded for the UAV, with the Spectrosense+GPS, the CropCircle, and the Sentinel-2 imagery following. The Spectrosense+GPS and UAV seemed to perform better and in a similar way, most probably due to the scanning orientation, which was the top side of the canopy at close proximity. Even though the UAV is classed as a remote sensor, it provides high spatial resolution. Although no pattern was noted in the correlation coefficient evolution, it is clear that the NDVI correlates the best with wine grape quality characteristics in the mid-late season. The exploratory analysis acted as an evaluation for performing predictive analytics on the dataset.

The dataset was then used for training machine learning algorithms, evaluating linear and nonlinear regression models, including OLS, Theil–Sen, and the Huber regression models and Ensemble Methods based on Decision Trees. The regression algorithms used, both linear and nonlinear regression analysis,

were performed using those highly correlated NDVI data to evaluate their performance in assessing the wine grapes' quality characteristics. The regression models between NDVI data from all four proximal and remote sensors and total soluble solids presented different degrees of accuracy, depending on the model fitted, the sensor used, and the growth stage assessed. The UAV and the Spectrosense+GPS data proved to be more accurate in predicting the sugars out of all wine grape quality characteristics, especially during the mid-late season with full canopy growth, in Berries pea-sized and the Veraison growth stages.

All regression methods that were applied, both linear and nonlinear, the OLS, Theil–Sen, and Huber regression models, Decision Trees and Ensemble methods based on Decision Trees, including AdaBoosting, Random Forests, and Extra Trees performed similarly in wine grapes quality parameters prediction, with the best-fitted models achieving a coefficient of determination of R^2 = 0.61. These results confirm the findings of Bhatnagar and Gohain (2020), who used decision tree and random forest-based machine learning approaches to estimate crop yield by comparing their values with NDVI values, and they concluded with the result of R = 0.67. Therefore, the implementation of machine learning techniques resulted in similar results as linear models. However, more precise wine grape quality predictions were obtained when NDVI data were collected close to the harvest date, although promising results were obtained for the early season, as noted by Ballesteros et al. (2020). The fact that the ensemble methods performed in some cases slightly worse than the linear methods could be due to the limited dataset size in combination with the use of 5-fold cross-validation, which reduces the training set to 80% of the total dataset size. However, this is considered necessary to provide reliable results. On the other hand, it could be discussed whether stacking learning techniques instead of boosting or bagging could lead to better performances (Wolpert, 1992; van der Laan et al., 2007).

CONCLUSION

In this paper, the use of machine learning techniques to estimate wine grape quality characteristics is investigated. An alternative approach investigating the combination of a selection of methods extensively used regression methods to more complex methods that deal better with outliers, predicted wine grape quality characteristics using NDVI data, collected at different growth stages from proximal and remote sensing, is proposed. This study proved that advanced sensing techniques may have many applications, especially with the help of the increasing computing power, allowing for more complex machine learning techniques to find patterns and correlations between canopy reflectance data and specific crop quality characteristics.

The descriptive statistical analysis showed that the NDVI data from the UAV, Spectrosense+GPS, and the CropCircle, during Berries pea-sized and the Veraison stage, mid-late season with full canopy growth, have the strongest correlations with the sugar content for both years. At the same time, Sentinel-2 imagery indicated less reliability for wine grapes' quality characteristics predictions. The predictive analysis indicated that regression models between NDVI data from all four proximal and remote sensors and total soluble solids presented different degrees of accuracy, depending on the model fitted, the sensor used, and the growth stage assessed. All regression methods that were applied, both linear and nonlinear, performed similarly in wine grapes quality parameters prediction. The UAV and the Spectrosense+GPS data proved to be more accurate in predicting the sugars out of all wine grape quality characteristics, especially closer to the harvesting period. Although correlation is not significant, it seems enough to predict wine grape quality with satisfied approximation.

The investigation of a selection of methods, including OLS, Theil–Sen, and Huber regression models, Decision Trees, AdaBoost, Random Forests, and Extra Trees, for the assessment of wine grape quality characteristics using spectral vegetation indices, presents a great potential for machine learning techniques to be used as an alternative method, to the currently widely used linear regression processes. Ensemble methods presented similar results to regression analysis, while dealing better with the outliers and ensuring robustness through cross-validation techniques. This research will be extended by assessing stacking learning techniques instead of boosting and bagging as a new ensemble method and exploring if they could lead to better performances. Finally, given the perennial nature of grapevines and the various environmental and endogenous factors determining quality, seasonal calibration for quality prediction should be considered in future research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors contributed to conception and design of the study. AK organized and realized the data collection and the data pre-processing and preparation, and drafted the manuscript outline. BE-G, EV, and IM contributed to the statistical analysis and machine learning implementation. EV and IM wrote the first draft of the introductory sections of the manuscript. AK, BE-G, and SF completed the results and discussion sections. All authors contributed to the article and approved the submitted version.

REFERENCES

Acevedo-Opazo, C., Tisseyre, B., Guillaume, S., and Ojeda, H. (2008). The potential of high spatial resolution information to define within-vineyard zones related to vine water status. *Precis. Agric.* 9, 285–302. doi: 10.1007/s11119-008-9073-1

Anastasiou, E., Balafoutis, A., Darra, N., Psiroukis, V., Biniari, A., Xanthopoulos, G., et al. (2018). Satellite and proximal sensing to estimate the yield and quality of table grapes. *Agriculture* 8:94. doi: 10.3390/agriculture8070094

Balafoutis, A., Beck, B., Fountas, S., Vangeyte, J., Wal, T. V. D., Soto, I., et al. (2017). Precision agriculture technologies positively contributing to GHG

emissions mitigation, farm productivity and economics. *Sustainability* 9:1339. doi: 10.3390/su9081339

Ballesteros, R., Intrigliolo, D. S., Ortega, J. F., Ramírez-Cuesta, J. M., Buesa, I., and Moreno, M. A. (2020). Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. *Precis. Agric.* 21, 1242–1262. doi: 10.1007/s11119-020-09717-3

Baluja, J., Diago, M. P., Goovaerts, P., and Tardaguila, J. (2012). Assessment of the spatial variability of anthocyanins in grapes using a fluorescence sensor: relationships with vine vigour and yield. *Precis. Agric.* 13, 457–472. doi: 10.1007/s11119-012-9261-x

Bhatnagar, R., and Gohain, G. B. (2020). "Crop yield estimation using decision trees and random forest machine learning algorithms on data from terra (EOS AM-1) & Aqua (EOS PM-1) satellite data," in *Machine Learning and Data Mining in Aerospace Technology: Studies in Computational Intelligence. Vol. 836*. eds. A. Hassanien, A. Darwish and H. El-Askary (Cham: Springer), 107–124.

Breiman, L. (2004). "Random forests," in *Machine Learning. Vol.* 45. 5–32.

Cortez, P., Teixeira, J., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). *Discovery Science*, Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 66–79.

Dai, Z. W., Ollat, N., Gomès, E., Decroocq, S., Tandonnet, J. P., Bordenave, L., et al. (2011). Ecophysiological, genetic, and molecular causes of variation in grape berry weight and composition: a review. *Am. J. Enol. Vitic.* 62, 413–425. doi: 10.5344/ajev.2011.10116

Fountas, S., Anastasiou, E., Balafoutis, A., Koundouras, S., Theoharis, S., and Theodorou, N. (2014). The influence of vine variety and vineyard management on the effectiveness of canopy sensors to predict winegrape yield and quality. In Proceedings of the International Conference of Agricultural Engineering, Zurich, Switzerland. 6–10.

Freund, Y., and Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. 550 EuroCOLT. 551.

García-Estévez, I., Quijada-Morín, N., Rivas-Gonzalo, J. C., Martínez-Fernández, J., Sánchez, N., Herrero-Jiménez, C. M., et al. (2017). Relationship between hyperspectral indices, agronomic parameters and phenolic composition of *Vitis vinifera* cv. Tempranillo grapes: hyperspectral indices, agronomic parameters and phenolic composition of *V. vinifera*. *J. Sci. Food Agric.* 97, 4066–4074. doi: 10.1002/jsfa.8366

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1

Google Earth Pro (2021). Map showing location of the wine grapes commercial vineyard. Available at: https://earth.google.com/web/@37.80431086,22.69389566,1360.70594953a,0d,35y,359.9995h,0t,0r?utm_source=earth7&utm_campaign=vine&hl=en

Hall, A., Lamb, D. W., Holzapfel, B. P., and Louis, J. P. (2011). Within-season temporal variation in correlations between vineyard canopy and winegrape composition and yield. *Precis. Agric.* 12, 103–117. doi: 10.1007/s11119-010-9159-4

He, M., Kimball, J. S., Maneta, M. P., Maxwell, B. D., Moreno, A., Beguería, S., et al. (2018). Regional crop gross primary productivity and yield estimation using fused landsat-MODIS data. *Remote Sens.* 10:372. doi: 10.3390/rs10030372

Heremans, S., Dong, Q., Zhang, B., Bydekerke, L., and Orshoven, J. (2015). Potential of ensemble tree methods for early-season prediction of winter wheat yield from short time series of remotely sensed normalized difference vegetation index and in situ meteorological data. *J. Appl. Remote Sens.* 9:97095. doi: 10.1117/1.JRS.9.097095

Herrera, J., Guesalaga, A., and Agosin, E. (2003). Shortwave–near infrared spectroscopy for non-destructive determination of maturity of wine grapes. *Meas. Sci. Technol.* 14, 689–697. doi: 10.1088/0957-0233/14/5/320

Huber, P. (1973). Robust regression: asymptotics, conjectures and monte carlo. *Ann. Stat.* 1, 799–821. doi: 10.1214/aos/1176342503

Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: a survey. *Comput. Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016

Khaliq, A., Comba, L., Biglia, A., Aimonino, D. R., Chiaberge, M., and Gay, P. (2019). Comparison of satellite and UAV-based multispectral imagery for vineyard variability assessment. *Remote Sens.* 11:436. doi: 10.3390/rs11040436

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: a review. *Sensors* 18:2674. doi: 10.3390/s18082674

Magney, T. S., Eitel, J. U., Huggins, D. R., and Vierling, L. A. (2016). Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. *Agric. For. Meteorol.* 217, 46–60. doi: 10.1016/j.agrformet.2015.11.009

Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., and Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* 121, 57–65. doi: 10.1016/j.compag.2015.11.018

Petersen, L. K. (2018). Real-time prediction of crop yields from MODIS relative vegetation health: a continent-wide analysis of Africa. *Remote Sens.* 10:1726. doi: 10.3390/rs10111726

Prasetyo, Y., Sukmono, A., Aziz, K. W., and Aji, B. J. P. S. (2018). Rice productivity prediction model design based on linear regression of spectral value using NDVI and LSWI combination on landsat-8 imagery. *IOP Conf. Ser. Earth Environ. Sci.* 165:12002. doi: 10.1088/1755-1315/165/1/012002

Sen, P. (1968). Estimates of the regression coefficient based on kendall's tau. *J. Am. Stat. Assoc.* 63, 1379–1389. doi: 10.1080/01621459.1968.10480934

Sun, L., Gao, F., Anderson, M. C., Kustas, W. P., Alsina, M. M., Sanchez, L., et al. (2017). Daily mapping of 30 m LAI and NDVI for grape yield prediction in California vineyards. *Remote Sens.* 9:317. doi: 10.3390/rs9040317

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., and Gemtos, T. A. (2013). Management zones delineation using fuzzy clustering techniques in grapevines. *Precis. Agric.* 14, 18–39. doi: 10.1007/s11119-012-9275-4

Taylor, J. A., McBratney, A. B., and Whelan, B. M. (2007). Establishing Management Classes for Broadacre Agricultural Production. *Agron. J.* 99, 1366–1376. doi: 10.2134/agronj2007.0070

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* 6:25. doi: 10.2202/1544-6115.1309

Wolpert, D. (1992). Stacked generalization. *Neural Netw.* 5, 241–259. doi: 10.1016/S0893-6080(05)80023-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kasimati, Espejo-Garcia, Vali, Malounas and Fountas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 29 June 2021
doi: 10.3389/fpls.2021.695749

Grape Leaf Black Rot Detection Based on Super-Resolution Image Enhancement and Deep Learning

Jiajun Zhu†, Man Cheng†, Qifan Wang, Hongbo Yuan and Zhenjiang Cai*

College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, China

The disease spots on the grape leaves can be detected by using the image processing and deep learning methods. However, the accuracy and efficiency of the detection are still the challenges. The convolutional substrate information is fuzzy, and the detection results are not satisfactory if the disease spot is relatively small. In particular, the detection will be difficult if the number of pixels of the spot is $<32 \times 32$ in the image. In order to effectively address this problem, we present a super-resolution image enhancement and convolutional neural network-based algorithm for the detection of black rot on grape leaves. First, the original image is up-sampled and enhanced with local details using the bilinear interpolation. As a result, the number of pixels in the image increase. Then, the enhanced images are fed into the proposed YOLOv3-SPP network for detection. In the proposed network, the IOU (Intersection Over Union, IOU) in the original YOLOv3 network is replaced with GIOU (Generalized Intersection Over Union, GIOU). In addition, we also add the SPP (Spatial Pyramid Pooling, SPP) module to improve the detection performance of the network. Finally, the official pre-trained weights of YOLOv3 are used for fast convergence. The test set test_pv from the Plant Village and the test set test_orchard from the orchard field were used to evaluate the network performance. The results of test_pv show that the grape leaf black rot is detected by the YOLOv3-SPP with 95.79% detection accuracy and 94.52% detector recall, which is a 5.94% greater in terms of accuracy and 10.67% greater in terms of recall as compared to the original YOLOv3. The results of test_orchard show that the method proposed in this paper can be applied in field environment with 86.69% detection precision and 82.27% detector recall, and the accuracy and recall were improved to 94.05 and 93.26% if the images with the simple background. Therefore, the detection method proposed in this work effectively solves the detection task of small targets and improves the detection effectiveness of the grape leaf black rot.

Keywords: small targets, grape black rot, super-resolution, convolutional neural network, deep-learning

OPEN ACCESS

Edited by:
Spyros Fountas, Agricultural University of Athens, Greece

Reviewed by:
Jun Liu, Weifang University of Science and Technology, China
Sanjaya Tripathy, Birla Institute of Technology, Mesra, India

****Correspondence:***
Hongbo Yuan
yuanhongbo222@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *15 April 2021*
Accepted: *07 June 2021*
Published: *29 June 2021*

Citation:
Zhu J, Cheng M, Wang Q, Yuan H and Cai Z (2021) Grape Leaf Black Rot Detection Based on Super-Resolution Image Enhancement and Deep Learning. Front. Plant Sci. 12:695749. doi: 10.3389/fpls.2021.695749

INTRODUCTION

Grapes are one of the most commonly grown economic fruits in the world, which are often used in the production of wine, fermented beverages, and raisins (Kole et al., 2014). The larger the area used for the cultivation of grapes, the larger is the scale of the disease affecting the grapes and consequently, the probability of economic loss is higher as well. Generally, the early stages of

grape diseases are evident on the leaves. Therefore, the leaves can be used for the identification and diagnosis of diseases during the early stage. Black rot is one of the most common grape diseases in the world (Molitor and Berkelmann-Loehnertz, 2011). Black rot is a fungal disease that exhibits a black spot on the grape leaves. This spot is relatively smaller as compared to the size of the leaves. This disease usually appears during the moist spring season and early summer. The black spot affects a wide area of the leaves (Pearson and Goheen, 1989). Currently, a manual method is mainly used for the identification of this disease. In this method, the farmers use their extensive experience to make a rough identification of the disease. However, it is notable that this approach not only requires a lot of manual labor but is also susceptible to the subjective factors (Chen et al., 2020). In order to ensure the grape production and economic well-being of the farmers, rapid and effective detection of black rot on grape leaves is important for the farming industry.

Currently, the machine vision technologies are widely used in various fields for detection and classification tasks. In the early stages of research on grape leaf diseases using machine learning, Agrawal et al. (2017) used SVM for the classification of grape diseases using leaves. The proposed method included image resizing, image enhancement, and image smoothing to save the memory and reduce the processing time. Waghmare et al. (2016) proposed the local binary patterns and machine learning for the detection and classification of grape diseases. The authors uniformly resize the images to 226 226 before further processing. In addition, the images are transformed from the RGB to HSV color space and the background subtraction is used to remove the unwanted background in the images. Es-Saady et al. (2016) proposed the automatic identification of plant diseases using leaves based on a serial combination of two SVM classifiers. The authors used different colors as the classification criterion in the first classifier. Then, in the second classifier, the shape and texture features of colored leaves were used for classification. Although the traditional machine learning algorithms have made some achievements in grape leaf disease detection, these methods require manual feature extraction. In addition, the accuracy of disease detection needs to be improved as well.

Recently, deep learning has been extensively used for the purpose of detection and classification in various applications. Felzenszwalb et al. (2008) proposed a region-based convolutional neural network (CNN) for target detection. There are few region-based detection algorithms presented in literature, such as RCNN (Girshick et al., 2014), Fast-RCNN (Girshick, 2015), and Faster-RCNN (Ren et al., 2017). Similarly, the well-known end-to-end detection algorithms include SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017, 2018). The CNNs have been used for the detection of grape diseases. Wagh et al. (2019) proposed an automatic grape disease identification system based on the AlexNet. This method has the ability to detect the bacterial spots and powdery mildew with an accuracy of 98.23%. Ji et al. (2020) proposed a unified model based on multiple CNNs for automatic identification of grape leaf diseases. This method was used for the classification of black rot, esca, isariopsis leaf spot, and healthy images. The average validation accuracy of this method is 99.17% and the test accuracy is 98.57%. Liu et al. (2020) proposed an improved CNN for the identification of grape leaf diseases. The authors use this technique to identify anthracnose, brown spot, mites, black rot, downy mildew, and leaf blight. This method utilizes the depth-separable convolution instead of standard convolutional layers. As a result, this method achieves higher convergence speed and accuracy. Xie et al. (2020) proposed a rapid detector for grape leaf diseases based on deep learning. This technique automatically extracts the disease spot features and has the ability to detect four common grape leaf diseases with high accuracy and fast detection speed. Alessandrini et al. (2021) proposed a grapevine leaves dataset for early detection and classification of Esca disease in vineyards through machine learning.

However, it is noteworthy that the existing networks, such as AlexNet, RCNN, and Fast-RCNN, suffer from non-negligible miss-detections and low recall for the small spots. For instance, if the spot pixels are <32 32 (Bosquet et al., 2018), or when the image resolution is not high. According to the definition of the international organization SPIE, a small target is a target area <80 pixels in a 256×256 image, that is, the target whose pixel proportion is $<0.12\%$ of the total image pixels. The small target recognition is still a challenge.

The super-resolution is an image processing method which is commonly used in the field of remote sensing (Xie et al., 2017; Arun et al., 2019), feature extraction, non-linear mapping, and image reconstruction. This technique has the ability to make small targets on the original image clear, even after the application of convolution. Bai et al. (2018a,b) applied the super-resolution for small target detection. The results show that this method effectively enhances the information of small targets, while improving the detection accuracy. Noh et al. (2019) proposed an accurate monitoring of feature super-resolution for small target detection. This method demonstrates the importance of using the appropriate high-resolution target features that share the same relative field of perception as the low-resolution input features to provide direct monitoring. Rabbi (2020) proposed remote sensing image small target detection based on an end-to-end edge-enhanced neural network and target detector network. The authors present the stochastic resonant network which has a faster response and yields the best results for small targets on satellite images.

In order to improve the detection accuracy of low-resolution small targets in the grape black rot spot detection, in this work, we propose a super-resolution image enhancement and deep learning-based detection of black rot in grape leaves. The proposed method used an improved loss function for performing detections. In addition, we also propose the spatial pyramid pooling (SPP) module in the detection network, which effectively increases the reception range of the backbone features and significantly separates the most important contextual features. Moreover, the proposed method improves the target recall and accuracy as compared with the max pooling technique.

The major contributions of this work are as follows.

(1) We perform enhancement of grape leaves by using the bilinear interpolation.

(2) We improve the YOLOv3 network. The IOU in the original yolo YOLOv3 network is replaced with GIOU. In addition, we also add the SPP module to improve the detection performance of the network.
(3) We perform experiments and analysis to evaluate the effectiveness of the super-resolution image enhancement and improved YOLOv3 network for grape black rot detection.

MATERIALS AND METHODS

Data Set and Test Environment Setup

The data used to perform experiments in this work is the open dataset Plant Village. We select 1,180 images of grapevine leaf black rot for disease detection. We use LabelImg for annotating the diseased parts of the leaves. The average number of diseases present in an image is around 15, with more than 17,000 detection targets present in total. Before starting the training process, we divide 1,180 images into training and test sets. We select 1,072 images for training the network and 108 images as the test set for evaluating the network, which was named test_pv. In addition, 108 images of grape leaves with black rot spots in the orchard environment were collected as an extra test set, which was named test_orchard. We further divide the training set into two parts during the process of network training, namely training set and validation set. The division ratio of training and validation sets is 9:1. In the convolutional neural network, the training set is used for model fitting, and the validation set is a separate sample set in the process of model training, which can be used to adjust the super parameters of the model and to preliminarily evaluate the ability of the model. The test set is used to evaluate the generalization ability of the final model. In this work, the number of epochs is 200, the input batch is 8, the learning rate is 0.001, and the size of the input image is 256 $\times$ 256. The coco dataset format is used for the dataset used in this work. We use the pre-trained model weights to accelerate the convergence. We conduct the simulations on Windows 10 based on the pytorch deep learning framework. The computer on which the tests are conducted contains 8GB GPU GeForce GTX 1070Ti and an AMD Ryzen 5 1600X Six-Core Processor. The Python language is used for programming.

Super-Resolution Enhancement of Grape Leaf Images

In order to improve the resolution of the original image, we use a software method to produce a single high-quality and high-resolution image from a set of low-quality, low-resolution

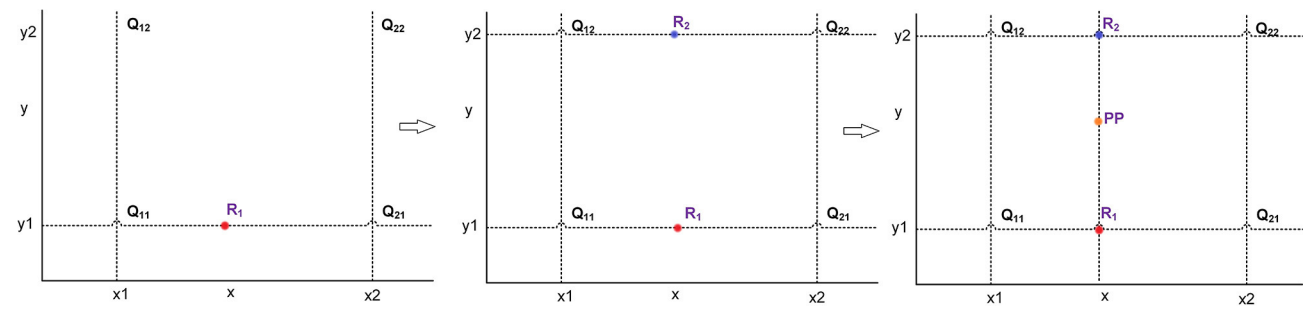

FIGURE 1 | The process of bilinear interpolation.

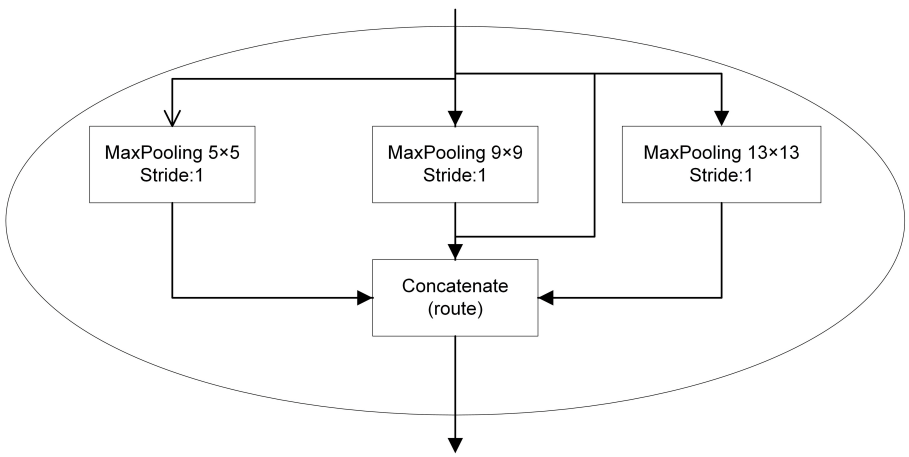

FIGURE 2 | The SPP structure in the improved YOLOv3 proposed in this work.

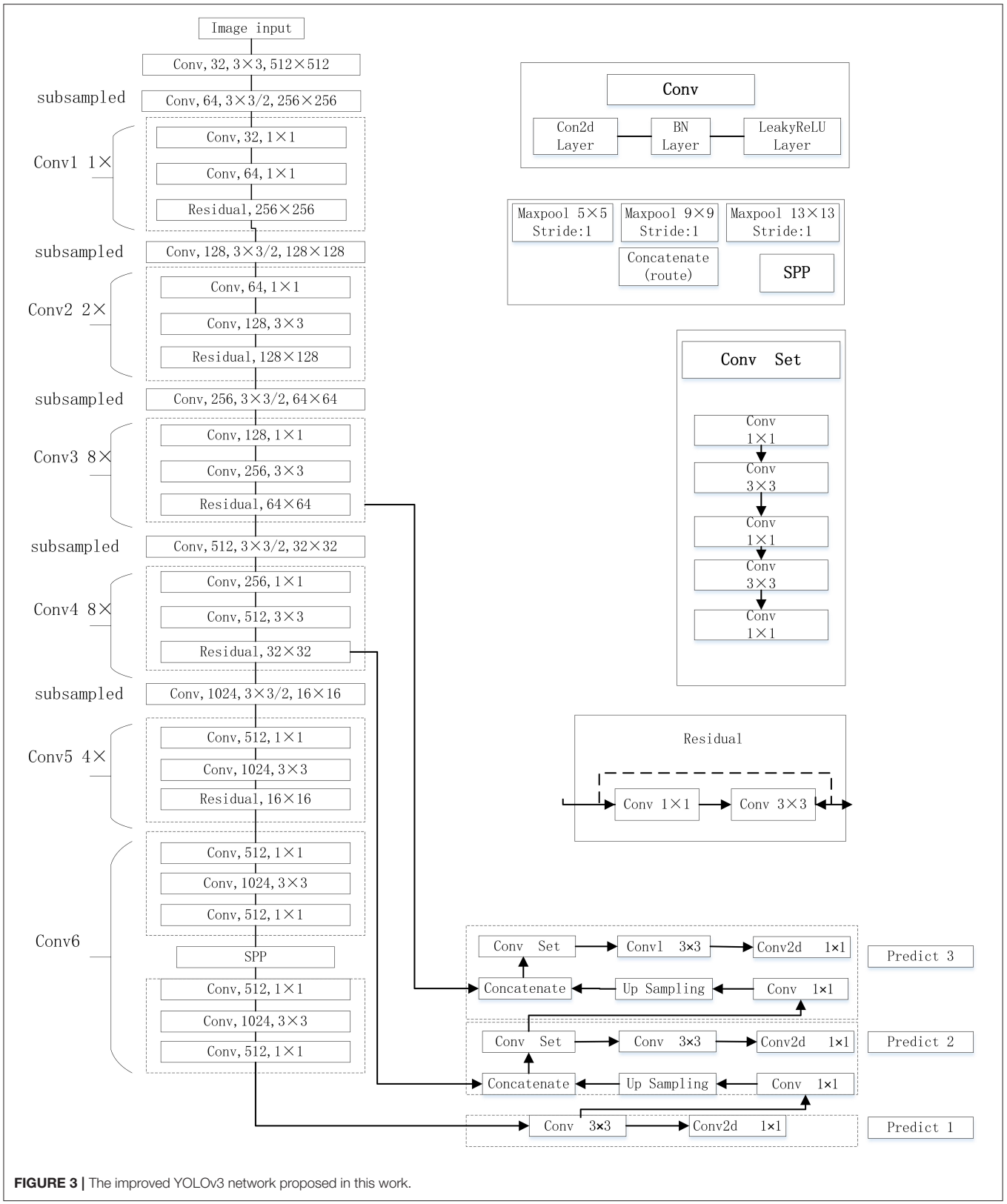

FIGURE 3 | The improved YOLOv3 network proposed in this work.

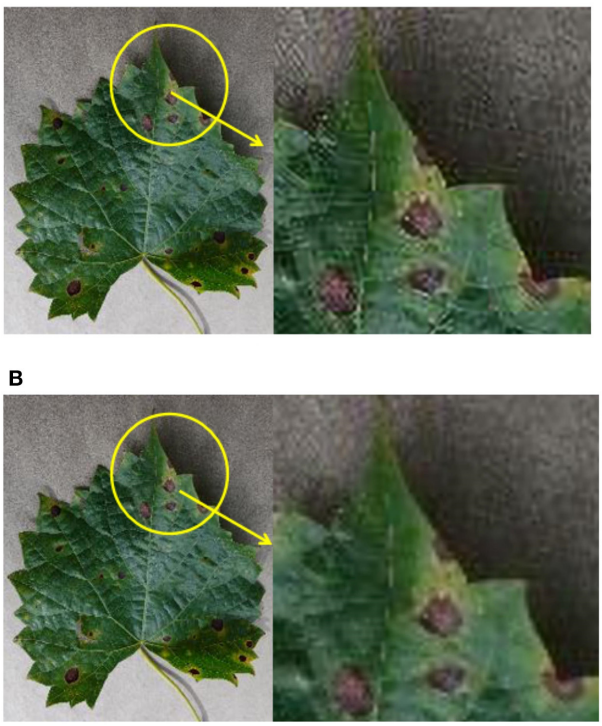

FIGURE 4 | The comparison of the original image and the enhanced image. **(A)** The original image. **(B)** Super-resolution enhanced image.

images. This method of transforming images is known as the super-resolution reconstruction (Shen et al., 2009). There are two types of image super-resolution reconstruction techniques. In the first technique, we synthesize a high-resolution image from multiple low-resolution images. In second technique, we acquire a high-resolution image from a single low-resolution image. The super-resolution techniques are divided into three categories, namely interpolation-based super-resolution, reconstruction-based super-resolution, and learning-based super-resolution. The interpolation-based super-resolution is relatively simple, and it is also widely used in various techniques. In reconstruction-based super-resolution, the main idea is to map the low-resolution images to high-resolution images. However, this technique is computationally expensive and requires a lot of computational resources. The learning-based super-resolution is implemented on the basis of convolutional neural networks (CNNs). In this work, we use the widely used interpolation method for image super-resolution. This technique is able to achieve results similar to the learning-based super-resolution images without requiring extensive computational resources. The bilinear interpolation (BL) is used for image super-resolution, which means that the original image is up-sampled for scaling operation. This is a technique for image scaling, which is mainly divided into two linear interpolation steps. First, we interpolate in the x-direction to find R_1 and R_2. The R_1 is obtained according to the values of Q_{11} and Q_{21}, and the R_2 is obtained according to the values of Q_{12} and Q_{22}. Then, the result of x-direction interpolation R1 and R2 are used to find the output of y-direction interpolation PP. The process of bilinear interpolation is shown in **Figure 1**.

In **Figure 1**, $Q_{11}(x_1, y_1)$, $Q_{12}(x_1, y_2)$, $Q_{21}(x_2, y_1)$, and $Q_{22}(x_2, y_2)$ represent the original pixel points. The first interpolation R1 along x-axis is calculated by using Q11 and Q21. The second interpolation R2 along x-axis is calculated by using Q12 and Q22. The interpolation result PP(x, y) along y-axis is calculated by using R1 and R2. The interpolation calculation process is shown in (1). The pixel points in the image are added in such a way that the image is transformed from low resolution to high resolution and super-resolution image is obtained.

$$\begin{cases} f(R_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \\ f(R_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \\ f(PP) = \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \end{cases} \quad (1)$$

Where, $f(Q_{11})$, $f(Q_{21})$, $f(R_1)$, $f(R_2)$, and f(PP) represent the value of corresponding points, respectively.

Deep Learning Method for Grape Leaf Spot Detection

In order to improve the accuracy of grapevine leaf black rot spot detection, we deign an improved YOLOv3 network. The YOLO (Redmon et al., 2016) is a typical CNN model used in target recognition, and YOLO v3 (Redmon and Farhadi, 2018) is the third version of the YOLO, which has advantages over v1 and v2 for the detection of small targets. In the improved network, we replace the loss function IOU (Intersection Over Union, IOU) with GIOU (Generalized Intersection Over Union, GIOU), and GIOU provides better regression boxes during the training process. In addition, we introduce SPP (Spatial Pyramid Pooling, SPP) before the output of feature layer, which is conducive to the detection of large differences in the target size in the image. We enhance the images by applying the linear interpolation before using the image as an input to the CNN, which is super-resolution of the image and can improve the effect of target detection.

Loss Function GIOU

In the original YOLOv3 target detection network, the engagement ratio IOU of the bounding box and the ground truth is used as the loss function. In the improved network proposed in this work, the IOU is replaced by GIOU in the deep learning network. The GIOU is calculated as shown in (2).

$$\mathrm{GIOU} = \mathrm{IOU} - \frac{|A_c \quad - U|}{|A_c|} \quad (2)$$

where, Ac denotes the area of the minimum closure region of the two boxes, i.e., the ground truth and the predicted bounding box, and U denotes the intersection area of the two boxes. By using the GIOU as a loss function, we avoid the problem caused when the two target boxes have no overlap. So, the gradient is continuously updated and better regression boxes are available during the training process.

SPP Module

The SPP works on the idea of the spatial pyramid (He et al., 2014). On the basis of SPP module, we achieve the fusion of local and global features. This feature fusion enhances the expressiveness of the feature map which is conducive in the detection of large differences in the target size in the image. In the YOLOv3-SPP network proposed in this work, the SPP module consists of four parallel branches with kernel sizes of 5 × 5, 9 × 9, 13 × 13, and 16 × 16. In this work, the input of SPP model is the 16 × 16 size image after subsampling, and the output is the fusion of four parallel branches. The SPP module is placed after the 16 × 16 convolutional layer and before the output. This structure is shown in **Figure 2**.

Improved YOLOv3-SPP Network Architecture

The YOLOv3-SPP network proposed in this work is implemented by improving the original YOLOv3. The original YOLOv3 network uses Darknet-53 as the backbone network. The Darknet-53 mainly consists of 5 residual blocks. This structure uses the idea of residual neural network, and the idea of FPN (Feature Pyramid Networks, FPN). The up-sampling fusion is adopted to detect the target independently by fusing multiple feature maps at three different scales, including 16 × 16, 32 × 32, and 64 × 64. The size of the minimum prediction frame is 8 × 8 (image size divided by grid size 512/64), which effectively obtains the feature information at low and high levels. This end-to-end network is not only more accurate but is computationally efficient as well. In this work, the SPP module is introduced in the Conv6 layer of the YOLOv3. The YOLOv3-SPP network structure is shown in **Figure 3**.

Evaluation Indicators

We use the precision (P) and recall (R) as evaluation metrics. The precision of any algorithm is computed as

$$P = \frac{TP}{TP + FP} \times 100\% \tag{3}$$

where, TP represents the true positives, which is expressed as the number of manually labeled grape disease pixels that overlap with pixels in the region automatically detected by the model as grape disease. FP represents the false positives, which is expressed as the number of pixels in the region manually considered as background, but automatically detected by the model as grape leaf region pixels. We calculate recall by using the following expression.

$$R = \frac{TP}{TP + FN} \times 100\% \tag{4}$$

where, FN denotes the false negatives, which indicates the number of pixels that are manually labeled as the grape leaf area pixels, but are detected by the model as background area pixels.

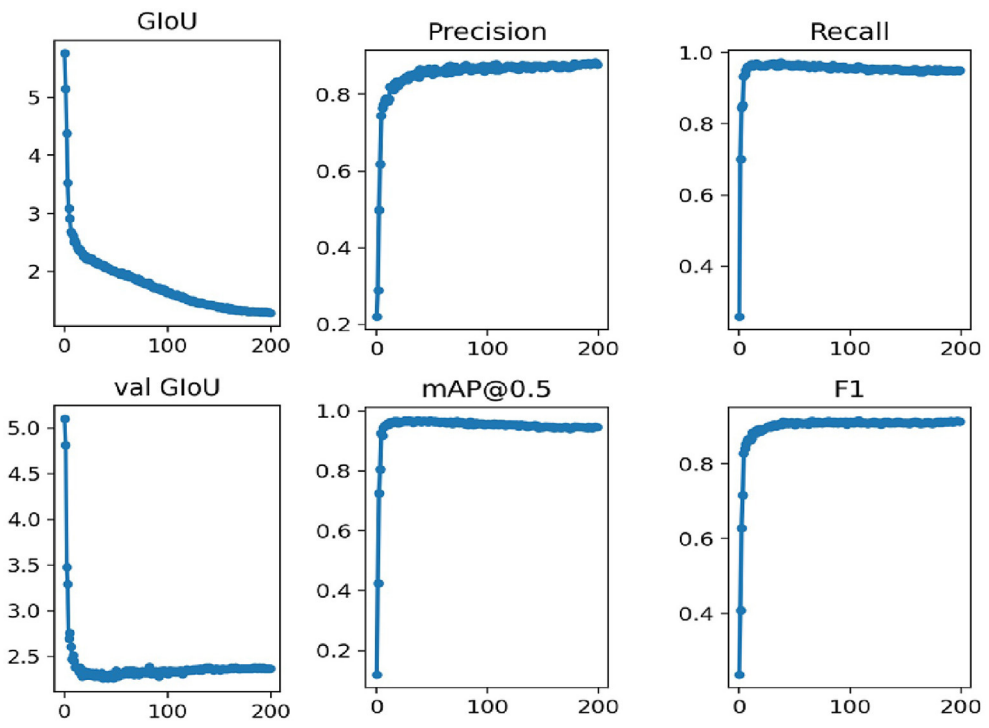

FIGURE 5 | The training performance of the original dataset YOLOv3-SPP network.

