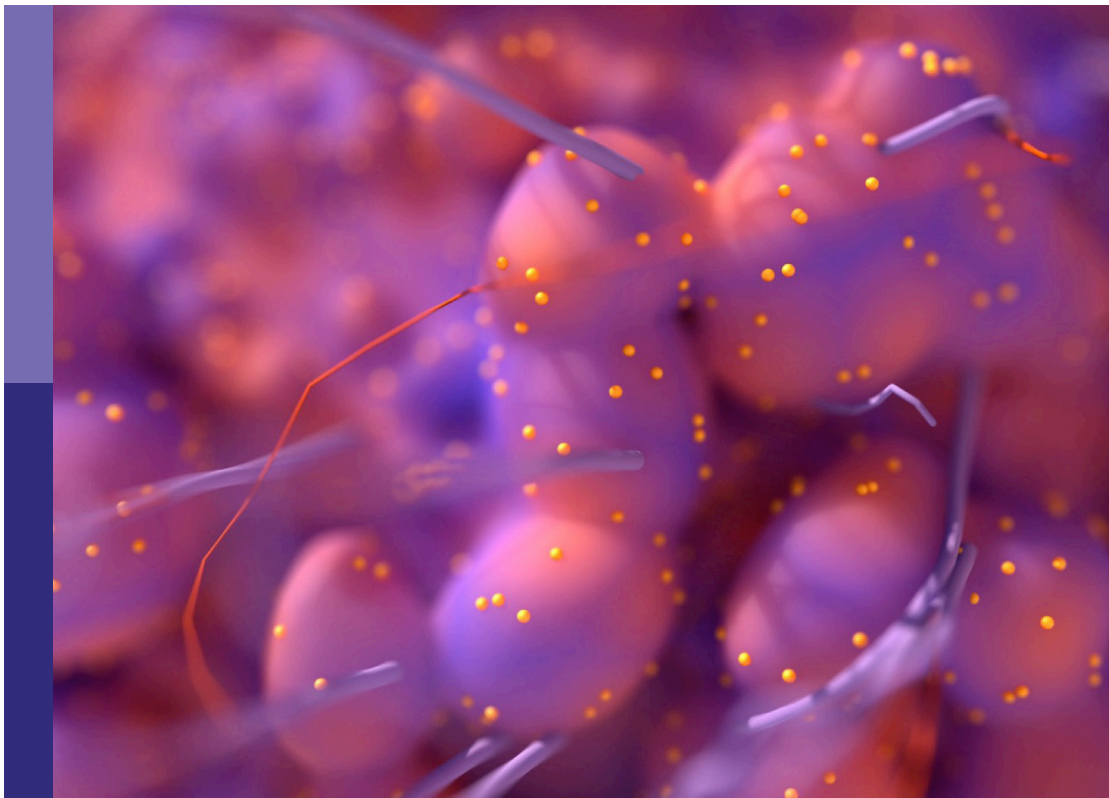
frontiers | Research Topics

Quantitative imaging: revolutionizing cancer management with biological sensitivity, specificity, and AI integration

Edited by
Harrison Kim, Eric Edward Sigmund and Kyung Hyun Sung

Published in
Frontiers in Oncology



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 978-2-8325-7484-3
DOI 10.3389/978-2-8325-7484-3

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Quantitative imaging: revolutionizing cancer management with biological sensitivity, specificity, and AI integration

Topic editors

Harrison Kim — University of Alabama at Birmingham, United States

Eric Edward Sigmund — New York University, United States

Kyung Hyun Sung — Department of Radiology, UCLA Health System, United States

Citation

Kim, H., Sigmund, E. E., Sung, K. H., eds. (2026). *Quantitative imaging: revolutionizing cancer management with biological sensitivity, specificity, and AI integration*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7484-3

Table of contents

06 Radiomic biomarkers of locoregional recurrence: prognostic insights from oral cavity squamous cell carcinoma preoperative CT scans
Xiao Ling, Gregory S. Alexander, Jason Molitoris, Jinhyuk Choi, Lisa Schumaker, Phuoc Tran, Ranee Mehra, Daria Gaykalova and Lei Ren

23 A hybrid deep learning scheme for MRI-based preliminary multiclassification diagnosis of primary brain tumors
Zhichao Wang, Chuchu He, Yan Hu, Haifeng Luo, Chao Li, Xiandong Wu, Yang Zhang, Jingjing Li and Jun Cai

35 A nomogram based on conventional and contrast-enhanced ultrasound radiomics for the noninvasively prediction of axillary lymph node metastasis in breast cancer patients
Chao Sun, Xuantong Gong, Lu Hou, Di Yang, Qian Li, Lin Li and Yong Wang

46 Revealing tumor cells and tissues with high selectivity through folic acid-targeted nanofluorescence probes responsive to acidic microenvironments
Jing Li, Hongyi He, Shuyan Liu, Xining Li and Fengfeng Wu

58 Imaging of colorectal adenomas with pseudoinvasion and malignant polyps using two-photon excitation microscopy
Maria-Alexandra Florea, Lucian George Eftimie, Remus Relu Glogojeanu, Radu Hristu, George A. Stanciu and Mariana Costache

70 Comparative evaluation of imaging methods for prognosis assessment in esophageal squamous cell carcinoma: focus on diffusion-weighted magnetic resonance imaging, computed tomography and esophagography
Yang Li, Xiaohua Su, Yuguang Shang, Hui Liu, Weishuai Wang, Andu Zhang and Gaofeng Shi

83 The application of mixed reality technique in oromaxillo-facial reconstruction with the perforator flap for malignant tumor patients
Yixiu Liu, Jian Wu, Daide Liu, Dalan Xiang, Xiaoyue Wu and Ting Wang

94 Radiomics based on ^{18}F-FDG PET/CT for prediction of pathological complete response to neoadjuvant therapy in non-small cell lung cancer
Jianjing Liu, Chunxiao Sui, Haiman Bian, Yue Li, Ziyang Wang, Jie Fu, Lisha Qi, Kun Chen, Wengui Xu and Xiaofeng Li

106 Spatiotemporal correlation enhanced real-time 4D-CBCT imaging using convolutional LSTM networks
Hua Zhang, Kai Chen, Xiaotong Xu, Tao You, Wenzheng Sun and Jun Dang

116 Development and validation of a CT-based nomogram for accurate hepatocellular carcinoma detection in high risk patients
Yingying Liang, Hongzhen Wu and Xinhua Wei

125 Pilot study: radiomic analysis for predicting treatment response to whole-brain radiotherapy combined temozolomide in lung cancer brain metastases
Yichu Sun, Fei Liang, Jing Yang, Yong Liu, Ziqiang Shen, Chong Zhou and Youyou Xia

134 Assessment of pathological grade and variants of bladder cancer with a continuous-time random-walk diffusion model
Wei Wang, Jingyun Wu, Qi Shen, Wei Li, Ke Xue, Yuxin Yang and Jianxing Qiu

144 Precision lung cancer screening from CT scans using a VGG16-based convolutional neural network
Hua Xu, Yuanyuan Yu, Jie Chang, Xifeng Hu, Zitong Tian and Ouwen Li

155 Weakly supervised large-scale pancreatic cancer detection using multi-instance learning
Shyamapada Mandal, Keerthiveena Balraj, Hariprasad Kodamana, Chetan Arora, Julie M. Clark, David S. Kwon and Anurag S. Rathore

168 Assessing microvascular invasion in HBV-related hepatocellular carcinoma: an online interactive nomogram integrating inflammatory markers, radiomics, and convolutional neural networks
Yun Zhong, Lingfeng Chen, Fadian Ding, Wenshi Ou, Xiang Zhang and Shangeng Weng

181 Assessment of synthetic MRI to distinguish Warthin's tumor from pleomorphic adenoma in the parotid gland: comparison of two methods of positioning the region of interest for synthetic relaxometry measurement
Jiabin Sun, Xinping Kuai, Dawei Huang, Xinghua Ji, Chuanhai Jia and Shengyu Wang

190 Habitat-based CT radiomics enhances the ability to predict spread through air spaces in stage T1 invasive lung adenocarcinoma
Xiuhua Peng, Hongxing Zhao, Shiyong Wu, Dan Jia, Miaomiao Hu, Biping Guo, Jinliang Hu and Pengliang Xu

202 Evaluating peritumoral and intratumoral radiomics signatures for predicting lymph node metastasis in surgically resectable non-small cell lung cancer
Ran Xu, Kaiyu Wang, Bo Peng, Xiang Zhou, Chenghao Wang, Tong Lu, Jiaxin Shi, Jiaying Zhao and Linyou Zhang

218 ^{18}F-FDG PET/CT detection of primary mammary-like adenocarcinoma of the vulva with multiple metastases: a case report
Qing Song, Biyun Zhang and Yu Gu

224 TriSwinUNETR lobe segmentation model for computing DIR-free CT-ventilation
Gabriela Roque Oliveira Nomura, Aaron T. Luong, Ananya Prakash, Annabelle Alemand, Tanish Bhowmick, Alisa Ali, Jaimie Ren, Basil Rehani, Girish Nair, Richard Castillo, Yevgeniy Vinogradskiy and Edward Castillo

235 Leveraging automated time-lapse microscopy coupled with deep learning to automate colony forming assay
Anusha Klett, Dennis Raith, Paula Silvestrini, Matías Stingl, Jonas Bermeitinger, Avani Sapre, Martin Condor, Roman Melachrinos, Mira Kusterer, Alexandra Brand, Guido Pisani, Evelyn Ullrich, Marie Follo, Jesús Duque-Afonso and Roland Mertelsmann

251 Pilot study of an arterial enhancement fraction-based model for progression prediction in HCC undergoing chemoembolization
Bin Chai, Dongqiao Xiang, Guofeng Zhou and Chuansheng Zheng

263 Retrospective BReast Intravoxel Incoherent Motion Multisite (BRIMM) multisoftware study
Dibash Basukala, Artem Mikheev, Xiaochun Li, Judith D. Goldberg, Nima Gilani, Linda Moy, Katja Pinker, Savannah C. Partridge, Debosmita Biswas, Masako Kataoka, Maya Honda, Mami Iima, Sunitha B. Thakur and Eric E. Sigmund

276 A deep learning aided bone marrow segmentation of quantitative fat MRI for myelofibrosis patients
Humera Tariq, Lubomir Hadjiiski, Dariya Malyarenko, Moshe Talpaz, Kristen Pettit, Gary D. Luker, Brian D. Ross and Thomas L. Chenevert

289 Leveraging a foundation model zoo for cell similarity search in oncological microscopy across devices
Gabriel Kalweit, Anusha Klett, Paula Silvestrini, Jens Rahnfeld, Mehdi Naouar, Yannick Vogt, Diana Infante, Rebecca Berger, Jesús Duque-Afonso, Tanja Nicole Hartmann, Marie Follo, Elitsa Bodurova-Spassova, Michael Lübbert, Roland Mertelsmann, Joschka Boedecker, Evelyn Ullrich and Maria Kalweit

308 Establishment of a CT-based radiomic feature robustness databank for OPC patients via image perturbation in a multi-institutional study: a practical method to safeguard model generalizability
Yongqiang Wang, Alex Zwanenburg, Jiang Zhang, Xinzhi Teng, Sai-Kit Lam, Jin Cao, Zongrui Ma, Ta Zhou, Yuanpeng Zhang, Hong Ge and Jing Cai

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 April 2024
DOI 10.3389/fonc.2024.1380599

Check for updates

OPEN ACCESS

EDITED BY
Harrison Kim,
University of Alabama at Birmingham, United States

REVIEWED BY
Lubomir Hadjiyski,
University of Michigan, United States
Hongyi Pan,
Northwestern University, United States

*CORRESPONDENCE
Lei Ren
lren@som.umaryland.edu

RECEIVED 01 February 2024
ACCEPTED 04 April 2024
PUBLISHED 23 April 2024

CITATION
Ling X, Alexander GS, Molitoris J, Choi J, Schumaker L, Tran P, Mehra R, Gaykalova D and Ren L (2024) Radiomic biomarkers of locoregional recurrence: prognostic insights from oral cavity squamous cell carcinoma preoperative CT scans.
Front. Oncol. 14:1380599.
doi: 10.3389/fonc.2024.1380599

COPYRIGHT
© 2024 Ling, Alexander, Molitoris, Choi, Schumaker, Tran, Mehra, Gaykalova and Ren. 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.

Radiomic biomarkers of locoregional recurrence: prognostic insights from oral cavity squamous cell carcinoma preoperative CT scans

Xiao Ling[1], Gregory S. Alexander[2], Jason Molitoris[1], Jinhyuk Choi[3], Lisa Schumaker[4], Phuoc Tran[1], Ranee Mehra[4], Daria Gaykalova[5,6,7] and Lei Ren[1]*

[1]Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, MD, United States, [2]Department of Radiation Oncology, Thomas Jefferson University, Philadelphia, PA, United States, [3]Department of Breast Surgery, Kosin University Gospel Hospital, Busan, Republic of Korea, [4]Marlene and Stewart Greenebaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD, United States, [5]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, United States, [6]Department of Otorhinolaryngology-Head and Neck Surgery, Marlene and Stewart Greenebaum Comprehensive Cancer Center, University of Maryland Medical Center, Baltimore, MD, United States, [7]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, United States

Introduction: This study aimed to identify CT-based imaging biomarkers for locoregional recurrence (LR) in Oral Cavity Squamous Cell Carcinoma (OSCC) patients.

Methods: Computed tomography scans were collected from 78 patients with OSCC who underwent surgical treatment at a single medical center. We extracted 1,092 radiomic features from gross tumor volume in each patient's pre-treatment CT. Clinical characteristics were also obtained, including race, sex, age, tobacco and alcohol use, tumor staging, and treatment modality. A feature selection algorithm was used to eliminate the most redundant features, followed by a selection of the best subset of the Logistic regression model (LRM). The best LRM model was determined based on the best prediction accuracy in terms of the area under Receiver operating characteristic curve. Finally, significant radiomic features in the final LRM model were identified as imaging biomarkers.

Results and discussion: Two radiomics biomarkers, Large Dependence Emphasis (LDE) of the Gray Level Dependence Matrix (GLDM) and Long Run Emphasis (LRE) of the Gray Level Run Length Matrix (GLRLM) of the 3D Laplacian of Gaussian (LoG σ=3), have demonstrated the capability to preoperatively distinguish patients with and without LR, exhibiting exceptional testing specificity (1.00) and sensitivity (0.82). The group with LRE > 2.99 showed a 3-year recurrence-free survival rate of 0.81, in contrast to 0.49 for the group with

LRE ≤ 2.99. Similarly, the group with LDE > 120 showed a rate of 0.82, compared to 0.49 for the group with LDE ≤ 120. These biomarkers broaden our understanding of using radiomics to predict OSCC progression, enabling personalized treatment plans to enhance patient survival.

KEYWORDS

oral cavity squamous cell carcinoma, outcome prediction, biomarker, recurrence, classification, logistic regression, CT, radiomics

1 Introduction

Oral Cavity Squamous Cell Carcinoma (OSCC) is the most common malignancy in the head and neck region and is characterized by a poor prognosis (1). Surgery is the primary treatment of OSCC, followed by cisplatin-based chemotherapy and/or radiotherapy depending on pathologic features and individualized risk of recurrence. Regional recurrence is the most common cause of failure after treatment of oral carcinoma (2, 3). Despite advancements in surgical techniques and adjuvant therapies, the 5-year overall survival rate hovers between 45-50%, contingent upon the stage and metastasis status of the disease (4, 5). Locoregional recurrence (LR), as indicated by prior studies (6, 7), represents a significant clinical challenge, with some patients cohorts demonstrating extremely high rates of LR even following surgery and appropriate adjuvant therapy (8). Given that disease recurrence is devastating for patients, and adjuvant therapies are associated with significant economic and quality of life detriment, identification of patients at higher risk of LR who would benefit most from adjuvant treatment is paramount. It is, therefore, relevant to identify patients who are at a higher risk of locoregional recurrence before their primary surgery to guide treatment plans and increase the therapeutic window. By utilizing noninvasive imaging information and cutting-edge machine learning algorithms, post-treatment failure can be better screened, enabling medical professionals to tailor treatment plans accordingly.

Histopathologic factors are used for OSCC diagnosis and prognosis staging evaluation (9). Studies (10–15) reported that tumor size, depth of invasion (DOI), stromal, vascular, and nerve invasion are significantly different between the groups with and without metastasis. The dysregulation of specific miRNAs in OSCC, such as miRNA-184 (16), miR-31 (17), and miR-27b (18), are implicated in malignant transformation and disease progression. Other proteins and peptides, such as Leukotriene A4 hydrolase (LTA4H) and its peptide, Pep8_LTA4H, among other proteins and peptides, may distinguish individuals with metastasis (N+) from individuals metastasis-free (N0) (19). Studies (20–22) implicated that the amplification of CCND1 and overexpression of cyclin D1 are significantly correlated with OSCC metastasis. Soluble factors, such as IL- 1β,TNF- α, and MIP- 1β, that can be detected in saliva, may also play a significant role in detecting metastasis (23). The other study (24) reveals an association between primary site recurrence and a high ratio of ITGA3/CD9. Elevated levels of squamous cell carcinoma antigen (SCC-Ag) in serum are significantly associated with tumor progression (25). While these parameters evaluated in pre-clinical settings hold promise in enhancing disease detection, prognosis, and personalized treatment, those findings need to be confirmed by larger and more rigorous studies. One of the limitations of histopathologic biopsy is that it may not capture the full heterogeneity of the tumor due to sampling bias (26). Furthermore, factors such as DOI are only available on the resection specimen. The extraction and analysis of biomarkers, such as H&E staining, tissue microarray, and sequencing, can be technically complex and expensive, requiring specialized resources, which may restrict their practicality in specific circumstances. Additionally, challenges with reproducibility and standardization across laboratories and the potential for false positives and negatives further complicate their practicality (27, 28). Furthermore, validating a biomolecule-based assay, from its initial discovery to clinical implementation, is often arduous and lengthy. A significant number of potential markers prove to be ineffective across various populations (29). Additionally, it is crucial to reduce the overall processing time to avoid LR in patients who require adjuvant therapy. Also, while some biomarkers may indicate the presence of a disease, they might not offer actionable information for treatment plans, thereby restricting their practical clinical use (30, 31). Lastly, employing genetic and other biomolecular markers raises ethical, legal, and societal concerns (32). This is a primary reason these biomarkers have not been introduced in clinical settings and lack FDA approval.

Imaging-based biomarkers have been investigated for different modalities, such as Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI), by extracting quantitative imaging features known as radiomics features. In contrast to biomolecule-based assays, imaging techniques are non-invasive. The imaging information is readily available from routine diagnostic scans without incurring additional costs. Moreover, imaging provides unique 3D information about neoplasm. These radiomics features can be

leveraged to develop predictive models for survival and treatment failure (33–42). The rationale behind this approach is that these images capture crucial information about the neoplasm phenotype and microenvironment (43). In fact, the American College of Radiology has developed a standardized Neck Imaging Reporting and Data System (NI-RADS) (44) to manage and surveil the posttreatment course. Studies (45, 46) demonstrated a strong association between NI-RADS category and treatment failure in HNSCC patients. Over recent decades, imaging factors have demonstrated their capacity to furnish accurate prognostic information for posttreatment recurrence screening. A study (47) found a significant association between PET/CT radiomic features and Head and Neck locoregional recurrence. Our pilot study (48) demonstrated two potential Radiomic overall survival biomarkers. However, the identification of non-invasive factors for 2-year locoregional recurrence after primary surgery in OSCC patients remains lacking. The distinction between our study and similar research lies in the emphasis on the susceptibility/risk associated with the biomarker, specifically an increased likelihood of developing locoregional recurrence (LR) within 2 years post-surgery. Consequently, we focus more on specificity and sensitivity to minimize the incidence of false positives and false negatives.

This study aimed to identify CT-based imaging risk factors for locoregional recurrence in patients with OSCC at an academic health network serving a diverse population, which enabled the development of machine learning classifiers that could accurately distinguish patients with locoregional recurrence from those without prior to treatment. A retrospective study design was used, with high-dimensional radiomics, pathological, and clinical information collected from this diverse cohort of OSCC cases. The primary endpoint was 2-year locoregional recurrence (defined as locoregional recurrence occurring within 2 years of surgery). The findings of this study lay the foundation for the implementation of pre-treatment screening for LR and risk assessment using non-invasive risk factors in this diverse patient population, which could ultimately impact the management of high-risk OSCC patients by helping physicians customize treatment planning and reduce the chance of distant metastasis.

2 Materials and methods

2.1 Data preparation and overall workflow

The workflow outlining our approach is illustrated in Figure 1. In this workflow, the neoplasm volume serves as the region of interest (ROI) from which all radiomics features are computed. The contouring of the ROI was performed manually by experienced Radiation Oncologists, not directly involved in the study, using the Varian Medical System Eclipse software environment. These features underwent a selection process to minimize redundancy and were combined with clinical data. A logistic regression model, optimized via five-repeated 10-fold cross-validation, was then applied. The model's predictive performance was evaluated using the Area Under the ROC Curve (AUC). All statistical analyses were performed using R programming language, with a significance level (alpha) set at 0.05 for all tests.

2.2 Radiomic feature acquisition and extraction

Uniformity in voxel sizes is essential for precise and dependable feature calculations in radiomics (49). Given the original CT scan resolutions varied significantly, from 0.3×0.3×0.5 mm to 1.3×1.3×5 mm, we resampled the voxel values to a uniform resolution of 1×1×1mm using interpolation. For this purpose, we employed the Bspline algorithm (50), suitable for smoothly adjusting voxel values. Concurrently, to maintain the binary nature of the tumor masks, we

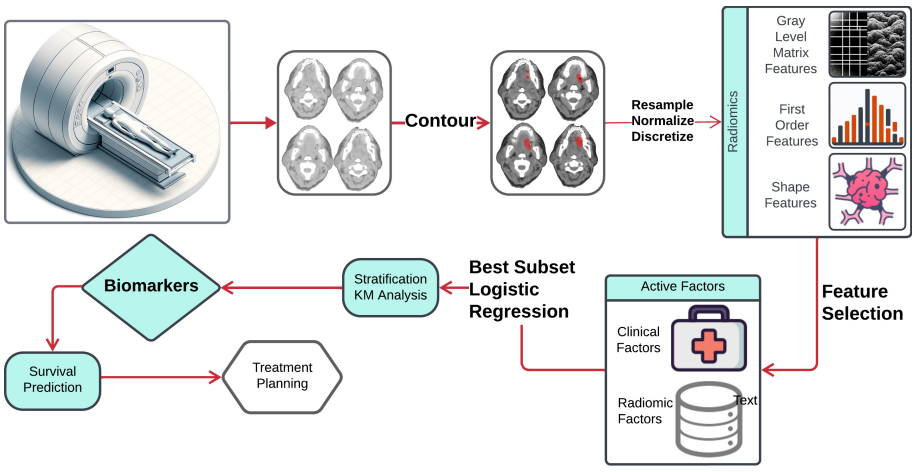


FIGURE 1
Schematic flowchart illustrating the steps from CT image acquisition and radiomics feature extraction, through the process of machine learning techniques to risk factors.

applied the nearest neighbor interpolation algorithm (51, 52), which assigns the mask label of the closest voxel in the original grid to each voxel in the interpolation grid. Another critical step employed was normalization, computed by $\frac{x-\mu}{\sigma}$, where x represents voxel intensity, μ is the average intensity, and σ is the standard deviation of intensity. As demonstrated in a previous study (49), this normalization minimizes variance and boosts feature robustness of radiomic features, especially against different discretization levels, ensured normalized intensities generally spanned from [-3, 3] after outlier exclusion (53, 54). These intensities were then scaled to a range of approximately [-300, 300]. Moreover, to capture detailed textural information, we discretized the intensities within the ROI using a uniform bin width of 5, starting from a normalized minimum HU value of 0. We chose a bin width of 5 to ensure an adequate number of bins (between 1 and 400), allowing for the capture of more detailed textural information (55). This discretization assigns new values to each voxel according to the formula: floor((original intensity)/5) + 1. This method not only suppresses noise but also enhances the robustness of radiomic features by smoothing out minor variations.

Medical images provide insights into the phenotypic traits of neoplasms. These images typically contain data from tens of thousands of voxel intensities per neoplasm, leading to a scenario where the number of features (p) greatly exceeds the sample size (n). In our study, we extracted features from each image set using the PyRadiomics library in Python. According to the Imaging Biomarker Standardization Initiative (IBSI) (56), we extracted features across six categories: shape, first-order statistics, gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM) (57), grey level size zone matrix (GLSZM) (58), gray level dependence matrix (GLDM) (59), and neighborhood grey tone difference matrix (NGTDM) (60). Additional calculations were performed on images processed with wavelet, Laplacian of Gaussian (LoG), square, square root, logarithm, exponential, and gradient filters, culminating in 1,092 features.

2.3 Feature selection and modeling

Logistic Regression Models (LRM) were used in assessing the features of discriminative power. Algorithms are available in the R stats, glmulti package. A logistic model can be mathematically described as follows: $\log\frac{\pi(x)}{1-\pi(x)} = \beta x$, where $\pi(x)$ is the probability of recurrence, and β the coefficient vector for the independent variables x (61). The guidelines outlined in (62) recommend that the number of predictors used in fitting LRM should not exceed 10% of the events in the sample. In our cohort, with 10% of events in the training sample, the optimal number of predictors for model fitting should be 2 to 3, as 10% of the total recurrence equals 2.1. Thus, we aimed to limit the final Logistic Regression Model (LRM) to a maximum of 3 degrees of freedom. We used the Best Subset Selection (BSS) Modeling strategy to identify the most effective LRM based on validation performance to achieve this. BSS, known for its efficiency in finding the most parsimonious model, outperforms methods like stepwise selection and Lasso, although its high computational demand is a limitation. For instance, fitting LRMs with 2, 3, or 4 degrees of freedom using 17 radiomics and six clinical features requires BSS to estimate a minimum of 41,262 coefficients, making exhaustive evaluation impractical. Therefore, we reduced the number of input variables before employing BSS.

Radiomics data often faces the challenge of high multicollinearity, where variables are highly correlated, affecting the significance of individual variables in the model. For example, sphericity, minor axis length, and elongation show strong multicollinearity. Multicollinearity can lead to the phenomenon where a variable is not deemed significant when correlated features are also present in the model. Figure 2 uses a color scheme where white represents no correlation, blue represents a perfect negative correlation, and red represents a perfect positive correlation. The heatmap illustrates the correlation coefficients prior to the feature selection process, revealing the initial relationships between features. The heatmap revealed numerous red and blue shades, indicating strong positive and negative correlations, respectively, among the data. There is now a substantial body of research on mitigating multicollinearity, such as Principal Component Analysis (PCA), Sparse PCA (63), and Kernel PCA (KPCA) (64). We employed Recursive Feature Elimination (RFE) (65), an iterative procedure to refine the input data for BSS LRM. Using RFE with repeated 10-fold cross-validation (Table 1), we narrowed down from 1,092 radiomic features to a subset of eight active features (Figure 3). We then fitted degree-2 LRMs, considering all combinations of these eight features across 5,000 data shuffles. The models were trained on 63 samples (80% of the cohort) using 10-fold cross-validation and evaluated based on AUC. The remaining 15 samples (20% of the cohort) were used for prognostic validation. We also used the ROC to visualize the classifiers' performance. Figure 4 shows that the 3,480 models out of 5,000 data shufflings unveiled the most distinctive factor among the extensive array of 1,092 radiomic features: specifically, Large Dependence Emphasis (LDE) of the Gray Level Dependence Matrix (GLDM) and Long Run Emphasis (LRE) of the Gray Level Run Length Matrix (GLRLM) of the 3D Laplacian of Gaussian ($LoG_{\sigma=3}$) filtered ROIs (Figures 5, 6). Statistical evidence was quantified by the p-value of the hypotheses' tests. If the p-value is less than $\alpha = 0.05$, then the null hypothesis will be rejected in favor of the alternative hypothesis.

2.4 Assessment of the prognosis

We stratified the cohort into high and low LRE and LDE subgroups according to the threshold selected by using respective median. Kaplan-Meier curve analyses were conducted to assess the impacts of the biomarkers on RFS. RFS was defined as the time from surgery to locoregional recurrence.

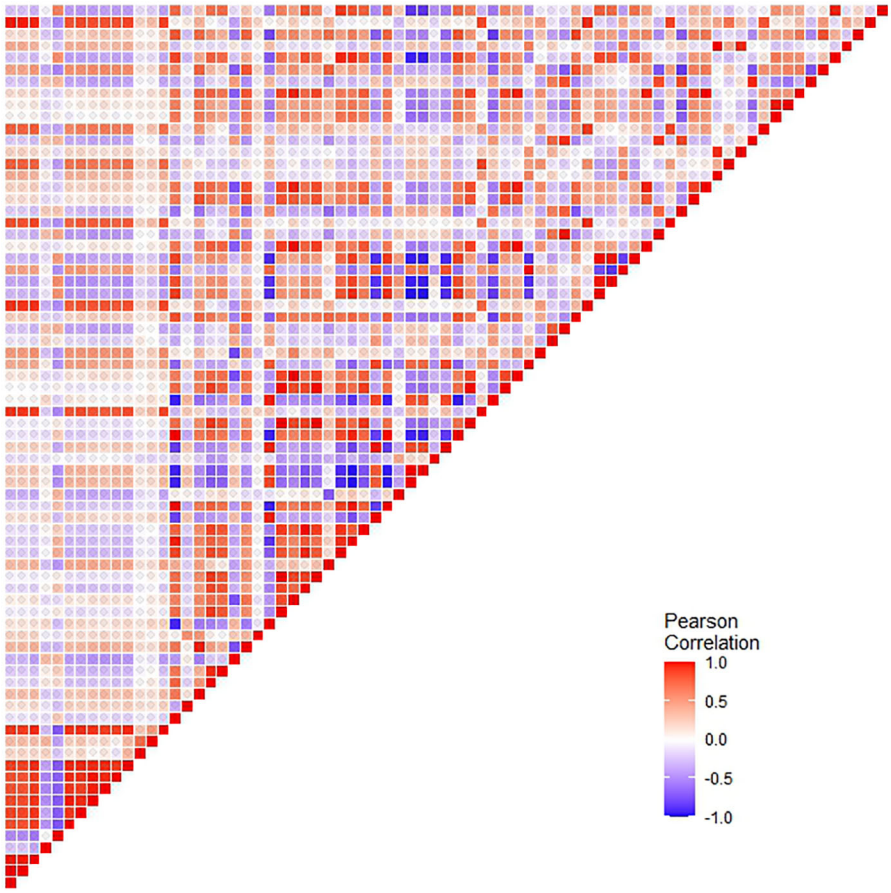


FIGURE 2
Correlation coefficient heatmaps: The diagonal heatmap illustrates the pairwise correlations among radiomics features before pruning. The color scale represents the strength of the correlation, with blue indicating negative correlation, red indicating positive correlation, and white representing no correlation.

TABLE 1 Recursive feature selection: 10 fold cross-validated repeated 5 times.

Variables	ROC	Sens	Spec	ROCSD	SensSD	SpecSD	Selected
1	0.792	0.507	0.802	0.228	0.360	0.169	
2	0.819	0.597	0.845	0.177	0.347	0.148	
3	0.843	0.620	0.888	0.170	0.332	0.130	
4	0.817	0.623	0.875	0.179	0.331	0.135	
5	0.827	0.573	0.867	0.169	0.339	0.139	
6	0.831	0.570	0.872	0.171	0.375	0.131	
7	0.842	0.543	0.878	0.166	0.392	0.136	
8	0.844	0.543	0.868	0.169	0.379	0.152	*
142	0.809	0.417	0.891	0.196	0.327	0.149	
1092	0.758	0.293	0.929	0.207	0.291	0.120	

ROCSD, standard deviation of ROC; SenSD, standard deviation of Sensitivity; SpecSD, Standard deviation of Specificity.

2.5 Constructing nomogram

The primary end-point of the analysis was the time to peritoneal locoregional recurrence. The follow-up duration to peritoneal LR was calculated from the date of surgery to the date when peritoneal LR was diagnosed or to the last follow-up, and information about the survival status and recurrence type was also documented. Finally, a radiomic nomogram was constructed. A model containing both radiomic and clinical factors was also constructed for comparison.

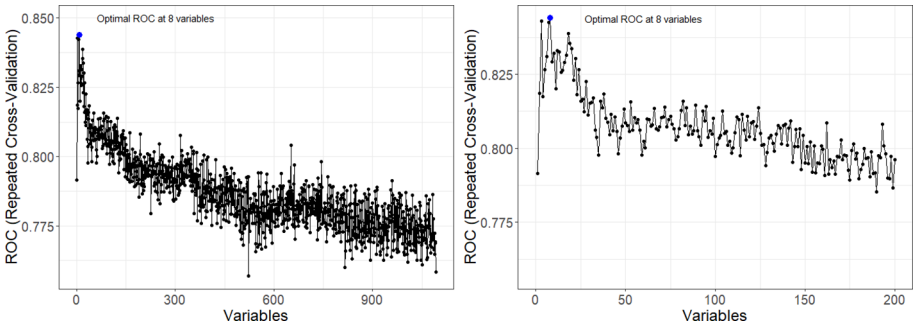


FIGURE 3
Recursive Feature Elimination (RFE) curve from five repeated 10-fold cross-validation. The plot showcases a notable increase in AUC at the retention of eight informative features, followed by a decline in performance as non-informative features are incorporated. Left panel provides a comprehensive view of the AUC achieved across varying numbers of variables. Right panel offers a closer examination of the AUC within the range of 0 to 200 features.

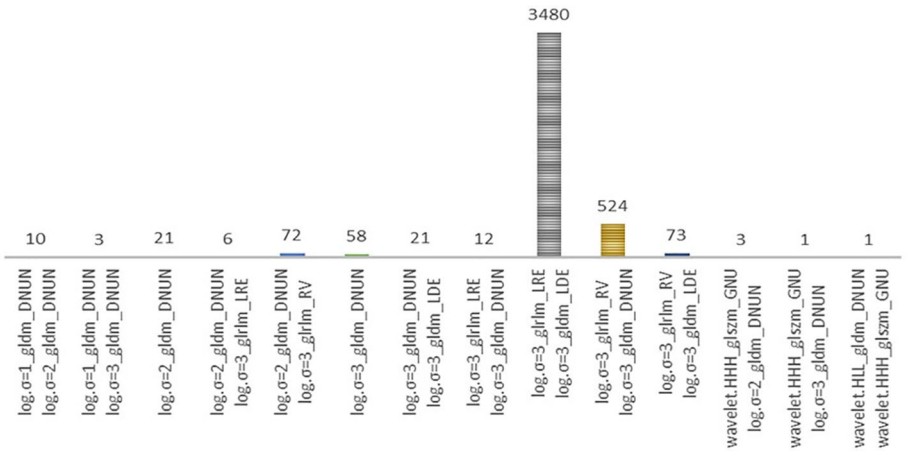


FIGURE 4
This frequency bar plot visualizes the counts of various degree-2 logistic models derived from the best-subset of eight radiomics features previously identified. These models were generated through 5000 iterations of data shuffling. Distribution of models during the selection of the final optimal set of risk radiomics features. These models comprise the risk radiomics features selected by the Best Subset Logistic Regression procedure from among the 8 RFE-selected features. Out of 5000 different partitions, 3480 models are based on two features, log.sigma.3.0.mm.3D_glrlm_LongRunEmphasis and log.sigma.3.0.mm.3D_gldm_LargeDependenceEmphasis.

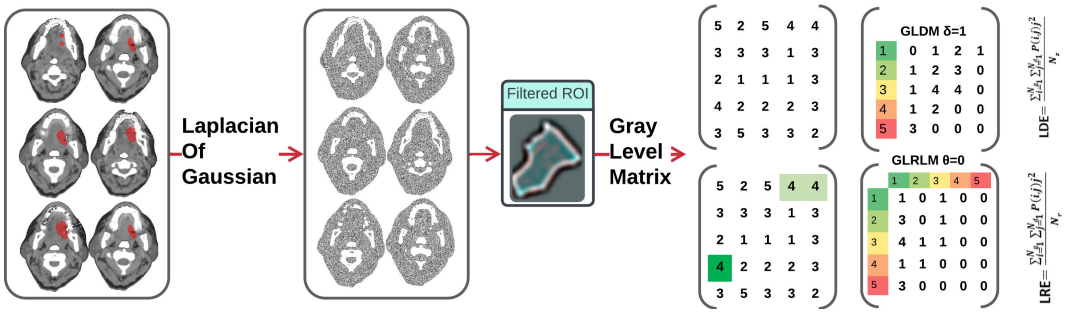


FIGURE 5
Axial contrast-enhanced CT image of the oral cavity with a red region of interest (ROI) indicating the squamous cell carcinoma. The images were filtered by Laplacian of Gaussian (LoG) and analyzed to extract LDE and LRE features.

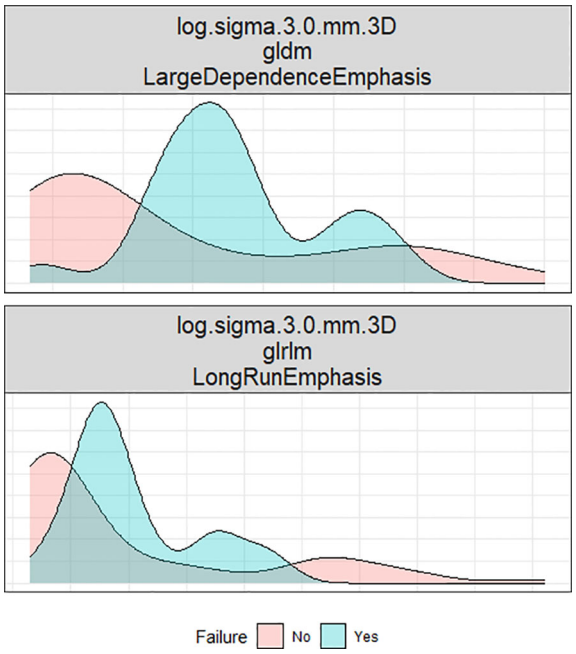


FIGURE 6
This density plot illustrates the distribution of two radiomic factors across two groups: recurrence and non-recurrence.

2.6 Code availability

The code used in this study is available at https://doi.org/10.5281/zenodo.10460030.

3 Results

3.1 Oral cavity squamous cell carcinoma cohort features

This retrospective biomarker analysis examines a group of oral cavity squamous cell carcinoma (OSCC) patients who underwent surgical/curative/elective neck/selective neck resection at the institution between 2006 and 2017. The study involved 78 patients, with 21 experiencing locoregional recurrence (LR), while 57 remained disease-free within a 2-year period after the end of the initial treatment course. Demographic and clinicopathological features of patients are detailed in Table 2. The mean age at the initial surgery was 60, ranging from 30 to 98 years. The median follow-up time for recurrence-free survival (RFS) was 56.2 months. A locoregional recurrence was defined as a positive biopsy in the primary site or the cervical lymphatic region after treatment. We collected six clinical characteristics of interest, including age, gender, tobacco usage, alcohol consumption, T-stage, N-stage, and race. All patients were in the first 2 years of follow-up after surgery. Patients were categorized into four T stages (1, 2, 3, and 4) based on the size and extent of the primary tumor. Smoking and alcohol status were self-reported and coded as 1 for Yes and 2 for No. The missing values for smoking and alcohol status were hard coded as 3 due to their substantial representation within the dataset. Smoking status revealed that 60% of the total cohort were smokers, with this number rising to 72% in the LR subgroup. For alcohol consumption, 40% of the total group reported alcohol use, compared to 48% in the LR subgroup. All patients underwent surgery treatment as primary treatment, along with chemoradiotherapy (CRT) or radiotherapy (RT). The endpoint in this study was 2-year LR status, defined as whether an LR happened within 2 years after curative treatment. Here,

TABLE 2 Frequency and significance of demographic and clinicopathological characteristics.

	Total (%)	LR (%)
Gender		
Male	45 (58%)	14 (67%)
Female	33 (42%)	7 (33%)
Race		
EA	69 (88%)	17 (81%)
AA	9 (12%)	4 (19%)
Smoking		
Yes	47 (60%)	15 (72%)
No	22 (28%)	3 (14%)
Unknown	9 (12%)	3 (14%)
Alcohol		
Yes	31 (40%)	10 (48%)
No	33 (42%)	9 (43%)
Unknown	14 (18%)	2 (9%)
T stage		
T1	26 (33%)	4 (19%)
T2	21 (27%)	4 (19%)
T3	13 (17%)	6 (29%)
T4	18 (23%)	7 (33%)
N stage		
N0	49 (63%)	11 (52%)
N1	10 (13%)	3 (14%)
N2	19 (24%)	7 (34%)
Treatment		
Sx	45 (58%)	8 (38%)
Sx + RT	18 (23%)	5 (24%)
Sx + CRT	15 (19%)	8 (38%)
Registry Sites		
Buccal Mucosa	11 (14%)	1 (5%)
Floor of Mouth	6 (8%)	2 (10%)
Gingiva	13 (17%)	2 (10%)
Retromolar trigone	1 (1%)	0 (0%)
Tongue	47 (60%)	16 (75%)

Total, the entire cohort; LR, the locoregional recurrent cohort.

the proportion increased in the LR subgroup (17% in the total group, 29% in LR), indicating a higher prevalence of this intermediate stage in the LR subgroup. T3 tumors constituted 17% of the cases. The proportion of T3 tumors increases to 29% in the LR subgroup. T4 tumors, which represent the most advanced stage of tumor size and extent, accounted for 23% of the total cohort. The representation of T4 tumors is notably higher in the LR subgroup, constituting 33%. In total, T3 and T4 stages comprise only 31% of the entire group, in contrast to 62% in the LR subgroup.

3.2 Radiomic factors selection and validation

A preliminary feature selection algorithm identified 8 radiomic factors of discriminative power in LR depicted in Figure 7. To ascertain the independence of these two radiomic factors from clinical factors and their potential as clinical alternatives, we investigated their interaction with clinical factors, smoking, alcohol I(ETOH), N stage, and T stage. Logistic regression modeling then incorporated the radiomic features with demographic and clinicopathological characteristics to define the final radiomic risk factors. In this comparative analysis of five logistic regression models in Table 3, denoted as Models 1 through 5, we have assessed their performance based on a range of statistical metrics. The Akaike Information Criterion (AIC) is employed as a model selection criterion. At the same time, accuracy (ACC), area under the receiver operating characteristic curve (AUC), sensitivities (Sens), and specificities (Spec) are utilized to evaluate the models' predictive capabilities. It is evident that the 2nd model (clinical-only) exhibits the highest AIC of 80, suggesting a worse fit to the training data (80%) compared to the other models. When considering the training measurements, the model demonstrates the lowest accuracy (0.75) and AUC values (0.46). In regard to the testing measurements on the held-out 20% data, the 2nd model shows the lowest accuracy (0.6) and AUC (0.67) among all models, showcasing its deficiency in generalization. Moreover, the model consistently maintains an unbalanced sensitivity (0.45) and specificity (1), highlighting its inability to make accurate predictions while minimizing false positives and false negatives. These findings collectively underscore the suboptimal performance of the 2nd model and establish it as the least favorable choice when contrasted with the other models in this analysis.

Numerous studies have identified smoking and drinking as risk factors for OSCC patients. Further analysis indicated that including clinical factors didn't significantly enhance the model's explanatory power (based on deviance analysis in Table 4 via chi-square test, P=0.43). Table 4 presents the Analysis of Deviance results for eight pairs of nested model comparisons, testing the null hypothesis that additional factors have no effects on outcome. Our analysis yielded robust evidence (via χ2-test, p-values< 0.0001) supporting the significance of adding radiomic factors in each pair. Augmenting radiomics to include Smoke, ETOH, and T (the first pair) decreased the deviance by 27.68, indicating a significantly better fit of the larger model to the data. The larger model's AUC showed a noteworthy improvement over the smaller model, as depicted in Figure 8.

3.3 Risk stratification and prognostic ability

Patients were stratified into high- and low-end groups for recurrence-free survival based on the median value of two factors (Figure 9). The Kaplan-Meier curves provide compelling evidence of a significant difference in RFS between the high- and low-end groups (Log-rank p< 0.05). Furthermore, the AUCs of logistic regression, incorporating radiomic factors (0.93), corroborate the significant enhancement in discriminative power when compared to clinical factors-only models (Figure 8). Nomograms were

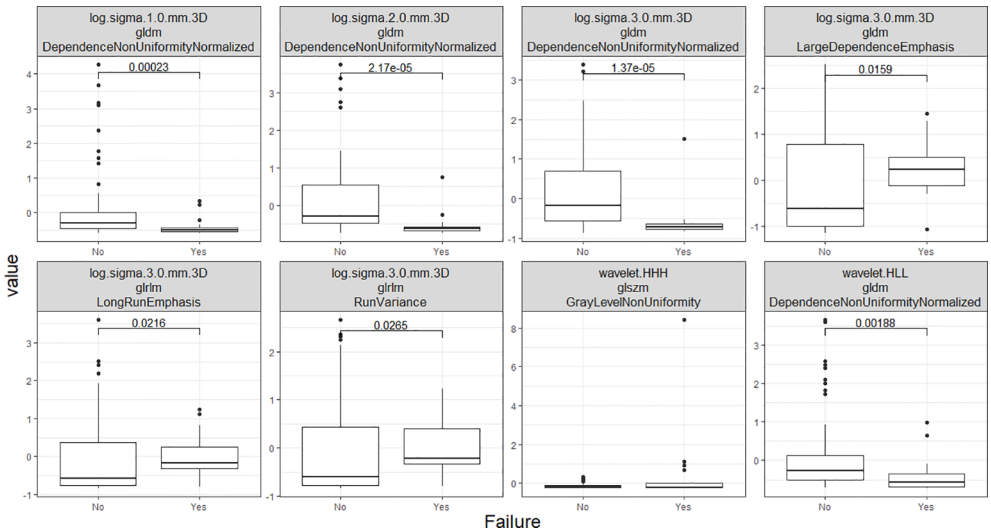


FIGURE 7
This set of box plots presents a comparative analysis of eight distinct radiomics features. The selection process involved repeated 5x10-fold cross-validation RFE.

TABLE 3 Logistic regression models with different factor inclusions.

							Training				Testing			
	Factor	Est.	se	z	*p*-value	AIC	ACC	AUC	Sens	Spec	ACC	AUC	Sens	Spec
1	Intercept	2.646	0.796	3.324	0.001	57	0.79	0.83	0.57	0.87	0.87	0.84	0.82	1.00
	LRE	1.323	0.429	3.086	0.002									
	LDE	-0.051	0.015	-3.486	0.000									
2	Intercept	0.679	0.481	1.411	0.158	80	0.75	0.46	0.23	0.93	0.60	0.67	0.45	1.00
	Non-Smoke	0.160	0.718	0.223	0.823									
	Non-ETOH	0.234	0.666	0.351	0.726									
	T2	-0.856	0.604	-1.418	0.156									
	T3	0.446	0.634	0.704	0.482									
	T4	0.976	0.648	1.505	0.132									
3	Intercept	2.360	0.869	2.715	0.007	58	0.80	0.86	0.57	0.87	0.87	0.86	0.82	0.87
	LRE	1.404	0.465	3.020	0.003									
	LDE	-0.053	0.016	-3.404	0.001									
	Non-Smoke	0.586	0.758	0.773	0.439									
4	Intercept	2.048	0.911	2.248	0.025	58	0.80	0.85	0.53	0.90	0.87	0.89	0.82	0.87
	LRE	1.600	0.532	3.008	0.003									
	LDE	-0.060	0.018	-3.357	0.001									
	Non-Smoke	0.191	0.825	0.232	0.817									
	Non-ETOH	1.125	0.851	1.323	0.186									
5	Intercept	2.076	1.008	2.060	0.039	63	0.76	0.83	0.45	0.87	0.87	0.89	0.82	0.87
	LRE	1.696	0.579	2.931	0.003									
	LDE	-0.063	0.019	-3.308	0.001									
	Non-Smoke	0.247	0.861	0.287	0.774									
	Non-ETOH	1.140	0.888	1.283	0.199									
	T2	0.355	0.765	0.464	0.643									
	T3	-0.161	0.833	-0.193	0.847									
	T4	0.897	0.793	1.131	0.258									

ACC measures overall correctness, AUC assesses discrimination ability, Sen measures the ability of the model to correctly identify positive, Spec measures the ability of the model to correctly identify negative, and AIC indicates how close fitted values to expected values. In this case, the 1st model (AIC=57) is considered more efficient in explaining the observed variation in the data than 2nd model.

TABLE 4 Analysis of deviance for various logistic regression models.

Model Comparison	Res.Df.	Res.Dev.	Df.	Dev.	p-value
Radiomics + Smoke + ETOH + T	70	56.74			
Smoke + ETOH + T	72	84.42	-2	-27.68	<0.0001
Radiomics + Smoke	74	60.47			

(Continued)

TABLE 4 Continued

Model Comparison	Res.Df.	Res.Dev.	Df.	Dev.	p-value
Smoke	76	89.33	-2	-28.86	<0.0001
Radiomics + ETOH	74	59.34			
ETOH	76	90.13	-2	-30.79	<0.0001
Radiomics + T	72	59.24			
T	74	84.78	-2	-25.53	<0.0001
Radiomics + ETOH + Smoke	73	58.99			
ETOH + Smoke	75	89.10	-2	-30.11	<0.0001
Radiomics + T + Smoke	71	58.60			
T + Smoke	73	84.65	-2	-26.04	<0.0001
Radiomics + T + ETOH	71	56.88			
T + ETOH	73	84.45	-2	-27.57	<0.0001
Radiomics + Smoke + ETOH + T	70	56.74			
Radiomics	75	61.55	-5	-4.81	0.4398

Comparisons assess the impact of radiomics inclusion. The first tests the superiority of the full model (radiomics and clinical) over the clinical model alone, while the rest evaluates the model with or without radiomics. Significant p-values favor the full model in all comparisons. The table details degrees of freedom (Res.Df.), residual deviance (Res.Dev.), changes in degrees of freedom (Df.), changes in deviance (Dev.), and associated p-values.

constructed with radiomic and clinical factors respectively in Figure 10. Notably, the addition of clinical features in the full model demonstrates minimal influence on the predicted RFS probability when compared to the radiomic feature-only model.

3.4 Radiomic features uncover hidden textural patterns

Tumor heterogeneity is widely acknowledged as a significant factor associated with tumor progression. The quantification of tumor heterogeneity has assumed a pivotal role in pathological assessments. Radiomic texture analysis presents distinct advantages, such as non-invasiveness and cost-effectiveness, compared to conventional pathological evaluations. Multiple studies (66–69) have underscored the prognostic potential of GLDM (Gray Level Dependence Matrix) and GLRLM (Gray Level Run Length Matrix) features in the evaluation of tumor progression. These features, GLDM and GLRLM, quantify the degree of local variation within an image (70). LRE (Long Run Emphasis) serves as a metric for assessing the distribution of long run lengths, with higher values indicative of longer run lengths and coarser structural textures. Conversely, LDE (Long Dependence Emphasis) quantifies the distribution of large dependencies, with elevated values denoting larger dependencies and more homogeneous textures. In light of the findings presented in Table 5, it is noteworthy that, when maintaining LDE at a constant value, each unit increment in LRE corresponds to a 79% (1-0.21) decrease in the odds of recurrence as opposed to non-recurrence. Conversely, when keeping LRE at a fixed value, every unit increasing in LDE results in a 5.8% increase in the odds of recurrence. It is essential to recognize that the estimate for the intercept represents the log odds of a patient with hypothetical zero values for LRE and LDE experiencing recurrence, which is calculated to be 0.095. This observation underscores a robust association between these two radiomic factors and the risk of locoregional recurrence.

Figure 11 compares two sets of images processed using the Laplacian of Gaussian (LoG) filter. The first three rows depict LoG-filtered results on an Oral Squamous Cell Carcinoma (OSCC) occurring in the tongue area, with varying sigma (σ) values. The last three rows display results for a non-OSCC area of the tongue. Lower sigma values highlight finer structures, while higher sigma values accentuate larger clusters in the tissue (71). We observed a pattern in the OSCC images: with increasing sigma, there is a reduction in highlights (white regions) in the filtered images, contrary to the non-OSCC images, which maintain a consistent level of highlights. This suggests that squamous cells in OSCC may be more homogeneous than normal cells. Furthermore, we observed circular artifacts in the OSCC images when filtered with larger sigma ($\sigma > 0.4$mm), whereas the normal set presents relatively random structures. to normal tissues. These findings support a fundamental histopathological principle: tumor tissues typically exhibit a more anaplastic and infiltrative pattern than normal tissues (72). This occurs because tumor cells grow unregulated and clonally, leading to the loss of normal differentiation and organization characteristic of healthy tissues.

4 Discussion

Several studies (57, 73, 74) have demonstrated the statistical significance of the discriminating ability of radiomic features. The

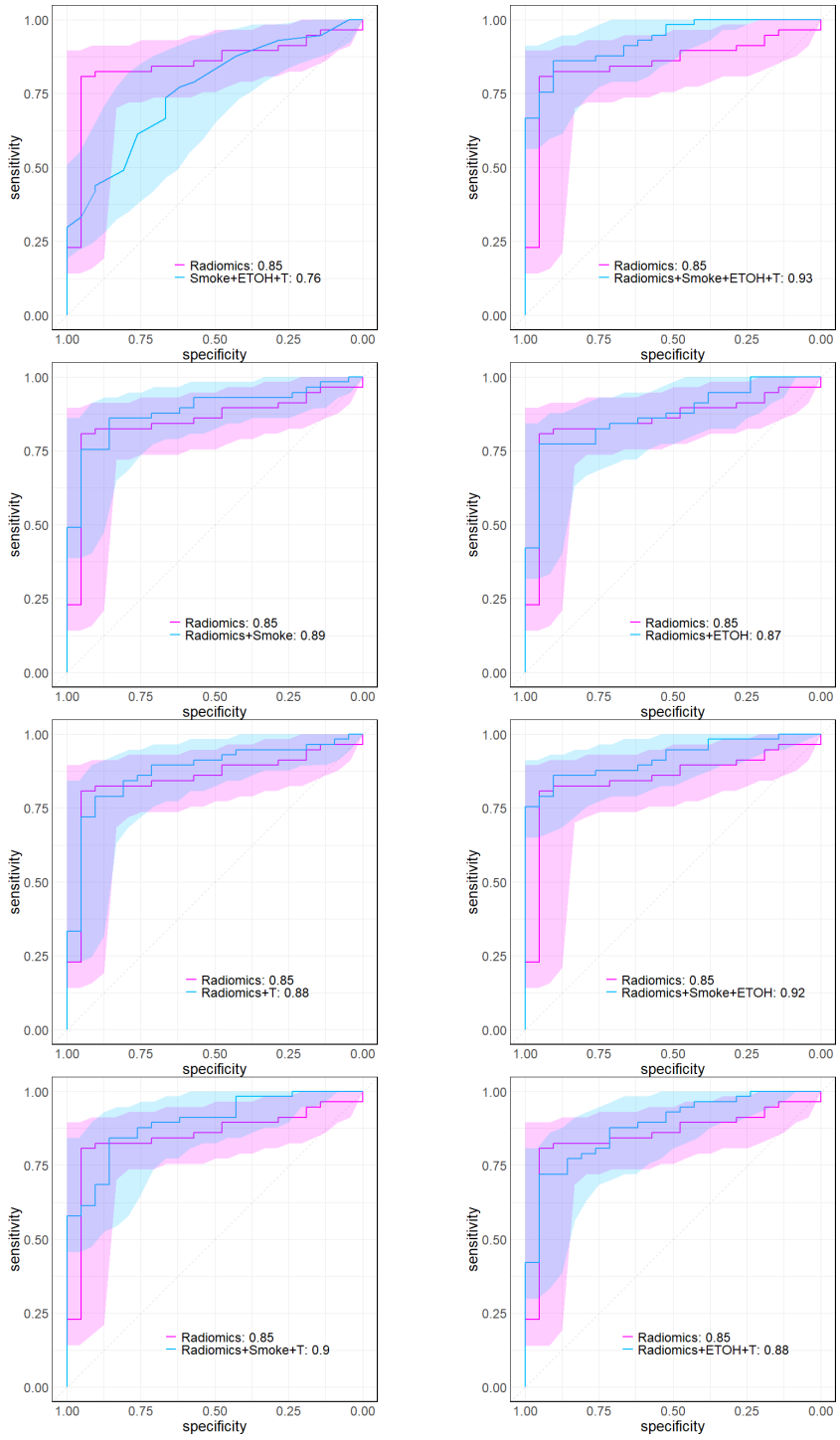


FIGURE 8
These plots illustrate the logistic ROC, delineating comparisons between the full model, which incorporates previously identified radiomic and clinical features, and alternative combinations. The numbers next to each model in the legend give the AUC.

Laplacian of an image highlights regions of rapid intensity change (75). LDE and LRE measure the distribution of low gray-level values, with a higher value indicating a greater concentration of low gray-level values in the tumor CT scan. SRLGLE measures the joint distribution of shorter run lengths with lower gray-level values despite the universal adoption of CT modality in OSCC diagnosis, automatic imaging prognostic evaluation is lacking and subjective. We present a fully automated prognostic evaluation tool to preoperatively detect locoregional failure in oral cavity cancer. The present study aimed to assess the prognostic capabilities of radiomic features in OSCC locoregional recurrence. Our findings demonstrate that analyzing radiomics from pre-treatment CT scans offers valuable insights into risk factors for locoregional failure and serves as prognostic

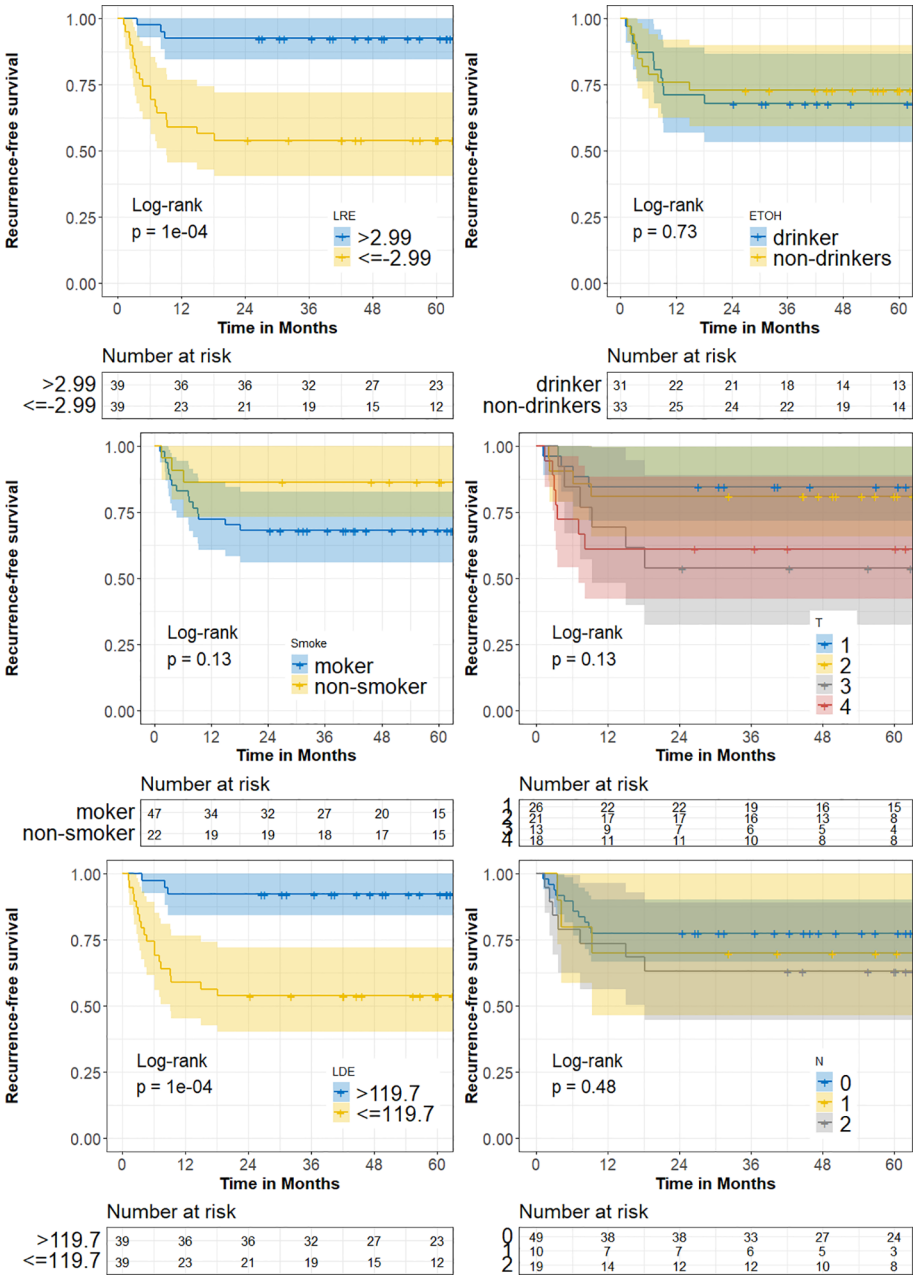


FIGURE 9
The Kaplan-Meier survival curves demonstrate a significant contrast (p = 1e-04) in recurrence-free survival (RFS) between high/low-end radiomic risk groups.

biomarkers in this patient population. Non-invasive risk factors play a crucial role in personalizing treatment planning, particularly in OSCC, due to the involvement of critical neck surgeries. It is well-known that neck surgeries potentially significantly impact the quality of a patient's life. Thoughtful treatment planning has the potential to mitigate the side effects of unnecessary neck surgery. Key findings of our study include two significant radiomic risk factors: Large Dependence Emphasis (LDE) of the Gray Level Dependence Matrix (GLDM) and Long Run Emphasis (LRE) of the Gray Level Run Length Matrix (GLRLM) of the 3D Laplacian of Gaussian (LoG σ=3) filtered ROIs.

The AUC showed a stable and approximate value of 0.8 with a sensitivity 0.8 and specificity of 0.8 at the optimal threshold, indicating good prognosis accuracy of the classifier. These results highlight the potential of radiomic features, as a biomarker indicator for treatment failure prognosis. Our study standardized voxel spacing in CT images across patients for precise feature

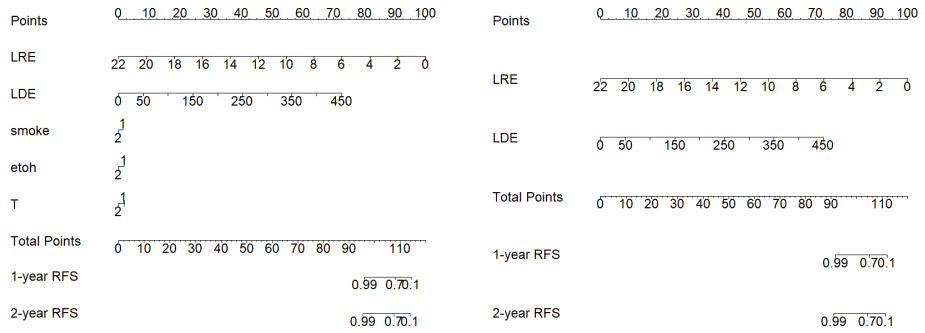


FIGURE 10
Side-by-side comparison of nomograms illustrating RFS probability predictions for the cohort. On the left, the "Full Model" includes both radiomic and clinical factors, while on the right, the "radiomics-Only Model" consists of two identified radiomic risk factors exclusively.