RESULTS

Super-Resolution-Based Image Enhancement Results

The size of an image in the original dataset is 256 × 256, and the average image size is around 20 kb. After the application of super-resolution technique, we enhance the input image to a size of 512 × 512. Now, the average image size is 100 kb. **Figure 4** shows the comparison between the original image and the resulting enhanced image. It is evident from **Figure 4** that the local parts of the image are equally magnified 4 times. The clarity of the image after the super-resolution is significantly higher than that of the original image.

Original Image YOLOv3-SPP Detection Results

In this work, we use the annotated images to train the network. The network is trained for 200 epochs which takes around 6 h. The training results are presented in **Figure 5**.

In **Figure 5**, the GIOU denotes the loss function uses in this work, the val GIOU denotes the validated loss function, map@0.5 denotes the average accuracy, and F1 represents an evaluation metric which combines the effect of precision and recall. It is evident from **Figure 5** that the network converges rapidly for the first 100 iterations. In addition, the loss function GIOU decreases rapidly until it gets flat around the 100th iteration. The evaluation metrics, i.e., precision, recall, map@0.5, and F1 also become flat at this stage.

Enhanced Image YOLOv3-SPP Detection Results

The enhanced image dataset with annotation information is fed into the YOLOv3-SPP network for training. Please note that the network parameters are the same as used in the case of training the network on original images. This training process consumes around 7 h. The training results are presented in **Figure 6**. It is evident form **Figure 6** that the model is continuously optimized during the epochs, and the loss function decreases rapidly in the initial iteration until it gets flat. This is consistent with training the network using original images which indicates that the improved YOLOv3-SPP network converges rapidly regardless of whether it uses the original images or the enhanced images as the input data. The trend of precision is similar to the case of training the network on original images. However, other evaluation metrics, i.e., recall, map@0.5, and F1 rise rapidly and reach a smooth state at the beginning of the training and show a small decrease after 100 iterations.

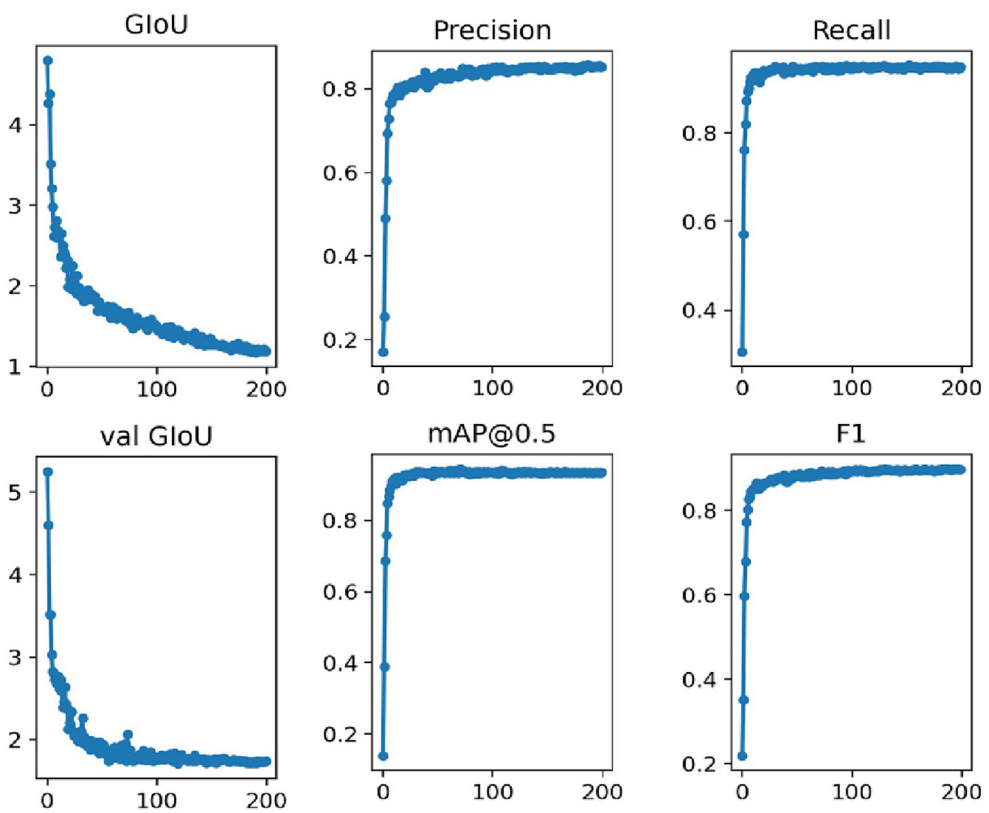

FIGURE 6 | The training performance of YOLOv3-SPP network by using the enhanced dataset.

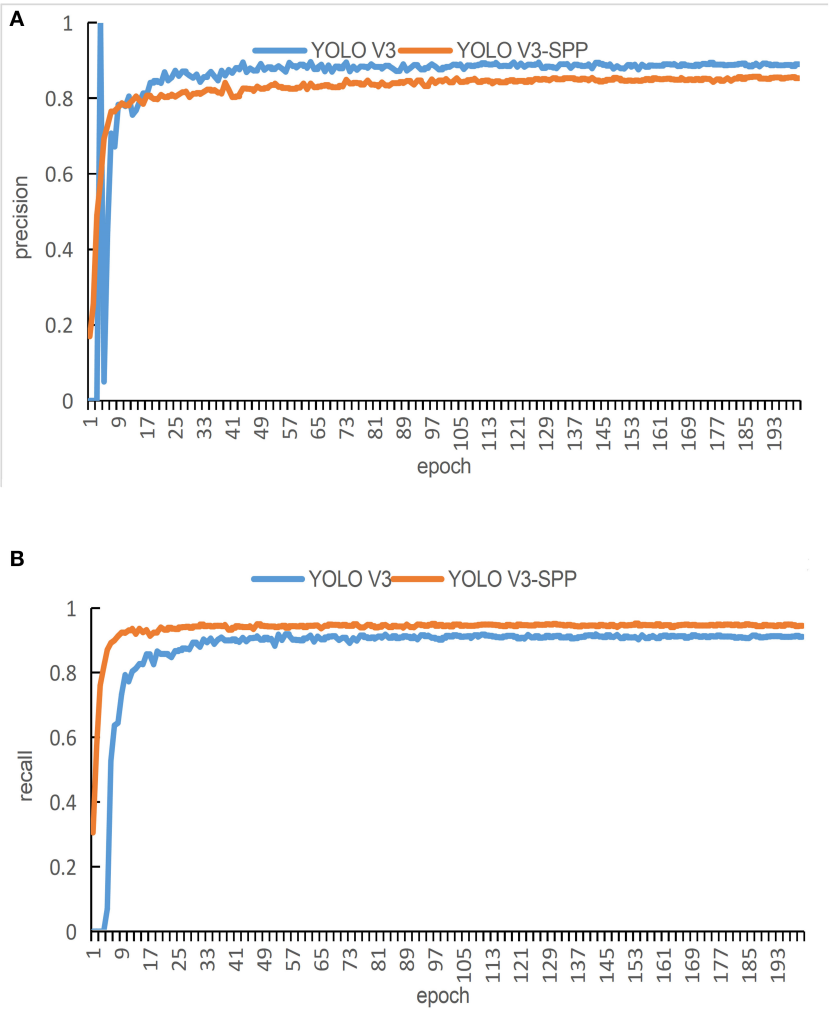

FIGURE 7 | The training results of the detection algorithm by using the original images. **(A)** A comparison of the original and improved YOLOv3 in terms of precision for detection. **(B)** A comparison of the original and improved YOLOv3 in terms of recall for detection.

DISCUSSION

Comparison of the Effect of YOLOv3 Network Improvement Before and After

In order to evaluate the performance of the proposed network in terms of detection accuracy, we train the original YOLOv3 network and the YOLOv3-SPP network by using the original images. The precision and recall of both techniques are compared in **Figure 7**. **Figure 7A** presents the comparison of precision and **Figure 7B** presents the comparison of recall of both algorithms.

It is evident from **Figure 7A** that the red curve converges faster than the blue curve. In addition, the blue curve shows huge fluctuations at the beginning epochs. These fluctuations continue for around 100 epochs. After 100 epochs in the training process, both curves reach a relatively stable state. The curve at the final epochs shows that the blue curve maintains an interval of about 2% from the red curve, i.e., the blue curve is more precise than the red curve. Therefore, in terms of the detection, the original YOLOv3 performs slightly better than the proposed YOLOv3-SPP.

It is evident form **Figure 7B** that the red curve converges faster as compared with the blue curve. There is an interval of about 3% between the two curves after reaching a steady state. During the entire training process of 200 epochs, the precision of the red curve is always greater than the blue curve. In terms of the detection, the proposed YOLOv3-SPP performs better than the original YOLOv3.

In order to further verify the detection accuracy of the original and the proposed YOLOv3, we present the recognition results of the test_pv in **Table 1**. There are 108 images with 1,532 spots of grape leaf black rot in the test_pv set. The results show that the proposed YOLOv3-SPP successfully identifies 1,427 spots. Contrary, the original YOLOv3 only identifies a total of 1,283 spots. It is also noteworthy that the proposed YOLOv3-SPP misidentified 87 spots, which is 58 less than the misdetections of original YOLOv3. The number of FN of YOLOv3-SPP was

TABLE 1 | The detection results of the test_pv before and after the improvement of YOLOv3 network.

Original image of the test_pv with 1,532 objectives	TP	FP	FN	Precision	Recall
YOLOv3	1,283	145	259	89.85%	83.75%
YOLOv3-SPP	1,427	87	105	94.25%	93.15%

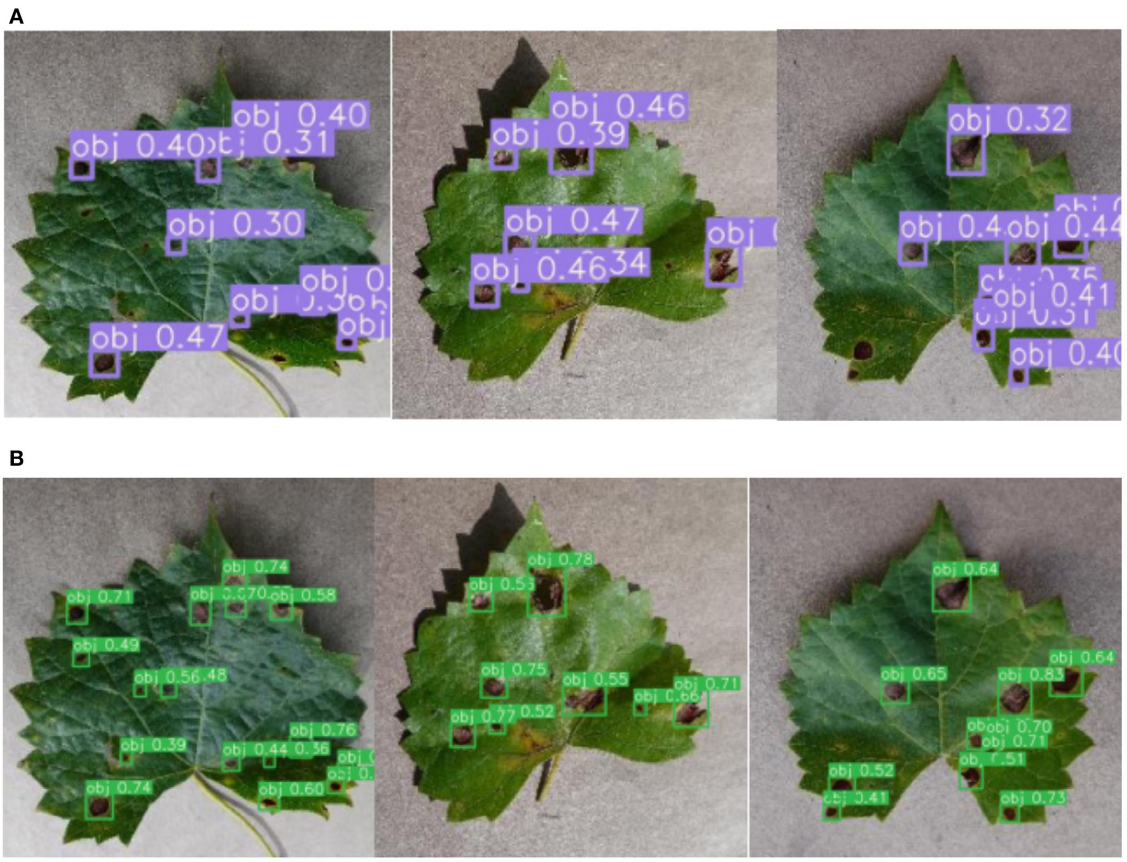

FIGURE 8 | The comparison of detection results before and after the improvement of YOLOv3 network. **(A)** The detection results of the original YOLOv3 for original images. **(B)** The detection results of the improved YOLOv3-SPP for original images.

105, which was 154 less than that of YOLOv3. The proposed YOLOv3-SPP achieves the detection accuracy of 94.25% and a recall of 93.15%. These results are 4 and 9% higher than the original YOLOv3, respectively.

Figure 8 shows the detection results of the original and the improved algorithm. **Figure 8A** presents the recognition results of the original YOLOv3 and **Figure 8B** presents the recognition results of the improved YOLO V3-SPP. It is evident from **Figure 8** that the recognition results of the improved algorithm are significantly better than the original algorithm. It is noteworthy that the targets missed by the original network are also recognized by the improved network. This shows that the addition of SPP module and the replacement of the loss function in the improved algorithm improve the recall of the detected targets effectively. Although the accuracy of the improved network is slightly lower than the original network. This indicates that the overall performance of the proposed YOLOv3-SPP algorithm is higher than that of the original YOLO V3, and the results of the improved method for grapevine leaf black rot detection are satisfactory.

Comparison of Different Super-Resolution Algorithms for Image Enhancement

In this work, the BL is used to enhance the images. The enhanced images are then used for target detection. In this work, before we select the BL as a choice for this work, we compare the two other different super-resolution methods, i.e., the nearest interpolation and the enhanced deep residual networks (EDSR) (Lim et al., 2017) with BL. The nearest interpolation is a traditional and simplest interpolation method which is used to enhance the image by directly copying the values of neighboring pixels. On the other hand, EDSR has the ability to handle multiple

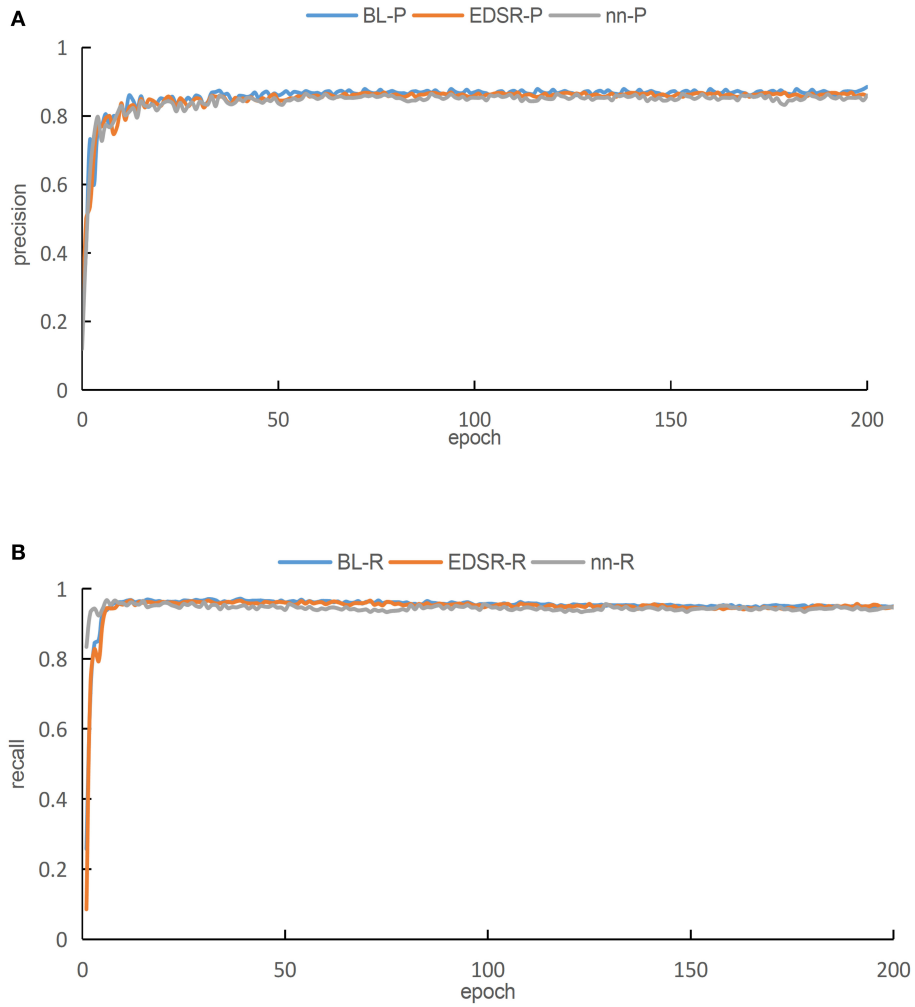

FIGURE 9 | The training results of YOLOv3-SPP network after image enhancement by using different super-resolution algorithms. **(A)** A comparison of accuracies for different super resolutions. **(B)** A comparison of recall for different super resolutions.

scaling factors (scales) of super-resolution simultaneously in a single network. The training process of EDSR models using multiple scales significantly improve the performance. However, EDSR-type architectures require bicubic interpolation. This interpolation technique is computationally expensive and needs more storage space as well.

In this work, we enhance the original images in the training dataset comprising 1,072 images by using all the three aforementioned super-resolution methods. These images are input into the proposed YOLOv3-SPP network for training. The training results are shown in **Figure 9**. Please note that **Figure 9A** presents a graph comparing the accuracies of different super-resolutions. As presented in **Figure 9**, the blue curve represents the result of BL, the red curve represents the result of EDSR, and the green curve represents the results of nearest interpolation.

As presented in **Figure 9A**, it is evident that when the images of the three different interpolation methods are fed into the proposed YOLOv3-SPP network separately, after 50 iterations, the precisions of all the networks tend to be stable. After 50 iterations, the precision of all three training methods fluctuates minutely, but the overall trend is stable. The comparison shows that the blue curve is better than the other two curves after achieving steady state. On the other hand, the green curve has a lower precision than the other two algorithms after becoming steady state. Therefore, we conclude that the BL technique

TABLE 2 | The evaluation of test_pv by using different super-resolution methods.

Evaluation indicators	BL	Nearest interpolation	EDSR
TP	1,448	1,430	1,450
FP	65	78	70
FN	84	102	82
Precision	95.79%	94.83%	95.39%
Recall	94.52%	93.34%	94.65%

has the highest accuracy rate P as compared with the other two super-resolution enhancement methods under the same parameters. As presented in **Figure 9B**, when the images of the three different interpolation methods are fed into the YOLOv3-SPP network separately, the resulting recalls approach 1 after 20 iterations approximately.

Please note that the red and blue curves are almost equal, however, the green curve is slightly below the other two curves. After processing the original image by using the three super-resolution image enhancement methods, the recalls approach 1, which are all higher than directly training the network using the original images. This indicates that the super-resolution enhancement of the images before using it as an input of the CNN for detection is better than using the original images directly. The evaluation metrics, i.e., precision and recall, make it evident that the network performs best when the image is enhanced using the BL technique as compared to the nearest interpolation. As compared to the EDSR method, the result of image enhancement performed using BL method has higher precision. and almost equal recall. However, the BL method has no complex residual convolution and is relatively less computationally intensive. We enhance 108 images in tese_pvaccording to the three

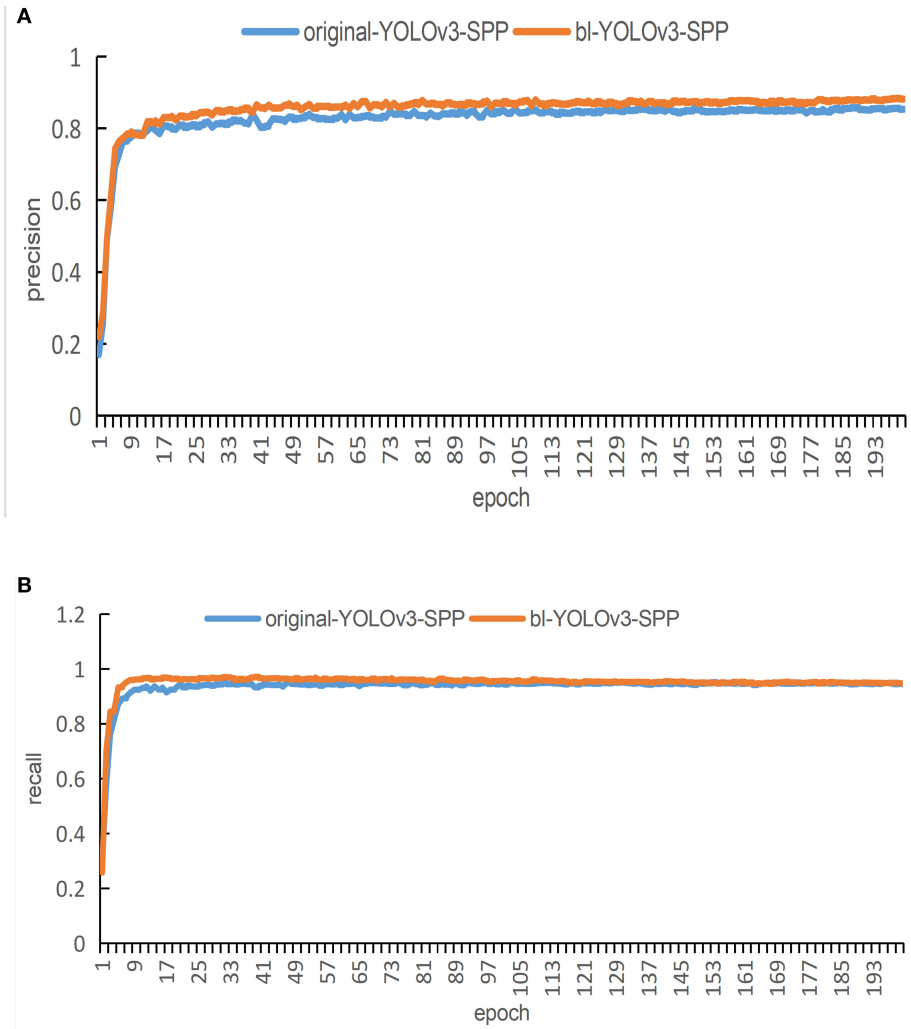

FIGURE 10 | The training results of the network before and after image enhancement. **(A)** A comparison of the detection accuracy under super-resolution image and original image. **(B)** A comparison of the detection recalls under the super-resolution image and the original image.

TABLE 3 | The detection results on test_pv set.

YOLOv3-SPP	Objectives	TP	FP	FN	Precision	Recall
Original image of test_pv	1,532	1,427	87	105	94.25%	93.15%
Super Resolution of test_pv	1,532	1,448	65	84	95.79%	94.52%

aforementioned methods and then use the resultant images to perform detections using the proposed YOLOv3-SPP network. The corresponding results are shown in **Table 2**. Please note that the BL method has the highest recognition precision for the spots and the recall rate is only 0.13% lower than the EDSR method. The results of simulations performed using test_pv show that all the image enhancement methods perform well, however, considering the overall performance using the evaluation metrics shows that the BL method is superior.

Comparison of the Detection Effect of Super-Resolution Image and Original Image Under the Improved Network

The training results of the proposed YOLO V3-SPP network by using the original and the enhanced images are presented in **Figure 10**. The blue and red curves in **Figure 10** represent the training results for original and improved images, respectively.

Figure 10A shows the comparison of detection accuracy under the super-resolution image and the original image. The results show that the red curve and the blue curve almost converge simultaneously, however, the red curve after smoothing is close to 0.9. The blue curve lags behind the red by about 1.2%. Both curves have occasional fluctuations during the process of convergence. After reaching 140 epochs, both curves smooth out. After enhancing the images using super-resolution, the accuracy of the network is higher than the accuracy obtained by using the unenhanced images.

Figure 10B shows the comparison of detection accuracy under the super-resolution images and the original images. The results show that the two curves rapidly stabilize at the beginning of the training process. Please note that the blue curve is lower than the red curve for first 100 epochs, but after 100 iterations both curves are relatively stable. The figure shows that both curves exhibit good recall. This indicates that after enhancing the images using super-resolution, the performance of the network reflects that its recall is not lower than that of the unenhanced images.

The test result of the proposed YOLOv3-SPP trained with the original and the enhanced images in test_pv are shown in **Table 3**. The YOLOv3-SPP trained using the enhanced images correctly identifies a total of 1,448 spots, that is 21 more spots than the YOLOv3-SPP trained using the original images. In addition, the number of false identifications is reduced by 22, and the number of miss identifications decreased by 21. In terms of precision and recognition rate, the proposed YOLOv3-SPP network trained on the enhanced images improved the performance by 1.54 and 1.37%, respectively.

Figure 11 shows the recognition results before and after the images are enhanced. **Figure 11A** shows the detection results of YOLOv3-SPP trained using the unenhanced images. **Figure 11B** presents the detection results of YOLOv3-SPP trained using the enhanced images. As presented in **Figure 11**, the proposed

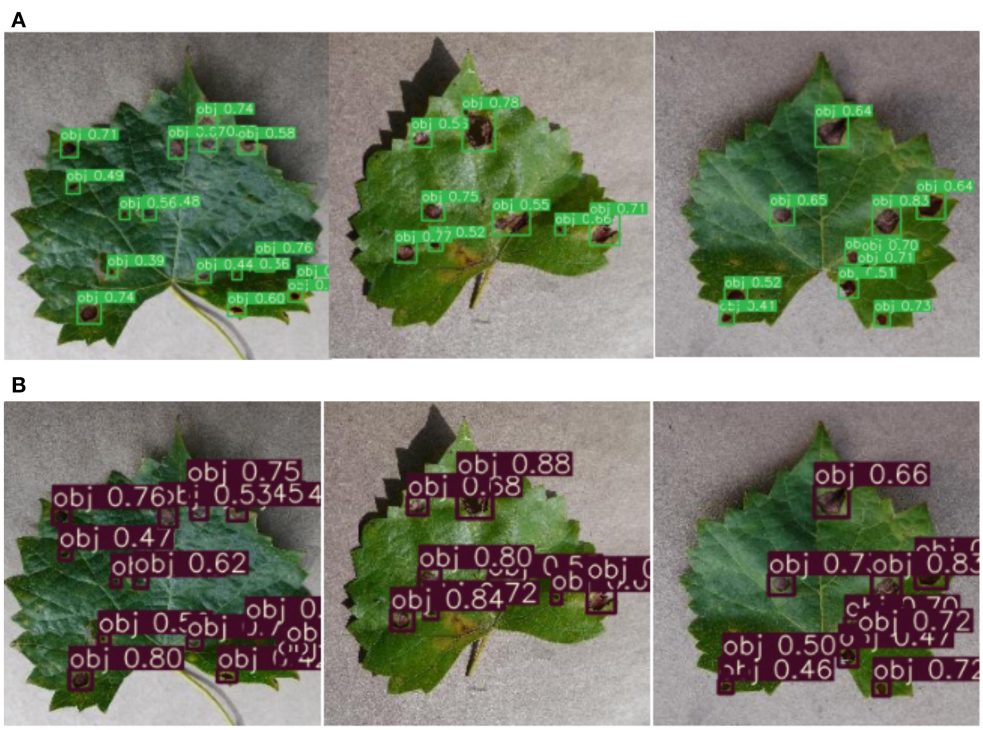

FIGURE 11 | The detection results before and after the data is enhanced for test_pv. **(A)** The detection results of YOLOv3-SPP trained on the unenhanced images. **(B)** The detection results of YOLOv3-SPP trained on the enhanced images.

YOLOv3-SPP trained using the enhanced images to detect the black rot of grape leaves has significantly better results than the recognition results of the network trained using the unenhanced images. This indicates that the introduction of super-resolution images for grape leaf black rot detection improves the accuracy of the target detectors without reducing the recall rate.

The Detection Effect in the Orchard Environment

The method of super-resolution image enhancement and deep learning can improve the detection effect of grape leaf black rot, which has been proved in the test_pv data set. An additional test set, test_orchard, was used to test the effectiveness of the proposed method in the orchard environment. There are 108 images of grape leaves with 1,275 spots of grape leaf black rot from different orchard environments. The results of spot identification by YOLOv3-SPP were shown in **Figure 12**. It can be seen from **Figure 12** that not only the detection precision of the spot was improved, but also the recall was improved after enhancing the images using super-resolution. Some undetected spots in the original image were identified on the enhanced images. The statistical data of identification results of the test_orchard was shown in **Table 4**. There are 1,275 spots in the test set, and 1,028 spots were recognized after inputting the original images into the YOLOv3-SPP, of which 186 spots were misidentified. The number of unrecognized spots was 247. The precision of disease spot detection was 84.68% and the recall was 80.63% for the original images of test_orchard. However, a total of 1,049 disease spots were identified for the images after enhancement by super-resolution, of which 161 spots were misidentified. The number of unrecognized spots was 247. The precision of disease spot detection was 86.69% and the recall was 82.27% for the super-resolution images of test_orchard. The number of misidentified and unrecognized spots decreased, and the precision and recall increased 2.01 and 1.64%, respectively after the images were enhanced by super-resolution.

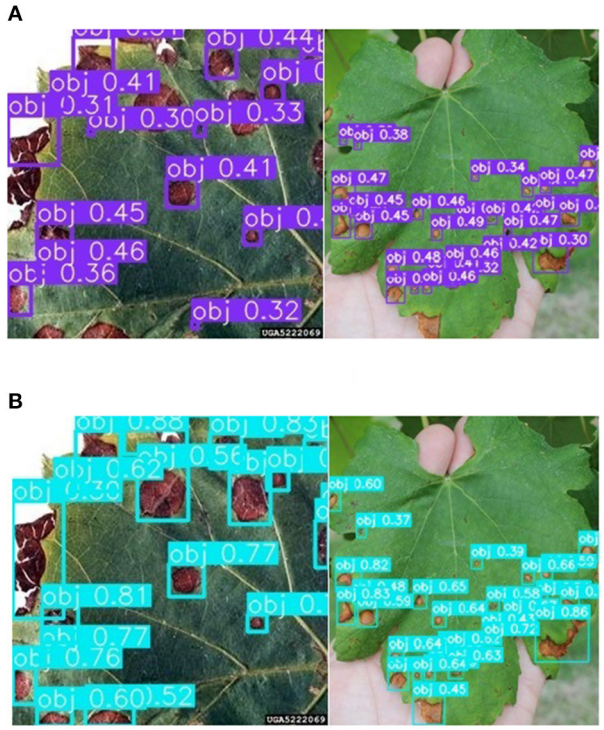

FIGURE 12 | The detection results before and after the data is enhanced for test_orchard. **(A)** The detection results of YOLOv3-SPP trained on the unenhanced images. **(B)** The detection results of YOLOv3-SPP trained on the enhanced images.

The detection precision and recall of test_orchard were lower than that of test_pv, because the images of test_orchard are from orchards, while the images of test_pv from Plant Village, which photograph indoors. The environment of orchards is complex compared to the indoor. The images of the test_orchard were classified into single-leaf and multi-leaf based on the number of grape leaves in the images, to compare the influence of different image acquisition ways on the detection effect. The detection results of the multi-leaf images were shown in **Figure 13**, and it can be seen that the leaf slits to be misrecognized as disease spots, which reduced the precision. The statistical results of the detection were shown in **Table 5** for the two category images. There are 701 spots in single-leaf images totally, 601 spots were detected, 71 disease spots were wrong identified and 91 spots were missed. The detection precision was 89.57% and the recall was 87.02% for the single-leaf images. A total of 574 disease spots in multi-leaf images and 135 spots were missed detection. Among the identified 439 spots, 90 spots were misidentified. The detection precision was 82.99% and the recall was 76.48% for the multi-leaf images. The detection precision and recall of single-leaf images are 6.76 and 10.54% higher than that of multi-leaf images.

This shows that the acquisition of single grape leaf image is more conductive to detection.

The background of images also affected the detection effect, so the set of test_orchard was divided into two subsets, simple background images and complex background images based on the status of the background. The images are considered to be complex background, which concludes fruits, branches, and soil except for grape leaf, otherwise, they are considered simple background. The detection results of the complex background images were shown in **Figure 14**, and it can be seen that the rotten fruits, gray branches, and soil to be misrecognized as disease

TABLE 4 | The detection results on test_orchard set.

YOLOv3-SPP	Objective	TP	FP	FN	Precision	Recall
Original image test_orchard set	1,275	1,028	186	247	84.68%	80.63%
Super-resolution test_orchard set	1,275	1,049	161	226	86.69%	82.27%

FIGURE 13 | Detection results of multi-leaf images.

TABLE 5 | The statistical results of the detection of single-leaf and multi-leaf images for test_orchard.

BL-YOLOv3-SPP	Objective	TP	FP	FN	Precision	Recall
Single-leaf	701	610	71	91	89.57%	87.02%
Multi-leaf	574	439	90	135	82.99%	76.48%

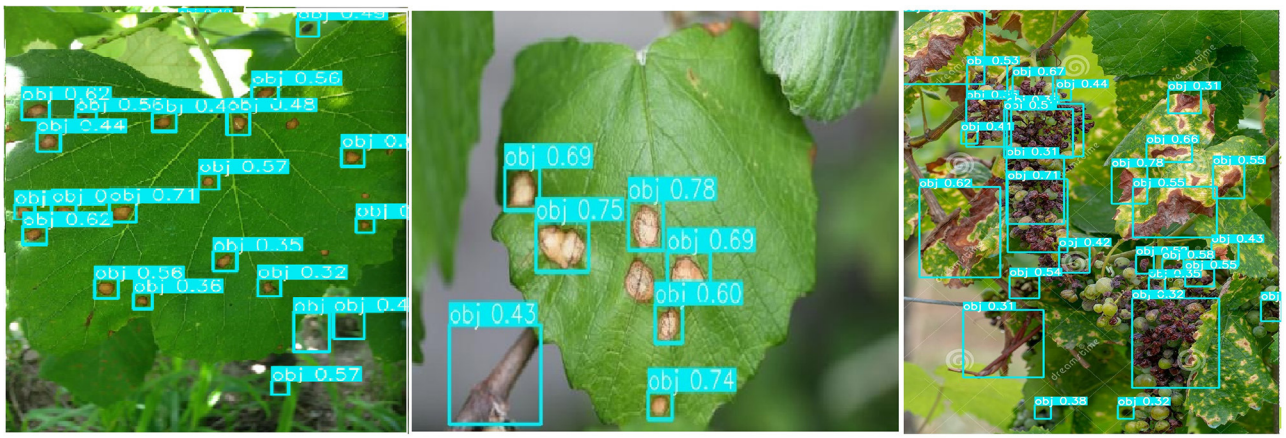

FIGURE 14 | Detection results of grapevine leaf black rot with complex background.

TABLE 6 | The statistical results of the detection of different background for test_orchard.

BL-YOLOv3-SPP	Objective	TP	FP	FN	Precision	Recall
Simple background	712	664	42	48	94.05%	93.26%
Complicated background	563	385	119	178	76.39%	68.38%

spots, which also reduced the precision. The statistical results of the detection were shown in **Table 6** for the test_orchard images in the different backgrounds. There are 712 spots in simple background images totally, 664 spots were detected, 42 disease spots were wrong identified and 48 spots were missed. The detection precision was 94.05% and the recall was 93.26% for the simple background images, which was close to the detection effect of test_pv. A total of 563 disease spots in complex background images and 178 spots were missed detection. Among the identified 386 spots, 119 spots were misidentified. The detection precision was 76.39% and the recall was 68.38% for the complex background images. The detection precision and recall of simple background images are 17.66 and 24.9% higher than that of complex background images. It can also be seen from the data in **Table 6**, that the influence of background in the image is greater than that of multiple leaves in the image for the grape leaf black rot.

The method proposed in this work can be used for the detection of grape leaf black rot in the natural environment through the test and analysis of the images collected from

orchards, and the detection effect is satisfactory especially in the case of the simple background. At the same time, the analysis results also provide a reference for the field image acquisition, that is, to avoid other objects appearing in the image except for grape leaf.

CONCLUSIONS

In this work, we propose an improved YOLOv3-SPP model for the detection of black rot of grape leaves. This method replaces the loss function in the original YOLOv3 with GIOU. In addition, we also add the SPP module. We enhance the training images of YOLOv3-SPP by using BL super-resolution method. Two test sets from Plant Village dataset and orchards are performed on the model. The results show that the YOLOv3-SPP network performs better for grape leaf black rot detection and has a precision of 95.79% and recall of 94.52% for the test set from Plant Village. For the orchards test set, the precision is 86.69% and the recall is 82.27%, it also has better performs than the original images of the test set. In addition, the precision and recall are improved to 94.05 and 93.26% for those images without fruits, branches, and soil in the background. Moreover, the image enhancement of the training set using the BL method improves the results in terms of precision and recall. The current method requires image enhancement and then trains the deep learning network. In future work, we will attempt to combine these steps.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JZ, MC, and HY: conceived the idea and proposed the method. JZ and QW: contributed to the preparation of equipment and acquisition of data, and wrote the code and tested the method. JZ, QW, and MC: validation results. JZ and HY: wrote the paper. JZ, HY, and ZC: revised the paper. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 32001412) and the Key Research and Development Program of Hebei Province (19227206D).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.695749/full#supplementary-material

REFERENCES

Agrawal, N., Singhai, J., and Agarwal, D. K. (2017). "Grape leaf disease detection and classification using multi-class support vector machine," in *2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)* (Bhopal). doi: 10.1109/RISE.2017.8378160

Alessandrini, M., Rivera, R. C. F., Falaschetti, L., and Pau, D. (2021). A grapevine leaves dataset for early detection and classification of Esca disease in vineyards through machine learning. *Data Brief* 35:1068009. doi: 10.1016/j.dib.2021.106809

Arun, P. V., Buddhiraju, K. M., Porwal, A., and Channusot, J. (2019). Cnn based spectral super-resolution of remote sensing images. *Signal Process.* 169:107394. doi: 10.1016/j.sigpro.2019.107394

Bai, Y., Zhang, Y., and Ding, M. (2018a). "Sod-mtgan: small object detection via multi-task generative adversarial network," in European Conference on Computer Vision. Computer Vision – ECCV *2018* (Munich), 210–226. doi: 10.1007/978-3-030-01261-8_13

Bai, Y., Zhang, Y., Ding, M., and Ghanem, B. (2018b). "Finding tiny faces in the wild with generative adversarial network," in *Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE). doi: 10.1109/CVPR.2018.00010

Bosquet, B., Mucientes, M., and Brea, V. M. (2018). "STDnet: a convnet for small target detection," in *Proc. Brit. Mach. Vis. Conf*. Available online at: http://www.bmva.org/bmvc/2018/contents/papers/0897.pdf

Chen, C., Zhou, G., Chen, A., Yi, J., and Hu, Y. (2020).Identification of tomato leaf diseases based on combination of abck-bwtr and b-arnet. *Comput. Electr. Agric.* 178:105730. doi: 10.1016/j.compag.2020.105730

Es-Saady, Y., Massi, I. E., Yassa, M. E., Mammass, D., and Benazoun, A. (2016). "Automatic recognition of plant leaves diseases based on serial combination of two SVM classifiers," in *2016 International Conference on Electrical and Information Technologies* (ICEIT) (Tangier: IEEE). doi: 10.1109/EITech.2016.7519661

Felzenszwalb, P., Mcallester, D., and Ramanan, D. (2008). "A discriminatively trained, multiscale, deformable part model," in *IEEE Conference on Computer Vision and Pattern Recognition* (Anchorage, AK: IEEE). doi: 10.1109/CVPR.2008.4587597

Girshick, R. (2015). "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.* (Washington, DC), 1440–1448. doi: 10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (Columbus, OH). doi: 10.1109/CVPR.2014.81

He, K., Zhang, X., Ren, S., and Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904–1916. doi: 10.1109/TPAMI.2015.2389824

Ji, M., Zhang, L., and Wu, Q.(2020) Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks. *Inform. Process. Agric.* 7, 418–426. doi: 10.1016/j.inpa.2019.10.003

Kole, D. K., Ghosh, A., and Mitra, S. (2014). "Detection of Downy Mildew disease present in the grape leaves based on fuzzy set theory," in *Advanced Computing, Networking and Informatics, Vol. 1*, eds M. Kumar Kundu, D. Mohapatra, A. Konar, and A. Chakraborty (Cham: Springer), 377–384. doi: 10.1007/978-3-319-07353-8_44

Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M. (2017). "Enhanced deep residual networks for single image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Honolulu, HI). doi: 10.1109/CVPRW.2017.151

Liu, B., Ding, Z., Tian, L., He, D., and Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Front. Plant Sci.* 11:1082. doi: 10.3389/fpls.2020.01082

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., and Fu, C. Y., et al. (2016). "SSD: single shot multibox detector," in *European Conference on Computer Vision.Computer Vision – ECCV 2016* (Amsterdam), 21–37. doi: 10.1007/978-3-319-46448-0_2

Molitor, D., and Berkelmann-Loehnertz, B. (2011). Simulating the susceptibility of clusters to grape black rot infections depending on their phenological development. *Crop Protect.* 30, 1649–1654. doi: 10.1016/j.cropro.2011.07.020

Noh, J., Bae, W., and Lee, W. (2019). "Better to follow, follow to be better: towards precise supervision of feature super-resolution for small object detection," in *International Conference on Computer Vision (ICCV)* (Seoul). doi: 10.1109/ICCV.2019.00982

Pearson, R. C., and Goheen, A. C. (1989). Compendium of grape diseases. *Mycologia* 81:176. doi: 10.2307/3759482

Rabbi, J. (2020). *Tiny Object Detection in Remote Sensing Images:End-to-End Super-Resolution and Object Detection with Deep Learning*. Thesis. University of Alberta, Edmonton, AB, Canada. doi: 10.20944/preprints202003.0313.v3

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Computer Vision and Pattern Recognition* (Boston, MA: IEEE). doi: 10.1109/CVPR.2016.91

Redmon, J., and Farhadi, A. (2017). "Yolo9000: better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI), 6517–6525. doi: 10.1109/CVPR.2017.690

Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. *arXiv [Preprint]*. Available online at: https://arxiv.org/abs/1804.02767

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Shen, H., Ng, M. K., Li, P., and Zhang, L. (2009).Super-resolution reconstruction algorithm to modis remote sensing images. *Comput. J.* 52, 90–100. doi: 10.1093/comjnl/bxm028

Wagh, T. A., Samant, R. M., Gujarathi, S. V., and Gaikwad, S. B. (2019). Grapes leaf disease detection using convolutional neural network. *Int. J. Comput. Appl.* 178, 7–11. doi: 10.5120/ijca2019918982

Waghmare, H., Kokare, R., and Dandawate, Y. (2016). "Detection and classification of diseases of Grape plant using opposite colour Local Binary Pattern feature and machine learning for automated Decision Support System," in *International Conference on Signal Processing and Integrated Networks* (Noida: IEEE). doi: 10.1109/SPIN.2016.7566749

Xie, Q., Yao, G., and Liu, P. (2017). Super-resolution reconstruction of satellite video images based on interpolation method. *Procedia Comput. Sci.* 107, 454–459. doi: 10.1016/j.procs.2017.03.089

Xie, X., Ma, Y., Liu, B., He, J., and Wang, H. (2020). A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Front. Plant Sci.* 11:751. doi: 10.3389/fpls.2020.00751

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhu, Cheng, Wang, Yuan and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 01 July 2021
doi: 10.3389/fpls.2021.630425

TheLNet270v1 – A Novel Deep-Network Architecture for the Automatic Classification of Thermal Images for Greenhouse Plants

Md. Parvez Islam[1], Yuka Nakano[2], Unseok Lee[1], Keinichi Tokuda[1] and Nobuo Kochi[1]*

[1]Agricultural AI Research Promotion Office, RCAIT, National Agriculture and Food Research Organization (NARO), Tsukuba, Japan, [2]Institute of Vegetable and Flower Research, NARO, Tsukuba, Japan

The real challenge for separating leaf pixels from background pixels in thermal images is associated with various factors such as the amount of emitted and reflected thermal radiation from the targeted plant, absorption of reflected radiation by the humidity of the greenhouse, and the outside environment. We proposed TheLNet270v1 (thermal leaf network with 270 layers version 1) to recover the leaf canopy from its background in real time with higher accuracy than previous systems. The proposed network had an accuracy of 91% (mean boundary F1 score or BF score) to distinguish canopy pixels from background pixels and then segment the image into two classes: leaf and background. We evaluated the classification (segment) performance by using more than 13,766 images and obtained 95.75% training and 95.23% validation accuracies without overfitting issues. This research aimed to develop a deep learning technique for the automatic segmentation of thermal images to continuously monitor the canopy surface temperature inside a greenhouse.

Keywords: deep learning, network architecture, classification, segmentation, thermal image

OPEN ACCESS

Edited by:
Reza Ehsani, University of California, Merced, United States

Reviewed by:
Pouria Sadeghi-Tehran, Rothamsted Research, United Kingdom
Yiannis Ampatzidis, University of Florida, United States

****Correspondence:***
Md. Parvez Islam
p.islam@affrc.go.jp

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *17 November 2020*
Accepted: *02 June 2021*
Published: *01 July 2021*

Citation:
Islam MP, Nakano Y, Lee U, Tokuda K and Kochi N (2021) TheLNet270v1 – A Novel Deep-Network Architecture for the Automatic Classification of Thermal Images for Greenhouse Plants. Front. Plant Sci. 12:630425. doi: 10.3389/fpls.2021.630425

INTRODUCTION

Leaf surface and internal structure changes are due to adverse growth, stomatal resistance, diseases, leaf angles, depth of the canopy, and water stress conditions, which alter the absorbance-reflection process of solar radiation (Lili et al., 1991; Kraft et al., 1996; Raza et al., 2015). Thermography detected this reflected (emitted) long-wave infrared (8–14 μm), then converted it into thermal images, and a false-color gradient demonstrated the temperature level of the plant leaves of canopies (Chaerle and Van Der Straeten, 2000). **Figure 1** shows the working principle of a thermal camera.

Over the last few years, the advancement of fast computing power, low-cost imaging systems with image processing software, and deep learning (DL) techniques have allowed for nondestructive disease diagnosis and detection of various stress conditions of plants in a timely manner (Liu and Wang, 2020). The DL based on a convolution neural network (CNN) is the successor of traditional machine learning approaches that can learn features with greater precision and accuracy by activating maximum networkability (Christopher et al., 2018). Bengio (2009) compared CNN-based DL with the Neocortex of the human brain, which learns response-based features dynamically from images.

CNN-based DL acquires hierarchical features and emphasizes nonlinear filters of the depth of the deep network structure for learning, and after that solves problem-specific tasks such as image classification, semantic segmentation (pixel-based classification), object detection, video processing, speech recognition, and natural language processing (Simonyan and Zisserman, 2014; Singh et al., 2018; Liu and Wang, 2020). Khan et al. (2020) classified deep network architectures into seven classes: spatial exploitation, depth, multi-path, width, feature-map exploitation, channel boosting, and attention-based CNNs. **Figure 2** demonstrates the classification of various deep network architectures along with the proposed TheLNet270v1. Shin et al. (2016) stated that a filter termed as a channel in a CNN can extract different levels of information (from fine-grained to coarse-grained) based on their sizes (small to large sizes). Simonyan and Zisserman (2015) and Khan et al. (2020) stated that the deep DL architecture has an advantage over the shallow depth DL architecture, which can learn complex representations at different levels of abstraction and thus increase the classification accuracy. According to Szegedy et al. (2015), branching within layers can abstract features with various spatial

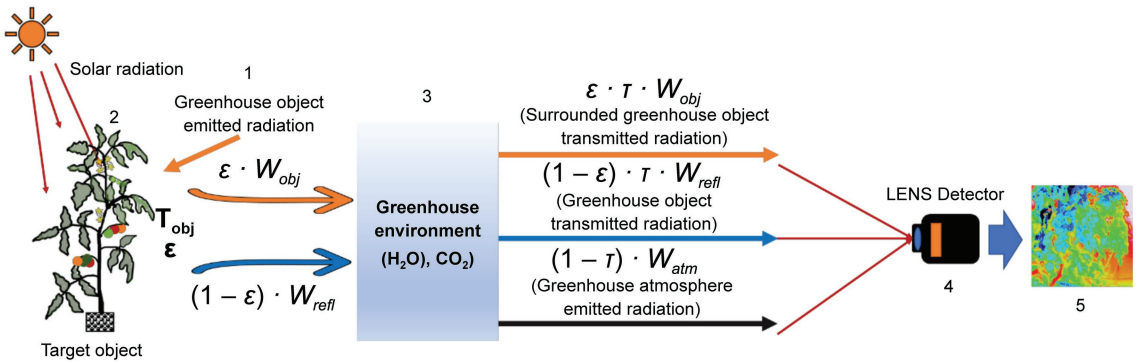

FIGURE 1 | A schematic representation of a thermal camera working principle. 1: surroundings, 2: object, 3: atmosphere, 4: thermal camera, and 5: thermal image.

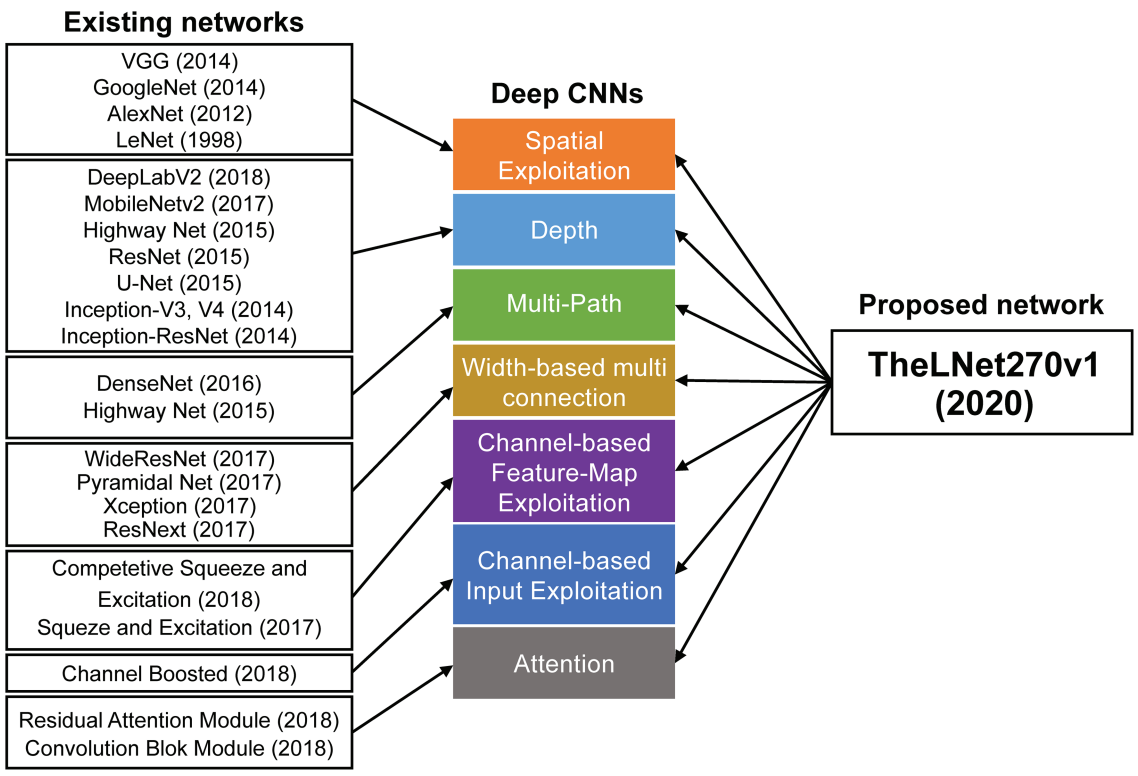

FIGURE 2 | Classification of available and proposed deep network architecture.

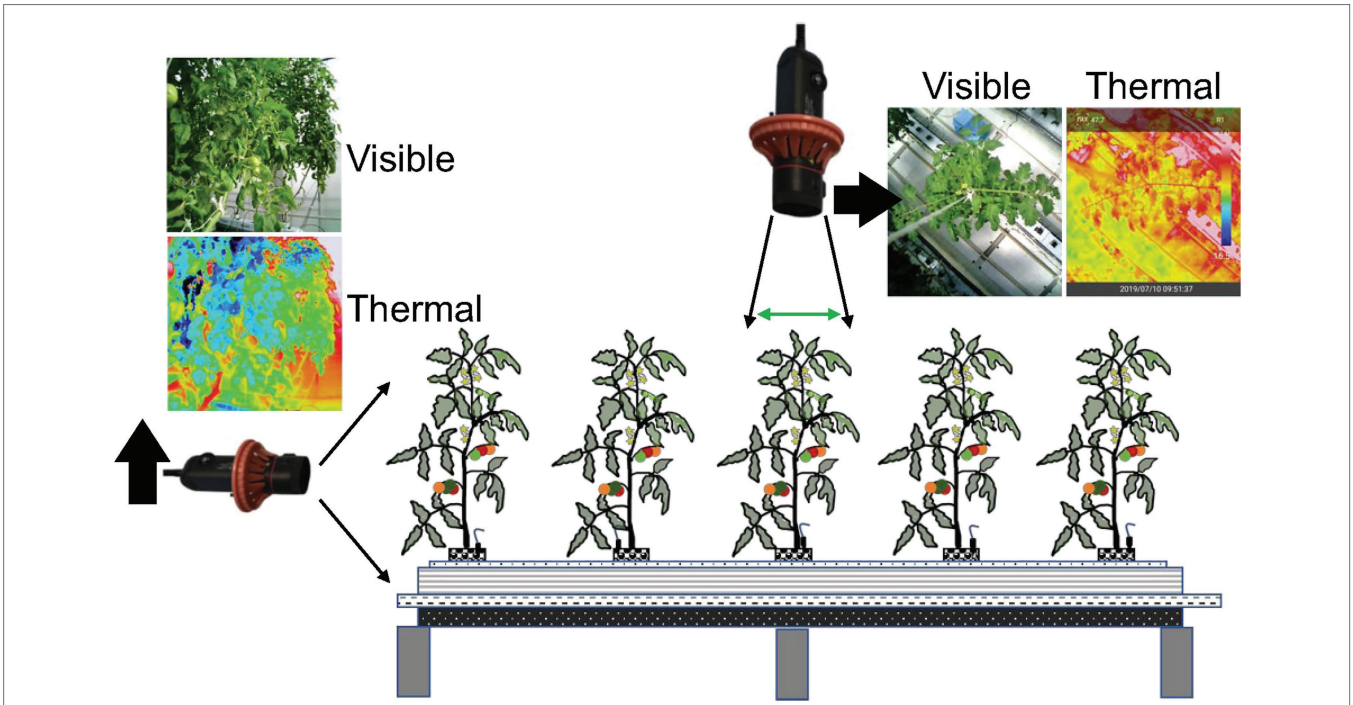

FIGURE 3 | Thermal image acquisition technique.

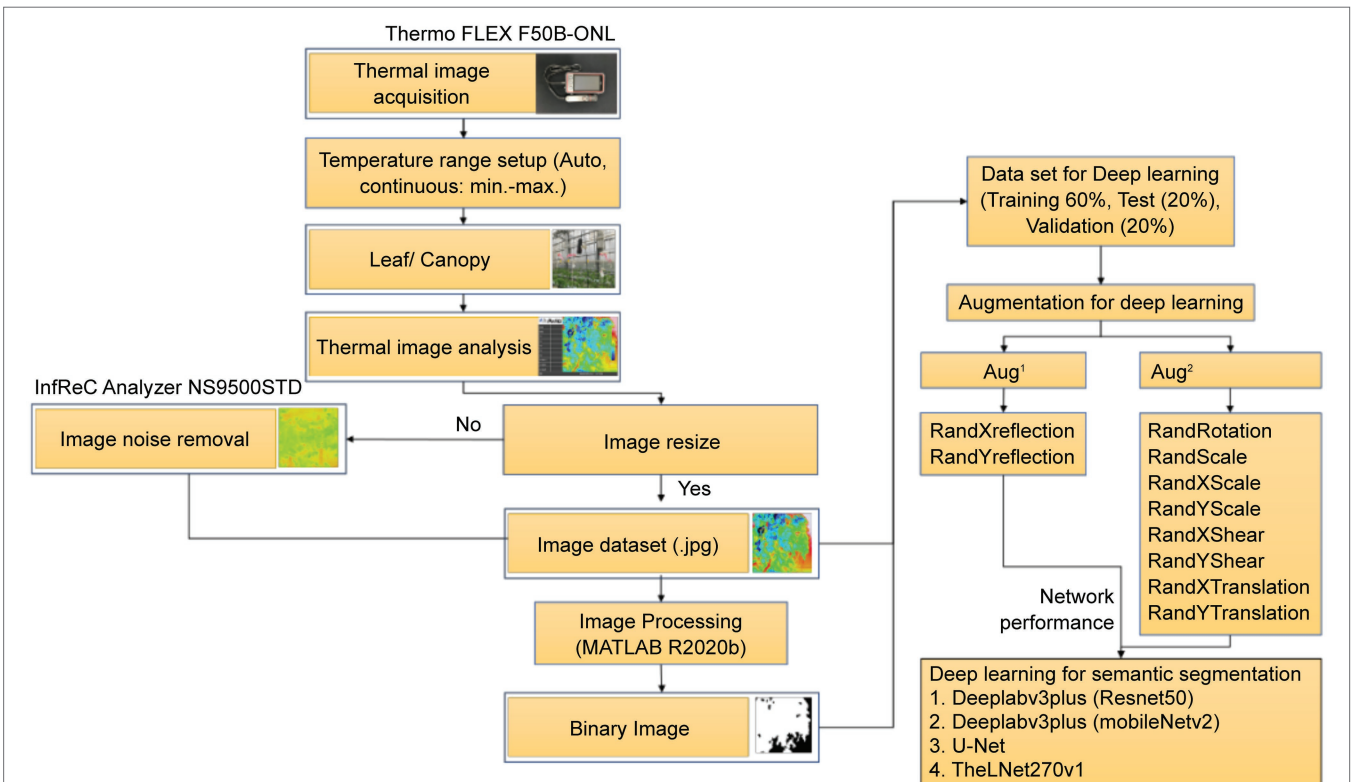

FIGURE 4 | Schematic diagram of the image dataset preparation.

scales. Srivastava et al. (2015), Dong et al. (2016), Larsson et al. (2016), Mao et al. (2016), Dauphin et al. (2017), Huang et al. (2017), Tong et al. (2017), and Kuen et al. (2018) proposed multi-paths or shortcut connections that connect one layer with another layer by skipping some intermediate layers. This allows overpassing some information to another layer and reduces the vanishing gradient problem, which causes a higher training error.

Li et al. (2019) proposed an edge-conditioned convolution neural network for thermal image segmentation with SODA (segment objects in day and night) benchmark for evaluating the thermal image segmentation performance. They used

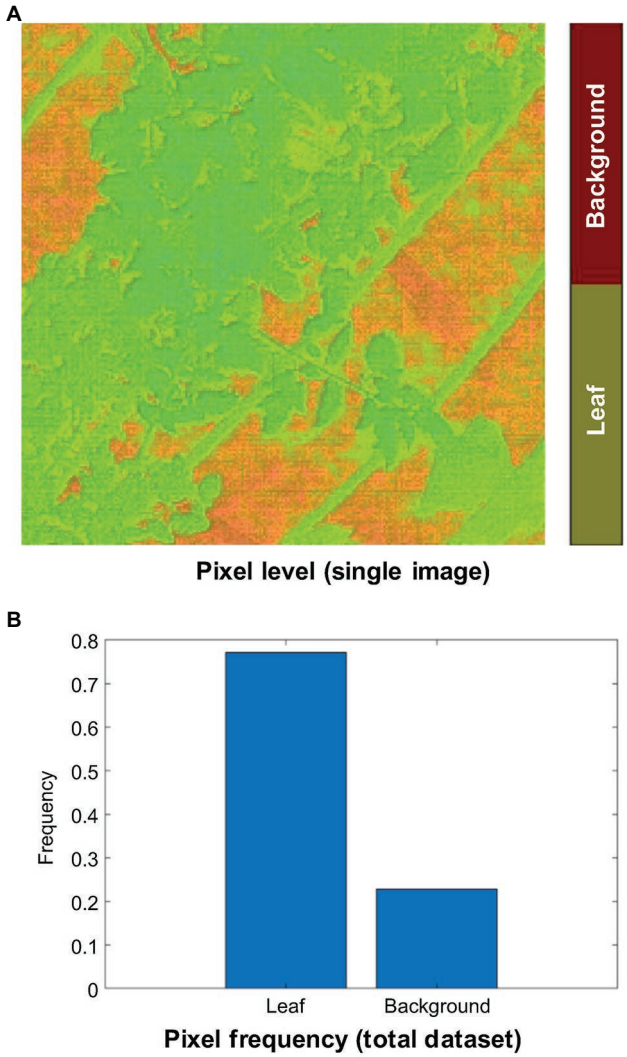

FIGURE 5 | Pixel classification of the thermal image dataset. **(A)** Manual annotation. **(B)** Frequency level of the annotated leaf and background pixels within total thermal image dataset.

TABLE 1 | Technical specification of the thermal camera (Thermo FLEX F50B-ONL).

Field of view, °	Focus	Spectral range, μm	Frame rate, Hz	Sensitivity, °C	Accuracy
70°×70°	Focus free	8~14	7.5	0.05°C at 30°C	±2°C or ±2% for 0–40°C (other conditions: ±4°C or ±4%)

TABLE 2 | The number of image datasets used for the DL analysis.

Condition	Original dataset	Binary dataset	Training	Validation	Test
Total image number	13,766	13,766	8,260	2,753	2,753

TABLE 3 | Properties of the augmented datasets for DL analysis.

Augmentation options		Training option	Visualization
RandXreflection	1	Aug[1]	**Figure 6A**
RandYreflection	1		
RandRotation	[–90 90]	Aug[2]	**Figure 6B**
RandScale	[1 1]		
RandXScale	[0.8 1.2]		**Figure 6C**
RandYScale	[0.8 1.2]		
RandXShear	[–20 20]		**Figure 6D**
RandYShear	[–20 20]		
RandXTranslation	[–10 20]		**Figure 6E**
RandYTranslation	[–10 20]		

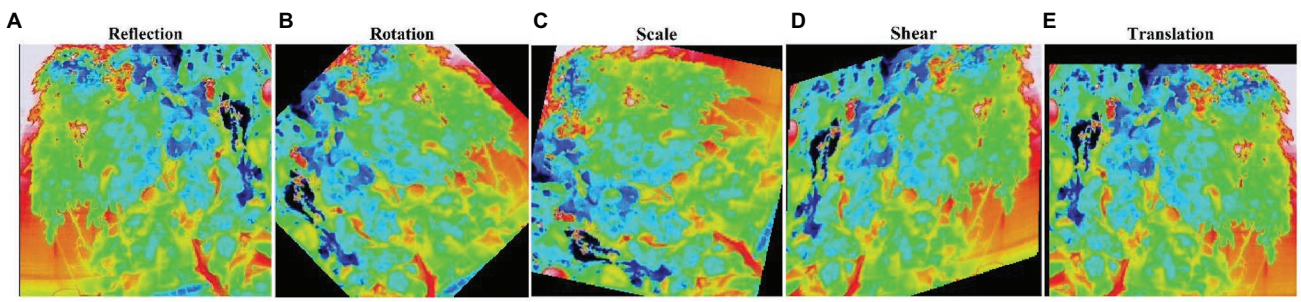

FIGURE 6 | Augmented image dataset. The statistics of the image datasets used for the deep learning (DL) network analysis are shown in **Table 2**. **(A)** Reflection; **(B)** Rotation; **(C)** Scale; **(D)** Shear; and **(E)** Translation.

manually annotated synthetically generated thermal images for training the network, which achieved 61.9% mean intersection over union (IoU), a lightly better than network trained with DeepLabv3 algorithm. Choi et al. (2016) developed a CNN-based thermal Image Enhancement technique for improving low-resolution thermal camera recognition tasks. The lightweight structure of the shallow convolutional neural network requires less CPU memory. In their architecture, they cropped a low-resolution thermal image with a uniform stride and used a bi-cubic interpolation method to upscale it. Chen et al. (2019) revealed a Fletcher-Reeves algorithm-based CNN model for hyperspectral image classification with 80.7% accuracy which outperforms other traditional CNN due to the advances in batch computing adaptability and convergence speed. Grbovic et al. (2019) reported a thermal-RGB image-based wheat-ears detection system for automatic counting wheat wars under outdoor conditions. They applied blocks of convolutional layers, each with an activated function for counting the wheat ears, which achieved 75.63 and 68.46% F1 Score for segmenting thermal and RGB images. Furthermore, achieved 89.22% accuracy for counting the wheat ears. Another study reported by Zhang et al. (2019) used a group of neurons termed as a capsule or vector for replacing traditional neurons and achieved equivariance by successfully encoding spatial information and properties of an input image. Bhattarai and Martínez-Ramón (2020) identified and classified objects in real time from thermal cameras carried by firefighters. The detection accuracy reported by authors varied from 70 to 95%, which depends on the depth of the convolution network layer.

Fuentes-Pacheco et al. (2019) developed a CNN with an encoder–decoder function which used top-view RGB images of fig plants and achieved a mean 93.85% segmentation accuracy under variable visual fig leaves the appearance and complex background. There are various DL architectures, such as LeNet, AlexNet, VGG, GoogleNet, YOLOv, Inception, and SqueezeNet, which are widely used for image classification and object detection. However, ResNet, U-Net, DeepLabv3, and MobileNet are mostly used for semantic segmentation (pixel) – based image (RGB) classification (Boulent et al., 2019; Saleem et al., 2019). In agriculture, the high or low thermal dynamic changes during sunny–cloudy–rainy days and nights make it difficult to spatially process bulk thermal images, such as separation of leaf/canopy pixels from background pixels (Cho et al., 2017; Salgadoe et al., 2019). To solve this classification challenge, the author proposed a new DL architecture with several components [convolutions, grouped convolution, transposed convolution, batch normalization, rectified linear unit (ReLU), max pooling, depth concatenation, element-wise addition, 2D crop, softmax, and classification output layer]. The aim of this study was to develop a DL architecture and demonstrate the learning ability of the DL architecture to separate the leaf/leaf canopy from a greenhouse background (ground, windows, roof, etc.) in thermal images under various environmental conditions (sunny, cloudy, and rainy: day or/and night).

MATERIALS AND METHODS

Thermal Image Acquisition System

The study was conducted in the greenhouse of the Vegetable and Flower Research Division, National Agriculture and Food Research Organization (NARO) in Tsukuba, Ibaraki, Japan. The Japanese cultivar "CF Momotaro York" (Takii Seeds Co., Ltd., Kyoto, Japan) of tomato (*Solanum lycopersicum*) grown in a Rockwool system was used for this experiment. The image data collection period ran from October 16,

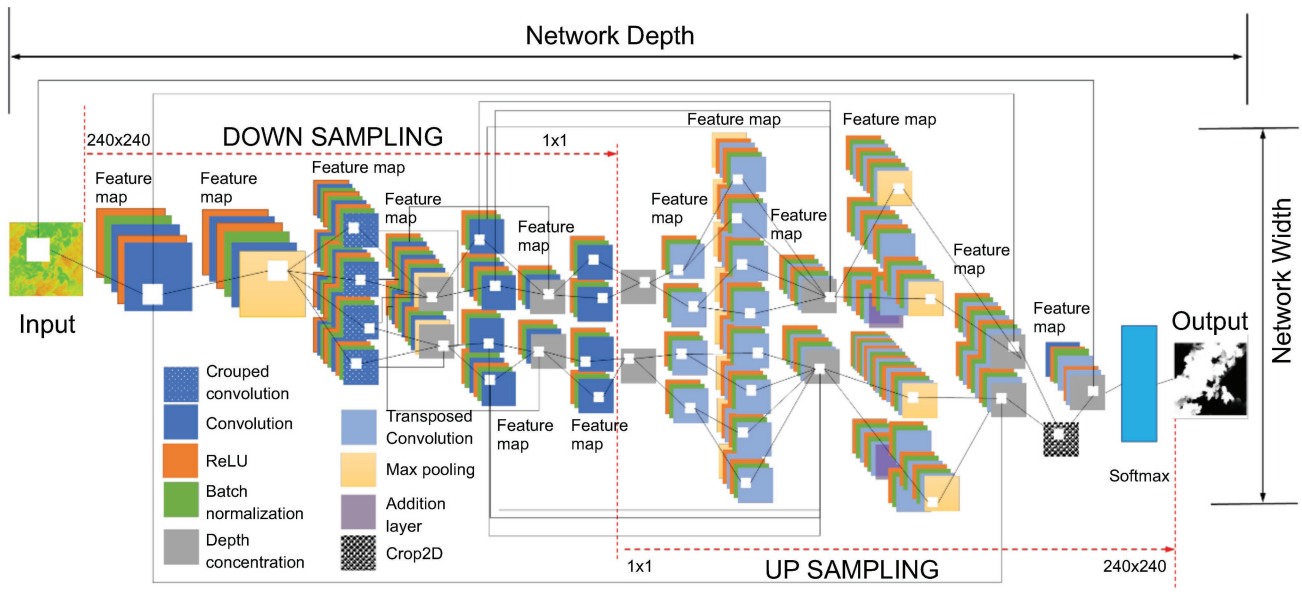

FIGURE 7 | The basic network architecture of the TheLNet270v1.

2019 to September 30, 2020. The air temperature and relative humidity at 1.2 m above the ground surface ranged between 8.6 and 37.5°C, 32 and 96% from October 16, 2019 to April 16, 2020. The air temperature and relative humidity at 1.2 m above the ground surface ranged between 9.6 and 39.3°C, 34 and 95% from August 7, 2020 to October 28, 2020. Thermal images with 1040 × 780 pixel resolution (screen) were obtained, as shown in **Figure 3**, using a compact long-wave thermal camera [Thermo FLEX F50B-ONL (Nippon Avionics Co., Ltd., Yokohama, Japan)] under various environmental conditions at a minimum distance of 0.3 m from the top and maximum 2 m from the side of the targeted tomato plant.

All images were stored in a 24-bit thermal image format. The emissivity range of the thermal camera is 0.1 to 1. In this experiment, the emissivity of the tomato leaf was considered to be 0.98 (López et al., 2012). The technical specifications of the thermal camera are listed in **Table 1**.

TABLE 4 | Characteristics of the TheLNet270v1 architecture.

Layers name	Total number of layers
Image input	1
Convolution	31
ReLU	83
Batch normalization	71
Max pooling	15
Transposed convolution	49
Addition layer	2
Grouped convolution	4
Depth concatenation	11
Crop2D	1
Softmax	1
Pixel-classification (output)	1

TABLE 5 | Hyperparameter optimization parameter.

Training options: adam	Execution environment: parallel
Learn rate schedule: piecewise	Validation patience: Inf
Shuffle: every-epoch	Epsilon:1e-8
Verbose: false	Initial learn rate: 1.0000e-03
Validation frequency:10	Learn rate drop factor: 0.1000
Gradient decay factor: 0.9000	Learn rate drop period: 10
Squared gradient decay factor: 0.9990	Gradient threshold method: l2norm
L2 regularization: 1.0000e-04	Verbose frequency: 50
Gradient threshold: Inf	Dispatch in background: 0
Sequence padding value: 0	Reset input normalization: 1
Sequence length: longest	Sequence padding direction: right

Image Dataset Preparation

Figure 4 demonstrates the schematic diagram of the image dataset preparation for network analysis.

In total, 13,766 thermal images were obtained during this experiment. The thermal images were resized into their original spatial resolution (240 × 240 pixels), and denoising (manipulation of scale and emissivity) was performed by a thermal imaging processing software (InfReC Analyzer NS9500STD for F50, Nippon Avionics Co., Ltd.) to meet the network input dimension (240 × 240 pixels) requirements. Furthermore, Image Segmenter (Image Processing and Computer Vision Toolbox, MATLAB R2020a) was used to convert the pixels of each thermal image into two groups manually: leaf (255) and background (0) as shown in **Figure 5A**. These pixel values were stored in binary images. The frequency levels of the leaf and background pixels within the total thermal image datasets were 77 and 23%, respectively (**Figure 5B**). In this experiment, 60% of the randomly selected images (thermal images and binary images) were used for training, 20% for validation, and 20% for test purposes.

The image dataset was augmented to increase the amount and type of variation within the training image data to prevent overfitting and generalizing the model performance (**Figures 6A–E**). **Table 2** shows the number of image datasets used for deep learning analysis. First, we augmented the image data, including random reflection in the X and Y directions [(aug^1)]. This dataset was used for the network performance study. Furthermore, for comparative analysis, we also augmented the thermal image dataset with the other four options (aug^2), as shown in **Table 3**.

Network Architecture

Figure 7 demonstrates the basic network architecture of the TheLNet270v1, which is a combination of the semantic segmentation-based network (convolution layers) and classification-based network (softmax). The convolution layer of the proposed network extracts the higher-level features from input images with multiple smaller filter sizes (3 × 3 × 3 × 32). The smaller filter size of the convolution layer has a strong generalization ability when the same types of objects within an image are conglutinated with each other (Zhang et al., 2020). This capability effectively improves network learning performance. According to Nair and Hinton (2010), the ReLUs activation function added non-linearities to the model, converted values less than zero to zero for each element of the input, transformed the summed weighted input from the node into output, and allowed models to learn faster with higher accuracy. The batch normalization layer increases the network stability and normalizes the output of a previous activation layer by subtracting the

TABLE 6 | The image datasets for performance study and comparative analysis.