TABLE 5 Summary of effects in model with entire cohort.

Coef.	Estimate	95% CI		OR	95% OR CI		p-value
Intercept	-2.355	-3.759	-0.952	0.095	0.019	0.331	0.001
LRE	-1.557	-2.404	-0.709	0.211	0.076	0.433	0.000
LDE	0.056	0.028	0.084	1.058	1.032	1.093	0.000

SE, standard error; z, z-value; CI, Confidence Interval.

calculation and applied gray-level normalization to enhance feature comparability. These steps are crucial for consistent radiomics analysis. Combining Correlation Analysis, Recursive Feature Selection, and Logistic Regression Best Subset Selection, our feature selection process effectively reduced feature space dimensionality while retaining critical prognostic information. This approach helps mitigate bias, overfitting, and multicollinearity in high-throughput data analysis.

The cost of missing a positive diagnosis (Type 2 error) is often higher than false alarms. On the contrary, since neck dissection significantly decreases the quality of life, reducing the false positive rate (Type 1 error) shall be necessary. In fact, study (76) demonstrate that ROC plots in the context of imbalanced datasets can be deceptive. Therefore, our modeling emphasis was placed on increasing sensitivity and specificity with due consideration to AUC. The threshold for positive event classification plays a pivotal role in predictive accuracy. While a threshold of 0.5 is commonly employed in default, this value is often suboptimal for practical applications in real-world studies, particularly in clinical settings where the distribution of positive cases may have an inherent prevalence, thereby elevating the risks of Type I and Type II errors. Both types of errors are of concern in the study since both overtreatment and undertreatment may lead to escalating healthcare costs and potential harm to patients. To mitigate these risks, we propose adopting a threshold that aligns with the natural prevalence of our cohort, specifically a value of 0.28 for this cohort, for final classification. This calibrated threshold aims to optimize two key metrics: high sensitivity, crucial for minimizing Type II errors and thereby maximizing the identification of LR, and high precision, vital for minimizing Type I errors to reduce false alarms. A number of studies (77, 78) have shown that the chance of an OSCC postoperative locoregional disease being diagnosed positive (Sensitivity) after surgery is only 29%. Our approach demonstrated a noteworthy testing AUC of 0.84, prioritizing both high sensitivity (0.82) and specificity (1). This significantly reduces Type I and Type II errors in post-treatment disease screening tests, effectively minimizing overtreatment and undertreatment.

Our study underscores radiomics' promise in OSCC classification, yet it's crucial to consider its limitations. The small sample size and the classification study's nature might influence our model's radiomics feature stability. For a low-biased, variance classification model with two effects, at least 20 events per training set are advisable, necessitating 27 events for a training set comprising 75% of the sample. This requirement could limit our model's flexibility, potentially impacting the diagnostic capability of the radiomics. Moreover, our analysis only involved radiomics features from CT imaging. Future research should explore features from various imaging techniques, like CT and MRI, to heighten prediction precision. Notably, the observed correlation between certain radiomics features and overall survival hints that these features may mirror the tumors' molecular traits. Upcoming studies should integrate genetic data, such as TP53 (13) mutations and P16 overexpression (28), with radiomics to more comprehensively characterize head and neck squamous cell carcinoma and offer a non-invasive, multimodal approach to OSCC outcome prediction. It is important to acknowledge the

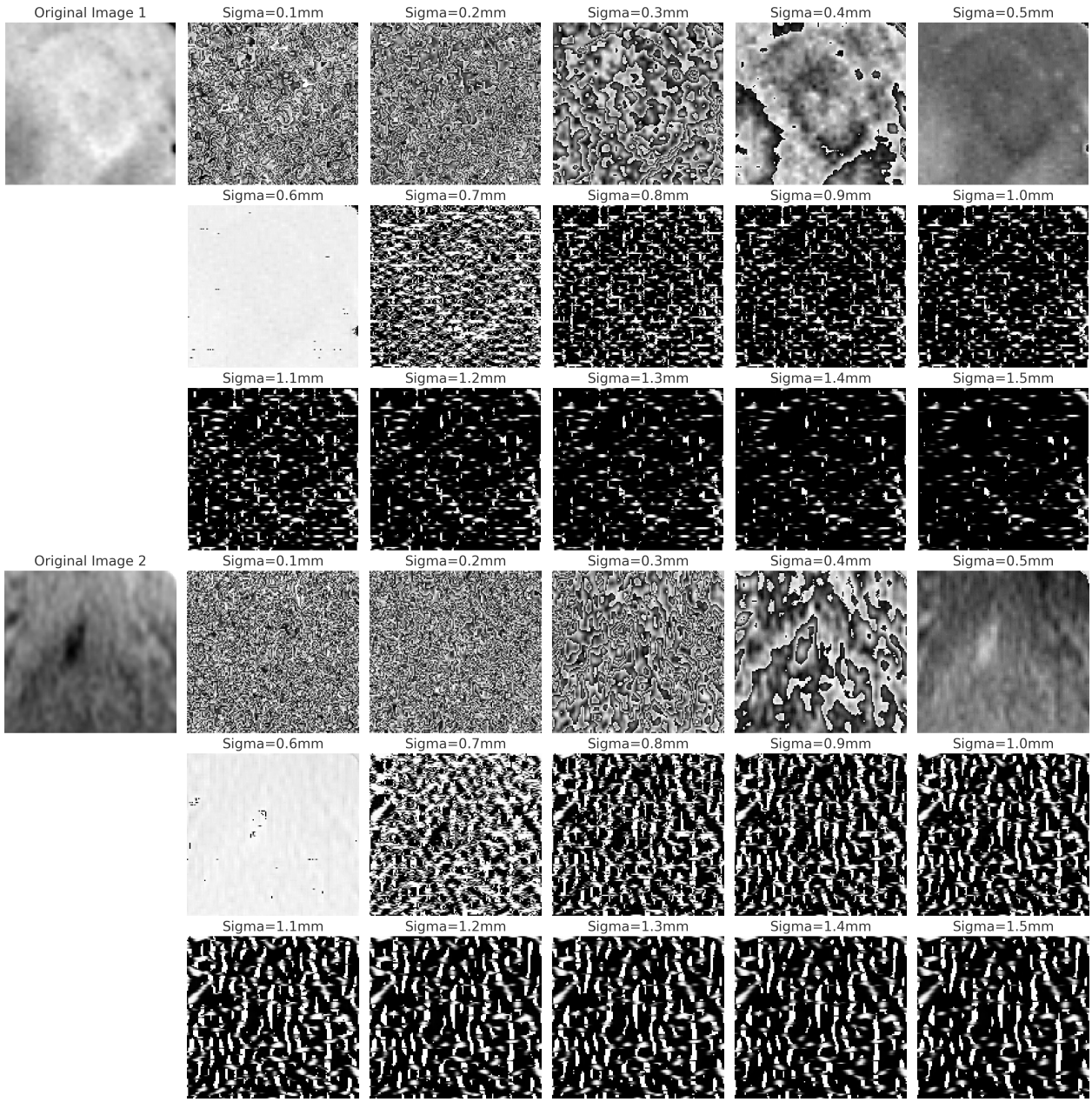


FIGURE 11
The figure displays two sets of LoG filtered images. The first three rows show OSCC images in the tongue region at varying sigma values, illustrating changes in the primary ROI. The last three rows depict normal tissue in the same region, also filtered at corresponding sigma values, to highlight contrasts between OSCC and normal textures.

challenges posed by the sensitivity of the data involved. To our knowledge, there is no public dataset available that could be directly applied to our validation needs. As a result, we are actively seeking to collaborate with multiple institutions to gather data for validation purposes, aiming to mitigate the limitation and ensure the robustness of findings.

In conclusion, our study demonstrated the potential of radiomics as an effective tool to predict treatment response in OSCC patients. Incorporating radiomics analysis into clinical practice could improve decision support and enhance patient stratification, reducing both over-treatment and under-treatment to improve outcomes. Moreover, processing the ROI at the level of small tiles provides an additional non-invasive avenue for assessing the spatial heterogeneity within the tumor. The findings from the study pave the way for future investigations through a larger clinical trial to further evaluate the clinical efficacy of radiomics biomarkers for overall survival prediction for OSCC patients.

Data availability statement

The datasets presented in this article are not readily available due to the governance of patient data by the Institutional Review Board. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by University of Maryland School of Medicine's Institutional Review Board (approval number 00007145). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

XL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. GA: Methodology, Writing – review & editing. JM: Data curation, Resources, Writing – review & editing. JC: Data curation, Resources, Writing – review & editing. LS: Data curation, Resources, Writing – review & editing. PT: Writing – review & editing, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision. RM: Resources, Writing – review & editing. DG: Conceptualization, Investigation, Resources, Supervision, Writing – review & editing, Data curation. LR: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. XL and LR was supported by the National Institutes of health (NIH) grants R01EB032680, R01CA279013, R01EB028324, and U54CA273956. DG was supported by a Research Scholarship Grant, RSG-21-020-01-MPC from the American Cancer Society, and R01DE027809 from the National Institute of Health.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1380599/full#supplementary-material

References

1. da Silva SD, Hier M, Mlynarek A, Kowalski LP, Alaoui-Jamali MA. Recurrent oral cancer: current and emerging therapeutic approaches. *Front Pharmacol*. (2012) 3:149. doi: 10.3389/fphar.2012.00149

2. Yuen APW, Wei WI, Wong YM, Tang KC. Elective neck dissection versus observation in the treatment of early oral tongue carcinoma. *Head Neck: J Sci Specialties Head Neck*. (1997) 19:583–8. doi: 10.1002/(ISSN)1097-0347

3. Contrera KJ, Huang AT, Shenson JA, Tang C, Roberts D, Myers JN, et al. Primary and recurrent regional metastases for lateralized oral cavity squamous cell carcinoma. *Surg Oncol*. (2022) 44:101804. doi: 10.1016/j.suronc.2022.101804

4. Licitra L, Locati L, Bossi P. Head and neck cancer. *Ann Oncol*. (2004) 15:iv267–73. doi: 10.1093/annonc/mdh937

5. Duprez F, Berwouts D, De Neve W, Bonte K, Boterberg T, Deron P, et al. Distant metastases in head and neck cancer. *Head Neck*. (2017) 39:1733–43. doi: 10.1002/hed.v39.9

6. Woolgar JA, et al. Survival, metastasis and recurrence of oral cancer in relation to pathological features. *Ann R Coll Surgeons Engl*. (1995) 77:325.

7. Woolgar JA, Rogers S, West CR, Errington RD, Brown JS, Vaughan ED. Survival and patterns of recurrence in 200 oral cancer patients treated by radical surgery and neck dissection. *Oral Oncol*. (1999) 35:257–65. doi: 10.1016/S1368-8375(98)00113-4

8. Leemans CR, Tiwari R, Nauta JJ, Waal IVD, Snow GB. Recurrence at the primary site in head and neck cancer and the significance of neck lymph node metastases as a prognostic factor. *Cancer*. (1994) 73:187–90. doi: 10.1002/(ISSN)1097-0142

9. Fuller C, Camilon R, Nguyen S, Jennings J, Day T, Gillespie MB. Adjunctive diagnostic techniques for oral lesions of unknown Malignant potential: Systematic review with meta-analysis. *Head Neck*. (2015) 37:755–62. doi: 10.1002/hed.v37.5

10. Woolgar J. T2 carcinoma of the tongue: the histopathologist's perspective. *Br J Oral Maxillofac Surg*. (1999) 37:187–93. doi: 10.1054/bjom.1999.0034

11. Shingaki S, Suzuki I, Nakajima T, Kawasaki T. Evaluation of histopathologic parameters in predicting cervical lymph node metastasis of oral and oropharyngeal carcinomas. *Oral Surg Oral Med Oral Pathol*. (1988) 66:683–8. doi: 10.1016/0030-4220(88)90318-0

12. Crissman JD, Liu W, Gluckman JL, Cummings G. Prognostic value of histopathologic parameters in squamous cell carcinoma of the oropharynx. *Cancer*. (1984) 54:2995–3001. doi: 10.1002/(ISSN)1097-0142

13. Fukano H, Matsuura H, Hasegawa Y, Nakamura S. Depth of invasion as a predictive factor for cervical lymph node metastasis in tongue carcinoma. *Head Neck: J Sci Specialties Head Neck*. (1997) 19:205–10. doi: 10.1002/(ISSN)1097-0347

14. Spiro RH, Huvos AG, Wong GY, Spiro JD, Gnecco CA, Strong EW. Predictive value of tumor thickness in squamous carcinoma confined to the

tongue and floor of the mouth. *Am J Surg*. (1986) 152:345–50. doi: 10.1016/0002-9610(86)90302-8

15. Jones KR, Lodge-Rigal RD, Reddick RL, Tudor GE, Shockley WW. Prognostic factors in the recurrence of stage I and II squamous cell cancer of the oral cavity. *Arch Otolaryngology–Head Neck Surg*. (1992) 118:483–5. doi: 10.1001/archotol.1992.01880050029006

16. Menini M, De Giovanni E, Bagnasco F, Delucchi F, Pera F, Baldi D. Salivary micro-RNA and oral squamous cell carcinoma: A systematic review. *J Pers Med*. (2021) 11:101. doi: 10.3390/jpm11020101

17. Liu C-J, Lin S-C, Yang C-C, Cheng H-W, Chang K-W. Exploiting salivary miR-31 as a clinical biomarker of oral squamous cell carcinoma. *Head Neck*. (2012) 34:219–24. doi: 10.1002/hed.21713

18. Momen-Heravi F, Trachtenberg A, Kuo W, Cheng Y. Genomewide study of salivary microRNAs for detection of oral cancer. *J Dental Res*. (2014) 93:86S–93S. doi: 10.1177/0022034514531018

19. Carnielli CM, Macedo CC, De Rossi T, Granato DC, Rivera C, Domingues RR, et al. Combining discovery and targeted proteomics reveals a prognostic signature in oral cancer. *Nat Commun*. (2018) 9:3598. doi: 10.1038/s41467-018-05696-2

20. Goto H, Kawano K, Kobayashi I, Sakai H, Yanagisawa S. Expression of cyclin D1 and GSK-3β and their predictive value of prognosis in squamous cell carcinomas of the tongue. *Oral Oncol*. (2002) 38:549–56. doi: 10.1016/S1368-8375(01)00121-X

21. Myo K, Uzawa N, Miyamoto R, Sonoda I, Yuki Y, Amagasa T. Cyclin D1 gene numerical aberration is a predictive marker for occult cervical lymph node metastasis in TNM Stage I and II squamous cell carcinoma of the oral cavity. *Cancer: Interdiscip Int J Am Cancer Soc*. (2005) 104:2709–16. doi: 10.1002/cncr.21491

22. Chien H-T, Cheng S-D, Liao C-T, Wang H-M, Huang S-F. Amplification of the EGFR and CCND1 are coordinated and play important roles in the progression of oral squamous cell carcinomas. *Cancers*. (2019) 11:760. doi: 10.3390/cancers11060760

23. Lee LT, Wong YK, Hsiao HY, Wang YW, Chan MY, Chang KW. Evaluation of saliva and plasma cytokine biomarkers in patients with oral squamous cell carcinoma. *Int J Oral Maxillofac Surg*. (2018) 47:699–707. doi: 10.1016/j.ijom.2017.09.016

24. Nagata M, Noman AA, Suzuki K, Kurita H, Ohnishi M, Ohyama T, et al. ITGA3 and ITGB4 expression biomarkers estimate the risks of locoregional and hematogenous dissemination of oral squamous cell carcinoma. *BMC Cancer*. (2013) 13:1–9. doi: 10.1186/1471-2407-13-410

25. Kato H, Tamai K, Morioka H, Nagai M, Nagaya T, Torigoe T. Tumor-antigen TA-4 in the detection of recurrence in cervical squamous cell carcinoma. *Cancer*. (1984) 54:1544–6. doi: 10.1002/(ISSN)1097-0142

26. Takes RP, Baatenburg de Jong RJ, Keuning J, Hermans J, Schuuring E, Van Krieken HJ. Protein expression of cancer associated genes: biopsy material compared to resection material in laryngeal cancer. *Anticancer Res*. (1998) 18:4787–91.

27. Bayramoglu N, Kaakinen M, Eklund L, Heikkila J. Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks, in: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. (Venice, Italy: IEEE International Conference on Computer Vision Workshops (ICCVW)). (2017). pp. 64–71.

28. Rana A, Yauney G, Lowe A, Shah P. Computational histological staining and destaining of prostate core biopsy RGB images with generative adversarial neural networks, in: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. (Orlando, FL: IEEE International Conference on Machine Learning and Applications (ICMLA)). (2018). pp. 828–34.

29. Redston M, Noffsinger A, Kim A, Akarca FG, Rara M, Stapleton D, et al. Abnormal TP53 predicts risk of progression in patients with Barrett's esophagus regardless of a diagnosis of dysplasia. *Gastroenterology*. (2022) 162:468–81. doi: 10.1053/j.gastro.2021.10.038

30. Flaherty KT, Gray R, Chen A, Li S, Patton D, Hamilton SR, et al. The molecular analysis for therapy choice (NCI-MATCH) trial: lessons for genomic trial design. *JNCI: J Natl Cancer Institute*. (2020) 112:1021–9. doi: 10.1093/jnci/djz245

31. Blucher AS, Mills GB, Tsang YH. Precision oncology for breast cancer through clinical trials. *Clin Exp Metastasis*. (2022) 39:71–8. doi: 10.1007/s10585-021-10092-0

32. Prudente S, Dallapiccola B, Pellegrini F, Doria A, Trischitta V. Genetic prediction of common diseases. Still no help for the clinical diabetologist! *Nutrition Metab Cardiovasc Dis*. (2012) 22:929–36. doi: 10.1016/j.numecd.2012.04.010

33. Martens RM, Koopman T, Noij DP, Pfaehler E, Übelhör C, Sharma S, et al. Predictive value of quantitative 18F-FDG-PET radiomics analysis in patients with head and neck squamous cell carcinoma. *EJNMMI Res*. (2020) 10:1–15. doi: 10.1186/s13550-020-00686-2

34. Salmanpour MR, Hosseinzadeh M, Rezaeijo SM, Rahmim A. Fusion-based tensor radiomics using reproducible features: application to survival prediction in head and neck cancer. *Comput Methods Programs Biomedicine*. (2023) 240:107714. doi: 10.1016/j.cmpb.2023.107714

35. Nie K, Hu P, Zheng J, Zhang Y, Yang P, Jabbour SK, et al. Incremental value of radiomics in 5-year overall survival prediction for stage II–III rectal cancer. *Front Oncol*. (2022) 12:779030. doi: 10.3389/fonc.2022.779030

36. Zheng J, Chakraborty J, Chapman WC, Gerst S, Gonen M, Pak LM, et al. Preoperative prediction of microvascular invasion in hepatocellular carcinoma using quantitative image analysis. *J Am Coll Surgeons*. (2017) 225:778–88. doi: 10.1016/j.jamcollsurg.2017.09.003

37. Attiyeh MA, Chakraborty J, Gazit L, Langdon-Embry L, Gonen M, Balachandran VP, et al. Preoperative risk prediction for intraductal papillary mucinous neoplasms by quantitative CT image analysis. *Hpb*. (2019) 21:212–8. doi: 10.1016/j.hpb.2018.07.016

38. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128

39. Yu Y, Tan Y, Xie C, Hu Q, Ouyang J, Chen Y, et al. Development and validation of a preoperative magnetic resonance imaging radiomics–based signature to predict axillary lymph node metastasis and disease-free survival in patients with early-stage breast cancer. *JAMA network Open*. (2020) 3:e2028086. doi: 10.1001/jamanetworkopen.2020.28086

40. Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer. *EBioMedicine*. (2018) 34:76–84. doi: 10.1016/j.ebiom.2018.07.029

41. Zhang Z, Huang M, Jiang Z, Chang Y, Torok J, Yin FF. 4D radiomics: impact of 4D-CBCT image quality on radiomic analysis. *Phys Med Biol*. (2021) 66(4):045023.

42. Zhang Z, Huang M, Jiang Z, Chang Y, Lu K, Yin FF. Patient-specific deep learning model to enhance 4D-CBCT image for radiomics analysis. *Phys Med Biol*. (2022) 67(8):085003.

43. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

44. Aiken AH, Rath TJ, Anzai Y, Branstetter BF, Hoang JK, Wiggins RH, et al. ACR Neck Imaging Reporting and Data Systems (NI-RADS): a white paper of the ACR NI-RADS Committee. *J Am Coll Radiol*. (2018) 15:1097–108. doi: 10.1016/j.jacr.2018.05.006

45. Dinkelborg P, Ro SR, Shnayien S, Schaafs LA, Koerdt S, Kreutzer K, et al. Retrospective evaluation of NI-RADS for detecting postsurgical recurrence of oral squamous cell carcinoma on surveillance CT or MRI. *Am J Roentgenol*. (2021) 217:198–206. doi: 10.2214/AJR.20.24209

46. Hsu D, Chokshi FH, Hudgins PA, Kundu S, Beitler JJ, Patel MR, et al. Predictive value of first posttreatment imaging using standardized reporting in head and neck cancer. *Otolaryngology–Head Neck Surg*. (2019) 161:978–85. doi: 10.1177/0194599819865235

47. Vallieres M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep*. (2017) 7:10117. doi: 10.1038/s41598-017-10371-5

48. Ling X, Alexander GS, Molitoris J, Choi J, Schumaker L, Mehra R, et al. Identification of CT-based non-invasive radiomic biomarkers for overall survival prediction in oral cavity squamous cell carcinoma. *Sci Rep*. (2023) 13:21774. doi: 10.1038/s41598-023-48048-x

49. Shafiq-ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. (2017) 44:1050–62. doi: 10.1002/mp.12123

50. Bagher-Ebadian H, Siddiqui F, Liu C, Movsas B, Chetty IJ. On the impact of smoothing and noise on robustness of CT and CBCT radiomics features for patients with head and neck cancers. *Med Phys*. (2017) 44:1755–70. doi: 10.1002/mp.12188

51. Hosseini SA, Shiri I, Hajianfar G, Ghafarian P, Karam MB, Ay MR. The impact of preprocessing on the PET-CT radiomics features in non-small cell lung cancer. *Front Biomed Technol*. (2021) 8:261–72. doi: 10.18502/fbt.v8i4.7754

52. Thévenaz P, Blu T, Unser M. Image interpolation and resampling. *Handb Med imaging Process Anal*. (2000) 1:393–420.

53. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic resonance Imaging*. (2004) 22:81–91. doi: 10.1016/j.mri.2003.09.001

54. Giannini V, Panic J, Regge D, Balestra G, Rosati S. Could normalization improve robustness of abdominal MRI radiomic features? *Biomed Phys Eng Express*. (2023) 9:055002. doi: 10.1088/2057-1976/ace4ce

55. Larue RT, van Timmeren JE, de Jong EE, Feliciani G, Leijenaar RT, Schreurs WM, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol*. (2017) 56:1544–53. doi: 10.1080/0284186X.2017.1351624

56. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295:328–38. doi: 10.1148/radiol.2020191145

57. Galloway MM. Texture analysis using gray level run lengths. *Comput Graphics image Process*. (1975) 4:172–9. doi: 10.1016/S0146-664X(75)80008-6

58. Thibault G, Fertil B, Navarro C. Texture indexes and gray level size zone matrix: application to cell nuclei classification, in: International Conference on Pattern Recognition and Information Processing (PRIP'09), Minsk, Belarus, (2009). pp. 140–5.

59. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. *Comput Vision Graphics Image Process*. (1983) 23:341–52. doi: 10.1016/0734-189X(83)90032-4

60. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans systems man Cybernetics*. (1989) 19:1264–74. doi: 10.1109/21.44046

61. Menard S. *Applied Logistic Regression Analysis*. London UK: Sage (2002). doi: 10.4135/9781412983433

62. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. (1996) 49:1373–9. doi: 10.1016/S0895-4356(96)00236-3

63. Ling X, Brooks JP. L1-norm regularized L1-norm best-fit line problem. (2020).

64. Ling X, Bui A, Brooks P. Kernel l1-norm principal component analysis for denoising. *Optimization Lett*. (2023). doi: 10.1007/s11590-023-02051-3

65. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn*. (2002) 46:389–422. doi: 10.1023/A:1012487302797

66. Kim H, Park J, Kim Y, Kim K. GLRLM feature by using CT texture analysis is prognostic factor, evaluation of pancreatic cancer heterogeneity. *Ann Oncol*. (2018) 29: ix63–4. doi: 10.1093/annonc/mdy432.053

67. Novitasari DCR, Lubab A, Sawiji A, Asyhar AH. Application of feature extraction for breast cancer using one order statistic, GLCM, GLRLM, and GLDM. *Adv Science Technol Eng Syst J (ASTESJ)*. (2019) 4:115–20. doi: 10.25046/astesj

68. Filipczuk P, Fevens T, Krzyżak A, Obuchowicz A. GLCM and GLRLM based texture features for computer-aided breast cancer diagnosis. *J Med Inf Technol*. (2012) 19.

69. Durgamahanthi V, Anita Christaline J, Shirly Edward A. GLCM and GLRLM based texture analysis: application to brain cancer diagnosis using histopathology images. In: *Intelligent Computing and Applications: Proceedings of ICICA 2019*. Singapore: Springer (2021). p. 691–706.

70. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Trans systems man cybernetics*. (1973) 6:610–21. doi: 10.1109/TSMC.1973.4309314

71. Depeursinge A, Andrearczyk V, Whybra P, Griethuysen JV, Müller H, Schaer R, et al. The image biomarker standardization initiative II: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2023) 295:328–38.

72. Troncone G, Vigliar E. Histopathology of the Tumors. In: *Practical Medical Oncology Textbook* (Switzerland: Springer Nature) (2021). p. 33–41.

73. Xie D, Wang TT, Huang SJ, Deng JJ, Ren YJ, Yang Y, et al. Radiomics nomogram for prediction disease-free survival and adjuvant chemotherapy benefits in patients with resected stage I lung adenocarcinoma. *Trans Lung Cancer Res*. (2020) 9:1112. doi: 10.21037/tlcr

74. Mazzaschi G, Quaini F, Milanese G, Madeddu D, Bocchialini G, Ampollini L, et al. Differentially regulated high-throughput CT imaging features correlate to distinct tumor immune contextures portraying a radiomic signature with prognostic impact on surgically resected NSCLC. *Ann Oncol*. (2019) 30:ii1. doi: 10.1093/annonc/mdz072.001

75. Marr D. Representing the image: zero-crossings and the raw primal sketch. In: *Vision*. WH Freeman, San Francisco, CA (1982). p. 54–68.

76. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One*. (2015) 10:e0118432. doi: 10.1371/journal.pone.0118432

77. Jones A, Phillips D, Helliwell T, Roland N. Occult node metastases in head and neck squamous carcinoma. *Eur Arch oto-rhino-laryngology*. (1993) 250:446–9. doi: 10.1007/BF00181087

78. Roepman P, Wessels LF, Kettelarij N, Kemmeren P, Miles AJ, Lijnzaad P, et al. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet*. (2005) 37:182–6. doi: 10.1038/ng1502

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 30 April 2024
DOI 10.3389/fonc.2024.1363756

Check for updates

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Andrea Bianconi,
University Hospital of the City of Health and Science of Turin, Italy
Kandala N. V. P. S. Rajesh,
VIT-AP University, India

*CORRESPONDENCE
Jingjing Li
63931204@qq.com
Jun Cai
caijun0540@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 31 December 2023
ACCEPTED 15 April 2024
PUBLISHED 30 April 2024

CITATION
Wang Z, He C, Hu Y, Luo H, Li C, Wu X, Zhang Y, Li J and Cai J (2024) A hybrid deep learning scheme for MRI-based preliminary multiclassification diagnosis of primary brain tumors.
Front. Oncol. 14:1363756.
doi: 10.3389/fonc.2024.1363756

COPYRIGHT
© 2024 Wang, He, Hu, Luo, Li, Wu, Zhang, Li 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.

A hybrid deep learning scheme for MRI-based preliminary multiclassification diagnosis of primary brain tumors

Zhichao Wang[1,2†], Chuchu He[1†], Yan Hu[1,2], Haifeng Luo[1], Chao Li[1], Xiandong Wu[1], Yang Zhang[1], Jingjing Li[1*] and Jun Cai[1*]

[1]Department of Oncology, The First Affiliated Hospital of Yangtze University, Jingzhou, Hubei, China, [2]Hubei Key Laboratory of Precision Radiation Oncology, Wuhan, China

Objectives: The diagnosis and treatment of brain tumors have greatly benefited from extensive research in traditional radiomics, leading to improved efficiency for clinicians. With the rapid development of cutting-edge technologies, especially deep learning, further improvements in accuracy and automation are expected. In this study, we explored a hybrid deep learning scheme that integrates several advanced techniques to achieve reliable diagnosis of primary brain tumors with enhanced classification performance and interpretability.

Methods: This study retrospectively included 230 patients with primary brain tumors, including 97 meningiomas, 66 gliomas and 67 pituitary tumors, from the First Affiliated Hospital of Yangtze University. The effectiveness of the proposed scheme was validated by the included data and a commonly used data. Based on super-resolution reconstruction and dynamic learning rate annealing strategies, we compared the classification results of several deep learning models. The multi-classification performance was further improved by combining feature transfer and machine learning. Classification performance metrics included accuracy (ACC), area under the curve (AUC), sensitivity (SEN), and specificity (SPE).

Results: In the deep learning tests conducted on two datasets, the DenseNet121 model achieved the highest classification performance, with five-test accuracies of 0.989 $\pm$ 0.006 and 0.967 $\pm$ 0.013, and AUCs of 0.999 $\pm$ 0.001 and 0.994 $\pm$ 0.005, respectively. In the hybrid deep learning tests, LightGBM, a promising classifier, achieved accuracies of 0.989 and 0.984, which were improved from the original deep learning scheme of 0.987 and 0.965. Sensitivities for both datasets were 0.985, specificities were 0.988 and 0.984, respectively, and relatively desirable receiver operating characteristic (ROC) curves were obtained. In addition, model visualization studies further verified the reliability and interpretability of the results.

Conclusions: These results illustrated that deep learning models combining several advanced technologies can reliably improve the performance, automation, and interpretability of primary brain tumor diagnosis, which is crucial for further brain tumor diagnostic research and individualized treatment.

KEYWORDS

brain tumor classification, MRI images, deep learning, transfer learning, model interpretability

1 Introduction

Brain and central nervous system (CNS) tumors are among the most deadly cancers and have a high incidence. In the United States, approximately 80,000 people were diagnosed with brain or CNS tumors in 2021, and 18,600 died from these diseases (1). Common brain tumors include gliomas, meningiomas, and pituitary tumors (approximately 23%, 38%, and 17% of primary brain and CNS tumors, respectively) (2, 3). Primary intracranial tumors arise from various sites, including brain tissue, meninges, pituitary gland, cranial nerves, and vascular tissue. Available treatment options include surgical resection, radiotherapy, and chemotherapy. Therefore, accurate early diagnosis is essential for individualized treatment and prognostic assessment.

Magnetic resonance imaging (MRI) is a widely employed technique for the preliminary diagnosis of brain tumors, which can provide clear visualization of the nervous system structure and local lesions. Clinical application of MRI-based manual diagnosis can be influenced by professional level, work pressure, and degree of automation. In recent years, artificial intelligence has achieved significant progress in medicine (4, 5). Numerous studies have investigated the potential of machine learning combined with radiomics in brain tumor detection, molecular and genetic diagnosis (6–8). Nevertheless, further improvements are needed regarding the level of automation, reproducibility, and feature extraction performance of machine learning in radiomics to address the limitations associated with its inherent flaws (9–11).

As deep learning has demonstrated powerful adaptive feature extraction and end-to-end advantages in various fields, intelligent tumor diagnosis has also been widely researched and clinical application. Afshar analyzed the advantages of deep learning-based radiomics, such as freedom from prior knowledge and target area outlining, and end-to-end training (11). Lao validated the potential of deep learning in feature extraction and overall survival prediction based on 112 glioma patients (12). However, while acknowledging the significance of preliminary brain tumors diagnosis, the recent report highlighted the presence of the Smart Hans phenomenon in automated classification studies, where the model achieved better results without specifically focusing on the tumor region (13). This study revealed this previously overlooked bias, and provided valuable guidance for subsequent research. On one hand, deep learning models integrating multiple cutting-edge technologies can adaptively extract local information and assign appropriate weights, showing better performance than undifferentiated manual feature extraction based on the entire slice. On the other hand, the model visualization is crucial for interpreting whether it indeed focuses on the tumor area, which is an important verification of diagnostic reliability.

Therefore, this study proposed a hybrid deep learning scheme that integrated several advanced technologies for automated preliminary diagnosis of tumors. By focusing on the study of primary brain tumors, this study provided an important foundation for further extensions, such as brain metastasis prediction and pathological classification. The main contributions of this work are summarized as follows:

- In model construction, super-resolution reconstruction and dynamic learning rate strategies were applied to improve image quality and training efficiency.
- Based on the advantages of deep learning (DL) in feature extraction, machine learning models were further combined to improve classification performance.
- To assess the generalization performance and alleviate the interpretability problem, this study utilized t-SNE and Score-Grad techniques and verified the effectiveness of the scheme based on our institute and public datasets.

The remaining parts are organized as follows. Section 2 (Materials and methods) introduces the patient population and methodologies employed in the proposed scheme. The datasets and results of the experiments are presented in Section 3 (Results). Based on the displayed results, Section 4 (Discussion) analyzes the performance of the proposed scheme in detail. Finally, conclusions are summarized in Section 5 (Conclusions).

2 Materials and methods

2.1 Patient population

This retrospective study was approved by the ethics committees of the First Affiliated Hospital of Yangtze University, and informed consent was waived. Patients were enrolled based on the following

criteria: 1) available postoperative pathological diagnosis results; 2) MRI examinations performed in our hospital within 2 weeks before surgery; 3) available medical records. Exclusion criteria were as follows: 1) history of preoperative treatment (radiation, chemotherapy, or other treatments); 2) unavailable contrast-enhanced T1-weighted sequence; 3) presence of MRI artifacts or tumors too small to seriously affect tumor imaging.

2.2 Image acquisition and preprocessing

Contrast-enhanced MRI scans were performed at our institution using two 1.5 T scanners (Philips prodiva) and one 3.0 T scanner. MRI examinations included T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), fluid-attenuated inversion recovery (Flair), diffusion-weighted imaging (DWI), and contrast-enhanced T1WI (CE-T1WI). The CE-T1WI sequence included transverse, sagittal, and coronal views, and several slices near the largest tumor level were acquired in each view. In addition, multi-sequence modal analysis can be considered in further studies. The scan parameter settings were as follows: a repetition time (TR) and echo time (TE) of 5.5ms and 2.4ms, a pixel matrix of 256*236, a slice thickness and slice gap of 1mm and 0.5mm, and a deflection angle of 15 degrees.

In addition, image resolution can be affected by factors such as hardware configuration, acquisition time, and radiation exposure, which can potentially hinder accurate diagnosis and treatment. In recent years, super-resolution reconstruction technology based on artificial intelligence has been widely studied in image preprocessing and data enhancement, with strong evidence of its effectiveness (14–17). In this study, the generative adversarial network (GAN) model supported by the Onekey platform was applied to learn the mapping from low-resolution to high-resolution, thereby improving the spatial resolution of the MRI slice in detail (18). The model was trained on millions of medical images, which enabled high-quality image preprocessing, including denoising, artifact removal, and intensity values normalization. The resolution of MRI slice was increased by a factor of 4, resulting in a transformation from 1*1 pixels to 0.25*0.25 (Figure 1). Although the image enhancement was reflected in small pixel changes, it provided the deep learning model with accurate feature information and fine tumor boundaries.

2.3 Deep learning model construction

The workflow of this study is illustrated in Figure 2. After image acquisition and preprocessing, the model construction was employed. This model consisted of a deep learning network, a feature adaptation module, and a full-connected classification layer. The deep learning networks were pre-trained using real images, including AlexNet, VGG16, ResNet18, ResNet50, DenseNet121, DenseNet169, GoogleNet, MobileNetV2, and MobileNetV3. The feature adaptation module consisted of two fully-connected layers and was connected behind the deep learning network. Its output was called deep learning (DL) feature with a dimension of 128. The DL features were then input into the classification layer to obtain the final tumor type. 5-fold cross-validation was implemented in this work to avoid overfitting of the deep learning model and ensure the reliability of the results.

Model training is critical to classification results and involves fine-tuning of hyperparameters, especially the learning rate. A fixed learning rate can lead to non-convergence or a local optimal solution. Inspired by recent successful applications of various dynamic learning rate strategies, this study applied them to improve model training efficiency and classification performance (19–22). It is recommended to adopt a larger learning rate in early training and reduce the learning rate with iteration. Therefore, based on the results of several previous tests, the dynamic learning

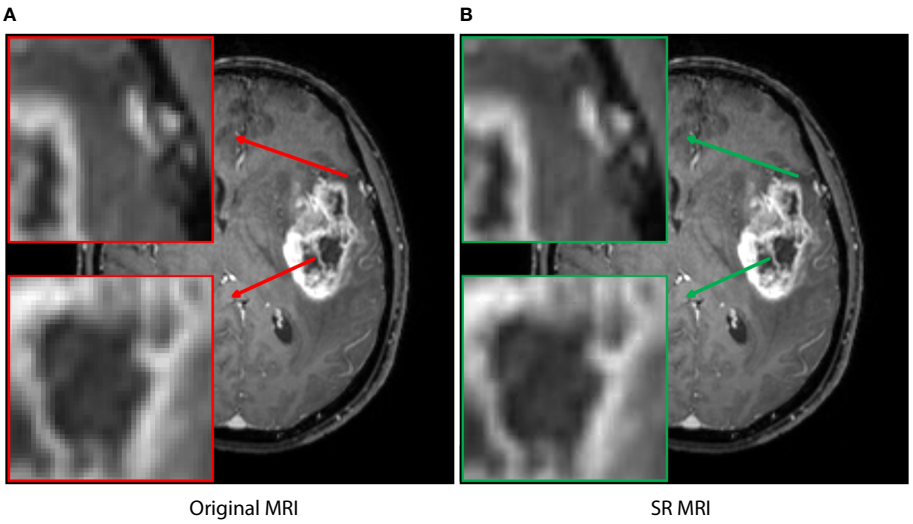


FIGURE 1
The super-resolution reconstruction result based on a generative adversarial network model. The reconstructed image **(B)** is not only very similar to the original transverse image **(A)**, but also has more reasonable edges and finer textures. The boxes represent local enlarged images, with red and green arrows representing the original MRI and SR MRI, respectively.

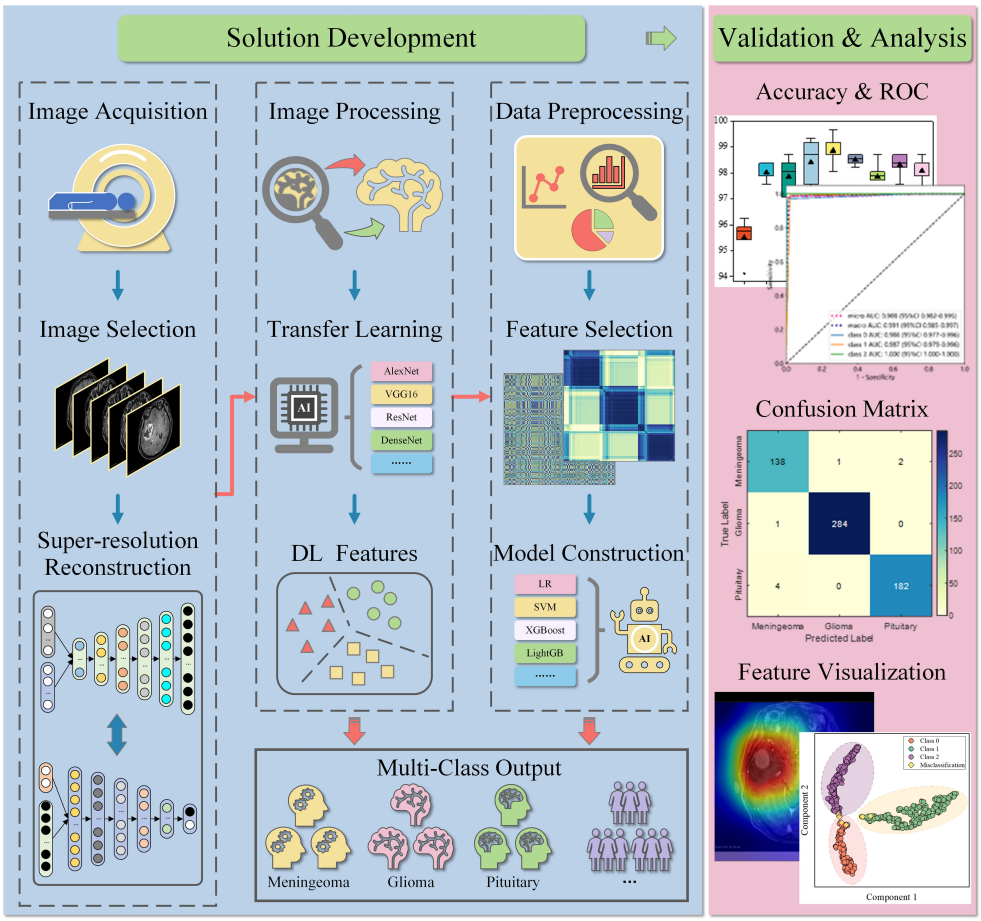


FIGURE 2
Workflow of the study. MRIs were retrospectively collected and selected, then pre-processed and input into the deep learning model. In deep learning diagnosis, the model can directly output common tumor types. In the hybrid classification scheme, the extracted DL features were further used for machine learning construction after preprocessing and selection. The performance of both schemes was verified and analyzed on the test set. LR, logistic regression; SVM, support vector machines; XGBoost, eXtreme gradient boosting; LightGB, light gradient boosting machine.

annealing rate of $lr = 0.01/(1 + 10 * p)^{0.75}$ was applied in model training. The p changed linearly from 0 to 1 with iteration. The batch size and momentum were set to 32 and 0.9, and models were trained for 200 epochs. During model training, cross-entropy loss was calculated, parameters were optimized based on stochastic gradient descent algorithm and backpropagation algorithm.

2.4 Deep learning feature and machine learning construction

As shown in Figure 2, the hybrid scheme integrates deep learning features and machine learning to improve classification accuracy. In this study, 128-dimensional DL features were first applied for data preprocessing, including data format validation, statistical outlier detection, and z-score standardization. Subsequently, Pearson's correlation coefficient was calculated for preliminary feature evaluation and selection, with the threshold set at 0.9. Features were filtered in the training set based on the least absolute shrinkage and selection operator (LASSO), and non-zero items in high-dimensional features were determined as available inputs. Ultimately, non-redundant low-dimensional features can be used to construct machine learning classifier, and perform preliminary diagnose of brain tumors.

2.5 Model performance evaluation

In the deep learning scheme, various key metrics such as accuracy, AUC, sensitivity and specificity were calculated to evaluate the classification performance. The multi-classification performance was demonstrated based on the confusion matrix. More importantly, feature visualization was performed to explore the reliability and interpretability of the model. In the hybrid scheme, accuracy, sensitivity and specificity were used to compare the classification performance of machine learning, including LR, NaiveBayes, SVM, RandomForest, ExtraTrees, XGBoost, LightGBM, Adaptive Boosting (AdaBoost), Multi-layer Perceptron (MLP). The ROC curves were compared using the DeLong test to analyze multi-classification performance.

2.6 Statistical analysis

Statistical analysis was performed using SPSS (version 26.0). Continuous variables were described as mean ± standard deviation, while categorical variables were presented as frequencies and percentages. Continuous variables were analyzed using Student's t test or analysis of variance. Chi-square test or Fisher's exact test was used to compare categorical variables. P value < 0.05 was considered statistical significance. Data preprocessing and feature evaluation, LASSO regression analysis, and DeLong test were performed using Python (version 3.11).

3 Results

3.1 General patient characteristics

Between January 2018 and December 2022, a total of 230 patients with common brain tumors were enrolled in this study. They were initially divided into meningiomas (97 cases), gliomas (66 cases), and pituitary tumors (67 cases) according to the pathological results, which were labeled as 0, 1, 2 in this study. This dataset was labeled as BT-YU in this study. In one data division of 5-fold cross-validation, the baseline clinical characteristics in the training and test cohorts were presented in Table 1. These characteristics included histologic diagnosis and demographic information. No significant differences were observed in any of the detailed characteristics between the two cohorts (all P > 0.05).

3.2 The performance of various deep learning models

In this section, several deep learning models were compared, including AlexNet, VGG16, ResNet18, ResNet50, DenseNet121, DenseNet169, GoogleNet, MobileNetV2, and MobileNetV3. Besides, a commonly used dataset CE-MRI was applied as an auxiliary test to further verify the effectiveness of the models (23). CE-MRI is a T1-weighted contrast-enhanced MRI image set with a total of 3064 images, including meningiomas (708 slices), gliomas (1426 slices) and pituitary tumors (930 slices). The images are a combination of transverse, sagittal and coronal, with a resolution of 512 × 512. The classification of CE-MRI was consistent with the BT-YU dataset. In the 5-fold cross-validation test, the influence of random effects and overfitting were avoided, and the results are shown in Figure 3; Table 2. Further, Supplementary Tables S1, S2 show the detail classification performance. Table 2 also presents the results of state-of-the-art models on the CE-MRI dataset. Among them, references (23) and (27) represent methods combining radiomics and machine learning, references (24) (25), and (26) represent cutting-edge convolutional neural network (CNN) models, and reference (28) represents an improved vision transformer model.

Figure 3 visually shows the results of the five tests. The DenseNet121 model achieved the highest accuracy and relatively low standard deviation in both datasets. In addition, Table 2 quantitatively describes the average statistical indicators of the tests. The classification performance of the DenseNet121 model reached the optimal level, with an accuracy of 0.989 ± 0.006, AUC of

TABLE 1 Patient characteristics.

Characteristics	Entire Cohort (n=230)	Training Cohort (n=185)	Test Cohort (n=45)	*P* value
Age, year	54.63 ± 12.53	54.79 ± 12.61	53.96 ± 12.32	0.820
Tumor size, cm	4.17 ± 1.60	4.13 ± 1.57	4.34 ± 1.78	0.233
Gender, No. (%)				0.309
Male	92	77	15	
Female	138	108	30	
Tumor position (%)				0.271
Frontal Lobe	78	64	14	
Parietial Lobe	20	19	1	
Occipital Lobe	7	4	3	
Temporal & Insular Lobe	40	29	11	
cerebellum	15	12	3	
pituitary	67	54	13	
Others	3	3	0	
MRI Images	2901	2334	567	

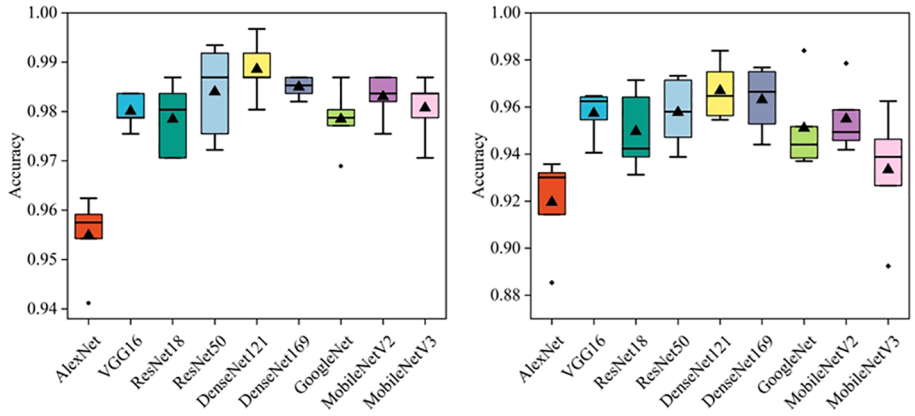


FIGURE 3
The diagnostic results of deep learning models in five tests.

TABLE 2 The diagnostic results of deep learning models.

Datasets	Models	ACC	AUC	SEN	SPE
CE-MRI	AlexNet	0.955 ± 0.008	0.994 ± 0.002	0.949 ± 0.007	0.953 ± 0.008
	VGG16	0.980 ± 0.004	0.997 ± 0.002	0.979 ± 0.005	0.980 ± 0.004
	ResNet18	0.978 ± 0.008	0.999 ± 0.001	0.976 ± 0.008	0.977 ± 0.008
	ResNet50	0.984 ± 0.010	0.998 ± 0.002	0.981 ± 0.011	0.983 ± 0.010
	DenseNet121	0.989 ± 0.006	0.999 ± 0.001	0.987 ± 0.007	0.988 ± 0.006
	DenseNet169	0.985 ± 0.002	0.999 ± 0.001	0.983 ± 0.003	0.984 ± 0.002
	GoogleNet	0.978 ± 0.007	0.998 ± 0.001	0.975 ± 0.008	0.978 ± 0.007
	MobileNetV2	0.983 ± 0.005	0.999 ± 0.000	0.982 ± 0.006	0.983 ± 0.005
	MobileNetV3	0.981 ± 0.006	0.998 ± 0.001	0.978 ± 0.007	0.980 ± 0.007
	Radiomics (23)	0.9128	——	——	——
	CNN (24)	0.9780	0.9890	0.9640	0.9830
	VGG16 (25)	0.9800	0.9900	0.9800	0.9800
	CNN (26)	0.9870	——	0.9860	0.9870
	BoF-SURF + KNN (27)	0.9870	——	0.9840	0.9860
	RanMerFormer (28)	0.9886	——	0.9846	0.9939
BT-YU	AlexNet	0.920 ± 0.021	0.979 ± 0.009	0.914 ± 0.031	0.918 ± 0.022
	VGG16	0.957 ± 0.010	0.994 ± 0.004	0.958 ± 0.012	0.957 ± 0.011
	ResNet18	0.950 ± 0.017	0.990 ± 0.005	0.947 ± 0.020	0.949 ± 0.018
	ResNet50	0.958 ± 0.015	0.993 ± 0.006	0.957 ± 0.020	0.957 ± 0.016
	DenseNet121	0.967 ± 0.013	0.994 ± 0.005	0.966 ± 0.014	0.967 ± 0.013
	DenseNet169	0.963 ± 0.014	0.996 ± 0.003	0.963 ± 0.014	0.963 ± 0.014
	GoogleNet	0.951 ± 0.019	0.992 ± 0.006	0.948 ± 0.023	0.950 ± 0.020
	MobileNetV2	0.955 ± 0.015	0.994 ± 0.003	0.952 ± 0.019	0.954 ± 0.015
	MobileNetV3	0.933 ± 0.026	0.985 ± 0.013	0.930 ± 0.037	0.933 ± 0.028

(Continued)

TABLE 2 Continued

- indicates that the indicator is not included in the listed literature.

0.999 ± 0.001, sensitivity and specificity of 0.987 ± 0.007 and 0.988 ± 0.006 in CE-MRI; the accuracy in BT-YU was 0.967 ± 0.013, AUC was 0.994 ± 0.005, sensitivity and specificity were 0.966 ± 0.014 and 0.967 ± 0.013, respectively.

To examine the multi-classification performance in detail, Figure 4 shows the confusion matrix of the DenseNet121 model

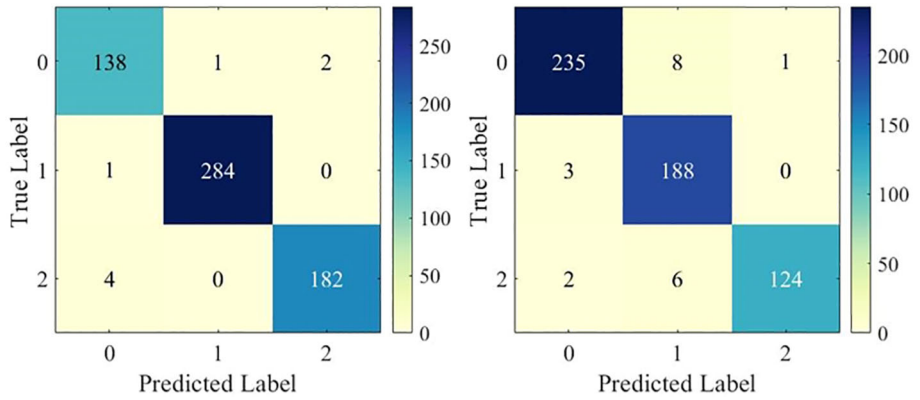


FIGURE 4
Confusion matrix of DenseNet121 model.

in one test. It can be observed that the model showed excellent multi-classification performance for brain tumors. However, as shown in Table 2, the metrics of DenseNet121 model were not all optimal. For example, its AUC on BT-YU was slightly lower than that of the DenseNet169 model. Therefore, actual applications require model selection based on specific conditions.

3.3 The performance of various hybrid models

To further improve accuracy of the Densenet121 model, this section combined deep learning features and machine learning classifiers for diagnostic testing. The models include LR, NaiveBayes, SVM, RandomForest, ExtraTrees, XGBoost, LightGBM, AdaBoost, and MLP. After feature selection and preprocessing, the 128-dimensional deep learning features were applied to construct machine learning classifier. Table 3 shows the learning effect of the model on the deep learning features in one test. In addition, Figure 5 shows the multi-classification ROC curve of each model in the BT-YU dataset. In summary, the LightGBM model maintained the highest accuracy on both datasets, which were 0.989 and 0.984 respectively. Compared to the original Densenet121 model, the accuracy of this hybrid scheme was improved to some extent, especially on the BT-YU dataset. The AUC of the LightGBM model was slightly lower than the original Densenet121 model, but the sensitivity and specificity were optimal (0.985 and 0.988 in CE-MRI dataset; 0.985 and 0.984 in BT-YU dataset). Further, Supplementary Tables S3, S4 show the detail classification performance.

3.4 Feature and model visualization

The classification performance of brain tumors in this study depends on the adaptive extraction of features from MRI images. To improve the interpretability of the classification results, this section first implemented feature visualization based on test sets. T-distributed stochastic neighbor embedding (t-SNE) technology can effectively preserve data similarity and local structure information during dimensionality reduction, while also alleviating the congestion problem (29–31). T-SNE was applied for dimensionality reduction of 128-dimensional features of DenseNet121 model, and cluster analysis was performed in 2-dimensional space for intuitive visualization (Figure 6). There were several misclassified features distributed in concentrated areas of the three clusters. More importantly, the deep learning features in both datasets were intuitively distinguished, providing useful features for multi-classification.

Score-weighted class activation mapping (Score-CAM) is a model visualization technology that obtains weights by the forward pass score of each activation map on the target class. Score-CAM effectively reduces gradient dependence, thus providing excellent visual representation (32–34). Therefore, Score-CAM was introduced to visualize the attention weights of the DenseNet121 model to demonstrate decision support for classification. Figure 7 shows representative sample heatmaps in the two datasets. The red core area indicates a large weight area that contributes significantly to the model classification. Besides, this area matched well with the tumor area, which further confirmed the effectiveness of feature extraction.

4 Discussion

In this study, we proposed a hybrid intelligent scheme that combined deep learning and machine learning for automated diagnosis of brain tumor types based on CE-MRI images. This solution integrated several advanced technologies to improve feature extraction performance, such as super-resolution reconstruction, dynamic learning rate strategy, convolutional neural network and machine learning. Our proposed scheme demonstrated excellent accuracy, AUC, sensitivity, and specificity. Besides, the generalization performance, robustness, and interpretability of the hybrid scheme were further verified on two datasets.

TABLE 3 The diagnostic results of machine learning models.

Datasets	Models	ACC	AUC	SEN	SPE
CE-MRI	—*	0.987	0.997	0.985	0.986
	LR	0.987	0.996	0.983	0.986
	NaiveBayes	0.786	0.794	0.692	0.764
	SVM	0.987	0.996	0.983	0.986
	RandomForest	0.982	0.994	0.978	0.981
	ExtraTrees	0.985	0.994	0.982	0.984
	XGBoost	0.985	0.993	0.981	0.984
	LightGBM	0.989	0.993	0.985	0.988
	AdaBoost	0.987	0.995	0.984	0.986
	MLP	0.984	0.996	0.981	0.983
BT-YU	—*	0.965	0.985	0.962	0.964
	LR	0.980	0.996	0.983	0.981
	NaiveBayes	0.979	0.985	0.982	0.979
	SVM	0.980	0.992	0.983	0.981
	RandomForest	0.980	0.989	0.983	0.981
	ExtraTrees	0.982	0.992	0.985	0.982
	XGBoost	0.982	0.986	0.984	0.982
	LightGBM	0.984	0.990	0.985	0.984
	AdaBoost	0.982	0.991	0.985	0.982
	MLP	0.982	0.996	0.985	0.982

*: The horizontal line indicates the results of the original DenseNet21 model in this test.

Brain tumor diagnosis has always been a meaningful and challenging clinical research hotspot. Previous studies have focused on traditional radiomics solutions for predicting glioblastoma, brain metastasis, and isocitrate dehydrogenase (IDH) mutations (35–37). The process of traditional radiomics includes image acquisition, reconstruction and preprocessing, region of interest delineation, manual feature extraction, feature selection, machine learning construction. Meißner developed a radiomics classifier to predict intracranial BRAF V600E mutation status in patients with melanoma brain metastases, and achieved an AUC value of 0.92 (38). Zhao implemented the World Health Organization (WHO) classification of meningiomas based on radiomics and clinical information, and the AUC value reached 0.860 (95% CI, 0.788–0.923) (39). However, recent research has revealed that traditional radiomics may have inherent limitations that restrict its application in clinical and more complex tasks (40, 41). The level of automation for ROI delineation remains a challenge, and reproducibility may be difficult to ensure in testing and prospective applications, leading to low generalization performance. In addition, manual feature extraction is difficult to comprehensively analyze image features, making the test results potentially accidental. Humphries developed a deep learning algorithm using full-resolution axial images as input to diagnose emphysema patterns, which provided important guidance for other tumor diagnosis (42). Therefore, comprehensive analysis of tumor-related slices, as well as adaptive extraction of tumor region features without pre-definition, may achieve more efficient data processing and more robust information mining.

Deep learning has powerful image processing and feature extraction capabilities, which provides an effective technical support for the above limitations. Convolutional neural networks, with many image processing advantages such as local receptive field, weight sharing and down-sampling, are able to extract local and key features, and are widely used in computer vision and target detection (43–46). Bhattacharjee implemented automatic multi-classification diagnosis of full-slice lung and kidney CT images based on the Xception model, and achieved accuracies of 99.39% and 100% respectively (47). Ziegelmayer verified the excellent robustness of deep learning features relative to radiomics features based on CT scans of 60 patients with hepatocellular carcinoma and hepatocolon cancer metastasis (48). Similar to these studies, our deep learning-based model showed good performance in tumor region detection and feature extraction. In addition, the combination of deep learning features and machine learning is considered an effective strategy to further improve the accuracy.

Therefore, to improve the limitations of traditional radiomics, this study first introduced several deep learning models to directly analyze MRI images. These models include AlexNet, VGG16, ResNet18, ResNet50, DenseNet121, DenseNet169, GoogleNet, MobileNetV2, and MobileNetV3. In addition, the proposed solution integrated super-resolution reconstruction technology and dynamic learning rate annealing technology to ensure the quality of image preprocessing and model training.

In the deep learning model test, the DenseNet121 model achieved the best comprehensive classification performance, and its accuracy reached 0.989 ± 0.006 and 0.967 ± 0.013 in the two datasets. It can be seen from Table 2 that in terms of accuracy, this work was slightly better than the comparison models, which may be caused by the comprehensive factors of image preprocessing, training strategy and model construction. Of course, the accuracy of most models was also close to 0.99, and the results of VGG16, ResNet, GoogleNet and MobileNet models in this work were not significantly lower, which indicated that accuracy, model complexity and training time should be comprehensively considered in model selection. This also inspired us to look for another effective solution to improve accuracy instead of changing deep learning model.

To further improve the accuracy of diagnosis, this study developed a hybrid deep learning scheme for multi-classification of brain tumors. This hybrid scheme essentially applied machine learning to replace the final classification layer of the deep learning model, thus combining the advantages of deep feature extraction and machine learning classification. The machine learning models include LR, NaiveBayes, SVM, RandomForest, ExtraTrees, XGBoost, LightGBM, AdaBoost, and MLP. In the test based on the features extracted by DenseNet121 model, the LightGBM model, as a promising machine learning model, achieved an

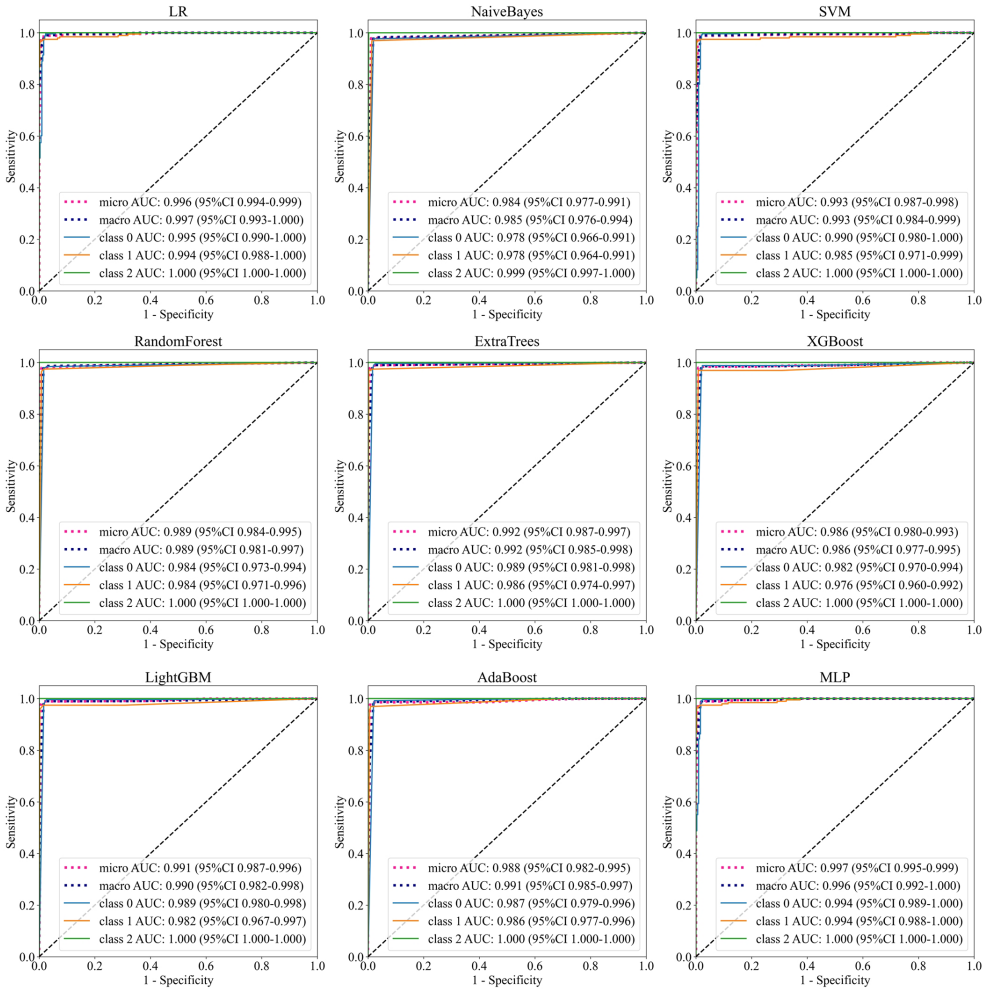


FIGURE 5
ROC curves of machine learning models on BT-YU dataset.