Image data	Original image	Binary image	Total image	Comments
Original dataset	13,766 × 1	13,766 × 1	27,532	-
Augmented dataset[1]	13,766 × 2	13,766 × 2	55,064	Original condition + XYReflection
Augmented dataset[2]	13,766 × 6	13,766 × 6	82,596	Original condition + XYReflection + Rotation + Scale + XYScale + XYShear + XY Translations

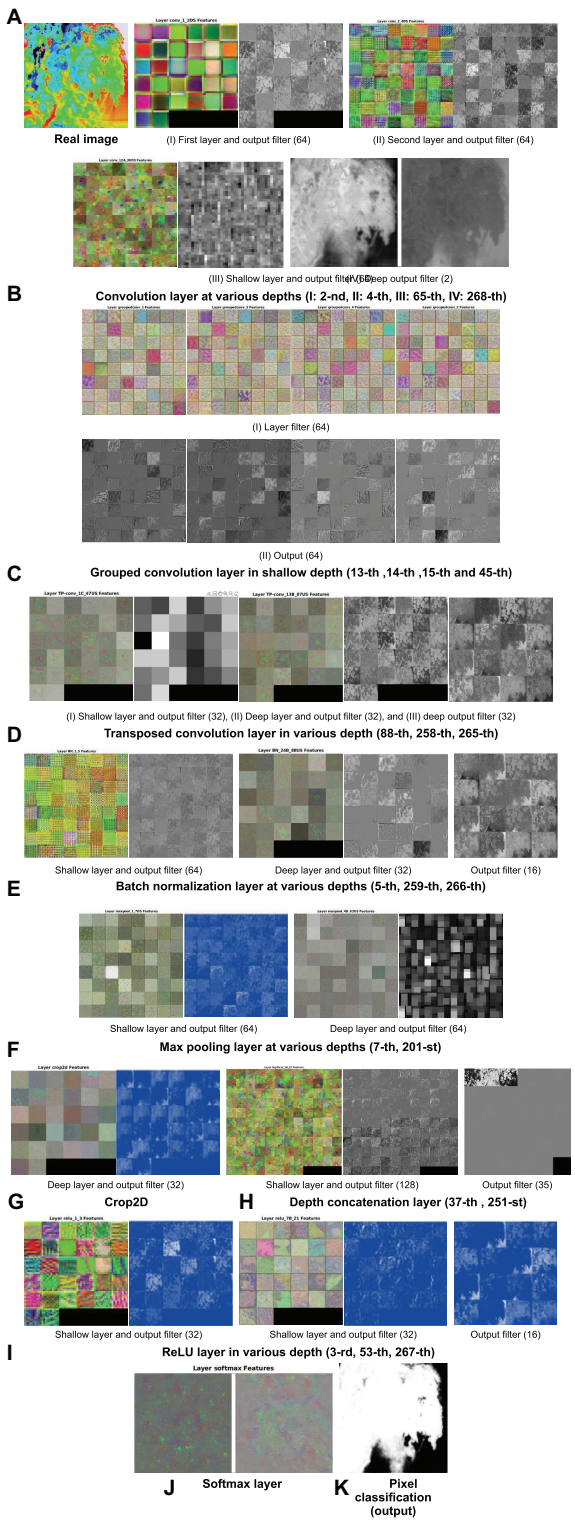

FIGURE 8 | Feature maps and visualization of the network. **(A)** Visualization of the first and second layer. **(B)** Convolution layer at various depths. **(C)** Grouped convolution layer at various depths. **(D)** Transposed convolution layer at various depths. **(E)** Batch normalization layer at various depths. **(F)** Max pooling layer at various depths. **(G)** Visualization of the crop2D. **(H)** Visualization of the depth concatenation layer. **(I)** ReLU layer at various depths. **(J)** Visualization of the softmax layer. **(K)** Visualization of the output.

batch mean and dividing by the batch SD (Ioffe and Szegedy, 2015). Krizhevsky et al. (2012) introduced grouped convolution for training AlexNet with less powerful GPUs with limited RAM. It is also termed as convolutions in parallel as this layer separates input channels into groups by applying sliding convolution filters (vertically and horizontally), computing the input and weights, adding a bias, and finally combining the convolutions for each group independently (Xavier and Bengio, 2010; He et al., 2016a). We included grouped convolution to increase the width of the network without hampering computational power. According to Scherer et al. (2010) and Zhang et al. (2020), the max-pooling layer simplifies the network complexity by compressing and extracting the main features, ensuring feature position and rotation invariance, and rotation reduced computing time. A 2D image cropping layer crops images at the center to explore contextual features (Blaschke, 2010). The last convolution layer has two outputs corresponding to two classes with a ReLU activation followed by a batch normalization layer with 16 filters. The output of the last convolution layer is fed into the softmax layer for calculating the probability of the output classification layer. Finally, these expanded features are passed to the classification layer for classification (Krizhevsky et al., 2012). Therefore, the depth of the DL architecture is fixed to 270 layers and accurately optimized based on training performance. The characteristics of the TheLNet270v1 architecture are shown in **Table 4**.

Network Parameters

The TheLNet270v1 was trained on a FUJITSU SHIHO Supercomputer equipped with TESLA V100-SXM2 32GB and CUDA version 10.2, DL, and parallel computing toolbox (MATLAB R2020a). The adaptive moment estimation (ADAM) algorithm was used to optimize the network weights. The transfer learning parameters applied for training the TheLNet270v1 were as follows: training option: Adam; validation frequency: 10; mini-batch size: 50/70/90/128/156/220/240/260/290/320; max epoch: 5/12/20/30/40; learn rate schedule: piecewise; shuffle: every-epoch; initial learn rate: 0.001; epsilon: 1e-08. ADAM was used to optimize the network weights. **Table 5** shows the hyperparameter optimization parameter for the TheLNet270v1 training.

Comparative Analysis and Evaluation Metrics

Currently, MobileNetv2 is widely used in low-powered mobile devices for image recognition or classification tasks because of its simple network architecture and lower computational complexity (Wong et al., 2020). He et al. (2016a) first introduced ResNet with cross-layer connectivity in a CNN, which sped up the convergence of deep neural networks, solved the vanishing gradient problem by actively deploying special skip connections and a batch normalization layer and 20 and 8 times deeper than AlexNet and VGG. On the other hand, U-Net is mostly used in high-powered fixed devices because of its complex network architecture. It is widely used for biomedical image segmentation and classification purposes (Ronneberger et al., 2015).

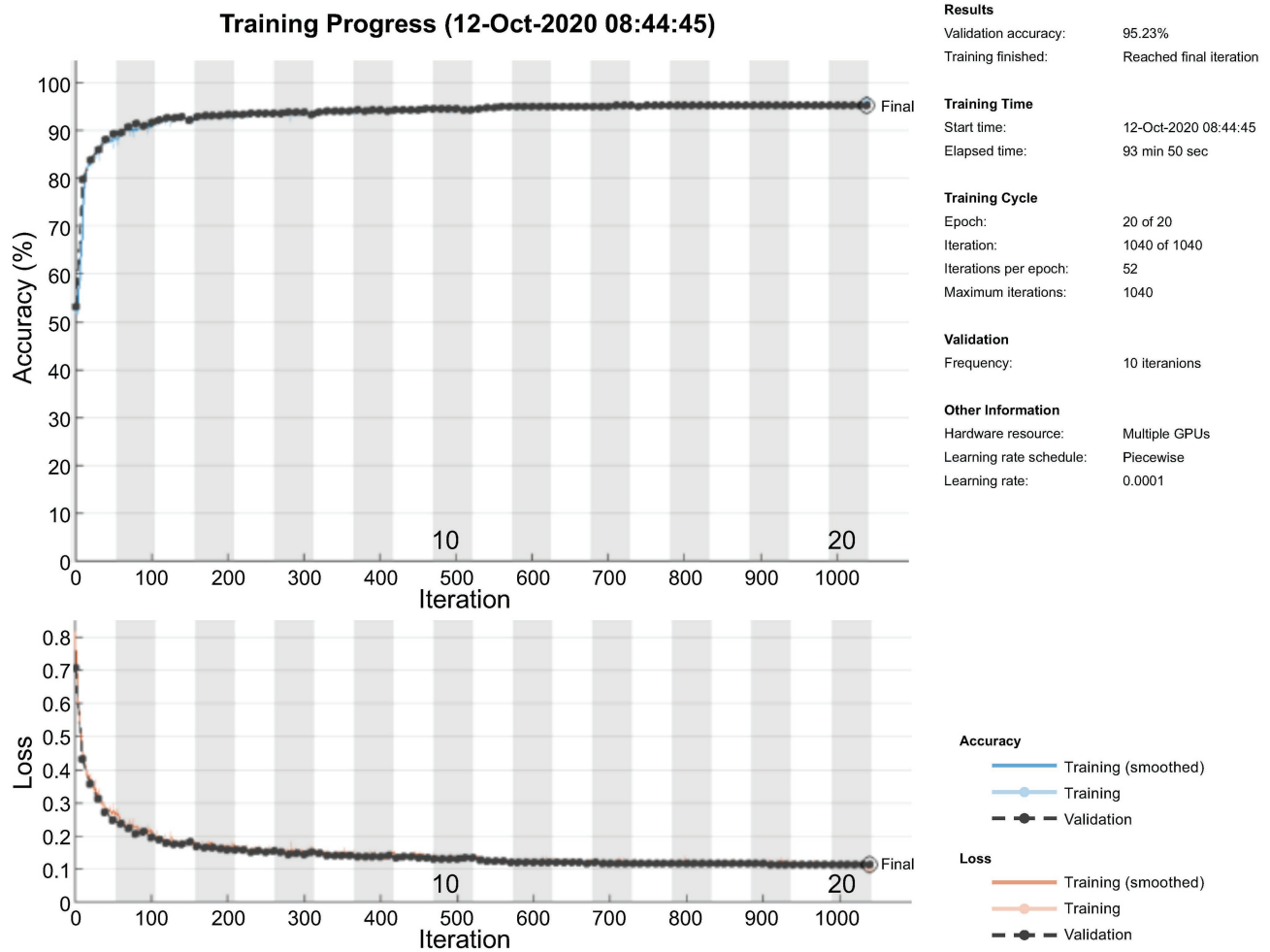

FIGURE 9 | Monitoring of overfitting.

The bottleneck layer between the contracting and expanding paths of the U-Net architecture increased the network depth and was regularized by dropout to solve the overfitting issue during the network learning process (Krizhevsky et al., 2012; He et al., 2016b). Giusti et al. (2013) stated that Deeplabv3plus employs atrous convolution or dilated convolutions in parallel or in cascade to extract dense features at multiple scales with better-stored information capability. TheLNet270v1 is designed so that it can be used in both low-powered mobile or high-powered fixed devices. There are several performance metrics such as training/validation/test accuracy (shows the percentage of correctly classified pixels), global accuracy (measuring ratio of correctly classified pixels to the total number of pixels), mean accuracy (measuring the percentage of correctly identified pixels for each class), confusion metrics, validation loss, training time, IoU/Jaccard index (measuring the amount of overlap per predicted class), weighted IoU (measuring the average IoU of each class), BF score (Boundary F1 – measuring the quality of the predicted boundary with the ground truth boundary), etc. are used for quantifying TheLNet270v1 accuracy and network efficiency. The same performance metrics were also evaluated on Deeplabv3plus (with a pretrained network MobileNetv2 and ResNet-50) and U-Net for comparative analysis.

RESULTS AND DISCUSSION

Image datasets are augmented into two categories for network training. The augmented dataset[1] and augmented dataset[2], as shown in **Table 6**, are both used for performance study and comparative analysis.

Feature Extraction and Activation for Visualization

Features extracted and visualized from the different depths of the TheLNet270v1 layers after completing the training are shown in **Figure 8**. Typical looking filters starting from the first layer in **Figure 8B(I)** show the colorful smooth pixels of each of the 64 filters, to noisy pixels in **Figure 8B(II)**,

and then slightly visible some features in **Figure 8B(III)**. The last convolution layer in **Figure 8B(IV)** finally represents the visible pixel class. In **Figures 8C(I,II)**, identical features of the grouped convolution layer in shallow depth are shown at different positions of an image. **Figures 8D–I** reveal different structures of the feature maps within each filter and layer, and visualizations show that the feature map is activated on the foreground tomato leaf image, not the background objects. Finally, softmax (**Figure 8J**) gives a discrete probability for each class (leaf/leaf canopy and background), which is between 0 and 1, and the result is visualized in the pixel classification output layer (**Figure 8K**), where 1 (white color) means leaf/leaf canopy and 0 means background (black color).

Performance Metrics

Figure 9 shows the accuracy and loss of the training and validation datasets used to monitor the network overfitting

TABLE 7 | The performance metrics for image datasets.

Accuracy	Global accuracy, %	Mean accuracy, %	Mean IoU, %	Weighted IoU, %	Mean BFScore, %
Train metrics	94.85	91.99	86.50	90.33	86.42
Validation metrics	94.82	91.69	86.43	90.28	86.34
Test metrics	94.81	91.75	86.20	90.26	86.36

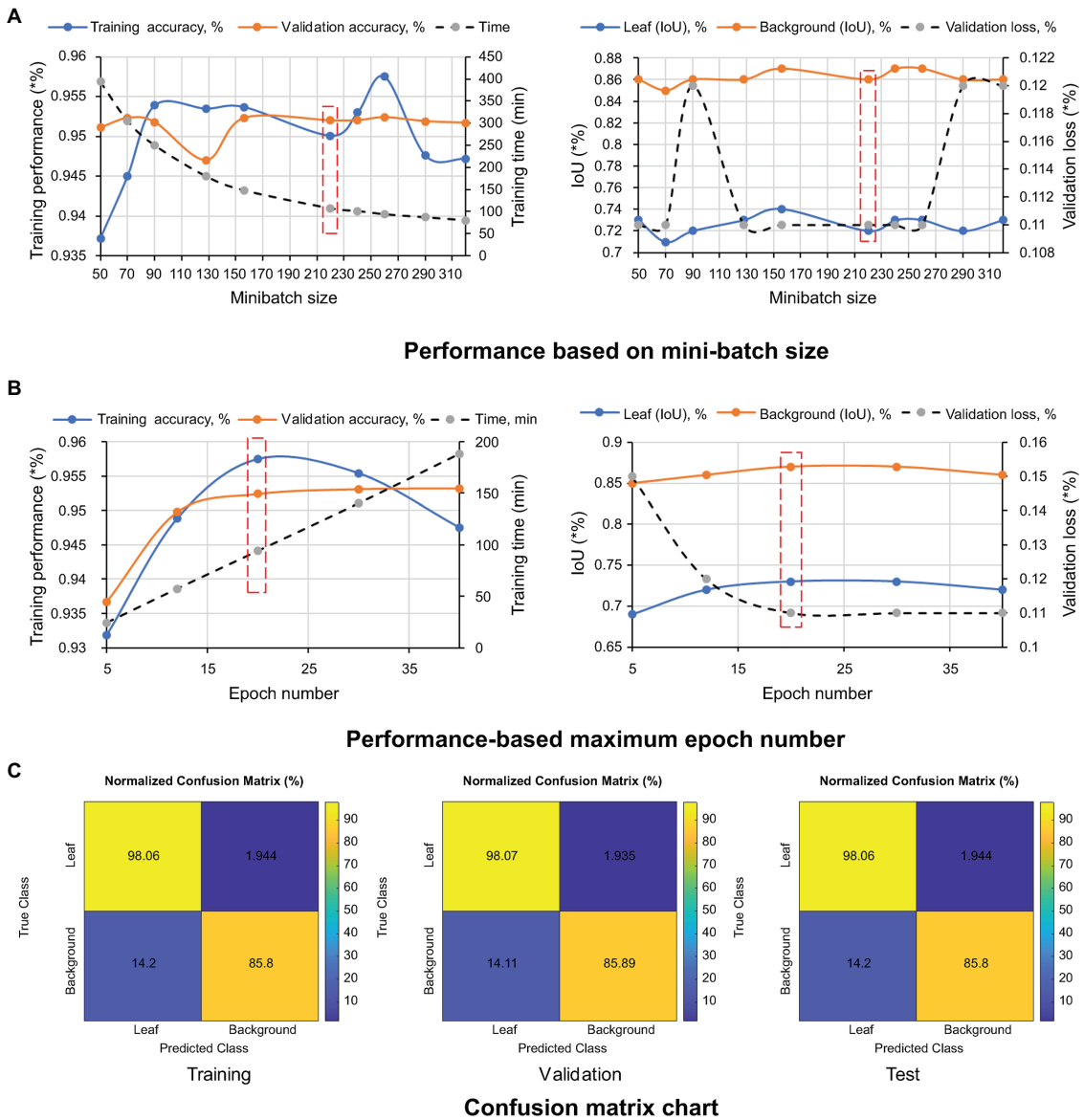

FIGURE 10 | Performance metrics of TheLNet270v1. **(A)** Performance based on mini-batch size. **(B)** Performance-based maximum epoch number. **(C)** Confusion matrix chart.

TABLE 8 | The intersection over union (IoU) and BFScore for each class.

Accuracy	Train metrics			Validation metrics			Test metrics		
	Accuracy	IoU	Mean BFScore	Accuracy	IoU	Mean BFScore	Accuracy	IoU	Mean BFScore
Leaf	97.27	93.57	90.99	97.24	93.54	91.03	97.29	93.56	91.10
Background	86.72	79.42	81.78	86.70	79.33	81.60	86.21	78.84	81.56

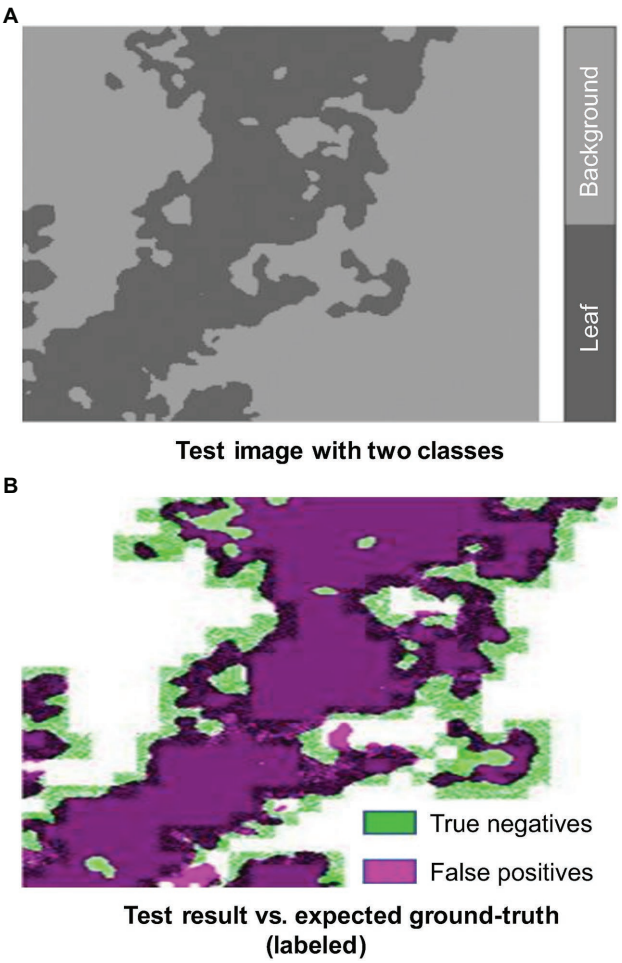

FIGURE 11 | Test image vs. expected ground truth. **(A)** Test image with two classes. **(B)** Test results vs. expected ground-truth (labeled).

issue. It is clearly visible that the model performs well on both training and validation data sets.

The pixel-level classification of thermal images by TheLNet270v1 was investigated. A validation accuracy of 95.22% was achieved with a minibatch size of 320, max epoch of 20, and training time of 94.15 min, shown in **Figures 10A,B**. Under the same conditions, the maximum IoU of 74 and 87% for leaf and background was achieved. During this time, a minimum validation loss of 12% was observed. The confusion matrix is given in terms of percentage and absolute number. It can be seen from the confusion chart in **Figure 10C** that the higher classification accuracies of 98.07, 98.06, and 98.07% for leaf and 85.89, 85.80, and 85.51% for the background achieved with the training, validation, and testing datasets (**Table 6**) and demonstrated that the network was well-trained.

Table 7 shows the test results of several other performance metrics such as global accuracy, mean accuracy, weighted IoU, and BF score. A higher value indicates better network performance.

The classification accuracy of each class (leaf and background) is described in **Table 8**.

Figure 11A shows an example of a test image successfully segmented into two classes, in which the dark color area represents leaf and light color background. **Figure 11B** shows a tiny presence of false positives (magenta color). However, the boundary between leaf and background is marked as green color (true negatives), which described that further refinement is possible if we retrain the network with more image data or images with higher resolutions.

Comparative Metrics

It is evident from **Table 9** that the TheLNet270v1 has a maximum depth layer of 270 with a lower total number of network parameters of 2e + 11, which is lower than Deeplabv3plus (ResNet50) and Deeplabv3plus (MobileNetv2). However, U-Net has a minimum of 46 layers with a higher total number of network parameters of 6e + 06 than TheLNet270v1. However, the training time for all networks (20 epoch, 220 minibatch sizes, and augmented dataset[1]) slightly differed.

Figure 12, Δ Performance (Eq. 1) demonstrated each evaluation metric's positive and negative values with different image datasets. A negative value indicates an increase in the network performance, while a positive value is decreasing. The longer red arrow in the image indicates the volatile nature of the network due to the increase in the image dataset. From this, it is clear that Deeplabv3 (MobileNetv2) and TheLNet270v1 both show stable network performance despite increasing the number of images in the augmented dataset, as described in **Table 6**.

$$\Delta\, Performance, \% = \left(\begin{pmatrix} evaluation\ metrics\ for \\ augmented\ data^1 - \\ evaluation\ metrics \\ for\ augmented\ data^2 \end{pmatrix} / 100 \right) \times 100 \quad (1)$$

Test results vs. expected ground-truth (labeled) on the image-basis test dataset with IoU histogram are shown in

TABLE 9 | Comparative statistics of the various network architectures.

Network name	Total layers	Total neurons	Total weights	Total biases	Total parameters	Training time, min
Deeplabv3plus (ResNet50)	206	4.00E + 07	6.00E + 06	4.00E + 04	4.00E + 12	80.51
Deeplabv3plus (MobileNetv2)	186	4.00E + 07	7.00E + 06	4.00E + 04	4.00E + 12	94.29
U-Net	46	8.00E + 07	7.00E + 06	3.00E + 03	6.00E + 06	94.16
TheLNet270v1	270	5.97E + 07	1.63E + 06	3.00E + 03	2.00E + 11	94.15

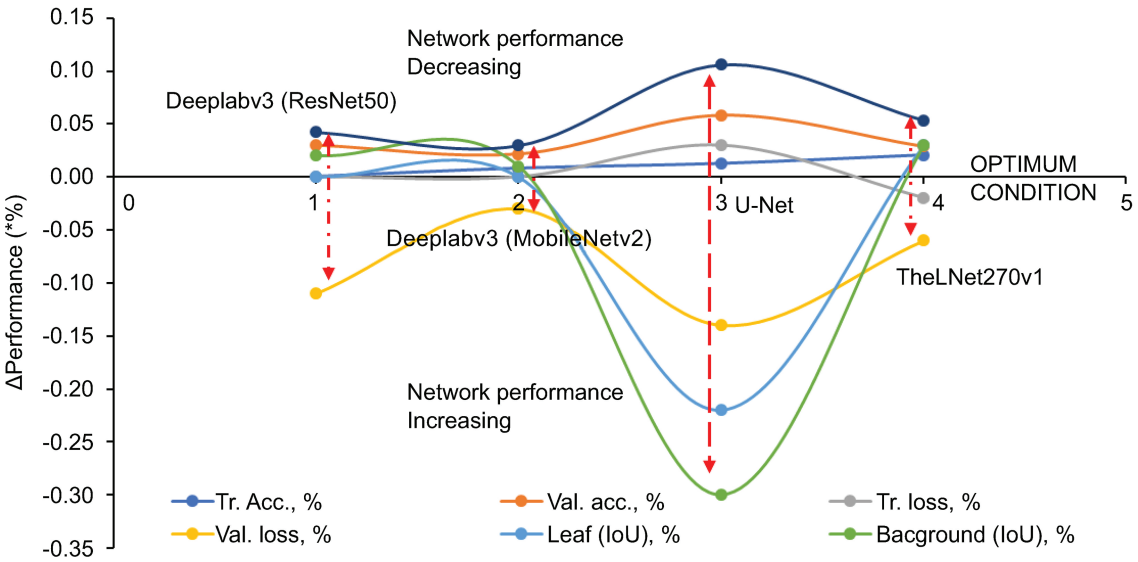

FIGURE 12 | Δ Performance of the evaluation metrics.

TABLE 10 | The mean IoU for comparative analysis.

Network name	Augmented data[1]	Augmented data[2]	Augmented data[1]	Augmented data[2]	Visualization
	Leaf		Background		
Deeplabv3plus (ResNet50)	0.74	0.74	0.87	0.85	**Figure 13A**
Deeplabv3plus (MobileNetv2)	0.72	0.72	0.86	0.85	**Figure 13B**
U-Net	0.44	0.66	0.52	0.82	**Figure 13C**
TheLNet270v1	0.73	0.7	0.87	0.84	**Figure 13D**

Figures 13A–D [Deeplabv3plus (ResNet50), Deeplabv3plus (MobileNetv2), U-Net, and TheLNet270v1], and the mean IoU of each class, as described in **Table 10**. The mean IoU of the leaf and background classes is indicated by the top bar in the image histogram. **Figure 13** and **Table 10** show that the difference in mean IoU is clearly noticeable for the network trained with augmented dataset[1] and augmented dataset[2]. No noticeable changes are occurring for networks trained with different types of data sets. However, U-Net demonstrated IoU improvement with an increasing number of image datasets. The results revealed that leaves that counted the maximum number of pixels had lower IoU than the background with the least number of pixels. Further increasing the number of images within the same pattern or adding high-resolution images can improve the network performance (Zhang et al., 2018).

Prediction Results

The prediction results of the independent image datasets are shown in **Figure 14**. **Figure 14A** represents the early morning with a sunny condition, **Figure 14B** represents the midday with a sunny–cloudy condition, and **Figure 14C** represents the midnight condition. These three sets of images were captured during September 2020 and were used to verify the network prediction efficiency. It is visualized that the TheLNet270v1 has better prediction ability compared with other networks.

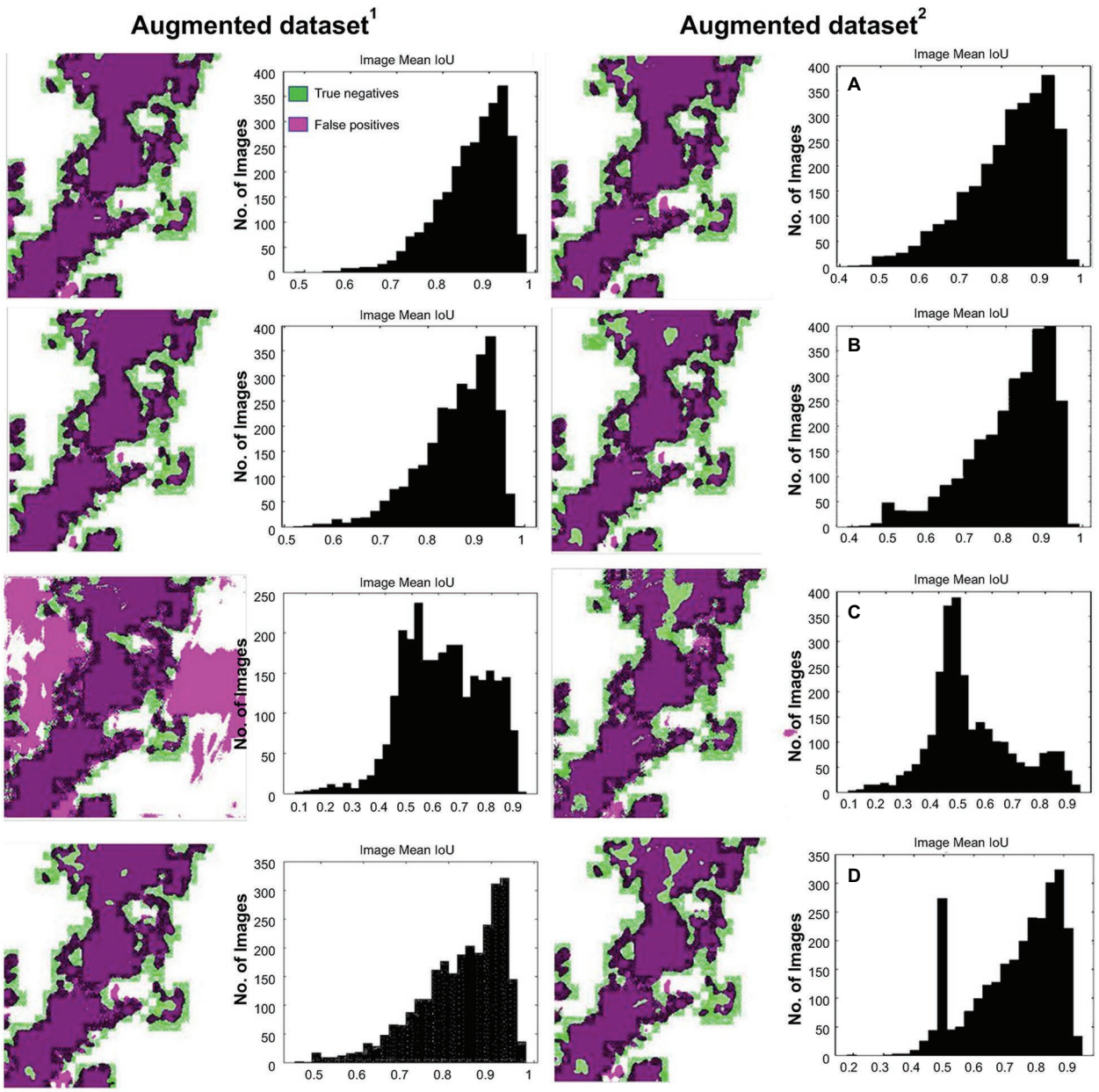

FIGURE 13 | Test result vs. expected ground-truth (labeled) with intersection over union (IoU) histogram. **(A)** Deeplabv3plus (ResNet50). **(B)** Deeplabv3plus (MobileNetv2). **(C)** U-Net. **(D)** TheLNet270v1.

Network Performance Verification With the IPPN Plant Phenotyping Image Dataset

As we described earlier, thermal images of the greenhouse-grown tomato plants were used for training the TheLNet270v1. This network successfully classified leaf/canopy and its background with higher accuracy, as shown in **Figures 11, 14**. We further investigated TheLNet270v1 performance using the IPPN plant phenotyping image dataset (leaf segmentation challenge component of the CVPPP workshop: CVPPP2017LSC-2017 and CVPPP2017LCC-2017; Minervini et al., 2015), which included pot-cultivated *Arabidopsis thaliana*. First, we predicted the TheLNet270v1 output using image data from CVPPP2017LCC-2017 and CVPPP2017LSC-2017, as shown in **Figures 15A,C**. Subsequently, we trained the network with the CVPPP2017LSC-2017 image dataset (total images: 236, RGB) and then predicted again with the same image data from CVPPP2017LCC-2017 (**Figures 15B,D**). It is clearly visible that the TheLNet270v1 output, which is almost identical to the manually segmented binary image, is shown in **Figure 15**. **Table 11** shows the TheLNet270v1 performance metrics.

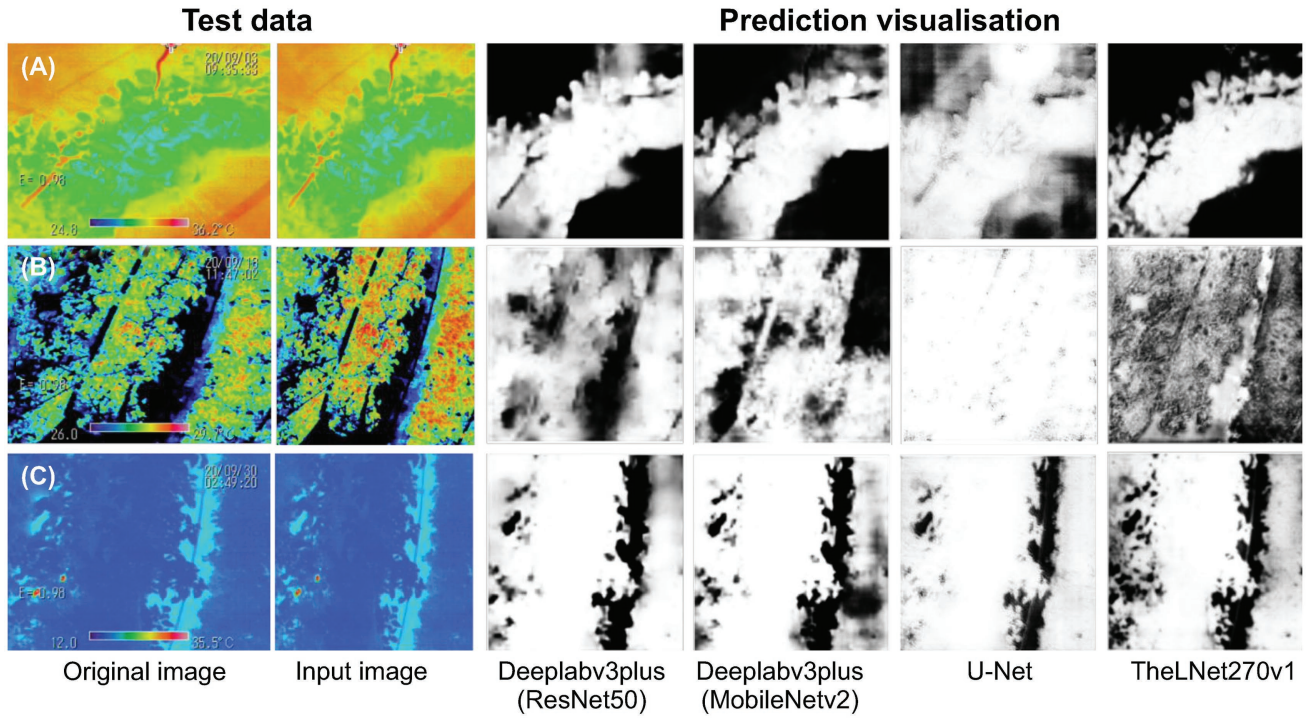

FIGURE 14 | Prediction visualization of the single independent test image. **(A)** The early morning with a sunny condition. **(B)** The midday with a sunny–cloudy condition. **(C)** The midnight condition.

TABLE 11 | The performance metrics for image datasets.

Accuracy	Mean IoU, %	Weighted IoU, %	Mean BFScore, %
Train metrics	49.86	99.72	99.59
Validation metrics	49.84	99.68	99.61
Test metrics	49.91	99.81	99.81

IoU, intersection over union.

CONCLUSION

This study introduced TheLNet270v1, a highly compact deep neural network (for mobile and non-mobile image classification) for classifying thermal images captured inside a greenhouse and demonstrating a higher classification accuracy. This paper also concludes a comparative analysis with other widely cited pre-trained networks for pixel-based classification, such as Deeplabv3plus (ResNet50), Deeplabv3plus (MobileNetv2), and U-Net, and found that TheLNet270v1 achieved a significantly better balance between accuracy and network efficiency. In our future work, we will apply the TheLNet270v1 network for on-site training, and output will be used for 24 h to monitor the relationships between plant growth and environmental conditions of the greenhouse. This network is suitable for the image with 240 × 240 pixels. However, to make it suitable for different pixel sizes, we consider modifying this network depending on the different image sizes in our future study.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MI performed the architecture development and analysis and wrote the first draft of the manuscript. MI, NK, and UL conducted the field experiment. KT and NK verified the experimental results. All authors contributed to manuscript revision and approved the submitted version.

FUNDING

This work was supported by NARO under the "Environment optimization control system in plant factory facility using AI technology" Program (no. C11).

ACKNOWLEDGMENTS

This research is the output of patented technology "A leaf temperature acquisition device, a crop growing system, a method for acquiring leaf temperature and a program for acquiring

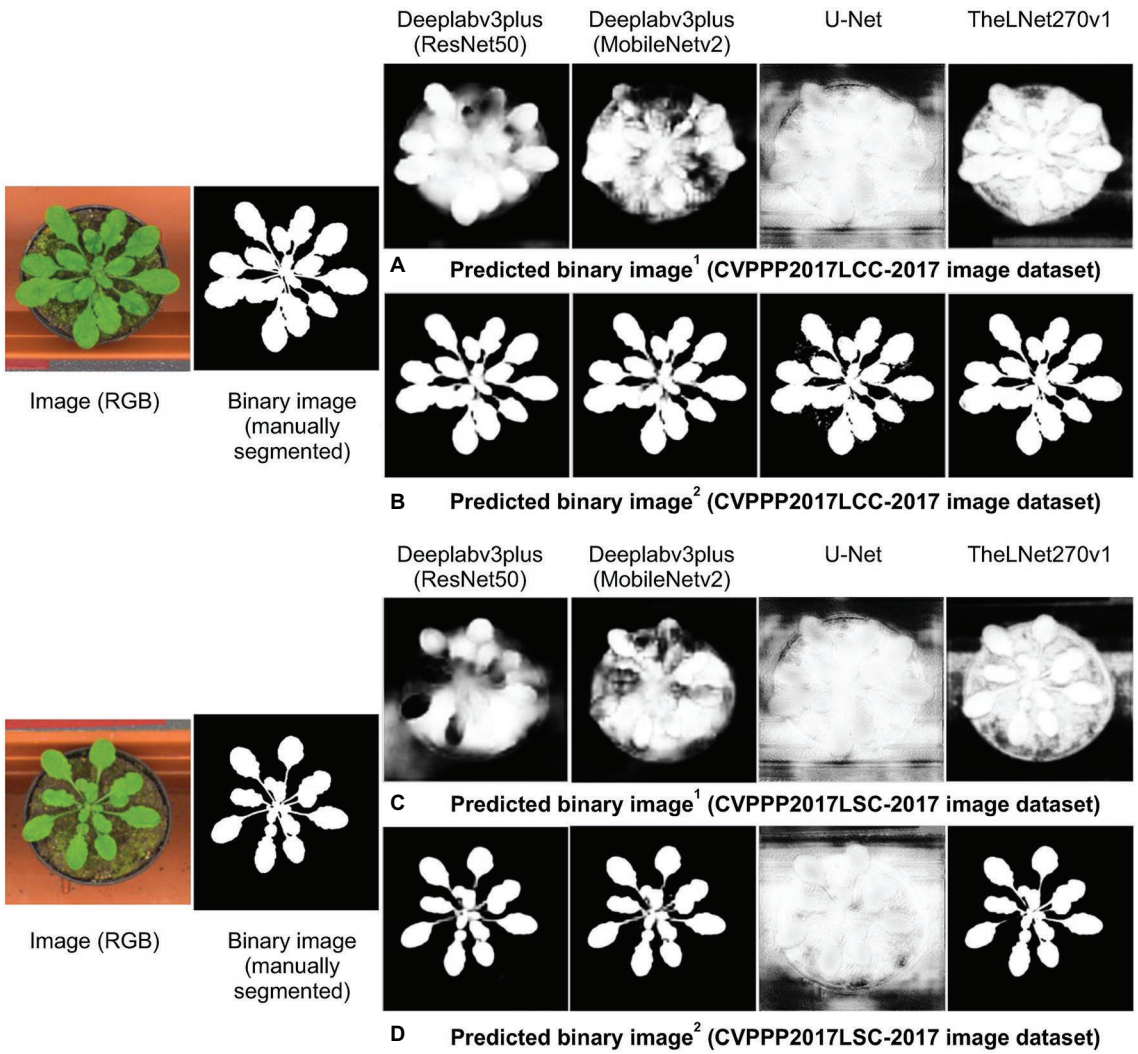

FIGURE 15 | Verification of the prediction capability. **(A)** Comparison of manually segmented data with the predicted image from augmented 1 CVPPP2017LCC-2017 data. **(B)** Comparison of manually segmented data with the predicted image from augmented 2 CVPPP2017LCC-2017 data. **(C)** Comparison of manually segmented data with the predicted image from augmented 1 CVPPP2017LSC-2017 data. **(D)** Comparison of manually segmented data with the predicted image from augmented 2 CVPPP2017LSC-2017 data.

leaf temperature," Japanese patent no. 2020-138804. The author gratefully acknowledges the technical support from Nippon Avionics Co., Ltd., Japan. We would like to acknowledge Tadahisa Higashide, NARO Institute of Vegetable and Flower Research.