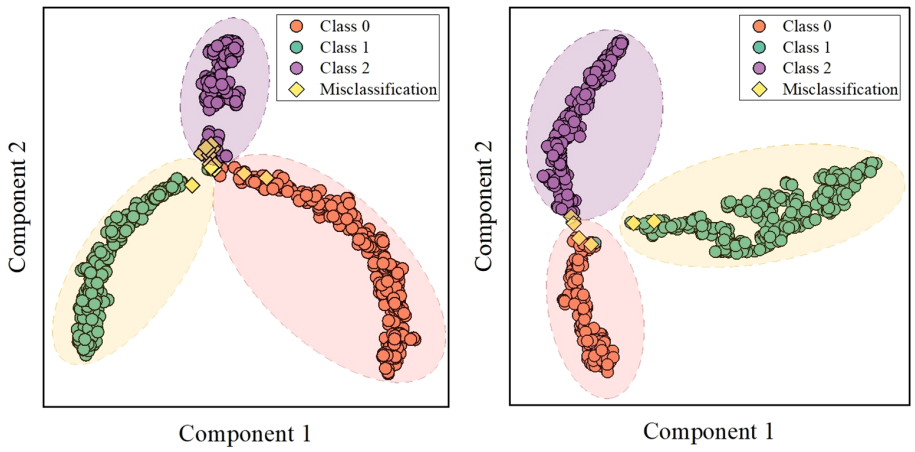


FIGURE 6
Feature visualization of DenseNet121 model based on t-SNE.

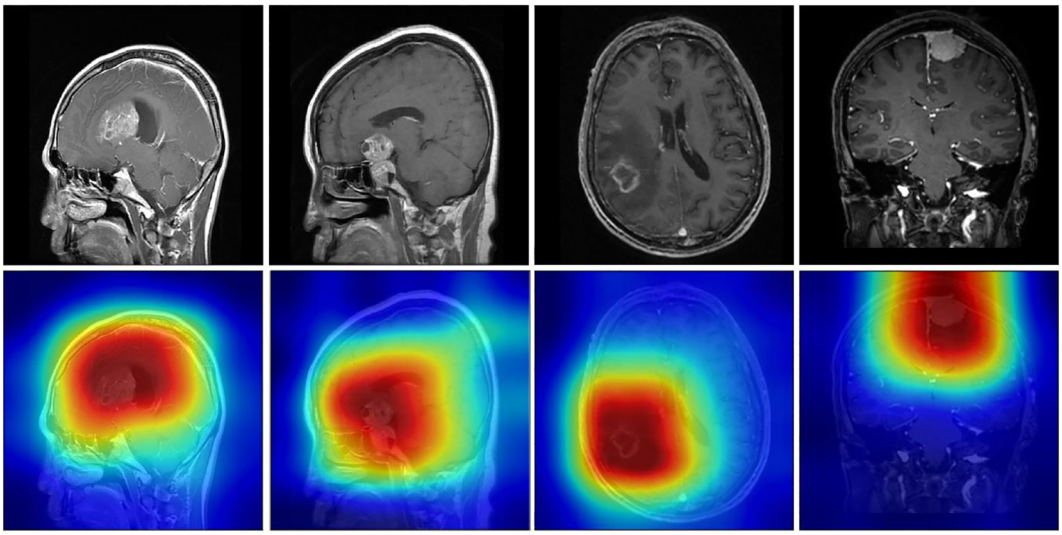


FIGURE 7
Model attention heatmaps on representative samples.

accuracy of 0.989 and 0.984, a sensitivity of 0.985, and a specificity of 0.988 and 0.984 in the two datasets. Although the AUC of the LightGBM model was not optimal, its ROC curves shown in Figure 5 were relatively ideal. Therefore, this model was considered to have the best comprehensive multi-class performance. Since the original accuracy of the CE-MRI dataset was already high, the enhancement effect of the hybrid scheme was more obvious in the BT-YU dataset.

In addition, this study utilized the entire MRI as input to improve automation, and it was crucial to explore feature visualization and model focus areas. Therefore, T-SNE and Score-Grad technologies were employed in the visualization research. Figure 6 shows the 2-dimensional clustering of test sample features, which well illustrates the difference in extracted features between different tumors. Figure 7 displays the Score-Grad heat map of representative samples, indicating that the focus areas of the model are well consistent with the tumor and peritumoral areas, thus reliably explaining the classification results. In summary, this hybrid scheme can focus on key areas of brain tumors and extract pivotal features, thereby providing decision support for targeted diagnosis and treatment plans.

Of course, we are also aware that various current imaging approaches can be used to non-invasively identify brain tumors (35, 36, 39). In addition, there are many classifications and subtypes of brain tumors, such as brain metastases, gliomas, and meningiomas. It is extremely challenging to develop a single validated solution to diagnose all classifications. Instead, the purpose of this study is not to provide the only effective method, but to demonstrate the potential application value of deep learning in automatic processing and interpretability, and to verify the feasibility of improving accuracy through deep learning and machine learning. Based on our results, the next step is to solve the problem of multi-center validation and detailed classification of brain tumors, such as single brain metastases, glioma subtypes, meningioma grading.

5 Conclusions

This study investigated a hybrid deep learning scheme for automated primary brain tumors diagnosis. The scheme integrated various advanced technologies and used two datasets to verify the classification performance and interpretability. By combining super-resolution reconstruction and dynamic learning rate annealing technologies, the deep learning model achieved high classification accuracy. Furthermore, based on deep feature transfer and machine learning models, the performance of brain tumor diagnosis can be significantly improved. In addition, through t-SNE cluster analysis and Score-Grad attention visualization, the efficient classification results of the model can be intuitively verified and explained. In conclusion, this study highlighted the importance of integrating multiple advanced technologies to extract robust deep learning features, which had important reference significance for the development of automated radiomics for brain tumors.

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 authors.

Ethics statement

The studies involving humans were approved by the ethics committees of the First Affiliated Hospital of Yangtze University, and informed consent was waived. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in

accordance with the national legislation and institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because This retrospective study was approved by the ethics committees of the First Affiliated Hospital of Yangtze University, and informed consent was waived.

Author contributions

ZW: Conceptualization, Funding acquisition, Methodology, Writing – original draft. CH: Conceptualization, Data curation, Project administration, Writing – original draft. YH: Resources, Supervision, Validation, Writing – review & editing. HL: Resources, Supervision, Validation, Writing – review & editing. CL: Formal analysis, Writing – review & editing. XW: Visualization, Writing – review & editing. YZ: Data curation, Writing – review & editing. JL: Investigation, Software, Writing – review & editing. JC: Formal analysis, Investigation, Resources, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Open Research Fund of Hubei Key Laboratory of Precision Radiation Oncology (jzfs016), and the Natural Science Foundation of Hubei Province (2023AFC037).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1363756/full#supplementary-material

References

1. Miller KD, Ostrom QT, Kruchko C, Patil N, Tihan T, Cioffi G, et al. Brain and other central nervous system tumor statistics, 2021. *CA Cancer J Clin*. (2021) 71:381–406. doi: 10.3322/caac.21693

2. Ostrom QT, Price M, Neff C, Cioffi G, Waite KA, Kruchko C, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2015-2019. *Neuro-oncology*. (2022) 24:v1–v95. doi: 10.1093/neuonc/noac202

3. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015. *Neuro-oncology*. (2018) 20:1–86. doi: 10.1093/neuonc/noy131

4. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. *New Engl J Med*. (2023) 388:1981–90. doi: 10.1056/NEJMra2301725

5. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol*. (2022) 19:132–46. doi: 10.1038/s41571-021-00560-7

6. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. *Am J Neuroradiol*. (2018) 39:208–16. doi: 10.3174/ajnr.A5391

7. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur Radiol*. (2017) 27:4188–97. doi: 10.1007/s00330-016-4637-3

8. Xu J, Meng Y, Qiu K, Topatana W, Li S, Wei C, et al. Applications of artificial intelligence based on medical imaging in glioma: Current state and future challenges. *Front Oncol*. (2022) 12:892056. doi: 10.3389/fonc.2022.892056

9. Zhu Y, Man C, Gong L, Dong D, Yu X, Wang S, et al. A deep learning radiomics model for preoperative grading in meningioma. *Eur J Radiol*. (2019) 116:128–34. doi: 10.1016/j.ejrad.2019.04.022

10. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and deep learning methods for radiomics. *Med Phys*. (2020) 47:e185–202. doi: 10.1002/mp.13678

11. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. *IEEE Signal Proc Mag*. (2019) 36:132–60. doi: 10.1109/MSP.2019.2900993

12. Lao J, Chen Y, Li Z, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep-UK*. (2017) 7:10353. doi: 10.1038/s41598-017-10649-8

13. Wallis D, Buvat I. Clever Hans effect found in a widely used brain tumour MRI dataset. *Med Image Anal*. (2022) 77:102368. doi: 10.1016/j.media.2022.102368

14. Hou M, Zhou L, Sun J. Deep-learning-based 3D super-resolution MRI radiomics model: superior predictive performance in preoperative T-staging of rectal cancer. *Eur Radiol*. (2023) 33:1–10. doi: 10.1007/s00330-022-08952-8

15. Tanno R, Worrall DE, Kaden E, Ghosh A, Grussu F, Bizzi A, et al. Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. *NeuroImage*. (2021) 225:117366. doi: 10.1016/j.neuroimage.2020.117366

16. Zhang S, Yuan Z, Wang Y, Bai Y, Chen B, Wang H. REUR: a unified deep framework for signet ring cell detection in low-resolution pathological images. *Comput Biol Med*. (2021) 136:104711. doi: 10.1016/j.compbiomed.2021.104711

17. Ahmad W, Ali H, Shah Z, Azmat S. A new generative adversarial network for medical images super resolution. *Sci Rep-UK*. (2022) 12:9533. doi: 10.1038/s41598-022-13658-4

18. Huang Y, Zhu T, Zhang X, Li W, Zheng X, Cheng M, et al. Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: a multicenter, retrospective study. *EClinicalMedicine*. (2023) 58:101899. doi: 10.1016/j.eclinm.2023.101899

19. Jiao J, Zhao M, Lin J, Liang K. Residual joint adaptation adversarial network for intelligent transfer fault diagnosis. *Mech Syst Signal Pr*. (2020) 145:106962. doi: 10.1016/j.ymssp.2020.106962

20. Guan H, Fu C, Zhang G, Li K, Wang P, Zhu Z. A lightweight model for efficient identification of plant diseases and pests based on deep learning. *Front Plant Sci*. (2023) 14:1227011. doi: 10.3389/fpls.2023.1227011

21. Houssein EH, Mohamed O, Samee NA, Mahmoud NF, Talaat R, Al-Hejri AM, et al. Using deep DenseNet with cyclical learning rate to classify leukocytes for leukemia identification. *Front Oncol*. (2023) 13:1230434. doi: 10.3389/fonc.2023.1230434

22. Hu Q, Wei X, Liang X, Zhou L, He W, Chang Y, et al. In-process vision monitoring methods for aircraft coating laser cleaning based on deep learning. *Opt Laser Eng*. (2023) 160:107291. doi: 10.1016/j.optlaseng.2022.107291

23. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, et al. Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS One*. (2015) 10:e0140381. doi: 10.1371/journal.pone.0140381

24. Khan MSI, Rahman A, Debnath T, Karim MR, Nasir MK, Band SS, et al. Accurate brain tumor detection using deep convolutional neural network. *Comput Struct Biotec*. (2022) 20:4733–45. doi: 10.1016/j.csbj.2022.08.039

25. Khaliki MZ, Başarslan MS. Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. *Sci Rep-UK*. (2024) 14:2664. doi: 10.1038/s41598-024-52823-9

26. Amou MA, Xia KW, Kamhi S, Mouhafid M. A novel MRI diagnosis method for brain tumor classification based on CNN and bayesian optimization. *Healthcare-Basel*. (2022) 10:494. doi: 10.3390/healthcare10030494

27. Mohammed ZF, Mussa DJ. Brain tumour classification using BoF-SURF with filter-based feature selection methods. *Multimed Tools Appl*. (2024). doi: 10.1007/s11042-024-18171-6

28. Wang J, Lu SY, Wang SH, Zhang YD. RanMerFormer: Randomized vision transformer with token merging for brain tumor classification. *Neurocomputing*. (2024) 573:127216. doi: 10.1016/j.neucom.2023.12721

29. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. (2008) 9:2579–605.

30. Yang R, Wu J, Sun L, Lai S, Xu Y, Liu X, et al. Radiomics of small renal masses on multiphasic CT: accuracy of machine learning–based classification models for the differentiation of renal cell carcinoma and angiomyolipoma without visible fat. *Eur Radiol*. (2020) 30:1254–63. doi: 10.1007/s00330-019-06384-5

31. Zhou L, Rueda M, Alkhateeb A. Classification of breast cancer nottingham prognostic index using high-dimensional embedding and residual neural network. *Cancers*. (2022) 14:934. doi: 10.3390/cancers14040934

32. Wang H, Wang Z, Du M, Yang F, Zhang Z, Ding S, et al. (2020). "Score-CAM: Score-weighted visual explanations for convolutional neural networks", in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 10662 Los Vaqueros Circle Los Alamitos, CA USA: IEEE Computer Society. doi: 10.1109/CVPRW50498.2020

33. Ibrahim R, Shafiq MO. Augmented Score-CAM: High resolution visual interpretations for deep neural networks. *Knowl-Based Syst*. (2022) 252:109287. doi: 10.1016/j.knosys.2022.109287

34. Hu H, Xu W, Jiang T, Cheng Y, Tao X, Liu W, et al. Expert-level immunofixation electrophoresis image recognition based on explainable and generalizable deep learning. *Clin Chem*. (2023) 69:130–9. doi: 10.1093/clinchem/hvac190

35. Cheong EN, Park JE, Park SY, Jung SC, Kim HS. Achieving imaging and computational reproducibility on multiparametric MRI radiomics features in brain tumor diagnosis: phantom and clinical validation. *Eur Radiol*. (2023) 69:2008–23. doi: 10.1007/s00330-023-10164-7

36. Kniep HC, Madesta F, Schneider T, Hanning U, Schönfeld MH, Schön G, et al. Radiomics of brain MRI: utility in prediction of metastatic tumor type. *Radiology*. (2019) 290:479–87. doi: 10.1148/radiol.2018180946

37. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. (2019) 69:127–57. doi: 10.3322/caac.21552

38. Meißner AK, Gutsche R, Galldiks N, Kocher M, Jünger ST, Eich ML, et al. Radiomics for the noninvasive prediction of the BRAF mutation status in patients with melanoma brain metastases. *Neuro-oncology*. (2022) 24:1331–40. doi: 10.1093/neuonc/noab294

39. Zhao Z, Nie C, Zhao L, Xiao D, Zheng J, Zhang H, et al. Multi-parametric MRI-based machine learning model for prediction of WHO grading in patients with meningiomas. *Eur Radiol*. (2023) 34:2468–79. doi: 10.1007/s00330-023-10252-8

40. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *Radiol Med*. (2021) 126:1296–311. doi: 10.1007/s11547-021-01389-x

41. Zhou H, Bai H, Jiao Z, Cui B, Wu J, Zheng H, et al. Deep learning-based radiomic nomogram to predict risk categorization of thymic epithelial tumors: A multicenter study. *Eur J Radiol*. (2023) 168:111136. doi: 10.1016/j.ejrad.2023.111136

42. Humphries SM, Notary AM, Centeno JP, Strand MJ, Crapo JD, Silverman EK, et al. Deep learning enables automatic classification of emphysema pattern at CT. *Radiology*. (2020) 294:434–44. doi: 10.1148/radiol.2019191022

43. Bianconi A, Rossi LF, Bonada M, Zeppa P, Nico E, De Marco R, et al. Deep learning-based algorithm for postoperative glioblastoma MRI segmentation: a promising new tool for tumor burden assessment. *Brain inf*. (2023) 10:26. doi: 10.1186/s40708-023-00207-6

44. Martini ML, Oermann EK. Intraoperative brain tumour identification with deep learning. *Nat Rev Clin Oncol*. (2020) 17:200–1. doi: 10.1038/s41571-020-0343-9

45. Minaee S, Boykov YY, Porikli F, Plaza AJ, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: A survey. *IEEE T Pattern Anal*. (2021) 44:3523–42. doi: 10.1109/TPAMI.2021.3059968

46. Chen Z, Lin L, Wu C, Li C, Xu R, Sun Y. Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. *Cancer Commun*. (2021) 41:1100–15. doi: 10.1002/cac2.12215

47. Bhattacharjee A, Rabea S, Bhattacharjee A, Elkaeed E, Murugan R, Selim HMRM, et al. A multi-class deep learning model for early lung cancer and chronic kidney disease detection using computed tomography images. *Front Oncol*. (2023) 13:1193746. doi: 10.3389/fonc.2023.1193746

48. Ziegelmayer S, Reischl S, Harder F, Makowski M, Braren R, Gawlitza J. Feature robustness and diagnostic capabilities of convolutional neural networks against radiomics features in computed tomography imaging. *Invest Radiol*. (2022) 57:171–7. doi: 10.1097/RLI.0000000000000827

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 10 May 2024
DOI 10.3389/fonc.2024.1400872

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Xin-Wu Cui,
Huazhong University of Science and Technology, China
Anum Kazerouni,
University of Washington, United States

*CORRESPONDENCE
Yong Wang
drwangyong77@163.com
Lin Li
linlin77216@sina.com
Qian Li
754427296@qq.com

†These authors have contributed equally to this work

RECEIVED 14 March 2024
ACCEPTED 25 April 2024
PUBLISHED 10 May 2024

CITATION
Sun C, Gong X, Hou L, Yang D, Li Q, Li L and Wang Y (2024) A nomogram based on conventional and contrast-enhanced ultrasound radiomics for the noninvasively prediction of axillary lymph node metastasis in breast cancer patients.
Front. Oncol. 14:1400872.
doi: 10.3389/fonc.2024.1400872

COPYRIGHT
© 2024 Sun, Gong, Hou, Yang, Li, Li 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.

A nomogram based on conventional and contrast-enhanced ultrasound radiomics for the noninvasively prediction of axillary lymph node metastasis in breast cancer patients

Chao Sun[1†], Xuantong Gong[1†], Lu Hou[2], Di Yang[1], Qian Li[3*], Lin Li[4*] and Yong Wang[1*]

[1]Department of Ultrasound, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [3]Department of Ultrasound, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China, [4]Department of Diagnostic Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Background: This study aimed to investigate whether quantitative radiomics features extracted from conventional ultrasound (CUS) and contrast-enhanced ultrasound (CEUS) of primary breast lesions can help noninvasively predict axillary lymph nodes metastasis (ALNM) in breast cancer patients.

Method: A total of 111 breast cancer patients with 111 breast lesions were prospectively enrolled. All the included patients received presurgical CUS screening and CEUS examination and were randomly assigned to the training and validation sets at a ratio of 7:3 (n = 78 versus 33). Radiomics features were respectively extracted based on CUS and CEUS using the *PyRadiomics* package. The max-relevance and min-redundancy (MRMR) and least absolute shrinkage and selection operator (LASSO) analyses were used for feature selection and radiomics score calculation in the training set. The variance inflation factor (VIF) was performed to check the multicollinearity among selected predictors. The best performing model was selected to develop a nomogram using binary logistic regression analysis. The calibration and clinical utility of the nomogram were assessed.

Results: The model combining CUS reported ALN status, CUS radiomics score (CUS-radscore) and CEUS radiomics score (CEUS-radscore) exhibited the best performance. The areas under the curves (AUC) of our proposed nomogram in the training and external validation sets were 0.845 [95% confidence interval (CI), 0.739-0.950] and 0.901 (95% CI, 0.758-1). The calibration curves and decision curve analysis (DCA) demonstrated the nomogram's robust consistency and clinical utility.

Conclusions: The established nomogram is a promising prediction tool for noninvasive prediction of ALN status. The radiomics features based on CUS and CEUS can help improve the predictive performance.

KEYWORDS

axillary lymph node, breast cancer, radiomics, conventional ultrasound, contrast-enhanced ultrasound

1 Introduction

Breast cancer is the most common cancer and the leading cause of tumor-related mortality in female patients worldwide (1). Though about 98.6% of breast cancer patients could survive for 5 years after the diagnosis, this rate would decrease to 84.4% in the presence of axillary lymph node metastasis (ALNM) (2, 3). Axillary lymph node (ALN) status is an independent prognostic indicator for disease-free survival and overall survival in early-stage patients with breast cancer (4). The correct preoperative staging of ALN status is of important clinical significance for the optimization of clinical decision.

Currently, axillary lymph node dissection (ALND) is the widely recognized method for identifying metastatic ALNs, which is invasive and associated with a series of complications including lymphedema, nerve injury, abnormal function (5, 6). To minimize unnecessary body damage, the sentinel lymph node biopsy (SLNB) has become the preferred approach for evaluating ALN status. However, SLNB is also accompanied with a series of side effects such as infection, allergies to tracer agents, skin staining, longer surgical times and higher surgical trauma (7–9).

Conventional ultrasound imaging (CUS) is a commonly recommended method for preoperative assessment of the ALN status, owing to its convenience, radiation-free and non-invasive advantages. However, the value of CUS in identifying ALNM is limited, with a sensitivity ranging from 48% to 87% and a specificity ranging from 55% to 97% (10). For the ALN involved with micro-metastasis, it will become imperceptible on CUS images. To tackle this issue, recent studies have attempted to exploit the ultrasonic features of primary breast lesions to predict ALNM (11–13). These ultrasonic features encompassed both the morphological information derived from CUS and the functional information derived from CEUS (11–13). However, the predictive performances based on these features are not excellent. The development of more effective assessment methods for noninvasive prediction of ALNM is imperative.

Radiomics is a widely applied technique for extracting high-throughput features from medical images. A great amount of high-dimensional features including shape, intensity and texture information that were unable to observe with the naked eye can be obtained and objectively analyzed using radiomics method (14, 15). Based on CUS images of primary breast tumor, ultrasound radiomics has been applied to establish prediction model for ALN status in several studies (16–18).

However, there is currently no study reporting the utility of radiomics features derived from CEUS in predicting ALN status in patients diagnosed with breast cancer. On the basis of routine CUS screening among breast cancer patients, it is convenient to perform CEUS and obtain the perfusion information regarding the primary breast lesions. In this study, we aimed to investigate whether quantitative radiomics features derived from both CUS and CEUS could help improve the predictive performance of ALNM.

2 Materials and methods

This was a single-center prospective study approved by the ethics committee of the cancer hospital of the Chinese Academy of Medical Sciences. The enrolled patients had given their informed consent to participate this study.

2.1 Patient selection and sample size estimation

From March 2019 to January 2022, a total of 111 breast lesions from 111 female patients were enrolled in this study based on the following inclusion criteria: (1) patients who were aged ≥ 18 and diagnosed with breast malignant cancers by postoperative pathology results. (2) with the presence of measurable lesions (≥1) proven by conventional ultrasound (CUS) and contrast enhanced ultrasound (CEUS) performed before any interventional treatment, including core needle biopsy, neoadjuvant chemotherapy, or surgery. (3) with complete baseline CUS images of breast lesions and ipsilateral axillary lymph node (ALN) assessment based on CUS features. (4) with complete baseline CEUS videos of breast lesions. (5) with clearly verified ALN status by pathology after sentinel lymph node biopsy (SLNB) or axillary lymph node dissection (ALND). The exclusion criteria were as follows: (1) currently has or had a history of malignant tumors besides breast cancer. (2) allergic to ultrasound contrast agents or other contraindication for ultrasound contrast agent application. The enrolled patients were assigned numbers and subjected to stratified random sampling based on the pathological status of ALNs in order to create a training set and an external validation set at a ratio of 7:3 (Figure 1).

The sample size estimation was based on the reported axillary lymph node metastasis (ALNM) incidence in breast cancer patients

and the principle of 10 outcome events per variable (19–21). Using an estimated ALNM incidence of 0.45 in the study population and for three predictors, we aimed to enroll 67 breast cancer patients but actually enrolled 70 in the training set.

2.2 Clinical and pathological information collection

The clinical and pathological characteristics of the enrolled patients, including age, breast tumor size (measured by CUS), breast tumor histological type, pathological ALN status, breast tumor receptor (estrogen receptor, ER; progesterone receptor, PR; human epidermal growth factor receptor, Her-2) status, were collected from the medical records system. All the breast lesions were surgically removed, following by ALND. The breast tumor histology and pathological ALN status were documented according to the postoperative pathological results.

2.3 CUS examination procedure and CUS-reported ALN status

The CUS was performed by two senior sonographers (5 years' experience in CUS diagnosis of breast tumor) using Philips EPIQ5 ultrasonic diagnostic equipment (Philips, Bothell, WA) with high frequency linear array probes to choose the target breast lesion and best sonographic sections for the observation. The whole breast underwent CUS screening and the identified breast lesions were graded based on the second edition of the American College of Radiology Breast Imaging Reporting and Data System for US (22). The breast lesion with BI-RADS grade 4C or BI-RADS grade 5 was considered as suspiciously malignant. The lesion size was measured on CUS images. If multiple breast lesions were suspiciously malignant in a patient, the biggest one was selected as the target lesion. The maximum transverse and longitudinal sections of each target lesion were respectively captured and stored, eventually obtaining two CUS images per target lesion. The ALN status was also evaluated by CUS. An ALN was defined as CUS-reported positive ALN when it presented one of the following features: irregular cortical thickness of greater than 3 mm, longest-to-shortest axis ratio less than 2, or absence of a fatty hilum (23). CUS images of target lesions and ALNs were initially stored in a Digital Imaging and Communications in Medicine (DICOM) format and subsequently converted to a Joint Photographic Experts Group (JPEG) format for further analysis.

2.4 CEUS examination procedure

The CEUS was also performed by two senior sonographers (5 years' experience in CEUS diagnosis) using the same ultrasonic

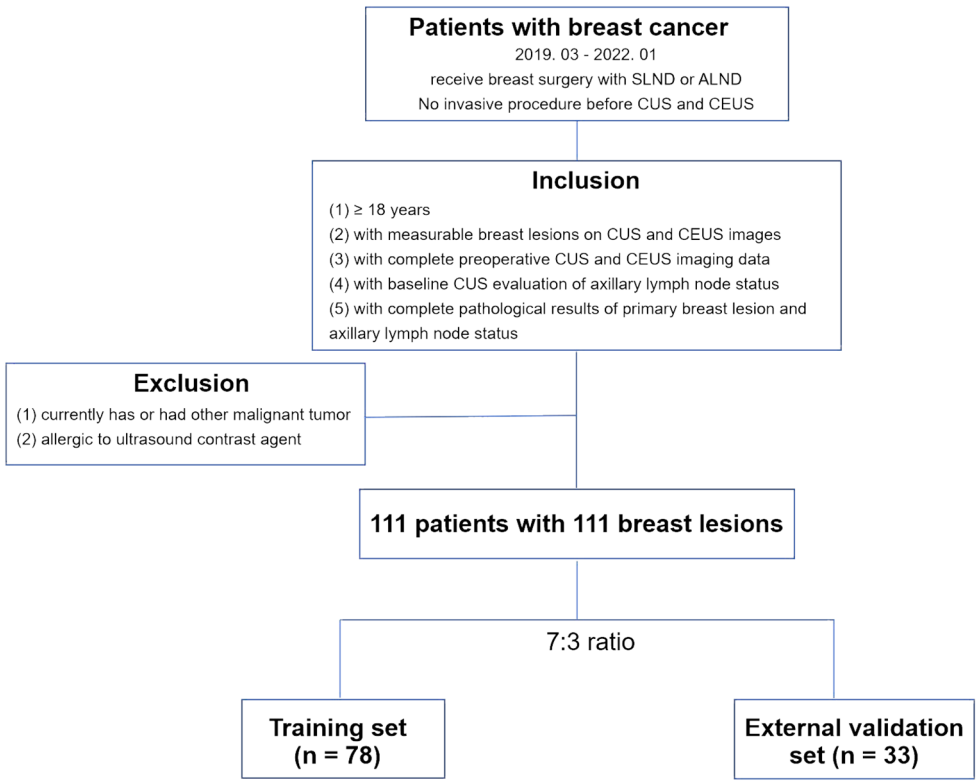


FIGURE 1
Flowchart for the patient selection criteria. CUS, conventional ultrasound; CEUS, contrast-enhanced ultrasound. SLNB, sentinel lymph node biopsy; ALND, axillary lymph node dissection.

diagnostic equipment mentioned in CUS examination. First, the lyophilized powder of contrast agent (Sono Vue, Bracco SpA, Milan, Italy) was reconstituted by adding 5 mL of 0.9% saline and shaking to form a homogeneous microbubble suspension. Second, the real-time imaging of double-frame CEUS mode was activated after the proper ultrasonic section of the target lesion was displayed by CUS. Then, a bolus of 4.8-mL suspension of the contrast agent was administered via antecubital vein. The continuous storage of CEUS imaging and chronograph were initiated immediately following the injection of the contrast agent, lasting for a duration of 3 minutes. The CEUS imaging data was initially stored as a dynamic video with a DICOM format and subsequently converted to an Audio Video Interleaved (AVI) format for further analysis.

2.5 Region of interest segmentation

Next, two CUS images (JPEG format) and one CEUS video (AVI format) of each target lesion was used for region of interest (ROI) segmentation. For CUS images, the ROI was delineated around the boundary of the target lesion. If the hyperechoic halo was present in CUS images, the boundary was positioned outside hyperechoic halo. For CEUS videos, a series of continuous frames of the CEUS video formed the CEUS images for target lesion. A rectangular box was outlined in the frame that can clearly displayed the tumor area and a computer vision algorithm was subsequently used to track and draw the ROI sketches ac nross the other frames in the video (24). The example of target lesion delineation based on CUS images and CEUS video frames were listed in Figure 2.

2.6 Feature extraction, selection and radiomics score calculation

The radiomics features were extracted from the ROI delineated based on CUS and CEUS images, respectively. An open-source *pyradiomics* package (http://github.com/Radiomics/pyradiomics) was used to extract shape features, first-order statistical features, and texture features from both the original images and transformed images, respectively. The transformed images were generated by performing 2D discrete wavelet decomposition and reconstruction, or filtering by the Laplacian of Gaussian method with different sigma parameters. Specifically, the shape features extracted from CEUS images were excluded due to the limitations of using a rectangular box delineation, which could not accurately capture the shape information of target lesion.

Next, the max-relevance and min-redundancy (MRMR) and least absolute shrinkage and selection operator (LASSO) analyses were used to respectively select the most effective CUS and CEUS feature subsets for the prediction of ALNM in the training set. Finally, two sets of radiomics scores based on CUS or CEUS images were respectively constructed with corresponding selected features.

2.7 Development and validation of nomogram

Before prediction model construction, the multicollinearity was analyzed by assessing the variance inflation factor (VIF) among involved variables. Multicollinearity was considered to exist when a VIF value was above 3. The clinical characteristic with $P < 0.05$ in the univariate analysis was incorporating with radiomics scores based on CUS and CEUS images to construct a nomogram as a quantitative tool to predict ALNM using multivariable logistic regression analysis. The discriminatory ability of the model was evaluated using receiver operating characteristic (ROC) curve analysis and the area under curve (AUC) in the training data and validation date. The Delong algorithm was used to compare AUC of different models ($P < 0.05$). The predictive accuracy of the model was evaluated by calibration curve. A decision curve analysis was performed to determine the clinical usefulness.

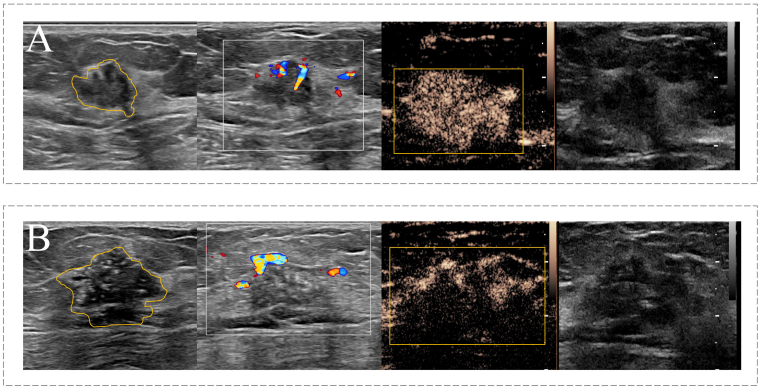


FIGURE 2
Example of the ROI segmentation in CUS image and CEUS videos. **(A)** The ROI segmentation on the CUS image (Left) and CEUS frame (Right) of malignant breast lesion with ALNM. **(B)** The ROI segmentation on the CUS image (Left) and CEUS frame (Right) of malignant breast lesion without ALNM.

2.8 Statistical analysis

R software (ver.1.4.1717, R Development Core Team) and SPSS 22.0 software (IBM Corporation, NY, USA) were used for statistical analysis. The χ^2 test or Fisher's exact test were used for the comparison of classification variables, whereas the independent-sample *t* test was used for the comparison of continuous variables. A *P* value<0.05 was considered statistically significant. SPSS was used for binary logistic regression analysis and a series of packages in R software were used to develop the predictive model and test the diagnostic performance of the model. The corresponding packages included the rsample, mRMRe, glmnet, caret, corrplot, survival, ggplot2, rms, pROC, tidyverse, rmda and ggDCA packages.

3 Results

3.1 Baseline characteristics of training set and external validation set

A total of 111 patients with 111 breast malignant lesions were enrolled in this study from March 2019 to January 2022. Table 1 summarizes the baseline clinical characteristics of 78 patients in the training set and 33 patients in the validation set. These baseline characteristics included the age of each enrolled patient, size, histology type and receptor status of each breast tumor, as well as the pathological status of ALNM. A total of 27 (34.6%) patients with ALNM were included in the training set and 12 (36.4%) patients with ALNM were included in the external validation set. Both sets had comparable ALN prevalence rate (p = 0.86). Additionally, there are no significant differences in the other baseline characteristics between two sets.

TABLE 1 Baseline characteristics of patients in training set and external validation set.

Characteristic	Training set (*n* = 78)	External validation set (*n* = 33)	P value
Age (year)	50.3 ± 10.76	53 ± 7.91	0.068
Pathological ALN status			0.86
Positive	27 (34.6%)	12 (36.4%)	
Negative	51 (65.4%)	21 (63.6%)	
Histological type			0.349
IDC/ILC/IDLC	69 (88.5%)	27 (81.8%)	
DCIS	9 (11.5%)	6 (18.2%)	
Receptor status			
ER			0.143
+	56 (71.8%)	28 (84.8%)	
-	22 (28.2%)	5 (15.2%)	
PR			0.112
+	55 (70.5%)	28 (84.8%)	
-	23 (29.5%)	5 (15.2%)	
Her-2			0.833
-/+	47 (60.3%)	20 (60.6%)	
++	17 (21.8%)	5 (15.2%)	
+++	14 (17.9%)	8 (24.2%)	
Primary tumor size (cm)	2.3 ± 1.41	2.3 ± 1.20	0.689
CUS-reported ALN status			
Suspicious	27 (34.6%)	14 (42.4%)	0.436
Unsuspicious	51 (65.4%)	19 (57.6%)	

ALN, axillary lymph node; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; IDLC, mixed invasive ductal and lobular carcinoma. DCIS, ductal carcinoma in situ; ER, estrogen receptor; PR, progesterone receptor. Her-2, human epidermal growth factor receptor 2; CUS, conventional ultrasound.

3.2 Radiomics features selection and radiomics score calculation

A total of 1530 radiomics features (CUS-rad-features) were extracted from CUS images, while 1395 radiomics features (CEUS-rad-features) were extracted from CEUS images. Next, top 30 features were respectively selected from these CUS-rad-features and CEUS-rad-features using MRMR algorithm. Finally, two CUS-rad-features and two CEUS-rad-features with non-zero coefficients were respectively identified by LASSO regression model (Table 2, Figure 3). The radiomics scores based on CUS images and CEUS images were respectively calculated using the final selected features to generate CUS-radscore and CEUS-radscore for model construction.

TABLE 2 List of the selected radiomics features extracted from CUS and CEUS images via LASSO analyses.

Image source	Image Type	Feature Class	Feature Name	Coefficient
CUS	Log_sigma	GLSZM	GrayLevelNonformity	0.0014277
CUS	Log_sigma	firstorder	InterquartileRange	-0.027179
CEUS	Wavelet_LH	GLCM	ClusterProminence	0.0134133
CEUS	Wavelet_LL	GLSZM	SmallAreaLowGrayLevel Emphasis	8.9129928

CUS, conventional ultrasound; CEUS, contrast enhanced ultrasound. GLSZM, Gray Level Size Zone Matrix; GLCM, Gray Level Co-occurrence Matrix.

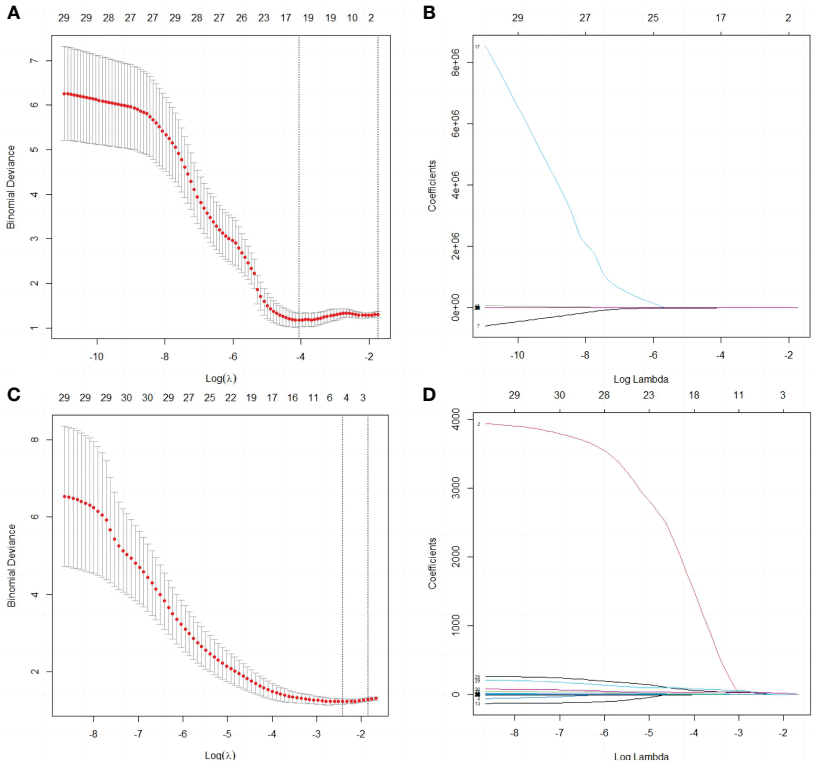


FIGURE 3
Selection of radiomics features by the LASSO analyses. **(A)** Selection of the tuning parameter λ in the LASSO analysis among candidate CUS radiomics features via 10- fold cross-validation based on the 1 standard error of the minimum criteria (1 – SE criteria). The value of λ that derived the minimum average binomial deviance was used to select features. Dotted vertical lines were drawn at the values using the minimum criteria and the 1 – SE criteria. **(B)** LASSO coefficient profiles of the 30 candidate CUS radiomics features. **(C)** Selection of the tuning parameter λ in the LASSO analysis among candidate CEUS radiomics features via 10- fold cross-validation based on the 1 standard error of the minimum criteria (1 – SE criteria). The value of λ that derived the minimum average binomial deviance was used to select features. Dotted vertical lines were drawn at the values using the minimum criteria and the 1 – SE criteria. **(D)** LASSO coefficient profiles of the 30 candidate CEUS radiomics features.

3.3 Development and validation of the nomogram

The baseline characteristics that were accessible prior to the surgical operation included age, primary tumor size and CUS-reported ALN status. The multivariate analyses further showed that CUS-reported ALN status was statistically related to ALNM (Table 3). Thus, a total of three predictors including CUS-reported ALN status, CUS-radscore and CEUS-radscore were obtained. The VIFs of these three predictors ranged from 1.269 to 1.439, indicating no multicollinearity existed among them. Next, three prediction models were respectively established by incorporating different amounts of predictors using logistic regression analysis (Model 1: CUS-reported ALN status; Model 2: CUS-reported ALN status + CUS-radscore); Model 3: CUS-reported ALN status + CUS-radscore + CEUS-radscore). The AUC values of three models in both training and external validation set were summarized (Table 4). Model 3 showed better performance than model 2 or model 1 in both the training set (AUC: 0.845 vs. 0.826 or 0.773, P = 0.4581 and P < 0.01) and external validation set (AUC: 0.901 vs. 0.889 or 0.821, P = 0.738 and P = 0.0283). Additionally, a better performance was also observed by adding CUS-radscore on the basis of CUS-reported ALN status in the training set (AUC: 0.826 vs. 0.773, P < 0.001) and the external validation set (AUC: 0.889 vs. 0.821, P = 0.013).Therefore, a nomogram was constructed using model 3 (Figure 4). The ROC curves of both the training and external validation set all showed excellent results (Figure 5). By incorporating three predictors, model 3 yielded an AUC value of 0.845 [95% confidence interval (CI), 0.739-0.950] with a sensitivity of 0.74.1% and a specificity of 92.2% in the training set, and an AUC

TABLE 3 The multivariate logistic analysis to identify independent predictor for ALNM among baseline characteristics in the training set.

Variables	Odds ratio	95% CI	P value
Age	1.026	0.967-1.088	0.403
Primary tumor size	1.496	0.986-2.270	0.058
CUS-reported ALN status (Suspicious)	12.719	3.808-42.489	< 0.01

ALNM, axillary lymph node metastasis; CUS, conventional ultrasound; CI, confidence interval.

TABLE 4 Performance of different models in training and external validation set.

Model	Training set		External validation set	
	AUC	95% CI	AUC	95% CI
1	0.773	0.672-0.875	0.821	0.682-0.961
2	0.826	0.719-0.932	0.889	0.750-1.00
3	0.845	0.739-0.950	0.901	0.758-1.00

Model 1, the model incorporating CUS-reported ALN status alone as the predictor; Model 2, the model incorporating CUS-reported ALN status and CUS-radscore as the predictors. Model 3, the model incorporating CUS-reported ALN status, CUS-radscore and CEUS-radscore as the predictors; AUC, area under curves; CI, confidence interval.

of 0.901 (95% CI, 0.758-1) with a sensitivity of 91.7% and a specificity of 85.7% in external validation set.

Calibration curves of the nomogram based on the training and external validation set were plotted to evaluate the consistency between the predicted probability of ALNM and actual pathological results of ALN (Figure 6). The calibration curves of our established nomogram showed a good fitting with the ideal curve in both training and external validation set. The decision curve analysis displayed a positive net benefit for the nomogram when a threshold probability was greater than 0.1, indicating a good clinical utility (Figure 7).

3.4 Representative examples of the nomogram in clinical practice

Then, we listed two examples of the clinical practice of our proposed nomogram. Patient 1, a 46-year-old woman, was assessed without ALNM by CUS, while an ALNM probability over 0.7 was calculated by the nomogram (Figure 8A). Pathology results confirmed metastatic ALNs were present in this patient. Patient 2, a 49-year-old woman, was assessed with ALNM by CUS, while an ALNM probability around 0.42 was calculated by the nomogram (Figure 8B). Pathology results confirmed metastatic ALNs were absent in this patient.

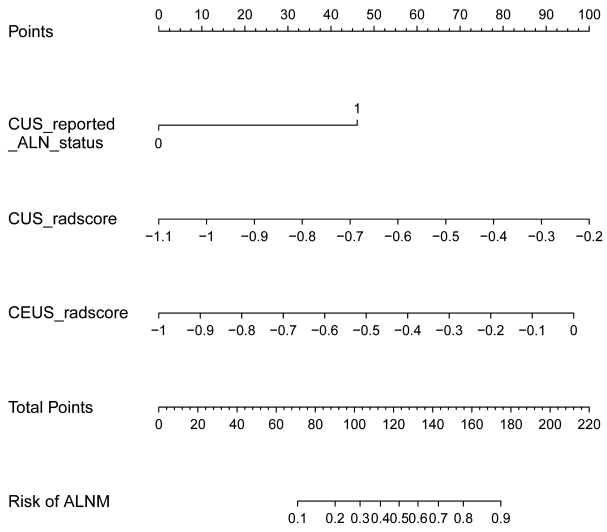


FIGURE 4
The nomogram developed in the training set using CUS-reported ALN status, CUS-radscore, and CEUS-radscore as predictors. The nomogram plot provides a visual way to calculate the risk of ALNM for breast cancer patients.

4 Discussion

In this study, we developed and validated a nomogram to predict ALNM by combining CUS-reported ALN status and radiomics features that were derived from both CUS and CEUS images. A good performance was achieved by our proposed nomogram, with an AUC of 0.845 in the training set and 0.901 in the external validation set.

The ALN status holds significant prognostic value for breast cancer. For invasive breast cancer accompanied with ALNM, ALND is an important treatment option in clinical practice. However, ALND may not suitable for all the patients with breast cancer due its associated complications (5, 6). SLN serves as the first station for lymph node metastases of primary tumors. SLND is considered as an alternative to ALND to determine the ALN status. For SLN negative patients, ALND is not recommended (25, 26). However, SLND is also associated with potential side effects and longer operation time, as well as a certain rate of false negative (7–9, 27). In the background of precision medicine, it is urgent to more effective method for uninvasive prediction of ALNM.

CUS is the routine method for noninvasive assessment of ALN status in breast cancer patients. The presence of irregular cortical thickness, longest-to-shortest axis ratio less than 2, and the absence of a fatty hilum were considered as ultrasonic features suggestive for ALNM on CUS images (23). However, it is difficult for CUS to achieve high accuracy depend on these CUS features for identifying ALNM, especially in pathological N1 patients (10, 28). Recently, several studies have found that the CUS features of primary breast lesions, including tumor size, margin, location and echogenicity were correlated with the tumor biological behavior and thus can help predict ALN status (29–33). Additionally, other studies have reported a potential association between CEUS findings and the prognosis of breast cancers (34, 35). CEUS is also a noninvasive imaging modality that can be conveniently performed on the basis of CUS screening. Different from contrast enhanced CT or MR imaging, CEUS used true intravascular contrast agents without deposition into extravascular space and was capable of reflecting the micro-vascular distribution. In common, the proliferation of vessels

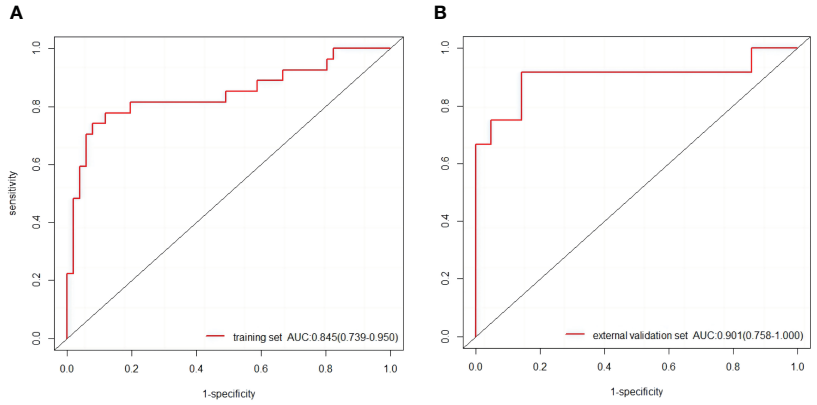


FIGURE 5
(A) The ROC curves of the nomogram in the training set. **(B)** The ROC curves of the nomogram in the external validation set.

would drive the aggressive growth of tumors (36). It was reported that coarse or twisted vessels and enhanced range of the primary tumors were independent predictors for metastatic ALNs (12, 13). The findings of these studies provide support for the feasibility of utilizing radiomics features based on CUS and CEUS to predict the ALN status.

Radiomics can objectively extract and quantitatively analyze features from medical images. In this study, radiomics features were separately extracted from CUS images and CEUS videos. An auto-tracer ROI segmentation technology was applied to outline target region on a continuous series of frames of CEUS video for each target lesion (24). Specifically, Gray Level Size Zone GLSZM (GLSZM) feature was selected as the preferred option among either CUS-radiomics or CEUS-radiomics features. GLSZM measures the size of homogeneous zones for each gray level in an image (37). Previous study demonstrated CLSZM was the optimum texture feature for breast lesion characterization (37, 38). The feature selection results obtained in our study have demonstrated that GLSZM is also optimum feature for the predicting ALNM. It is worth noting that the finally selected CEUS-radiomics features were all generated from images after wavelet transformation. The wavelet transformation is a mathematical algorithm which can mine the hidden patterns from various data, which is not visible to the naked eye (39). Compared with applying CUS-reported ALN status alone, the introduction of either CUS-radiomics features (CUS-radscore) or combined radiomics-features (CUS-radscore and CEUS-radscore) significantly increased the predictive performance. Furthermore, the combination of CUS-radscore, CEUS-radscore and CUS-reported ALN status achieved the highest AUC value, demonstrating the potential value of CEUS for ALNM prediction. Despite an improvement in AUC values for either the training set or external validation set was observed after the incorporation of the CEUS-radscore, the result of Delong test did not reveal a significant difference in AUC values between model 2 (CUS-reported ALN status + CUS-radscore) and model 3 (CUS-reported ALN status + CUS-radscore + CEUS-radscore).This could potentially be attributed to the relatively limited sample size in this study and the N stage of enrolled patients. Given that CUS-reported ALN status exhibits lower diagnostic efficacy in pathological N1 patients (10, 28), integrating predictors derived from CUS or CEUS

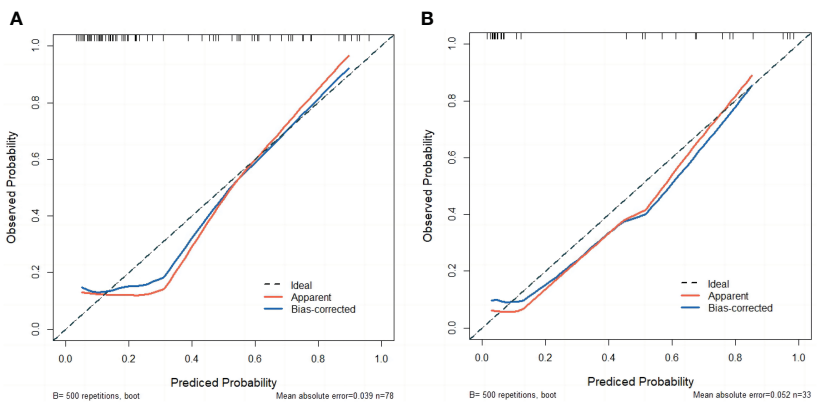


FIGURE 6
(A) The calibration curve of the nomogram in the training set. **(B)** The calibration curve of the nomogram in the external validation set.

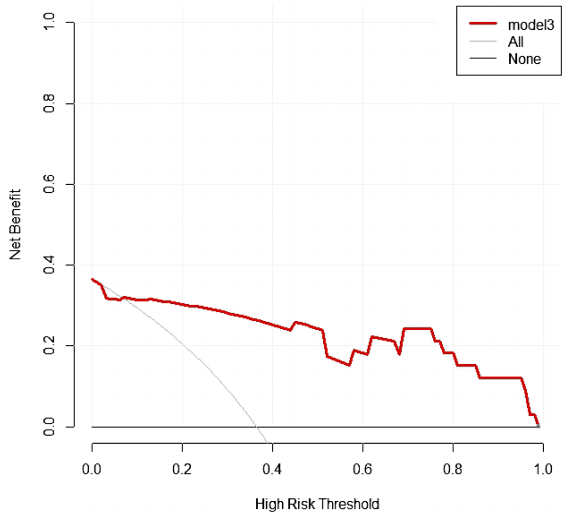


FIGURE 7
Decision curve analysis in external validation set.

radiomics into constructing a predictive model holds greater potential for improving AUC values among pathological N1 patients. However, the relatively small sample size of this study restricted the screening of pathological N1 patients for further analysis. A multi-center study with a larger sample size would help to better elucidate the value of CEUS-radscore in predicting ALNM. In clinical practice, the proposed nomogram in our study provides a practical method to quantitatively evaluate the ALNM probability among breast cancer patients. The ALNM probability calculated by the nomogram can serve as a valuable reference for clinicians in determining the necessity of implementing ALND.

There are some limitations of this study that should be mentioned. First, as we mentioned above, the sample size is relatively small and a multi-center study with larger sample size need to be further implemented. Second, a rectangular box was outlined for the ROI segmentation based on CEUS videos. Thus, the shape features derived from CEUS videos could not be used as candidate features. Third, certain ultrasonic sections for target lesions were used to present the whole 3-dimensional lesion for radiomics feature extraction. Fourth, the inter-observer and intra-observer agreement during ROI delineation was not assessed in this study.

5 Conclusions

This study has established and validated a nomogram for the prediction of ALNM in patients with breast cancer. On the basis of CUS-reported ALN status, the introduction of CUS and CEUS radiomics features derived from primary breast lesions can further improve the predictive performance. Our proposed nomogram is very important in guiding clinical decision and avoid unnecessary invasive operation.

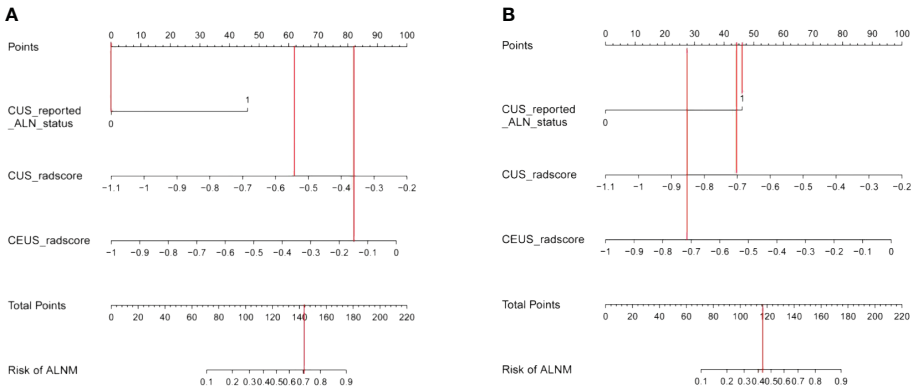


FIGURE 8
(A) Examples of nomogram evaluation of ALNM in patients with false-negative CUS findings. **(B)** Examples of nomogram evaluation of ALNM in patients with false-positive CUS findings.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the ethics committee of the Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

CS: Writing – original draft, Methodology, Formal analysis, Conceptualization. XG: Writing – review & editing, Methodology, Data curation. LH: Writing – review & editing, Software, Methodology. DY: Writing – review & editing, Resources, Investigation, Conceptualization. QL: Writing – review & editing, Methodology. LL: Writing – review & editing. YW: Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Beijing Hope Run Special Fund of Cancer Foundation of China (No. LC2019A01), China postdoctoral science foundation (No. 2017M620683), and Youth Research Fund of Peking Union Medical College (No. 2017320015).

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.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

2. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin*. (2014) 64:252–71. doi: 10.3322/caac.21235

3. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. *CA Cancer J Clin*. (2019) 69:438–51. doi: 10.3322/caac.21583

4. Park KU, Caudle A. Management of the axilla in the patient with breast cancer. *Surg Clin North Am*. (2018) 98:747–60. doi: 10.1016/j.suc.2018.04.001

5. Sakorafas GH, Peros G, Cataliotti L, Vlastos G. Lymphedema following axillary lymph node dissection for breast cancer. *Surg Oncol*. (2006) 15:153–65. doi: 10.1016/j.suronc.2006.11.003

6. Warmuth MA, Bowen G, Prosnitz LR, Chu L, Broadwater G, Peterson B, et al. Complications of axillary lymph node dissection for carcinoma of the breast: a report based on a patient survey. *Cancer*. (1998) 83:1362–8. doi: 10.1002/(sici)1097-0142(19981001)83:7<1362::aid-cncr13>3.0.co;2-2

7. Wilke LG, McCall LM, Posther KE, Whitworth PW, Reintgen DS, Leitch AM, et al. Surgical complications associated with sentinel lymph node biopsy: results from a prospective international cooperative group trial. *Ann Surg Oncol*. (2006) 13:491–500. doi: 10.1245/ASO.2006.05.013

8. Schrenk P, Wayand W. Sentinel-node biopsy in axillary lymph-node staging for patients with multicentric breast cancer. *Lancet*. (2001) 357:122. doi: 10.1016/S0140-6736(00)03550-9

9. Tee SR, Devane LA, Evoy D, Rothwell J, Geraghty J, Prichard RS, et al. Meta-analysis of sentinel lymph node biopsy after neoadjuvant chemotherapy in patients with initial biopsy-proven node-positive breast cancer. *Br J Surg*. (2018) 105:1541–52. doi: 10.1002/bjs.10986

10. Alvarez S, Añorbe E, Alcorta P, López F, Alonso I, Cortés J. Role of sonography in the diagnosis of axillary lymph node metastases in breast cancer: a systematic review. *AJR Am J Roentgenol*. (2006) 186:1342–8. doi: 10.2214/AJR.05.0936

11. Luo Y, Zhao C, Gao Y, Xiao M, Li W, Zhang J, et al. Predicting axillary lymph node status with a nomogram based on breast lesion ultrasound features: performance in N1 breast cancer patients. *Front Oncol*. (2020) 10:581321. doi: 10.3389/fonc.2020.581321

12. Zhu AQ, Li XL, An LW, Guo LH, Fu HJ, Sun LP, et al. Predicting axillary lymph node metastasis in patients with breast invasive ductal carcinoma with negative axillary ultrasound results using conventional ultrasound and contrast-enhanced ultrasound. *J Ultrasound Med*. (2020) 39:2059–70. doi: 10.1002/jum.15314

13. Du Y, Yi CB, Du LW, Gong HY, Ling LJ, Ye XH, et al. Combining primary tumor features derived from conventional and contrast-enhanced ultrasound facilitates the prediction of positive axillary lymph nodes in Breast Imaging Reporting and Data System category 4 Malignant breast lesions. *Diagn Interv Radiol*. (2023) 29:469–77. doi: 10.4274/dir.2022.22534

14. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin*. (2019) 69:127–57. doi: 10.3322/caac.21552

15. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

16. Zhang W, Wang S, Wang Y, Sun J, Wei H, Xue W, et al. Ultrasound-based radiomics nomogram for predicting axillary lymph node metastasis in early-stage breast cancer. *Radiol Med*. (2024) 129:211–21. doi: 10.1007/s11547-024-01768-0

17. Gao Y, Luo Y, Zhao C, Xiao M, Ma L, Li W, et al. Nomogram based on radiomics analysis of primary breast cancer ultrasound images: prediction of axillary lymph node tumor burden in patients. *Eur Radiol*. (2021) 31:928–37. doi: 10.1007/s00330-020-07181-1

18. Qiu X, Jiang Y, Zhao Q, Yan C, Huang M, Jiang T. Could ultrasound-based radiomics noninvasively predict axillary lymph node metastasis in breast cancer? *J Ultrasound Med*. (2020) 39:1897–905. doi: 10.1002/jum.15294

19. Harrell F. *Regression modelling strategies with application to linear models, logistic regression, and survival analysis. 1ed*. New York: Springer-Verlag New York (2001). doi: 10.1007/978-1-4757-3462-1

20. Cavaller L, Goupille C, Arbion F, Vilde A, Body G, Ouldamer L. Metastatic profiles and survival differences between infiltrating ductal carcinoma and infiltrating lobular carcinoma in invasive breast cancer. *J Gynecol Obstet Hum Reprod*. (2024) 53:102740. doi: 10.1016/j.jogoh.2024.102740

21. Duraker N, Hot S, Akan A, Nayır PÖ. A comparison of the clinicopathological features, metastasis sites and survival outcomes of invasive lobular, invasive ductal and mixed invasive ductal and lobular breast carcinoma. *Eur J Breast Health*. (2020) 16:22–31. doi: 10.5152/ejbh.2019.5004

22. Mendelson EB, Böhm-Vélez M, Berg WA. ACR BI-RADS® Ultrasound. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology (2013).

23. Koelliker SL, Chung MA, Mainiero MB, Steinhoff MM, Cady B. Axillary lymph nodes: US-guided fine-needle aspiration for initial staging of breast cancer–correlation with primary tumor size. *Radiology*. (2008) 246:81–9. doi: 10.1148/radiol.2463061463

24. Zhang K, Zhang L, Yang MH. Fast compressive tracking. *IEEE Trans Pattern Anal Mach Intell*. (2014) 36:2002–15. doi: 10.1109/TPAMI.2014.2315808

25. Galimberti V, Cole BF, Viale G, Veronesi P, Vicini E, Intra M, et al. Axillary dissection versus no axillary dissection in patients with breast cancer and sentinel-node micrometastases (IBCSG 23-01): 10-year follow-up of a randomised, controlled phase 3 trial. *Lancet Oncol*. (2018) 19:1385–93. doi: 10.1016/S1470-2045(18)30380-2

26. Meretoja TJ, Leidenius MH, Heikkilä PS, Boross G, Sejben I, Regitnig P, et al. International multicenter tool to predict the risk of nonsentinel node metastases in breast cancer. *J Natl Cancer Inst*. (2012) 104:1888–96. doi: 10.1093/jnci/djs455

27. Pesek S, Ashikaga T, Krag LE, Krag D. The false-negative rate of sentinel node biopsy in patients with breast cancer: a meta-analysis. *World J Surg*. (2012) 36:2239–51. doi: 10.1007/s00268-012-1623-z

28. Zhang YN, Wang CJ, Xu Y, Zhu QL, Zhou YD, Zhang J, et al. Sensitivity, specificity and accuracy of ultrasound in diagnosis of breast cancer metastasis to the axillary lymph nodes in Chinese patients. *Ultrasound Med Biol*. (2015) 41:1835–41. doi: 10.1016/j.ultrasmedbio.2015.03.024

29. Akissue de Camargo Teixeira P, Chala LF, Shimizu C, Filassi JR, Maesaka JY, de Barros N. Axillary lymph node sonographic features and breast tumor characteristics as predictors of Malignancy: A nomogram to predict risk. *Ultrasound Med Biol*. (2017) 43:1837–45. doi: 10.1016/j.ultrasmedbio.2017.05.003

30. Bae MS, Shin SU, Song SE, Ryu HS, Han W, Moon WK. Association between US features of primary tumor and axillary lymph node metastasis in patients with clinical T1-T2N0 breast cancer. *Acta Radiol*. (2018) 59:402–8. doi: 10.1177/0284185117723039

31. Li XL, Xu HX, Li DD, He YP, Yue WW, Xu JM, et al. A risk model based on ultrasound, ultrasound elastography, and histologic parameters for predicting axillary lymph node metastasis in breast invasive ductal carcinoma. *Sci Rep*. (2017) 7:3029. doi: 10.1038/s41598-017-03582-3

32. Guo Q, Dong Z, Zhang L, Ning C, Li Z, Wang D, et al. Ultrasound features of breast cancer for predicting axillary lymph node metastasis. *J Ultrasound Med*. (2018) 37:1354–3. doi: 10.1002/jum.14469

33. Zhang H, Sui X, Zhou S, Hu L, Huang X. Correlation of conventional ultrasound characteristics of breast tumors with axillary lymph node metastasis and ki-67 expression in patients with breast cancer. *J Ultrasound Med*. (2019) 38:1833–40. doi: 10.1002/jum.14879

34. Wan CF, Du J, Fang H, Li FH, Zhu JS, Liu Q. Enhancement patterns and parameters of breast cancers at contrast-enhanced US: correlation with prognostic factors. *Radiology*. (2012) 262:450–9. doi: 10.1148/radiol.11110789

35. Moon WK, Im JG, Noh DY, Han MC. Nonpalpable breast lesions: evaluation with power Doppler US and a microbubble contrast agent-initial experience. *Radiology*. (2000) 217:240–6. doi: 10.1148/radiology.217.1.r00oc03240

36. Yilmaz R, Demir AA, Kaplan A, Sahin D, Ozkurt E, Dursun M, et al. Magnetic resonance imaging features of idiopathic granulomatous mastitis: is there any contribution of diffusion-weighted imaging in the differential diagnosis? *Radiol Med*. (2016) 121:857–66. doi: 10.1007/s11547-016-0666-x

37. Osapoetra LO, Chan W, Tran W, Kolios MC, Czarnota GJ. Comparison of methods for texture analysis of QUS parametric images in the characterization of breast lesions. *PloS One*. (2020) 15:e0244965. doi: 10.1371/journal.pone.0244965

38. Stamatopoulou V, Parisot P, De Vleeschouwer C, Lafontaine DLJ. Use of the iNo score to discriminate normal from altered nucleolar morphology, with applications in basic cell biology and potential in human disease diagnostics. *Nat Protoc*. (2018) 13:2387–406. doi: 10.1038/s41596-018-0044-3

39. Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. *AJNR Am J Neuroradiol*. (2010) 31:809–16. doi: 10.3174/ajnr.A2061

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 12 June 2024
DOI 10.3389/fonc.2024.1404148

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Samaneh Zolghadri,
Islamic Azad University of Jahrom, Iran
Dalin Zhang,
Stanford University, United States

*CORRESPONDENCE
Xining Li
hzlixining@yeah.net
Fengfeng Wu
hzwufengfeng@163.com

†These authors have contributed equally to this work

RECEIVED 20 March 2024
ACCEPTED 30 May 2024
PUBLISHED 12 June 2024

CITATION
Li J, He H, Liu S, Li X and Wu F (2024) Revealing tumor cells and tissues with high selectivity through folic acid-targeted nanofluorescence probes responsive to acidic microenvironments.
Front. Oncol. 14:1404148.
doi: 10.3389/fonc.2024.1404148

COPYRIGHT
© 2024 Li, He, Liu, Li and Wu. 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.

Revealing tumor cells and tissues with high selectivity through folic acid-targeted nanofluorescence probes responsive to acidic microenvironments

Jing Li[1†], Hongyi He[2†], Shuyan Liu[3†], Xining Li[4*] and Fengfeng Wu[5,6*]

[1]Neurobiology Laboratory, Wannan Medical College, Wuhu, China, [2]College of Pharmacy, Hubei University of Science and Technology, Xianning, China, [3]Department of Obstetrics and Gynecology, Second Hospital of Jilin University, Changchun, China, [4]School of Medicine, Huzhou University, Huzhou, China, [5]Department of Orthopedics and Rehabilitation, Huzhou Hospital of Zhejiang University School of Medicine, Huzhou, China, [6]Department of Orthopedics and Rehabilitation, Huzhou Central Hospital, Huzhou, China

Tumor-specific fluorescent probes must fulfill the dual requirements of targeted accumulation within tumors and high-resolution imaging capabilities. To achieve both tumor-targeted accumulation and high-resolution imaging performance, we developed a composite comprising an acid-responsive bodipy conjugated to amphiphilic PEG-b-PLA polymer, along with folic acid (FA)-modified PEG-b-PLA as a targeting moiety for active tumor-specific accumulation. Finally, a novel assembly of hybrid fluorescent nanoparticles was successfully synthesized by integrating these two components, demonstrating exceptional responsiveness to acidic conditions for fluorescence excitation and remarkable tumor-targeted accumulation capabilities. We conducted comprehensive *in vitro* and *in vivo* investigations employing techniques such as analysis of physicochemical properties, fluorescence-based probes detection at varying pH levels, assessment of *in vitro* cytotoxicity, evaluation of cellular uptake capacity, analysis of lysosomal co-localization imaging, examination of tumor fluorescence images *in vivo*, and investigation of biological distribution patterns. The results demonstrated that the acid-responsive nanofluorescence probe we designed and synthesized possesses desirable physical and chemical characteristics, including a small particle size and low cytotoxicity. Moreover, it exhibits rapid real-time response to acidic environments and displays enhanced fluorescence intensity, enabling the real-time tracking of probe entry into tumor cells as well as intracellular lysozyme accumulation. We achieved highly specific *in vivo* tumor visualization by combining nanoprobes targeting folate receptor. Through imaging cervical tumor mice, we demonstrated the precise imaging performance and high targeted accumulation of FA-targeted nanofluorescence probes in tumor tissue. Furthermore, we confirmed the *in vivo* safety of the FA-

targeted nanofluorescence probe through biological distribution analysis. These findings highlight the potential widespread application of FA-targeted acid-responsive nanofluorescence probes for selective imaging of tumor cells and tissues.

KEYWORDS

tumor cells, tissues, folic acid, nanofluorescence, probes, acidic microenvironments

1 Introduction

Fluorescence imaging has emerged as a convenient and safe tracer technique for labeling various tissues in organisms (1–3). To date, numerous fluorescent probes have been developed, with 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene (BDY) garnering significant attention owing to its robust ultraviolet absorption, exceptional physical stability, high fluorescence peak intensity, adjustable emission properties, and superior quantum yield (4). Despite the extensive development of BDY-based probes for diverse biomedical applications (5), their tumor targeting distribution and specific imaging efficiency often fall short of ideal expectations. The reason can be attributed to the challenges encountered by BDY, a small molecule fluorescent probe, which are similar to those faced by other small molecule fluorescent probes. For instance, the majority of small molecule fluorescent probes exhibit hydrophobic properties and are readily absorbed by the reticuloendothelial system following intravenous administration (6), consequently leading to a reduced half-life in blood circulation. Furthermore, the utilization of small molecule fluorescent probes often necessitates their conjugation with various biological entities such as antibodies, peptides, and nucleic acids to achieve a certain level of targeting specificity (7–9). However, such binding interactions significantly compromise the sensitivity of *in vivo* probe detection. Consequently, there is an urgent imperative to develop highly tumor-sensitive fluorescent probes possessing superior tumor-targeting properties.

Currently, nanoparticles are widely employed for the efficient transportation of fluorescent probes in various applications (10). In comparison to conventional biomarker techniques, nanoprobes offer several advantages including uniform particle size distribution, stable optical properties, minimal cytotoxicity, excellent water solubility, and remarkable biocompatibility (11). Therefore, they exhibit enhanced resolution, heightened fluorescence intensity, and superior detection capabilities. Specifically, PLA-b-PEG nanocarriers possess the advantages of excellent biocompatibility, biodegradability, and prolonged *in vivo* circulation time, rendering them extensively employed for the transportation of fluorescent probes. Generally, nanoparticles can passively target tumor sites through their enhanced permeability and retention effect by delivering various probes (12, 13). Additionally, many types of nanoparticles are surface-modified with specific targeting groups to enhance their accumulation in tumors while reducing uptake in normal tissues (14–17). The same approach can also be employed for the delivery of fluorescent probes, thereby endowing fluorescent nanoparticles with not only enhanced tumor tissue targeting capabilities but also heightened sensitivity in detecting tumor tissues. The achievement of this goal necessitates addressing two pivotal issues: firstly, the fluorescence emission of nanoparticles in normal tissues should be suppressed until their internalization by tumor tissue and subsequent exposure to acidic environments within the tumor, thereby enabling significant specific fluorescence emission. Another crucial aspect that necessitates attention is the effective integration of tumor-specific highly expressed substances as targeting moieties with nanoprobes, thereby augmenting their accumulation within tumors while minimizing uptake in normal tissue. To address these dual challenges concurrently, this study proposes the utilization of two carriers to fabricate multifunctional composite nanoparticles. One carrier is required to load a substantial quantity of acid-sensitive components, while the other carrier needs to accommodate a significant number of targeted groups.

To achieve this objective, we synthesized two polymers by combining an acid-responsive dye (BDY) or folate receptor-targeting ligand (FA) with amphiphilic polyethylene glycol-b-polylactic acid (PEG-b-PLA) polymer. Subsequently, these components were assembled into composite fluorescent nanoparticles, and their physical and chemical properties, pH-dependent fluorescence emission behavior, *in vitro* cytotoxicity spectrum, cellular uptake efficiency in tumor cells, lysosomal co-localization imaging ability, as well as comprehensive analysis of *in vitro* and *in vivo* tumor fluorescence images were investigated.

2 Materials and methods

2.1 Materials

The following reagents were obtained from Sigma-Aldrich Co. (St. Louis, MO, USA): N-hydroxysuccinimide (NHS), N,N-diisopropylethylamine (DIPEA), 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ), N,N'-dicyclohexylcarbodiimide (DCC), dimethylformamide (DMF), tetrahydrofuran (THF). Additionally,

FA and N-acetyl-L-cysteine were also purchased from Sigma-Aldrich Co. Always-on probes were procured from GlpBio (Cat. No.: GC66041, USA).

2.2 Synthesis of pH-activatable nanoprobes

Dissolve 2,4-dimethylpyrrole and the corresponding 4-substituted benzaldehyde in dichloromethane, followed by conducting a condensation reaction. After stirring the resulting reddish solution overnight, DDQ was introduced and the mixture was stirred for 2 h. Subsequently, DIPEA and BF_3OEt_2 were added to the reaction mixture. After an additional 1 h of reaction time, the mixture was concentrated under reduced pressure and then redissolved in EtOAc before being washed with water. The aqueous layer was extracted with EtOAc, while the combined organic layers were dried, filtered, and subsequently concentrated under reduced pressure. Finally, purification of the residue using silica gel column chromatography yielded BDY1.

The synthesis of BDY2 was achieved through the Vilsmeier-Haack reaction. In brief, a mixture of DMF and $POCl_3$ was stirred at 0°C for 5 min, followed by warming to room temperature and stirring for an additional 30 min. Subsequently, BDY2 in dichloroethane was added to the mixture and stirred at 50°C for 2 h. After cooling to room temperature, the mixture was washed with $NaHCO_3$ and water. The organic layers were then evaporated and dried under vacuum conditions. Finally, purification of the crude product was accomplished using column chromatography resulting in the isolation of pure BDY2.

The synthesis of BDY2-PEG-b-PLA was conducted as follows: NH_2-PEG-b-PLA and BDY2 were dissolved in THF and stirred at room temperature overnight. Subsequently, $NaBH_4$ was added and the mixture was stirred for an additional 12 h. The resulting product was purified through dialysis in methanol for 24 h followed by freeze-drying for subsequent use.

2.3 Preparation of FA-targeted pH-activatable nanoprobes

At 0°C, DCC and NHS were added to DMF dissolved with FA and stirred for 12 h. Subsequently, NH_2-PEG-b-PLA was added and the mixture was stirred for an additional 12 h. The resulting solution was concentrated, precipitated in ether, collected, and dried under vacuum. Hybrid micelles were also prepared using co-assembly methods. Briefly, BDY2-PEG-b-PLA and FA-PEG-b-PLA were dissolved in THF followed by sonication for 2 min. The polymer solutions were then dropwise added to deionized water while vigorously stirring the suspension for 10 h until no THF residue could be detected.

2.4 *In vitro* release

The nanoprobes were suspended in phosphate buffer saline (PBS) with a pH of 7.4, containing Tween-80 at a concentration of 1.0 wt%. Subsequently, the solution was added to a dialysis bag with a molecular weight cutoff of 3.5 kDa for release testing purposes. Briefly, the sealed dialysis bag was placed inside a vial and immersed in PBS. The vial was then subjected to shaking on a table at a frequency of 1 Hz and maintained at 37°C. At regular intervals of every 2 h, medium samples were withdrawn from the vial and replaced with an equal volume of fresh buffer solution. The amount of BDY released was determined through UV-vis analysis using a detection wavelength set at 500 nm.

2.5 Fluorescence detection of probes

The fluorescence intensity of always-on probes and pHAN was measured in acidic (pH=4.4), neutral (pH=7.4), and basic (pH=8.4) PBS solutions, respectively. The excitation wavelength used was 488 nm, and emission spectra were obtained from 515 nm onwards. Unmixed images were generated by employing authentic spectral patterns while considering the background.

2.6 Cell line and cell culture

The human cervical epithelial cell line (HCerEpic cells) and human cervical carcinoma cell line (HeLa cells) were obtained from the cell bank of the Shanghai Institutes for Biological Sciences. The cells were cultured at 37°C in a humidified atmosphere with 5% (v/v) CO_2 using Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum, as well as streptomycin (100 U/mL) and penicillin (100 mg/mL).

2.7 Cytotoxicity test

Cells were seeded in 96-well plates at a density of 1×10^6 cells/well and incubated in DMEM for 24 h. Subsequently, the medium was replaced with various concentrations of probes and further incubated for 48 h. Following this, 3-(4,5-Dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide (MTT) solution (5 mg/mL) in PBS was added to each well and the plates were incubated for an additional 4 h at 37°C. The culture medium containing MTT was then removed, and dimethyl sulfoxide was added to dissolve the formazan crystals in each well. Finally, the plates were shaken for 10 min and the absorbance of formazan product was measured at 570 nm using a microplate reader (Bio-RAD Model 680).

2.8 Confocal laser scanning microscopy

Sterile coverslips were placed in a six-well plate at a density of 2×10^4 cells/well, and HeLa cells were cultured at 37°C for 24 h until they adhered to the well. The probes were then co-cultivated with HeLa cells for 0–8 h at 37°C. After washing the cells three times with PBS (pH=7.4), probe uptake was observed using CLSM (FluoView FV1000, Olympus). To investigate the co-localization phenomenon between the probe and lysosome, the prepared LysoTracker

solution was added to the treated cells and incubated at 37°C for 30 min. Subsequently, the cells were fixed with 4% paraformaldehyde at 4°C for 10 min and washed twice with PBS. The nuclei were stained by adding a solution of 4',6-diamidino-2-phenylindole (DAPI) in PBS, followed by gentle washing to remove excess dyes. Finally, CLSM was employed to capture images of each culture dish after co-culturing the cells and probes for 4 h. In order to evaluate the tumor cell targeting function of FA on the surface of FA-pHAN, we pre-incubated the cells with free FA (4 mM) for 30 min prior to introducing FA-pHAN. Subsequently, we added FA-pHAN to block the binding between the probes' FA and its receptors on tumor cells, thereby impeding the targeted effect of FA-pHAN on tumor cells. CLSM was employed to capture images of each culture dish after co-culturing the cells and probes for 4 h.

2.9 *In vivo* fluorescence imaging

Animal experiments were conducted in accordance with the National Institutes of Health standards and the Guidelines for the Care and Use of Experimental Animals, and were approved by the Animal Experiment Ethics Committee of Huzhou University. Cervical tumors were induced in Kunming mice through subcutaneous injection of 2×10^6 HeLa cells suspended in 200 μL of PBS. Comparative experiments were then conducted using nude mice bearing HeLa cervical tumors to directly compare always-on probes, pHAN, and FA-pHAN for specificity monitoring at a post-injection time point of 2 h. For tumor imaging, pHAN (10 mg/kg) or FA-pHAN (10 mg/kg) was intravenously administered to the tumor-bearing mice. To compare probe accumulation in normal and tumor tissues, we injected the same probe into two different sites of each mouse-namely, the tumor tissue on one leg and the corresponding normal tissue on the opposite leg. Subsequently, euthanasia anesthesia was used to sacrifice the mice. The excitation wavelength for probe activation was set at 488 nm, and emission spectra were obtained at 515 nm. Unmixed images were generated using actual spectral patterns and backgrounds. By comparing average fluorescence intensity between the tumor site and whole body excluding the tumor site, we determined the ratio of tumor to normal tissue (T=N).

2.10 Biological distribution

After 24 h of administration, the major organs and tumors were dissected, weighed, and their measurements recorded. Each tissue sample was placed in an EP tube and homogenized with physiological saline using a homogenizer. Subsequently, methanol and chloroform were added separately and vortexed for 5 min. The lower liquid phase was collected after centrifugation. Following natural evaporation, dimethyl sulfoxide was added for re-dissolution. The upper clear liquid phase was then quantified based on the established fluorescence standard curve to calculate the Bodipy content in each sample.

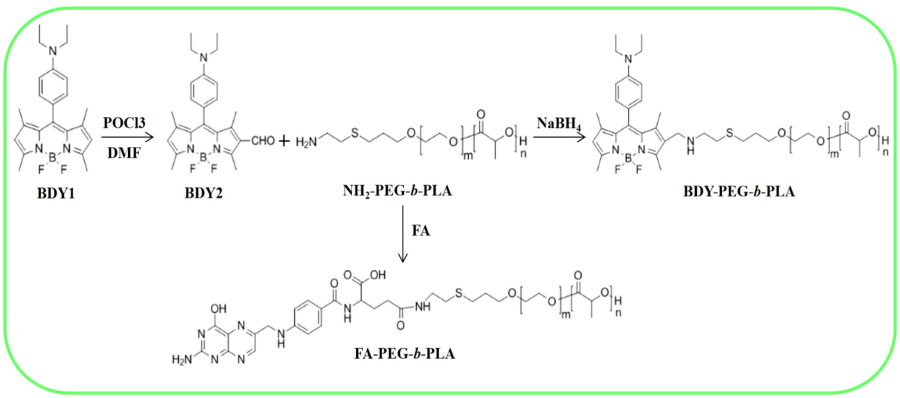
2.11 Statistical analysis

All data are presented as the mean ± standard deviation and were analyzed using one-way analysis of variance followed by Tukey's *post-hoc* test. A *P*-value of less than 0.05 was considered to indicate a statistically significant difference.

3 Results

3.1 Physical and chemical properties of probes

As depicted in Scheme 1, we synthesized two copolymers, namely BDY2-PEG-b-PLA and FA-PEG-b-PLA, for the assembly of FA-pHAN. BDY2 was utilized to react with the amino group of NH_2-PEG-b-PLA, forming a schiff base which could be subsequently reduced to stable carbon nitrogen bonds by $NaBH_4$ treatment, resulting in the formation of BDY2-PEG-b-PLA. Additionally, employing the DCC/NHS method, we conjugated the target



SCHEME 1
Design strategy for nanoprobes with pH-responsiveness and tumor targeting.

molecule FA onto NH_2-PEG-b-PLA polymer to generate FA-PEG-b-PLA. As illustrated in Scheme 2, an acid-responsive component comprising BDY coupled with amphiphilic PEG-b-PAL polymer was prepared and further modified with FA as an active targeting moiety to yield another component possessing active targeting capability. Finally, fluorescent nanoparticles were assembled from these two components and exhibited both acid-responsive fluorescence excitation ability and tumor-targeted accumulation performance. As depicted in Figures 1A, B, both pHAN and FA-pHAN exhibit typical spherical morphology with diameters of 56 nm and 62 nm, respectively, while demonstrating exceptional monodispersity in aqueous solutions. The specific physical and chemical parameters of the two copolymers and nanoparticles are presented in Figure 1C. The BDY2 compound, which contains aldehyde groups, was synthesized via the Vilsmeier-Haack reaction with a yield of 80%. Composite nanoparticles were prepared by mixing BDY2-PEG-b-PLA with FA-PEG-b-PLA to achieve pH-responsive properties and tumor tissue targeting. The mixed nanoparticles had a FA content of 20wt%. These results further demonstrate that the nanoprobes we developed meet the requirements of an ideal drug delivery system. The diameters of pHAN and FA-pHAN, as measured by DLS, remained relatively constant over a period of 2 weeks (see Figure 1D). The release of pHAN and FA-pHAN is depicted in Figure 2, illustrating a gradual and sustained release of BDY from both materials. Within 14 h, approximately 60% of BDY was released cumulatively. Subsequently, the release rate for both probes decelerated, eventually reaching a plateau after 16 h. To validate the pH-dependent fluorescence emission characteristics of the probe, we conducted measurements of fluorescence intensity across various pH values. The probe exhibited maximum absorption at approximately 515 nm. As depicted in Figures 3A, B, the fluorescence intensity of the always-on probes demonstrated no discernible correlation with the pH levels present within the buffer. The fluorescence intensity of pHAN, in contrast, exhibits a decrease as the pH increases, indicating an absence of fluorescence at physiological pH and a progressively increasing trend at lower pH levels.

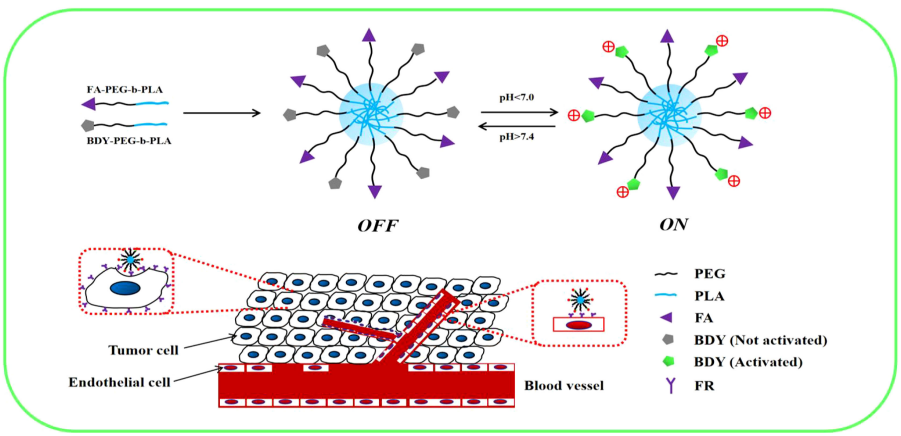
3.2 Cytotoxicity of probes

The MTT method was employed to investigate the potential cytotoxicity of probes on HCerEpic cells and HeLa cells at doses ranging from 0.064 to 200 μg/mL, as depicted in Figure 4. Following a co-culture with pHAN and FA-pHAN for 48 h, the viability of HCerEpic cells and HeLa cells remained above 90% irrespective of whether pHAN or FA-pHAN was administered at low or high doses, thereby indicating negligible *in vitro* cytotoxicity associated with pHAN and FA-pHAN during cell experiments.

3.3 Cell uptake of probes

As depicted in Figures 5A, B, the cells were co-cultured with pHAN for a duration of 0–8 h. As the phagocytosis of pHAN by the cells progressed, there was a significant enhancement in fluorescence signals at 4 h of incubation compared to 0 h, and stronger fluorescence signals were observed at 8 h of incubation compared to 4 h. Conversely, as illustrated in Figures 5C, D, nearly identical green fluorescence was detected in the cells after co-culturing with the always-on probes for a duration of 0–8 h. These findings suggest that pHAN exhibits substantial intracellular accumulation and fluorescence activation starting from approximately 4 h post-ingestion by tumor cells.

To elucidate the efficacy of targeting functional group FA in facilitating cellular uptake of probes, we incubated cells with pHAN or FA-pHAN for 4 h and assessed intracellular fluorescence signals. As depicted in Figures 6A, B, the fluorescence intensity observed in the FA-pHAN group was significantly higher compared to that of the pHAN group, indicating that FA labeling can expedite cell internalization of pHAN. Pre-incubation of free FA in the cell culture medium resulted in reduced uptake of FA-pHAN by cells, exhibiting a fluorescence intensity similar to that observed in the pHAN uptake group. This data further substantiates that enhanced cellular uptake of FA-pHAN is attributed to specific interactions between FA and its receptors on tumor cells.



SCHEME 2
Synthesis process of BDY2-PEG-b-PLA and FA-PEG-b-PLA.

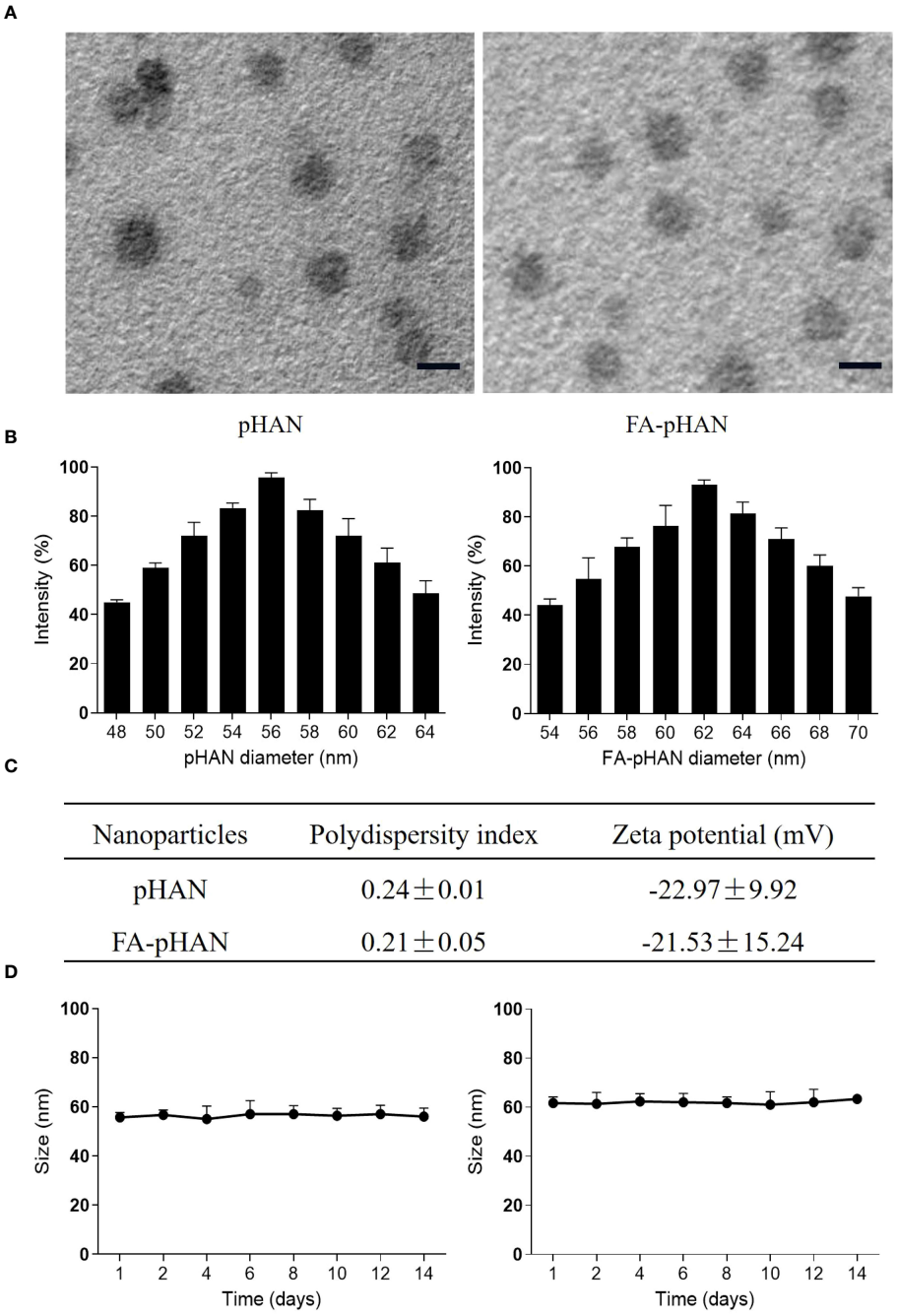


FIGURE 1
(A) TEM images of pHAN and FA-pHAN, with a scale bar of 50 nm. **(B)** Diameter of pHAN and FA-pHAN. **(C)** Polydispersity index and zeta potential measurements of the nanoparticles. **(D)** The size of pHAN and FA-pHAN undergoes changes over time. pHAN, pH-activatable nanoprobes; FA-pHAN, FA-targeted pH-activatable nanoprobes.

3.4 Lysosome colocalization imaging of probes

Upon engulfment by cells, pHAN gradually accumulates within lysosomes, where its green fluorescence is activated due to the acidic nature of these organelles. The resulting yellow fluorescence arises from the superimposition of the green fluorescence emitted by pHAN probes and the red fluorescence emitted by lysosomes. As depicted in Figures 7A, B, negligible green fluorescence was observed in cells after co-culturing with pHAN for 0 h. However, as the co-culture time extended to 4 h and 8 h, enhanced green and yellow fluorescence became evident in HeLa cells. Subsequent colocalization studies revealed that these yellow fluorescent spots were predominantly localized within acidic lysosomes.

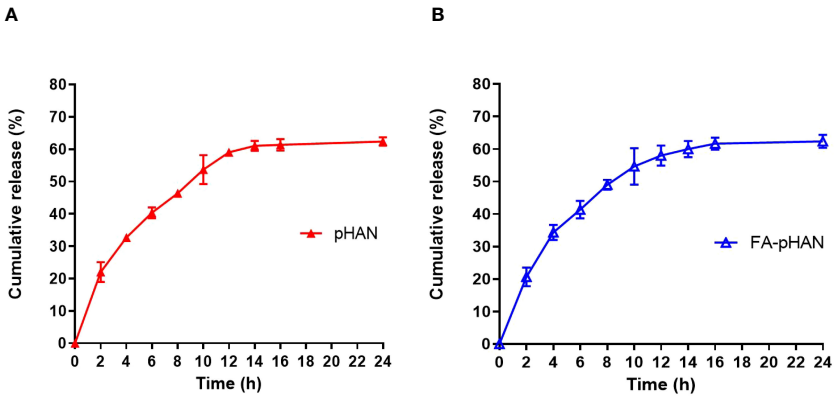


FIGURE 2
Cumulative release profile of the fluorescent dye BDY from pHAN **(A)** and FA-pHAN **(B)**. pHAN, pH-activatable nanoprobes; FA-pHAN, FA-targeted pH-activatable nanoprobes.

3.5 Fluorescence images of tumor bearing mice

Fluorescence imaging was conducted 2 h post intravenous administration of each probe group in mice with subcutaneously transplanted HeLa cells. As depicted in Figure 8, a significant fluorescence signal was observed exclusively at the tumor site in mice injected with pHAN, while no fluorescence emission was detected in normal tissues. Conversely, the always-on probes group exhibited substantial fluorescence signals both in tumor and normal tissues. The fluorescence intensity observed within the tumor tissue of FA-pHAN injection group mice increased by a factor of 11 compared to the pHAN group, whereas no notable fluorescence signal was discernible in normal tissue.

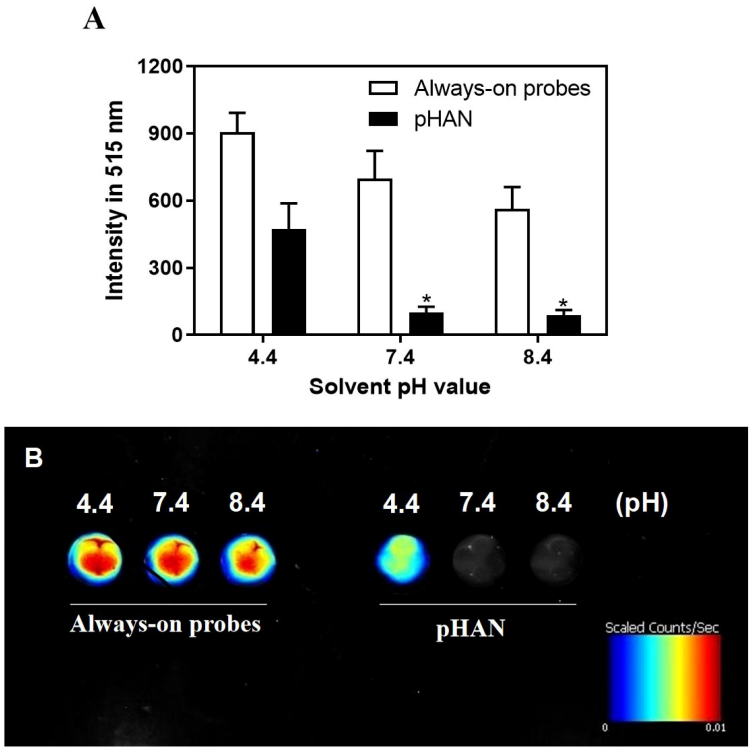


FIGURE 3
(A) Fluorescence intensity comparison between always-on probes and pHAN at different pH values, showing changes in fluorescence spectra under acidic conditions (pH=4.4), neutral conditions (pH=7.4), and basic conditions (pH=8.4). * $P < 0.05$ compared to the pH 4.4 group for statistical significance analysis purposes only. **(B)** Fluorescence imaging comparison between always-on probes and pHAN. pHAN, pH-activatable nanoprobes; FA-pHAN, FA-targeted pH-activatable nanoprobes.

10.3389/fonc.2024.1404148

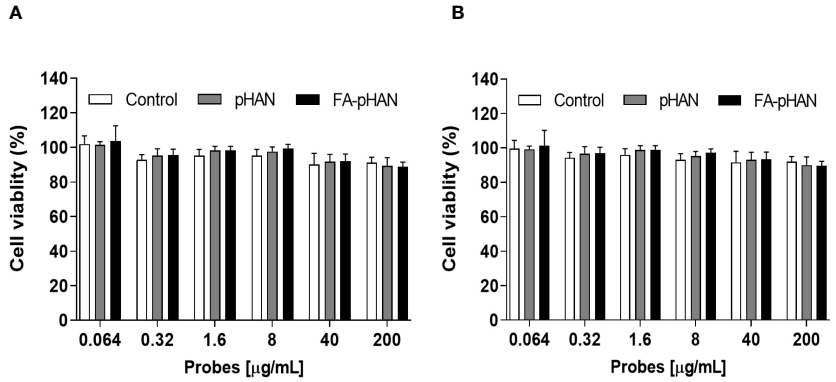


FIGURE 4
In vitro cytotoxicity of pHAN and FA-pHAN against HCerEpic cells **(A)** and HeLa cells **(B)** at 48 h. pHAN, pH-activatable nanoprobes; FA-pHAN, FA-targeted pH-activatable nanoprobes.

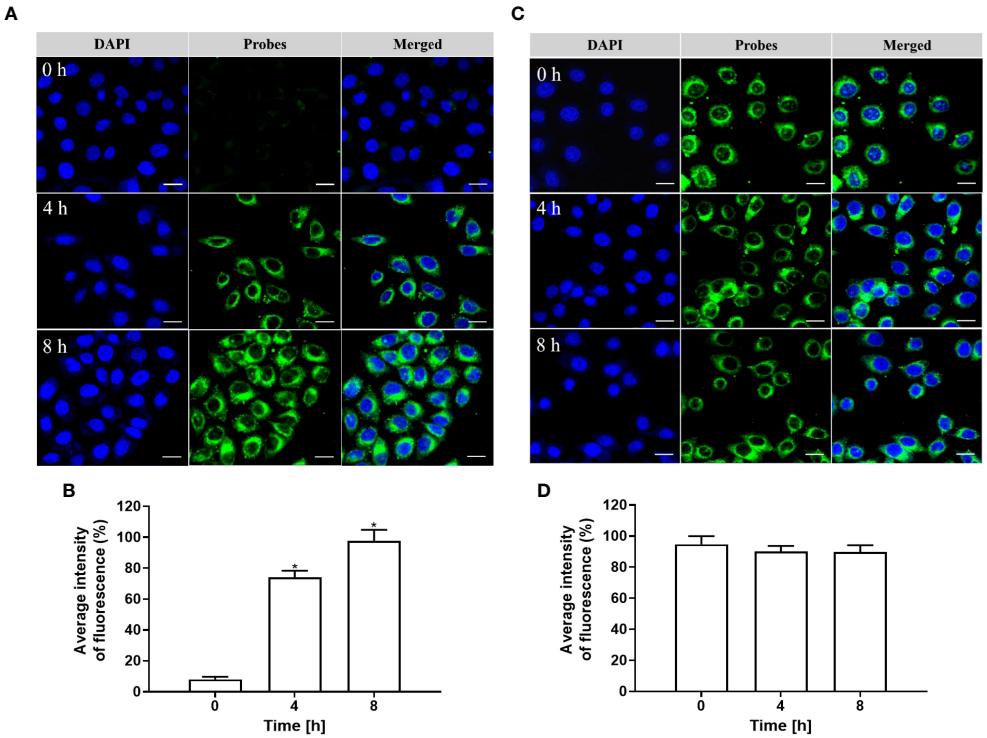


FIGURE 5
(A) CLSM images of HeLa cells cultured at 37°C with pHAN for 0 h, 4 h, and 8 h are shown. The green channel represents pHAN fluorescence, while the blue channel shows DAPI staining. A scale bar of 20 μm is included for reference. **(B)** Quantitative analysis was performed on HeLa cells cultured at 37°C with pHAN for a duration of 0–8 h (*$P < 0.05$ compared to the 0 h group). **(C)** CLSM images depict HeLa cells cultured at 37°C with always-on probes for time points of 0 h, 4 h, and 8 h. The green channel corresponds to always-on probes fluorescence, while the blue channel indicates DAPI staining. A scale bar of 20 μm is provided as a size reference. **(D)** Quantitative analysis was conducted on HeLa cells cultured at a temperature of 37°C using always-on probes over a period ranging from 0–8 h.V pHAN, pH-activatable nanoprobes.

3.6 Biological distribution of probes in major organs

To investigate the *in vivo* activity of the probe, we conducted an examination of the organ distribution of the drug in animal subjects. In this study, mice were administered the drug for a period of 24 h. The findings, depicted in Figure 9, reveal a higher presence of permanently open probes in the liver and spleen. When compared to always-on probes, there was a slight reduction in the distribution of pHAN within these organs, while FA-pHAN exhibited a significant decrease. Notably, there was a slight increase in pHAN content within tumor tissue and a substantial increase in FA-pHAN content within tumor tissue when compared to that observed with the always-on probe. This observation suggests that targeted accumulation of nanoprobes featuring specific targeting groups at tumor sites may serve to diminish

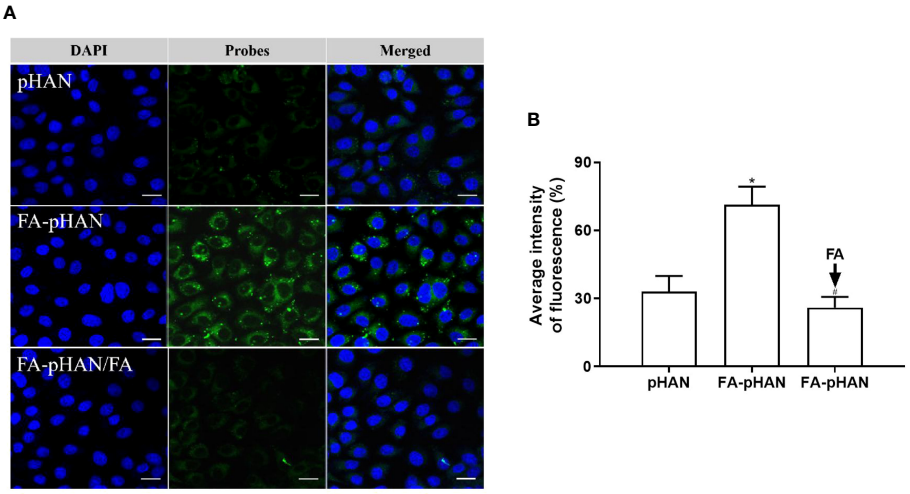


FIGURE 6
(A) CLSM images of HeLa cells incubated with pHAN, FA-pHAN, or competitive assay with free FA added 10 min prior to FA-pHAN administration. Green channel represents pHAN or FA-pHAN, while blue channel shows DAPI staining. Scale bar: 20 μm. **(B)** Quantitative analysis of HeLa cells treated with pHAN, FA-pHAN, or FA-pHAN/FA for 4 h. *$P < 0.05$ compared to the pHAN group; # $P < 0.05$ compared to the FA-pHAN group. pHANm, pH-activatable nanoprobes; FA-pHAN, FA-targeted pH-activatable nanoprobes.

probe distribution within other organs while concurrently enhancing their accumulation within tumor tissues.

4 Discussion

Recently, multifunctional nanoparticles have garnered significant attention in various applications such as biosensors, nanoprobes, and targeted drug delivery. These endeavors primarily stem from the imperative need for precise control of drug release to enhance biological specificity in diagnosis and treatment. To accomplish this objective, researchers have been dedicated to developing stimulus-responsive nano platforms that can be triggered by environmental cues including pH variations, enzyme expression levels, redox reactions, and light sources (18–20). Among these activation signals, the pH difference between tumor tissue and normal tissue or various organelles within tumor cells has garnered widespread attention as a stimulation mode for inducing fluorescence activation of nanoprobes (21). Several pH-responsive nanosystems have been reported to enhance the sensitivity or therapeutic effectiveness of tumor imaging (21–24). In addition, folate receptors are overexpressed on the surface of many malignant tumor cells (25, 26), while they are rarely or only slightly expressed in normal cells (27). Therefore, by modifying the vector's surface with FA as the targeting group based on the

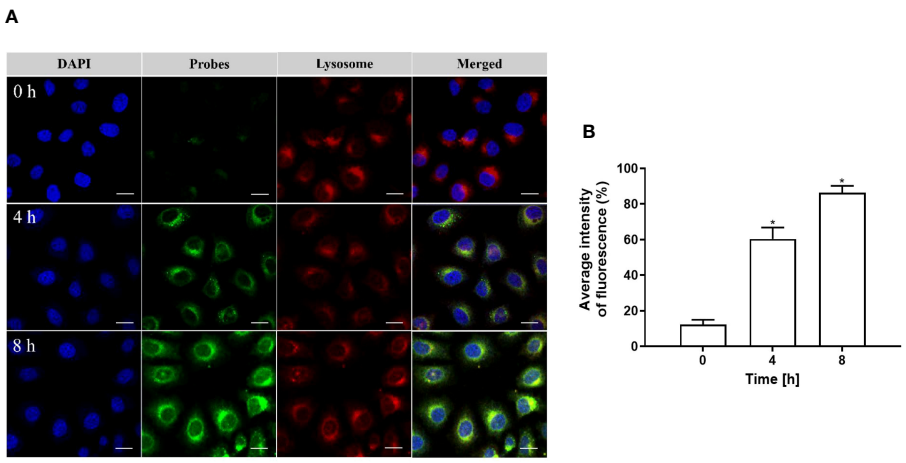


FIGURE 7
(A) CLSM images depicting the co-localization of pHAN with lysosomes in HeLa cells cultured at 37°C for 0 h, 4 h, and 8 h. The green channel represents pHAN, the red channel represents lysosomes, and the blue channel represents DAPI-stained nuclei. Scale bar: 20 μm. **(B)** Quantitative analysis demonstrating the degree of co-localization between pHAN and lysosomes. *$P < 0.05$ compared to the 0 h group. pHAN, pH-activatable nanoprobes.

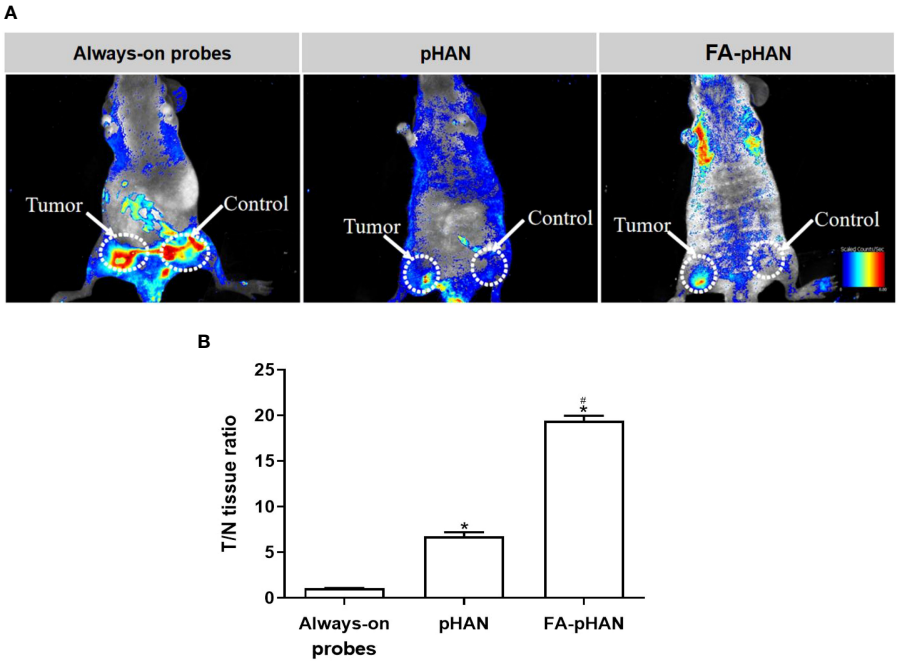


FIGURE 8
(A) Overlaid fluorescent images of HeLa-tumor-bearing mice at 2 h post-injection of always-on probes, pHAN and FA-pHAN. **(B)** Near-infrared fluorescence intensity ratio between tumor and normal tissues (T=N ratio) as a function of probes injection. *$P < 0.05$ compared to the pHAN group; # $P < 0.05$ compared to the FA-pHAN group. pHAN, pH-activatable nanoprobes; FA-pHAN, FA-targeted pH-activatable nanoprobes.

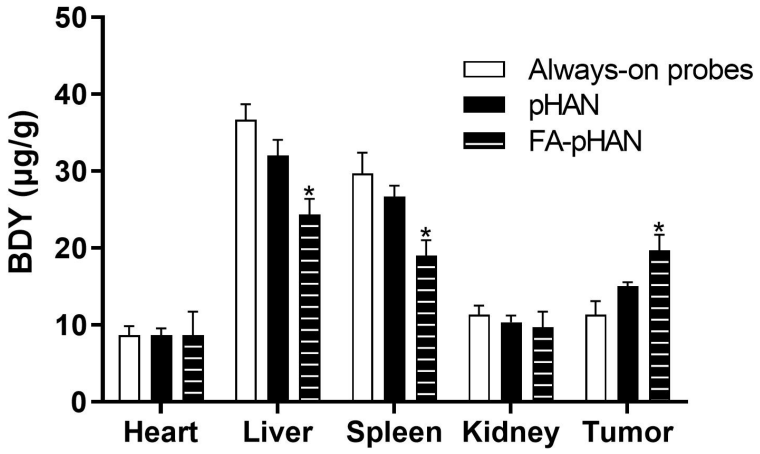


FIGURE 9
The biodistribution of mice at 24 h post-injection of always-on probes, pHAN and FA-pHAN. *$P < 0.05$ compared to the always-on probes group. pHAN, pH-activatable nanoprobes; FA-pHAN, FA-targeted pH-activatable nanoprobes.

characteristics of folic acid receptor expression, targeted delivery to tumor cells with overexpressed folic acid receptors can be achieved. This approach enables successful drug delivery to tumor cells, thus minimizing toxicity to normal cells and enhancing drug efficacy (28). However, further research is required to achieve targeted transport of nanoprobes to tumor cells and their specific activation within these cells. Therefore, our objective is to develop nanoprobes that possess both pH-responsiveness and tumor cell targeting capabilities.

The BDY exhibits strong visible light absorption, high fluorescence quantum yield, and excellent photostability (4). In this study, we employed BDY with aniline in the middle position as the substrate for pHAN. To verify the pH-dependent fluorescence characteristics of the probe, we measured the fluorescence intensity emitted by both pHAN and always-on probes at various pH values. Interestingly, only pHAN showed significant changes in fluorescence intensity upon altering the pH. Conversely, the fluorescence intensity of pHAN is significantly influenced by pH,

with no observable fluorescence at physiological pH; however, a notable increase in fluorescence occurs at lower pH levels. These findings suggest that pHAN can serve as a responsive imaging tool in acidic tumor environments. Additionally, we assessed the potential cytotoxicity of nanoprobes on cells using the MTT method and determined that any potential toxic effects from both pHAN and FA-pHAN were negligible *in vitro* cell experiments. To investigate the intracellular uptake ability of always-on probes and pHAN, CLSM was employed to detect the cellular uptake of probes in HeLa cells. It was observed that upon co-culturing with pHAN for 0 h, minimal green fluorescence was detected within the cells compared to the control sample. However, as the co-culture time extended to 4 h and particularly 8 h, a significant increase in green fluorescence intensity was observed in HeLa cells. Conversely, when co-cultured with always-on probes for 0–8 h, consistent levels of green fluorescence were detected within the cells. These findings suggest that pHAN may gradually accumulate within specific acidic organelles subsequent to internalization by tumor cells, resulting in its fluorescence being progressively stimulated by the acidic microenvironment. This hypothesis will be validated in forthcoming investigations. Lysosomes are intracellular compartments housing acid hydrolases responsible for waste degradation and cellular debris breakdown (29). Therefore, it is imperative to conduct further investigation into the lysosomal localization of pHAN. To facilitate observation, nuclear staining of HeLa cells was performed using DAPI, which emits blue fluorescence. Additionally, specific labeling of lysosomes was achieved using LysoTracker ® Red DND 99, emitting red fluorescence. Our findings demonstrate a gradual accumulation of pHAN within lysosomes upon cellular engulfment and its activation in response to acidic environments, resulting in an increasing emission of green fluorescence. Subsequently, the green fluorescence emitted by the probe and the red fluorescence originating from the lysosome overlap, resulting in a distinct yellow fluorescence signal. As anticipated, these prominent yellow spots predominantly localized within acidic lysosomal compartments at both 4 h and 8 h time points. This data strongly suggests that a significant proportion of pHAN molecules are internalized by tumor cells and subsequently sequestered within acidic lysosomes, thereby facilitating efficient emission of high-intensity fluorescence in response to acidity and ultimately achieving pH-responsive fluorescent behavior.

In the design of multifunctional nanoprobes, while the responsiveness to the tumor acidic environment is crucial, it is equally important to ensure their efficient accumulation at the tumor site. Insufficient accumulation may hinder effective discrimination between tumor and normal tissue, thereby compromising both sensitive imaging and targeted functionality. Hence, simultaneous achievement of tumor tissue-sensitive imaging and targeting functions remains a pivotal challenge that necessitates immediate attention. To accomplish this objective, we synthesized FA-pHAN based on the pHAN architecture. In the subsequent experimental investigation, CLSM was employed to observe the fluorescence activation level of FA-pHAN within tumor cells. The findings from this study demonstrate that following a 4 h incubation period with nanoprobes, the fluorescence signal emitted by FA-pHAN is significantly higher compared to non-targeted pHAN. However, early addition of free FA in the cell culture medium diminishes the cellular uptake of FA-pHAN, indicating that enhanced internalization of FA-pHAN by cells is attributed to the interaction between FA and tumor cells. The results suggest that FA-pHAN exhibits excellent tumor cell targeting ability and holds great potential as a promising candidate for tumor-targeted imaging. Our *in vivo* tumor imaging studies have further validated the precise delivery of FA-pHAN to the target site, with its fluorescence being effectively activated under acidic conditions within the tumor microenvironment, while no such changes were observed in normal tissues, thus enabling successful differentiation between healthy and diseased tissues. Furthermore, our biodistribution analysis demonstrated that FA-pHAN exhibits a targeted accumulation in tumor tissue compared to always-on probes or pHAN. This also leads to a reduction in probe accumulation in the liver and spleen, thereby minimizing potential side effects associated with *in vivo* administration and improving the targeted anti-cancer efficacy.

In summary, we have successfully engineered and synthesized pHAN with exceptional pH responsiveness, enabling significant activation within acidic organelles. This nanoplatform offers an efficient ON/OFF system for tumor imaging and drug delivery. Furthermore, to achieve tumor targeting functionality, we further developed FA-pHAN based on this platform. *In vitro* and *in vivo* studies unequivocally demonstrated that the fluorescence signal delivery efficiency of FA-pHAN surpassed that of non-targeted pHAN in both tumor cells and tissues. Our findings underscore the remarkable tumor cell targeting ability and specific imaging capability of FA-pHAN, positioning it as a promising candidate for real-time monitoring of tumor cells and tissues.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal study was approved by The Ethics of Animal Research Ethics Committee of Huzhou University. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

JL: Methodology, Investigation, Formal analysis, Writing – original draft. HH: Methodology, Investigation, Formal analysis, Writing – original draft. SL: Methodology, Investigation, Formal analysis, Writing – original draft. XL: Investigation, Formal

analysis, Conceptualization, Methodology, Resources, Writing – review & editing. FW: Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Key Projects of Huzhou Science and Technology Bureau in Zhejiang Province (Project No. 2021GZ52) and Zhejiang Province Public Welfare Technology Application Research Program Project (Project No. LGF21H060001) and Huzhou City Science and Technology Commissioner Project (2021KT07).

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.

References

1. Hesemans E, Buttiens K, Manshian BB, Soenen SJ. The role of optical imaging in translational nanomedicine. *J Funct Biomater*. (2022) 13:137. doi: 10.3390/jfb13030137

2. Cui FZ, Liu JH, Pang SW, Li B. Recent advance in tumor microenvironment-based stimuli-responsive nanoscale drug delivery and imaging platform. *Front Pharmacol*. (2022) 13:929854. doi: 10.3389/fphar.2022.929854

3. Refaat A, Yap ML, Pietersz G, Walsh APG, Zeller J, Del Rosal B, et al. *In vivo* fluorescence imaging: success in preclinical imaging paves the way for clinical applications. *J Nanobiotechnol*. (2022) 20:450. doi: 10.1186/s12951–022-01648–7

4. Rybczynski P, Smolarkiewicz-Wyczachowski A, Piskorz J, Bocian S, Ziegler-Borowska M, Kędziera D, et al. Photochemical properties and stability of BODIPY dyes. *Int J Mol Sci*. (2021) 22:6735. doi: 10.3390/ijms22136735

5. Antina E, Bumagina N, Marfin Y, Guseva G, Nikitina L, Sbytov D, et al. BODIPY conjugates as functional compounds for medical diagnostics and treatment. *Molecules*. (2022) 27:1396. doi: 10.3390/molecules27041396

6. Zan CF, An J, Wu ZF, Li SJ. Engineering molecular nanoprobes to target early atherosclerosis: Precise diagnostic tools and promising therapeutic carriers. *Nanotheranostics*. (2023) 7:327–44. doi: 10.7150/ntno.82654

7. Turner MA, Lwin TM, Amirfakhri S, Nishino H, Hoffman RM, Yazaki PJ, et al. The use of fluorescent anti-CEA antibodies to label, resect and treat cancers: A review. *Biomolecules*. (2021) 11:1819. doi: 10.3390/biom11121819

8. Xie RX, Wu ZJ, Zeng FX, Cai HW, Wang D, Gu L, et al. Retro-enantio isomer of angiopep-2 assists nanoprobes across the blood-brain barrier for targeted magnetic resonance/fluorescence imaging of glioblastoma. *Signal Transduct Target Ther*. (2021) 6:309. doi: 10.1038/s41392–021-00724-y

9. Vilímová L, Hervé-Aubert K, Chourpa L. Formation of miRNA nanoprobes-conjugation approaches leading to the functionalization. *Molecules*. (2022) 27:8428. doi: 10.3390/molecules27238428

10. Li W, Kaminski Schierle GS, Lei BF, Liu YL, Kaminski CF. Fluorescent nanoparticles for super-resolution imaging. *Chem Rev*. (2022) 122:12495–543. doi: 10.1021/acs.chemrev.2c00050

11. Prieto-Montero R, Katsumiti A, Cajaraville MP, López-Arbeloa I, Martínez-Martínez V. Functionalized fluorescent silica nanoparticles for bioimaging of cancer cells. *Sensors (Basel)*. (2020) 20:5590. doi: 10.3390/s20195590

12. Tian HL, Zhang TT, Qin SY, Huang Z, Zhou L, Shi JY, et al. Enhancing the therapeutic efficacy of nanoparticles for cancer treatment using versatile targeted strategies. *J Hematol Oncol*. (2022) 15:132. doi: 10.1186/s13045–022-01320–5

13. Sharifi M, Cho WC, Ansariesfahani A, Tarharoudi R, Malekisarvar H, Sari S, et al. An updated review on EPR-based solid tumor targeting nanocarriers for cancer treatment. *Cancers (Basel)*. (2022) 14:2868. doi: 10.3390/cancers14122868

14. Mahmoud K, Swidan S, El-Nabarawi M, Teaima M. Lipid based nanoparticles as a novel treatment modality for hepatocellular carcinoma: a comprehensive review on targeting and recent advances. *J Nanobiotechnol*. (2022) 20:109. doi: 10.1186/s12951–022-01309–9

15. Yao HJ, Sun L, Li JC, Zhou XF, Li R, Shao RG, et al. A novel therapeutic siRNA nanoparticle designed for dual-targeting CD44 and Gli1 of gastric cancer stem cells. *Int J Nanomed*. (2020) 15:7013–34. doi: 10.2147/IJN.S260163

16. Li WP, Little N, Park JH, Foster CA, Chen JW, Lu JQ. Tumor associated fibroblasts-targeting nanoparticles for enhancing solid tumor therapy: progresses and challenges. *Mol Pharm*. (2021) 18:2889–905. doi: 10.1021/acs.molpharmaceut.1c00455

17. Beh CY, Prajnamitra RP, Chen LL, Hsieh PCH. Advances in biomimetic nanoparticles for targeted cancer therapy and diagnosis. *Molecules*. (2021) 26:5052. doi: 10.3390/molecules26165052

18. Rahim MA, Jan N, Khan S, Shah H, Madni A, Khan A, et al. Recent advancements in stimuli responsive drug delivery platforms for active and passive cancer targeting. *Cancers (Basel)*. (2021) 13:670. doi: 10.3390/cancers13040670

19. Jin MZ, Jin WL. The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct Target Ther*. (2020) 5:166. doi: 10.1038/s41392–020-00280-x

20. Firouzabadi BM, Gigliobianco MR, Joseph JM, Censi R, Martino PD. Design of nanoparticles in cancer therapy based on tumor microenvironment properties. *Pharmaceutics*. (2022) 14:2708. doi: 10.3390/pharmaceutics14122708

21. Siriwibool S, Kaekratoke N, Chansaenpak K, Siwawannapong K, Panajapo P, Sagarik K, et al. Near-infrared fluorescent pH responsive probe for targeted photodynamic cancer therapy. *Sci Rep*. (2020) 10:1283. doi: 10.1038/s41598–020-58239–5

22. Farjadian F, Ghasemi S, Akbarian M, Hoseini-Ghahfarokhi M, Moghoofei M, Doroudian M. Physically stimulus-responsive nanoparticles for therapy and diagnosis. *Front Chem*. (2022) 10:952675. doi: 10.3389/fchem.2022.952675

23. Qiu KQ, Du Y, Liu JY, Guan JL, Chao H, Diao JJ. Super-resolution observation of lysosomal dynamics with fluorescent gold nanoparticles. *Theranostics*. (2020) 10:6072–81. doi: 10.7150/thno.42134

24. Wang LJ, Zhou Q, Yang HY. A facile fabrication of lysosome-targeting pH fluorescent nanosensor based on PEGylated polyester block copolymer. *Polymers (Basel)*. (2022) 14:2420. doi: 10.3390/polym14122420

25. Martín-Sabroso C, Torres-Suárez AI, Alonso-González M, Fernández-Carballido A, Fraguas-Sánchez AI. Active targeted nanoformulations via folate receptors: State of the Art and future perspectives. *Pharmaceutics*. (2022) 14:14. doi: 10.3390/pharmaceutics14010014

26. Jurczyk M, Jelonek K, Musiał-Kulik M, Beberok A, Wrześniok D, Kasperczyk J. Single- versus dual-targeted nanoparticles with folic acid and biotin for anticancer drug delivery. *Pharmaceutics*. (2021) 13:326. doi: 10.3390/pharmaceutics13030326

27. Hu YW, Chen DY, Napoleon JV, Srinivasarao M, Singhal S, Savran CA, et al. Efficient capture of circulating tumor cells with low molecular weight folate receptor-specific ligands. *Sci Rep*. (2022) 12:8555. doi: 10.1038/s41598–022-12118–3

28. Mahalunkar S, Yadav AS, Gorain M, Pawar V, Braathen R, Weiss S, et al. Functional design of pH-responsive folate-targeted polymer-coated gold nanoparticles for drug delivery and *in vivo* therapy in breast cancer. *Int J Nanomedicine*. (2019) 14:8285–302. doi: 10.2147/IJN.S215142

29. Webb BA, Aloisio FM, Charafeddine RA, Cook J, Wittmann T, Barber DL. pHLARE: A new biosensor reveals decreased lysosome pH in cancer cells. *Mol Biol Cell*. (2021) 32:131–42. doi: 10.1091/mbc.E20–06-0383

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 14 June 2024
DOI 10.3389/fonc.2024.1394493

Check for updates

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
He N. Xu,
University of Pennsylvania, United States
Francesk Mulita,
General University Hospital of Patras, Greece

*CORRESPONDENCE
Radu Hristu
radu.hristu@upb.ro
George A. Stanciu
stanciu@physics.pub.ro

†These authors have contributed equally to this work

RECEIVED 01 March 2024
ACCEPTED 20 May 2024
PUBLISHED 14 June 2024

CITATION
Florea M-A, Eftimie LG, Glogojeanu RR, Hristu R, Stanciu GA and Costache M (2024) Imaging of colorectal adenomas with pseudoinvasion and malignant polyps using two-photon excitation microscopy. *Front. Oncol.* 14:1394493. doi: 10.3389/fonc.2024.1394493

COPYRIGHT
© 2024 Florea, Eftimie, Glogojeanu, Hristu, Stanciu and Costache. 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.