REFERENCES

Bengio, Y. (2009). Learning deep architectures for AI. Found Trends®. Mach. Learn. Available at: https://www.iro.umontreal.ca/~lisa/pointeurs/TR1312.pdf (Accessed October 14, 2020).

Bhattarai, M., and MartíNez-Ramón, M. (2020). A deep learning framework for detection of targets in thermal images to improve firefighting. *IEEE Access.* 8, 88308–88321. doi: 10.1109/ACCESS.2020.2993767

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS J. Photogramm.* 65, 2–16. doi: 10.1016/j.isprsjprs.2009.06.004

Boulent, J., Foucher, S., Théau, J., and St-charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10:941. doi: 10.3389/fpls.2019.00941

Chaerle, L., and Van Der Straeten, D. (2000). Imaging techniques and the early detection of plant stress. *Trends Plant Sci.* 5, 495–501. doi: 10.1016/S1360-1385(00)01781-7

Chen, C., Ma, Y., and Ren, G. A. (2019). Convolutional neural network with Fletcher–Reeves algorithm for hyperspectral image classification. *Remote Sens.* 11, 1325. doi: 10.3390/rs11111325

Cho, Y., Julier, S. J., Marquardt, N., and Bianchi-Berthouze, N. B. (2017). Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging. *Biomed. Opt. Express* 8, 4480–4503. doi: 10.1364/BOE.8.004480

Choi, Y., Kim, N., Hwang, S., and Kweon, I. S. (2016). "Thermal image enhancement using convolutional neural network," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'16*; October 09–14, 2016; Daejeon.

Christopher, M., Belghith, A., Bowd, C., Proudfoot, J. A., Goldbaum, M. H., Weinreb, R. N., et al. (2018). Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci. Rep.* 8:16685. doi: 10.1038/s41598-018-35044-9

Dauphin, Y. N., Fan, A., Auli, M., and Grangier, A. (2017). "Language modeling with gated convolutional networks," in *Proceedings of the 34th International Conference on Machine Learning, PMLR'17*; August 06–11, 2017; Sydney.

Dong, C., Loy, C. C., He, K., and Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 295–307. doi: 10.1109/TPAMI.2015.2439281

Fuentes-Pacheco, J., Torres-Olivares, J., Roman-Rangel, E., Cervantes, S., Juarez-Lopez, P., Hermosillo-Valadez, J., et al. (2019). Fig plant segmentation from aerial images using a deep convolutional encoder-decoder network. *Remote Sens.* 11:1157. doi: 10.3390/rs11101157

Giusti, A., Ciresan, D. C., Masci, J., Gambardella, L. M., and Schmidhuber, J. (2013). "Fast image scanning with deep max-pooling convolutional neural networks," in *Proceedings of the IEEE International Conference on Image processing, ICIP'13*; September 05–18, 2013; Melbourne.

Grbovic, Z., Panic, M., Marko, O., Brdar, S., and Crnojevic, V. (2019). "Wheatear detection in RGB and thermal images using deep neural networks," in *Proceedings of the International Conference on Machine Learning and Data Mining, MLDM'19*; July 20–25, 2019; New York.

He, K., Zhang, X., Ren, S., and Jian, S. (2016b). "Identity mappings in deep residual networks," in *Proceedings of the European Conference on Computer Vision*; October 11–14, 2016. Amsterdam: ECCV.

He, K., Zhang, X., Ren, S., and Sun, J. (2016a). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'16*; June 27–30, 2016; Las Vegas.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR'17*; July 21–26, 2017; Honolulu, HI, United States.

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine learning*; July 07–09, 2015. France: PMLR. Available at: http://proceedings.mlr.press/v37/ioffe15.pdf (Accessed October 19, 2020).

Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* 53, 5455–5516. doi: 10.1007/s10462-020-09825-6

Kraft, M., Weigel, H., Mejer, G., and Brandes, F. (1996). Reflectance measurements of leaves for detecting visible and non-visible ozone damage to crops. *J. Plant Physiol.* 148, 148–154. doi: 10.1016/S0176-1617(96)80307-5

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Neural Information Processing Systems Conference*; December 03–08, 2012. NV: NIPS.

Kuen, J., Kong, X., Wang, G., and Tan, Y. P. (2018). "DelugeNets: Deep networks with efficient and flexible cross-layer information inflows," in *Proceedings of the IEEE International Conference on Computer Vision Workshop, ICCVW'17*; October 22–29, 2017; Venice, Italy.

Larsson, G., Maire, M., and Shakhnarovich, G. (2016). "Fractalnet: Ultra-deep neural networks without residuals," in *Proceedings of the International Conference on Learning Rerepresentations, Toulon, ICLR'17*; April 24–26, 2016; Toulon.

Li, C., Xia, W., Yan, Y., Luo, B., and Tang, J. (2019). Segmenting objects in day and night:edge-conditioned CNN for thermal image semantic segmentation. arXiv [Preprint]. Available at: https://arxiv.org/abs/1907.10303 (Accessed July 16, 2020).

Lili, Z., Duchesne, J., Nicolas, H., Rivoal, R., and Breger, P. (1991). Détection infrarouge thermique des maladies du blé d'hiver (Infrared detection of winter-wheat diseases). *EPPO Bull.* 21, 659–672. doi: 10.1111/j.1365-2338.1991.tb01300.x

Liu, J., and Wang, X. (2020). Early recognition of tomato gray leaf spot disease based on MobileNetv2-Yolov3 model. *Plant Methods* 16:83. doi: 10.1186/s13007-020-00624-2

López, A., Molina-Aiz, F. D., Valera, D. L., and Peña, A. (2012). Determining the emissivity of the leaves of nine horticultural crops by means of infrared thermography. *Sci. Hortic.* 137, 49–58. doi: 10.1016/j.scienta.2012.01.022

Mao, X., Shen, C., and Yang, Y.-B. (2016). "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proceedings of the Advances in Neural Information Processing Systems, NIPS'16* (Barcelona).

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2015). Finely grained annotated datasets for image-based plant phenotyping. *Pattern Recogn. Lett.* 81, 80–89. doi: 10.1016/j.patrec.2015.10.013

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning, ICML'10*; June 21–24, 2010; Haifa.

Raza, S. E., Prince, G., Clarkson, J. P., and Rajpoot, N. M. (2015). Automatic detection of diseased tomato plants using thermal and stereo visible light images. *PLoS One* 10:e0123262. doi: 10.1371/journal.pone.0123262

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*. eds. N. Navab, J. Hornegger, W. Wells and A. Frangi (Cham: Springer), 234–241.

Saleem, M. H., Potgieter, J., and Mahmood Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plan. Theory* 8:468. doi: 10.3390/plants8110468

Salgadoe, A. S. A., Robson, A. J., Lamb, D. W., and Schneider, D. (2019). A non-reference temperature histogram method for determining Tc from ground-based thermal imagery of orchard tree canopies. *Remote Sens.* 11:714. doi: 10.3390/rs11060714

Scherer, D., Müller, A., and Behnke, S. (2010). "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks, ICANN 2010, Lecture Notes in Computer Science*. eds. K. Diamantaras, W. Duch and L. S. Iliadis (Berlin, Heidelberg: Springer), 92–101.

Shin, H.-C. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolution neural networks for computer aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* 35, 1285–1298. doi: 10.1109/TMI.2016.2528162

Simonyan, K., and Zisserman, A. (2014). "Two-stream convolutional networks for action recognition in videos," in *Proceedings of the 28th Conference on Neural Information Processing Systems, NIPS'14*; December 08–13, 2014; Montréal, Canada.

Simonyan, K., and Zisserman, A. (2015). "A very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations, ICLR'15*; May 07–09, 2015; Sandiego.

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Srivastava, R. K., Greff, K., and Schmidhuber, J. (2015). "Highway networks," in *Proceedings of the International Conference on Machine learning, ICML'15*; July 06–11, 2015; Lille.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going Deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVP'15*; June 06–12, 2015; Boston.

Tong, T., Li, G., Liu, X., and Gao, Q. (2017). "Image super-resolution using dense skip connections," in *Proceedings of the IEEE international conference on computer vision, ICCV'17*; October 22–29, 2017 (Venice, Italy).

Wong, A., Famouri, M., and Shafiee, M. J. (2020). AttendNets: tiny deep image recognition neural networks for the edge via visual attention condensers. arXiv [Preprint]. Available at: https://arxiv.org/abs/2009.14385v1 (Accessed July 16, 2020).

Xavier, G., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*; May 13–15, 2010; Sardinia: AISTATS.

Zhang, X., He, L., Majeed, Y., Whiting, M. D., Karkee, M., and Zhang, Q. (2018). A precision pruning strategy for improving efficiency of vibratory mechanical harvesting of apples. *Trans. ASABE* 61, 1565–1576. doi: 10.13031/trans.12825

Zhang, Q., Liu, Y., Gong, C., Chen, Y., and Yu, H. (2020). Applications of deep learning for dense scenes analysis in agriculture: a review. *Sensors* 20:1520. doi: 10.3390/s20051520

Zhang, W., Tang, P., and Zhao, L. (2019). Remote sensing image scene classification using CNN-CapsNet. *Remote Sens.* 11:494. doi: 10.3390/rs11050494

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Islam, Nakano, Lee, Tokuda and Kochi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 25 August 2021
doi: 10.3389/fpls.2021.613507

Predictive Chromatography of Leaf Extracts Through Encoded Environmental Forcing on Phytochemical Synthesis

Junelle Rey C. Bacong† *and Drandreb Earl O. Juanico**†

DataSc/ense TechnoCoRe, Technological Institute of the Philippines, Quezon City, Philippines

Environment fluctuations can influence a plant's phytochemical profile via phenotypic plasticity. This adaptive response ensures a plant's survival under fluctuating growth conditions. However, the resulting plant extract composition becomes unpredictable, which is a problem for highly standardized medicinal applications. Here we demonstrate, for the first time, the feasibility of tracking the changes in the phytochemical profile based on real-time measurements of a few environment and extract-preparation variables. As a result, we predicted the chromatograms of *Blumea balsamifera* extracts through an imputation-augmented convolutional neural network, which uses the image-transformed temporal measurements of the variables. We developed a sensor network that collected data in a greenhouse and a training algorithm that concurrently generated a data representation of the implicit plant-environment interactions leading to the mutable chromatograms of leaf extracts. We anticipate the generic applicability of the method for any plant and recognize its potential for addressing the standardization problems in plant therapeutics.

Keywords: plant-environment interactions, phenotypic plasticity, phytochemical profiling, plant therapeutics, plant extracts, *Blumea balsamifera*, convolutional neural network

OPEN ACCESS

Edited by:
Yiannis Ampatzidis,
University of Florida, United States

Reviewed by:
Luigi Milella,
University of Basilicata, Italy
Supratim Basu,
New Mexico Consortium,
United States

****Correspondence:***
Drandreb Earl O. Juanico
reb.juanico@tip.edu.ph

†ORCID:
Junelle Rey C. Bacong
orcid.org/0000-0003-4258-1969
Drandreb Earl O. Juanico
orcid.org/0000-0003-3439-3167

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *02 October 2020*
Accepted: *26 July 2021*
Published: *25 August 2021*

Citation:
Bacong JRC and Juanico DEO (2021) Predictive Chromatography of Leaf Extracts Through Encoded Environmental Forcing on Phytochemical Synthesis. Front. Plant Sci. 12:613507. doi: 10.3389/fpls.2021.613507

INTRODUCTION

Plants may be thought of as factories that synthesize highly complex and unusual substances for various medical and non-medical applications (Mishra and Tiwari, 2011; Nikam et al., 2012). These complex phytochemical mixtures in herbal or plant-derived medicines have been shown to have advantages over the single molecules that are isolated or synthetically modified from natural sources (Rodriguez-Concepcion et al., 2006; Carmona and Soares Pereira, 2013; Ekor, 2014). This has led to a tremendous increase in the use of herbal products and supplements over the past three decades, as many people around the world have resorted to using these products for treating various health-related concerns (Calixto, 2000; WHO, 2004; Ekor, 2014). However, the production of herbal medicines is a gradual and meticulous process. It involves three basic steps: (i) identification of herbs based on macroscopical and microscopical features; (ii) evaluation of drugs for the confirmation of their identity and purity; and (iii) standardization (Kunle et al., 2012; Newmaster et al., 2013). The standardization of herbal formulations encompasses all of the quality control measures taken during the manufacturing process such as sample preparation and phytochemical evaluation, as well as microbial, biological, and toxicity testing (Calixto, 2000; Rodriguez-Concepcion et al., 2006; Kunle et al., 2012; Newmaster et al., 2013).

Additionally, guidelines and protocols are utilized to ensure the safety, quality, and efficacy of all herbal products and formulations (WHO, 1998; Harvey, 2008; Sahoo et al., 2010; Newmaster et al., 2013).

In the Philippines, there are 10 herbal plants that are recommended by the Department of Health for medical applications and potential product commercialization. These 10 medicinal plants have already been scientifically and clinically validated. In fact, these plants are listed under the Republic Act No. 8423 and by the Philippine Institute of Traditional and Alternative Health Care as recommended for use in treating specific physiological problems (Ammakiw and Odiem, 2013; Boy et al., 2018). An example of which is *Blumea balsamifera* (locally known and referred to hereafter as "sambong"), is a shrub that grows across Southeast Asia, India, and China, known for managing urolithiasis and other kidney problems (Ammakiw and Odiem, 2013; Montealegre and De Leon, 2017; Boy et al., 2018). However, despite proven therapeutic effects, the herbal products derived from such medicinal plants remain difficult to commercialize because of the inconsistent use of extraction methods and the variable content in different batches of these herbal formulations (Sahoo et al., 2010; Carmona and Soares Pereira, 2013). As such, the primary goal of the standardization of herbal formulations is to ensure a reproducible quality of herbal products (Calixto, 2000; Rodriguez-Concepcion et al., 2006; Sahoo et al., 2010; Kunle et al., 2012).

A very important aspect in the standardization of plant-derived medicinal products is the phytochemical evaluation. This involves the identification and relative quantification of bioactive compounds in the herbal extracts. The evaluation is conducted by analyzing the phytochemical profile of the extracts obtained from tedious chromatographic and spectroscopic procedures. Such procedures involve the use of highly-technical setups such as liquid and gas chromatography in conjunction with mass or ultraviolet spectroscopy (LC/GC-MS, LC/GC-UV), capillary electrophoresis, nuclear magnetic resonance spectral analysis, attenuated total reflection, and Fourier transform infrared spectroscopic imaging, among others (Dias et al., 2012; Seger et al., 2013; Huck, 2015). However, despite the use of these modern chemical and analytical procedures, the determination and isolation of bioactive metabolites in plant materials remains challenging (Calixto, 2000). Such difficulty arises from the plants' inherent phenotypic plasticity in response to stress and their environment, resulting to significant variability in their phytochemical make-up. For instance, raw herbal materials cultivated and collected from the same area of vegetation may have different phytochemical profiles and may thereby exhibit different bioactivities. Pérez-Balibrea et al. (2008) showed that the light treatment of sprouting broccoli (*Brassicaceae*) seeds increases the concentration of health-promoting phytochemicals, such as vitamin C, glucosinolates, and phenolic compounds. Odjegba and Alokolaro (2013) simulated the effects of a drought and varying salinity conditions in *Acalypha wilkesiana* plants, which resulted in a decrease in the quantity of alkaloids, flavonoids, and tannins in the extracts, as well as an increase in the saponin production levels. Due to their plasticity, plants can adjust their responses to a multitude of biotic and abiotic stresses. Therefore, changes in environmental conditions such as temperature, humidity, sunlight, rainfall, and soil conditions, as well as diurnal and seasonal cycles, can promote significant variability in the phytochemical make-up of raw herbal materials.

The complex nature of plant extracts makes the development of herbal products a difficult task. A large analytical effort and high-quality manufacturing skills are needed to produce standardized and quality controlled herbal formulations (Cravotto et al., 2010; Carmona and Soares Pereira, 2013). One approach to studying the complexity of these plant extracts is through chemometrics, which aims to understand metabolomic or chromatographic data using multivariate data analysis (Parker et al., 2009; Turi et al., 2015). Chemomemtric analysis denotes the application of statistical tools such as principal component analysis (PCA) (Le Gall et al., 2003; Want et al., 2010; Worley and Powers, 2013; Wolfender et al., 2015), support vector machines (SVMs) (Zheng et al., 2009; Gromski et al., 2015), and multivariate regression models (Brown et al., 2012; Das et al., 2017; Ballesteros-Vivas et al., 2019) to examine and validate the phytochemistry of organic extracts based on their chromatographic or metabolomic profiles. Unsupervised analytical techniques such as PCA and SVMs have been used to determine the secondary metabolites that contribute to the specific bioactivity of a plant extract (Le Gall et al., 2003; Zheng et al., 2009; Want et al., 2010; Worley and Powers, 2013; Gromski et al., 2015; Wolfender et al., 2015). Multivariate regression models, a type of supervised statistical technique, have been used to correlate the extraction parameters, such as the solvent type and pH, with the concentrations of specific metabolites in the plant extracts (Brown et al., 2012; Das et al., 2017; Ballesteros-Vivas et al., 2019).

However, these types of chemometric tools usually require the cumbersome process of choosing specific features that may be suboptimal for a given task. Artificial intelligence technologies such as deep learning (LeCun et al., 2015; Schmidhuber, 2015) have generated new methods over recent years that permit the determination of the most suitable set of features within the training process, without any involvement from the investigator (Zhang et al., 2017). In natural product research wherein the volume of data sets is typically very large, deep learning methodologies have shown promising results (Chen et al., 2018; Sarker and Nahar, 2018). For instance, artificial neural networks (ANNs) (Dahmoune et al., 2015; Eftekhari et al., 2018) were trained to determine the non-linear relationship between the laboratory and extraction parameters as the inputs and the metabolite concentrations as the outputs. ANNs were also used to predict the bioactivity of plant extracts given the relative concentration of their secondary metabolites (Hosu et al., 2014; Das et al., 2017). Moreover, convolutional neural networks (CNNs) that are typically used for extracting features and classifying spatial and grid-structured data such as images have been applied to the 2D HSQC spectra of compounds from marine and terrestrial organisms for the characterization of their metabolic profiles (Zhang et al., 2017; Reher et al., 2020). This particular CNN tool leverages the wealth of these spectral data sets constructed over the past four decades from natural product research (Zhang et al., 2017). CNNs were also used to analyze

LC-MS data, particularly in classifying the true and false peaks in the LC-MS spectra (Kantz et al., 2019). The cumbersome process does not justify the prediction performance of these chemometric tools.

A missing piece of information is likely the source of the unexplained variability in existing predictive techniques applied to the chromatographic characterization of plants. Few studies have accounted for the gross effects of the environment on the phytochemical compositions of herbal extracts, which also makes the standardization of herbal formulations difficult to achieve. Most previous studies have focused primarily on characterizing specific groups of phytochemicals, such as phenolic compounds, and their related bioactivity in the extracts (Le Gall et al., 2003; Zheng et al., 2009; Want et al., 2010; Brown et al., 2012; Worley and Powers, 2013; Dahmoune et al., 2015; Gromski et al., 2015; Wolfender et al., 2015; Das et al., 2017; Eftekhari et al., 2018; Ballesteros-Vivas et al., 2019). In this work, we present a novel method for predicting the chromatogram of sambong leaf extracts using sensor data collected from the environment in which the plant has been exposed for over 1 month. We used deep learning technology, particularly CNNs, to correlate the abiotic stresses, such as the changes in temperature, humidity, ambient light, soil pH, and soil moisture, with the supposed chromatograms of the leaf extracts. Herein, we show that the environmental forcing on phytochemical synthesis can be encoded using CNNs. As a result, the trained network model can be used to accurately predict the entire chromatographic profile of plant extracts based on different time-varying environmental parameters, as well as using the controlled laboratory variables. Unlike previous studies that have focused only on analyzing specific groups of compounds, our work predicts the entire phytochemical profile that represents the synergistic contributions of each putative metabolite in the extract. As such, this method can be used to evaluate the phytochemical composition of herbal extracts without undergoing tedious laboratory and chromatographic procedures. Our work on predictive chromatography offers a fast, accurate, and high-throughput alternative for phytochemical evaluation, which is an integral component of standardizing herbal formulations. To our knowledge, this study is the first to consider the extraction of temporal information from environmental data using CNNs to predict the chromatogram of a plant extract.

MATERIALS AND METHODS

The underlying workflow of the predictive chromatography is graphically outlined in **Figure 1**. The following methods are comprised of separate data collections for input and output data sets. Subsequent pre-processing procedures were applied to both the input and output data sets prior to the neural network training and model evaluation.

Collection of Input Data Sets

An in-house remote environmental monitoring system (REMS) was installed in Los Baños, Laguna to monitor and record the real-time data for soil pH, soil moisture, ambient temperature, relative humidity, and light intensity over a 1-month study period (see **Supplementary Figures 1, 2**). The REMS consists of a plurality of sensing instruments that are made from off-the-shelf sensors and meters for detecting temperature and humidity (DHT22), soil moisture (DFRobot SKU SEN0193), ambient light (Adafruit TSL2591), and soil pH (Fisher Science Education PH700 Rapitest pH meter). These instruments are linked together via an expansion port that facilitates data transmission from the sensors. The aggregated environmental data from the linked instruments are then sent to a database server.

Collection of Output Data Sets

Chemicals

The naringenin standards were sourced from Sigma-Aldrich. The solvents used for extraction, namely ethyl acetate, methanol, and n-hexane, were all HPLC grade and were obtained from RCI Labscan. LC-MS-grade methanol, formic acid and acetonitrile with 0.1% (v/v) formic acid were purchased from Scharlau. Ultrapure water (18.2 MΩ·cm resistivity at 25 , < 10 ppb total organic carbon, passed through a 0.22–μm polyvinylidene difluoride filter) was generated from a Milli-Q Integral 5 water purification system.

Plant Cultivation and Harvesting

The sambong planting materials including the seedlings, garden soil, and pots were all obtained from Los Baños, Laguna, Philippines. All plants for the treatment experiment were obtained using the cutting method. After a rooting period of 50–60 d, healthy plants were transferred to 2-L pots containing garden soil. These plants were kept in the greenhouse for another 15–20 d to adapt and acclimatize. After this period, the plants were divided into 10 separate pots according to their respective treatments. The environmental and post-harvest processing parameters were randomized across the plant samples via a Plackett-Burman design (see **Supplementary Table 1**). Pots were placed either under sunlight or under a high-density polyethylene woven shade net (55–60% sun-shading). Pots were watered daily to maintain their respective soil moisture content, as indicated in **Supplementary Table 1**.

Sample Preparation and Liquid Chromatography

During harvest, the collected sambong leaves were washed with water, dried in a convection oven at 70°C for 5 h, ground, and stored at −20°C before use. Samples were extracted with either methanol (E1), ethyl acetate (E2), or n-hexane (E3). Each sample was prepared in six replicates. Extracts were filtered and passed through 0.2–μm polytetrafluoroethylene filters prior to LC analysis.

Ultra-high-performance liquid chromatography (UPLC) (Want et al., 2010) was performed using a Waters ACQUITY I-Class UPLC with ACQUITY photodiode array (PDA) eλ Detector. A reverse-phase Waters ACQUITY HSS C18 column (2.1-mm internal diameter×100-mm length; 1.8–μm particle size) was used and maintained at 30°C. The mobile phases consisted of 0.1% formic acid in ultrapure water (A) and 0.1% formic acid in acetonitrile (B). A gradient elution was performed at a flow rate of 0.4 mL/min with an injection volume of 2

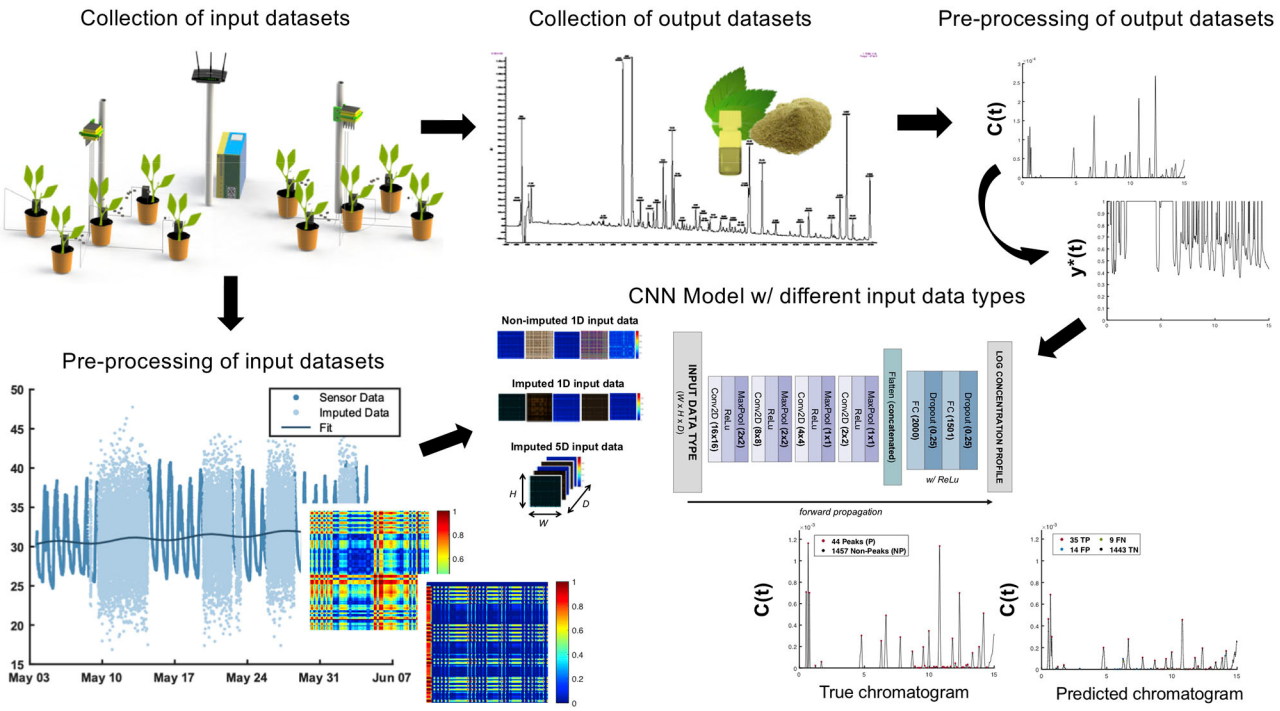

FIGURE 1 | A method of predicting the phytochemical profile of plant extracts. Predictive chromatography of sambong leaf extracts obtained using the environmental parameters such as temperature and humidity, ambient light, soil moisture, and soil pH. The training of the CNN model proceeds from the collection and pre-processing of input and output data sets obtained from REMS and LC-UV chromatography, respectively. Using the images of the environmental time-series data as inputs, the CNN model will be able to predict the relative percentage concentration profile of an extract taken from a specific sambong plant.

μL. The gradient was as follows: 20% B (0–3 min), 20–50% B (3–20 min), 50–100% B (20–22 min), 100–20% B (22–23 min), and 20% B (23–25 min). A PDA detector was used to scan the UV absorbance in the wavelength range of 200–700 nm and at a single wavelength channel of 285 nm. UV absorbances were acquired for only up to 20 min during the UPLC run time. A 40–μg/mL solution of naringenin was used as an external standard for relative quantification. All LC-UV data were acquired using MassLynx (Waters Corporation, Milford, USA).

Pre-processing of the Input Data Sets

Although the REMS was programmed to collect data approximately every 2 ms, this automated data collection may be compromised due to power interruptions, as well as other logistical and hardware concerns. We applied a stochastic regression imputation (Wang and Oates, 2015) using a stochastic fitting function to fill in the missing values in any of the environmental data sets due to these logistical issues (see **Supplementary Figure 3**). For comparison, we used both the non-imputed and imputed input data sets for training the CNN model. Non-imputed input data are sensor data that contain missing values or NaNs due to interruptions in data collection. Imputed input data are those with missing values that have been replaced or imputed with stochastic variables.

Moreover, the environmental time-series data $X = \{x_1, x_2, \ldots, x_N\}$ collected from the REMS must be normalized because it does not possess the same range as the output values. To achieve this, we applied technical indicators used in financial stock market chart analysis (Dash and Dash, 2016) such as William's R and stochastic oscillators to transform the range $[0, 1]$ while preserving any seasonality trends and auto-regressive features in the time-series data. This normalization procedure resampled our initial observation $X(t)$ in a uniform a set of $\tilde{X}(t) = \{\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_N}\}$, where each $\tilde{x_i}$ represents data collected every minute ranging between 0 and 1.

Finally, a tempo-spatial transformation (Wang and Oates, 2015; Fawaz et al., 2019) known as Gramian Angular Summation (Difference) Fields (GASF and GADF) was used to convert the resulting normalized time-series data to a 128 × 128 image (see **Supplementary Figure 4**). Upon resampling $\tilde{X}(t)$ to a 128-vector, each pixel in the resulting image therefore contains about 6 h of environmental data. As a result, k has an upper-bound value of $k_{\max} = \text{ceil}\left(\frac{n}{128}\right)$ that can be used for data augmentation. We considered multiple combinations of $k \in \{5, 20, 30, 60, 720, 1440, 4320, 7200\}$ min and $d \in \{720, 1440, 4320\}$ min for these equations to increase the number of pairwise training data for the neural network by about 12-fold (see **Supplementary Methods**).

Pre-processing of the Output Data Sets

The typical outputs for CNNs are in the range of $[0, 1]$. However, raw chromatographic data sets, specifically LC-UV data, have

absorbance units that are above the order of 10^5. Therefore, a pre-processing procedure must be conducted before training the CNN model. For instance, a min-max normalization could be applied to these output data sets to achieve the desired range. However, chromatograms are not free from noises and disturbances from the environment. Baseline drifts, for example, are caused by column or temperature changes during elution. As a result, min-max normalization could wrongly identify the minimum and maximum peak height of the signals with baseline drifts. Furthermore, the peak heights in the original chromatogram will not be preserved because the CNN model will only predict values in the range of [0, 1]. Peak heights are very important features of a chromatogram because they relate to the relative concentrations of different metabolites in the extract.

In this work, we first corrected the baseline drifts by using the BEADS algorithm (Ning et al., 2014). We then transformed the raw data $A(t)$ to its relative concentration profile $C(t)$ (see **Supplementary Figure 5**). The relative concentration of each metabolite was calculated as mg naringenin equivalents per 100 mg of dried leaves (mg/100 mg or %). The normalization of $C(t)$ is derived from its area under the curve, which is basically the concentration of all of the detected metabolites in the extracts. The resulting normalized relative % concentrations $C(t)$ will already be in the desired range of [0,1], but they are typically on the order of 10^{-4} and 10^{-3}. This order of $C(t)$ may lead to vanishing gradients and slow convergence during the training of the neural network. To mitigate these problems, we scaled up $C(t)$ to the order of 10^{-1} by taking the log normalized % relative concentration $y^*(t)$. Because this log transformation has a unique inverse, the % concentration $C(t)$ can be easily obtained from the predicted log concentration profile $y^*(t)$ of the model from any given sample.

Input Data Types and the CNN Model

In CNNs, the input data can be generalized to a spatial data set or an image of the form $W \times H \times D$, where W, H, and D refer to the width, height, and depth of the input. In this work, we formed three types of input training data with varying depths: (1) non-imputed 1d data, (2) imputed 1d data, and (3) imputed 5d data (see **Supplementary Figure 6**).

Because we have a total of five environmental parameters to correlate per one output chromatogram of a sample extract, we horizontally concatenated each 128×128 image of the parameter to form input data with a depth of $D = 1$ (1d), or of the size $640 \times 128 \times 1$. To compare the model performances achieved using different input data structures, we also stacked the five 128×128 images to form $D = 5$ (5d) input data with dimensions of $128 \times 128 \times 5$. These three types of input data were trained separately using the same CNN model (see **Supplementary Figure 7**). The CNN is composed of four convolution layers for extracting pertinent features from the input images, as well as two fully-connected layers for correlating these features to the log relative % concentration profile of the samples. A total of 6,048 pairwise input-output data were obtained after performing data augmentation on the input data set. A model was trained using 85% of the pairwise data set (randomly selected) and evaluated using the remaining 15%. The metrics used for model evaluation were the cross-correlation, R^2, and the Matthew's Correlation Coefficient (*MCC*) (Boughorbel et al., 2017). During training, we used the mean absolute error for the cost function and RMSProp for the optimization algorithm.

RESULTS

Model Evaluation for the Different Input Data Types

An example of a predicted chromatogram produced using each input data type is shown in **Figure 2**. By inspection, **Figure 2D** is the least similar to the test chromatogram (**Figure 2A**) among the other input data types. Although it contains outliers beyond the 10^{-3} range of the test chromatogram (**Figure 2A**), it was still able to recover the peak located around $t = 4.8$ min, as shown in the inset plot. Among the three input data types, the imputed 1d input data type (**Figure 2C**) yielded the most visually similar profile, as shown in **Figure 2A**.

To generalize this observation, we measured the degree of similarity between the test and predicted profiles using a cross-correlation. As shown in **Figure 3**, the imputed 1d input data type obtained the highest average cross-correlation of $\mu_{xcorr} = 0.798 \pm 0.163$ (s.d). This indicates that the predictions from the model obtained using the imputed 1d input data have a high degree of similarity to the test samples. At the extreme end is the imputed 5d input data type, which demonstrated the lowest average cross-correlation of $\mu_{xcorr} = 0.013 \pm 0.011$ (s.d). This very low average cross-correlation can be attributed to the outliers observed in **Figure 2D**. If these points were to be filtered from the raw predictions of the model, the cross-correlation for the 5d input data will increase dramatically to $\mu_{xcorr} = 0.771 \pm 0.170$ (s.d).

In **Figure 4**, we quantified the accuracy of the predictions by measuring the coefficient of determination, or R^2, between the test and predicted profiles. Unlike the cross-correlation that measures the overall similarity of two signals based on their phase difference, the R^2-value measures the accuracy of the predicted $y^*(t)$. We observed in **Figures 4C,F** that the predictions from the imputed 5d input data type have a higher mean R^2 [$\mu_{R^2} = 0.426 \pm 0.183$ (s.d)] despite having the lowest μ_{xcorr} compared to the non-imputed input data type [$\mu_{R^2} = 0.394 \pm 0.173$ (s.d)]. This can be attributed to the presence of outliers in the predicted chromatograms. These outliers skew the resulting regression model away from the non-outlier data points, thereby increasing R^2.

Interestingly, the model with 5d input data performed poorly compared to the model that uses 1d input data, despite both containing the same amount of temporal information from the sensor data. This suggests that the predictive performance of a model does not only depend on data integrity, but also on the structure of the input layer. More complicated structures of the input layer require complex combinations of filters and weights of the CNN. In 5d input data sets, two additional 128×128 images were stacked in addition to the usual 3d inputs (representing the RGB channels in images) for the CNN. A different architecture might be required to attain an equal or greater performance than

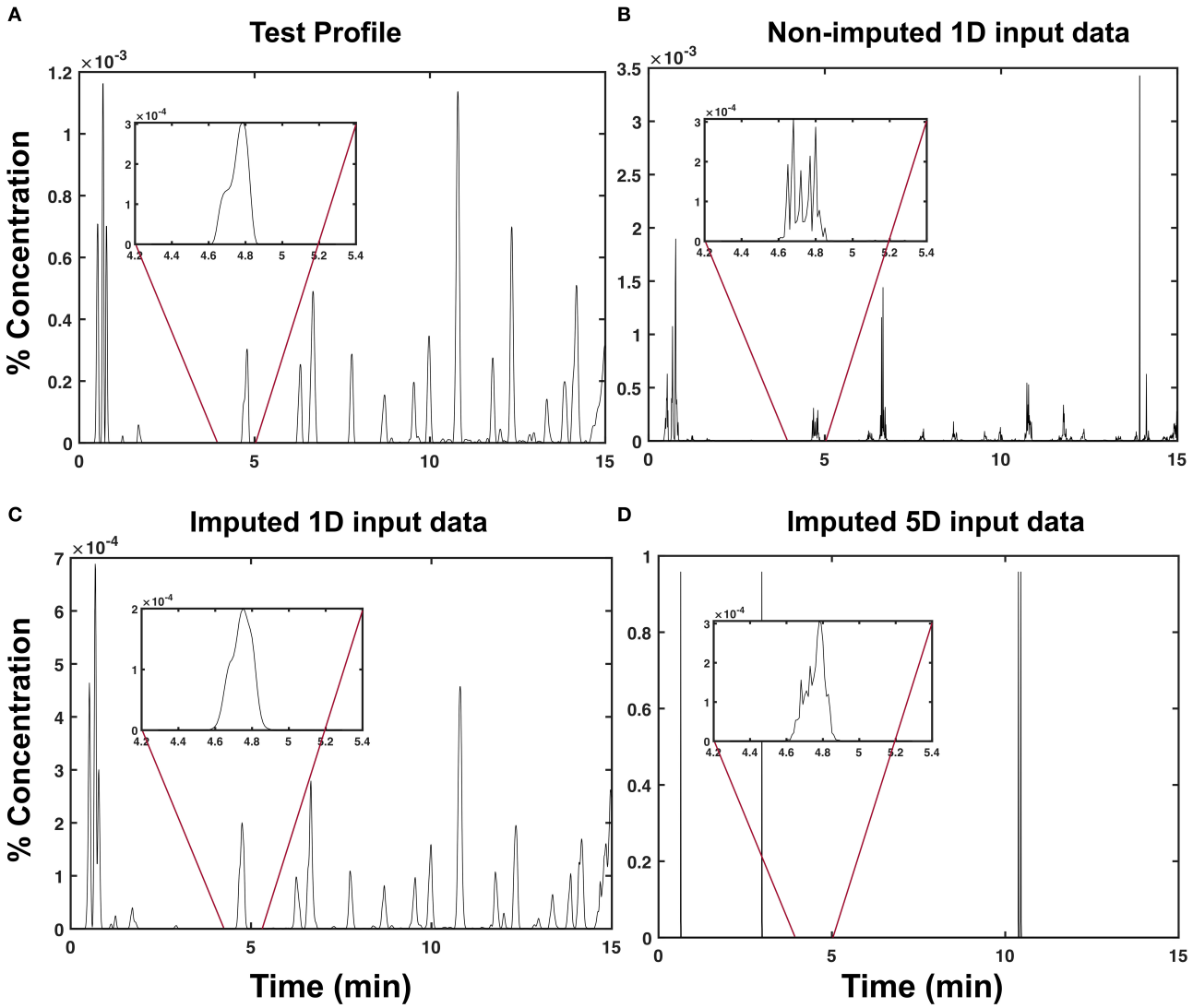

FIGURE 2 | Predicted chromatograms of sambong leaf extracts. Sample chromatographic profiles ($C(t)$ in % concentration) of a test extract obtained via **(A)** LC-UV chromatography, **(B)** model prediction using non-imputed 1d input data, **(C)** model prediction using imputed 1d input data, and **(D)** model prediction using imputed 5d input data.

achieved by the 1d inputs. Nonetheless, it can be observed from **Figures 4B,E** that using the imputed 1d input data type for the given neural network yields the most accurate predictions among the three input data types [$\mu_{R^2} = 0.470 \pm 0.192$ (s.d)].