Imaging of colorectal adenomas with pseudoinvasion and malignant polyps using two-photon excitation microscopy

Maria-Alexandra Florea[1,2†], Lucian George Eftimie[1,3,4†], Remus Relu Glogojeanu[4†], Radu Hristu[3*], George A. Stanciu[3*] and Mariana Costache[2]

[1]Pathology Department, Central University Emergency Military Hospital, Bucharest, Romania, [2]Pathology Department, University of Medicine and Pharmacy, Carol Davila', Bucharest, Romania, [3]Center for Microscopy-Microanalysis and Information Processing, National University of Science and Technology Politehnica Bucharest, Bucharest, Romania, [4]Department of Special Motricity and Medical Recovery, The National University of Physical Education and Sports, Bucharest, Romania

Introduction: Although the incidence and mortality rates of colorectal cancer exhibit significant variability, it remains one of the most prevalent cancers worldwide. Endeavors to prevent colorectal cancer development focus on detecting precursor lesions during colonoscopy. The diagnosis of endoscopically resected polyps relies on hematoxylin and eosin staining examination. For challenging cases like adenomatous polyps with epithelial misplacement, additional diagnostic methods could prove beneficial.

Methods: This paper aims to underscore stromal changes observed in malignant polyps and polyps with pseudoinvasion, leveraging two-photon excitation microscopy (TPEM), a technique extensively employed in the medical field in recent years.

Results and discussions: Both the subjective and quantitative analysis of TPEM images revealed distinct distributions and densities of collagen at the invasion front in malignant polyps compared to areas of pseudoinvasion. TPEM holds potential in discerning true invasion in malignant polyps from pseudoinvasion, offering enhanced visualization of local stromal changes.

KEYWORDS

pseudoinvasion, malignant polyps, multiphoton microscopy, second harmonic generation, collagen

1 Introduction

Colorectal cancer stands as the most prevalent malignancy affecting the gastrointestinal tract and ranks as the second leading cause of cancer-related mortality worldwide (1). Its incidence has been on the rise across both well-developed and medium to low-income countries in recent years, highlighting its significance as a global public health concern (2). Early detection of colorectal tumors is paramount in mitigating future morbidity and mortality rates.

Colorectal cancer presents a heterogeneous nature (3), both molecularly and morphologically. Most colorectal adenocarcinomas originate from precursor lesions, such as conventional adenomas or serrated polyps, which are typically marked by dysplasia (4). Colonoscopy emerges as the most effective method for detecting colonic polyps and cancers (5), aiming to identify and remove polyps with the potential to progress to invasive cancers. Advances in endoscopic techniques have significantly enhanced visualization of intestinal lesions and facilitated the removal of more complex lesions (6).

As per the WHO classification of digestive system tumors (2019), carcinoma *in situ* (Tis) denotes cancer cell invasion within the lamina propria, while T1 indicates invasion through the muscularis mucosae into the submucosa (7). Malignant polyps are characterized by tumoral cell invasion through the muscularis mucosa into the submucosa but not beyond it (8). The prevalence of malignant polyps among endoscopically removed polyps in screening programs can be as high as 11% (9). Features such as tubulovillous or villous architecture, size exceeding 10 mm, and high-grade dysplasia signify advanced adenomas and elevate the risk of malignant transformation (7).

Pseudoinvasion, also referred to as epithelial misplacement, is commonly observed in large pedunculated polyps, particularly in the sigmoid colon, resembling invasive carcinoma (10). It is characterized by the presence of dysplastic glands beneath the muscularis mucosa in the polyp head, stalk, or deeper regions. Unlike true infiltration, pseudoinvasion involves displaced adenomatous glands protruding through weakened areas of the muscularis mucosa, likely due to factors such as repeated stalk twisting, ischemia, or prior biopsy (11-12). Polyps exhibiting pseudoinvasion should be managed as ordinary adenomas, given their lack of malignant features.

The diagnosis of pseudoinvasion primarily relies on morphological examination, performed by an experienced pathologist. Deeper-level sections through the paraffin block and seeking a second opinion from a pathologist with expertise in digestive pathology are recommended for equivocal cases (13). Various methods, including immunohistochemistry - both epithelial (p53, E-chaderin) and stromal markers (collagen IV, MMP-1) and three-dimensional reconstruction or infrared spectroscopic techniques (12, 14), have been explored for the differential diagnosis of these entities. However, their utility in distinguishing challenging cases remains limited (15).

To highlight the local stromal alterations in both adenomas with epithelial misplacement and malignant polyps, we propose employing a nonlinear optical imaging technique. Specifically, we demonstrate the efficacy of two-photon excitation microscopy (TPEM) (16–19), a method that has gained prominence in clinical microscopy (20) in recent years and has been applied in endoscopic procedures (21, 22). TPEM simultaneously utilizes two nonlinear optical contrast mechanisms—second harmonic generation (SHG) and two-photon excitation fluorescence (TPEF)—to acquire images. SHG involves coherent optical second-order nonlinear effects (23), generating a new photon with double the energy of the initial photons upon interaction with non-centrosymmetric structures like collagen (24–28). TPEF microscopy (16), on the other hand, involves the simultaneous absorption of two photons with a total energy enough to produce a transition to an excited state. The subsequent spontaneous emission generates a photon with a slightly smaller energy than the total energy of the two excitation photons. It provides complementary information on cell and tissue morphology (29).

In this study, we employ both TPEF and SHG microscopy techniques to capture images from benign and malignant regions. Our aim is to compare collagen distribution patterns in normal tissue, benign polyps with pseudoinvasion, and malignant polyps at the invasive front.

2 Materials and methods

2.1 Sample preparation

The dataset utilized for this study comprises 14 polyps collected from 10 patients via polypectomy procedures conducted during colonoscopies and colectomies at the Central University Emergency Military Hospital (Bucharest, Romania). Nine polyps were identified as malignant, out of which five exhibited also pseudoinvasion. Additionally, there were four cases of adenomatous polyps showing pseudoinvasion, and one case of intramucosal adenocarcinoma. The samples were diagnosed by the Pathology Department and formalin-fixed paraffin-embedded (FFPE) blocks were stored in the histology laboratory at the Central University Emergency Military Hospital. Written informed consent was obtained from the patients and all samples were anonymized before analysis. All experiments were performed according to the relevant guidelines and regulations and in accordance with the Declaration of Helsinki. Samples obtained by cutting the FFPE blocks were sectioned (4–5 µm thick) and stained with hematoxylin and eosin (H&E). All samples were processed together in a single batch to reduce the impact of variations in histological stain recipes or procedures (30) that could lead to color discrepancies.

2.2 Microscopy techniques

All slides were scanned using a bright-field Aperio LV1 IVD Whole Slide Scanner (Leica Biosystems) equipped with a 20X objective lens. Expert pathologists identified regions of invasion and pseudoinvasion in malignant polyps, and areas of pseudoinvasion in adenomatous polyps. Following their

annotations, these regions were further imaged by TPEM and compared qualitatively and quantitatively to identify collagen signatures that distinguish between the two types. Selected regions of interest (ROIs) were examined using a two-channel Leica TCS SP laser scanning confocal microscope configured for nonlinear imaging. The detailed description of the TPEM imaging system can be found elsewhere (31). The excitation source was a Ti:Sapphire laser (Coherent Chameleon Ultra II) tuned to a wavelength of 860 nm, with a pulse width of 140 fs and a repetition rate of 80 MHz. Laser power levels below 15 mW, as measured at the objective focus, were employed during scanning. SHG and TPEF signals were simultaneously collected in epi detection mode. A 10x magnification objective with a numerical aperture of 0.3 was employed to focus the excitation laser beam onto the samples and to collect TPEF and backward-generated SHG (BSHG) signals. The spectrally resolved detection setup inherent to the Leica TCS SP was utilized for capturing BSHG (425 to 435 nm) and TPEF (450 to 700 nm) signals on separate channels. Composite TPEM images were generated, with BSHG represented in the green channel and TPEF in the magenta channel. As the microscope objective's maximum field-of-view is 1 x 1 mm^2, image tiles were acquired and stitched together to create a mosaic covering a larger tissue area.

It is important to note that in H&E-stained tissue sections, the TPEF signal originates from regions containing eosin. Eosin binds nonspecifically to proteins within the cytoplasm, cell membrane borders, red blood cells, and extracellular structures. Conversely, the BSHG signal (Figure 1) originates from collagen fibers within the tissue, irrespective of the presence of stain.

2.3 Automated image analysis

Using the histogram analysis toolbox in ImageJ, four parameters associated with the distribution of pixel intensities in SHG images – Mean, Standard Deviation, Skewness, and Kurtosis – were calculated. While mean and standard deviation are well-recognized statistical moments, skewness and kurtosis characterize the distribution's shape. To discern between collagen structures with distinct distributions, one additional parameter was considered. Pixels with SHG intensity surpassing a threshold were tallied, yielding the total collagen area ratio (TC-ratio), expressed as the ratio of this count to the total image area. Image thresholding was carried out using the automated thresholding toolbox in ImageJ, which encompasses 16 distinct thresholding methods. We visually examined how these algorithms worked on a set of TPEM images and found that the Triangle thresholding method provided the most effective results.

A second set of parameters, derived from the Gray-Level Co-Occurrence Matrix (GLCM), provides information about the spatial relationships between pixel intensities in an image. ImageJ's GLCM texture plugin was used to compute the matrix for adjacent pixels in horizontal, vertical, and two diagonal directions. Information about collagen organization in SHG images can be extracted from the GLCM using the following parameters: Contrast, Homogeneity, Energy, Entropy and Correlation. Contrast and Homogeneity measure the local variations in an image, while Energy and Entropy can be used to give a quantitative measurement of the mutual orientation of collagen fibers (32). Correlation is based on the statistical analysis of pixel value dependence and can be used to assess periodicity within an image. Here, we employ a distance of one pixel between neighboring pixels, and we calculate the average values obtained from the horizontal, vertical, and two diagonal directions for each of the five parameters derived from the GLCM.

SHG images were utilized to assess collagen fiber orientation through FFT analysis with a custom-written ImageJ script. FFT power spectra images generated for each acquired SHG image (insets in Figure 1), were initially binarized and then fitted with an ellipse. A collagen orientation index (N) was computed based on the lengths of the minor (S) and major (L) axes of the ellipse: N = 1 – S/L (33). Consequently, collagen orientation can be expressed by an index ranging from 0 (indicating random fibers) to 1 (indicating aligned fibers).

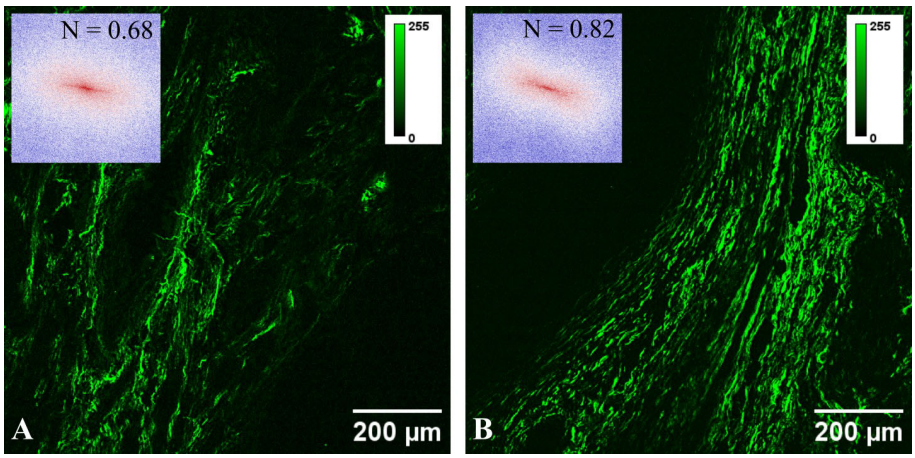


FIGURE 1
Typical SHG images for **(A)** pseudoinvasion and **(B)** invasion areas. The inset in each panel represents the FFT power spectrum and the corresponding collagen orientation index.

An alternative approach to objectively quantify organization in an image involves fractal analysis. The SHG images underwent analysis after thresholding with the Triangle method, hence in a binary format, utilizing an ImageJ plugin (i.e., Fractal box count) to compute the fractal dimension.

For the statistical analysis, we employed the two-way unpaired Student's t-test using Prism 10 (GraphPad Software, USA). Normality was evaluated using the D'Agostino and Pearson test. A p-value of 0.05 was used as the threshold for determining statistical significance in all tests.

3 Results

We examined a total of 14 colonic polyps, capturing whole slide images for each specimen. Additionally, corresponding TPEM images were acquired and analyzed for selected ROIs. Due to space constraints, only the most representative images are included in the manuscript. However, all images captured for this experiment, as well as high-resolution versions of the images presented here, are available in a public repository (DOI 10.17605/OSF.IO/JDRBN).

Figure 2 displays typical images for an adenomatous polyp with low-grade dysplasia and epithelial misplacement. In Figure 2A, the lower half of the polyp, including the pedicle and implantation base, is depicted. Figure 2B illustrates a lateral margin of the pedicle, showcasing non-dysplastic glands and the underlying connective stroma, albeit with some artifacts induced by electro resection. Histologic examination utilizing overlaying BSHG and TPEF images reveals collagen fibers highlighted in the green channel, forming short and thin bands of various orientations. The magenta channel highlights non-dysplastic colonic glands with honeycomb arrangement and fibro-connective tissue from the base of the polyp (Figure 2C).

A pseudoinvasive area is identified in this polyp (Figures 2D, F), characterized by groups of dysplastic glands surrounded by lamina propria within the connective stroma of the polyp pedicle. Microhemorrhages and deposits of hemosiderin are observed near dysplastic glands. Surrounding misplaced glands, there is a diminished collagenous reaction, with short collagen fibers which lack consistent alignment, interspersed with smooth muscle cells (TPEM, Figures 2E, G).

Similar observations are made in Figure 3, where a pedunculated adenomatous polyp with villous architecture, low-grade dysplasia, and an area of epithelial misplacement beneath the

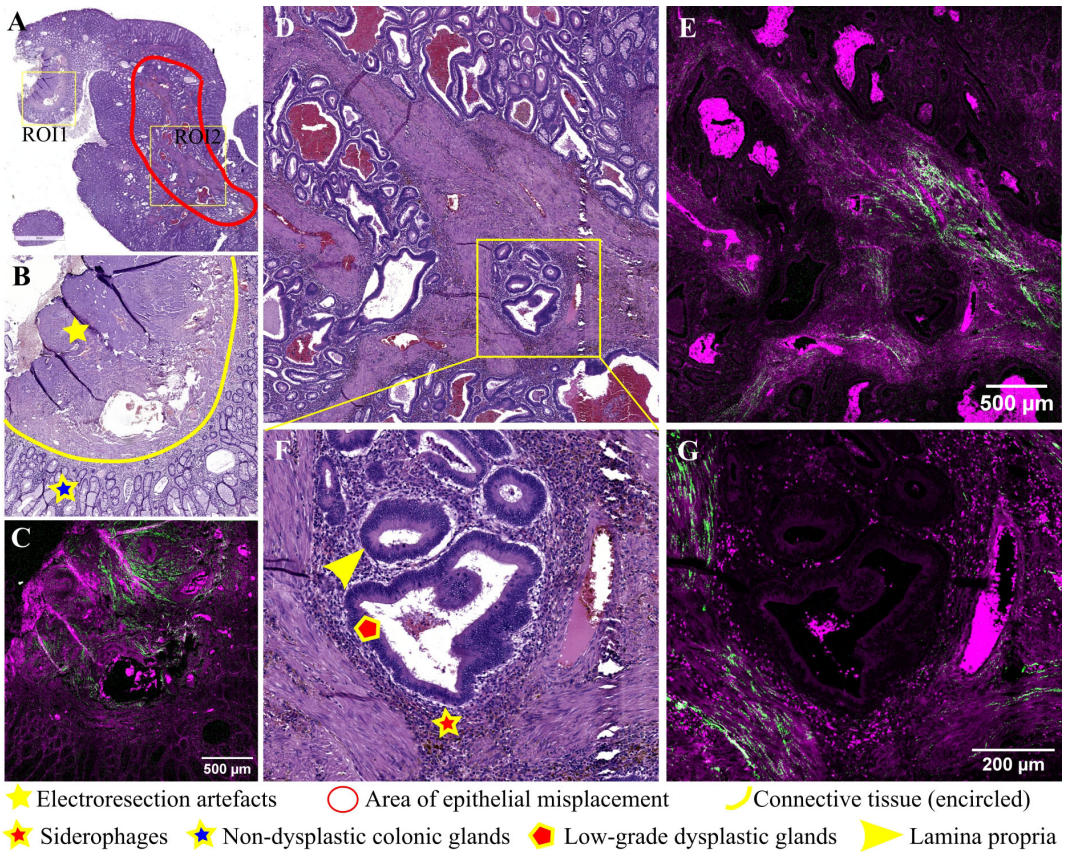


FIGURE 2
Adenomatous colonic polyp with pseudoinvasion area: **(A)** Large scale bright-field scanned image with rectangular ROIs selected for TPEM imaging; **(B)** Bright-field microscopy image of H&E-stained ROI1; **(C)** Corresponding TPEM image for ROI1; **(D)** Bright-field microscopy image of H&E-stained ROI2 corresponding to a pseudoinvasion area (low magnification); **(E)** Corresponding TPEM image for ROI2; **(F)** Bright-field microscopy image for a pseudoinvasion area – high magnification; **(G)** Corresponding TPEM image.

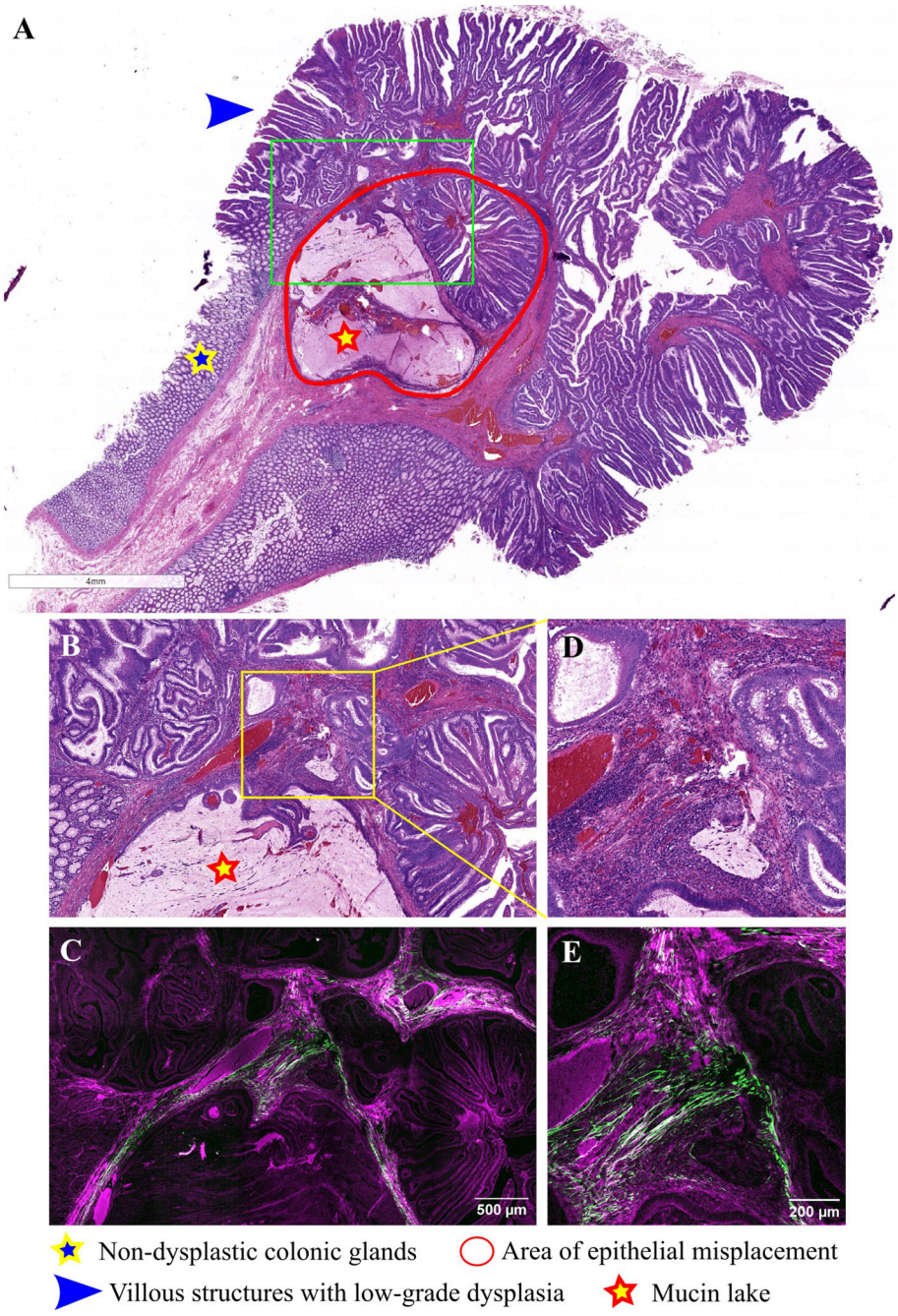


FIGURE 3
Pedunculated adenomatous colonic polyp with epithelial misplacement: **(A)** Large scale bright-field scanned image with a rectangular ROI selected for TPEM imaging; **(B)** Bright-field microscopy image of H&E-stained ROI corresponding to the area of epithelial misplacement (low magnification); **(C)** Corresponding TPEM image for the ROI; **(D)** Bright-field microscopy image of the area of epithelial misplacement (high magnification); **(E)** Corresponding TPEM image.

muscularis mucosae is depicted. It is composed of a rounded group of villous structures, along with two mucin lakes, with dysplastic epithelium at the periphery of the mucin lake, but without any floating tumoral cells and lacking a desmoplastic stroma (Figures 3A, B, D). TPEM images (Figures 3C, E) via the BSHG channel reveal a comparable distribution of collagen fibers as seen in the previous case.

Representative images for another adenomatous polyp, displaying irregular, atypical glands infiltrating the submucosa and inducing a desmoplastic reaction, are presented in Figure 4. In this case, the diagnosis was adenocarcinoma arising in a tubulovillous adenoma, invasive in the submucosa. TPEM images (Figures 4C, E) show a more prominent fibrous reaction in the submucosa, with collagen fibers tending to accumulate more at the

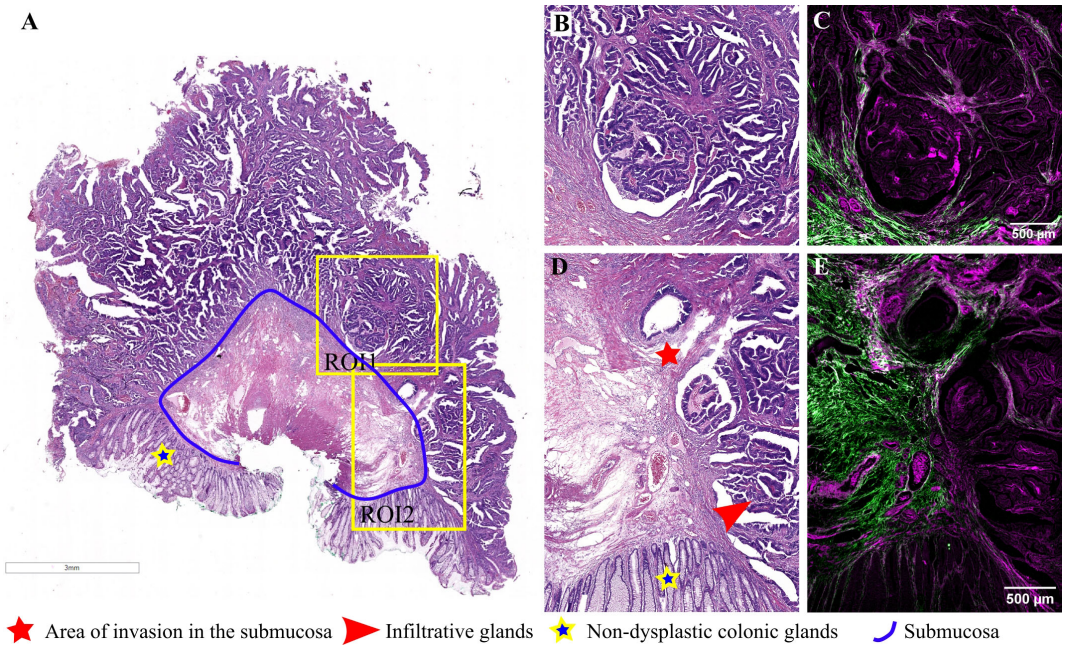


FIGURE 4
Adenocarcinoma invading the submucosa developed on a colonic adenomatous polyp: **(A)** Large scale bright-field scanned image with rectangular ROIs selected for TPEM imaging. **(B)** Bright-field microscopy image of H&E-stained tissue for ROI1; **(C)** Corresponding TPEM image for ROI1; **(D)** Bright-field microscopy image of H&E-stained tissue for ROI2; **(E)** Corresponding TPEM image for ROI2.

invasion front and less between invasive glands. At higher magnification, irregular and atypical glands infiltrating the submucosa, accompanied by a desmoplastic reaction, are depicted (Figures 4B, D). Collagen fibers at the invasive front appear organized in longer fascicles, with the collagen predominantly localized at the periphery of the tumor (TPEM, Figures 4C, E).

In the current study, another noteworthy polyp examined is a villous adenoma with an area of invasive adenocarcinoma, as evidenced by irregular, back-to-back glands infiltrating through the muscularis mucosae into the submucosa (Figures 5A, B). TPEM images reveal a prominent collagenous reaction surrounding malignant glands, characterized by longer fascicles of collagen

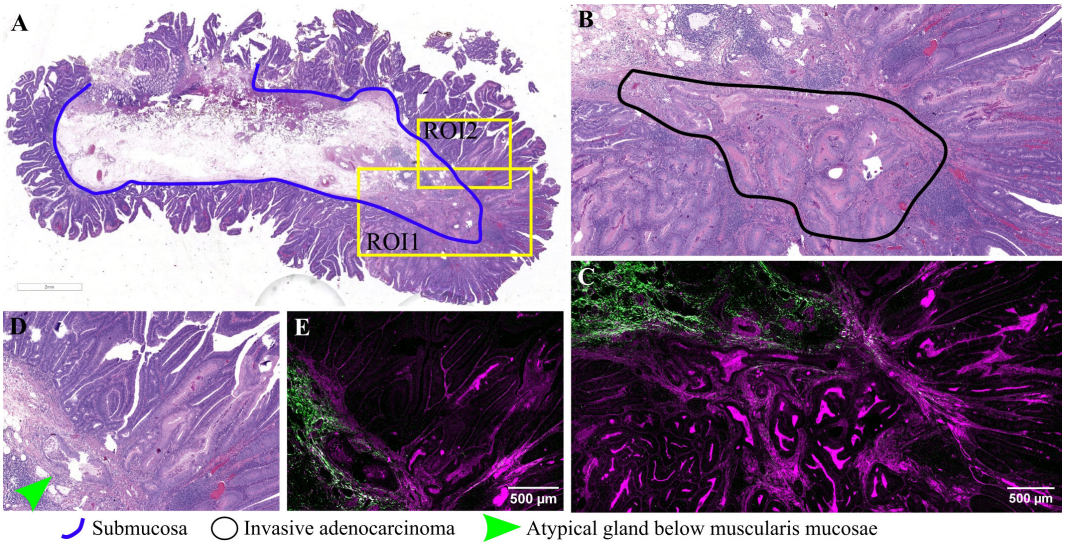


FIGURE 5
Adenocarcinoma arising in an adenomatous polyp: **(A)** Large scale bright-field scanned image with rectangular ROIs selected for TPEM imaging; **(B)** Bright-field microscopy image of H&E-stained tissue for ROI1; **(C)** TPEM image for ROI1; **(D)** Bright-field microscopy image of H&E-stained tissue for ROI2; **(E)** TPEM image for ROI2.

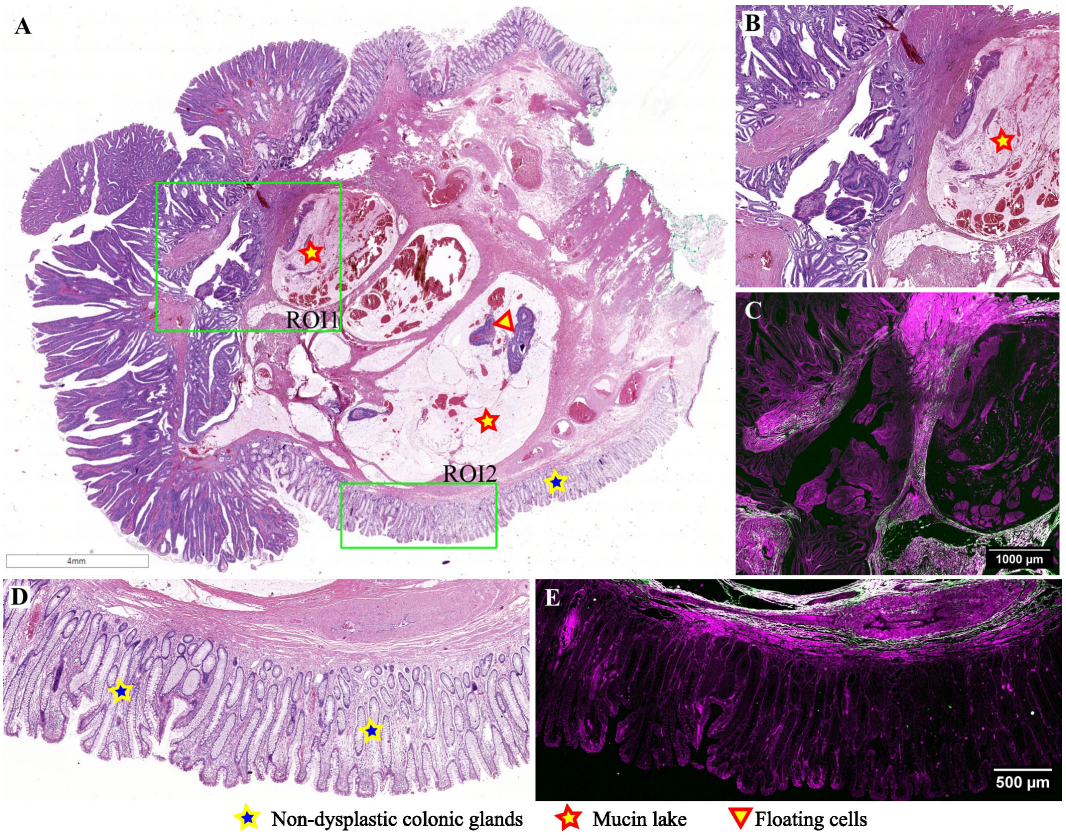


FIGURE 6
Adenomatous polyp with an area of invasive mucinous adenocarcinoma: **(A)** Large scale bright-field scanned image with rectangular ROIs selected for TPEM imaging; **(B)** Bright-field microscopy image of H&E-stained tissue for ROI1; **(C)** TPEM image for ROI1; **(D)** Bright-field microscopy image of H&E-stained tissue for ROI2, corresponding to healthy colonic tissue; **(E)** TPEM image for ROI2.

fibers exhibiting a tendency to align parallel to the long axis of the tumor (Figures 5C, E).

Figure 6 presents an intriguing case of a pedunculated adenomatous polyp, wherein displaced atypical glands infiltrate through the muscularis mucosae, alongside mucin pools containing floating tumoral cells (Figures 6A, B). This case is diagnosed as mucinous adenocarcinoma arising in an adenomatous polyp because an area of invasive carcinoma was identified on another serial section from this polyp (not depicted in this image). An image selected from the lateral margins of this polyp depicts non-dysplastic colonic mucosa, muscularis mucosae, and a portion of a mucinous pool in the submucosa (Figure 6D). TPEM images illustrate the presence of a few short collagen fibers in the submucosa, below the muscularis mucosae in non-tumoral tissue (Figure 6E), along with a greater abundance of collagen fibers surrounding mucin lakes, arranged in longer fascicles.

As collagen features exhibited noticeable differences in the invasion and pseudoinvasion areas, we sought to assess whether quantitative parameters derived from TPEM images could differentiate between these regions. Since the visual examination of TPEM images returned differences in malignant polyps and polyps with pseudoinvasion which related to the collagen distribution outlined in the SHG images, the quantitative analysis was further performed only on the SHG channel in the TPEM images. For a comprehensive quantitative image analysis, we chose 88 images from invasion areas and 65 images from pseudoinvasion areas with significant collagen content from the acquired 1 x 1 mm^2 SHG image tiles. These image sets were then used to examine the distribution and organization of collagen in the SHG images using texture analysis methods detailed in the Methods section.

The results in Figure 7 indicate that, except for one parameter (i.e., fractal dimension), all the computed parameters detected statistically significant differences between SHG images acquired on malignant polyps and polyps with pseudoinvasion.

4 Discussion

Differentiating between epithelial misplacement and true invasion in an adenomatous polyp presents a significant challenge, as it helps prevent the overdiagnosis of the lesion and unnecessary surgical interventions, thus impacting the patient's quality of life significantly.

Histopathologically, malignant glands often exhibit irregular outlines, complex architecture (such as back-to-back glands, cribriform structures, or solid nests of atypical cells), intraluminal necrosis (dirty necrosis), and high-grade cytologic atypia—characterized by enlarged nuclei, hyperchromatism, prominent

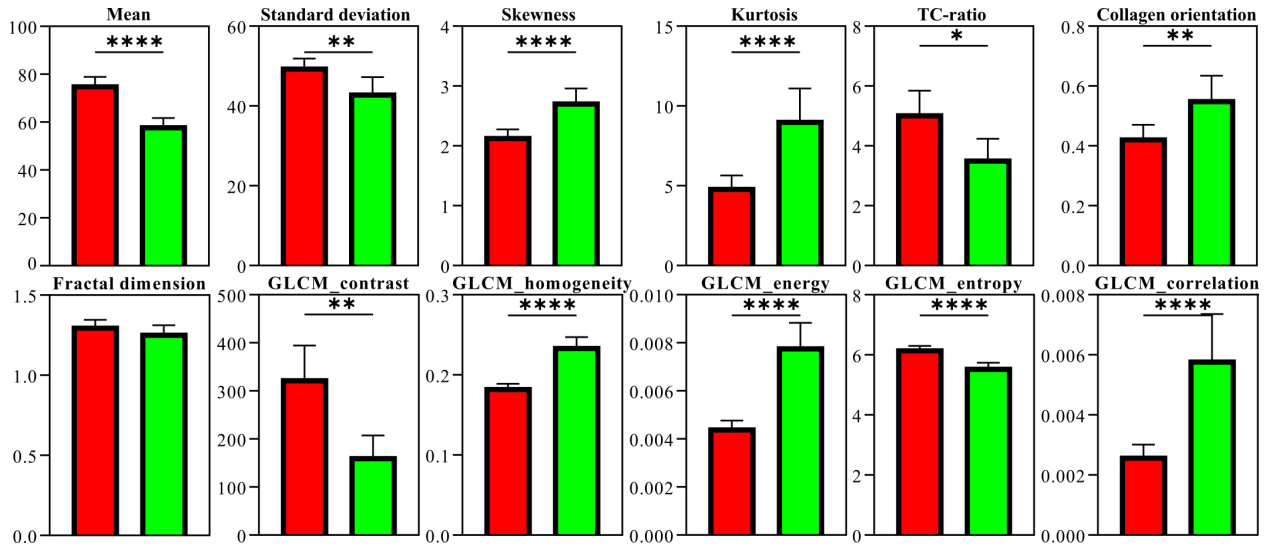


FIGURE 7
Parameters computed from SHG image sets for malignant polyps (red) and polyps with pseudoinvasion (green). The number of measurements was considered the number of image tiles for which the parameters were computed, hence 88 for invasion areas and 65 for pseudoinvasion areas. The vertical error bars represent the 95% confidence intervals (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$).

nucleoli, and atypical mitosis. The presence of desmoplasia, a stromal reaction surrounding cancerous cells, is a critical feature that, in the context of dysplasia, supports the diagnosis of invasive carcinoma. These histological changes are evident in our H&E-stained tissue samples from the polyps depicted in Figures 4A, B, D; Figures 5A, B, D; and Figures 6A, B, D.

We utilized a nonlinear optical method—two-photon microscopy—to elucidate local stromal changes in both malignant adenomas and adenomas with pseudoinvasion. The results are depicted as two-channel TPEM images, with SHG represented on the green channel and TPEF on the magenta channel. Relevant findings are derived from analyzing the SHG channel, which accentuates the collagen architecture in the examined samples. The samples imaged under the TPEM microscope were prepared for bright-field microscopy, hence the TPEF images contain nonspecific information collected on eosin fluorescence.

The desmoplastic reaction entails alterations in the extracellular matrix, a component vital for cellular communication, adhesion, and proliferation (34). This reaction was observed in three out of the five cases examined via TPEM (Figures 4, 5, and 6C, E), manifesting as an accumulation of collagen surrounding infiltrative glands. Collagen possesses properties such as fiber density, distribution, orientation, and organization, which serve as crucial markers in cancer research.

Conversely, pseudoinvasion is characterized by several histopathological features, including rounded displaced glands lacking infiltrative contours, surrounded by normal stroma of the lamina propria, alongside inflammatory cells and fragmented smooth muscle bundles exhibiting similar morphological changes to those observed in the surface epithelium (10). These specific pseudoinvasion changes were evident in the polyp depicted in Figures 2D, F, as well as Figures 3B, D, on H&E-stained sections. Notably, there is an absence of desmoplastic reaction around displaced glands, with inflammation, hemorrhage, and hemosiderin deposits observed in the stalk surrounding the glands. TPEM images highlighted a reduced presence of collagen fibers around pseudoinvasion areas, indicating a fibro-muscular reaction around prolapsed glands rather than true desmoplasia.

Displaced glands within the polyp may undergo dilation and accumulate mucus, unable to reach the lumen due to the entrapment of surrounding glands. Subsequently, these glands may rupture into the stalk, resulting in the formation of mucinous pools, which typically exhibit a rounded outline. In such cases, the epithelium is situated at the periphery of the mucinous pool rather than floating within it, as observed in invasive cancers (35). These distinctive features are discernible in the polyp depicted in Figure 6. Distinguishing invasion from epithelial misplacement in an adenomatous polyp with high-grade dysplasia poses a greater challenge (7). Isolated glands lacking surrounding lamina propria, poor differentiation, and evidence of vascular invasion lean towards a diagnosis of adenocarcinoma.

Merely assessing an image obtained from an H&E-stained section makes it difficult to differentiate the desmoplastic stroma associated with submucosal invasion in malignant polyps from the fibro-muscular reaction present in polyps with pseudoinvasion. In both scenarios, collagen fibers tend to exhibit an eosinophilic compact appearance. However, our findings suggest that TPEM images, particularly the SHG images, offer superior visualization of collagen fibers compared to conventional microscopy. Analysis of collagen in the TPEM images unveils a distinct collagen architecture in the invasive area compared to regions of pseudoinvasion (Figure 8). Notably, our cases revealed more prominent collagen deposits in malignant polyps, particularly at the invasion front,

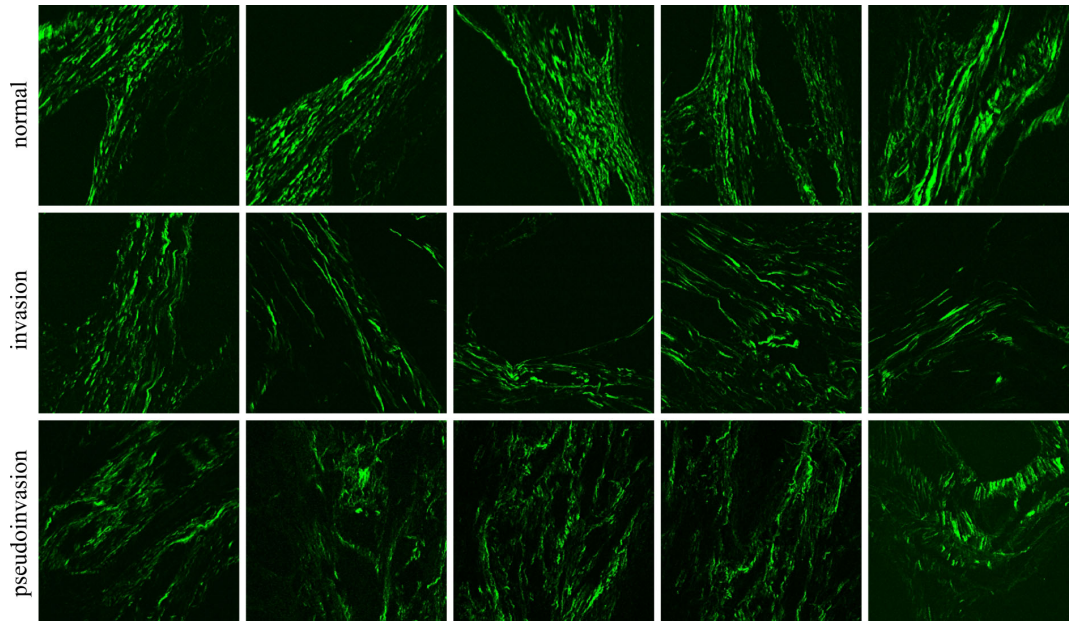


FIGURE 8
SHG images of normal, invasion and pseudoinvasion areas. Each image is 500 x 500 μm^2.

exhibiting a unique organizational pattern compared to normal tissue from the base of an adenomatous polyp or areas of epithelial misplacement (pseudoinvasion). Such architectural changes imply a potential role of this reaction in impeding tumor extension.

The quantitative findings (Figure 7) can be partially associated with the qualitative observations from the TPEM images (Figures 2–6). The lower mean pixel intensities observed in polyps with pseudoinvasion, showing statistically significant differences compared to malignant polyps, suggest a reduced SHG signal. This reduction could stem from a decrease in collagen density, a decrease in the SH generation of individual collagen fibers, or a combination of both factors. To address this question, we calculated another parameter, namely TC-ratio, which assesses the total collagen area in an SHG image and specifically considers changes in collagen density. Notably, a statistically significant decrease in TC-ratio was also observed for polyps with pseudoinvasion. This outcome aligns with the notion of decreased collagen content in cases of pseudoinvasion.

The Standard Deviation of pixel intensities serves to measure the extent of dispersion within a dataset. Typically, a reduced standard deviation in the distribution of pixel intensities may indicate lower contrast, implying a more consistent image. In the case of polyps with pseudoinvasion, there is a statistically significant decrease in standard deviation compared to malignant polyps. This finding correlates with the diminished collagen content in polyps exhibiting pseudoinvasion.

As for Skewness, both distributions exhibit a positive skew, characterized by longer right tails. A higher skewness in pixel distributions within an image typically implies lower pixel values, thus resulting in a low-intensity (dark) SHG image (36). In our context, this may suggest either a reduced intensity of SHG signals from collagen or a decreased collagen density in the imaged region. The elevated Skewness value observed in polyps with pseudoinvasion aligns with the observation of lower collagen content.

Both distributions exhibit elevated positive Kurtosis. A statistically significant rise in Kurtosis has been noted in polyps with pseudoinvasion, signifying that pixel values are distributed in closer proximity to the mean compared to a normal distribution, with minimal variance attributed to infrequent extreme deviations. The lower Standard Deviation, associated with higher Kurtosis, indicates a narrower distribution of pixel values around the mean and suggests sharper features (collagen fibers) in the image.

While the previously discussed first-order statistical parameters, computed through the image histogram, are directly associated with the gray level distribution of pixel intensities, second-order statistical parameters for images are contingent on the spatial arrangements and correlation of pixel intensities.

Contrast is linked to the disparity between adjacent GLCM elements, measuring the uneven distribution of values within the GLCM. Elevated GLCM contrast signifies highly contrasting images, often associated with features distributed without a preferred alignment (37). A statistically significant increase in contrast was observed for malignant polyps compared to polyps with pseudoinvasion, possibly attributed to the presence of aligned collagen fibers in the former scenario. Homogeneity and contrast typically exhibit an inverse correlation, a trend that is consistent with our findings as well.

Energy identifies irregularities in textures, with a maximum value of one when the gray level distribution is consistent and uniform. Lower energy values arise when there is an increase in small entries in the GLCM, which may point to heterogeneous images. In our case, the results regarding energy and homogeneity suggest that the GLCM values exhibit a lower level of uniformity.

Similar to Contrast and Homogeneity, Energy and Entropy also show an inverse correlation. A higher entropy value suggests a more complex texture in the image.

The findings from the GLCM analysis should be viewed in the context of the broader perception of an image, which extends beyond pixel intensity values to include factors such as context and surrounding patterns. This broader perspective can lead to varied visual interpretations, even when statistical measures suggest homogeneity. Although an image might exhibit high overall homogeneity, local variations or structures within it can make it visually compelling. For example, high homogeneity might stem from consistency across much of the image, while other areas may still display distinctive textures or patterns. Despite potential discrepancies between GLCM results and straightforward visual interpretations, these findings indicate changes in the collagen architecture between areas of true invasion and pseudoinvasion.

Correlation serves as an indicator of linear dependencies among gray levels in an image. Both malignant polyps and those with pseudoinvasions exhibited low Correlation values. A low Correlation typically signifies independent adjacent gray levels, indicating the absence of a significant regular pattern in the image.

Fractal analysis was conducted on binary images derived from image thresholding. Despite the prevalence of binary fractal analysis, we faced here one of its drawbacks, such as the requirement for image binarization through thresholding. In our case, no statistically significant differences were observed in the fractal dimension between malignant polyps and polyps with pseudoinvasions.

The collagen orientation index determined through FFT analysis exhibits a statistically significant elevated value for polyps with pseudoinvasion, indicating a more random distribution of collagen fibers compared to malignant polyps.

Previously, TPEM microscopy has been employed to characterize collagen alterations in various epithelial tumors, including breast, ovarian, gastric, colorectal, pancreatic, lung, bladder, thyroid, and skin cancer. Studies have revealed that as tumors progress, they tend to exhibit a higher quantity, lower organization, and increased linearity of collagen fibers (38). These findings align closely with the observations made in the two cases of malignant polyps described herein. However, in the context of colorectal pathology, TPEM has not been utilized for evaluating pseudoinvasion.

The quantitative imaging approach adopted in this study could pave the way for incorporating machine learning (ML) and deep learning (DL) techniques into colorectal cancer diagnosis (39) with significant implications for the advancement of TPEM technology. Over the past 20 years, TPEM has undergone significant development and has evolved to encompass diverse versions applicable in clinical settings (40). Despite its current advantages, TPEM has not achieved widespread adoption, partly due to challenges faced by end users, namely pathologists and surgeons, in interpreting the data. Consequently, they require retraining to effectively analyze TPEM images. ML and DL methodologies can automate the interpretation of TPEM images, assisting pathologists and surgeons in making more accurate and consistent diagnoses. ML and DL algorithms trained on large datasets of annotated TPEM and H&E-stained images could identify patterns and features indicative of colorectal cancer, potentially even discovering new markers not previously recognized by human experts. Moreover, the integration of ML and DL approaches with high-throughput TPEM systems capable of generating images of entire histological slides [e.g., widefield SHG microscopy (41, 42), and compact higher harmonic generation microscopy for unprocessed tissue (43)] will enable direct comparisons between H&E-stained slides and their TPEM counterparts (44). This integration could simplify TPEM images interpretation, rendering it a promising avenue for further exploration. It could also pave the way for personalized treatment plans based on the specific characteristics of each tumor, ultimately enhancing patient outcomes and revolutionizing colorectal cancer care.

By integrating TPEM microscopy with quantitative parameters derived from image texture analysis, including histogram analysis, gray level co-occurrence matrix, fractal analysis, and FFT analysis applied to SHG collagen images obtained from malignant polyps and polyps with pseudoinvasions, we have demonstrated the capability to distinguish between true invasion and pseudoinvasion sites based on collagen distribution in colonic polyps. Among the 12 parameters examined, 11 parameters exhibited statistically significant differences between the two considered classes. Our quantitative findings align with the observed alterations in collagen fiber organization, indicating a random organization in pseudoinvasion areas and a more structured distribution in true invasion sites.

5 Conclusions

Using two-photon excitation microscopy techniques, we identified significant qualitative and quantitative alterations in the ultrastructure of collagen within the two tumor types under investigation. There were significant differences in the orientation and quantity of collagen fibers between true invasion and pseudoinvasion in colonic polyps. Two-photon excitation microscopy demonstrated clear superiority in visualizing stromal changes when compared to conventional H&E images. Our research indicates that the integration of two-photon excitation microscopy with quantitative image analysis has the potential to distinguish true invasion in malignant polyps from pseudoinvasion, based on distinctive content and distribution of collagen fibers. This approach proves particularly valuable in cases presenting diagnostic challenges and may serve as a useful method in the early detection of colorectal cancer.

Data availability statement

The datasets generated during and/or analyzed during the current study are available in a public repository (DOI: 10.17605/OSF.IO/JDRBN).

Ethics statement

The studies involving humans were approved by Central University Emergency Military Hospital (Bucharest, Romania).

The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MF: Validation, Writing – review & editing, Writing – original draft, Methodology, Investigation. LE: Writing – review & editing, Visualization, Validation, Methodology. RG: Data curation, Writing – review & editing, Visualization, Investigation. RH: Visualization, Formal analysis, Writing – original draft, Validation, Investigation. GS: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. MC: Project administration, Writing – review & editing, Supervision.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The use of the Chameleon Vision II (Coherent) Ti: Sapphire laser was possible due to European Regional Development Fund through Competitiveness Operational Program 2014–2020, Priority axis 1, Project No. P_36_611, MySMIS code 107066 – INOVABIOMED. This work was supported as well by the UEFISCDI grant PN-III-P4-PCE-2021–0444 (RESONANO).

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.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/CAAC.21660

2. Guren MG. The global challenge of colorectal cancer. *Lancet Gastroenterol Hepatol*. (2019) 4:894–5. doi: 10.1016/S2468–1253(19)30329–2

3. Keum NN, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol*. (2019) 16:713–32. doi: 10.1038/S41575–019-0189–8

4. Strum WB. Colorectal adenomas. *N Engl J Med*. (2016) 374:1065–75. doi: 10.1056/NEJMRA1513581

5. Young PE, Womeldorph CM. Colonoscopy for colorectal cancer screening. *J Cancer*. (2013) 4:217–26. doi: 10.7150/JCA.5829

6. Kaltenbach T, Anderson JC, Burke CA, Dominitz JA, Gupta S, Lieberman D, et al. Endoscopic removal of colorectal lesions: recommendations by the US multi-society task force on colorectal cancer. *Am J Gastroenterol*. (2020) 115:435–64. doi: 10.14309/AJG.0000000000000555

7. Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology*. (2020) 76:182–8. doi: 10.1111/his.13975

8. Aarons CB, Shanmugan S, Bleier JIS. Management of Malignant colon polyps: Current status and controversies. *World J Gastroenterol*. (2014) 20:16178–83. doi: 10.3748/WJG.V20.I43.16178

9. Bujanda L, Cosme A, Gil I, Arenas-Mirave JI. Malignant colorectal polyps. *World J Gastroenterol*. (2010) 16:3103–11. doi: 10.3748/WJG.V16.I25.3103

10. Muto T, Bussey HJ, Morson BC. Pseudo-carcinomatous invasion in adenomatous polyps of the colon and rectum. *J Clin Pathol*. (1973) 26:25–31. doi: 10.1136/JCP.26.1.25

11. Tae JB, Han DS, Ahn SB, Cho HS, Eun CS, Jeon YC, et al. Pseudoinvasion in an adenomatous polyp of the colon mimicking invasive colon cancer. *Gut Liver*. (2009) 3:130–3. doi: 10.5009/GNL.2009.3.2.130

12. Shepherd NA, Griggs RKL. Bowel cancer screening-generated diagnostic conundrum of the century: Pseudoinvasion in sigmoid colonic polyps. *Modern Pathol*. (2015) 28:S88–94. doi: 10.1038/MODPATHOL.2014.138

13. Tanizawa T, Seki T, Nakano M, Kamiyama R. Pseudoinvasion of the colorectal polypoid tumors: Serial section study of problematic cases. *Pathol Int*. (2003) 53:584–90. doi: 10.1046/j.1440-1827.2003.01527.x

14. Nallala J, Griggs R, Lloyd GR, Stone N, Shepherd NA. Infrared spectroscopic analysis in the differentiation of epithelial misplacement from adenocarcinoma in sigmoid colonic adenomatous polyps. *Clin Pathol*. (2022) 15. doi: 10.1177/2632010X221088960

15. Yantiss RK, Bosenberg MW, Antonioli DA, Odze RD. Utility of MMP-1, p53, E-cadherin, and collagen IV immunohistochemical stains in the differential diagnosis of adenomas with misplaced epithelium versus adenomas with invasive adenocarcinoma. *Am J Surg Pathol*. (2002) 26:206–15. doi: 10.1097/00000478–200202000–00007

16. Denk W, Strickler JH, Webb WW. Two-photon laser scanning fluorescence microscopy. *Sci (1979)*. (1990) 248:73–6. doi: 10.1126/science.2321027

17. Franken PA, Hill AE, Peters CW, Weinreich G. Generation of optical harmonics. *Phys Rev Lett*. (1961) 7:118. doi: 10.1103/PhysRevLett.7.118

18. Hansen WP, Fine S. Optical second harmonic generation in biological systems. *Appl Optics*. (1971) 10:2350–3. doi: 10.1364/AO.10.002350

19. Hristu R, Eftimie LG, Stanciu SG, Glogojeanu RR, Gheorghita P, Stanciu GA. Assessment of extramammary paget disease by two-photon microscopy. *Front Med (Lausanne)*. (2022) 9:839786. doi: 10.3389/FMED.2022.839786

20. Fast A, Lal A, Durkin AF, Lentsch G, Harris RM, Zachary CB, et al. Fast, large area multiphoton exoscope (FLAME) for macroscopic imaging with microscopic resolution of human skin. *Sci Rep*. (2020) 10. doi: 10.1038/S41598–020-75172–9

21. Dilipkumar A, Al-Shemmary A, Kreiß L, Cvecek K, Carlé B, Knieling F, et al. Label-free multiphoton endomicroscopy for minimally invasive *in vivo* imaging. *Advanced Sci*. (2019) 6(8):1801735. doi: 10.1002/advs.201801735

22. Cho HJ, Chun HJ, Kim ES, Cho BR. Multiphoton microscopy: An introduction to gastroenterologists. *World J Gastroenterol*. (2011) 17:4456–60. doi: 10.3748/WJG.V17.I40.4456

23. Campagnola PJ, Dong C. Second harmonic generation microscopy: principles and applications to disease diagnosis. *Laser Photon Rev*. (2011) 5:13–26. doi: 10.1002/lpor.200910024

24. Chen X, Nadiarynkh O, Plotnikov S, Campagnola PJ. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nat Protoc*. (2012) 7:654–69. doi: 10.1038/nprot.2012.009

25. Tilbury K, Hocker J, Wen BL, Sandbo N, Singh V, Campagnola PJ. Second harmonic generation microscopy analysis of extracellular matrix changes in human idiopathic pulmonary fibrosis. *J BioMed Opt*. (2014) 19:86014. doi: 10.1117/1.JBO.19.8.086014

26. Adur J, Pelegati VB, Thomaz AA, Baratti MO, Andrade LALA, Carvalho HF, et al. Second harmonic generation microscopy as a powerful diagnostic imaging modality for human ovarian cancer. *J Biophotonics*. (2014) 48:37–48. doi: 10.1002/jbio.201200108

27. Hristu R, Eftimie LG, Stanciu SG, Tranca DE, Paun B, Sajin M, et al. Quantitative second harmonic generation microscopy for the structural characterization of capsular collagen in thyroid neoplasms. *BioMed Opt Express*. (2018) 9:3923–36. doi: 10.1364/BOE.9.003923

28. Bueno JM, Avila FJ, Hristu R, Stanciu SG, Eftimie L, Stanciu GA. Objective analysis of collagen organization in thyroid nodule capsules using second harmonic generation microscopy images and the Hough transform. *Appl Opt*. (2020) 59:6925–31. doi: 10.1364/AO.393721

29. Zipfel WR, Williams RM, Christiet R, Nikitin AY, Hyman BT, Webb WW. Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation. *Proc Natl Acad Sci USA*. (2003) 100:7075–80. doi: 10.1073/PNAS.0832308100

30. Eftimie LG, Glogojeanu RR, Tejaswee A, Gheorghita P, Stanciu SG, Chirila A, et al. Differential diagnosis of thyroid nodule capsules using random forest guided selection of image features. *Sci Rep*. (2022) 12. doi: 10.1038/s41598-022-25788-w

31. Hristu R, Stanciu SG, Dumitru A, Eftimie LG, Paun B, Tranca DE, et al. PSHG-TISS: A collection of polarization-resolved second harmonic generation microscopy images of fixed tissues. *Sci Data*. (2022) 9:376. doi: 10.1038/S41597–022-01477–1

32. Adur J, Carvalho HF, Cesar CL, Casco VH. Nonlinear optical microscopy signal processing strategies in cancer. *Cancer Inform*. (2014) 13:67–76. doi: 10.4137/CIN.S12419

33. Osman OS, Selway JL, Harikumar PE, Stocker CJ, Wargent ET, Cawthorne MA, et al. A novel method to assess collagen architecture in skin. *BMC Bioinf*. (2013) 14. doi: 10.1186/1471–2105-14–260

34. Frantz C, Stewart KM, Weaver VM. The extracellular matrix at a glance. *J Cell Sci*. (2010) 123:4195–200. doi: 10.1242/jcs.023820

35. Molavi D, Argani P. Distinguishing benign dissecting mucin (stromal mucin pools) from invasive mucinous carcinoma. *Adv Anat Pathol*. (2008) 15:1–17. doi: 10.1097/PAP.0b013e31815e52aa

36. Kumar V, Gupta P. Importance of statistical measures in digital image processing. *International Journal of Emerging Technology and Advanced Engineering* (2012) 2(8):56–62.

37. Mostaço-Guidolin LB, Ko AC-T, Wang F, Xiang B, Hewko M, Tian G, et al. Collagen morphology and texture analysis: From statistics to classification. *Sci Rep*. (2013) 3:1–10. doi: 10.1038/srep02190

38. de Andrade Natal R, Adur J, Cesar CL, Vassallo J. Tumor extracellular matrix: lessons from the second-harmonic generation microscopy. *Surg Exp Pathol*. (2021) 4:7. doi: 10.1186/s42047–021-00089–0

39. Bousis D, Verras G-I, Bouchagier K, Antzoulas A, Panagiotopoulos I, Katinioti A, et al. The role of deep learning in diagnosing colorectal cancer. *Gastroenterol REVIEW-PRZEGLAD GASTROENTEROLOGICZNY*. (2023) 18:266–73. doi: 10.5114/pg.2023.129494

40. Pena AM, Baldeweck T, Decencière E, Koudoro S, Victorin S, Raynaud E, et al. *In vivo* multiphoton multiparametric 3D quantification of human skin aging on forearm and face. *Sci Rep*. (2022) 12. doi: 10.1038/s41598-022-18657-z

41. Mirsanaye K, Castaño LU, Kamaliddin Y, Golaraei A, Augulis R, Kontenis L, et al. Machine learning-enabled cancer diagnostics with widefield polarimetric second-harmonic generation microscopy. *Sci Rep*. (2022) 12. doi: 10.1038/S41598–022-13623–1

42. Padrez Y, Golubewa L, Kulahava T, Vladimirskaja T, Semenkova G, Adzerikho I, et al. Quantitative and qualitative analysis of pulmonary arterial hypertension fibrosis using wide-field second harmonic generation microscopy. *Sci Rep*. (2022) 12. doi: 10.1038/S41598–022-11473–5

43. Kok SD, Schaap PMR, Dommelen van L, Huizen van LMG, Dickhoff C, Dijkum Nieveen-van EM, et al. Compact portable higher harmonic generation microscopy for the real time assessment of unprocessed thyroid tissue. *J Biophotonics*. (2023) 17: e202300079. doi: 10.1002/jbio.202300079

44. Enache A, Eftimie LG, Graur AM, Glogojeanu RR, Geambaşu A, Voinea OC, et al. Assessment of cerebral tumors and metastases by two-photon excitation microscopy. *ROMANIAN J OF MILITARY Med*. (2023) 126:418–23. doi: 10.55453/rjmm.2023.126.4.11

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 03 July 2024
DOI 10.3389/fonc.2024.1397266

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Nima Gilani,
New York University, United States
Bilgin Kadri Aribas,
Bülent Ecevit University, Türkiye

*CORRESPONDENCE
Andu Zhang
zhangandu@hebmu.edu.cn
Gaofeng Shi
gaofengs1962@hebmu.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 07 March 2024
ACCEPTED 17 June 2024
PUBLISHED 03 July 2024

CITATION
Li Y, Su X, Shang Y, Liu H, Wang W, Zhang A and Shi G (2024) Comparative evaluation of imaging methods for prognosis assessment in esophageal squamous cell carcinoma: focus on diffusion-weighted magnetic resonance imaging, computed tomography and esophagography.
Front. Oncol. 14:1397266.
doi: 10.3389/fonc.2024.1397266

COPYRIGHT
© 2024 Li, Su, Shang, Liu, Wang, Zhang and Shi. 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.

Comparative evaluation of imaging methods for prognosis assessment in esophageal squamous cell carcinoma: focus on diffusion-weighted magnetic resonance imaging, computed tomography and esophagography

Yang Li[1†], Xiaohua Su[2†], Yuguang Shang[3], Hui Liu[1], Weishuai Wang[4], Andu Zhang[3*] and Gaofeng Shi[1*]

[1]Department of Computed Tomography and Magnetic Resonance Imaging, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China, [2]Department of Oncology, Hebei General Hospital, Shijiazhuang, China, [3]Department of Radiotherapy, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China, [4]CS Service AP, Siemens Healthineers Digital Health Technology (Shanghai) Co., Ltd. Beijing Branch, Beijing, China

Objective: To identify the most sensitive imaging examination method to evaluate the prognosis of esophageal squamous cell carcinoma (ESCC).

Materials and methods: Thirty patients with esophageal squamous cell carcinoma (ESCC) participated in the study and underwent chemoradiotherapy (CRT). They were divided into two groups based on their survival status: the survival group and non-survival group. The diagnostic tests were utilized to determine the most effective imaging examination method for assessing the prognosis.

Results: 1. There were no significant differences in tumor length shown on esophagography or computed tomography (CT) or the maximal esophageal wall thickness shown on CT at the specified time points between the two groups. 2. The tumor length on diffusion-weighted imaging (DWI) in the survival group was significantly lower than in the non-survival group at the end of the sixth week of treatment (P=0.001). The area under the ROC curve was 0.840 (P=0.002), and the diagnostic efficiency was moderately accurate. 3. The apparent diffusion coefficient (ADC) values of the survival group were significantly higher than those in the non-survival group at the end of the fourth week and sixth week of treatment (both P<0.001). Areas under the curve were 0.866 and 0.970, with P values of 0.001 and <0.001 and good diagnostic accuracy. Cox regression

analyses indicated the ADC at the end of the sixth week of treatment was an independent risk factor.

Conclusions: Compared with esophagography and CT, DW-MRI has certain advantages in predicting the prognosis of ESCC.

KEYWORDS

esophageal cancer, radiotherapy, diffusion-weighted imaging, esophagography, computer tomography

1 Introduction

Esophageal cancer (EC) represents a severe malignancy mainly due to its poor prognosis and survival rate, ranking sixth among all cancers in terms of mortality and eighth among the most commonly occurring cancers on a global scale (1).Since EC patients can only be diagnosed when they present with symptoms such as dysphagia, dysphagia, anemia, or weight loss, chemoradiotherapy (CRT) is widely used in cases of unresectable EC (2, 3). Thus, it has been deemed necessary to conduct a close assessment of the efficacy of CRT, which is very important for the adjustment of individualized treatment strategies for patients. Various methods are used to assess tumor response, and esophagography and computed tomography (CT) are routine tools for the evaluation of esophageal tumor response to treatment (4, 5). The criteria that are currently widely used in China to evaluate the efficacy of radiotherapy for EC follow the evaluation criteria for the efficacy of EC proposed by Professor Wan Jun in 1989, which are based on esophagography.

Esophagography is a noninvasive and inexpensive examination that details the structure and function of the esophageal mucosa and can determine the scope of esophageal lesions. Although it has wide availability, low cost, and rapid performance, there are also apparent limitations of this examination: esophagography cannot be used to evaluate the thickness of the esophageal wall or regional lymph node metastasis (6). However, the internationally standard evaluation criteria for the therapeutic effect of solid tumors, the Response Evaluation Criteria in Solid Tumors (RECIST) standard version 1.1 (7), are not entirely suitable for evaluating the efficacy of esophageal tumors after CRT.

CT can clearly show the thickness of the esophageal wall, tumor invasion and lymph node, and distant metastasis. However, because esophageal tumors originate in cavity organs, the esophageal structure still exists after CRT. There is edema in the wall of the esophagus after radiotherapy; CT is unable to differentiate between viable tumors, inflammatory changes, and scar tissue (8). CT is not sensitive enough to accurately evaluate treatment response. Functional imaging technology can compensate for the deficiency of morphological imaging technology and reflect the functional metabolism of tumor cells before morphological changes; it also has advantages for evaluating the efficacy of malignant tumors (9, 10).

Diffusion-weighted magnetic resonance imaging (DW-MRI) is an evolving imaging technique that contributes considerably and positively to the treatment of EC (11, 12). The apparent diffusion coefficient (ADC) is calculated for each pixel in an image, exhibited by a parametric map (13). The ADC is a reliable and reproducible value that serves as a promising noninvasive indicator that can be used to assess tumor aggressiveness as well as tumor responses to CRT (14, 15). The potential of the ADC value as a helpful marker has been well documented in various studies, which have highlighted its ability to predict treatment response and the survival probabilities of patients with esophageal squamous cell carcinoma (ESCC) (16).

This study analyzed the length of lesions shown on esophagography, CT, and DWI, the maximal esophageal wall thickness shown on CT, and the ADC values measured in DWI at specified time points, combined with diagnostic tests, between the survival group and the non-survival group to determine the most sensitive imaging examination method to evaluate the prognosis of patients with ESCC and provide valuable reference information for clinical work.

2 Material and methods

2.1 Patient selection criteria

All enrolled patients were diagnosed with ESCC by pathology and had an Eastern Cooperative Oncology Group of 0~2 (ECOG, which used to evaluate the patient's performance status). All patients had no previous history of cancer or diseases that may affect the completion of treatment. No age limits were set. No distant metastases were found during routine imaging studies (MRI for the brain; CT for the lung, liver, and bone). All patients were first-time radiotherapy recipients. There were no MRI test contraindications, and patients approved all examinations.

2.2 Study population

A total of 30 patients with ESCC who were admitted to our hospital between February 2017 and June 2017 met the inclusion

TABLE 1 Characteristic of patients in survival group and non-survival group.

Characteristics	Survival group (n=12)	Non-Survival group (n=18)	Z/t	*P*
Gender, n (%)			4.678	<0.001#
Male	6 (50.0)	17 (94.4)		
Female	6 (50.0)	1 (5.6)		
Age	53-79 (66)	53-87 (70)	-0.811*	0.424*
Location, n (%)			3.202	0.002#
Cervical	1 (8.3)	0 (0)		
Upper	1 (8.3)	3 (16.7)		
Middle	9 (75.0)	14 (77.8)		
Lower	1 (8.3)	1 (5.6)		
T stage, n (%)				
T1-2	5 (41.7)	5 (27.8)	3.336	0.002#
T3	2 (16.7)	1 (5.6)		
T4	5 (41.7)	12 (66.7)		
N stage, n (%)			4.842	<0.001#
N0	2 (16.7)	0 (0.0)		
N1	5 (41.7)	6 (33.3)		
N2	5 (41.7)	12 (66.7)		
TNM stage, n (%)			3.981	<0.001#
I	2 (16.7)	0 (0.0)		
II	4 (33.3)	2 (11.1)		
III	6 (50.0)	16 (88.9)		
GTV volume (cm^3), range (median)	18.65-272.31 (74.17)	25.14-284.20 (117.71)	-0.721*	0.477*
Dose (Gy), range (median)	50.4-60 (60)	50.4-60 (60)	0.335	0.787#

#, Mann-Whitney U test; *, Student's t-test.

criteria. All patients were classified according to the 7th edition of the TNM staging system [International Union for Cancer Control (UICC)]. As of the follow-up date, all patients were divided into a survival group and a non-survival group according to their survival status. Details of the patients in the two groups are shown in Table 1.

2.3 Delineation of the target volume and organs at risk

Based on the CT images, the gross tumor volume of the primary tumor (GTV-p) and gross tumor volume of the metastatic lymph nodes (GTV-n) were outlined according to the department protocol on ESCC tissues. The OARs were also outlined in a manner consistent with international guidelines. Specifically, and in principle, CT images revealed that the standard GTV-p was a tumor size of more than 5 mm wide or an esophageal diameter of more than 10 mm with esophageal wall stiffness or full-wall thickening; the clinical target volume of the primary tumor (CTV-p) was contoured by extending 0.5 cm around the GTV-p in the axial direction and 2.0 cm in both the superior and inferior directions. Finally, the planning target volume of the primary tumor (PTV-p) was outlined around the CTV-p with a positive margin of 0.5 cm in the axial direction and 1.0 cm in both the superior and inferior directions. Afterwards, the GTV-n was defined as a paraesophageal lymph node with a short-axis diameter greater than 1.0 cm, and for lymph nodes in particular regions, such as those in the paraesophageal region or tracheoesophageal groove and cardiophrenic angle lymph nodes, the standard guideline was a short-axis diameter greater than 0.5 cm (17, 18). The PTV-n is the result of uniformly extending the GTV-n by 1.0 cm.

2.4 Treatment plan and delivery

The prescription radiation doses for the CRT group ranged from 50.4 to 60 Gy (median 60 Gy). All patients received intensity-

modulated radiotherapy (IMRT), and all treatment plans included 1.8–2.0 Gy per fraction, 5 fractions per week, with a treatment time of 6 weeks. Each treatment plan required that the dose received by 95% of the PTV (PTV D95) be more than 100% of the prescription dose. The OAR doses were limited to lung $V_5 \leq 65\%$, $V_{20} \leq 30\%$, and $V_{30} \leq 20\%$, $heart_{mean} \leq 30$ Gy and $cord_{max}<45$ Gy. The treatment plan was completed by a physiotherapist, as required, and confirmed by a superior physician. All patients completed the treatment plan.

2.5 Chemotherapy

The standard chemotherapy regimens were as follows: FP: cisplatin 25 mg/m2 × 3 days, 5-FU 450–500 mg/m2 × 5 days; and TP: cisplatin 25 mg/m2 × 3 days, paclitaxel 135-150 mg/m2 × 1 day.

2.6 Observation indicators

The length of lesions shown on esophagography, chest CT and DWI, the maximal esophageal wall thickness shown on chest CT and the ADC values shown on DWI of the two groups were measured before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment and the end of the sixth week of treatment.

2.6.1 Tumor length on esophagography

All patients were examined with esophagography at the specified time points. The tumor length on esophagography was measured on the axis image.

2.6.2 Tumor length and maximal esophageal wall thickness on chest CT

All patients underwent CT at the indicated time points. Tumor length on CT was measured on a section showing the tumor in its entirety in the sagittal position. The maximum esophageal wall thickness of the tumor in the same horizontal region on the transverse section was measured based on the location of the tumor prior to CRT and the corresponding anatomic landmarks.

2.6.3 Tumor length on DWI and ADC measurements

All patients underwent DW-MRI at the specified time points. The MRI examination involved a Siemens 3.0 T MRI scanner (Siemens Healthineers), an 18-channel body coil and a scanning sequence, including T1-weighted imaging (T1WI), T2-weighted imaging (T2WI) and DWI sequences. The b values (dispersion −sensitive gradient) were as follows (19–21): 0 and 600 s/mm^2. The length of the lesions was measured on the axis image based on the DWI display signal. The images segmentation and measurement were performed using ITK-SNAP software version 3.8.0 (http://www.itksnap.org). Before delineating the tumor boundaries, we carefully reviewed the images across different sequences and selected the images of b-value with max image contrast between the lesion and background tissue. To address the issue of image clarity, we adjusted the window width and window level to show the lesion better (Table 2). Then, the lesion with high signal intensity was delineated layer by layer on the selected relatively high b-value. The ADC measurement process as follows: Using DWI images with b-values of 0 and 600 s/mm^2, we reconstructed the ADC maps. Both the reconstructed ADC maps and the original b=600 images were imported into the ITK-SNAP software. In ITK-SNAP, we meticulously delineated the tumor boundaries on the relatively high b-value images (b=600) layer by layer. These delineated regions of interest (ROIs) were then overlaid onto the ADC maps to measure the ADC values accurately.

All of the above imaging data were analyzed by two experienced radiologists (with 10 and 8 years of experience in clinical radiology, respectively) who reached a consensus. In order to assess the intra-observer reproducibility of the measurements, a concordance analysis of 50 images (10 images per observation indicator) was performed using the values measured by two radiologists. The ICC interpretations were as follows: excellent (ICC ≥ 0.90); good (0.75 ≤

TABLE 2 MRI scan sequences and parameters.

Parameters	Sequences		
	T1WI (Axial)	T2WI (Axial)	DWI (Axial)
TR (ms)	4.00	3000	8200
TE (ms)	1.29	81	49
FOV Read (mm)	380	380	300
FOV Phase (%)	81.3	100	68
Slice Thickness (mm)	3.5	6	3
Distance Factor (%)	20	20	25
Base Resolution	320	320	100
Phase Resolution (%)	75	/	100
Phase Oversampling (%)	30	37.5	40
Slicer Oversampling (%)	33.3	/	/
Average	1	1	2
b-value (s/mm^2)	/	/	600
Phase encoding	A>>P	R>>L	A>>P
Blade coverage (%)	/	100	/
Bandwidth (Hz)	1040	781	2272
RF Pulse Mode	Fast	Fast	Normal
Turbo Factor	/	43	/
Echo spacing (ms)	/	3.66	0.54
Fat-Water Contrast	Dixon	SPAIR	Fat Saturation

ICC < 0.90); moderate (0.50 ≤ ICC < 0.75); and poor (ICC < 0.50), and the ICCs are 0.98 (tumor length on esophagography), 0.83 (tumor length on CT), 0.98 (maximal esophageal wall thickness on CT), 0.94 (tumor length on DWI) and 0.90 (ADC), respectively, with the p values < 0.05, and the above results indicate good consistency and repeatability. Final analysis was performed using the average of the values measured by the two radiologists.

2.7 Statistical methods

Statistical analysis was performed using GraphPad software (GraphPad Prism v7.0, GraphPad Software). The normality of quantitative data was assessed using a Kolmogorov-Smirnov test. Categorical variables and continuous data conforming to a nonnormal distribution were analyzed using the nonparametric Mann-Whitney U test, while continuous data conforming to a normal distribution were analyzed using Student's t-test. The variables at different time points were determined using the repeated measures analysis of variance (ANOVA). Cox regression models and Kaplan-Meier analyses were conducted to estimate overall survival (OS), and the log-rank test was applied to assess differences between groups. P<0.05 was considered statistically significant. Receiver operating characteristic (ROC) curves were plotted for the observation indicators, and the area under the curve (AUC) was calculated. The optimal cutoff values were chosen as follows: when Youden's index (YI=sensitivity+specificity-1) was the maximum, AUC values between 0.9 and 1.0 were deemed "accurate"; 0.7-0.9 "moderately accurate"; and 0.5-0.7 "uninformative" (22, 23).

3 Results

3.1 OS and local control

As of January 1, 2021, all patients were followed up for over 3 years, no patients were lost to follow-up, and the total follow-up rate was 100%. There were 12 patients in the survival group and 18 in the non-survival group. OS was evaluated from the date of radiotherapy to the date of death or the last follow-up. All patients' 1- and 3-year OS rates were 80.0% and 43.3%, respectively, with a median survival time of 30 months. All patients' 1- and 3-year LC rates were 83.2% and 50.6%, respectively, with a median LC time of 31 months.

3.2 Tumor length shown on esophagography

Tumor lengths shown on esophagography in all patients before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment, and at the end of the sixth week of treatment were 6.71 ± 2.63 cm, 5.47 ± 2.55 cm, 3.98 ± 1.62 cm, and 3.01 ± 1.65 cm, respectively. Tumor length varied at different points in the survival group (F=16.897, P<0.001) and in the non-survival group (F=37.782, P<0.001). With increased radiotherapy sessions, tumor length showed a sustained and rapid downward trend in both groups, and no significant difference in tumor length was shown on esophagography between the two groups at the specified time points (Figure 1A). Diagnostic tests were carried out at the specified time points to assess tumor length on esophagography.

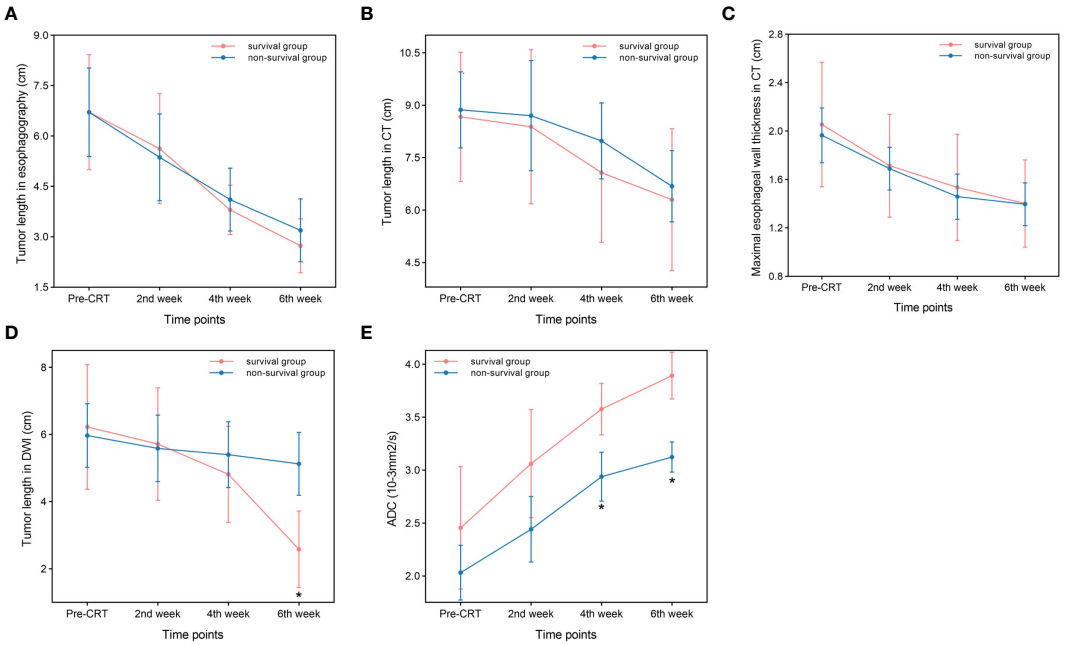


FIGURE 1
Comparison between the survival group and non-survival group. **(A)** Tumor length in esophagography. **(B)** Tumor length in CT. **(C)** Maximal esophageal wall thickness in CT. **(D)** Tumor length in DWI (P=0.001,at the end of the sixth week). **(E)** ADC. (P<0.001,at the end of the fourth and sixth week) *p<0.05.

The ROC curve showed that the P value was greater than 0.05, and the area under the curve was between 0.484-0.569 (Table 3), indicating that the diagnostic effectiveness of tumor length measured on esophagography was low.

3.3 Tumor length shown on chest CT

Tumor lengths on chest CT in all patients before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment, and the end of the sixth week of treatment were 8.79 ± 2.45 cm, 8.58 ± 3.24 cm, 7.62 ± 2.59 cm, and 6.53 ± 2.53 cm, respectively. Tumor length varied at different points in the survival group (F=6.820, P=0.001) and in the non-survival group (F=7.005, P=0.002). With an increase in radiotherapy sessions, tumor length on CT showed a sustained and rapid downward trend in both groups, and there was no significant difference in tumor length shown on CT between the two groups at the specified time points (Figure 1B). Diagnostic tests were carried out at the specified time points for tumor length on CT. The ROC curve showed that the P value was greater than 0.05, and the area under the curve was between 0.495 and 0.602 (Table 3), indicating that the diagnostic effectiveness of tumor length shown on chest CT was low.

TABLE 3 ROC curve analysis.

Observation indicators	AUC	P	95% CI	
			lower	Upper
Tumor length in esophagography (cm)				
before CRT	0.484	0.882	0.258	0.710
2nd week	0.484	0.882	0.263	0.704
4th week	0.569	0.525	0.357	0.782
6th week	0.523	0.832	0.310	0.736
Tumor length in CT (cm)				
before CRT	0.521	0.849	0.299	0.743
2nd week	0.495	0.966	0.270	0.720
4th week	0.602	0.352	0.383	0.821
6th week	0.539	0.719	0.311	0.768
The maximal esophageal wall thickness in CT (cm)				
before CRT	0.500	1.000	0.264	0.736
2nd week	0.495	0.966	0.263	0.727
4th week	0.498	0.983	0.268	0.727
6th week	0.465	0.751	0.244	0.686
Tumor length in DWI (cm)				
before CRT	0.479	0.849	0.249	0.709
2nd week	0.477	0.832	0.249	0.705
4th week	0.569	0.525	0.350	0.789
6th week	0.840	0.002	0.700	0.980
ADC ($10^{-3}mm^2/s$)				
before CRT	0.576	0.485	0.340	0.813
2nd week	0.745	0.025	0.560	0.931
4th week	0.866	0.001	0.740	0.992
6th week	0.970	<0.001	0.918	1.000

2nd week, the end of the second week.4th week, the end of the fourth week. 6th week, the end of the sixth week.

3.4 Maximal esophageal wall thickness shown on chest CT

The maximal esophageal wall thicknesses shown on chest CT in all patients before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment, and at the end of the sixth week of treatment were 2.00 ± 0.61 cm, 1.70 ± 0.49 cm, 1.49 ± 0.52 cm, and 1.40 ± 0.44 cm, respectively. The maximal esophageal wall thickness varied at different points in the survival group (F=17.775, P<0.001) and in the non-survival group (F=58.602, P<0.001). With an increase in the number of radiotherapy sessions, the maximal esophageal wall thickness shown on CT initially showed a sustained and rapid downward trend, which slowed in the fifth week and sixth week in both groups, and there was no significant difference in the maximal esophageal wall thickness shown on chest CT between the two groups at the specified time points (Figure 1C). The ROC curve showed that the P value was greater than 0.05, and the area under the curve was between 0.465-0.500 (Table 3), indicating that the diagnostic effectiveness of the maximal esophageal wall thickness shown on CT was low.

3.5 Tumor length shown on DW-MRI

Tumor lengths shown on DW-MRI in all patients before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment, and at the end of the sixth week of treatment were 6.07 ± 2.32 cm, 5.64 ± 2.23 cm, 5.17 ± 2.07 cm, and 4.11 ± 2.22 cm, respectively. Three patients did not have a high signal expression on DWI at the end of the sixth week of treatment, all of whom had long-term survival, with survival times of 39 months, 42 months, and 43 months. Tumor length shown on DWI varied at different points in the survival group (F=21.379, P<0.001) but not in the non-survival group (F=3.146, P=0.057). With an increase in the number of radiotherapy sessions, tumor length shown on DWI showed a continuous and rapid declining trend, which was more evident after the fourth week in the survival group. In addition, three patients did not have a high signal expression on DWI at the end of treatment. Tumor length on DWI increased slightly in the second week compared to pretreatment and then showed a slow downward trend in the non-survival group. All patients in this group had a high signal expression on DWI at the end of treatment (Figure 1D). There was no significant difference in tumor length shown on DWI between the two groups before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment, and tumor length in the survival group was significantly

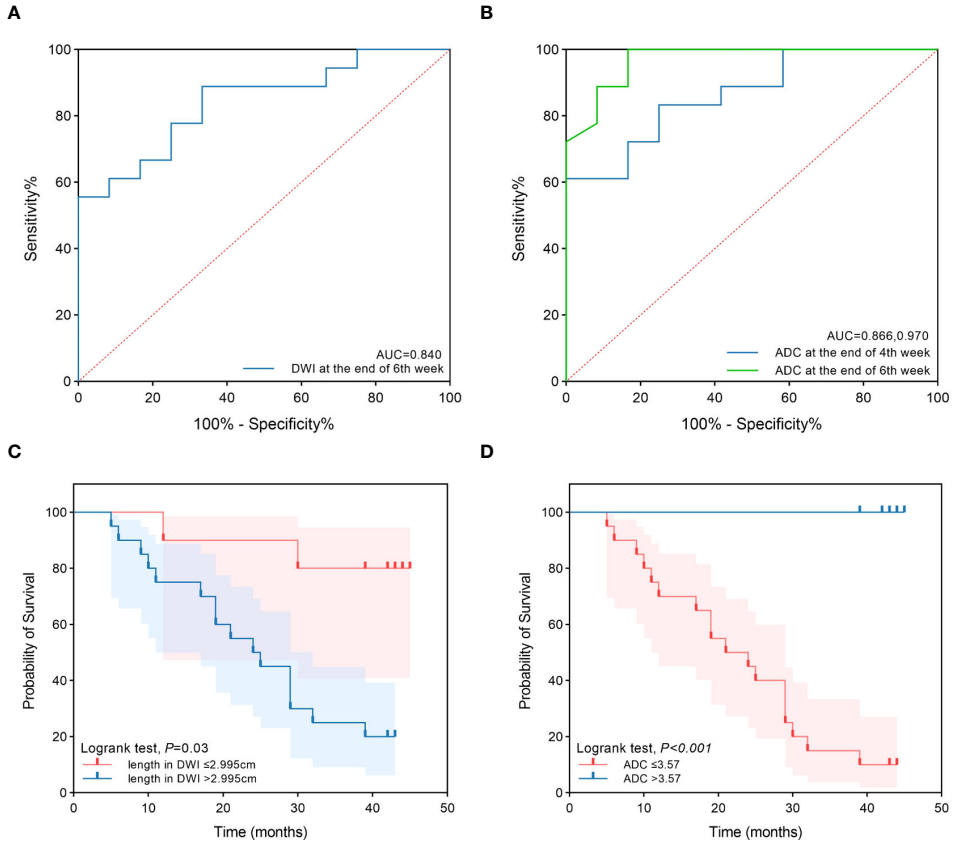


FIGURE 2
The ROC curves and survival curves. **(A)** The ROC curves of tumor length in DWI at the end of 6th week of treatment to predict prognosis. **(B)** The ROC curves of ADC at the end of 4th and 6th week of treatment to predict prognosis. **(C)** The survival curves of the two groups for tumor length in DWI at the end of 6th week ≤2.995cm and >2.995cm. **(D)** The survival curves of the two groups for ADC at the end of 6th week ≤3.57×10^{-3}mm^2/s and >3.57×10-3mm^2/s.

lower than that in the non-survival group at the end of the sixth week of treatment (t=-3.687, P=0.001). Diagnostic tests were carried out at the specified time points for tumor length on DWI. The ROC curve showed that with tumor length shown on DWI as the diagnostic index at the end of the sixth week of treatment, the P value was 0.002, and the area under the curve was 0.840. The diagnostic efficacy was moderately accurate, with a cutoff value of 2.995 cm, a sensitivity of 0.889, and a specificity of 0.667 (Figure 2A, Table 3). With tumor length shown on DWI at the end of the sixth week of treatment as the cutoff point to divide the whole group of patients into two groups (2.995 cm), there were ten patients with a tumor length ≤2.995 cm and 20 patients with a tumor length > 2.995 cm. The 1- and 3-year OS rates of the two groups were 90.0% and 80.0% and 75.0% and 25.0%, respectively, with medium survival times of 40 months and 24 months, respectively (χ^2 = 8.531, P=0.003) (Figure 2C).

3.6 ADC on DWI

The ADC values on DWI in all patients before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment and the end of the sixth week of treatment were 2.20 ± 0.72×10-3 mm^2/s, 2.69 ± 0.75×10-3 mm^2/s, 3.19 ± 0.53×10-3 mm^2/s, and 3.43 ± 0.49×10-3 mm^2/s, respectively. The ADC values differed at different time points in the survival group (F=18.939, P<0.001) and the non-survival group (F=45.122, P<0.001). With an increase in radiotherapy sessions, the ADC value showed a continuous upward trend in the survival and non-survival groups. However, the upward trend in the non-survival group slowed significantly from the fourth week. There were no significant differences in ADC values between the two groups before CRT and at the end of the second week of treatment. The ADC value of the survival group was significantly higher than that of the non-survival group at the end of the fourth week and sixth week of treatment (t=3.942, 6.592, P<0.001, P<0.001) (Figure 1E). Diagnostic tests were carried out at the specified time points for the ADC value. The ROC curve showed that with the ADC at the end of the fourth week of treatment and the end of the sixth week of treatment as the diagnostic index, the P values were 0.001 and <0.001, the areas under the curve were 0.866 and 0.970, and the diagnostic efficacies were moderately accurate and accurate, with cutoff values of 2.965×10-3 mm2/s and 3.570×10^{-3} mm^2/s, sensitivities of 1 and 0.833, and specificities of 0.611 and 1 (Figure 2B, Table 3). With the ADC at the end of the sixth week of treatment as the cutoff point to divide the whole group of patients

into two groups (3.570×10-3 mm^2/s), there were 20 patients with a tumor length ≤ 3.570×10-3 mm2/s and ten patients with a tumor length >3.570×10-3 mm^2/s. The 1- and 3-year OS rates of the two groups were 70.0% and 15.0% and 100.0% and 100.0%, respectively, with medium survival times of 23 months and 42 months, respectively (χ^2 = 18.843, P<0.001) (Figure 2D).

3.7 Cox regression analysis

Considering the potential clinical significance of the observation indicators for OS, we aimed to clarify the correlations of observation indicators with other traditional clinical features, including age, sex, tumor site, TNM stage, GTV, and prescription dose. The observation indicators were initially merged with other variables, and Cox analysis was subsequently performed. Then, univariate Cox analysis indicated that sex (*P*=0.035), TNM stage (*P*=0.051), tumor length on DWI at the end of the sixth week of treatment (*P*=0.001), and ADC at the end of the sixth week of treatment (*P* < 0.001) were all risk factors. Nonetheless, the ADC at the end of the sixth week of treatment (*P*=0.05) retained significance in the multivariate Cox regression analysis (Table 4).

3.8 Diagrams of typical cases

Figure 3 show the esophagography, CT, and DW images of one patient in the survival group before CRT, at the end of the second week of treatment, at the end of the fourth week of treatment, and at the end of the sixth week. Figure 4 shows the esophagography, CT, and DW images of one patient in the non-survival group at the specified time points.

4 Discussion

Malignant tumors are a severe primary disease that threatens human health and social development. EC is one of the leading causes of cancer-related death in China (24, 25); its incidence has prominent regional distribution characteristics, and ESCC is the primary tissue type (26, 27). It is of great significance to evaluate the therapeutic effect of malignant tumors objectively, quantitatively, and accurately, and prognostic indicators could guide individualized treatment decisions and thus improve the benefits of treatment. However, most methods, including CT, esophagography, endoscopic biopsy, and endoscopic ultrasonography (EUS), yield unsatisfactory results for tumor response to neoadjuvant CRT (nCRT) (28–30). Metabolic and functional imaging modalities such as 18F-fluorodeoxyglucose positron emission tomography integrated with CT (18F-FDG PET/CT) and DW-MRI may be more promising because they allow the biological and microstructural characterization of tumors and visualization of treatment-induced changes before volumetric changes become apparent (31, 32). However, PET/CT is expensive and not widely used, while the use of MRI in the treatment response evaluation of ECs has gained increasing interest.

TABLE 4 Univariate and multivariate Cox analysis for clinical characteristics.

Variables	Univariate Cox analysis				Multivariate Cox analysis			
	HR	95.0% CI		P	HR	95.0% CI		P
		Lower	Upper			Lower	Upper	
Sex	0.114	0.015	0.860	0.035	0.182	0.020	1.647	0.130
Age	1.034	0.967	1.105	0.327	–	–	–	–
TNM	3.857	0.996	14.931	0.051	2.171	0.454	10.392	0.332
Site	1.316	0.575	3.014	0.515	–	–	–	–
GTV volume	1.002	0.996	1.009	0.503	–	–	–	–
Dose	1.006	0.886	1.142	0.930	–	–	–	–
Tumor length in esophagography (6th)	1.010	0.785	1.299	0.940	–	–	–	–
Tumor length in CT(6th)	1.060	0.887	1.267	0.521	–	–	–	–
The maximal esophageal wall thickness in CT (6th)	1.050	0.389	2.833	0.923	–	–	–	–
Tumor length in DWI (6th)	1.489	1.168	1.900	0.001	0.937	0.637	1.378	0.742
ADC (6^{th})	0.069	0.019	0.252	<0.001	0.146	0.021	1.002	0.05

6th week, the end of the sixth week.

10.3389/fonc.2024.1397266

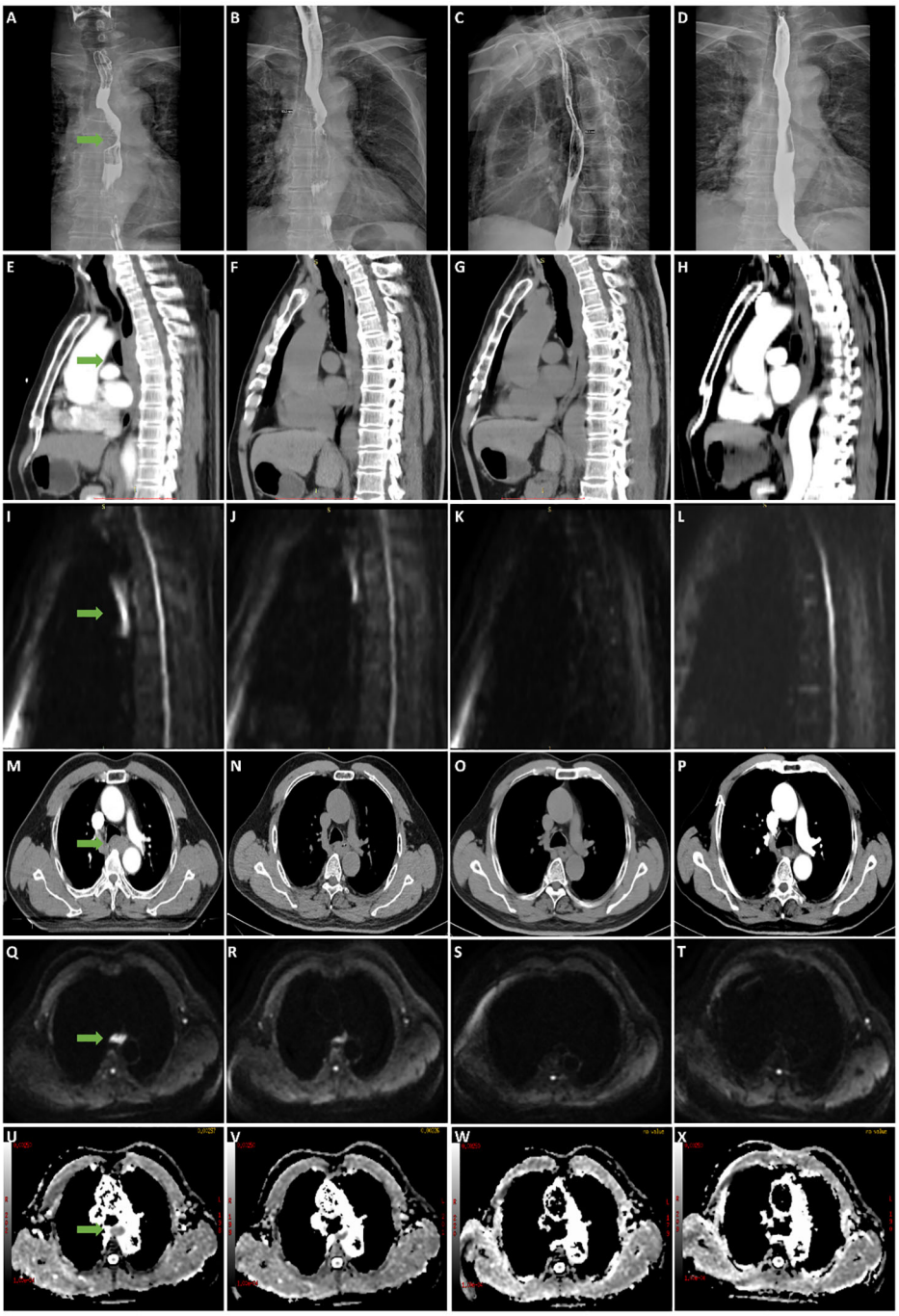


FIGURE 3
A 63-year-old male patient with stage III esophageal cancer in survival group was monitored for treatment response using esophagography, CT, and DWI at various stages: before CRT, and at the end of the 2nd, 4th, and 6th weeks of treatment. **(A–D)** Tumor in esophagography images. **(E–H)** Tumor in CT images (median sagittal section). **(I–L)** Tumor in DWI images (median sagittal section). **(M–P)** Tumor in CT images (transverse section). **(Q–T)** Tumor in DWI images (transverse section). **(U–X)** Tumor in ADC images (transverse section).

DWI can afford valuable markers to predict treatment response, as well as the survival of patients with ESCC, and the sustained high signal expression on DWI is a risk factor (33). However, high signal expression judgment is relatively subjective. Physicians with different levels of experience and qualifications may define high, slightly higher, and no signals differently. Hence, this study aimed to find a more objective, more straightforward, and more sensitive imaging examination method to judge the prognosis of ESCC and provide valuable reference information for clinical use.

Here, both the survival and the non-survival groups had similar trends in tumor length measured on esophagography before CRT and at the end of the second week, fourth week, and sixth week of treatment: a monotonically decreasing trend. However, there was no significant difference in tumor length measured on

10.3389/fonc.2024.1397266

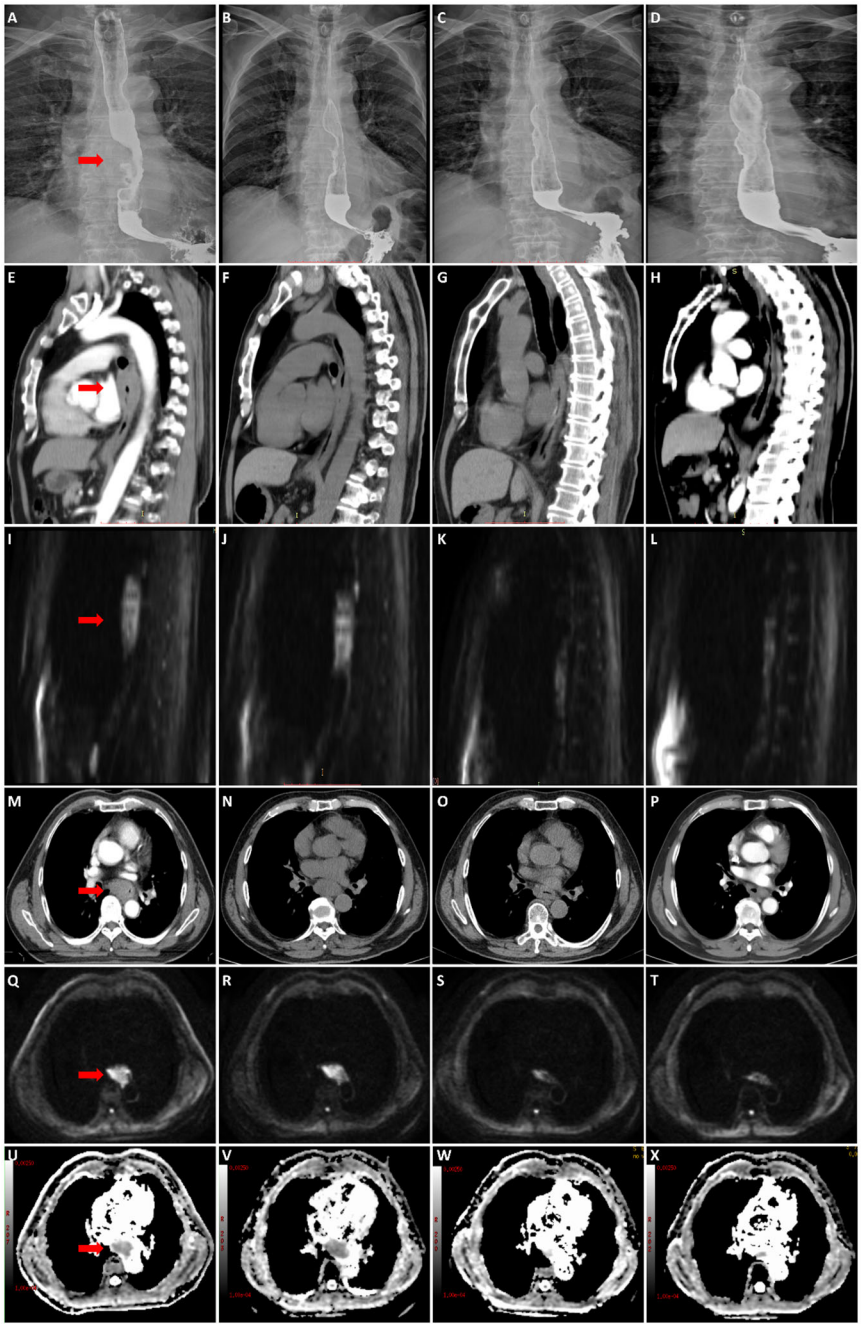


FIGURE 4
A 57-year-old male patient with stage III esophageal cancer in non-survival group was monitored for treatment response using esophagography, CT, and DWI at various stages: before CRT, and at the end of the 2nd, 4th, and 6th weeks of treatment. **(A–D)** Tumor in esophagography images. **(E–H)** Tumor in CT images (median sagittal section). **(I–L)** Tumor in DWI images (median sagittal section). **(M–P)** Tumor in CT images (transverse section). **(Q–T)** Tumor in DWI images (transverse section). **(U–X)** Tumor in ADC images (transverse section).

esophagography between the two groups at the four-time points. The ROC curve showed that the P-values were greater than 0.05, indicating that the prognostic diagnostic efficiency of tumor length is very low compared to that measured on esophagography. Similar to the trend of tumor length measured on esophagography, the trends of tumor length measured on CT and the maximal esophageal wall thickness measured on CT in the survival group were similar to those in the non-survival group, and the ROC curves showed that the P-values were greater than 0.05, indicating that according to tumor length measured on CT and the maximal esophageal wall thickness measured on CT, their prognostic diagnostic efficiencies were very low. Therefore, traditional

morphological imaging techniques are unreliable in the early evaluation of tumor response to CRT (34, 35), making it unreasonable to determine the downgrade or upgrade of treatment.

Tumor length measured on DWI at different points showed a sustained rapid declining trend, which was more pronounced after four weeks of treatment in the survival group, and a slow downward trend was observed in the non-survival group. Three patients did not have a high signal expression on DW images at the end of the sixth week of treatment, all of whom achieved long-term survival, and all the patients in the non-survival group had high signal expression at the end of radiotherapy. Tumor length measured on DWI in the survival group was significantly lower than that in the non-survival group at the end of the sixth week of treatment (P=0.001). The ROC curve showed that according to tumor length measured on DWI at the end of the sixth week of treatment as the diagnostic index (2.995 cm), the area under the curve was 0.840. The diagnostic efficiency was accurate, with a sensitivity of 0.889 and a specificity of 0.667. The whole group was divided into two groups according to tumor length measured on DWI at the end of the sixth week of treatment as the cutoff value (2.995 cm): 10 patients had a tumor length ≤2.995 cm, 20 patients had a tumor length >2.995 cm, and the 1- and 3-year survival rates in both subgroups were 90.0% and 80.0% and 75.0% and 25.0% (P=0.003), which indicated that tumor length measured on DWI could effectively predict prognosis. This is similar to the conclusions of other studies (19).

The ADC is inversely correlated with tissue cellularity. Cytotoxic therapy affects the permeability and integrity of the tumor cell membrane. It induces apoptosis, necrosis, and dissolution, leading to changes in tissue density and water molecule dispersion, causing increased ADC values. The ADC has emerged as a potential biomarker of response to cancer therapy (36, 37). Many studies (38–40) have confirmed that the ADC value increases significantly after CRT for EC: compared with those who did not respond well, there was a significant increase in the ADC values after antitumor therapy in those who did respond well. This study showed that the ADC value of the whole group of patients also showed a gradual upward trend after treatment. However, with an increase in the frequency of radiotherapy, the ADC value showed a continuous upward trend in the survival group. In contrast, the rising trend decreased from the end of the fourth week of treatment in the non-survival group. There were significant differences between the two groups at the end of the fourth week and sixth week of treatment, and the ADC value of the survival group was significantly higher than that of the non-survival group (P<0.001). The ROC curve showed that with the ADCs at the end of the fourth week of treatment and the end of the sixth week of treatment as the diagnostic indexes (2.965×10-3 mm^2/s and 3.570×10-3 mm^2/s), the P values were 0.001 and <0.001, and the areas under the curve were 0.866 and 0.970, with sensitivities of 1 and 0.833 and specificities of 0.611 and 1. The diagnostic efficacies were accurate; therefore, the diagnostic indicator of the ADC at the end of the sixth week of treatment was better. With the ADC at the end of the sixth week of treatment as the cutoff point to divide the whole group of patients into two groups (3.570×10-3 mm^2/s), the OS rate of the ADC>3.570×10-3 mm^2/s group was significantly better than that of the ADC≤ 3.570×10^{-3} mm^2/s group (P<0.001). Univariate and multivariate Cox regression analyses also indicated that the ADC at the end of the sixth week of treatment was a risk factor, similar to other reports (33). From the diagrams of the two patients, we can see the uniqueness and advantages of functional imaging technology, the maximum wall thickness of the patients in the survival group gradually decreased throughout the treatment. By the end of the treatment, no high signal was observed on the DW images, and the lesions were no longer visible. In contrast, patients in the non-survival group showed a reduction in lesions on esophageal esophagography by the end of the treatment, with a significant decrease in maximum esophageal wall thickness and a shorter tumor length compared to before CRT. However, despite these reductions, the lesions still exhibited high signals on DW images at the end of the treatment. similarly, Alicia S Borggrevet (11) reported that early changes on 18F-FDG PET/CT and DW-MRI during nCRT could help identify EC patients who could achieve pathologic complete response. However, these changes cannot be observed with morphological imaging technology, the integration of DW-MRI into clinical practice for the management of esophageal cancer can provide significant advantages over traditional imaging methods. It offers superior prognostic value, enabling more precise and early predictions of survival outcomes, thus facilitating optimized therapeutic strategies. Moreover, the non-invasive nature of DWI makes it a favorable alternative to more invasive diagnostic procedures, thereby reducing patient discomfort and associated risks.

The results of this study showed that when tumor length measured on DWI at the end of the sixth week of treatment and the ADC values measured at the end of the fourth week and sixth week of treatment were used as diagnostic indicators, the prognosis can be effectively judged. Their diagnostic efficacies are better than those of morphological examination methods. Moreover, DWI is a simple, reliable, convenient, and sensitive prediction method that can be used to judge prognosis effectively. The strength of this study is that it allows for an intuitive comparison of the advantages and disadvantages of the three imaging methods by using dense time points. However, this study has some limitations. The first is the number of cases in our study was relatively small, which may affect the universality of the results, and differences in scanning equipment and parameter settings may lead to data variability. Therefore, larger scale, multicenter studies are needed in the future to validate our findings and further clarify the role of DWI/ADC values in the prognosis evaluation of esophageal cancer radiotherapy. Additionally, due to the limitations of scanning conditions at that time, we did not collect data that could be used for advanced diffusion models, including IVIM and DKI, have to some extent corrected potential flaws in DWI, such as susceptibility to cellular or vascular system effects (41), or the influence of nonnormally distributed motion on ADC values (42). Many scholars have studied advanced diffusion models such as IVIM and have drawn meaningful conclusions (43–45). In the future, we will attempt to use advanced diffusion models such as IVIM and DKI to conduct more in-depth research in this field.

In summary, functional imaging technology can accurately reflect the actual treatment effect of tumors. As a relatively

economical and straightforward examination method, DW-MRI can allow the ADC to be measured directly on maps, reflecting the metabolic information of tissue cells, and can provide tumor-related information objectively and quantitatively, which is worthy of clinical application. Radiotherapists must be aware of the strengths and limitations of different imaging modalities in various clinical settings. If necessary, information from anatomic and functional imaging can be combined. A multimodality-based approach to imaging is essential in clinical practice to achieve the best possible outcome for patients with EC.

5 Conclusion

Compared with esophagography and CT, DW-MRI has certain advantages in predicting the prognosis of ESCC. ADC value as a non-invasive imaging biomarker, have the potential to predict the prognosis of esophageal cancer radiotherapy and provide valuable information for personalized treatment.

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. Requests to access these datasets should be directed to zhangandu@hebmu.edu.cn.

Ethics statement

The studies involving humans were approved by The Fourth Hospital of Hebei Medical. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YL: Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. XS: Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. YS: Data curation, Investigation, Methodology, Resources, Validation, Writing – original draft. HL: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. WW: Software, Methodology, Validation, Writing – original draft. AZ: Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing, Formal Analysis, Funding acquisition. GS: Conceptualization, Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by The Hebei Province Department of Science and Technology (Approval Number: 22377731D), Natural Science Foundation of Hebei Province (Approval Number: H2021206250) and Health Commission of Hebei Province (Approval Number:20210774).

Conflict of interest

Author WW was employed by Siemens Healthineers Digital Health Technology (Shanghai) Co., Ltd. Beijing Branch.

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.

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.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. Duan XF, Tang P, Yu ZT. Neoadjuvant chemoradiotherapy for resectable esophageal cancer: an in-depth study of randomized controlled trials and literature review. *Cancer Biol Med* (2014) 11(3):191–201. doi: 10.7497/j.issn.2095-3941.2014.03.005

3. Vlacich G, Samson PP, Perkins SM, Roach MC, Parikh PJ, Bradley JD, et al. Treatment utilization and outcomes in elderly patients with locally advanced esophageal carcinoma: a review of the national cancer database. *Cancer Med* (2017) 6(12):2886–96. doi: 10.1002/cam4.1250

4. Hirohata R, Hamai Y, Emi M, Kurokawa T, Yoshikawa T, Ohsawa M, et al. Prediction of the tumor response and survival based on computed tomography in esophageal squamous cell carcinoma after trimodality therapy. *Surg Today* (2021) 51(9):1496–505. doi: 10.1007/s00595-021-02277-7

5. Tsurumaru D, Hiraka K, Komori M, Shioyama Y, Morita M, Honda H. Role of barium esophagography in patients with locally advanced esophageal cancer: evaluation of response to neoadjuvant chemoradiotherapy. *Radiol Res Pract* (2013) 2013:502690. doi: 10.1155/2013/502690

6. Debi U, Sharma M, Singh L, Sinha A. Barium esophagogram in various esophageal diseases: A pictorial essay. *Indian J Radiol Imaging* (2019) 29(2):141–54. doi: 10.4103/ijri.IJRI_465_18

7. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026

8. de Gouw D, Klarenbeek BR, Driessen M, Bouwense S, van Workum F, Fütterer JJ, et al. Detecting pathological complete response in esophageal cancer after neoadjuvant therapy based on imaging techniques: A diagnostic systematic review and meta-analysis. *J Thorac Oncol* (2019) 14(7):1156–71. doi: 10.1016/j.jtho.2019.04.004

9. Maffazzioli L, Zilio MB, Klamt AL, Duarte JA, Mazzini GS, Campos VJ, et al. ADC as a predictor of pathologic response to neoadjuvant therapy in esophageal cancer: a systematic review and meta-analysis. *Eur Radiol* (2020) 30(7):3934–42. doi: 10.1007/s00330-020-06723-x

10. Hu X, Zhou T, Ren J, Duan J, Wu H, Liu X, et al. Response prediction using (18) F-FAPI-04 PET/CT in patients with esophageal squamous cell carcinoma treated with concurrent chemoradiotherapy. *J Nucl Med* (2023) 64(4):625–31. doi: 10.2967/jnumed.122.264638

11. Borggreve AS, Goense L, van Rossum P, Heethuis SE, van Hillegersberg R, Lagendijk J, et al. Preoperative prediction of pathologic response to neoadjuvant chemoradiotherapy in patients with esophageal cancer using (18)F-FDG PET/CT and DW-MRI: A prospective multicenter study. *Int J Radiat Oncol Biol Phys* (2020) 106(5):998–1009. doi: 10.1016/j.ijrobp.2019.12.038

12. Liu C, Sun R, Wang J, Ning F, Wang Z, Luo J, et al. Combination of DCE-MRI and DWI in predicting the treatment effect of concurrent chemoradiotherapy in esophageal carcinoma. *BioMed Res Int* (2020) 2020:2576563. doi: 10.1155/2020/2576563

13. Bozgeyik Z, Onur MR, Poyraz AK. The role of diffusion weighted magnetic resonance imaging in oncologic settings. *Quant Imaging Med Surg* (2013) 3(5):269–78. doi: 10.3978/j.issn.2223-4292.2013.10.07

14. Jafar MM, Parsai A, Miquel ME. Diffusion-weighted magnetic resonance imaging in cancer: Reported apparent diffusion coefficients, in-vitro and in-vivo reproducibility. *World J Radiol* (2016) 8(1):21–49. doi: 10.4329/wjr.v8.i1.21

15. Hirata A, Hayano K, Ohira G, Imanishi S, Hanaoka T, Murakami K, et al. Volumetric histogram analysis of apparent diffusion coefficient for predicting pathological complete response and survival in esophageal cancer patients treated with chemoradiotherapy. *Am J Surg* (2020) 219(6):1024–9. doi: 10.1016/j.amjsurg.2019.07.040

16. An D, Cao Q, Su N, Li W, Li Z, Liu Y, et al. Response prediction to concurrent chemoradiotherapy in esophageal squamous cell carcinoma using delta-radiomics based on sequential whole-tumor ADC map. *Front Oncol* (2022) 12:787489. doi: 10.3389/fonc.2022.787489

17. van Overhagen H, Laméris JS, Berger MY, van der Voorde F, Tilanus HW, Klooswijk AI, et al. Supraclavicular lymph node metastases in carcinoma of the esophagus and gastroesophageal junction: assessment with CT, US, and US-guided fine-needle aspiration biopsy. *Radiology* (1991) 179(1):155–8. doi: 10.1148/radiology.179.1.2006268

18. Mizowaki T, Nishimura Y, Shimada Y, Nakano Y, Imamura M, Konishi J, et al. Optimal size criteria of malignant lymph nodes in the treatment planning of radiotherapy for esophageal cancer: evaluation by computed tomography and magnetic resonance imaging. *Int J Radiat Oncol Biol Phys* (1996) 36(5):1091–8. doi: 10.1016/s0360-3016(96)00425-7

19. Wang L, Liu L, Han C, Liu S, Tian H, Li Z, et al. The diffusion-weighted magnetic resonance imaging (DWI) predicts the early response of esophageal squamous cell carcinoma to concurrent chemoradiotherapy. *Radiother Oncol* (2016) 121(2):246–51. doi: 10.1016/j.radonc.2016.10.021

20. Liu G, Yang Z, Li T, Yang L, Zheng X, Cai L. Optimization of b-values in diffusion-weighted imaging for esophageal cancer: Measuring the longitudinal length of gross tumor volume and evaluating chemoradiotherapeutic efficacy. *J Cancer Res Ther* (2017) 13(5):748–55. doi: 10.4103/jcrt.JCRT_630_17

21. Li Z, Han C, Wang L, Zhu J, Yin Y, Li B. Prognostic value of texture analysis based on pretreatment DWI-weighted MRI for esophageal squamous cell carcinoma patients treated with concurrent chemo-radiotherapy. *Front Oncol* (2019) 9:1057. doi: 10.3389/fonc.2019.01057

22. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol* (2018) 63 (7):07TR01. doi: 10.1088/1361-6560/aab4b1

23. Swets JA. Measuring the accuracy of diagnostic systems. *Science* (1988) 240 (4857):1285–93. doi: 10.1126/science.3287615

24. Wu C, Li M, Meng H, Liu Y, Niu W, Zhou Y, et al. Analysis of status and countermeasures of cancer incidence and mortality in china. *Sci China Life Sci* (2019) 62(5):640–7. doi: 10.1007/s11427-018-9461-5

25. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in china: good or bad news from the 2018 global cancer statistics. *Cancer Commun (Lond)* (2019) 39 (1):22. doi: 10.1186/s40880-019-0368-6

26. Mao Y, Fu Z, Zhang Y, Dong L, Zhang Y, Zhang Q, et al. A six-microRNA risk score model predicts prognosis in esophageal squamous cell carcinoma. *J Cell Physiol* (2019) 234(5):6810–9. doi: 10.1002/jcp.27429

27. Mao Y, Fu Z, Zhang Y, Dong L, Zhang Y, Zhang Q, et al. A seven-lncRNA signature predicts overall survival in esophageal squamous cell carcinoma. *Sci Rep* (2018) 8(1):8823. doi: 10.1038/s41598-018-27307-2

28. van Rossum P, Goense L, Meziani J, Reitsma JB, Siersema PD, Vleggaar FP, et al. Endoscopic biopsy and EUS for the detection of pathologic complete response after neoadjuvant chemoradiotherapy in esophageal cancer: a systematic review and meta-analysis. *Gastrointest Endosc* (2016) 83(5):866–79. doi: 10.1016/j.gie.2015.11.026

29. Yip C, Cook GJ, Landau DB, Davies A, Goh V. Performance of different imaging modalities in assessment of response to neoadjuvant therapy in primary esophageal cancer. *Dis Esophagus* (2016) 29(2):116–30. doi: 10.1111/dote.12315

30. Griffin Y. Esophageal cancer: Role of imaging in primary staging and response assessment post neoadjuvant therapy. *Semin Ultrasound CT MR* (2016) 37(4):339–51. doi: 10.1053/j.sult.2016.02.003

31. van Rossum PS, van Lier AL, van Vulpen M, Reerink O, Lagendijk JJ, Lin SH, et al. Diffusion-weighted magnetic resonance imaging for the prediction of pathologic response to neoadjuvant chemoradiotherapy in esophageal cancer. *Radiother Oncol* (2015) 115(2):163–70. doi: 10.1016/j.radonc.2015.04.027

32. Fang P, Musall BC, Son JB, Moreno AC, Hobbs BP, Carter BW, et al. Multimodal imaging of pathologic response to chemoradiation in esophageal cancer. *Int J Radiat Oncol Biol Phys* (2018) 102(4):996–1001. doi: 10.1016/j.ijrobp.2018.02.029

33. Wang L, Han C, Zhu S, Shi G, Wang Q, Tian H, et al. Investigation of using diffusion-weighted magnetic resonance imaging to evaluate the therapeutic effect of esophageal carcinoma treatment. *Oncol Res Treat* (2014) 37(3):112–6. doi: 10.1159/000360210

34. Westerterp M, van Westreenen HL, Reitsma JB, Hoekstra OS, Stoker J, Fockens P, et al. Esophageal cancer: CT, endoscopic US, and FDG PET for assessment of response to neoadjuvant therapy–systematic review. *Radiology* (2005) 236(3):841–51. doi: 10.1148/radiol.2363041042

35. Schneider PM, Metzger R, Schaefer H, Baumgarten F, Vallbohmer D, Brabender J, et al. Response evaluation by endoscopy, rebiopsy, and endoscopic ultrasound does not accurately predict histopathologic regression after neoadjuvant chemoradiation for esophageal cancer. *Ann Surg* (2008) 248(6):902–8. doi: 10.1097/SLA.0b013e31818f3afb

36. Decker G, Mürtz P, Gieseke J, Träber F, Block W, Sprinkart AM, et al. Intensity-modulated radiotherapy of the prostate: dynamic ADC monitoring by DWI at 3.0 t. *Radiother Oncol* (2014) 113(1):115–20. doi: 10.1016/j.radonc.2014.07.016

37. Scialpi M, Martorana E, Scialpi P, D'Andrea A, Mancioli FM, Mignogna M, et al. MRI apparent diffusion coefficient (ADC): A biomarker for prostate cancer after radiation therapy. *Turk J Urol* (2021) 47(6):448–51. doi: 10.5152/tud.2021.21274

38. Zheng H, Ren W, Pan X, Zhang Q, Liu B, Liu S, et al. Role of intravoxel incoherent motion MRI in early assessment of the response of esophageal squamous cell carcinoma to chemoradiotherapy: A pilot study. *J Magn Reson Imaging* (2018) 48 (2):349–58. doi: 10.1002/jmri.25934

39. Heethuis SE, Goense L, van Rossum P, Borggreve AS, Mook S, Voncken F, et al. DW-MRI and DCE-MRI are of complementary value in predicting pathologic response to neoadjuvant chemoradiotherapy for esophageal cancer. *Acta Oncol* (2018) 57 (9):1201–8. doi: 10.1080/0284186X.2018.1473637

40. Yuan Z, Niu XM, Liu XM, Fu HC, Xue TJ, Koo CW, et al. Use of diffusion-weighted magnetic resonance imaging (DW-MRI) to predict early response to anti-tumor therapy in advanced non-small cell lung cancer (NSCLC): a comparison of intravoxel incoherent motion-derived parameters and apparent diffusion coefficient. *Transl Lung Cancer Res* (2021) 10(8):3671–81. doi: 10.21037/tlcr-21-610

41. Dyvorne HA, Galea N, Nevers T, Fiel MI, Carpenter D, Wong E, et al. Diffusion-weighted imaging of the liver with multiple b values: effect of diffusion gradient polarity and breathing acquisition on image quality and intravoxel incoherent motion parameters–a pilot study. *Radiology* (2013) 266(3):920–9. doi: 10.1148/radiol.12120686

42. Dijkstra H, Baron P, Kappert P, Oudkerk M, Sijens PE. Effects of microperfusion in hepatic diffusion weighted imaging. *Eur Radiol* (2012) 22(4):891–9. doi: 10.1007/s00330-011-2313-1

43. Song T, Yao Q, Qu J, Zhang H, Zhao Y, Qin J, et al. The value of intravoxel incoherent motion diffusion-weighted imaging in predicting the pathologic response to neoadjuvant chemotherapy in locally advanced esophageal squamous cell carcinoma. *Eur Radiol* (2021) 31(3):1391–400. doi: 10.1007/s00330-020-07248-z

44. Ding Y, Hazle JD, Mohamed AS, Frank SJ, Hobbs BP, Colen RR, et al. Intravoxel incoherent motion imaging kinetics during chemoradiotherapy for human papillomavirus-associated squamous cell carcinoma of the oropharynx: preliminary results from a prospective pilot study. *NMR BioMed* (2015) 28(12):1645–54. doi: 10.1002/nbm.3412

45. Liu N, Yang X, Lei L, Pan K, Liu Q, Huang X. Intravoxel incoherent motion model in differentiating the pathological grades of esophageal carcinoma: Comparison of mono-exponential and bi-exponential fit model. *Front Oncol* (2021) 11:625891. doi: 10.3389/fonc.2021.625891

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 July 2024
DOI 10.3389/fonc.2024.1437598

Check for updates

OPEN ACCESS

EDITED BY
Harrison Kim,
University of Alabama at Birmingham, United States

REVIEWED BY
Brian Swendseid,
Cooper Health System, United States
Jochen Weitz,
MKG Augsburg, Germany

*CORRESPONDENCE
Xiaoyue Wu
18996072620@163.com
Ting Wang
cacy.wang@163.com

†These authors have contributed equally to this work

RECEIVED 24 May 2024
ACCEPTED 08 July 2024
PUBLISHED 19 July 2024

CITATION
Liu Y, Wu J, Liu D, Xiang D, Wu X and Wang T (2024) The application of mixed reality technique in oromaxillo-facial reconstruction with the perforator flap for malignant tumor patients. *Front. Oncol.* 14:1437598. doi: 10.3389/fonc.2024.1437598

COPYRIGHT
© 2024 Liu, Wu, Liu, Xiang, Wu 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.

The application of mixed reality technique in oromaxillo-facial reconstruction with the perforator flap for malignant tumor patients

Yixiu Liu[1†], Jian Wu[1†], Daide Liu[2], Dalan Xiang[2], Xiaoyue Wu[1*] and Ting Wang[3*]

[1]Head and Neck Surgery, Chongqing University Cancer Hospital, Chongqing, China, [2]Department of Surgery, People's Hospital of Shizhu, Chongqing, China, [3]Internal Medicine-Oncology, Chongqing University Cancer Hospital, Chongqing, China

Objectives: The integration of quantitative imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) with mixed reality (MR) technology holds promise for enhancing the diagnosis, prognosis, and treatment monitoring of cancer. This study compares the characteristics and effects of MR and color Doppler ultrasound (CDU) in the localization of perforator blood vessels in the lower extremities.

Methods: Two techniques were used to locate the perforator vessels in 40 cases of maxillofacial defect repair using perforator flaps from the lower extremities. The number of perforator vessels located in the flap area and the actual number of perforator vessels explored during the surgery were recorded. The recognition rate was calculated and the operation time and blood loss were recorded for each case.