Peak Evaluation in the Predicted Profiles

The most important feature in a chromatogram is its peaks. A peak represents a metabolite, and the area under its curve is related to the concentration of that metabolite in the sample. To assess the performance of the CNN in terms of peak reconstruction and classification, we matched the peaks identified in the test and predicted chromatograms as shown in **Figure 5**. We only considered the predictions resulting from the 1d input data types because they both demonstrated a higher degree of similarity with the test chromatograms compared to that obtained using the 5d input data.

In matching the predicted peaks p' with the test peaks $\boldsymbol{P}$, we first classify a predicted peak p' as a true peak tp if it lies within a tolerance $n\sigma$ of the test peak p. This peak tolerance also addresses the peak shifts that may have occurred during the chromatography procedure, thereby making this classification of predicted peaks robust to such disturbances. Mathematically, the set of true peaks tp can be expressed as:

$$tp = \left\{ p' \in P' \,\middle|\, p - n\sigma \leq p' + n\sigma' \leq p + n\sigma,\ \forall\, n \in Z, p \in P \right\} \quad (1)$$

where $\pm n\sigma$ is the peak tolerance and σ, σ' are the standard deviations of the Gaussian curves approximated by the peaks p and p', respectively. From this definition, we could also identify the false positive peaks fp as the set of predicted peaks p' that

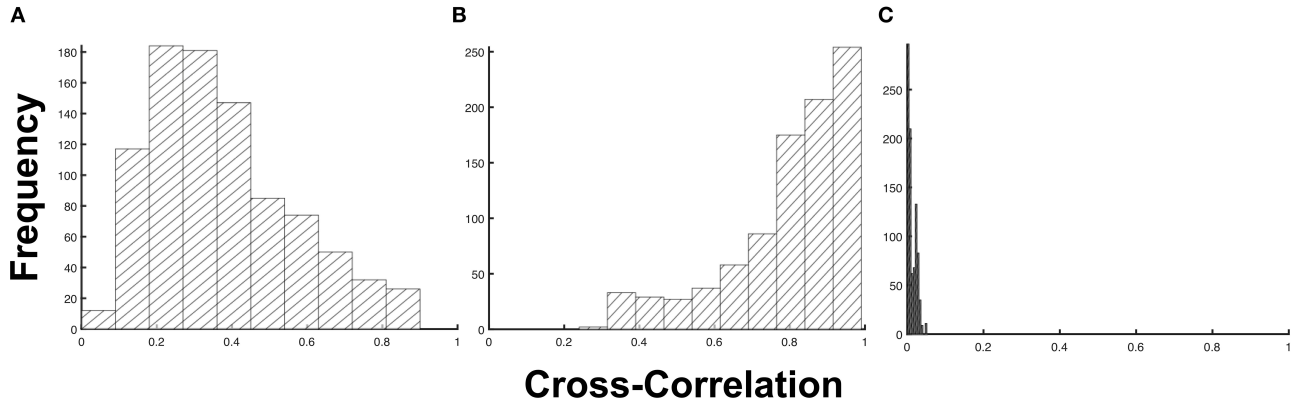

FIGURE 3 | Frequency distribution of the cross-correlation metric. The cross-correlation distribution for **(A)** non-imputed 1d input data ($p = 1.75 \times 10^{-4}$), **(B)** imputed 1d input data ($p = 1.09 \times 10^{-15}$), and **(C)** imputed 5d input data ($p = 0.89$) at a 5% level of significance.

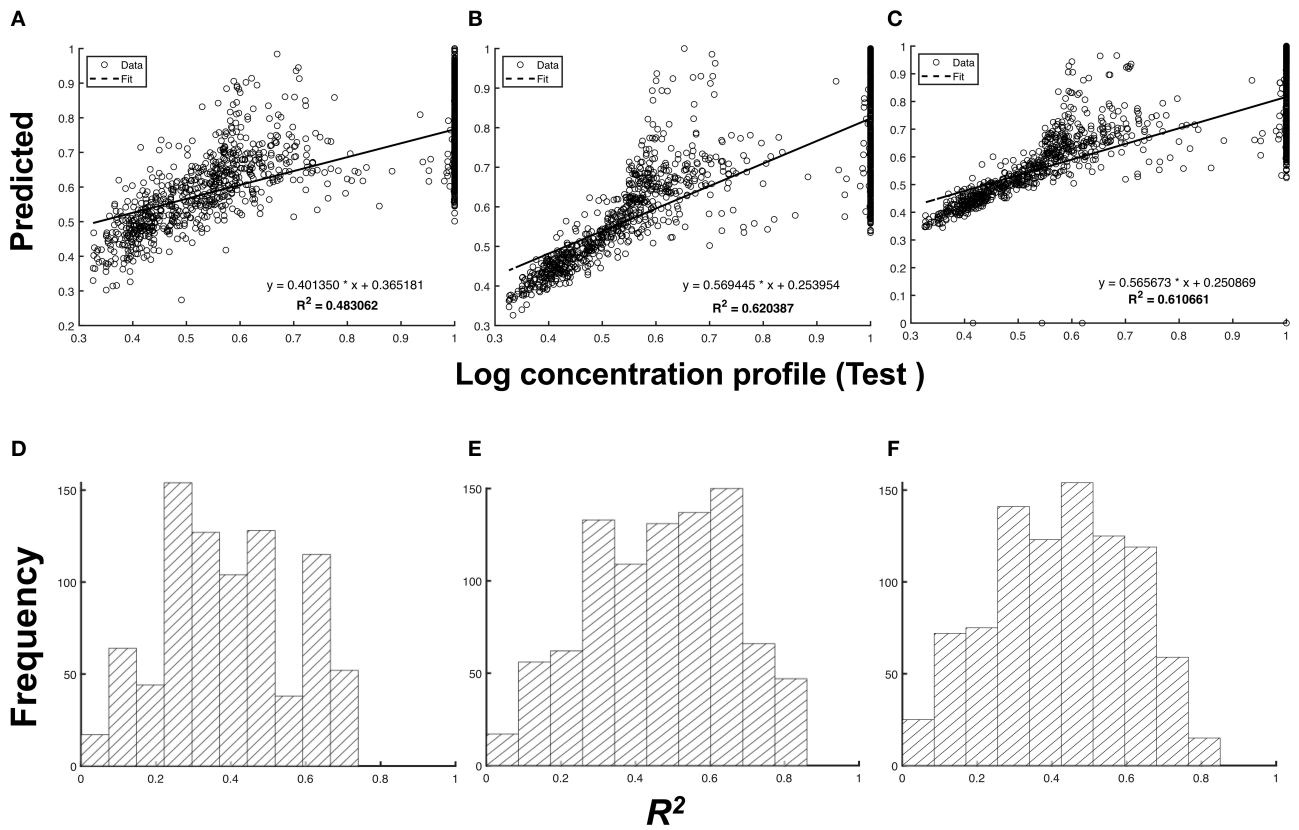

FIGURE 4 | Comparison between the predicted and test chromatogram across different types of input data. Scatterplot of the log-transformed chromatogram, predicted vs. the test chromatogram: **(A)** non-imputed 1d input data, **(B)** imputed 1d input data, and **(C)** imputed 5d input data. The frequency distribution of R^2 for all test samples for **(D)** non-imputed 1d input data ($p = 1.36 \times 10^{-4}$), **(E)** imputed 1d input data ($p = 2.50 \times 10^{-6}$), and **(F)** imputed 5d input data ($p = 1.89 \times 10^{-5}$) at a 5% significance level.

do not correspond to any peaks in the test profiles ($fp = \{p' \in \boldsymbol{P'} | \, p' + n\sigma' \notin tp\}$). Conversely, false negatives are those non-peaks in the predicted profile that should have been classified as true peaks by the model ($fn = \left\{np' \in \boldsymbol{NP'} \;\middle|\; np' \in tp\right\}$), while true negatives are those non-peak points in both the predicted and test profiles ($tn = \left\{np' \in \boldsymbol{NP'} \;\middle|\; np' \notin tp\right\}$). Using these definitions, we may evaluate the performance of the model in terms of the peak classification using Matthew's Correlation Coefficient

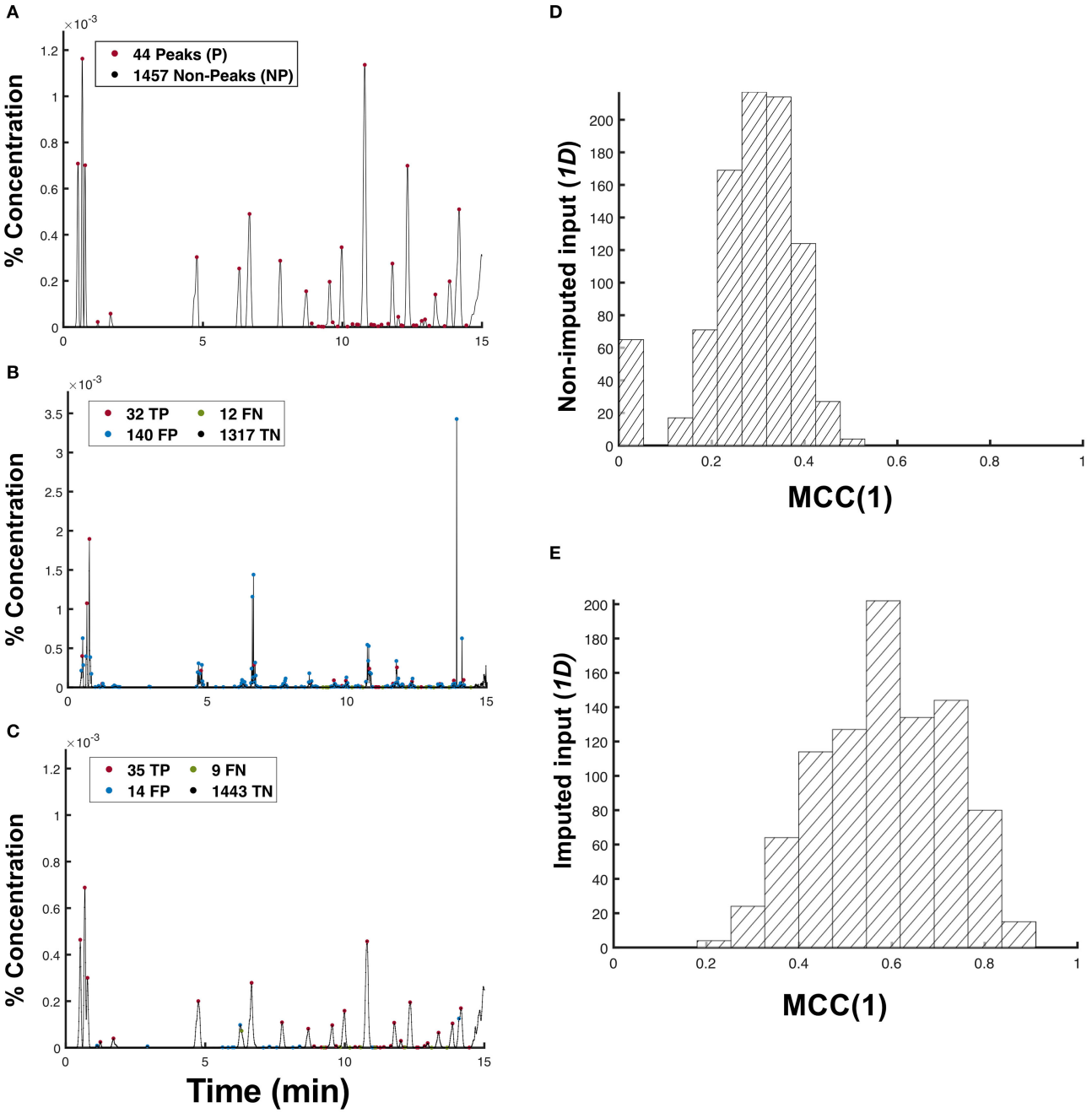

FIGURE 5 | Matching of peaks between the predicted and test chromatogram. Peak classifications for $N = 1,501$ points in the **(A)** test chromatogram, and the predictions from the **(B)** non-imputed and **(C)** imputed 1d input data types. The *MCC*(1) distribution for the **(D)** non-imputed input data ($p = 4.50 \times 10^{-3}$) and **(E)** imputed 1-channel input data type ($p = 1.89 \times 10^{-5}$) at a 5% significance.

(MCC) given by:

$$MCC\,(n) = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TP + FN) \times (TN + FP)}} \quad (2)$$

where TP, TN, FP, and FN are the cardinality of the sets tp, fp, tn, and fn, respectively. An MCC of $+1$ implies a perfectly correct predictor; an MCC of 0 is as good as a random guess; and an MCC of -1 implies a perfectly wrong predictor. We used MCC to evaluate the performance of our model because the distribution of the peak types in a chromatogram is imbalanced.

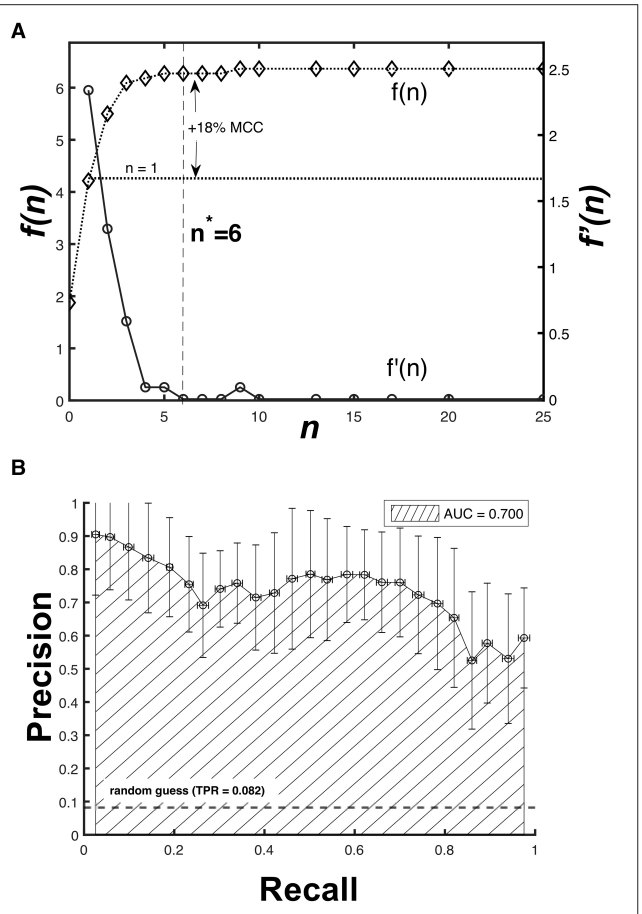

FIGURE 6 | Optimization results for the peak tolerance and peak matching accuracy. **(A)** The rate of change $f'(n)$ is monotonically decreasing for $1 \leq n \leq 6$, which suggests an optimum value $n^* = 6$ resulting to an 18% improvement in *MCC*. **(B)** The Precision-Recall *AUC* curve for *MCC*(6) shows 70% accuracy of the CNN model.

In **Figure 5**, the model obtained using a non-imputed data input has a higher *TP* compared to the model obtained using an imputed data input. However, its $MCC\,(1) = 0.3556$ is significantly lower compared to the latter model with $MCC(1) = 0.6736$. This huge difference in $MCC(1)$ is clearly a result of the presence of non-smooth peaks, as shown in **Figure 2B**. Because most peak detection algorithms function by using the first derivative test, those unwarranted sharp peaks in **Figure 2B** are classified as false positives peaks. The more false-positive or false-negative peaks that the model can classify, the lower its *MCC* value will be. This observation is evident in **Figures 5D,E**, wherein the non-imputed input data is shown to have obtained a significantly lower $\mu_{\mathrm{MCC}(1)} = 0.283 \pm 0.104$ (s.d) compared to the imputed 1d input data type with $\mu_{\mathrm{MCC}(1)} = 0.587 \pm 0.138$ (s.d).

As the peak classification hinges on the peak tolerance $n\sigma$, there exists a value $n = n^*$ such that the increase of $MCC(n)$ is no longer significant for $n > n^*$. We considered the ratio $f\,(n)$ as the basis of our optimization (Equation 3).

$$n^* = arg\ min\left\{f'\,(n)\,\middle|\, f\,(n) = \frac{\boldsymbol{\mu}_{MCC}}{\boldsymbol{\sigma}_{MCC}}\right\} \tag{3}$$

The solution for Equation 3 is shown in **Figure 6A**. Although $f(n)$ continues to increase for larger n values, the rate of change $f'(n)$ is monotonically decreasing for $1 \leq n \leq 6$, with $f'(6) \approx 10^{-3}$. This means that increasing n further corresponds to a diminishing gain in the peak classification accuracy. At the optimum value of $n^* = 6$, we obtained $\mu_{MCC(6)} = 0.691 \pm 0.110$ (s.d), which is a significant 18% improvement ($p < 0.001$, one-tailed t-test) from the previous mean we considered where $\mu_{MCC(1)} = 0.587 \pm 0.138$ (s.d).

Furthermore, we also cross-validated the model (60-fold) using different partitions of the imputed 1d input data set to obtain its overall performance. **Figure 6B** shows that the area under the curve (*AUC*) of the precision-recall plot for *MCC*(6) is equal to $AUC = 0.70$. A perfect classifier has an $AUC = 1$, suggesting that the CNN model is a sufficient predictor despite having only been trained with sensor data covering a month-long period. Increasing the amount of time over which a data set is available will further improve the predictive performance of the trained model.

Model Performance for Different Solvents

The appearance of the chromatogram is dependent on the solvent that is used to perform the extraction. In this work, the solvent system spans the extremes of dielectric constants, which could likely indicate that the chromatograms would represent subsets of very different extracted compounds. To determine if the model would perform better using a particular solvent system, we also assessed the model performance in relation to the type of solvent used in each sample. **Figure 7** summarizes the differences among the three solvents after cross-validating the model using the non-imputed and imputed 1d data sets (k-folds = 60 folds). Consistently, the results of the cross-validation showed a better performance for the model that uses imputed 1d data (see **Supplementary Table 2**).

In **Figure 7**, we observe differences in the average metric scores among the three solvents. This suggests that the model has a preference toward a specific solvent system. In particular, methanol (E1) has been shown to have the highest average metric scores for the imputed input data. The difference observed between ethyl acetate and n-hexane using this input data is significant ($p < 0.001$, one-tailed t-test), which implies that this trained model will more accurately predict the chromatograms of extracts with methanol compared to those with ethyl acetate or n-hexane (see **Supplementary Tables 3, 4**). Considering the model obtained using non-imputed input data, the difference between the methanol and ethyl acetate solvents is also significant. However, the model's preference for the best solvent system is inconsistent as it fluctuates between these two solvents depending on the metric that is being considered (see **Supplementary Tables 3, 5**).

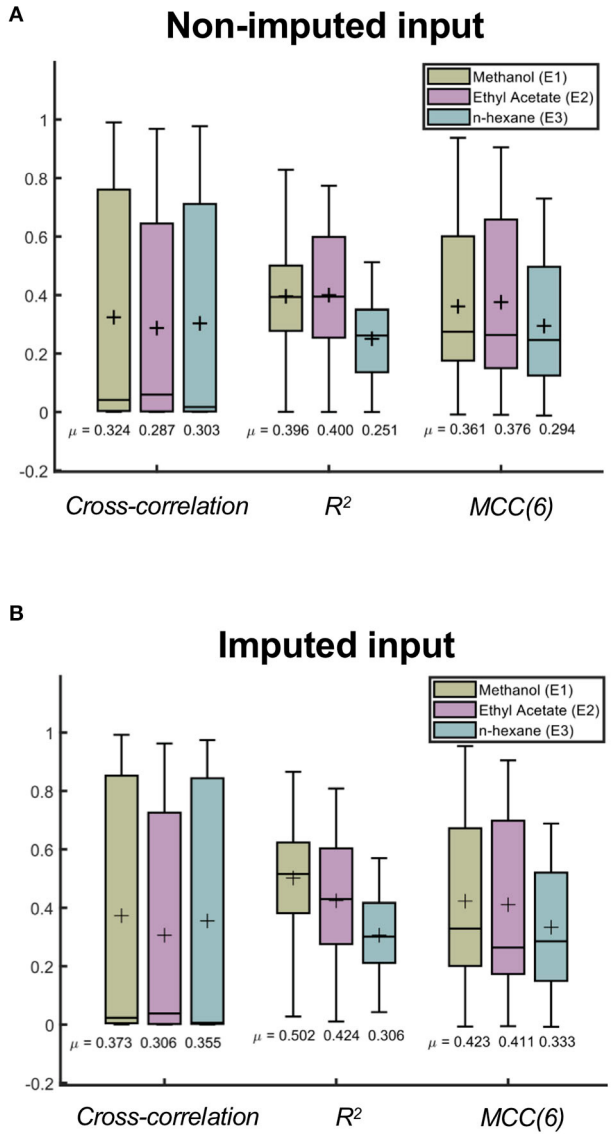

FIGURE 7 | Comparing the model performance across three solvent systems used during extraction. Randomized cross-validation results (60 folds) of the CNN model for different solvent systems in **(A)** non-imputed ($E1 = 24,996$ samples, $E2 = 27,147$ samples, and $E3 = 6,877$ samples) and **(B)** imputed ($E1 = 19,218$ samples, $E2 = 21,060$ samples, and $E3 = 5,122$ samples) 1d data sets. The cross mark represents the average metric score of each distribution.

DISCUSSION

Typically, controllable laboratory variables, such as solvent systems and ratios, are studied and standardized when evaluating the phytochemistry and bioactivity of herbal extracts. However, the plants' phenotypic plasticity in response to stress and their environment can also add significant variability to the phytochemical make-up of raw herbal materials. This inherent variability in plant extracts caused by plant-environment interactions make the standardization of herbal formulations, and other plant therapeutics challenging. Here, we have demonstrated the feasibility of tracking the changes in the phytochemical profile of plant extracts based on real-time measurements of a few environment and extract-preparation variables. As a result, we predicted the chromatograms of the *Blumea balsamifera* leaf extracts using an imputation-augmented convolutional neural network (CNN) that uses the image-transformed temporal measurements of the variables.

The methods that we have established in this work involve many data pre-processing steps that are inspired from multiple scientific disciplines. To pre-process the input data, the following steps were involved: (1) stochastic imputation for the missing sensor values usually applied in statistics involving real world data; (2) tempo-spatial transformation of the time series using GASFs and GADFs that are conceptually equivalent to the Gramian matrix in linear algebra; and (3) data augmentation using technical indicators that are commonly applied in stock chart analysis. The amalgamation of these seemingly unrelated techniques is what allowed us to normalize and use time-series data as an input for the CNN model that conventionally utilizes only spatial data sets. Moreover, our results also showed the importance of these pre-processing procedures, particularly imputing missing sensor data to improve the accuracy of the neural network model. Overall, deep learning strategies such as CNNs depend not only on the amount, but also on the quality of information that can be extracted from the data sets.

Furthermore, our methods can also address the baseline and peak shifts that commonly appear in chromatograms due to column or temperature changes during elution. Corrections in the baseline shifts are learned by the model as the training data sets, in particular the output chromatograms, undergo pre-processing using the BEADS algorithm. The peak shifts, conversely, can be resolved by optimizing the peak tolerance for each predicted peak in the profile. Aside from the physical disturbances that occur during elution, chromatograms are also affected by the choice of solvent used during extraction. From our results, we showed that there is a significant difference in the accuracy of the predictions obtained using different solvent types that span the extremes of dielectric constants. More specifically, we found that the trained model could more accurately predict the chromatogram of extracts when methanol was used. We found that methanol has the highest dielectric constant of 33 at 20 compared to ethyl acetate (6.08) and n-hexane (1.89). Although the scope of this work focuses primarily on environmental forcing as it effects phytochemical synthesis, we have also demonstrated the extent of this method in providing insights about the effect of solvent types on the predicted phytochemical profile of the extracts.

This novel approach for predictive chromatography highly depends on the volume and veracity of the training data sets, which include both the environmental and the laboratory parameters. If trained with a sufficient amount of data, this method could provide an alternative high-throughput chromatography procedure for the identification and relative quantification of bioactive compounds for plant therapeutics. Unlike other methods used for the phytochemical profiling of plant extracts, the trained CNN model would only rely on the time-varying environmental data obtained for an area of vegetation. Without requiring any tedious laboratory procedures,

this method would be able to accurately and rapidly predict the phytochemical profile of a particular plant extract using only the data collected by the REMS. In future studies, the author would recommend the use of a more comprehensive and robust environmental monitoring system for collecting data over longer cultivation periods to observe the effects of year-long seasonal patterns on the phytochemistry of sambong leaves.

Although the proposed technique used LC-UV chromatograms, it may also work with chromatograms generated by other spectral approaches, such as LC-MS. Furthermore, while we applied this method toward extracts taken from the leaf of a *Blumea balsamifera*, it is also possible to use the same framework for other plant species and for plant extracts from various parts of the plant, such as the root, seed, or fruit. For example, the same set of environmental time-series data could predict the chromatograms of extracts obtained from different plants or plant parts exposed to the same conditions, such as those present in a greenhouse.

Therefore, the proposed method may also function as an encoder of environmental forcing on phytochemistry across multiple herbal species. The encoder could be useful for controlling the growth and extract preparation conditions to intensify the expression of specific bioactive compounds, or the combinations thereof, in the extracts. The method can also detect the impact of climate change in the form of significant structural modifications to the phytochemical compositions of plant species over extended periods of time.

Moreover, the proposed method may also be used to discover previously unknown metabolites that contribute to the observed therapeutic effects of the herbal extracts. This accurate and high-throughput alternative to the tedious laboratory and chromatographic procedures could permit the fast screening of putative bioactive compounds across multiple herbal species. Therefore, the synthesis of both herbal formulations or single-molecule medicines could become much faster.

Another practical application of the proposed method is quality-assurance verification at the phytochemical level for plant-derived produce from farms. An automated system for identifying which environmental factors exert a significant impact on plant phytochemistry could provide valuable insights for optimizing produce characteristics. For example, farmers could use such insights to enhance their current farming practices to increase production and improve quality control of their products. Although in this study we saw the limitations of our method in terms of continuous power supply and robust sensing instrumentation, we believe that the fast pace technology advancement would address these limitations and enhance the practicability of our method, especially in the actual farm setting.

Lastly, this framework that may be used to attribute the environment's influence on a plant's ability to synthesize compounds will be useful in the analytical chemistry of natural products in the future. It provides a direct and scalable means to encode complex environmental influences on the chemical synthesis processes within a plant. This framework is direct because it considers the impact of a combination of multiple environmental factors simultaneously without referencing any particular molecular theory of forcing. It is scalable because the method could assimilate additional environmental factors to obtain more accurate and precise predictions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

DJ conceptualized and designed the study. JB performed the experiments and simulations. Both authors analyzed and interpreted the data and wrote the manuscript.

FUNDING

This project was funded by the Department of Science and Technology (DOST) through the Science for Change Program (S4CP)—Collaborative Research and Development to Leverage Philippine Economy (CRADLE) and monitored by DOST—Philippine Council for Health Research and Development (PCHRD) with Project No. 7787.

ACKNOWLEDGMENTS

We thank P. V. Nonat and N. J. C. Aguel for developing the instrumentation system for environmental data collection, and Isagani Padolina and Rowell Abogado of the Pascual Pharma Corp. (PPC) for providing us the chromatography data and pertinent analysis. DJ also acknowledges the support from the CHED DARETO Cycle 1 funding for MR-SUAVE.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.613507/full#supplementary-material

REFERENCES

Ammakiw, C. L., and Odiem, M. P. (2013). "Availability, preparation and uses of herbal plants in kalinga, philippines," in *1st Global Multidisciplinary eConference*, *Vol. 3*. (Macedonia: UNESCO's World Science Day Celebration (European Scientific Institute).

Ballesteros-Vivas, D., Álvarez-Riverab, G., Morantes, S. J., Sánchez-Camargo, A. d. P., Ibáñez, E., Parada-Alfonso, F., et al. (2019). An integrated approach for the valorization of mango seed kernel: efficient extraction solvent selection, phytochemical profiling and antiproliferative activity assessment. *Food Res. Int.* 126, 108616. doi: 10.1016/j.foodres.2019.108616

Boughorbel, S., Jarray, F., and El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews correlation coefficient metric. *PLoS ONE* 12:e0177678. doi: 10.1371/journal.pone.0177678

Boy, H. I. A., Rutilla, A. J. H., Santos, K. A., Ty, A. M. T., Yu, A. I., Mahboob, T., et al. (2018). Recommended medicinal plants as source of natural products: a review. *Digit. Chin. Medicine* 1, 131–142. doi: 10.1016/S2589-3777(19)30018-7

Brown, P. N., Turi, C. E., Shipley, P. R., and Murch, S. J. (2012). Comparisons of large (*Vaccinium macrocarpon Ait.*) and small (*Vaccinium oxycoccos L.*, *Vaccinium vitis-idaea L.*) cranberry in British Columbia by phytochemical determination, antioxidant potential, and metabolomic profiling with chemometric analysis. *Planta Med.* 78, 630–640. doi: 10.1055/s-0031-1298239

Calixto, J. B. (2000). Efficacy, safety, quality control, marketing and regulatory guidelines for herbal medicines (phytotherapeutic agents). *Braz. J. Med. Biol. Res.* 33, 179–189. doi: 10.1590/S0100-879X2000000200004

Carmona, F., and Soares Pereira, A. M. (2013). Herbal medicines: old and new concepts, truths and misunderstandings. *Rev. Brasileira de Farmacog.* 23, 379–385. doi: 10.1590/S0102-695X2013005000018

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., and Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discov. Today* 23, 1241–1250. doi: 10.1016/j.drudis.2018.01.039

Cravotto, G., Boffa, L., Genzini, L., and Garella, D. (2010). Phytotherapeutics: an evaluation of the potential of 1000 plants. *J. Clin. Pharm. Ther.* 35, 11–48. doi: 10.1111/j.1365-2710.2009.01096.x

Dahmoune, F., Remini, H., Dairi, S., Aoun, O., Moussi, K., and Bouaoudia-Madi, N. (2015). Ultrasound assisted extraction of phenolic compounds from *P. lentiscus L. leaves:* comparative study of artificial neural network (ann) versus degree of experiment for prediction ability of phenolic compounds recovery. *Ind. Crop. Prod.* 77, 251–261. doi: 10.1016/j.indcrop.2015.08.062

Das, A. B., Goud, V. V., and Das, C. (2017). Extraction of phenolic compounds and anthocyanin from black and purple rice bran (*Oryza sativa L.*) using ultrasound: a comparative analysis and phytochemical profiling. *Ind. Crop. Prod.* 95, 332–341. doi: 10.1016/j.indcrop.2016.10.041

Dash, R., and Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The J. Finance Data Sci.* 2, 42–57. doi: 10.1016/j.jfds.2016.03.002

Dias, D. A., Urban, S., and Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites* 2, 303–336. doi: 10.3390/metabo2020303

Eftekhari, M., Yadollahi, A., Ahmadi, H., Shojaeiyan, A., and Ayyari, M. (2018). Development of an artificial neural network as a tool for predicting the targeted phenolic profile of grapevine (*Vitis vinifera*) foliar wastes. *Front. Plant Sci.* 9:837. doi: 10.3389/fpls.2018.00837

Ekor, M. (2014). The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. *Front. Pharmacol.* 4:177. doi: 10.3389/fphar.2013.00177

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P. A. (2019). Deep learning for time series classification. *Data Min. Knowl. Discov.* 33, 917–963. doi: 10.1007/s10618-019-00619-1

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., and Turner, M. L. (2015). A tutorial review: metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Anal. Chimica Acta* 879, 10–23. doi: 10.1016/j.aca.2015.02.012

Harvey, A. L. (2008). Natural products in drug discovery. *Drug Discov. Today* 13, 894–901. doi: 10.1016/j.drudis.2008.07.004

Hosu, A., Cristea, V.-M., and Cimpoiu, C. (2014). Analysis of total phenolic, flavonoids, anthocyanins and tannins content in romanian red wines: prediction of antioxidant activities and classification of wines using artificial neural networks. *Food Chem.* 150, 113–118. doi: 10.1016/j.foodchem.2013.10.153

Huck, C. W. (2015). Advances of infrared spectroscopy in natural product research. *Phytochem. Lett.* 11, 384–393. doi: 10.1016/j.phytol.2014.10.026

Kantz, E. D., Tiwari, S., Watrous, J. D., Cheng, S., and Jain, M. (2019). Deep neural networks for classification of LC-MS spectral peaks. *Anal. Chem.* 91, 12407–12413. doi: 10.1021/acs.analchem.9b02983

Kunle, O. F., Egharevba, H. O., and Ahmadu, P. O. (2012). Standardization of herbal medicines—a review. *Int. J. Biodivers. Conserv.* 4, 101–112. doi: 10.5897/IJBC11.163

Le Gall, G., Colquhoun, I. J., Davis, A. L., Collins, G. J., and Verhoeyen, M. E. (2003). Metabolite profiling of tomato (*Lycopersicon esculentum*) using 1H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J. Agric. Food Chem.* 51, 2447–2456. doi: 10.1021/jf0259967

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 61, 436–444. doi: 10.1038/nature14539

Mishra, B. B., and Tiwari, V. K. (2011). Natural products: an evolving role in future drug discovery. *Eur. J. Med. Chem.* 46, 4769–4807. doi: 10.1016/j.ejmech.2011.07.057

Montealegre, C. M., and De Leon, R. L. (2017). Effect of blumea balsamifera extract on the phase and morphology of calcium oxalate crystals. *Asian J. Urol.* 4, 201–207. doi: 10.1016/j.ajur.2016.08.009

Newmaster, S. G., Grguric, M., Shanmughanandhan, D., Ramalingam, S., and Ragupathy, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med.* 11:222. doi: 10.1186/1741-7015-11-222

Nikam, P., Kareparamban, J., Jadhav, A., and Kadam, V. (2012). Future trends in standardization of herbal drugs. *J. Appl. Pharm. Sci.* 2, 38–44. doi: 10.7324/JAPS.2012.2631

Ning, X., Selesnick, I. W., and Duval, L. (2014). Chromatogram baseline estimation and denoising using sparsity (BEADS). *Chemom. Intell. Lab. Syst.* 139, 156–167. doi: 10.1016/j.chemolab.2014.09.014

Odjegba, V., and Alokolaro, A. (2013). Simulated drought and salinity modulates the production of phytochemicals in *Acalypha wilkesiana*. *J. Plant Stud.* 2, 105–112. doi: 10.5539/jps.v2n2p105

Parker, D., Beckmann, M., Zubair, H., Enot, D. P., Caracuel-Rios, Z., Overy, D. P., et al. (2009). Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by *Magnaporthe grisea*. *Plant J.* 59, 723–737. doi: 10.1111/j.1365-313X.2009.03912.x

Pérez-Balibrea, S., Moreno, D. A., and García-Viguera, C. (2008). Influence of light on health-promoting phytochemicals of broccoli sprouts. *J. Sci. Food Agric.* 88, 904–910. doi: 10.1002/jsfa.3169

Reher, R., Kim, H. W., Zhang, C., Mao, H. H., Wang, M., Nothias, L.-F. L., et al. (2020). A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. *J. Am. Chem. Soc.* 142, 4114–4120. doi: 10.1021/jacs.9b13786

Rodriguez-Concepcion, M., Avalos, J., Bonet, M. L., Boronat, A., Gomez-Gomez, L., Hornero-Mendez, D., et al. (2006). A global perspective on carotenoids: metabolism, biotechnology, and benefits for nutrition and health. *Prog. Lipid Res.* 70, 62–93. doi: 10.1016/j.plipres.2018.04.004

Sahoo, N., Manchikanti, P., and Dey, S. (2010). Herbal drugs: standards and regulation. *Fitoterapia* 81, 462–471. doi: 10.1016/j.fitote.2010.02.001

Sarker, S. D., and Nahar, L. (2018). "Chapter 1-an introduction to computational phytochemistry," in *Computational Phytochemistry*, eds S. D. Sarker and L. Nahar (Amsterdam: Elsevier), 1–41.