Results: The recognition rates of MR technology and CDU in perforating vessels of the lower limbs were 93.9% and 97.2%, respectively ($p > 0.05$). The operation time was 52-74 minutes, 65-88 minutes ($p > 0.05$). The average bleeding volumes were 24 and 56 ml ($p < 0.05$), respectively. All perforator flaps were alive. One flap had a crisis and recovered after emergency exploratory treatment. Thirty donor sites of the lower extremities were directly sutured, and wounds were closed by abdominal skin grafting in 10 cases.

Conclusion: MR technology for successfully identifying perforator vessels can shorten the operation time, reduce the amount of bleeding in the donor site, and reduce trauma to the donor site.

KEYWORDS

mixed reality, computed tomography angiography, perforator, oromaxillo-facial reconstruction, malignant tumors, cancer care

1 Introduction

Koshima and Soeda introduced the concept of the perforator flap in 1989, which differs from the conventional flap, which requires dissection from the main vessel to the distal branch vessels (1). Instead, the perforator flap utilizes retrodissection from the peripheral perforator vessels to the source vessel. By identifying the perforator vessels in the designated area, the flap can be harvested via dissection of the perforator vessel (2). This approach allows for successful outcomes even if the surgeon is not entirely familiar with the tissue structure in that region. Furthermore, the perforator flap method offers several advantages, including flexibility, ease of transfer, minimal damage to donor sites, and clinical reliability and efficacy (3–6). As a result, the perforator flap has become the preferred choice for repairing soft tissue defects and reconstructing the jaws in head, neck, and maxillofacial surgeries. However, the accuracy and predictability of clinical applications are limited by differences in the location, origin, and course of perforator vessels among individuals, which can affect the final repair outcome. Therefore, accurate identification of perforator vessels and appropriate flap design are crucial clinical considerations that must be addressed (7, 8).

The mixed reality (MR) technique is lately introduced that leverages computed tomography (CT) and magnetic resonance imaging (MRI) data in a software workstation to generate a 3D model, allowing it to be downloaded to a head-mounted holographic display (9). In clinical settings, holographic displays are used by operators to project the model onto a patient's body surface and subsequently match it to the relevant organ based on locators or fixed anatomical markers embedded in the tissue. The positioning error between the virtual model and patient is less than 1 mm (10), with negligible discrepancies of 2-3mm observed in the process of surgical anatomy (11). In this particular investigation, computed tomography angiography (CTA) data were used to construct a 3D model, and MR technology was used to facilitate real-time location of the lower limb perforator vessels during surgical procedures. In this study, we compared this new technique with color Doppler ultrasound (CDU) and evaluated its effects on the preparation and harvest of perforator flaps. This study introduces a novel and efficient approach for conducting lower limb perforator flap surgery, offering significant support for the application of mixed reality technology in the medical field. Our findings provide important insights for enhancing surgical outcomes, mitigating intraoperative risks for patients, and elevating overall medical quality. These discoveries hold promising implications for advancing the utilization of mixed reality technology in medicine.

2 Materials and methods

2.1 Clinical data

Our study included 40 patients (29 males and 11 females) aged between 32 and 75 years of age (mean age, 57.3 years) who underwent harvesting perforator flaps in the lower extremities at the Chongqing University Cancer Hospital between January 2019 and September 2020. These patients were randomly assigned to two groups of 20 patients each: the experimental group, which used MR technology with CTA data to locate perforator vessels, and the control group, which used CDU. All perforator flap surgeries were performed by the same surgeon. The inclusion criteria were as follows: (A) patients with malignant tumors in the oral and maxillofacial regions and clinically diagnosed at the middle or advanced stage; and (B) patients who were unable to undergo forearm free-flap reconstruction due to extensive tissue defects resulting from tumor excision. The exclusion criteria were as follows: (A) patients who had small primary lesion defects that could be sutured or repaired with a free-flap forearm reconstruction; (B) patients who had previous surgeries or trauma in both lower limbs; (C) patients who were unable to accept general anesthesia and surgeries for various reasons.

2.2 Lower limb CTA examinations of the experimental group

The scan was performed using a dual-source CT system (Siemens Somatom Drive; Siemens, Erlangen, Germany). Patients were scanned in a supine position with their feet first and their arms placed higher than the head in the same position as the surgery. The iodinated contrast agent iopromide 370 was injected at 4.0 ml/s for a total volume of approximately 90 ml. The arterial phase was triggered by monitoring the femoral artery with a trigger threshold of 100 HU, and the delay phase was 10 s. After the scanning was completed, the original image data of 1 mm layer thickness without horizontal axis spacing were transmitted to the workstation, and the image was post-processed by multi-plane reconstruction, maximum density projection, volume reconstruction, surface reconstruction, and other technologies.

2.3 3D reconstruction and import of experimental CTA data

We imported the obtained CTA digital imaging and communications in medicine (DICOM) data into a software workstation for staging and reconstruction of the original data. The selected region was used to determine the threshold value and generate the masking-out of the reconstructed area to construct a 3D model. A preliminary 3D model was then established and further edited by smoothing the soft tissue, bone tissue, and blood vessels. The resulting model was optimized by refining the triangular surfaces, removing any protrusions, and hollowing out blood vessels or cavity organs. The boundary contour was adjusted, and a 3D model was prepared. The 3D reconstruction model was imported into a head-mounted holographic display via the internet. The production and import of all the data were completed by the same technician.

2.4 The CDU examination and vascular localization towards the lower extremities of control group

Ultrasonic inspection was performed using a Philips Blood Vessel Detector (Philips, Best, Netherlands) at a 7.5 MHz frequency. The patients were instructed to adopt the corresponding posture according to the operator. In addition to ultrasonic measurement of perforator vessels in the donor sites of the lower limbs, the course, diameter, and blood flow dynamics of the perforator vessels were recorded. Marks were also made on the body surface projections of the perforator vessels. Both measurements and CDU marks were performed by the same medical technician.

2.5 Surgical procedures

The superior thigh of the perforator artery was selected as the donor site. In the experimental group, the flaps were designed with the perforating point of the perforator vessel as the center point, as determined by a head-mounted holographic display according to the scope, shape, and size of the defect area. In the control group, the flaps were designed with the perforating point of the vessel as the center point, as determined by CDU prior to surgery. After dissection of the skin and subcutaneous adipose tissue, the perforator vessels below the surface of the fascia lata were identified directly from the front to the rear. The perforator vessels were well preserved after dissection. After preparation of the tissue flaps was completed, the pedicles were dissected, and blood vessels were transected and ligated near the proximal end for anastomosis with the blood vessels at the primary lesion site. The defect in the donor area of the lower-extremity epidermis was closed or sutured with a skin graft.

2.6 Recording of measurement indicators

In this section, we validate the accuracy of the two techniques for localizing perforator vessels applied in the experimental and control groups. In addition, the number of localized perforator vessels and the actual perforator vessels within the flap area were recorded, and the recognition rate was calculated (recognition rate = number of located vessels/number of actual perforator vessels). The number of cases of direct closure suture or skin graft suture in the experimental and control groups, the time of completion of each operation, and blood loss volume in each group were also recorded.

2.7 Statistical analysis

The obtained data were statistically processed by means of SPSS statistical software.

3 Results

3.1 Lower limb perforator flap surgery outcomes and comparative analysis of MR and CDU techniques

Free perforator flaps from the lower limbs were successfully obtained in all 40 patients who underwent surgery, and all the flaps survived after surgery. One patient in the experimental group needed urgent re-exploration and salvage of the flap 4 h after surgery and succeeded in treating flap complications. Among the 20 patients in the experimental group, 18 underwent anterolateral thigh flap reconstruction, 2 underwent fibula osteocutaneous flap repair, 16 underwent direct suture closure in the lower limbs, and 4 underwent skin graft suturing. The operation time ranged from to 52-74 min, and the average blood loss volume was 24 ml. Among the 20 patients in the control group, anterolateral thigh flaps for reconstruction were adopted in 17 cases, fibula osteocutaneous flap repair was used in three cases, direct suture closure in the lower limbs in 14 cases, and skin graft in six cases. The operation time ranged from to 65-88 min, and the average blood loss was 56 ml. The identification rates of MR and CDU in perforator vessels of the lower limbs were 93.9% (31/33) and 97.2% (35/36), respectively. Statistical analyses showed that there was no significant difference in operation time between the two groups ($p > 0.05$). The recognition rate of MR technology is slightly lower than that of CDU, but there is no significant difference between the two groups ($p > 0.05$) (details are presented in Table 1).

3.2 Case presentation

A 56-year-old female underwent surgery for ameloblastoma in the left mandible 2 years previously. Specialist examination revealed that the patient had swelling on the left face and neoplasm in the area surrounding the left mandibular angle and ramus, approximately 5.0 × 4.0 cm measuring in size. The patient also had a mild limitation in mouth opening, and 36, 37, and 38 were not detected in the oral cavity. According to the patient's previous medical records and consultation report from the pathology department of our hospital, the pathological diagnosis of the patient was confirmed as ameloblastoma (Figure 1). The patient was diagnosed with ameloblastoma upon admission and scheduled to undergo partial mandibular resection and fibular osteocutaneous flap repair. She underwent CTA examination of the maxillofacial and the left calf prior to surgery. The obtained CTA data were imported into a software workstation for 3D reconstruction and subsequently downloaded onto a holographic display (Figures 2, 3). During the operation, the left calf was automatically positioned according to the holographic display, and the perforator vessels were dissected in real-time (Figure 4). Fibular osteocutaneous flaps were prepared to repair the mandible, and direct sutures were placed at the donor site (Figures 5, 6). After the operation, the flap fully survived, and

TABLE 1 Flap details of the experimental and control groups.

	Number (Cases)	Age (Y/O.)	Repair method (Cases)	Donor site management (Cases)	Operation time (min)	Average blood loss (mL)	Recognition rate (%)
Experiment	20	32-68	Anterolateral thigh flaps (18)	Direct suture (16)	52-74	24	93.9
			Fibula osteocutaneous flaps (2)	Skin graft repair (4)			
Control	20	35-75	Anterolateral thigh flaps (17)	Direct suture (14)	65-88	56	97.2
			Fibula osteocutaneous flaps (3)	Skin graft repair (6)			

during the 6-month follow-up, the patient expressed satisfaction with the reconstructive outcome.

4 Discussion

Since the 1990s, various techniques have been established to evaluate perforator vessel characteristics. Of these methods, the application of CTA and CDU is the most extensive (12, 13), but their effectiveness remains controversial. Many scholars believe that CTA is the gold standard for locating blood vessels (14), as it can precisely identify the course of vessels in muscles (15, 16) and shorten the operation time (17, 18). However, some scholars (19) also believe that CDU is superior to CTA in terms of the radiation dose and localization of superficial blood vessels. Currently, there is no consensus regarding the effectiveness of these two methods reported in literature. However, a common problem in clinical practice is that neither CTA nor CDU can provide real-time intraoperative guidance for perforator vessel localization. The 2D images from CDU only capture the superficial portion of perforator vessels on the body surface, lacking information on the 3D structural course of the vascular bundles. 3D CTA images require surgeons to match the reconstructed images with the actual surgical area during the operation based on their experience, which is time-consuming (20). Meanwhile, leaving the surgical field for a surgeon to view CTA and CDU images on the screen poses potential risks, including failure to promptly detect bleeding at the operative site or dropped instruments (21, 22).

Contextually, an ideal surgical navigation system should possess the following characteristics (23): (A) excessive additional work should not be introduced into the surgical process; (B) excessive invasive procedures should be avoided; and (C) rapid, real-time localization should be achieved and maintained in a sterile manner. Despite being a research hotspot in academia, current solutions for surgical navigation systems cannot meet the aforementioned requirements and, consequently, hinder the clinical application of navigation matching.

MR technology is a novel technique that has gained popularity in recent years, and it offers multiple advantages for its applications in the medical field (10, 24, 25): (A) It provides surgeons with intuitive and real-time imaging information to view both deep and superficial anatomical structures, lowering the difficulty in identifying tissue structures; (B) MR technology can superimpose images in real time onto the patient's anatomical structure and guide surgery using virtual 3D visualization information, with great convenience in operation; (C) for tumor patients, MR technology

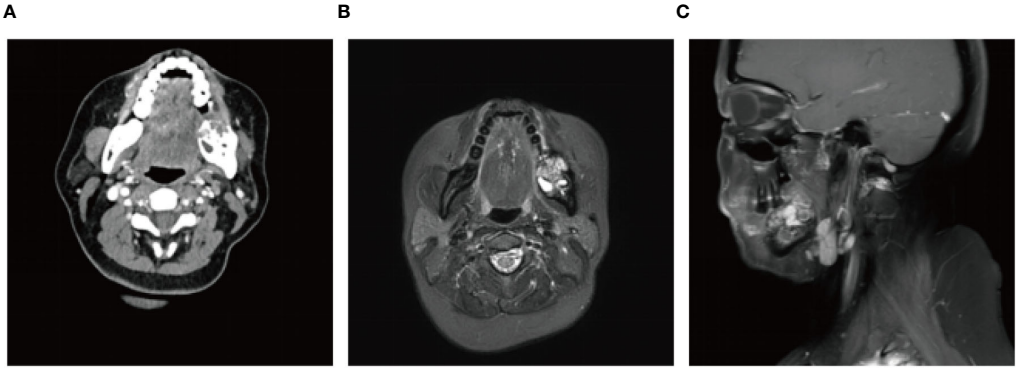


FIGURE 1
Preoperative maxillofacial imaging. **(A)** Cross-section plane of the preoperative computed tomography scan; **(B)** Cross-section plane of preoperative magnetic resonance; **(C)** Coronal plane of preoperative magnetic resonance.

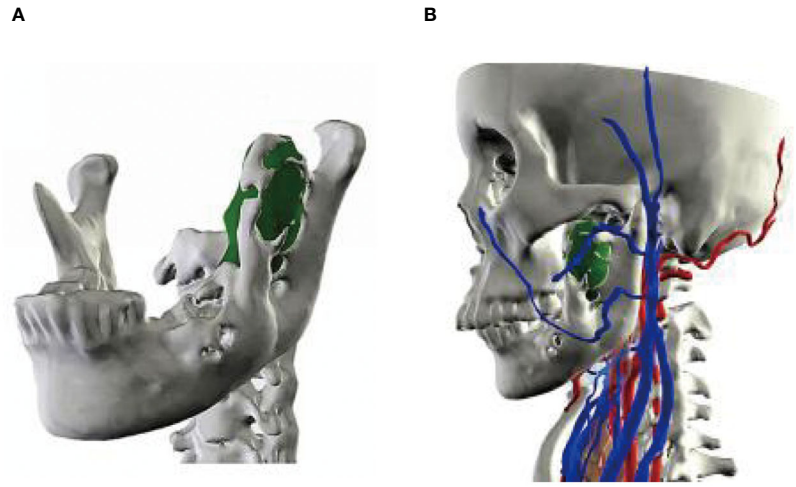


FIGURE 2
Maxillofacial 3D model. **(A, B)** Lateral views of the model.

can design tumor resection margins in advance, avoiding insufficient resection during surgery, which affects prognosis; (D) Through WIFI, MR technology transmits 3D models and matching images of patients to experts for remote communication and instruction during surgery; and (E) Compared with traditional positioning methods, surgeons can independently use HoloLens to project the 3D model of perforating vessels, touch and manipulate virtual objects, adjust the position, angle and scale of the 3D model, and overlap the 3D model on the human body to understand the shape and distribution of blood vessels and protect them during the flap preparation process, achieving aseptic operation.

After applying MR technology in the field of orthopedics, Lee et al. (26) demonstrated that displaying the anatomical structure of skin-covered areas aids surgeons in rapid localization during screw placement and reduces surgical risks, thus improving surgical accuracy. Shi et al. (27) were the first to apply MR technology in hepatectomy and achieved accurate matching between a 3D hologram model and the target organs. Thus, MR technology can be combined with other clinical treatment techniques. For example, when combined with a da Vinci robotic system, it can minimize trauma and surgical complications while achieving the goal of curing lesions (28, 29). Although MR technology has been applied in orthopedics, hepatobiliary surgery, and neurosurgery (30–32), there are relatively few reports on its application in the field of oral and maxillofacial surgery (33–35). Therefore, in the early stages of the study, the author first attempted to apply MR technology to clinical teaching and doctor-patient communication, which achieved good results and accumulated rich experience (Figure 7). In this study, we used MR technology to perform 3D

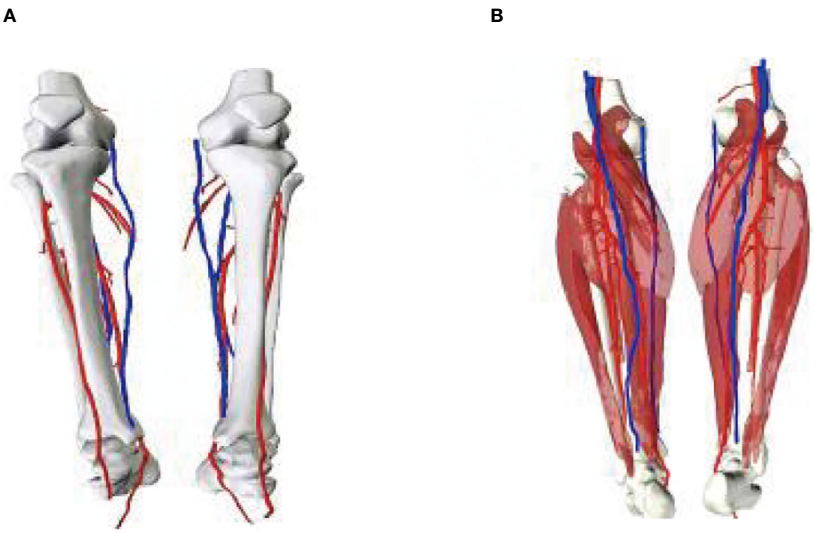


FIGURE 3
3D models of blood vessels in lower extremities. **(A)** Front; **(B)** Posterior views of the model.

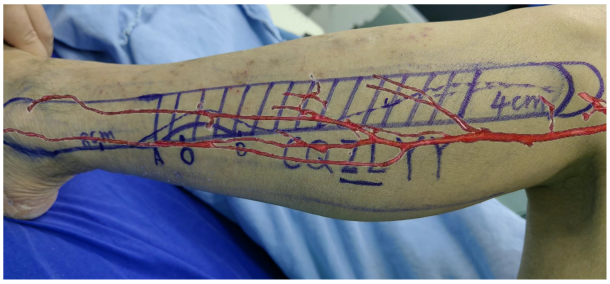


FIGURE 4
Intraoperative schematic diagram of mixed reality technology.

reconstruction of the maxillofacial region in 40 patients with malignant oral tumors. We found that the reconstructed 3D models were accurate and intuitive, and could be overlaid on the surgical site in real time with high precision. Compared to traditional imaging techniques, MR technology allows for easier understanding of organ anatomy, tumor shape, and location and expands the previously limited view of maxillofacial surgery in terms of depth and breadth, reducing judgment time and mental workload. In this study, the 3D reconstruction of the mandible in five patients was more distinct in terms of three-dimensional sense and boundary level compared to the reconstruction of soft tissues (Figure 8).

Recently, there has been an increasing interest among scholars in utilizing MR technology for the vascular localization of perforator flaps. This technology enables the observation of the origin, course, branching, and distribution of perforator vessels in 3D space and facilitates the reconstruction of precise 3D visualization models of perforator flap vessels. Bosc (36) employed MR technology to locate the perforator vessels in the lower abdomen for breast reconstruction. CTA data are

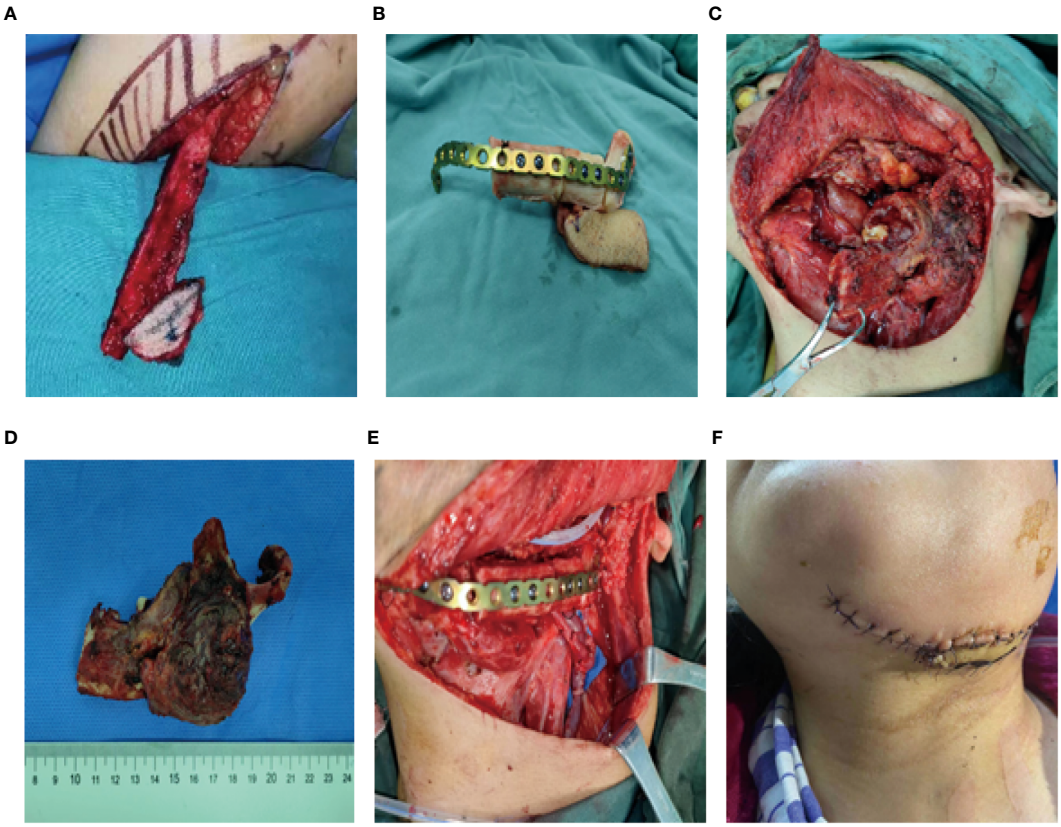


FIGURE 5
Repair and reconstruction of peroneal musculocutaneous flap. **(A–E)** Intraoperative reconstruction with fibula musculocutaneous flap; **(F)** Reconstruction outcome after the surgery.

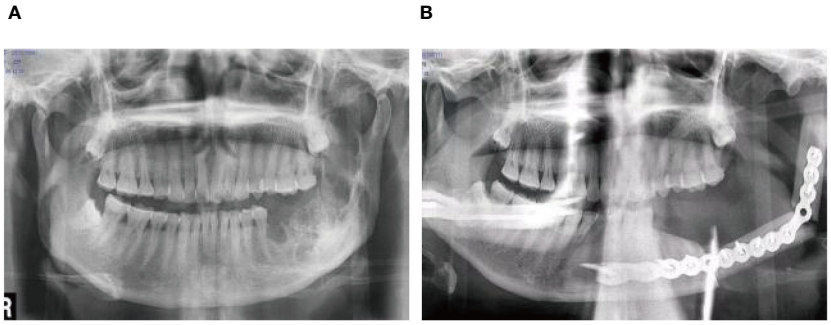


FIGURE 6
Repair and reconstruction of peroneal musculocutaneous flap. **(A)** Panoramic radiograph before treatment; **(B)** Panoramic radiograph after treatment.

reconstructed and imported into a holographic display, which allows the automatic matching of skin locators or anatomical structures during surgery. As a result, surgeons are able to perform a visual operation during the reconstruction process without having to direct their gaze towards a distant screen, available for a "fluoroscopic" view of the blood vessels. Pereira (37) utilized MR technology to anatomically locate blood vessels in the groin area of 60 patients and found that the positions of all vessels and lymph nodes corresponded to the actual operative location. MR technology can accurately locate the position of blood vessels and lymph nodes in the groin area, reducing flap harvesting time by 20% compared to traditional methods. In this study, we performed CTA examinations of the lower limbs and created a 3D model from the acquired data, which was then imported into a head-mounted holographic display. With the help of the display, the perforator vessels can be automatically located without relying on experience or spending extra time. Reverse dissection was performed on the basis of the course of the vessels. The time required for flap harvest was approximately 52-74 minutes, saving about 20% of the time (65-88 minutes) required by traditional methods. The results of this study are consistent with those of Pereira et al., but the difference was not statistically significant. The head-mounted display used was Microsoft HoloLens, which is comfortable and lightweight. However, long-time wearing of this device may cause a sense of dizziness, which may be related to the LED lighting and video sensor flicker of the head-mounted holographic display. This is consistent with researchers in other countries who believe that no fatigue or pain is associated with prolonged device use (24, 38).

In clinical practice in our hospital, ALT free flap surgery usually uses CDU for vascular localization, and CTA is only used in a small number of patients. However, although CDU is simple and low-cost for vascular localization, it has a large error. Therefore, the original intention of this study was to explore new ways to locate blood vessels. A large number of clinical studies have shown that the accuracy of CTA in vascular localization is higher than that of CDU (39), and it is known as the "gold standard" for vascular localization (40). Currently, there are no relevant literature reports on the clinical application of MR-superimposed CTA data for vascular identification. This study used MR technology to locate blood vessels based on CTA data. Although it increased costs and made the surgical process relatively complicated to a certain extent, it achieved three-dimensional visualization of perforating vessels, which can intuitively understand the shape and distribution of perforating vessels and the location of the exit point, which is conducive to preoperative flap design, intraoperative protection of perforating vessels and exit points, and reducing the possibility of flap crisis. Based on this, we believe that the use of MR based on CTA data in ALT has both

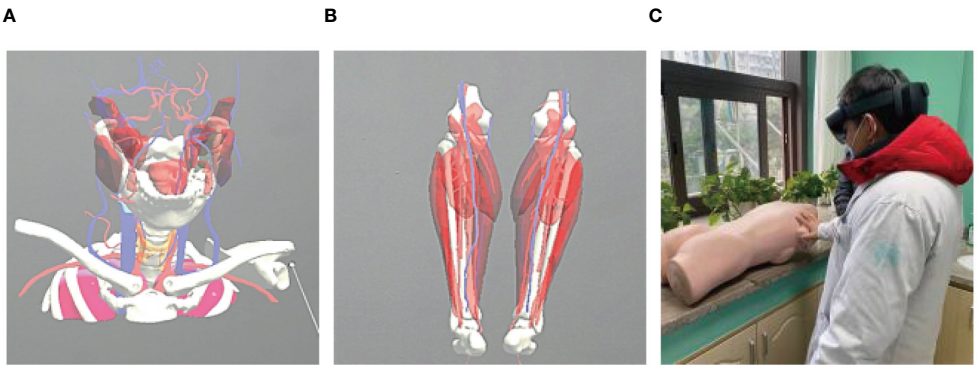


FIGURE 7
Application of mixed reality technology in teaching. **(A, B)** Video demonstration in teaching; **(C)** Teaching practice with holographic display.

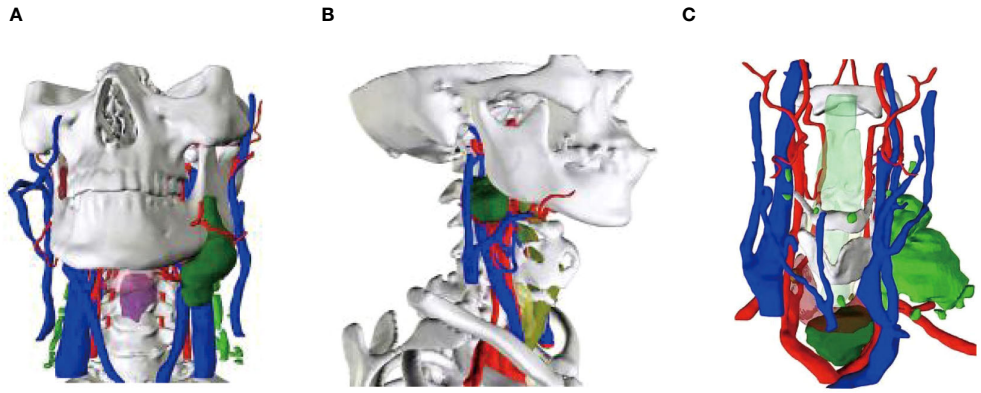


FIGURE 8
Mixed reality technology to rebuild the primary focus. **(A)** Buccal mucosa carcinoma reconstruction; **(B)** Oropharyngeal carcinoma reconstruction; **(C)** Thyroid malignancy reconstruction.

advantages and disadvantages. The advantage is that it can achieve three-dimensional visualization of perforating vessels, which is conducive to preoperative flap design, reduces the possibility of flap crisis, and reduces the actual flap production time. However, the disadvantage is that the process is relatively complicated and the cost is increased.

Moreover, this study validated the accuracy of MR technology and CDU in identifying lower limb vessels by anatomical dissection. The results showed that the identification rate of MR technology was 93.9%, which was slightly lower than that of the conventional CDU (97.2%). We believe this may be explained by the following reasons: (A) small sample size; (B) new technologies may have inherent positioning errors (MR errors plus CTA errors); (C) the curved contours of the lower limbs may cause vessel displacement relative to the skin, leading to measurement errors; and (D) variations in adipose tissue thickness among cases may also have contributed to errors. MR technology can provide complete perforator information in patients with a thick adipose layer. However, in areas where the adipose layer is thin, the display of the end of the perforating vessels may be unclear because of the influence of CTA data on the MR technology. Although MR technology had a lower identification rate than traditional positioning methods in this study, it can provide surgeons with 3D courses of perforating vessels in practical operations, achieving the effect of "fluoroscopic" and precise anatomy once identified. Meanwhile, the average blood loss during the dissection of perforator flaps was 24 ml using MR technology and 56 ml in the control group, and the difference between the two was statistically significant. Hence, we assumed that the successful identification of perforating vessels using MR technology may result in reduced flap harvesting time, decreased blood loss, fewer postoperative complications, and eventually, benefit patients.

There are many types of free flaps available to head and neck reconstructive surgeons (41), the most commonly used of which are the radial forearm flap (RFF) and the anterolateral thigh flap (ALT) (42). Since the 1980s, RFF and ALT have been widely recognized as versatile and reliable free perforator flaps (43, 44). With significant advances in microsurgical technology, the success rate of free flap transplantation has increased to more than 90% in most published case studies (45, 46). Las, D.E (47). counted 1530 free flaps in 1247 patients, and the incidence of partial and total flap necrosis was 5.5% and 4.4%, respectively. WZhou et al. (48) included 881 flap transplants of the head and neck. Only 26 of the 881 flaps failed (2.9%). In this study, age, diabetes, history of lateral neck surgery, donor site, selection of recipient vein, and postoperative anticoagulation were not related to the outcome of free flaps, which were mainly affected by preoperative radiotherapy. At the same time, Ranganath (49) conducted an electronic search using PubMed, EMBASE, and the Medline Database of Systematic Reviews (CDSR), including all papers published between 2000 and 2022, and combined the following keywords: (RFF), (ALT). The final meta-analysis included 16 studies (50–65) that evaluated flap success rates, with a success rate of 98.3% (460/468) in ALT patients and 97.3% (476/489) in RFF patients. In this study, one of the 40 perforator flaps had a flap crisis 24 hours after surgery. After timely inspection by the nurse and timely vascular exploration in the operating room by the doctor, the flap eventually survived. However, in the treatment of our hospital, there are still cases of flap necrosis, mainly in patients who underwent surgery after radiotherapy. The overall flap success rate is not much different from the success rate reported by the researchers. In the future work, we will further statistically analyze the data of the flap success rate in our hospital.

In summary, MR technology has the potential to be applied in the repair and reconstruction of oral and maxillofacial defects by using flaps. This technology has the capability to successfully identify perforator vessels, which results in reduced blood loss and shorter surgical time, and may emerge as a novel auxiliary tool for future microsurgery. Despite the promising results, several limitations should be acknowledged. Firstly, the sample size in our study was relatively small, which may limit the generalizability of our findings. Additionally, the retrospective nature of the study design and the lack of a randomized control trial may introduce bias and confounding variables. Moreover, the study focused exclusively on lower limb perforator flap surgery, and the applicability of MR technology in other surgical contexts remains to be explored. Future

research with larger sample sizes and prospective study designs is warranted to further validate the utility of MR technology in reconstructive surgery and to address these limitations.

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.

Ethics statement

This clinical trial has been conducted in accordance with the principles of the Declaration of Helsinki and has received ethical approval from the Ethics Committee of Chongqing University Cancer Hospital (No. CZLS2021177-A). All participants provided informed consent before participating in the study. The study design, procedures, risks, and benefits have been thoroughly explained to the participants. Participants have the right to withdraw from the study at any time without any consequences. Confidentiality of all participant information is strictly maintained throughout the study. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YL: Conceptualization, Investigation, Project administration, Visualization, Writing – original draft, Writing – review & editing. JW: Conceptualization, Funding acquisition, Investigation, Writing – original draft, Writing – review & editing. DL: Data curation, Methodology, Writing – original draft. DX: Software, Writing – original draft. XW: Supervision, Writing – review & editing. TW: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Chongqing Municipal Scientific Research Institutions Performance Incentive and Guidance Project (grant number cstc2021jxjl0064), the Key Laboratory Development Fund of Chongqing University Cancer Hospital (grant number cquchkfjj004) and the Project Supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (grant number KJQN202300124).

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.

References

1. Koshima I, Soeda S. Inferior epigastric artery skin flaps without rectus abdominis muscle. *Br J Plast Surg.* (1989) 42:645–8. doi: 10.1016/0007-1226(89)90075-1

2. Liu F, Gao Y, Xu B, Xiong S, Yi S, Sun J, et al. Peg10 amplification at 7q21. 3 potentiates large-cell transformation in cutaneous T-cell lymphoma. *Blood J Am Soc Hematol.* (2022) 139:554–71. doi: 10.1182/blood.2021012091

3. Abdelfattah U, Power HA, Song S, Min K, Suh HP, Hong JP. Algorithm for free perforator flap selection in lower extremity reconstruction based on 563 cases. *Plast Reconstr Surg.* (2019) 144:1202–13. doi: 10.1097/PRS.0000000000006167

4. D'Arpa S, Cordova A, Pignatti M, Moschella F. Freestyle pedicled perforator flaps: safety, prevention of complications, and management based on 85 consecutive cases. *Plast Reconstr Surg.* (2011) 128:892–906. doi: 10.1097/PRS.0b013e3182268c83

5. Lazzeri D, Huemer GM, Nicoli F, Larcher L, Dashti T, Grassetti L, et al. Indications, outcomes, and complications of pedicled propeller perforator flaps for upper body defects: A systematic review. *Arch Plast Surg.* (2013) 40:44–50. doi: 10.5999/aps.2013.40.1.44

6. Hallock GG. The propeller flap version of the adductor muscle perforator flap for coverage of ischial or trochanteric pressure sores. *Ann Plast Surg.* (2006) 56:540–2. doi: 10.1097/01.sap.0000210512.81988.2b

7. Tang JY, Qing LM, He JQ, Wu P,D, Zhou ZB, Liang JY, et al. Application of computer assisted technique to desgin polyfoliate perforator flappedicled on the descending branch of the circumflex femoral lateral artery. *Chin J Microsurg.* (2016) 39:123–6. doi: 10.3760/cma.j.issn.1001-2036.2016.02.006

8. Feng SQ, Xi WJ, Wang Y, Li H, Li K, Zhang YX. The comparison of color Doppler ultrasound and computed tomography angiography in the preoperative planning of lower extremity perforators. *Chin J Microsurg.* (2016) 39:26–32. doi: 10.3760/cma.j.issn.1001-2036.2016.01.008

9. Fick T, Meulstee JW, Köllen MH, Van Doormaal JAM, Van Doormaal TPC, Hoving EW. Comparing the influence of mixed reality, a 3d viewer, and mri on the spatial understanding of brain tumours. *Front Virtual Real.* (2023) 4:1214520. doi: 10.3389/frvir.2023.1214520

10. Bosc R, Fitoussi A, Hersant B, Dao TH, Meningaud JP. Intraoperative augmented reality with heads-up displays in maxillofacial surgery: A systematic review of the literature and a classification of relevant technologies. *Int J Oral Maxillofac Surg.* (2019) 48:132–9. doi: 10.1016/j.ijom.2018.09.010

11. Mavioso C, Araújo RJ, Oliveira HP, Anacleto JC, Vasconcelos MA, Pinto D, et al. Automatic detection of perforators for microsurgical reconstruction. *Breast (Edinburgh Scotland).* (2020) 50:19–24. doi: 10.1016/j.breast.2020.01.001

12. Fan XY, Xu YQ, Xu LJ, Dong KX, Su XX, Xu LQ, et al. Application of high frequency color ultrasound combined with contrastenhanced ultrasound in preoperative design and evaluation of perforator flap. *Chin J Microsurg.* (2013) 36:322–6. doi: 10.3760/cma.j.issn.1001-2036.2013.06.003

13. Hallock GG. Doppler sonography and color duplex imaging for planning a perforator flap. *Clin Plast Surg.* (2003) 30:347–57, v-vi. doi: 10.1016/S0094-1298(03)00036-1

14. He J, Cui H, Qing L, Wu P, Guliyeva G, Yu F, et al. Strategies for selecting perforator vessels for transverse and oblique diep flap in male pediatric patients: anatomical study and clinical applications. *Front Pediatr*. (2022) 10:978481. doi: 10.3389/fped.2022.978481

15. Peng W, Lü C, Zhou B, Song D, Li Z. [Application and prospect of preoperative computed tomographic angiography in deep inferior epigastric artery perforator flap for breast reconstruction]. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi*. (2020) 34:927–31. doi: 10.7507/1002-1892.201907017

16. Teunis T, Heerma van Voss MR, Kon M, van Maurik JF. Ct-angiography prior to diep flap breast reconstruction: A systematic review and meta-analysis. *Microsurgery*. (2013) 33:496–502. doi: 10.1002/micr.22119

17. Zhang X, Huang Y, Mu D. Efficiency in diep flap breast reconstruction: the real benefit of computed tomographic angiography imaging. *Plast Reconstr Surg*. (2021) 148:301e–2e. doi: 10.1097/PRS.0000000000008107

18. Cui L, Jiang WQ, Zhang DK, Wang GF, Han YD, Pu WW, et al. A three-dimensional visualization of the full-field surgical region based on thin-slice mri: A helpful approach for simultaneously guiding tumor resection and perforator flap elevation. *Front Surg*. (2022) 9:984892. doi: 10.3389/fsurg.2022.984892

19. Cina A, Salgarello M, Barone-Adesi L, Rinaldi P, Bonomo L. Planning breast reconstruction with deep inferior epigastric artery perforating vessels: multidetector ct angiography versus color doppler us. *Radiology*. (2010) 255:979–87. doi: 10.1148/radiol.10091166

20. Pelanis E, Kumar RP, Aghayan DL, Palomar R, Fretland ÅA, Brun H, et al. Use of mixed reality for improved spatial understanding of liver anatomy. *Minim Invasive Ther Allied Technol*. (2020) 29:154–60. doi: 10.1080/13645706.2019.1616558

21. Edwards PJ, Hawkes DJ, Hill DL, Jewell D, Spink R, Strong A, et al. Augmentation of reality using an operating microscope for otolaryngology and neurosurgical guidance. *J Image Guid Surg*. (1995) 1:172–8. doi: 10.1002/(ISSN)1522-712X

22. Li L, Yang J, Chu Y, Wu W, Xue J, Liang P, et al. A novel augmented reality navigation system for endoscopic sinus and skull base surgery: A feasibility study. *PloS One*. (2016) 11:e0146996. doi: 10.1371/journal.pone.0146996

23. Mijuskovic B, Tremp M, Heimer MM, Boll D, Aschwanden M, Zeindler J, et al. Color doppler ultrasound and computed tomographic angiography for perforator mapping in diep flap breast reconstruction revisited: A cohort study. *J Plast Reconstr Aesthet Surg*. (2019) 72:1632–9. doi: 10.1016/j.bjps.2019.06.008

24. Tepper OM, Rudy HL, Lefkowitz A, Weimer KA, Marks SM, Stern CS, et al. Mixed reality with hololens: where virtual reality meets augmented reality in the operating room. *Plast Reconstr Surg*. (2017) 140:1066–70. doi: 10.1097/PRS.0000000000003802

25. Ramakrishnan VR, Orlandi RR, Citardi MJ, Smith TL, Fried MP, Kingdom TT. The use of image-guided surgery in endoscopic sinus surgery: an evidence-based review with recommendations. *Int Forum Allergy rhinol*. (2013) 3:236–41. doi: 10.1002/alr.21094

26. Lee SC, Fuerst B, Tateno K, Johnson A, Fotouhi J, Osgood G, et al. Multi-modal imaging, model-based tracking, and mixed reality visualisation for orthopaedic surgery. *Healthc Technol Lett*. (2017) 4:168–73. doi: 10.1049/htl.2017.0066

27. Shi L, Luo T, Zhang L, Kang ZC, Chen J, Wu FY, et al. Preliminary use of HoloLens glasses in surgery of liver cancer. *J Cent South Univ (Medical Science)*. (2018) 43:500–4. doi: 10.11817/j.issn.1672-7347.2018.05.007

28. Golusiński W, Pieńkowski P, Majchrzak E. Robotic surgery (Da vinci xi system) in head and neck cancer - own experience. *Otolaryngol Pol*. (2019) 74:1–5. doi: 10.5604/01.3001.0013.5262

29. Liu HH, Li LJ, Shi B, Xu CW, Luo E. Robotic surgical systems in maxillofacial surgery: A review. *Int J Oral Sci*. (2017) 9:63–73. doi: 10.1038/ijos.2017.24

30. Marescaux J, Rubino F, Arenas M, Mutter D, Soler L. Augmented-reality-assisted laparoscopic adrenalectomy. *JAMA*. (2004) 292:2214–5. doi: 10.1001/jama.292.18.2214-c

31. Ma CY, Zhou YL, Li HL, Hou WX, Shao BF, Zhang SQ, et al. Application of mixed reality navigation and 3d visualization in complex hepatocellular carcinoma resection. *Chin Med J*. (2019) 99:279–83. doi: 10.3760/cma.j.issn.0376-2491.2019.04.008

32. Xu L, Sun Y, Chen L, Wang XW. Application of mixed reality technique in the resection of benign lateral skull base tumor. *J Shandong Univ (Health Sci)*. (2020) 58:31–8. doi: 10.6040/j.issn.1671-7554.0.2019.1139

33. Zhang ZY, Duan WC, Chen RK, Zhang FJ, Yu B, Zhan YB, et al. Preliminary application of mxed reality in neurosurgery: development and evaluation of a new intraoperative procedure. *J Clin Neurosci*. (2019) 67:234–8. doi: 10.1016/j.jocn.2019.05.038

34. Ma YF, Li YJ, Zhang T, Feng CJ, Pan JY, Liu Y. Precise localization excision of pulmonary nodules assisted by mixed reality technology. *Chin J Endoscopic Surg*. (2019) 12:179–81. doi: 10.3877/cma.j.issn.1674-6899.2019.03.012

35. Sun P, Zhao Y, Men J, Ma ZR, Jiang HZ, Liu CY, et al. Application of virtual and augmented reality technology in hip surgery: systematic review. *J Med Internet Res*. (2023) 25:e37599. doi: 10.2196/37599

36. Bosc R, Fitoussi A, Pigneur F, Tacher V, Hersant B, Meningaud JP. Identification of perforating vessels by augmented reality: application for the deep inferior epigastric perforator flap. *Ann Chir Plast Esthet*. (2017) 62:336–9. doi: 10.1016/j.anplas.2017.01.002

37. Pereira N, Kufeke M, Parada L, Troncoso E, Bahamondes J, Sanchez L, et al. Augmented reality microsurgical planning with a smartphone (Arm-ps): A dissection route map in your pocket. *J Plast Reconstr Aesthet Surg*. (2019) 72:759–62. doi: 10.1016/j.bjps.2018.12.023

38. Gregory TM, Gregory J, Sledge J, Allard R, Mir O. Surgery guided by mixed reality: presentation of a proof of concept. *Acta Orthop*. (2018) 89:480–3. doi: 10.1080/17453674.2018.1506974

39. Moore R, Mullner D, Nichols G, Scomacao I, Herrera F. Color doppler ultrasound versus computed tomography angiography for preoperative anterolateral thigh flap perforator imaging: A systematic review and meta-analysis. *J Reconstr Microsurg*. (2022) 38:563–70. doi: 10.1055/s-0041-1740958

40. Zhan Y, Zhu H, Geng P, Zou M, Qi J, Zhu Q. Revisiting the blood supply of the rectus femoris: A case report and computed tomography angiography study. *Ann Plast Surg*. (2020) 85:419–23. doi: 10.1097/SAP.0000000000002141

41. Young AMH, Bache S, Segaren N, Murphy S, Maraka J, Durrani AJ. Free flap selection and outcomes of soft tissue reconstruction following resection of intra-oral Malignancy. *Front Surg*. (2019) 6:53. doi: 10.3389/fsurg.2019.00053

42. Cai Y-C, Li C, D-f Z, YQ Z, RH S, CY S, et al. Comparative analysis of radial forearm free flap and anterolateral thigh flap in tongue reconstruction after radical resection of tongue cancer. *Orl*. (2019) 81:252–64. doi: 10.1159/000502151

43. Soutar DS, McGregor I. The radial forearm flap in intraoral reconstruction: the experience of 60 consecutive cases. *Plast reconstr Surg*. (1986) 78:1–8. doi: 10.1097/00006534-198607000-00001

44. Haerle M, Häfner H-M, Dietz K, Schaller H-E, Brunelli F. Vascular dominance in the forearm. *Plast Reconstr Surg*. (2003) 111:1891–8. doi: 10.1097/01.PRS.0000057529.76413.D7

45. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. (2003) 327:557. doi: 10.1136/bmj.327.7414.557

46. Yao C, Liu H-N, Wu H, Chen Y-J, Li Y, Fang Y, et al. Diagnostic and prognostic value of circulating micrornas for esophageal squamous cell carcinoma: A systematic review and meta-analysis. *J Cancer*. (2018) 9:2876. doi: 10.7150/jca.25351

47. Las DE, de Jong T, Zuidam JM, Verweij NM, Hovius SER, Mureau MAM. Identification of independent risk factors for flap failure: A retrospective analysis of 1530 free flaps for breast, head and neck and extremity reconstruction. *J Plastic Reconstr Aesthetic Surg*. (2016) 69:894–906. doi: 10.1016/j.bjps.2016.02.001

48. Zhou W, Zhang WB, Yu Y, Wang Y, Mao C, Guo CB, et al. Risk factors for free flap failure: A retrospective analysis of 881 free flaps for head and neck defect reconstruction. *Int J Oral Maxillofac Surg*. (2017) 46:941–5. doi: 10.1016/j.ijom.2017.03.023

49. Ranganath K, Jalisi SM, Naples JG, Gomez ED. Comparing outcomes of radial forearm free flaps and anterolateral thigh free flaps in oral cavity reconstruction: A systematic review and meta-analysis. *Oral Oncol*. (2022) 135:106214. doi: 10.1016/j.oraloncology.2022.106214

50. Ioannidis JPA, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ*. (2007) 335:914. doi: 10.1136/bmj.39343.408449.80

51. Weber SM, Kim JH, Wax MK. Role of free tissue transfer in skull base reconstruction. *Otolaryngology–Head Neck Surg*. (2007) 136:914–9. doi: 10.1016/j.otohns.2006.12.030

52. Cigna E, Rizzo MI, Greco A, De Virgilio A, Palmieri A, Maruccia M, et al. Retromolar trigone reconstructive surgery: prospective comparative analysis between free flaps. *Ann Surg Oncol*. (2015) 22:272–8. doi: 10.1245/s10434-014-3963-4

53. Knott PD, Seth R, Waters HH, Revenaugh PC, Alam D, Scharpf J, et al. Short-term donor site morbidity: A comparison of the anterolateral thigh and radial forearm fasciocutaneous free flaps. *Head Neck*. (2016) 38:E945–E8. doi: 10.1002/hed.24131

54. Oranges CM, Ling B, Tremp M, Wettstein R, Kalbermatten DF, Schaefer DJ. Comparison of anterolateral thigh and radial forearm free flaps in head and neck reconstruction. *In Vivo*. (2018) 32:893. doi: 10.21873/invivo.11325

55. Lin C-H, Chiu Y-H, Perng C-K, Liao W-C, Ma H. Experience with the use of free fasciocutaneous flap in through-and-through cheek-buccal defect reconstruction: surgical outcome and quality of life analysis. *Ann Plast Surg*. (2016) 76:S74–9. doi: 10.1097/SAP.0000000000000692

56. Farace F, Fois VEE, Manconi A, Puddu A, Stomeo F, Tullio A, et al. Free anterolateral thigh flap versus free forearm flap: functional results in oral reconstruction. *J Plastic Reconstr Aesthetic Surg*. (2007) 60:583–7. doi: 10.1016/j.bjps.2006.11.014

57. Aladimi MT, Han B, Li C, Helal H, Gao Z, Li L. Factors to consider when deciding on the type of free-flap reconstruction of head and neck soft tissue defects. *ORL*. (2017) 79:230–8. doi: 10.1159/000478103

58. Lu M, Sun G, Hu Q, Tang E, Wang Y. Functional assessment: free thin anterolateral thigh flap versus free radial forearm reconstruction for hemiglossectomy defects. *Medicina Oral Patología Oral y Cirugía Bucal*. (2015) 20:e757. doi: 10.4317/medoral.20727

59. Liu W-W, Li H, Guo Z-M, Zhang Q, Yang A-K, Liu X-K, et al. Reconstruction of soft-tissue defects of the head and neck: radial forearm flap or anterolateral thigh flap? *Eur Arch Oto-Rhino-Laryngol*. (2011) 268:1809–12. doi: 10.1007/s00405-011-1548-4

60. de Vicente JC, de Villalaín L, Torre A, Peña I. Microvascular free tissue transfer for tongue reconstruction after hemiglossectomy: A functional assessment of radial forearm versus anterolateral thigh flap. *J Oral Maxillofac Surg*. (2008) 66:2270–5. doi: 10.1016/j.joms.2008.01.018

61. Camaioni A, Loreti A, Damiani V, Bellioni M, Passali F, Viti C. Anterolateral thigh cutaneous flap vs. Radial forearm free-flap in oral and oropharyngeal reconstruction: an analysis of 48 flaps. *Acta otorhinolaryngol italica*. (2008) 28:7–12.

62. Novak CB, Lipa JE, Noria S, Allison K, Neligan PC, Gilbert RW. Comparison of anterolateral thigh and radial forearm free flap donor site morbidity. *Microsurgery*. (2007) 27:651–4. doi: 10.1002/micr.20425

63. Huang C-H, Chen H-C, Huang Y-L, Mardini S, Feng G-M. Comparison of the radial forearm flap and the thinned anterolateral thigh cutaneous flap for reconstruction of tongue defects: an evaluation of donor-site morbidity. *Plast Reconstr Surg*. (2004) 114:1704–10. doi: 10.1097/01.PRS.0000142476.36975.07

64. Thakker JS, Fernandes R. Evaluation of reconstructive techniques for anterior and middle skull base defects following tumor ablation. *J Oral Maxillofac Surg*. (2014) 72:198–204. doi: 10.1016/j.joms.2013.05.017

65. Wang X, Yan G, Zhang G, Li J, Liu J, Zhang Y. Functional tongue reconstruction with the anterolateral thigh flap. *World J Surg Oncol*. (2013) 11:303. doi: 10.1186/1477-7819-11-303

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 26 July 2024
DOI 10.3389/fonc.2024.1425837

Check for updates

OPEN ACCESS

EDITED BY
Sharon R. Pine,
University of Colorado Anschutz Medical Campus, United States

REVIEWED BY
Salvatore Annunziata,
Fondazione Policlinico Universitario A. Gemelli IRCCS, Italy
Zheng Yuan,
China Academy of Chinese Medical Sciences, China

*CORRESPONDENCE
Xiaofeng Li
xli03@tmu.edu.cn
Wengui Xu
wenguixy@yeah.net

†These authors have contributed equally to this work

RECEIVED 30 April 2024
ACCEPTED 09 July 2024
PUBLISHED 26 July 2024

CITATION
Liu J, Sui C, Bian H, Li Y, Wang Z, Fu J, Qi L, Chen K, Xu W and Li X (2024) Radiomics based on ^{18}F-FDG PET/CT for prediction of pathological complete response to neoadjuvant therapy in non-small cell lung cancer.
Front. Oncol. 14:1425837.
doi: 10.3389/fonc.2024.1425837

COPYRIGHT
© 2024 Liu, Sui, Bian, Li, Wang, Fu, Qi, Chen, Xu and Li. 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.

Radiomics based on ^{18}F-FDG PET/CT for prediction of pathological complete response to neoadjuvant therapy in non-small cell lung cancer

Jianjing Liu[1,2,3†], Chunxiao Sui[1,2†], Haiman Bian[2,4], Yue Li[2,5], Ziyang Wang[3], Jie Fu[1,2,3], Lisha Qi[2,6], Kun Chen[1,2], Wengui Xu[1,2*] and Xiaofeng Li[1,2*]

[1]Department of Molecular Imaging and Nuclear Medicine, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [2]National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China, [3]Department of Molecular Imaging and Nuclear Medicine, Tianjin Cancer Hospital Airport Hospital, Tianjin, China, [4]Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [5]Department of Lung Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [6]Department of Pathology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China

Purpose: This study aimed to establish and evaluate the value of integrated models involving ^{18}F-FDG PET/CT-based radiomics and clinicopathological information in the prediction of pathological complete response (pCR) to neoadjuvant therapy (NAT) for non-small cell lung cancer (NSCLC).

Methods: A total of 106 eligible NSCLC patients were included in the study. After volume of interest (VOI) segmentation, 2,016 PET-based and 2,016 CT-based radiomic features were extracted. To select an optimal machine learning model, a total of 25 models were constructed based on five sets of machine learning classifiers combined with five sets of predictive feature resources, including PET-based alone radiomics, CT-based alone radiomics, PET/CT-based radiomics, clinicopathological features, and PET/CT-based radiomics integrated with clinicopathological features. Area under the curves (AUCs) of receiver operator characteristic (ROC) curves were used as the main outcome to assess the model performance.

Results: The hybrid PET/CT-derived radiomic model outperformed PET-alone and CT-alone radiomic models in the prediction of pCR to NAT. Moreover, addition of clinicopathological information further enhanced the predictive performance of PET/CT-derived radiomic model. Ultimately, the support vector machine (SVM)-based PET/CT radiomics combined clinicopathological information presented an optimal predictive efficacy with an AUC of 0.925 (95% CI 0.869–0.981) in the training cohort and an AUC of 0.863 (95% CI 0.740–0.985) in the test cohort. The developed nomogram involving radiomics and pathological type was suggested as a convenient tool to enable clinical application.

Conclusions: The ^{18}F-FDG PET/CT-based SVM radiomics integrated with clinicopathological information was an optimal model to non-invasively predict pCR to NAC for NSCLC.

KEYWORDS

^{18}F-FDG PET/CT, radiomics, NSCLC, neoadjuvant therapy, pathological complete response

1 Introduction

Lung cancer is a significant contributor to cancer-related mortality globally, with non-small cell lung cancer (NSCLC) comprising approximately 85% of all lung cancer cases (1–3). Despite the advancements in the treatment options for metastatic NSCLC, progress in the early-stage setting is limited (4). Driven by the necessity to enhance survival outcomes, renewed interest is emerging in exploring neoadjuvant strategies for early-stage and locally advanced NSCLC (5, 6).

Neoadjuvant therapy (NAT) encompasses various systemic treatment modalities administered prior to surgery, such as neoadjuvant chemotherapy, radiotherapy, targeted therapy, and immunotherapy (7). These therapies aim to reduce tumor burden, improve surgical outcomes, and ultimately enhance long-term survival rates for cancer patients (8). In 2020, the International Association for the Study of Lung Cancer (IASLC) released a recommendation for the pathological evaluation of neoadjuvant therapy, which introduced a standardized definition for major pathological response (MPR) as less than or equal to 10% viable tumor, whereas pathological complete response (pCR) indicated the absence of viable tumor (5). On the one hand, by virtue of the pathological response evaluation to neoadjuvant therapy, tumor sensitivity to systemic therapy is assessed at an early stage, which serves as a guide in determining the appropriate postoperative treatment strategy (9). On the other hand, previous clinical studies also proved that the pathological response to neoadjuvant treatment was a strong predictor for both disease-free survival and overall survival (10). Therefore, a noninvasive and reliable approach was in urgent need to predict the pathologic response to neoadjuvant therapy before treatment, which was beneficial to select potentially responsive NSCLC patients to administrate neoadjuvant therapy and maximize the therapeutic efficacy.

Computed tomography (CT), contrast-enhanced CT, and ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) are now commonly used imaging modalities in the clinical management of patients with lung cancer (11, 12). Particularly, PET/CT, as a hybrid imaging method, which is able to simultaneously provide metabolic information and anatomical details, is widely used in almost every aspect of clinical practice, including diagnosis, staging, treatment evaluation, and survival prognostication (13, 14). Consistently, several traditional metabolic parameters derived from PET images, such as maximum standardized uptake value (SUVmax), SUVmean, SUVpeak, metabolic tumor volume (MTV), and total lesion glycolysis (TLG), were previously identified as potential biomarkers in molecular subtype classification, pathological patterns determination, and outcome prediction for patients with NSCLC (15–18). However, as semi-quantitative parameters, especially for single-pixel value SUV, those commonly used conventional PET metabolic parameters were not able to reflect the complex heterogeneity existed in the images. More advanced methodology with enhanced predictive capability is expected to improve the prospect of NSCLC.

With the enormous improvement in the computing techniques in the era of big data, artificial intelligence, such as radiomics and machine learning, is increasingly becoming prevalent in the field of medical imaging (11, 19). In radiomics, a high throughput of features, which reflect the intra-tumor and inter-tumor heterogeneity are extracted, and a subset of informative radiomic features are finally selected after using a series of mathematical algorithms (20–22). In the end, multiple types of machine learning models are established and used as classifiers (23, 24). It is worth noting that these radiomic features are able to provide comprehensive heterogeneity information, which are usually not captured by the naked eyes. Though previous studies reported the roles of radiomics and machine learning based on CT in predicting pCR to NAT for NSCLC (6, 25, 26), few studies regarding PET/CT-derived radiomics were currently available.

In the present investigation, a total of 25 machine learning models, which involved five different combinations of predictive feature sources (PET-based alone radiomics, CT-based alone radiomics, PET/CT-based radiomics, clinicopathological features, and PET/CT radiomics integrated with clinicopathological features) and five different machine learning classifiers, were established to select the optimal model for predicting pCR to NAT in NSCLC. Furthermore, a nomogram with a visually straightforward representation was also constructed to detect the potential application of the developed machine learning models in clinical practice. This established machine learning model was potentially predictive of pCR to NAT in NSCLC, which provided a non-invasive approach to optimize the efficacy of NAT and improve the personalized treatment for NSCLC.

2 Materials and methods

2.1 Study population

A total of 188 NSCLC patients who underwent PET/CT imaging prior to NAT and surgical resection from June, 2020 to July, 2022 were enrolled in the retrospective study according to inclusion and exclusion criteria (Supplementary Materials). Ultimately, 106 NSCLC patients were included in the analysis, which were randomly divided into a training cohort (n = 74) and a testing cohort (n = 32) with a ratio of 7:3. This retrospective study was approved by the Ethics Review Committee of Tianjin Medical University Cancer Institute and Hospital, and written informed consent was waived. All procedures performed on human participants were conducted in compliance with the declaration of Helsinki and relevant ethical guidelines.

2.2 NAT regimen and pathological assessment

All the included patients underwent three to four cycles of platinum-based neoadjuvant chemotherapy, some of which also received concurrent immunotherapy. Then, surgery was conducted on all patients within 4–6 weeks after NAT. The pathological response to NAT was evaluated based on biopsy after resection by two pathologists with over 10 years of experience. pCR was defined as absence of residual tumor in histopathological section after resection (27).

2.3 Image acquisition and calculation of conventional PET metabolic parameters

Before imaging, NSCLC patients were informed to fast for at least 6 h and maintain their blood glucose levels below 140 mg/dl. Then, each patient received an intravenous injection of 3.7–4.44 MBq/kg (0.1–0.12 mCi/kg) of ^{18}F-FDG. After resting for approximately 60 min, the acquisition of images was carried out using the GE Discovery Elite PET/CT scanner (GE Medical Systems). A low-dose CT scan (helical pitch 0.75:1, 3.75-mm slice thickness, 120 kV and 50–80 mAs) was first performed to provide anatomical correlation and for attenuation correction purpose. Then, a PET scan, consisting of eight-bed positions with each bed position requiring a 2-min duration with increments of 16.2 cm (3D mode), was followed from the top of the skull to the distal femur. All PET/CT images were independently reviewed by two experienced experts specialized in PET/CT imaging, and any disagreement in the interpretation was resolved by consensus. To determine the volume of interest (VOI), a commercial software (PET VCAR; GE Healthcare, USA) on GE Advantage Workstation 4.6 (AW 4.6) was employed by applying an isocontour threshold of 41% of the maximum SUV (SUVmax) method (28). Within the VOI, calculations of SUVmax, SUVmean, and SUVpeak were automatically performed. MTV was defined as a volumetric measurement of a lesion exhibiting significantly high ^{18}F-FDG uptake (29). TLG was another volumetric index that was calculated by multiplying MTV with SUVmean.

2.4 Image segmentation and feature extraction

Semi-automatic segmentation of VOI was performed on CT images and PET images using 3D Slicer (version: 4.11.20210226) software by two nuclear medicine physicians with more than 5-year experiences specialized in PET/CT imaging. Before feature extraction, the spacing of PET and CT images and their corresponding VOIs were resampled to $1 \times 1 \times 1$ mm^3. A total of 4,032 radiomic features were extracted for each of the included NSCLC patients using the Pyradiomics module in Python 3.7.0, including a set of 2,016 CT-based alone and 2,016 PET-based alone radiomic features. To normalize the data into a standardized intensity range, we employed Z-score normalization for each radiomic feature.

2.5 Feature selection

For radiomic feature selection, the interclass correlation coefficient (ICC) test was first performed to assess the intra-observer and inter-observer repeatability in radiomic feature extraction. Radiomic features with ICC ≥0.75 were indicative of good reproducibility and reliability (30), whereas features with ICC <0.75 were excluded from further analysis. Second, a Mann–Whitney U test was used to select features highly related to pathological response to NAT with a significance level of 0.05 ($p < 0.05$). Then, Pearson's rank correlation analysis was conducted to eliminate or avoid feature redundancy. Features with Pearson's correlation coefficients above 0.90 were potentially highly related, in which one of the paired two features with a lower AUC was excluded. Furthermore, Minimum Redundancy Maximum Relevance (MRMR) was also implemented to further select the most significant and independent features. In the end, the least absolute shrinkage and selection operator (LASSO) was employed to select features for constructing the LASSO equation and calculating the corresponding feature weights. By adjusting the regularization weight λ, LASSO effectively reduced the magnitude of regression coefficients toward zero and eliminated the coefficients of irrelevant features by setting them precisely to zero. To determine the optimal λ, 10-fold cross-validation with minimum criteria was used. Nonzero coefficient features were selected and fitted into the regression model forming a radiomics signature. A radiomic score (Rad_Score) was then computed for each patient by combining the retained features linearly weighted by their respective model coefficients. Three distinct radiomic models were developed depending on the source of the extracted radiomic features. Rad_CT model was a radiomic model based on CT-derived alone features, and Rad_PET model was a radiomic model based on PET-derived alone features. For Rad_PET/CT model, both CT- and

PET-based radiomic features were included to select a subset of predictive radiomic features to establish radiomic model. For feature selection of clinicopathological information, we followed a two-step procedure. First, univariate logistic regression analysis was conducted to identify significant features with a p-value <0.05. Then, the stepwise multivariate logistic regression analysis was performed on the aforementioned significant features to determine the independent indicator with a p-value <0.05, which were used as the predictive clinicopathological parameters to establish machine learning models for prediction of pCR to NAT for NSCLC.

2.6 Machine learning model construction

The imbalanced data between pCR and non-pCR groups (35:71) was corrected using synthetic minority over-sampling technique (SMOTE), in which the k-nearest neighbor algorithm was utilized to oversample the minority sample until achieving an equal number of cases in each group. After LASSO regression, the multiple sets of selected features were incorporated with five different types of machine learning classifiers, including Logistic Regression (LR), support vector machine (SVM), K-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM), and NaiveBayes (NB), to construct corresponding machine learning models. Apart from these machine learning model based on radiomic features, machine learning model based on predictive clinicopathological parameters were also established, which were referred as Cli_Pat model in the study. To assess the contribution of radiomics combined with clinicopathological information to prediction of pCR to NAT for NSCLC, an integrated model named Cli_Pat_Rad_PET/CT was also built to determine its outperformance in contrast with Rad_PET/CT.

2.7 Statistical analysis

The IBM SPSS Statistics 27.0.1 and Python 3.7.0 software were employed for statistical analysis. For clinical data, quantitative data that conformed to normal distribution were expressed as mean ± SD, and comparisons between the two groups were conducted using the two independent-sample t-test. Non-normally distributed quantitative information was represented as M (P25–P75), and comparisons between two groups were performed using the Mann–Whitney U test. Qualitative data were compared using either the $\chi 2$ test or Fisher's exact test. Univariate and multivariate logistic regression analyses were performed to select the significant clinicopathological parameters in the prediction of pCR to NAT for NSCLC. A two-sided p-value below 0.05 was considered statistically significant. The prediction results of each model were plotted on a receiver operator characteristic (ROC) curve, and the area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated to assess the prediction performance. Thus, a nomogram using logistic regression algorithm involving radiomics and significant clinicopathological indicators was developed to detect its potential application in clinical practice.

3 Results

3.1 Patient characteristics

A total of 106 patients with NSCLC were eligible and recruited according to inclusion criteria. Based on the pathological outcome to NAT, these included NSCLC patients were classified into two groups: pCR and non-pCR. The differences in the clinicopathologic characteristics between the two groups are presented in Table 1. As shown, among all the listed clinicopathologic characteristics, only pathological type (p = 0.012) exhibited statistical significance in distinguishing pCR from non-pCR. With regard to traditional PET metabolic parameters, including SUVmax, SUVpeak, SUVmean, MTV, and TLG, none of them was suggested as potential indicator to predict pathological outcome to NAT for NSCLC.

3.2 Radiomics for prediction of pCR to NAT

The flow chart of radiomics used in the study to predict pCR to NAT for NSCLC is presented in Figure 1. Briefly, the radiomics consist of VOI segmentation, feature extraction, feature selection, model construction and performance evaluation. A total of 4,032 radiomic features (2,016 CT-based and 2,016 PET-based radiomic features) were extracted for each lesion, including morphological, first-order, and texture features. Based on a different combination of selective feature source (PET-based alone radiomics, CT-based alone radiomics, PET/CT-based radiomics, clinicopathological features, and PET/CT radiomics integrated with clinicopathological features) and machine learning classifier (SVM, LR, KNN, LightGBM, and NaiveBayes), 25 machine learning models were finally constructed to select the optimal model for prediction of pCR to NAT. Figure 2 shows all radiomic features and corresponding p-value results. After radiomic feature extraction and reduction, LASSO regression finally selected nine features for Rad_CT model (Figure 3A), four features for Rad_PET model (Figure 3B), and seven features for Rad_PET/CT model (Figure 3C). Rad_Score was then computed for each patient by combining the selected features linearly, weighted by their respective coefficients. The formulas of Rad_Score for Rad_CT model, Rad_PET model, and Rad_PET/CT model are listed in Table 2.

3.3 Machine learning models based on radiomics

Based on the aforementioned radiomics model construction after LASSO regression, including Rad_CT model, Rad_PET model, and Rad_PET/CT model, a total of 15 machine learning models were developed by incorporating five different machine learning classifiers (SVM, KNN, LR, LightGBM, and NaiveBayes). ROC analyses were performed to evaluate the performance of these established machine learning models in the prediction of pathological response to NAT for NSCLC using the AUC as the main outcome. As indicated in Figure 4, LR-Rad_CT model (Figures 4A, D) with an AUC of 0.844 (training cohort) and 0.732 (testing cohort), KNN-Rad_PET model (Figures 4B, E) with an AUC of 0.773 (training cohort) and 0.729

TABLE 1 Demographic information and clinicopathological characteristics of patients.

Characteristics	Non-PCR (n = 71)	PCR (n = 35)	p-Value
Gender			0.999
Male	62 (87.3%)	31 (88.6%)	
Female	9 (12.7%)	4 (11.4%)	
Age	60.29 ± 8.07	62.57 ± 6.76	0.138
Smoking status			0.708
Current or former	57 (80.3%)	27 (77.1%)	
Never	14 (19.7%)	8 (22.9%)	
BMI	24.89 ± 3.11	25.13 ± 3.15	0.696
NLR	2.37 (1.71–2.99)	2.57 (1.79–3.19)	0.904
PLR	134.71 (104.98–176.10)	150.33 (113.11–182.03)	0.550
Tumor size	4.60 (3.75–6.15)	4.70 (3.05–6.40)	0.407
Tumor location			0.352
Superior lobe of left lung	22 (31.0%)	8 (22.9%)	
Inferior lobe of left lung	11 (15.5%)	4 (11.4%)	
Superior lobe of right lung	15 (21.1%)	14 (40.0%)	
Middle lobe of right lung	7 (9.9%)	2 (5.7%)	
Inferior lobe of right lung	16 (22.5%)	7 (20.0%)	
Pathological type			**0.012***
Adenocarcinoma	20 (28.2%)	3 (8.6%)	
Squamous carcinoma	46 (64.8%)	32 (91.4%)	
Large cell carcinoma	5 (7.0%)	0 (0.0%)	
Pathological stage			0.449
I	5 (7.0%)	5 (14.3%)	
II	18 (25.4%)	7 (20.0%)	
III	48 (67.6%)	23 (65.7%)	
PD-L1			0.845
Negative	25 (35.2%)	13 (37.1%)	
Positive	46 (64.8%)	22 (62.9%)	
Nodal metastasis			0.707
Negative	27 (38%)	12 (34.3%)	
Positive	44 (62%)	23 (65.7%)	
SUVmax	15.54 (11.90–21.13)	18.58 (12.55–23.04)	0.530
SUVpeak	13.10 (10.02–18.09)	15.97 (10.87–20.35)	0.631
SUVmean	9.31 (7.54–12.54)	11.04 (7.51–14.94)	0.321
MTV	21.91 (12.58–41.66)	19.41 (6.59–38.84)	0.721
TLG	205.26 (114.83–444.61)	220.92 (57.12–550.29)	0.938

A t-test was used for age and BMI; a Mann–Whitney U test was used for NLR, PLR, tumor size, SUVmax, SUVpeak, SUVmean, MTV, and TLG. A χ^2 test or Fisher's exact test was used for the rest. *$p < 0.05$.
BMI, body mass index; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; PD-L1, programmed death protein ligand 1; SUV, standardized uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.

(testing cohort) and LightGBM-Rad_PET/CT model (Figures 4C, F) with an AUC of 0.864 (training cohort) and 0.841 (testing cohort) were considered the optimal model in Rad_CT models, Rad_PET models, and Rad_PET/CT models, respectively. The decision curves and calibration curves are also depicted and shown in Supplemenatry Figure S1. Other measurements, including accuracy, sensitivity, specificity, PPV, and NPV, in training cohort and testing cohort are also calculated and demonstrated in Supplemenatry Tables S1–S3.

3.4 Machine learning models based on radiomics combined with clinicopathological information

Clinicopathological information was reported to potentially provide complementary information to radiomic models. The predictive clinicopathological parameters were selected based on univariate analysis and multivariate analysis. As shown in Supplemenatry Table S4, pathological type was found to be significantly related to pCR (p = 0.003), which was also an independent predictor (OR 0.786; 95% CI 0.689–0.897; p = 0.003). In the study, the clinicopathological information was also incorporated with machine learning classifiers to construct Cli_Pat machine learning models (Figures 5A, B). Combined Cli_Pat_Rad_PET/CT machine learning models were built by integrating the Rad_PET/CT radiomics with pathological type (Figures 5C, D). As indicated in the ROC curves of Figure 5, the combined Cli_Pat_Rad_PET/CT machine learning models outperformed the Cli_Pat machine learning models in prediction of pathological response to NAT for NSCLC. Among the five constructed combined Cli_Pat_Rad_PET/CT machine learning models, SVM-Cli_Pat_Rad_PET/CT model outperformed other models with an AUC of 0.923 in the training cohort and an AUC of 0.857 in the testing cohort, which further improved the predictive performance of Rad_PET/CT models. Therefore, the SVM-Cli_Pat_Rad_PET/CT model was selected for the following study. Furthermore, both the decision curve (Figure 5E) and the calibration curve (Figure 5F) analysis confirmed that the SVM-Cli_Pat_Rad_PET/CT exhibited the highest net benefit and the best calibration in predicting pCR status. Detailed information

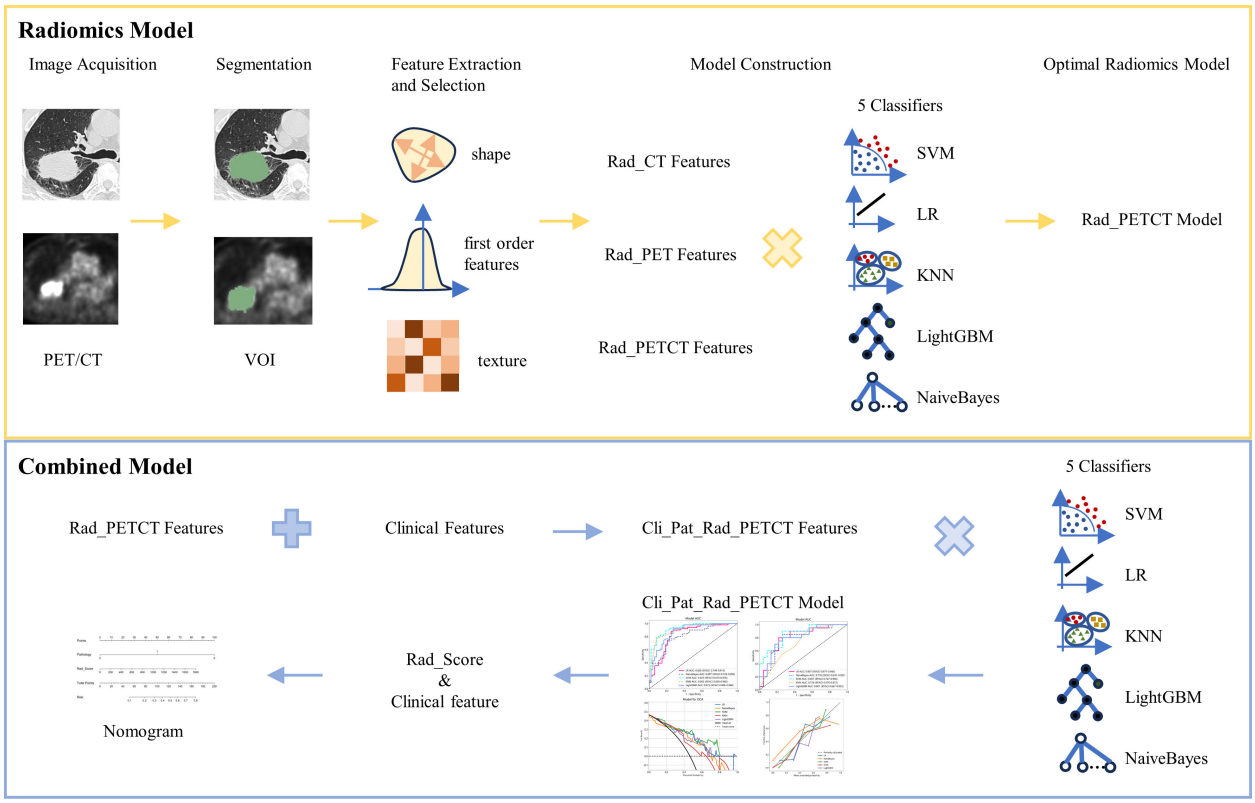


FIGURE 1
The workflow of radiomic and clinical analysis for image acquisition, segmentation, feature extraction and selection, and model and nomogram construction.

regarding the performance of all the constructed Cli_Pat machine learning models and Cli_Pat_Rad_PET/CT models were calculated and presented in Table 3 and Supplementary Table S5.

3.5 Nomogram construction

To detect the potential application of the developed PET/CT-derived radiomic models in predicting pCR to NAT for NSCLC, a nomogram using logistic regression algorithm was developed. As shown in Figure 6, both Rad_Score and clinicopathological predictor pathological type were involved in nomogram with a visually straightforward representation. In other words, the respective eligible point was endowed to Rad_Score and pathological type based on their different status, and the point for Rad_Score plus the point for pathological type was the total point. Ultimately, the risk of being predicted as pCR to NAT for individual NSCLC was deduced based on the obtained total point.

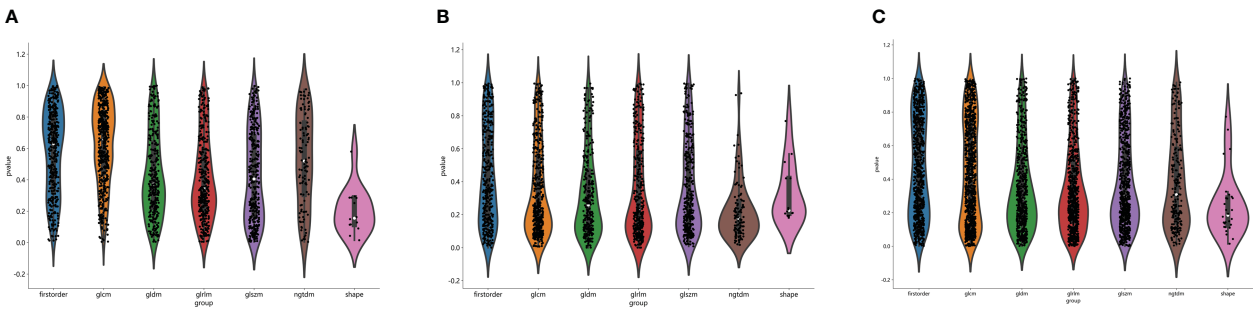


FIGURE 2
Distribution of radiomic features extracted from CT images only **(A)**, PET images only **(B)** and combined PET/CT images **(C)**, and corresponding p-value results in distinguishing PCR from non-PCR. glcm, gray-level co-occurrence matrix; gldm, gray-level dependence matrix; glrlm, gray-level run length matrix; glszm, gray level size zone matrix; ngtdm, neighborhood gray-tone difference matrix.

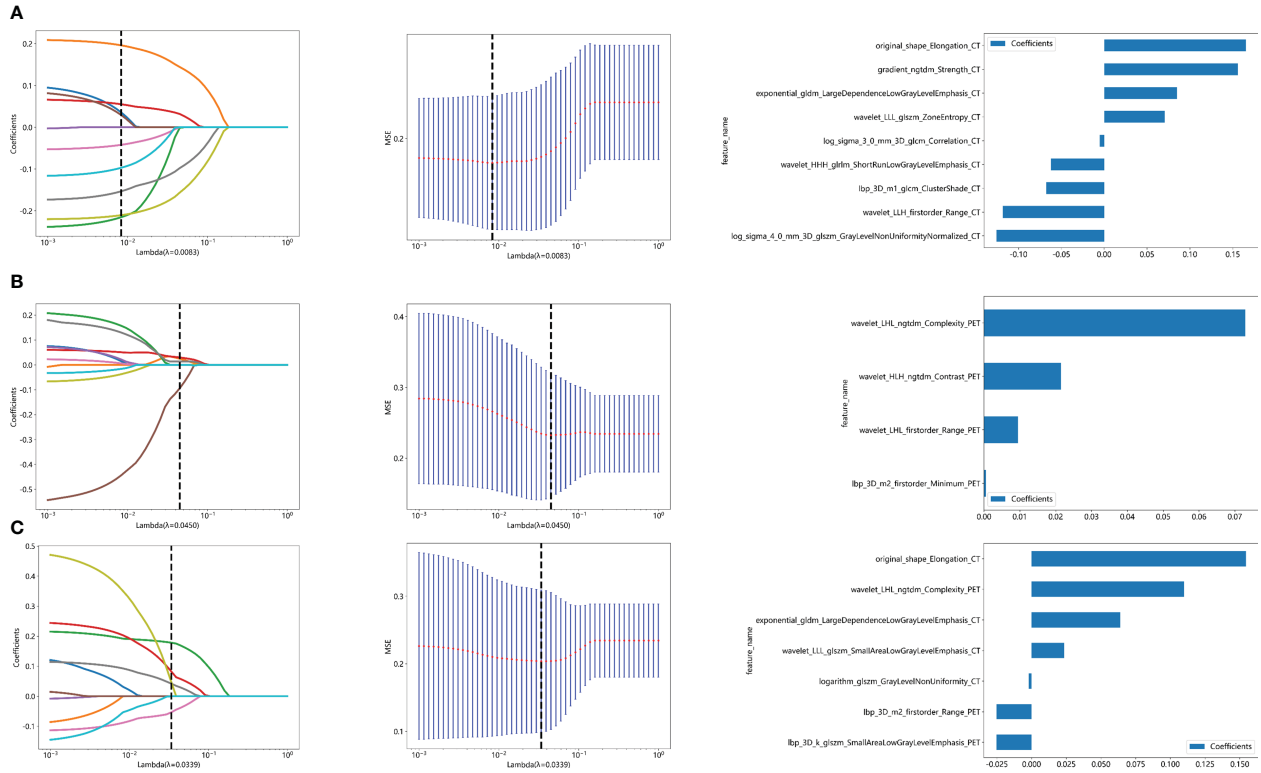


FIGURE 3
Radiomic feature selection used a LASSO algorithm, which was adjusted by a super parameter (λ), to achieve the purpose of screening the optimal features. The coefficients and mean standard error (MSE) of 10-fold cross validation and the histogram of coefficients of Rad_CT **(A)**, Rad_PET **(B)**, and Rad_PET/CT **(C)** models, respectively. **(A)** The vertical dashed line showed that the corresponding optimal λ value when obtaining the minimum deviation value was λ = 0.0083. Features with non-zero coefficients were screened out corresponding to the vertical lines in the plot, with a total of nine best features selected for Rad_CT model. Correspondingly, **(B)** the optimal λ value was 0.0450, and a total of four optimal features were selected for Rad_PET model. **(C)** The optimal λ value was 0.0339, and a total of seven optimal features were selected for Rad_PET/CT model.

4 Discussion

In the present study, a comprehensive radiomic analysis was performed to determine the role of PET/CT-derived machine learning models in the prediction of pathological response to NAT for NSCLC. To select an optimal model, a total of 25 machine learning models were established based on multiple combinations of different machine learning algorithms and different radiomic feature sources or clinicopathological indicators. Generally, PET/CT-derived radiomic models exhibited improved predictive performance than PET-based alone and CT-based alone radiomic models. The SVM-Rad_PET/CT model outperformed other machine learning models based on radiomic features. Furthermore, the SVM-Cli_Pat_Rad_PET/CT model was

TABLE 2 The establishment of Rad_Score formulas using selected radiomic features based on LASSO algorithm of the three Rad models.

Model Name	Formulas
Rad_CT	Rad_Score = 0.3314405037744722 + 0.070663 × wavelet_LLL_glszm_ZoneEntropy_CT + 0.165537 × original_shape_Elongation_CT - 0.062359 × wavelet_HHH_glrlm_ShortRunLowGrayLevelEmphasis_CT + 0.084869 × exponential_gldm_LargeDependenceLowGrayLevel Emphasis_CT + 0.156102 × gradient_ngtdm_Strength_CT − 0.067547 × lbp_3D_m1_glcm_ClusterShade_CT − 0.118425 × wavelet_LLH_firstorder_Range_CT − 0.125792 × log_sigma_4_0_mm_3D_glszm_GrayLevelNonUniformityNormalized_CT − 0.005096 × log_sigma_3_0_mm_3D_glcm_Correlation_CT
Rad_PET	Rad_Score = 0.3584026346819105 + 0.072922 × wavelet_LHL_ngtdm_Complexity_PET + 0.021449 × wavelet_HLH_ngtdm_Contrast_PET + 0.000527 × lbp_3D_m2_firstorder_Minimum_PET + 0.009493 × wavelet_LHL_firstorder_Range_PET
Rad_PET/CT	Rad_Score = 0.3638561342659228 + 0.154468 × original_shape_Elongation_CT + 0.109735 × wavelet_LHL_ngtdm_Complexity_PET − 0.025411 × lbp_3D_k_glszm_SmallAreaLowGrayLevelEmphasis_PET + 0.023428 × wavelet_LLL_glszm_SmallAreaLowGrayLevelEmphasis_CT − 0.002115 × logarithm_glszm_GrayLevelNonUniformity_CT + 0.063967 × exponential_gldm_LargeDependenceLowGrayLevelEmphasis_CT - 0.025217 × lbp_3D_m2_firstorder_Range_PET

CT, computed tomography; PET, positron emission tomography; PET/CT, positron emission tomography/computed tomography; GLSZM, gray-level size zone matrix; GLRLM, gray-level run length matrix; GLDM, gray-level dependence matrix; GLCM, gray-level co-occurrence matrix.

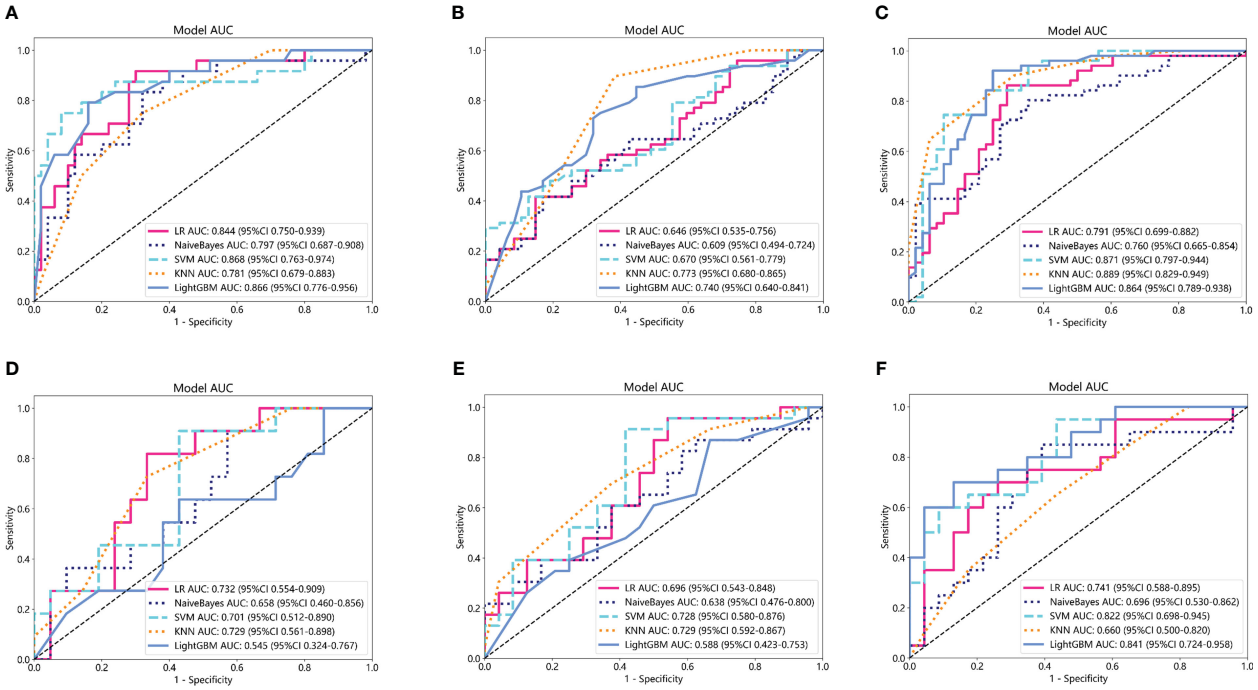


FIGURE 4
Comparison of ROC curves for the training cohorts of Rad_CT **(A)**, Rad_PET **(B)**, and Rad_PET/CT **(C)** models and testing cohorts of Rad_CT **(D)**, Rad_PET **(E)**, and Rad_PET/CT **(F)** models of five machine learning models. As indicated, the LR-Rad_CT model, KNN-Rad_PET model, and LightGBM-Rad_PET/CT model were the best predicting models, respectively. Among them, the LightGBM-Rad_PET/CT model with an AUC of 0.864 in the training cohort and 0.841 in the testing cohort were considered the optimal model for further analysis.

finally selected as the optimal model, which enhanced the predictive efficacy of the SVM-Rad_PET/CT model suggesting a complementary role provided by clinicopathological information.

Accurate prediction of pCR to NAT prior treatment was of significance in treatment decision making and survival prognostication for NSCLC (31). Though tremendous efforts were paid to identify potential biomarkers for predicting pCR to NAT, divergences remained to be resolved. Khorrami et al. assessed the role of clinicopathological variables in discerning pathological responses following NAT in NSCLC (25). Among all the variables, only lymphatic invasion exhibited statistical significance in distinguishing (MPR) from non-MPR (OR 0.052; 95% CI 0.007–0.23; $p = 0.0006$), while none of age, sex, pathological type, vascular infiltration, or tumor volume was found to be without significant differences between the MPR and non-MPR groups. Lin et al. revealed that gender was capable of predicting a good pathological response (GPR) to NAT ($p = 0.019$) (26). In our study, pathological type was significantly related to pCR ($p < 0.05$), which was also an independent predictor. The inconsistences in the determination of potential indicators were attributed to different pathological outcome settings and different patient cohorts with various clinical characteristics included in an individual study.

Besides clinical parameters, radiological indicator was also accepted as a non-invasive approach to predict pathological response to NAT for NSCLC. Particularly, several metabolic parameters based on PET/CT images were employed as potential radiological indicator to evaluate the response to NAT. Cui et al. revealed that SUVmax, SUVpeak, and peak SUV corrected for lean body mass (SULpeak) were significantly associated with pCR in patients with stage III NSCLC undergoing neoadjuvant immunochemotherapy therapy followed by surgery (32). Tao et al. discovered a negative correlation between the degree of pathological regression and the SULmax, SULpeak, MTV, and TLG of the preoperative PET/CT (33). Thus, the predictive capability of SUV parameters in determining pathological responses to NAT was verified in several studies for patients with a few types of tumors (34, 35), whereas our study proved that traditional PET/CT parameters (such as SUVmax, SUVmean, SUVpeak, MTV, and TLG) were unable to predict the pathological response to NAT ($p > 0.05$). Consistent with our results, Antunovic et al. also confirmed that SUVmax and TLG was not suggested as effective predictors of pCR to NAC for breast cancer patients (27). Among these PET metabolic semi-quantitative parameters, choosing a single-pixel value, such as SUV, is not able to comprehensively reflect the intra- and inter-tumoral heterogeneity (36–38).

Radiomics is an emerging hot topic in medical imaging, which is actually a high throughput of feature extraction to reflect the complex heterogeneity existing in the medical images, which is commonly not observed by the naked eye (39–42). Though Lin et al. established a combined radiomic model involving clinical features, radiomic features, and deep learning features for prediction of GPR

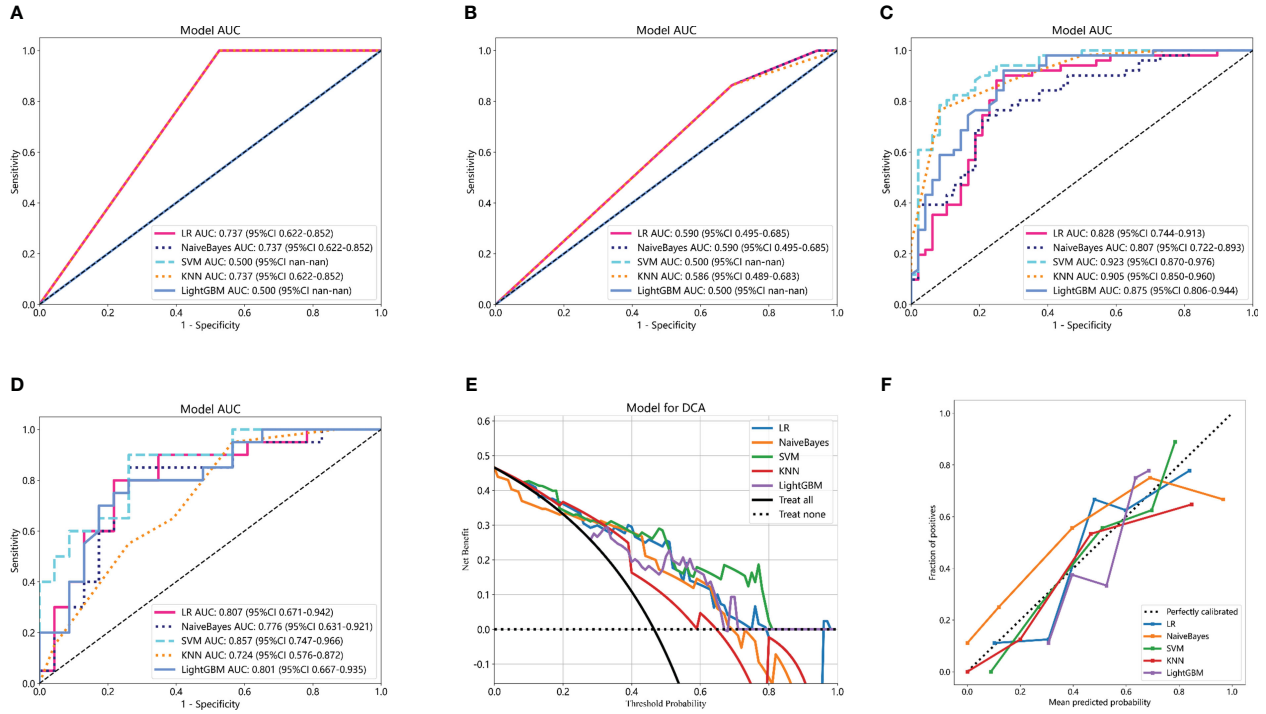


FIGURE 5
Comparison of ROC curves of five machine learning models for the training cohorts of Cli_Pat **(A)**, testing cohorts of Cli_Pat **(B)**, training cohorts of Cli_Pat_Rad_PET/CT **(C)**, and testing cohorts of Cli_Pat_Rad_PET/CT **(D)** models. The SVM-Cli_Pat_Rad_PET/CT model with an AUC of 0.923 in the training cohort and 0.857 in the testing cohort outperformed other Rad_PET/CT and Cli_Pat_Rad_PET/CT machine learning models. **(E)** The decision curve analysis of the Cli_Pat_Rad_PET/CT models. The SVM-Cli_Pat_Rad_PET/CT model had a higher net benefit in predicting pCR compared to the other four machine learning models. **(F)** The calibration curve of the Cli_Pat_Rad_PET/CT models. The 45° black dashed line represents the ideal prediction performance. The colorful lines of five machine learning models in which closer to the black dashed line represented the higher prediction accuracy. The SVM-Cli_Pat_Rad_PET/CT exhibited the best calibration in predicting pCR status.

to immunotherapy-based NAT for NSCLC, there is only a limited amount of radiomic features based on single-mode CT images (26). Despite prior research indicating the improvement of radiomics based on CT in predicting pCR to NAT for NSCLC, few studies involving PET/CT-derived radiomics were currently accessible. While PET/CT, as a dual-modality imaging technique, being capable of providing both anatomical and metabolic information, is expected to improve the performance of CT-based radiomic model in the prediction of pathological response to NAT for NSCLC (43–47). Thus, PET/CT-derived radiomics were

TABLE 3 Each evaluation index of Cli_Pat_Rad_PETCT model in five machine learning algorithms.

Model name	AUC	95% CI	Accuracy	Sensitivity	Specificity	PPV	NPV	
SVM	0.923	0.8704–0.9762	0.894	1.000	0.787	0.825	1.000	Train
SVM	0.857	0.7473–0.9664	0.896	1.000	0.826	0.828	1.000	Test
KNN	0.905	0.8504–0.9603	0.830	0.894	0.766	0.792	0.878	Train
KNN	0.724	0.6762–0.8723	0.792	1.000	0.636	0.706	0.878	Test
LR	0.828	0.7442–0.9131	0.787	0.830	0.745	0.765	0.814	Train
LR	0.807	0.6711–0.9420	0.833	0.958	0.708	0.767	0.944	Test
LightGBM	0.875	0.8062–0.9443	0.819	0.830	0.809	0.812	0.826	Train
LightGBM	0.801	0.6673–0.9354	0.833	0.833	0.870	0.833	0.833	Test
NaiveBayes	0.807	0.7223–0.8934	0.766	0.809	0.723	0.745	0.826	Train
NaiveBayes	0.776	0.6312–0.9213	0.833	0.792	0.875	0.864	0.833	Test

AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; SVM, support vector machine; KNN, K-nearest neighbors; LR logistic regression; LightGBM, light gradient boosting machine.

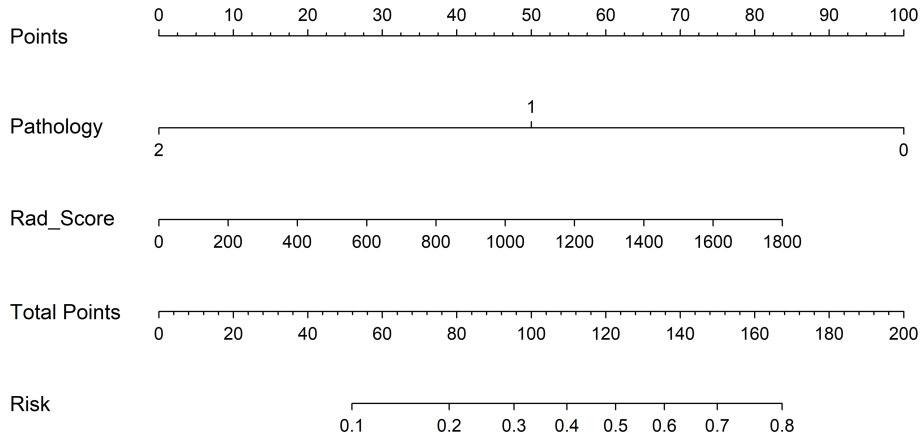


FIGURE 6
Clinical application of the nomogram in the differentiation of pCR and non-pCR in NSCLC patients. Locate on the pathology and Rad_Score coordinate axis. Calculate and sum the scores corresponding to each point; locate on the total point coordinate axis. The corresponding value on the bottom line is the probability of pathological response to pCR in patients with NSCLC after NAT.

increasingly performed to evaluate its power in the prediction of pCR to NAT for NSCLC. As expected, the PET/CT-derived radiomic models were superior to PET-based alone and CT-based alone radiomic models in our study (48, 49). Moreover, five machine learning classifiers were employed and compared to improve the prediction efficiency of the radiomic models. Among the five classifiers used in the present investigation, including LR, SVM, KNN, LightGBM, and NaiveBayes, SVM was found to be the optimal classifier, which was recommended to deal with nonlinear and high-dimensional classification issue with a small-to-medium sample size. Consistently, SVM was proven to be the optimal machine learning algorithm in various studies with respect to radiomics.

Nomogram was commonly used to detect potential application of a model involving multiple predictive indicators in clinical practice (50–52). As indicated in the results, the Cli_Pat_Rad_PET/CT model was proven to be the optimal model with an accuracy of 0.894 in the training cohort and an accuracy of 0.896 in the test cohort, respectively. Thus, a nomogram based on the Cli_Pat_Rad_PET/CT model was depicted in our study. The probability of being predicted as pCR was calculated based on both point of Rad_Score and point of pathological type, which were endowed according to their status. The developed nomogram simplified the procedure of prediction process with a straightforward visualization, which remarkably enhanced its feasibility to be conveniently applied in clinical practice (53, 54). In other words, NSCLC patients with a higher probability of being predicted as pCR to NAT, intensive administration is expected to maximize the therapeutic efficacy. In contrast, for NSCLC patients with a lower probability of being predicted as pCR to NAT, modified NAT or other effective treatment choice should be suggested in early stage of therapy process. Accurate prediction of pCR to NAT for NSCLC before treatment significantly contributing to therapeutic decision making promisingly improves the clinical outcome of NSCLC.

Although promising findings were obtained in this study, several limitations remained to be addressed. First, this was a retrospective study with a relatively small sample size, which was performed in one institution. A multi-center prospective study with an adequate sample size is warranted in the future to further verify the conclusion. Then, a subgroup analysis according to the type of NAT was not conducted due to the limited number of included NSCLC patients. Therefore, the status of PD-L1 was not involved to construct the combined machine learning model with clinicopathological information in the study because PD-L1 was usually assessed for NSCLC with immune checkpoint inhibitor treatment-based NAT. In the end, to improve the performance of artificial intelligence models based on radiological images in the prediction of pathological response to NAT, deep learning models involving VOI of both tumor lesion itself and peri-tumor region are considered a promising choice.

5 Conclusion

Machine learning models constructed based on PET/CT-derived radiomics were able to effectively predict pathological response to NAT prior treatment for NSCLC, and their predictive performances were further enhanced by the developed combined model involving PET/CT-derived radiomics and clinicopathological information. Therefore, the SVM-Cli_Pat_Rad_PET/CT model was potentially used a non-invasive tool to optimize personalized treatment and improve the clinical prospect of NAT for NSCLC.

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 authors.

Ethics statement

The studies involving humans were approved by Ethics Review Committee of Tianjin Medical University Cancer Institute and Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The written informed consent from all enrolled patients was waived with the confirmation of patient data confidentiality in the retrospective study.

Author contributions

JL: Writing – original draft, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. CS: Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. HB: Writing – original draft, Methodology, Formal analysis, Data curation. YL: Writing – original draft, Data curation. ZW: Writing – original draft, Data curation. JF: Writing – original draft, Data curation. LQ: Writing – original draft, Resources, Data curation. KC: Writing – original draft, Data curation. WX: Writing – review & editing, Supervision, Resources. XL: Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the National Natural Science Foundation of China (82272074), Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-010A), and institutional project (B2309).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1425837/full#supplementary-material

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

2. Monaco L, De Bernardi E, Bono F, Cortinovis D, Crivellaro C, Elisei F, et al. The "digital biopsy" in non-small cell lung cancer (NSCLC): a pilot study to predict the PD-L1 status from radiomics features of [18F]FDG PET/CT. *Eur J Nucl Med Mol Imaging*. (2022) 49:3401–11. doi: 10.1007/s00259-022-05783-z

3. Yuan Z, Yu X, Wu S, Wu X, Wang Q, Cheng W, et al. Instability mechanism of osimertinib in plasma and a solving strategy in the pharmacokinetics study. *Front Pharmacol*. (2022) 13:928983. doi: 10.3389/fphar.2022.928983

4. Yoo J, Lee J, Cheon M, Kim H, Choi YS, Pyo H, et al. Radiomics analysis of (18)F-FDG PET/CT for prognosis prediction in patients with stage III non-small cell lung cancer undergoing neoadjuvant chemoradiation therapy followed by surgery. *Cancers (Basel)*. (2023) 15:2012. doi: 10.3390/cancers15072012

5. Saw SPL, Ong BH, Chua KLM, Takano A, Tan DSW. Revisiting neoadjuvant therapy in non-small-cell lung cancer. *Lancet Oncol*. (2021) 22:e501–e16. doi: 10.1016/S1470-2045(21)00383-1

6. Liu C, Zhao W, Xie J, Lin H, Hu X, Li C, et al. Development and validation of a radiomics-based nomogram for predicting a major pathological response to neoadjuvant immunochemotherapy for patients with potentially resectable non-small cell lung cancer. *Front Immunol*. (2023) 14:1115291. doi: 10.3389/fimmu.2023.1115291

7. Beukinga RJ, Hulshoff JB, Mul VEM, Noordzij W, Kats-Ugurlu G, Slart R, et al. Prediction of response to neoadjuvant chemotherapy and radiation therapy with baseline and restaging (18)F-FDG PET imaging biomarkers in patients with esophageal cancer. *Radiology*. (2018) 287:983–92. doi: 10.1148/radiol.2018172229

8. Chen K, Wang J, Li S, Zhou W, Xu W. Predictive value of 18F-FDG PET/CT-based radiomics model for neoadjuvant chemotherapy efficacy in breast cancer: a multi-scanner/center study with external validation. *Eur J Nucl Med Mol Imaging*. (2023) 50:1869–80. doi: 10.1007/s00259-023-06150-2

9. Li P, Wang X, Xu C, Liu C, Zheng C, Fulham MJ, et al. (18)F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *Eur J Nucl Med Mol Imaging*. (2020) 47:1116–26. doi: 10.1007/s00259-020-04684-3

10. Umutlu L, Kirchner J, Bruckmann NM, Morawitz J, Antoch G, Ting S, et al. Multiparametric (18)F-FDG PET/MRI-based radiomics for prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. *Cancers (Basel)*. (2022) 14:1727. doi: 10.3390/cancers14071727

11. Yin X, Liao H, Yun H, Lin N, Li S, Xiang Y, et al. Artificial intelligence-based prediction of clinical outcome in immunotherapy and targeted therapy of lung cancer. *Semin Cancer Biol*. (2022) 86:146–59. doi: 10.1016/j.semcancer.2022.08.002

12. Dercle L, Sun S, Seban R-D, Mekki A, Sun R, Tselikas L, et al. Emerging and evolving concepts in cancer immunotherapy imaging. *Radiology*. (2023) 306:32–46. doi: 10.1148/radiol.210518

13. Chen L, Liu K, Zhao X, Shen H, Zhao K, Zhu W. Habitat imaging-based (18)F-FDG PET/CT radiomics for the preoperative discrimination of non-small cell lung cancer and benign inflammatory diseases. *Front Oncol*. (2021) 11:759897. doi: 10.3389/fonc.2021.759897

14. Fiz F, Masci C, Costa G, Sollini M, Chiti A, Ieva F, et al. PET/CT-based radiomics of mass-forming intrahepatic cholangiocarcinoma improves prediction of pathology data and survival. *Eur J Nucl Med Mol Imaging*. (2022) 49:3387–400. doi: 10.1007/s00259-022-05765-1

15. Hannequin P, Decroisette C, Kermanach P, Berardi G, Bourbonne V. FDG PET and CT radiomics in diagnosis and prognosis of non-small-cell lung cancer. *Transl Lung Cancer Res*. (2022) 11:2051–63. doi: 10.21037/tlcr-22-158

16. Li J, Ge S, Sang S, Hu C, Deng S. Evaluation of PD-L1 expression level in patients with non-small cell lung cancer by (18)F-FDG PET/CT radiomics and clinicopathological characteristics. *Front Oncol*. (2021) 11:789014. doi: 10.3389/fonc.2021.789014

17. Zhou J, Zou S, Kuang D, Yan J, Zhao J, Zhu X. A novel approach using FDG-PET/CT-based radiomics to assess tumor immune phenotypes in patients with non-small cell lung cancer. *Front Oncol*. (2021) 11:769272. doi: 10.3389/fonc.2021.769272

18. Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med*. (2018) 59:1321–8. doi: 10.2967/jnumed.117.199935

19. Mazzaschi G, Milanese G, Pagano P, Madeddu D, Gnetti L, Trentini F, et al. Integrated CT imaging and tissue immune features disclose a radio-immune signature with high prognostic impact on surgically resected NSCLC. *Lung Cancer*. (2020) 144:30–9. doi: 10.1016/j.lungcan.2020.04.006

20. Wang X, Yang W, Zhou Q, Luo H, Chen W, Yeung SJ, et al. The role of (18)F-FDG PET/CT in predicting the pathological response to neoadjuvant PD-1 blockade in combination with chemotherapy for resectable esophageal squamous cell carcinoma. *Eur J Nucl Med Mol Imaging*. (2022) 49:4241–51. doi: 10.1007/s00259-022-05872-z

21. Murakami Y, Kawahara D, Tani S, Kubo K, Katsuta T, Imano N, et al. Predicting the local response of esophageal squamous cell carcinoma to neoadjuvant chemoradiotherapy by radiomics with a machine learning method using (18)F-FDG PET images. *Diagnostics (Basel)*. (2021) 11:1049. doi: 10.3390/diagnostics11061049

22. Mu W, Jiang L, Shi Y, Tunali I, Gray JE, Katsoulakis E, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J Immunother Cancer*. (2021) 9:e002118. doi: 10.1136/jitc-2020-002118

23. Jimenez JE, Abdelhafez A, Mittendorf EA, Elshafeey N, Yung JP, Litton JK, et al. A model combining pretreatment MRI radiomic features and tumor-infiltrating lymphocytes to predict response to neoadjuvant systemic therapy in triple-negative breast cancer. *Eur J Radiol*. (2022) 149:110220. doi: 10.1016/j.ejrad.2022.110220

24. Ninatti G, Kirienko M, Neri E, Sollini M, Chiti A. Imaging-based prediction of molecular therapy targets in NSCLC by radiogenomics and AI approaches: A systematic review. *Diagnostics*. (2020) 10:359. doi: 10.3390/diagnostics10060359

25. Khorrami M, Jain P, Bera K, Alilou M, Thawani R, Patil P, et al. Predicting pathologic response to neoadjuvant chemoradiation in resectable stage III non-small cell lung cancer patients using computed tomography radiomic features. *Lung Cancer*. (2019) 135:1–9. doi: 10.1016/j.lungcan.2019.06.020

26. Lin Q, Wu HJ, Song QS, Tang YK. CT-based radiomics in predicting pathological response in non-small cell lung cancer patients receiving neoadjuvant immunotherapy. *Front Oncol*. (2022) 12:937277. doi: 10.3389/fonc.2022.937277

27. Antunovic L, De Sanctis R, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. *Eur J Nucl Med Mol Imaging*. (2019) 46:1468–77. doi: 10.1007/s00259-019-04313-8

28. Tong H, Sun J, Fang J, Zhang M, Liu H, Xia R, et al. A machine learning model based on PET/CT radiomics and clinical characteristics predicts tumor immune profiles in non-small cell lung cancer: A retrospective multicohort study. *Front Immunol*. (2022) 13:859323. doi: 10.3389/fimmu.2022.859323

29. Zhao X, Zhao Y, Zhang J, Zhang Z, Liu L, Zhao X. Predicting PD-L1 expression status in patients with non-small cell lung cancer using [18F]FDG PET/CT radiomics. *EJNMMI Res*. (2023) 13:4. doi: 10.1186/s13550-023-00956-9

30. Xie D, Xu F, Zhu W, Pu C, Huang S, Lou K, et al. Delta radiomics model for the prediction of progression-free survival time in advanced non-small-cell lung cancer patients after immunotherapy. *Front Oncol*. (2022) 12:990608. doi: 10.3389/fonc.2022.990608

31. Wang C, Ma J, Shao J, Zhang S, Liu Z, Yu Y, et al. Predicting EGFR and PD-L1 status in NSCLC patients using multitask AI system based on CT images. *Front Immunol*. (2022) 13:813072. doi: 10.3389/fimmu.2022.813072

32. Cui Y, Lin Y, Zhao Z, Long H, Zheng L, Lin X. Comprehensive (18)F-FDG PET-based radiomics in elevating the pathological response to neoadjuvant immunochemotherapy for resectable stage III non-small-cell lung cancer: A pilot study. *Front Immunol*. (2022) 13:994917. doi: 10.3389/fimmu.2022.994917

33. Tao X, Li N, Wu N, He J, Ying J, Gao S, et al. The efficiency of (18)F-FDG PET-CT for predicting the major pathologic response to the neoadjuvant PD-1 blockade in resectable non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. (2020) 47:1209–19. doi: 10.1007/s00259-020-04711-3

34. Yang L, Chang J, He X, Peng M, Zhang Y, Wu T, et al. PET/CT-based radiomics analysis may help to predict neoadjuvant chemotherapy outcomes in breast cancer. *Front Oncol*. (2022) 12:849626. doi: 10.3389/fonc.2022.849626

35. Giannini V, Mazzetti S, Bertotto I, Chiarenza C, Cauda S, Delmastro E, et al. Predicting locally advanced rectal cancer response to neoadjuvant therapy with (18)F-FDG PET and MRI radiomics features. *Eur J Nucl Med Mol Imaging*. (2019) 46:878–88. doi: 10.1007/s00259-018-4250-6

36. Polverari G, Ceci F, Bertaglia V, Reale ML, Rampado O, Gallio E, et al. 18F-FDG pet parameters and radiomics features analysis in advanced nsclc treated with immunotherapy as predictors of therapy response and survival. *Cancers*. (2020) 12:1163. doi: 10.3390/cancers12051163

37. Wang X, Xu C, Grzegorzek M, Sun H. Habitat radiomics analysis of pet/ct imaging in high-grade serous ovarian cancer: Application to Ki-67 status and progression-free survival. *Front Physiol*. (2022) 13:948767. doi: 10.3389/fphys.2022.948767

38. Beaumont J, Acosta O, Devillers A, Palard-Novello X, Chajon E, de Crevoisier R, et al. Voxel-based identification of local recurrence sub-regions from pre-treatment PET/CT for locally advanced head and neck cancers. *EJNMMI Res*. (2019) 9:90. doi: 10.1186/s13550-019-0556-z

39. Wang C, Xu X, Shao J, Zhou K, Zhao K, He Y, et al. Deep learning to predict EGFR mutation and PD-L1 expression status in non-small-cell lung cancer on computed tomography images. *J Oncol*. (2021) 2021:5499385. doi: 10.1155/2021/5499385

40. Wang C, Ma J, Shao J, Zhang S, Li J, Yan J, et al. Non-invasive measurement using deep learning algorithm based on multi-source features fusion to predict PD-L1 expression and survival in NSCLC. *Front Immunol*. (2022) 13:828560. doi: 10.3389/fimmu.2022.828560

41. Napel S, Mu W, Jardim-Perassi BV, Aerts H, Gillies RJ. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. *Cancer*. (2018) 124:4633–49. doi: 10.1002/cncr.31630

42. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol*. (2017) 72:3–10. doi: 10.1016/j.crad.2016.09.013

43. Leger MA, Routy B, Juneau D. FDG PET/CT for evaluation of immunotherapy response in lung cancer patients. *Semin Nucl Med*. (2022) 52:707–19. doi: 10.1053/j.semnuclmed.2022.04.010

44. Ding E, Lu D, Wei L, Feng X, Shen J, Xu W. Predicting tumor recurrence using metabolic indices of (18)F-FDG PET/CT prior to orthotopic liver transplantationfor hepatocellular carcinoma. *Oncol Lett*. (2020) 20:1245–55. doi: 10.3892/ol.2020.11681

45. Leithner D, Schoder H, Haug A, Vargas HA, Gibbs P, Haggstrom I, et al. Impact of comBat harmonization on PET radiomics-based tissue classification: A dual-center PET/MRI and PET/CT study. *J Nucl Med*. (2022) 63:1611–6. doi: 10.2967/jnumed.121.263102

46. Li Y, Zhang Y, Fang Q, Zhang X, Hou P, Wu H, et al. Radiomics analysis of [(18)F]FDG PET/CT for microvascular invasion and prognosis prediction in very-early- and early-stage hepatocellular carcinoma. *Eur J Nucl Med Mol Imaging*. (2021) 48:2599–614. doi: 10.1007/s00259-020-05119-9

47. Evangelista L, Cuocolo A, Pace L, Mansi L, Del Vecchio S, Miletto P, et al. Performance of FDG-PET/CT in solitary pulmonary nodule based on pre-test likelihood of Malignancy: results from the ITALIAN retrospective multicenter trial. *Eur J Nucl Med Mol Imaging*. (2018) 45:1898–907. doi: 10.1007/s00259-018-4016-1

48. Chen Y, Wang Z, Yin G, Sui C, Liu Z, Li X, et al. Prediction of HER2 expression in breast cancer by combining PET/CT radiomic analysis and machine learning. *Ann Nucl Med*. (2022) 36:172–82. doi: 10.1007/s12149-021-01688-3

49. Tixier F, Cheze-le-Rest C, Schick U, Simon B, Dufour X, Key S, et al. Transcriptomics in cancer revealed by Positron Emission Tomography radiomics. *Sci Rep*. (2020) 10:5660. doi: 10.1038/s41598-020-62414-z

50. Wang X, Xie T, Luo J, Zhou Z, Yu X, Guo X. Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment. *Breast Cancer Res*. (2022) 24:20. doi: 10.1186/s13058-022-01516-0

51. Cui L, Yu T, Kan Y, Dong Y, Luo Y, Jiang X. Multi-parametric MRI-based peritumoral radiomics on prediction of lymph-vascular space invasion in early-stage cervical cancer. *Diagn Interv Radiol*. (2022) 28:312–21. doi: 10.5152/dir.2022.20657

52. Mu W, Liang Y, Hall LO, Tan Y, Balagurunathan Y, Wenham R, et al. (18)F-FDG PET/CT habitat radiomics predicts outcome of patients with cervical cancer treated with chemoradiotherapy. *Radiol Artif Intell*. (2020) 2:e190218. doi: 10.1148/ryai.2020190218

53. Wang Y, Luo S, Jin G, Fu R, Yu Z, Zhang J. Preoperative clinical-radiomics nomogram for microvascular invasion prediction in hepatocellular carcinoma using [Formula: see text]F-FDG PET/CT. *BMC Med Imaging*. (2022) 22:70. doi: 10.1186/s12880-022-00796-4

54. Wang Q, Li C, Zhang J, Hu X, Fan Y, Ma K, et al. Radiomics models for predicting microvascular invasion in hepatocellular carcinoma: A systematic review and radiomics quality score assessment. *Cancers (Basel)*. (2021) 13:5864. doi: 10.3390/cancers13225864

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 05 August 2024
DOI 10.3389/fonc.2024.1390398

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Bin Li,
Sun Yat-sen University Cancer Center (SYSUCC), China
Zhenhui Dai,
Guangzhou University of Chinese Medicine, China

*CORRESPONDENCE
Wenzheng Sun
sunwenzheng@zju.edu.cn
Jun Dang
dangjun@cicams-sz.org.cn

†These authors have contributed equally to this work

RECEIVED 23 February 2024
ACCEPTED 01 July 2024
PUBLISHED 05 August 2024

CITATION
Zhang H, Chen K, Xu X, You T, Sun W and Dang J (2024) Spatiotemporal correlation enhanced real-time 4D-CBCT imaging using convolutional LSTM networks.
Front. Oncol. 14:1390398.
doi: 10.3389/fonc.2024.1390398

COPYRIGHT
© 2024 Zhang, Chen, Xu, You, Sun and Dang. 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.

Spatiotemporal correlation enhanced real-time 4D-CBCT imaging using convolutional LSTM networks

Hua Zhang[1,2†], Kai Chen[3†], Xiaotong Xu[1,2], Tao You[4], Wenzheng Sun[5*] and Jun Dang[6*]

[1]School of Biomedical Engineering, Southern Medical University, Guang Zhou, Guangdong, China, [2]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guang Zhou, Guangdong, China, [3]School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China, [4]Department of Radiation Oncology, The Affiliated Hospital of Jiangsu University, Zhenjiang, Jiangsu, China, [5]Department of Radiation Oncology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, [6]Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China

Purpose: To enhance the accuracy of real-time four-dimensional cone beam CT (4D-CBCT) imaging by incorporating spatiotemporal correlation from the sequential projection image into the single projection-based 4D-CBCT estimation process.

Methods: We first derived 4D deformation vector fields (DVFs) from patient 4D-CT. Principal component analysis (PCA) was then employed to extract distinctive feature labels for each DVF, focusing on the first three PCA coefficients. To simulate a wide range of respiratory motion, we expanded the motion amplitude and used random sampling to generate approximately 900 sets of PCA labels. These labels were used to produce 900 simulated 4D-DVFs, which in turn deformed the 0% phase 4D-CT to obtain 900 CBCT volumes with continuous motion amplitudes. Following this, the forward projection was performed at one angle to get all of the digital reconstructed radiographs (DRRs). These DRRs and the PCA labels were used as the training data set. To capture the spatiotemporal correlation in the projections, we propose to use the convolutional LSTM (ConvLSTM) network for PCA coefficient estimation. For network testing, when several online CBCT projections (with different motion amplitudes that cover the full respiration range) are acquired and sent into the network, the corresponding 4D-PCA coefficients will be obtained and finally lead to a full online 4D-CBCT prediction. A phantom experiment is first performed with the XCAT phantom; then, a pilot clinical evaluation is further conducted.

Results: Results on the XCAT phantom and the patient data show that the proposed approach outperformed other networks in terms of visual inspection and quantitative metrics. For the XCAT phantom experiment, ConvLSTM achieves the highest quantification accuracy with MAPE(Mean Absolute Percentage Error), PSNR (Peak Signal-to-Noise Ratio), and RMSE(Root Mean Squared Error) of 0.0459, 64.6742, and 0.0011, respectively. For the patient pilot clinical experiment, ConvLSTM also achieves the best quantification accuracy with that of 0.0934, 63.7294, and 0.0019, respectively. The quantification evaluation labels that we used are 1) the Mean Absolute Error (MAE), 2) the

Normalized Cross Correlation (NCC), 3)the Structural Similarity Index Measurement(SSIM), 4)the Peak Signal-to-Noise Ratio (PSNR), 5)the Root Mean Squared Error(RMSE), and 6) the Absolute Percentage Error (MAPE).

Conclusion: The spatiotemporal correlation-based respiration motion modeling supplied a potential solution for accurate real-time 4D-CBCT reconstruction.

KEYWORDS

ConvLSTM, PCA, radiation therapy, 4D-CBCT, spatiotemporal

1 Introduction

Stereotactic radiotherapy (SBRT) is commonly used in routine clinical radiation therapy circumstances, especially for early-stage cancer such as lung cancer (1). The high dose rate of the SBRT beam also brings high risk for moving targets (e.g., lung cancer). Hence, accurate image guidance plays a crucial role in precise lung SBRT. In clinical routine, the most common image guidance tool is the integrated 3D Cone Beam CT (CBCT) imaging system (2). However, conventional static 3D-CBCT is unable to provide qualified 4D lung motion during respiration.

Four-dimensional cone beam CT (4D-CBCT) imaging has been developed to address this issue. 4D-CBCT can supply temporal image sequences for moving organs such as the lung. Conventional analytical 4D-CBCT methods, such as the McKinnon–Bates (MKB) algorithm, are widely used in commercial linear accelerators. However, the image quality suffered from reduced contrast and the inevitable motion blurring induced by the time-averaged prior image (3). Another type of 4D-CBCT reconstruction method is the image deformation-based scheme (4). For these kinds of methods, the deformation vector fields (DVFs) calculation/estimation between the 0% phase and each other phase is critical to achieve the final accurate 4D-CBCT. The DVF optimization process is quite time consuming, and it raises a blind treatment risk for initiating radiation pneumonia (5). Both the above-mentioned analytical and deformable-based 4D-CBCT reconstructions all use the full 360° range acquired projections. Recently, online real-time CBCT estimation/reconstruction via single or only a few X-ray projections has attracted more interest. It benefits oncologists not only fast but also pretty low-dose real-time 4D-CBCT images compared with the conventional full projection-based 3D-CBCT (6).

The 2D- to 4D-CBCT estimation has been previously studied by many groups in the past decades. Li (7) proposed a motion model (MM) to predict 4D-CBCT via forward matching between 3D volumes and 2D X-ray projections. You (8) reported a motion model free deformation (MM-FD) scheme to introduce free deformation alignment for promoting 4D-CBCT estimation accuracy. One limitation of these iterative approaches is that they are quite time consuming. On the other aspect, Xu (6) reported a linear model for predicting 4D-CBCT via DRR (Digital Reconstructed Radiography) and validated it with digital and physical phantom experiments. However, the proposed linear model mismatches with the complex relationship between the intensity variation and the real breathing motion. Wei (9, 10) proposed a Convolutional Neural Network (CNN)-based framework to extract the motion feature from 2D DRRs to corresponding 3D-CBCT (e.g., one phase of 4D-CBCT). However, all of the aforementioned 4D-CBCT prediction strategies neglected the spatiotemporal correlation inherent in 4D-CBCT.

To address the issues, we propose a combined model that contains 1) a convolutional LSTM (ConvLSTM) and 2) a principal component analysis (PCA) model with prior 4D-CT to map a single 2D measured projection to one phase of 4D-CBCT. We evaluated the model's performance on both the XCAT phantom and pilot clinical data. Quantitative metrics are used for network performance quantification between our proposed method versus other state-of-the-art networks.

2 Methods

The overall workflow is illustrated in Figure 1. In the training stage, the 4D-DVFs are first derived from the 4D-CT (between 0% phase and other phases) via the voxel-by-voxel image registration algorithms (11–13). The DVFs then will be simply represented by the first few PCA coefficients. In our experiment, we chose the first three PCA coefficients. The PCA coefficient is further expanded to fully cover the potential possible motion range for simulation. We then performed random sampling and generated approximately 900 PCA coefficient groups. These groups will be used to create the corresponding 900 DVFs, which will in turn generate 900 deformed 4D-CT images with varying respiratory motions. Finally, a forward projection will be performed at a single angle for all 900 4D-CT images to acquire 900 DRRs. A ray-tracing algorithm (14, 15) is used in the forward projection simulation process. The generated DRRs will be used to train the ConvLSTM network, which has three output labels representing the first three PCA-modeled coefficients labels.