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Seger, C., Sturm, S., and Stuppner, H. (2013). Mass spectrometry and NMR spectroscopy: modern high-end detectors for high resolution separation techniques–state of the art in natural product HPLC-MS, HPLC-NMR, and CE-MS hyphenations. *Nat. Prod. Rep.* 30, 970–987. doi: 10.1039/c3np70015a

Turi, C. E., Finley, J., Shipley, P. R., Murch, S. J., and Brown, P. N. (2015). Metabolomics for phytochemical discovery: development of statistical approaches using a cranberry model system. *J. Nat. Prod.* 78, 953–966. doi: 10.1021/np500667z

Wang, Z., and Oates, T. (2015). "Imaging time-series to improve classification and imputation," in *Proceedings of the 17th International Conference on Artificial Intelligence* (Las Vegas, NV), 3939–3945.

Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., et al. (2010). Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protocols* 5, 1005. doi: 10.1038/nprot.2010.50

WHO (1998). *Quality Control Methods for Medicinal Plant Materials*. Geneva: World Health Organization.

WHO (2004). *WHO Guidelines on Safety Monitoring of Herbal Medicines in Pharmacovigilance Systems*. Technical Report. World Health Organization, Geneva.

Wolfender, J. L., Marti, G., Thomas, A., and Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. *J. Chromat. A* 1382, 136–164. doi: 10.1016/j.chroma.2014.10.091

Worley, B., and Powers, R. (2013). Multivariate analysis in metabolomics. *Curr. Metabolomics* 1, 92–107. doi: 10.2174/2213235X130108

Zhang, C., Idelbayev, Y., Roberts, N., Tao, Y., Nannapaneni, Y., and Duggan, B. M. (2017). Small molecule accurate recognition technology (SMART) to enhance natural products research. *Sci. Rep.* 7, 2045–2322. doi: 10.1038/s41598-017-13923-x

Zheng, L., Watson, D. G., Johnston, B. F., Clark, R. L., Edrada-Ebel, R., and Elseheri, W. (2009). A chemometric study of chromatograms of tea extracts by correlation optimization warping in conjunction with PCA, support vector machines and random forest data modeling. *Anal. Chimica Acta* 642, 257–265. doi: 10.1016/j.aca.2008.12.015

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bacong and Juanico. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 29 September 2021
doi: 10.3389/fpls.2021.693521

Hyperspectral Imaging Combined With Deep Transfer Learning for Rice Disease Detection

Lei Feng[1,2], *Baohua Wu*[1,2], *Yong He*[1,2] *and Chu Zhang*[3]*

[1] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2] Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou, China, [3] School of Information Engineering, Huzhou University, Huzhou, China

Various rice diseases threaten the growth of rice. It is of great importance to achieve the rapid and accurate detection of rice diseases for precise disease prevention and control. Hyperspectral imaging (HSI) was performed to detect rice leaf diseases in four different varieties of rice. Considering that it costs much time and energy to develop a classifier for each variety of rice, deep transfer learning was firstly introduced to rice disease detection across different rice varieties. Three deep transfer learning methods were adapted for 12 transfer tasks, namely, fine-tuning, deep CORrelation ALignment (CORAL), and deep domain confusion (DDC). A self-designed convolutional neural network (CNN) was set as the basic network of the deep transfer learning methods. Fine-tuning achieved the best transferable performance with an accuracy of over 88% for the test set of the target domain in the majority of transfer tasks. Deep CORAL obtained an accuracy of over 80% in four of all the transfer tasks, which was superior to that of DDC. A multi-task transfer strategy has been explored with good results, indicating the potential of both pair-wise, and multi-task transfers. A saliency map was used for the visualization of the key wavelength range captured by CNN with and without transfer learning. The results indicated that the wavelength range with and without transfer learning was overlapped to some extent. Overall, the results suggested that deep transfer learning methods could perform rice disease detection across different rice varieties. Hyperspectral imaging, in combination with the deep transfer learning method, is a promising possibility for the efficient and cost-saving field detection of rice diseases among different rice varieties.

Keywords: rice disease detection, hyperspectral imaging, fine-tuning, deep CORAL, deep domain confusion, saliency map

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Shangpeng Sun, McGill University, Canada
Yuxiong Wang, University of Illinois at Urbana-Champaign, United States

***Correspondence:**
Chu Zhang
chuzh@zjhu.edu.cn

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *11 April 2021*
Accepted: *26 August 2021*
Published: *29 September 2021*

Citation:
Feng L, Wu B, He Y and Zhang C (2021) Hyperspectral Imaging Combined With Deep Transfer Learning for Rice Disease Detection. Front. Plant Sci. 12:693521. doi: 10.3389/fpls.2021.693521

INTRODUCTION

Rice is one of the most vital crops for human beings and is critical for maintaining food supply. However, the growth of rice is subjected to stresses that are biological and abiotic in nature. Diseases are one of the major threats to rice, causing severe losses in quality and yield (Yang et al., 2019). There are various diseases threatening rice growth, such as bacterial leaf blight, blast, and sheath blight, which are the three major diseases of rice (Kumar et al., 2020; Molla et al., 2020). After being infected with these different diseases, the change in the inner chemical composition of rice varies from variety to variety, with external symptoms of rice leaf also varying.

To ensure the growth of rice, it is crucial to develop rapid and accurate detection methods. For the past decades, numerous rice disease detection methods have been proposed based on the inner and external changes in infected rice leaves (Awaludin et al., 2020; Lin et al., 2020; Shrivastava and Pradhan, 2020). Hyperspectral imaging (HSI) is one of the most studied detection approaches and is sensitive enough to capture the inner chemical difference between a healthy plant and an infected one (Abdulridha et al., 2020; Gao et al., 2020). Therefore, HSI was widely performed for rice disease detection at an early stage (Liu et al., 2018b; Zhu et al., 2019; Conrad et al., 2020).

However, there are still some problems in rice disease detection by HSI. Generally, the dataset of diseased samples is relatively small, which restricts the abilities of machine learning and deep learning methods (Zhao et al., 2018). Besides, the variation in rice varieties makes it challenging to develop a general model for disease detection. Commonly, a model established based on a variety of plants fails to obtain satisfactory results in another variety. In general, an individual model is constructed for each variety of a plant. In addition to the variation in variety, it is also hard to apply a model constructed with plants from a particular period or environmental condition to the same plants from another growth period or environmental condition (Conrad et al., 2020). The reason could be that spectral data of rice from different varieties or different growth periods have different feature spaces and data distributions. Besides, a model developed with data collected from a particular piece of equipment usually fail when applied to another because of differences in absorbance and wavelength shifts (Chen et al., 2020). Several methods were proposed to solve the problem, such as calibration transfer (Li et al., 2015; Liu et al., 2018a; Chen et al., 2020). Yan et al. applied a model built with a spectrometer to predict spectra collected from another with a method based on independent component analysis (Liu et al., 2018a). However, most research studies on calibration transfer methods have a requirement of a certain number of standard samples. Generally, the performance of a calibration transfer has a positive correlation with the number of standard samples (Chen et al., 2020). Overall, the problem of different data distribution (such as variety variation and equipment variation) has restricted the development of real-world applications.

In computer vision, distribution change or domain shift always exists because of many factors such as illumination, pose, background, and image quality (Gong et al., 2012). Recently, transfer learning methods have been used to solve this different kind of data distribution problem, which has been successfully applied to image classification, object detection, face recognition, semantic segmentation, etc. (Wang and Deng, 2018). Transfer learning (TL) is a method that tries to transfer the knowledge learned from one domain (called the source domain) to another different but related domain (called the target domain) (Yang, 2010). In general, transfer learning approaches can be categorized into four classes: instance-based TL, feature-based TL, parameter-based TL, and relational knowledge TL (Pan and Yang, 2010). The conception of these four kinds of TL was discussed in detail in a study by Pan (Pan and Yang, 2010). From another perspective, TL methods could be categorized into traditional and deep-learning-based methods (Tan et al., 2018; Wang and Deng, 2018). Wang et al. reviewed deep-learning-based TL methods in recent years (Wang and Deng, 2018). Many methods were proposed and tested on several standard domain adaption benchmarks in the field of computer vision, such as Office-31 (Saenko et al., 2010; Mingsheng et al., 2015) and Office-Caltech 10 (Gong et al., 2012; Mingsheng et al., 2015). Office-31 is the most used benchmark for transfer learning and consists of 4,652 images within 31 categories collected from three distinct backgrounds: Amazon (A), Webcam (W), and DLSR (D).

Several groups of researchers have introduced these TL methods into hyperspectral classification. Qiu et al. utilized two traditional transfer learning methods for detecting plastic pollution levels in different soil regions with a near-IR sensor (Qiu et al., 2020). Zhao et al. developed a multi-source deep transfer learning framework to classify hyperspectral images within the datasets of Indian Pines and Botswana (Zhao et al., 2020). Jiang et al. obtained satisfactory segmentation results of the hyperspectral image by combining TL and Markov random fields (Jiang et al., 2020). Compared with hyperspectral image classification, there were relatively fewer works focusing on classification based on spectra. In this study, the feasibility of deep transfer learning was investigated for rice disease classification with spectral data. Three deep transfer learning methods, namely, fine-tuning, deep CORrelation ALignment (deep CORAL) (Sun and Saenko, 2016), and deep domain confusion (DDC) (Tzeng et al., 2014), were introduced to address the problem stemming from the domain shift existing in different varieties of rice. The main contents of this survey are as follows: (1) a common CNN architecture was proposed and, respectively, trained with four different varieties of rice for disease-stressed rice classification; (2) the CNN trained on a specific variety of rice was directly applied to another three varieties of rice without fine-tuning (non-fine-tuning) and with fine-tuning; (3) since there are four domains (varieties of rice), deep CORAL methods were applied to all 12 shifts, taking one variety as the source domain and another variety as the target domain; (4) the DDC was applied to all 12 shifts and the performances of all deep transfer learning methods were compared; (5) the saliency map was used to intuitively visualize the key wavelength range captured by the CNN with and without transfer learning.

MATERIALS AND METHODS

Sample Preparation

To verify the effectiveness of the deep transfer learning methods among rice varieties, four varieties of rice were used in this study, namely, Zhongzheyou1 (which was provided by the China Rice Research Institute and Zhejiang Bedoknon Seed Co.), Jiuyou418 (which was selected and bred by the Northern Hybrid Japonica Rice Engineering Technology Center and Xuzhou Institute of Agricultural Science in Xuzhou, Jiangsu Province), Zhongzao39 (which was provided by the China Rice Research Institute), and Xiushui134 (which was jointly selected by the Jiaxing Academy of Agricultural Sciences, Yuyao Seed Management Station, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, and Zhejiang Jiaxing Crop High-Tech Breeding

TABLE 1 | Rice varieties and their corresponding labels and sample numbers.

Rice variety	Zhongzheyou1	Jiuyou418	Zhongzao39	Xiushui134
Encoding	01[a]	02	03	04
Total numbers	189	204	228	234
The training set	158	171	191	196
The validation set	16	17	19	19
The test set	15	16	18	19
CK[b]	45	51	60	54
RLB[c]	42	54	51	75
RB[d]	42	42	51	45
RSB[e]	60	57	66	60

[a] *01 represents the rice variety Zhongzheyou1; 02 represents the rice variety Jiuyou418; 03 represents the rice variety Zhongzao39; 04 represents the rice variety Xiushui134.*
[b] *CK represents total healthy samples.*
[c] *RLB represents total samples inoculated by rice leaf blight.*
[d] *RB represents total samples inoculated by rice blast.*
[e] *RSB represents total samples inoculated by rice sheath blast.*

Center). The four rice varieties are recorded as 01, 02, 03, and 04 for brevity, as shown in **Table 1**. A month after sowing the seeds into seed plots, the seedlings were transplanted into a laboratory greenhouse and fertilized and watered regularly.

To obtain inoculated samples, an *in vitro* inoculation method was applied. Rice blast and rice sheath blight are fungal diseases, while rice leaf blight is a bacterial disease. Therefore, the fungi of rice blast and rice sheath blight were cultured in a potato dextrose agar medium, and the bacteria of rice blight were cultured in conical flasks.

Healthy rice seedlings of the four varieties with similar growth were selected, washed, and transferred to sterilized plastic flat plates for disease inoculation. Approximately 20 plants from each variety of rice in the three- to five-leaf stage are used, and two to four leaves with visible symptoms per plant were used for analyses. To prevent the plants from drying, the roots were covered with distilled water-sterilized wipes. For rice blast and rice sheath blight inoculation, mycelial pellets were placed on the upper, middle, and lower parts of the leaves, with about three pellets per leaf. For rice leaf blight inoculation, a solution of bacteria was sprayed onto the leaf surface. After inoculation, the plates were covered with plastic wrap and then placed in a room with a temperature of about 26–28°C and a relative humidity of ∼80%, and healthy leaves were used as control. Leaves with visible symptoms were collected 4 days later.

Infected leaves cut from the plants were collected for hyperspectral image acquisition. The number of leaves used in this study is presented in **Table 1**. Three leaves were acquired in an image. In this study, the category value of the healthy samples (CK) was assigned as 0, and the category values of the samples inoculated with rice leaf blight (RLB), rice blast (RB), and rice sheath blight (RSB) were assigned as 1, 2, and 3, respectively. The representative images of healthy and disease samples are shown in **Supplementary Figure 1**. Regarding data splitting, the samples of each category were. randomly selected into the training set, the validation set and the testing set in a 10:1:1 ratio The number of plants from each rice variety in each category is listed in **Table 1**.

Hyperspectral Image Acquisition and Spectra Extraction

A visible/near-IR hyperspectral imaging system covering the spectral range of 379–1,024 nm was used to acquire hyperspectral images of healthy and infected leaves. The hyperspectral imaging system (as shown in **Supplementary Figure 2**) is formed by an imaging spectrograph (ImSpector V10E; Spectral Imaging Ltd., Oulu, Finland), a highly sensitive 8484-05G CCD camera (Hamamatsu, Hamamatsu City, Japan), and a long camera lens (OLES23; Specim, Spectral Imaging Ltd., Oulu, Finland). The illumination of the system is provided by 150-W tungsten halogen lamps (2,900 Lightsource; Illumination Technologies Inc., Liverpool, NY, United States). This hyperspectral imaging system conducts line scanning, and a moving plate driven by a stepper motor (Isuzu Optics Corp., Taiwan, China) is used to move the samples.

To acquire clear and non-deformable images, the distance between the camera lens and the moving plate, the exposure time of the camera, and the moving speed of the moving plate were adjusted to 13.7 cm, 0.17 s, and 0.7 ms/s, respectively. The acquired hyperspectral images were then corrected using the white reference image (acquired using a piece of pure white Teflon board with nearly 100% reflectance) and the dark reference image (acquired by covering the lens with a black lens cap with nearly 0% reflectance) according to the following equation:

$$I_C = \frac{I_R - I_D}{I_W - I_D} \quad (1)$$

where I_C is the corrected image, I_R is the raw image, I_W is the white reference image, and I_D is the dark reference image.

After image correction, each leaf was defined as a region of interest (ROI), and a wavelet transform (wavelet function: Daubechies 10; decomposition level: 3) was used to de-noise the pixel-wise spectra. The average spectrum of each leaf was calculated as a sample spectrum. The head and the tail of the

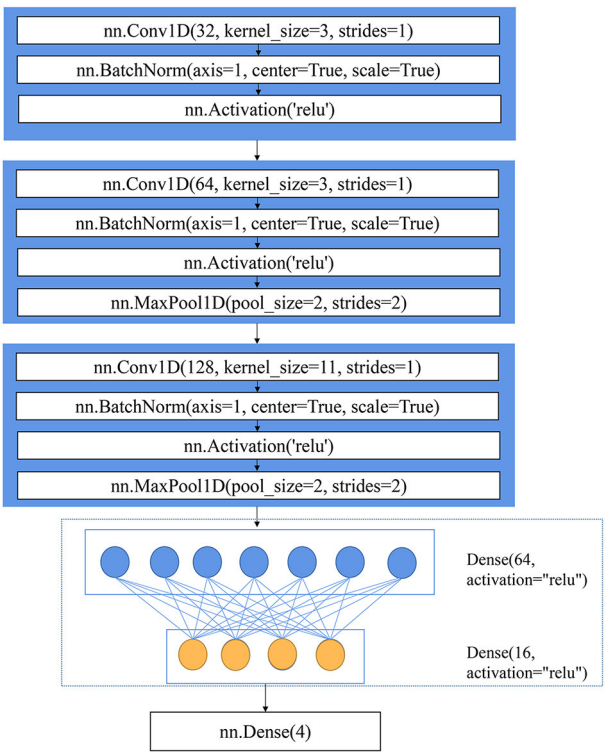

FIGURE 1 | The architecture of CNN.

spectra contained obvious noises and were then dropped, and the full spectra in the range of 448–947 nm (393 wavelengths in total) were used for analysis.

Model Establishment, Evaluation, and Software

A self-designed CNN architecture was developed for the classification task, as shown in **Figure 1**. The CNN consisted of three convolution layers, two fully connected layers, and an output layer. The structure (number of convolution layers/fully connected layers and kernel size) of the CNN in this manuscript was designed using a trial-and-error method. This one-dimensional CNN with different kernels performs convolution operations on the input data, thereby obtaining the global features of the data, while the pooling operation contributed to down-sampling the extracted features and reducing the amount of calculation (Zhong et al., 2021). The CNN could obtain different levels of features because of its stacked convolution layers. If the stride is less than or equal to the size of the convolution kernel, all spectral variables will participate in the convolution operation without losing some information. The stride was set to 1 so that the size of the output feature map of each convolution layer was consistent with the spectral variables. Spectral variables have rich information, and the extracted features are extremely subtle. The shallow semantics generated by the first two convolution layers should retain enough details, and the size of the convolution kernel should be appropriately small and be selected as 3. A large convolution kernel can retain more information in down-sampling. After down-sampling, the convolution kernel size increment was set to 11 (Cai et al., 2018). With the deepening of the CNN, the possibility of feature permutations and combinations increases, that is, the description of each key attribute should be more specific. Therefore, the number of channels should be increased to make the CNN more expressive and cover as many key attributes as possible. In this manuscript, the number of channels continued to double. The rectified linear unit (ReLU) was used as the activation function. The output of the CNN was followed by the softmax function to obtain the probabilities assigned to each class. Cross-entropy loss was used for the classification task. The function of "SoftmaxCrossEntropyLoss" provided by MXNET was used for the softmax operation and loss calculation. The last output layer of the CNN gave a four-value matrix for each sample; then, the prediction category (0, 1, 2, or 3, representing CK, RLB, RB, and RSB, respectively) was obtained according to the four values.

Before being fed into the CNN, the full spectra (393 wavelengths in total) of each class were further implemented with the standardization process. This standardization preprocessing method standardized each sample of the training set by removing the mean and scaling to unit variance, with the same standardization being performed on the test set by a utility class scale in scikit-learn. After standardization preprocessing, the shape of data was (number of samples, number of wavelengths). In order to feed into CNN, the shape of data needs to be reshaped to be (number of samples, 1, number of wavelengths).

To conduct fine-tuning, the CNN was trained on each variety of rice to obtain a decent discriminative performance. Next, the pre-trained parameters of the first three convolution blocks were fronzen and transferred to the CNN to predict another varietry of rice. Fine-tuning assumes that the pre-trained parameters contain the knowledge learned from the source domain and that this knowledge will be equally applicable to the target domain. Since there were four rice varieties, we conducted experiments on all the 12 shifts: 01→02, 01→03, 01→04, 02→01, 02→03, 02→04, 03→01, 03→02, 03→04, 04→01, 04→02, and 04→03.

The spectra extraction of HSI was conducted on Matlab R2019b (MathWorks, Natick, MA, United States). To evaluate the performance of the model, classification accuracy was used, which was the ratio of the correctly classified number of samples to the total sample number. Deep learning was conducted using Python 3 with an MXNET framework (Amazon; Seattle, WA, United States) with GPU acceleration. A computer with an Intel core-i7 8700k CPU, 16 GB of RAM, an NVidia GeForce a GTX1660 GPU (8GB RAM, CUDA cores 1408, CUDA version 9.2.148), and a 256 GB SSD was used for calculation.

Transfer Learning Methods

Fine-Tuning

Fine-tuning is a common technique in transfer learning that is widely used in computer vision and natural language processes. It can migrate the knowledge of a pre-trained network based on the source dataset to the target dataset (Oquab et al., 2014). Considerable publications underline the benefits of pre-training deep networks on large datasets (Käding et al., 2016; Zhang

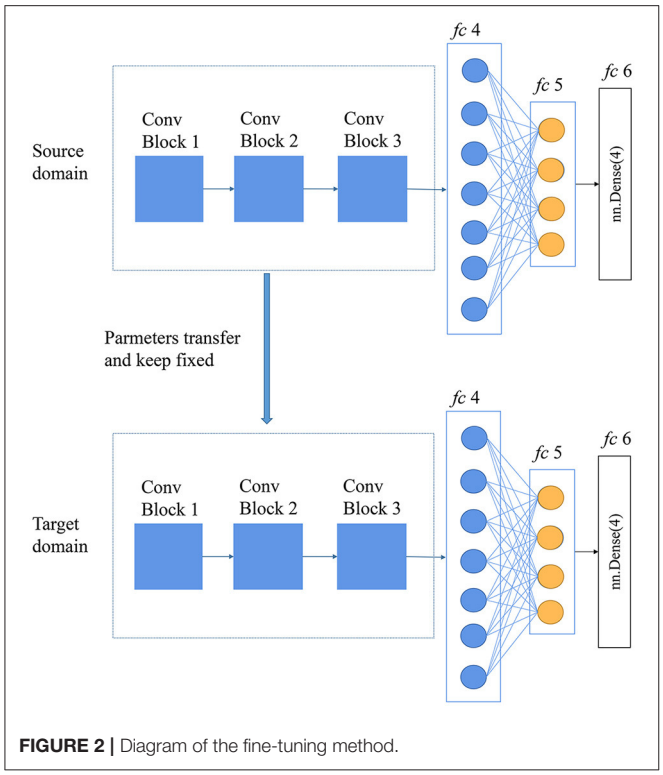

FIGURE 2 | Diagram of the fine-tuning method.

et al., 2019). To conduct fine-tuning, the network is first trained on the source dataset, and then, pre-trained parameters are transferred to the target task and kept fixed, with only a few layers (commonly the last few layers) trained on the target dataset (Oquab et al., 2014). In this study, the number of the frozen layers was determined using the "trial and error" method. All layers of the pre-trained CNN are frozen, except for the last three fully connected layers, as shown in **Figure 2**.

Deep CORAL

CORrelation Alignment is a feature-based transfer learning method proposed by Sun et al. (2016) that minimizes the distance between the source domain and the target domain by aligning the second-order statics of source and target distributions (Sun et al., 2016). Furthermore, Sun and Saenko (2016) incorporated CORAL into deep neural networks by constructing a loss function that minimized the difference between the source and target correlations, which is named CORAL loss. CORAL loss is defined as the distance between the second-order statistics (covariances) of the source and target features, the crucial formula of which is as follows (Sun and Saenko, 2016):

$$l_{\text{CORAL}} = \frac{1}{4\text{d}^2} \left\| C_S - C_T \right\|_F^2 \tag{2}$$

where $\|\bullet\|_F^2$ represents the squared matrix Frobenius norm; C_S and C_T represent the feature covariance matrices of the source domain and the target domain, respectively; d is the feature dimension in a specific layer in a neural network.

For generalization and simplicity, in this study, we apply CORAL loss to the *fc6* layer of the self-designed CNN, shown in **Figure 3**. In the training phase, the batch size was set as 40, and the base learning rate was 0.0001. The weight of the CORAL loss (λ) was set to 0.01 at the initial stage and then expanded to 0.1 and 1. Adding CORAL loss was of help to learn the feature representation that was discriminative and that minimized the distance between the source and the target domain.

Deep Domain Confusion

Directly training a classifier using only the source data often leads to inferior performance in the target domain (Tzeng et al., 2014). Deep domain confusion uses domain confusion loss based on the maximum mean discrepancy (MMD) to automatically learn a feature representation jointly trained to optimize for classification and domain invariance (Tzeng et al., 2014). Maximum mean discrepancy is a standard distribution distance between the embeddings of the probability distributions in a reproducing kernel Hilbert space (Gretton et al., 2012; Ghifary et al., 2014). The difference and relationship between CORAL and MMD were discussed in this study (Sun et al., 2016). The key point of DDC is to learn the feature representation that minimizes the distance between domains and is conducive to training a strong classifier at the same time. The approach to meet both these criterias (minimizing distance between domains and training a strong classifier) is to minimize the loss:

$$L = L_C\left(X_S, y\right) + \lambda MMD^2\left(X_S, X_T\right) \tag{3}$$

where $L_C\left(X_S, y\right)$ represents classification loss on the source domain and $MMD^2(X_S, X_T)$ represents the distance between the source data, X_S, and the target data, X_T. The hyperparameter λ determines the weight of distance loss.

In this study, the same CNN architecture was used in deep CORAL and DDC. Maximum mean discrepancy loss was calculated with the features of the last layer (*fc6*), which is shown in **Figure 4**. In the training phase, the batch size was set to 40, the base learning rate was 0.0001, and the weight of the MMD loss (λ) was set to 0.001 at the initial stage. Adding MMD loss into total loss was of help to learning the feature representation that was discriminative and minimized the distance between the source and the target domain.

Furthermore, the relationship between epochs and the accuracy of the training set of the target domain with deep CORAL and DDC are provided in Supplementary **Figure 3**, which illustrates the change in training accuracy and train loss as the change in epochs. The training process took about 286.85 s per 1,000 epochs.

RESULTS

Spectral Profile

To illustrate the difference among rice plants under different disease stress conditions, the average original spectra were plotted

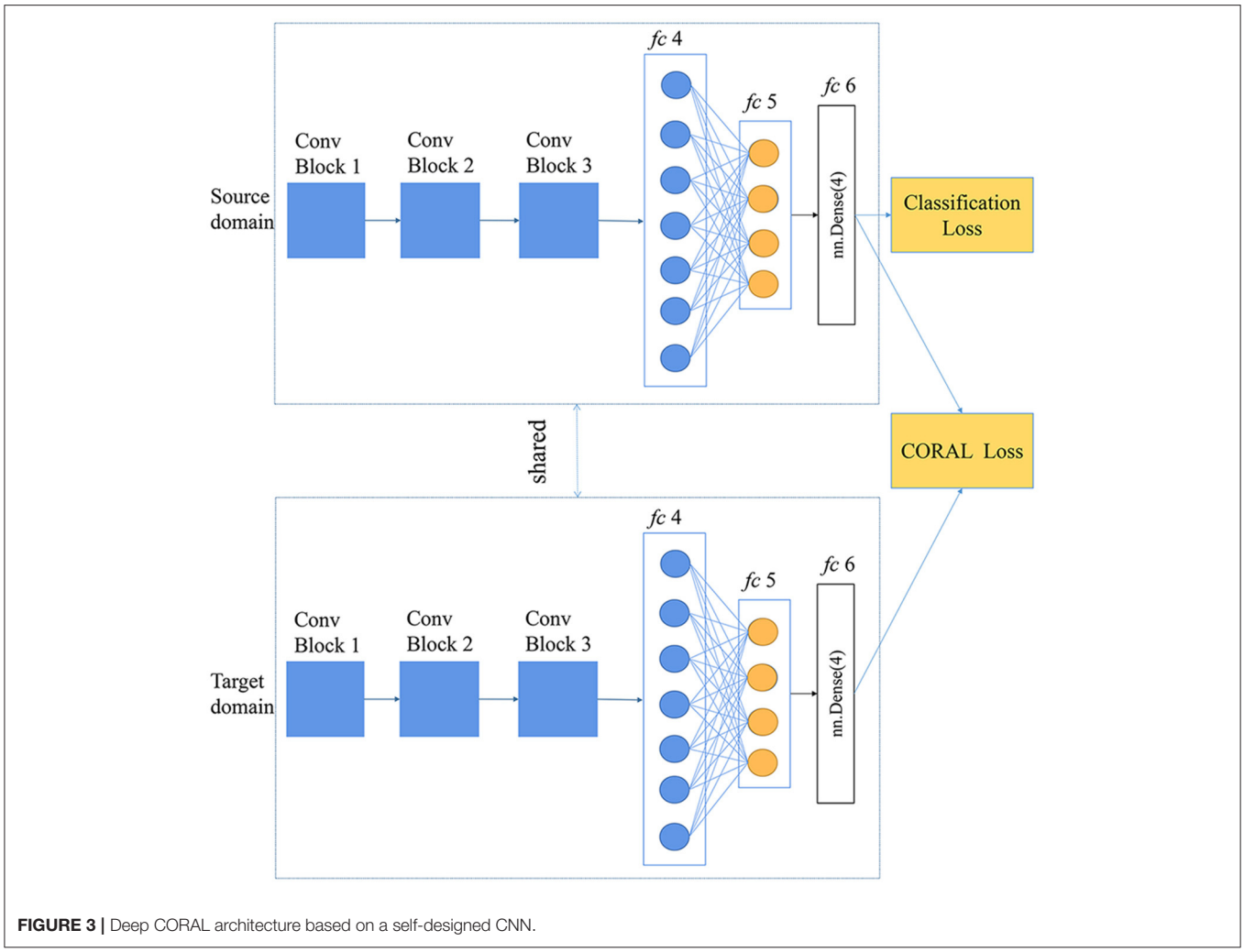

FIGURE 3 | Deep CORAL architecture based on a self-designed CNN.

for visualization. In **Figures 5A–D** present healthy and disease-stressed leaves of the four rice cultivars, corresponding to rice varieties 01, 02, 03, and 04, respectively. The change tendency of these four varieties was similar to the spectral profile of the other two varieties of rice in the previous study (Feng et al., 2020).

The change tendencies of the rice leaves under different disease stress conditions were similar. The distinct difference could be observed in some ranges of wavelengths, including the range from 450 to 500, 580 to 680, and 720 to 940 nm. The difference among each class is a foundation for developing a classifier for rice disease detection. Besides, the distinction among different cultivars suggested that a discrepancy exists among different cultivars of rice under the same disease stress conditions. To perform a quantitative analysis for disease detection in rice, further processing should be applied to the spectra.

Classification Models on Each Variety of Rice

The common CNN architecture was trained on four varieties of rice. The classification results are shown in **Table 2**. All the CNN models obtained an accuracy of over 97% for the training set and an accuracy of over 93% and over 87% for the validation set and the test set, respectively. These results indicated that the self-designed CNN was capable of extracting the features for discrimination of rice under different disease stress conditions. This provides the possibility for transferring the learned knowledge across different varieties.

Results With and Without Fine-Tuning

To verify the effectiveness of fine-tuning, the pre-trained CNN based on one rice variety was directly used to predict another rice variety without fine-tuning. The results of the prediction without fine-tuning are shown in **Table 3**.

Without fine-tuning, when 01 was the source domain, the accuracy of the test set of the target domains was lower than 69%, except for 03. When the CNN based on 02 was used to predict 04, it only obtained an accuracy of no more than 52% on both the training set and the test set. When 02 was transferred to 01 and 03, it also obtained unsatisfactory classification results. Besides, deterioration in classification results also occurred when CNNs

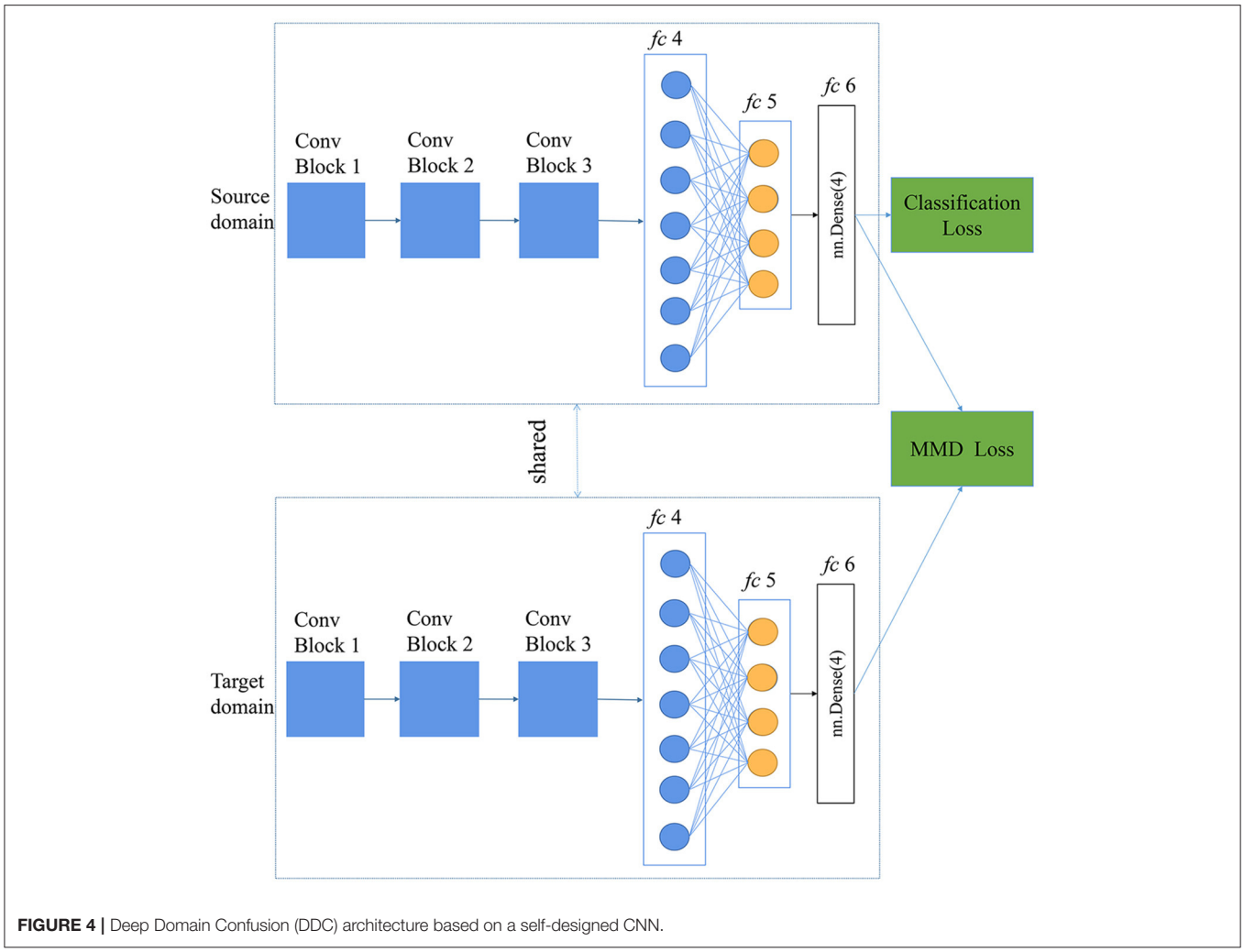

FIGURE 4 | Deep Domain Confusion (DDC) architecture based on a self-designed CNN.

trained on 03 were directly applied to predict 01, 02, and 04. This situation also happened when the source domain was 04.

After fine-tuning, the accuracy of the training set increased to more than 92%, while the accuracy of the validation set and the test set, respectively, increased to over 88 and 86%, with 10 out of the 12 transfer tasks obtaining an accuracy of over 87% on the test set. Compared with the results of the prediction without fine-tuning, the accuracy of the training set improved by 23–56.73%, while that of the test set improved by 18% on 10 out of the 12 transfer tasks. Besides, the accuracy after fine-tuning was close to the results of the CNN trained on each variety of rice, which suggested that fine-tuning was capable of transferring the feature representations learned from one variety of rice to another.

To further verify the effectiveness of fine-tuning, we also conducted fine-tuning with a smaller dataset of the target domain. The smaller training set only consisted of 20 samples of each class, and the validation set and the test set were kept the same. The training process was the same. As shown in **Table 3**, with a smaller dataset, the accuracy of the test set on nine out of all the transfer tasks is equal to or higher than that with the bigger dataset. Even though the accuracy of the test set of three transfer tasks declined, the amplitude of decline did not exceed 5.56%. The results indicated that fine-tuning could transfer the knowledge among different domains and that fine-tuning obtained satisfying results with a relatively small training set in the target domain.

Results With Deep CORAL and DDC

The results of transfer learning with deep CORAL are listed in **Table 4**. Eight transfer tasks have achieved an accuracy of over 75% on the test set of the target domain. Compared with the results of non-fine-tuning, the accuracy for the test set of the target domain obtained improvements of over 15% on 8 out of the 12 transfer tasks. Furthermore, five of all the transfer tasks obtained an improvement of over 25%. It indicated that CORAL loss contributed to learning features that work well on the target domain. As for transfer task 03→ 01, the accuracy of the validation set and the test set of the target domain was 93.75 and 86.67%, respectively, which was slightly lower than the accuracy (93.33%) of the fine-tuning method.

Regarding DDC, the overall performance was inferior to the performance of deep CORAL. However, compared with the

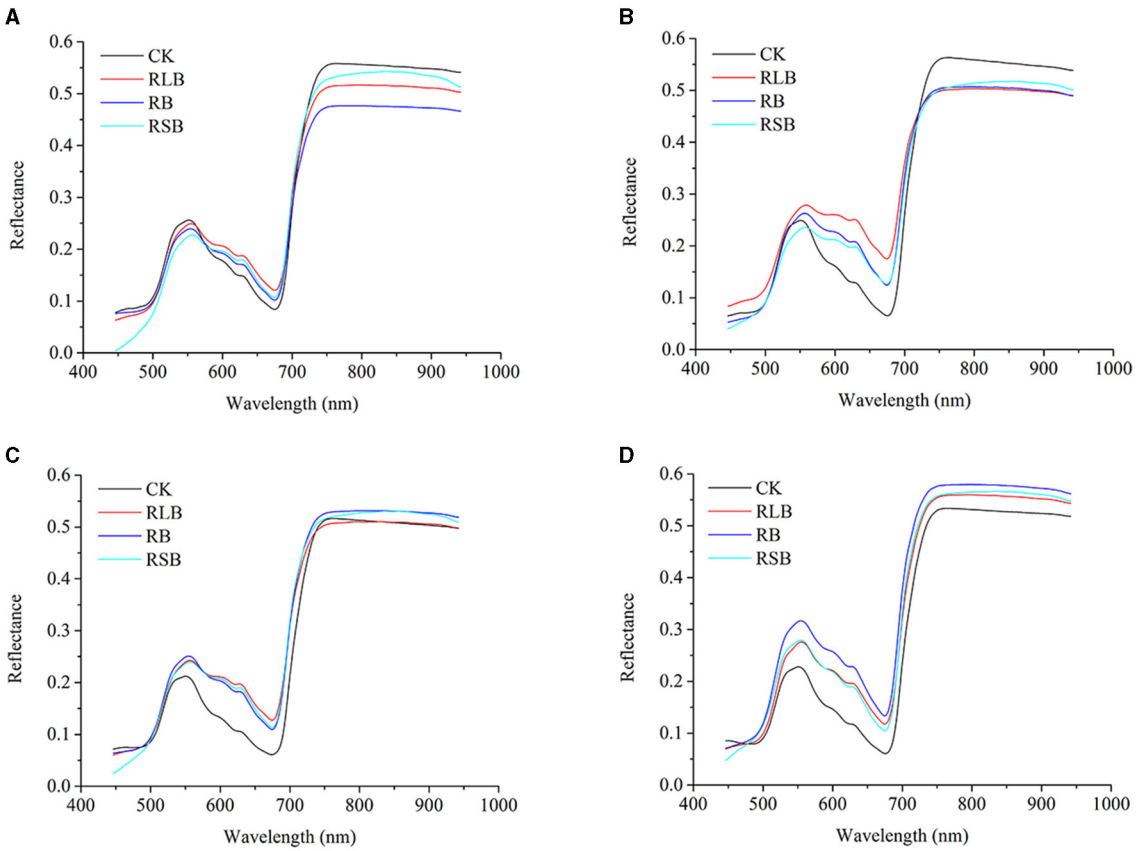

FIGURE 5 | (A–D) represent the reflectance of rice varieties 01 (Zhongzheyou1), 02 (Jiuyou418), 03 (Zhongzao39), and 04 (Xiushui134).

TABLE 2 | Classification results of the four varieties of rice.

Variety	Tr[a](%)	Val[b](%)	Te[c](%)
01	100	93.75	100
02	97.08	94.12	87.5
03	100	94.74	94.44
04	100	100	94.74

[a]*Tr is the training set;* [b]*Val is the validation set;* [c]*Te is the test set.*

results with no fine-tuning, the accuracy for the test set of the target domain was improved by a rate over 10% on 8 out of 12 transfer tasks. Therein, three of the transfer tasks obtained an improvement of over 30%. The improvements suggested that MMD loss was of help in reducing the distance between the source domain and the target domain.

To investigate the choices of the layer, the *fc5* layer was adapted to compute domain loss in both deep CORAL and DDC. The results are listed in **Table 5**. Regarding deep CORAL, 4 out of all the 12 transfer tasks obtained an accuracy of over 75% on the test set of the target domain. In general, the accuracy of the target domain of each transfer task was slightly lower than their corresponding accuracy with the *fc6* as the transfer layer. In terms of DDC, six transfer tasks obtained an accuracy of over 73% on the test set of the target domain, while there was the case for seven transfer tasks with the *fc6* as the transfer layer. Overall, the *fc6* layer was more conducive for transferring the knowledge learned from one cultivar of rice to another for rice disease detection.

Furthermore, the saliency map was applied to visualize the most informative wavelengths captured by the CNN model.

Saliency mapping is a technique for visualizing class models, and it is widely studied in computer vision. The method will numerically generate an image according to a learned classification network and a class of interest. The generated image is representative of the class in terms of the class scoring model (Simonyan et al., 2013). In computer vision, the magnitude of the class score defines the importance of the corresponding pixels for the class (Simonyan et al., 2013). We introduced this method into hyperspectral imaging analysis and the visualization of key wavelengths captured by a learned classification model. The interpretation of computing the class saliency using the class score derivative is that the magnitude of the derivative indicates which wavelength needs to be changed the least to affect the class score the most. To realize the visualization of the critical wavelength range, the class score was first computed by calculating the derivative (gradient) of the correctly classified class. Next, the wavelengths of each correctly classified sample

TABLE 3 | Classification results without and with fine-tuning.

S-T[a]	Non-fine-tuning			Fine-tuning			Finetuning with a smaller set[e]		
	Tr[b](%)	Val[c](%)	Te[d](%)	Tr(%)	Val(%)	Te(%)	Tr(%)	Val(%)	Te(%)
01–02	60.23	52.94	68.75	98.32	**88.24**	**87.50**	96.25	**88.24**	**87.50**
01–03	76.44	84.21	88.89	100.00	**89.47**	**88.89**	95.00	**89.47**	**88.89**
01–04	52.04	73.68	57.89	92.35	**89.47**	**89.47**	91.25	**89.47**	**89.47**
02–01	62.03	56.25	66.67	99.36	**93.75**	**86.66**	96.25	**93.75**	**86.67**
02–03	67.02	89.47	61.11	91.62	**89.47**	**88.89**	100.00	89.43	83.33
02–04	53.57	42.11	52.63	96.43	**94.74**	**89.47**	91.25	89.48	78.95
03–01	70.89	75.00	60.00	94.30	**93.75**	**93.33**	95.00	**93.75**	**93.34**
03–02	63.74	47.06	50.00	98.83	**88.24**	**87.50**	95.00	**88.23**	**87.50**
03–04	58.67	78.95	52.63	97.96	**100.00**	**89.47**	100.00	94.74	84.21
04–01	51.27	50.00	53.33	97.47	**93.75**	**86.67**	98.75	**87.50**	**86.67**
04–02	43.27	47.06	37.50	100.00	**88.24**	**87.50**	98.75	**88.24**	**87.50**
04–03	72.77	57.89	77.78	97.91	**100.00**	**88.89**	100.00	**94.74**	**88.89**

*[a]S-T means the source domain-the target domain; [b]Tr means the training set of the target domain; [c]Val means the validation set of the target domain; [d]Te means the test set of the target domain; [e]a small set only consisted of 20 samples of each class; the test set remained the same as in **Table 1**. The bold values means the accuracy rate of the test set was beyond 86%.*

TABLE 4 | Classification results of deep CORrelation ALignment (CORAL) and deep domain confusion (DDC) (*fc6*[a]).

S-T[b]	Deep CORAL						DDC					
	S[c]			T[d]			S			T		
	Tr[e](%)	Val[f](%)	Te[g](%)	Tr(%)	Val(%)	Te(%)	Tr(%)	Val(%)	Te(%)	Tr(%)	Val(%)	Te(%)
01–02	100.00	87.50	80.00	61.99	58.82	62.50	100.00	87.50	86.67	56.14	58.82	56.25
01–03	100.00	87.50	86.67	80.63	**84.21**	**83.33**	100.00	87.50	86.67	80.10	**84.21**	**83.33**
01–04	100.00	87.50	86.67	69.90	73.68	73.68	100.00	87.50	80.00	69.90	73.68	73.68
02–01	100.00	82.35	75.00	79.11	87.50	73.33	100.00	82.35	75.00	74.05	**81.25**	**80.00**
02–03	100.00	76.47	75.00	79.58	**78.95**	**77.78**	100.00	76.47	75.00	75.39	78.95	72.22
02–04	100.00	82.35	81.25	68.88	73.68	68.42	100.00	88.24	75.00	72.96	73.68	52.63
03–01	100.00	94.74	83.33	86.08	**87.50**	**86.67**	100.00	89.47	72.22	82.28	87.50	73.33
03–02	100.00	94.74	88.89	77.78	**76.47**	**75.00**	100.00	89.47	77.78	76.61	70.59	62.50
03–04	100.00	94.74	88.89	75.51	**84.21**	**84.21**	100.00	84.21	83.33	71.43	78.95	73.68
04–01	100.00	94.74	94.74	77.22	**81.25**	**80.00**	100.00	94.74	94.74	72.78	68.75	73.33
04–02	100.00	94.74	94.74	73.68	**76.47**	**75.00**	100.00	84.21	89.47	69.59	70.59	68.75
04–03	100.00	89.47	100.00	82.72	**84.21**	**77.78**	100.00	84.21	94.74	78.01	78.95	**77.78**

*[a]fc6 means that the last fully connected layer of the CNN was used for calculating domain loss (CORAL loss and maximum mean discrepancy, MMD, loss, as shown in **Figures 3** and **4**); [b]S-T means the source domain-the target domain; [c]S means the source domain; [d]T means the target domain; [e]Tr means the training set of the target domain; [f]Val means the validation set of the target domain; [g]Te means the test set of the target domain. The bold values means the accuracy rate of the test set was beyond 75%.*

were sorted by the absolute value of the corresponding gradient in descending order. Finally, the first 50 critical wavelengths of each sample of the same class were selected and counted according to the frequency of wavelengths. After obtaining the critical wavelengths and their corresponding frequencies, the wavelength saliency map was plotted for visualization. Here, the results of 01 were used for visualization. Since there were only 15 samples of the test set of 01, there were too few to visualize the critical wavelengths. The validation set and the test set of 01 were concatenated (called the combined set) and used for prediction with the CNN model in **Table 2** and the deep CORAL in **Table 4**. Then, the results were used for the visualization of the critical wavelength range. On the one hand, the combined set results of the CNN model built on the data of 01 alone (the prediction accuracy of the combined set was 96.77%) were used to visualize the critical wavelength range without transfer learning. On the other hand, the combined set results of 01 after achieving the transfer task 03→ 01 (the prediction accuracy of the combined set was 90.32%) with deep CORAL were used to visualize the key wavelengths after transfer learning.

Figure 6 shows the differences and connections between the spectral profile and the saliency map. In **Figures 6A–D** represent the frequencies of the critical wavelengths of classes 0, 1, 2, and 3, respectively. Without transfer learning, the wavelengths

TABLE 5 | Classification results of Deep CORAL and DDC (*fc5*[a]).

S-T[b]	Deep CORAL						DDC					
	S[c]			T[d]			S			T		
	Tr[e](%)	Val[f](%)	Te[g](%)	Tr(%)	Val(%)	Te(%)	Tr(%)	Val(%)	Te(%)	Tr(%)	Val(%)	Te(%)
01–02	100.00	87.50	80.00	58.48	58.24	56.25	100.00	87.50	86.67	60.23	58.82	56.25
01–03	100.00	93.75	86.67	80.63	**84.21**	**83.33**	100.00	87.50	86.67	80.10	**84.21**	**83.33**
01–04	100.00	87.50	93.34	65.82	**73.68**	**73.68**	100.00	81.25	86.67	70.41	**73.68**	**73.68**
02–01	100.00	82.35	75.00	75.95	**81.25**	**73.33**	92.40	82.35	68.75	71.52	**75.00**	**73.33**
02–03	100.00	76.47	75.00	76.44	**77.78**	**78.95**	100.00	76.47	75.00	78.53	73.68	72.22
02–04	100.00	82.35	75.00	72.45	68.42	68.42	95.32	82.35	75.00	71.43	73.68	57.89
03–01	100.00	89.47	83.33	82.91	**87.50**	**80.00**	100.00	89.47	77.78	82.91	**87.50**	**73.33**
03–02	100.00	89.47	83.33	74.85	70.59	68.75	100.00	94.74	83.33	73.68	76.47	62.50
03–04	100.00	94.74	88.89	84.69	**89.47**	**73.68**	100.00	94.74	88.89	73.98	**78.95**	**73.68**
04–01	100.00	89.47	94.74	71.52	73.33	68.75	100.00	100.00	94.74	69.62	68.75	66.67
04–02	100.00	94.74	94.74	74.85	68.75	58.82	100.00	84.22	94.74	66.08	64.71	62.50
04–03	100.00	89.47	89.47	83.77	**84.21**	**77.78**	100.00	100.00	89.47	82.20	**78.95**	**77.78**

[a]*fc5 means that the penultimate fully connected layer of the convolutional neural network (CNN) was used for calculating domain loss (CORAL loss and MMD loss, as shown in* ***Figures 3*** *and* ***4****);* [b]*S-T means the source domain-the target domain;* [c]*S means the source domain;* [d]*T means the target domain;* [e]*Tr means the training set of the target domain;* [f]*Val means the validation set of the target domain;* [g]*Te means the test set of the target domain. The bold values means the accuracy rate of the test set was beyond 73%.*

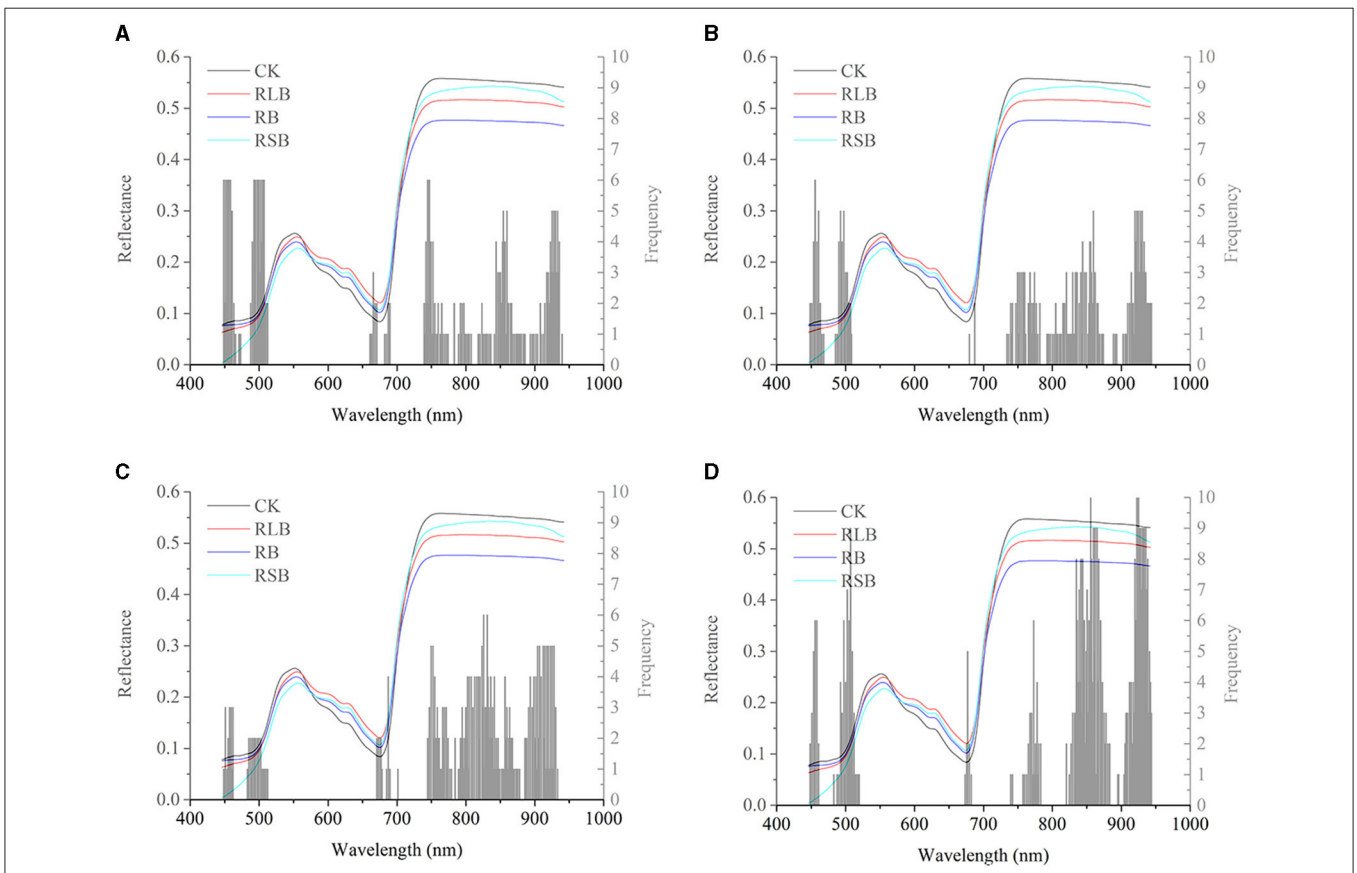

FIGURE 6 | Saliency map before transfer learning regarding the rice variety 01 (Zhongzheyou1). **(A–D)** represent the key wavelength ranges captured by CNN to correctly classify rice diseases with classes 0 (healthy samples, CK), 1 (rice leaf blight, RLB), 2 (rice blast, RB), and 3 (rice sheath blight, RSB), respectively.

of interest are located in almost the same ranges, including the range from 448 to 467 nm, the range from 485 to 510 nm, the range from 660 to 690 nm, and the range from 740 to 940 nm. Compared with the wavelength range discussed in section Spectral profile spectral profile, there were several common wavelength ranges, including the range from 448 to 467 nm, the range from 485 to 510 nm, and the range from 740 to 940 nm. It indicated that the result of the saliency map was largely consistent with the spectral profile.

In **Figure 7**, the wavelengths after transfer learning were located in a similar range as they were without transfer learning, including the ranges from 450 to 500 nm, the range from 518 to 525 nm, the range from 535 to 540 nm, the range from 600 to 750 nm, and the range from 760 to 910 nm. Overall, the range of feature wavelengths captured by the CNN with transfer learning had a significant overlap with the wavelength range captured by the CNN without transfer learning.

The intersection of the key wavelengths with and without transfer learning suggested that deep CORAL was able to learn the features that work well on both the source domain and the target domain. Furthermore, CORAL loss, as a part of a total loss, contributed to the reduction in the distance between the source domain and the target domain. Transfer learning also had the capability to learn knowledge that could be applied to a different domain.

DISCUSSION

Transfer learning is able to address the problem that a predictive model based on specific data cannot work well on another data source from a related domain but under different distribution conditions. According to the saliency map shown in **Figures 6**, **7**, it appears that the critical wavelength range is similar in cases with and without transfer learning. In this study, three different deep transfer learning methods were applied for disease detection among different cultivars of rice. The results showed that deep transfer learning methods could perform disease detection across different rice cultivars in an efficient manner. Among the methods used in this study, fine-tuning was an easily operative and effective transfer learning method, obtaining the best overall performance after transfer learning. It was easy to operate and obtain the accuracy for the target domain, which was equal to or just slightly lower than the accuracy of the CNN directly built on the corresponding target domain. Besides, deep CORAL, as a feature-based transfer learning method, has the ability to learn the feature representation that works well on both the source domain and the target domain. Overall, compared with the result of DDC, deep CORAL performed better. In the study of Sun et al. (2016), the CORAL-based method was also shown to be superior to the MMD-based method. The reason could be that CORAL, as an asymmetric transformation, tries to "bridge" the two domains, while MMD as a symmetric transformation tries to find a space that ignores the difference between the source and the target domain (Sun et al., 2016). Although deep CORAL has achieved good predictions on both the source domain and the target domain among most of the transfer tasks, the accuracy was not as good as the results of direct modeling and fine-tuning. This situation also existed in other studies. Some traditional transfer learning methods obtained an average accuracy lower than 60% on the standard Office dataset, such as geodesic flow kernel (GFK) (Gong et al., 2012) and subspace alignment (SA) (Fernando et al., 2013). The CORAL method obtained an accuracy lower than 65% on four of the transfer tasks and an average of 69.4% for all the six transfer tasks with the same dataset (Sun et al., 2016). Furthermore, deep CORAL obtained an average accuracy of 72.1% on the same dataset. Long et al. improved the average accuracy to 72.9% on the same dataset with deep adaption networks (Mingsheng et al., 2015).

To achieve better performance with transfer learning, a deep network should be carefully designed. The deep network should obtain good predictions based on the data of the source domain. Thus, the deep network allows the extraction of feature representations that are conducive to achieving classification. Given that the resnet architecture is one of the most excellent deep neural network architectures, deep transfer learning based on resnet was carried out for a comparison with the results based on the customized CNN architecture. To avoid overfitting, ResNet14 (the architecture is provided in **Supplementary Table 1**) was used for transfer learning, and the global average pooling layer was set as the transfer layer. The results based on ResNet14 are provided in **Supplementary Table 2**. Comparing with the results shown in **Table 4**, the results based on ResNet14 were similar or inferior to the results based on the customized CNN. Therefore, a well-designed CNN architecture was a prerequisite for a good transfer effect. In this case, since the spectral data were one-dimensional, a relatively shallow CNN architecture was suitable for the classification task.

In addition, the pair-wise transfer was widely studied in the field of computer vision, and it was shown to be effective for the classification of rice diseases in the above sections. Moreover, it was worth investigating whether a multi-task transfer strategy could achieve better performance. The multi-task transfer strategy could potentially improve the transfer performance for this strategy by increasing and enriching the training data. Considering the large number of transfer tasks, only the case of joint training of three rice transfers to one rice variety with deep CORAL was explored. Thus, the multi-task transfer strategy was implemented based on the self-designed CNN in section Material and Methods with *fc*6 as the transfer layer, which first jointly trained on three rice varieties (e.g., 02, 03, and 04) and then transferred to the fourth one (e.g., variety 01). The results of the multi-task transfer are provided in **Supplementary Table 3**. With the deep CORAL method, the accuracy of the test set of the target domain was improved from 86.67 to 93.33% compared with the result (03–01) shown in **Table 4**. In addition, the accuracy of the test set of the target domain was improved from 83.33 to 88.89% compared with the result (01–03) of deep CORAL shown in **Table 4**. A similar improvement occurred when the target domain was 02 or 04. Therefore, the multi-task transfer strategy could contribute to obtaining better performance, which alleviated the challenge of

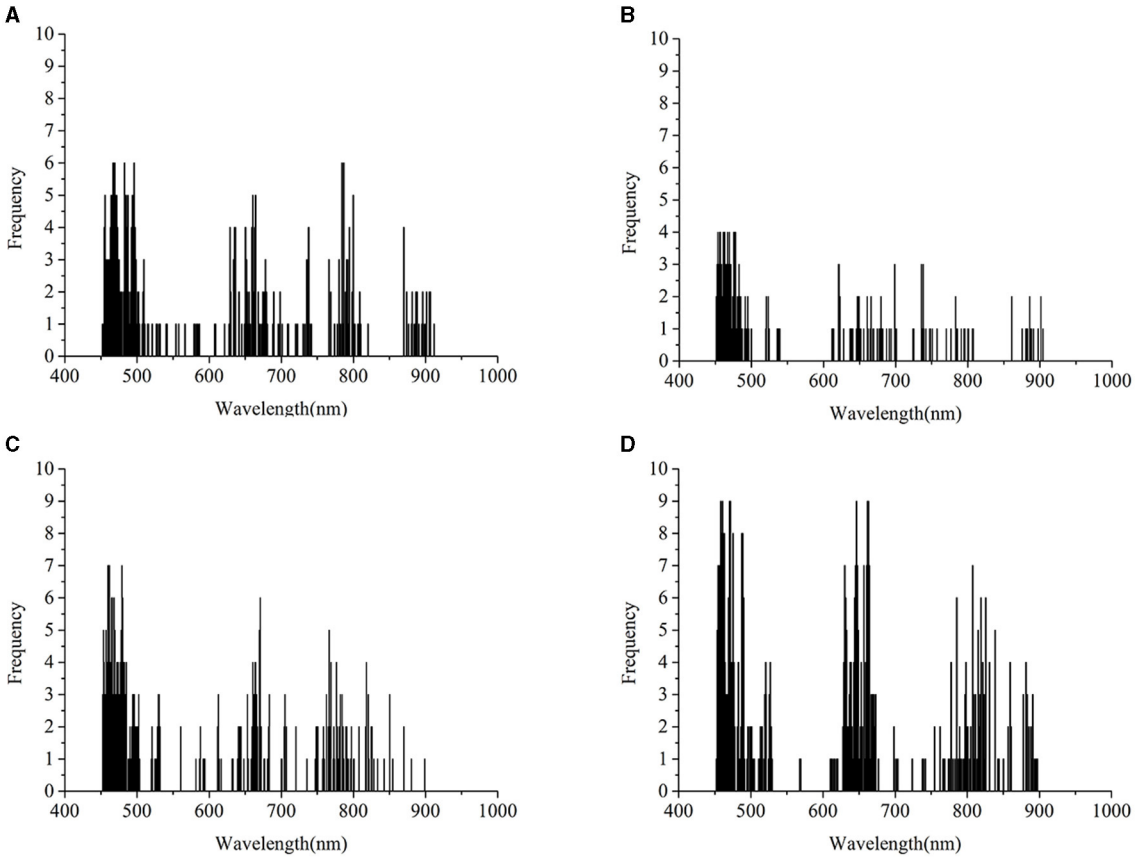

FIGURE 7 | Saliency map after transfer learning regarding the rice variety 01 (Zhongzheyou1), which is based on the results of 01 after achieving the transfer task 03→ 01 with deep CORAL. **(A–D)** Represent the key wavelength ranges captured by CNN to correctly classify rice diseases with classes 0 (healthy samples, CK), 1 (rice leaf blight, RLB), 2 (rice blast, RB), and 3 (rice sheath blight, RSB), respectively.

transfer learning with a limited amount of data. Furthermore, it is worth studying which varieties are selected and how they are selected for joint transfer when there are a large number of varieties of rice. Besides, it contributes to confusing domains by adding domain loss both in deep CORAL and DDC. According to some related studies (Tzeng et al., 2014; Sun and Saenko, 2016; Sun et al., 2016), the last few fully connected layers were usually applied for calculating domain loss. In this study, the choices for the layer for calculating domain loss were studied. The results showed that there was a difference in using a different layer for calculating domain loss. Therefore, this observation should be taken into account in transfer learning. Generally, CORAL and MMD are kernel tricks that try to measure the distance between the source domain and the target domain in another space. To improve the transferability of transfer learning in spectral analyses, some kernel tricks could be introduced and combined into the spectral data metric. In addition, Long et al. proposed a multiple kernel variant of MMD to realize transfer learning, which combined with the last three fully connected layers for domain loss. This method was beyond discussion in this study and could be investigated in future studies. Considering the balance of classification loss and domain loss, the hyperparameter λ played a key role in this aspect, which also causes difficulty for training the network in transfer learning to some extent. Some previous studies on computer vision have tended to set it to 1 according to previous experience (Mingsheng et al., 2015; Sun and Saenko, 2016), while some set λ to 0.25 (Tzeng et al., 2014). According to the experience of the authors, the setting of λ to 0.01 was better at the first stage of training, and then it was expanded to 0.1 and 1.

Considering practical application, among the previous studies that focused on rice disease detection with a spectroscopy technique, it was hard to find a network trained on a specific variety that could be directly applied to another variety. However, it will cost much time and energy to build and train a deep network for each cultivar of rice. Transfer learning is a promising tool for solving this problem. Transfer learning can learn feature representations that work well on different but related tasks. Since deep learning is good at automatically extracting features of different levels, it is quite desirable for combining deep learning networks and transfer learning together to realize transferability across different tasks. This study showed the feasibility of combining deep transfer learning with spectra data for rice disease detection. In future studies, more samples need to be

used to further improve performance. If a high-quality spectral database like ImageNet could be established and maintained, it will strongly promote the development of transfer learning with spectral analyses and contribute to the development of practical applications.

CONCLUSIONS

In this study, hyperspectral imaging was performed to acquire the information of four cultivars of rice under different disease stress conditions. Deep transfer learning was introduced for the first time to rice disease detection across different rice cultivars simultaneously. A self-designed CNN architecture was developed as a classification model and basic network of deep transfer learning. The transfer learning methods used in this study were fine-tuning, deep CORAL, and DDC. The results illustrated that the fine-tuning method was a relatively easy and efficient solution for rice disease detection across different rice cultivars. Deep CORAL was capable of transferring the knowledge learned from a specific variety to another variety and was superior to DDC in overall performance. In addition, when jointly training on three varieties of rice and then transferring to the fourth one, the accuracy of the target domains improved. This indicated that the multi-task transfer strategy could improve transfer performance, which increased and enriched the training data.

Nevertheless, there existed limitations in this study. Rice leaves were collected and explored from only four different cultivars. Rice samples of more cultivars were suggested to be collected to study the universality of deep transfer learning methods. The number of samples of each cultivar was no more than 250. The small size of a dataset may restrict the performance of deep learning methods and transfer learning combined with deep learning. More samples may lead to the better performance of deep transfer learning methods. The results of this study showed that it is feasible to combine spectral data with deep transfer learning for the classification of rice diseases and that the inclusion of more samples and the use of emerging transfer learning methods are worth further study. A relevant standard database of spectral data, like ImageNet in the field of computer vision, with different rice diseases could be organized and developed, which would be a valuable resource for researchers, educators, and students.

In future studies, transfer learning methods can be extended to more scenarios, such as different regions and different equipment. Transfer learning has great potential for rice disease detection and will contribute to the translation of relevant researches into practical applications in an efficient manner.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LF and CZ: conceptualization, formal analysis, and supervision. CZ and BW: data curation and validation. YH, LF, and CZ: funding acquisition and writing, project administration review and editing. CZ: investigation, methodology, and software. CZ and YH: resources. BW and LF: visualization and writing, original draft.

FUNDING

This study was supported by XPCC Science and Technology Projects of Key Areas (2020AB005) and the National Natural Science Foundation of China (31871526, 61705195).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.693521/full#supplementary-material

REFERENCES

Abdulridha, J., Ampatzidis, Y., Roberts, P., and Kakarla, S. C. (2020). Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. *Biosyst. Eng.* 197, 135–148. doi: 10.1016/j.biosystemseng.2020.07.001

Awaludin, N., Abdullah, J., Salam, F., Ramachandran, K., Yusof, N. A., and Wasoh, H. (2020). Fluorescence-based immunoassay for the detection of Xanthomonas oryzae pv. oryzae in rice leaf. *Anal. Biochem.* 610:113876. doi: 10.1016/j.ab.2020.113876

Cai, H., Zhu, L., and Han, S. (2018). ProxylessNAS: direct neural architecture search on target task and hardware. *arXiv preprint arXiv:1812.00332*.

Chen, L., Liu, D., Zhou, J., Bin, J., and Li, Z. (2020). Calibration transfer for near-infrared (NIR) spectroscopy based on neighborhood preserving embedding. *Anal. Lett.* 54, 1–19. doi: 10.1080/00032719.2020.1788572

Conrad, A. O., Li, W., Lee, D.-Y., Wang, G.-L., Rodriguez-Saona, L., and Bonello, P. (2020). Machine learning-based presymptomatic detection of rice sheath blight using spectral profiles. *Plant Phenom.* 2020, 8954085–8954085. doi: 10.34133/2020/8954085

Feng, L., Wu, B., Zhu, S., Wang, J., Su, Z., Liu, F., et al. (2020). Investigation on data fusion of multisource spectral data for rice leaf diseases identification using machine learning methods. *Front. Plant Sci.* 11:77063. doi: 10.3389/fpls.2020.577063

Fernando, B., Habrard, A., Sebban, M., and Tuytelaars, T. (2013). "Unsupervised visual domain adaptation using subspace alignment," in *2013 IEEE International Conference on Computer Vision* (Cambridge, MA: IEEE),2960-2967.

Gao, Z., Khot, L. R., Naidu, R. A., and Zhang, Q. (2020). Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging. *Comput. Electron. Agric.* 179:807. doi: 10.1016/j.compag.2020.105807

Ghifary, M., Kleijn, W. B., and Zhang, M. J. (2014). "Domain adaptive neural networks for object recognition," in *Pricai 2014: Trends in Artificial Intelligence,* eds. D.N. Pham and S.B. Park, 898–904.

Gong, B., Shi, Y., Sha, F., and Grauman, K. (2012). "Geodesic Flow Kernel for Unsupervised Domain Adaptation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition* (Cambridge, MA: IEEE), 2066–2073.

Gretton, A., Sriperumbudur, B., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., et al. (2012). "Optimal kernel choice for large-scale two-sample tests," in *Advances in Neural Information Processing Systems,* 1205–1213.

Jiang, X., Zhang, Y., Li, Y., Li, S., and Zhang, Y. (2020). Hyperspectral image classification with transfer learning and markov random fields. *IEEE Geosci. Remote Sens. Lett.* 17, 544–548. doi: 10.1109/LGRS.2019.2923647

Käding, C., Rodner, E., Freytag, A., and Denzler, J. (2016). "Fine-tuning deep neural networks in continuous learning scenarios," in *Computer Vision—ACCV 2016 Workshops*, eds. C.-S. Chen, J. Lu and K.-K. Ma (New Yoek, NY: Springer International Publishing), 588–605.

Kumar, A., Kumar, R., Sengupta, D., Das, S. N., Pandey, M. K., Bohra, A., et al. (2020). Deployment of genetic and genomic tools toward gaining a better understanding of rice-xanthomonasoryzae pv. oryzae interactions for development of durable bacterial *Blight Resist. Rice*. 11:1152. doi: 10.3389/fpls.2020.01152

Li, X., Cai, W., and Shao, X. (2015). Correcting multivariate calibration model for near infrared spectral analysis without using standard samples. *J. Near Infra. Spectroscopy* 23, 285–291. doi: 10.1255/jnirs.1165

Lin, F., Guo, S., Tan, C., Zhou, X., and Zhang, D. (2020). Identification of Rice sheath blight through spectral responses using hyperspectral images. *Sensors* 20:21. doi: 10.3390/s20216243

Liu, Y., Xu, H., Xia, Z., and Gong, Z. (2018a). Multi-spectrometer calibration transfer based on independent component analysis. *Analyst* 143, 1274–1280. doi: 10.1039/C7AN01555K

Liu, Z.-Y., Qi, J.-G., Wang, N.-N., Zhu, Z.-R., Luo, J., Liu, L.-J., et al. (2018b). Hyperspectral discrimination of foliar biotic damages in rice using principal component analysis and probabilistic neural network. *Precis. Agric.* 19, 973–991. doi: 10.1007/s11119-018-9567-4

Mingsheng, L., Yue, C., Jianmin, W., and Michael, J. (2015). *Learning Transferable Features with Deep Adaptation Network*. Long Beach, CL: PMLR.

Molla, K. A., Karmakar, S., Molla, J., Bajaj, P., Varshney, R. K., Datta, S. K., et al. (2020). Understanding sheath blight resistance in rice: the road behind and the road ahead. *Plant Biotechnol. J.* 18, 895–915. doi: 10.1111/pbi.13312

Oquab, M., Bottou, L., Laptev, I., and Sivic, J. (2014). "Learning and transferring mid-level image representations using convolutional neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1205–1212.

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Qiu, Z., Zhao, S., Feng, X., and He, Y. (2020). Transfer learning method for plastic pollution evaluation in soil using NIR sensor. *Sci. Total Environ.* 740:140118. doi: 10.1016/j.scitotenv.2020.140118

Saenko, K., Kulis, B., Fritz, M., and Darrell, T. (2010). "Adapting Visual Category Models to New Domains", in: *Computer Vision—ECCV 2010*, eds K. Daniilidis, P. Maragos and N. Paragios: Springer Berlin Heidelberg), 213–226.

Shrivastava, V. K., and Pradhan, M. K. (2020). Rice plant disease classification using color features: a machine learning paradigm. *J. Plant Pathol.* 3:683. doi: 10.1007/s42161-020-00683-3

Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv preprint arXiv:1312.6034*.

Sun, B. C., Feng, J. S., Saenko, K., and Aaai T. (2016). "Return of frustratingly easy domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence* (Palo Alto, CL: AAAI Press), 30.

Sun, B. C., and Saenko, K. (2016). "Deep CORAL: correlation alignment for deep domain adaptation," in *Computer Vision - Eccv 2016 Workshops, Pt Iii,* eds G. Hua and H. Jegou), 443–450.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). "A survey on deep transfer learning," in *Artificial Neural Networks and Machine Learning—ICANN 2018*, eds. V. Kurková, Y. Manolopoulos, B. Hammer, L. Iliadis and I. Maglogiannis (New York, NY: Springer International Publishing), 270–279.

Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., and Darrell, T. (2014). Deep domain confusion: maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*.

Wang, M., and Deng, W. H. (2018). Deep visual domain adaptation: a survey. *Neurocomputing* 312, 135–153. doi: 10.1016/j.neucom.2018.05.083

Yang, J., Chen, G., Li, L., Zou, Z., and Mao, C. (2019). Portable rice disease spores capture and detection method using diffraction fingerprints on microfluidic chip. *Micromachines* 10:289. doi: 10.3390/mi10050289

Yang, Q. (2010). Three challenges in data mining. *Front. Comput. Sci. China* 4, 324–333. doi: 10.1007/s11704-010-0102-7

Zhang, R., Xie, H., Cai, S., Hu, Y., Liu, G., k., et al. (2019). Transfer-learning-based Raman spectra identification. *J. Raman Spectrosc.* 51, 176–186. doi: 10.1002/jrs.5750

Zhao, X., Liang, Y., Guo, A. J. X., and Zhu, F. (2020). Classification of small-scale hyperspectral images with multi-source deep transfer learning. *Remote Sens. Lett.* 11, 303–312. doi: 10.1080/2150704X.2020.1714772

Zhao, Y. Y., Zhu, S. S., Zhang, C., Feng, X. P., Feng, L., and He, Y. (2018). Application of hyperspectral imaging and chemometrics for variety classification of maize seeds. *RSC Adv.* 8, 1337–1345. doi: 10.1039/C7RA05954J

Zhong, Y. W., Li, H. L., and Chen, L. L. (2021). Construction project risk prediction model based on EW-FAHP and one dimensional convolution neural network. *PLoS ONE* 16:246539. doi: 10.1371/journal.pone.0246539

Zhu, M. Y., Yang, H. B., and Li, Z. W. (2019). Early Detection and identification of rice sheath blight disease based on hyperspectral image and chlorophyll content. *Spectroscopy Spectr. Anal.* 39, 1898–1904.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Feng, Wu, He and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership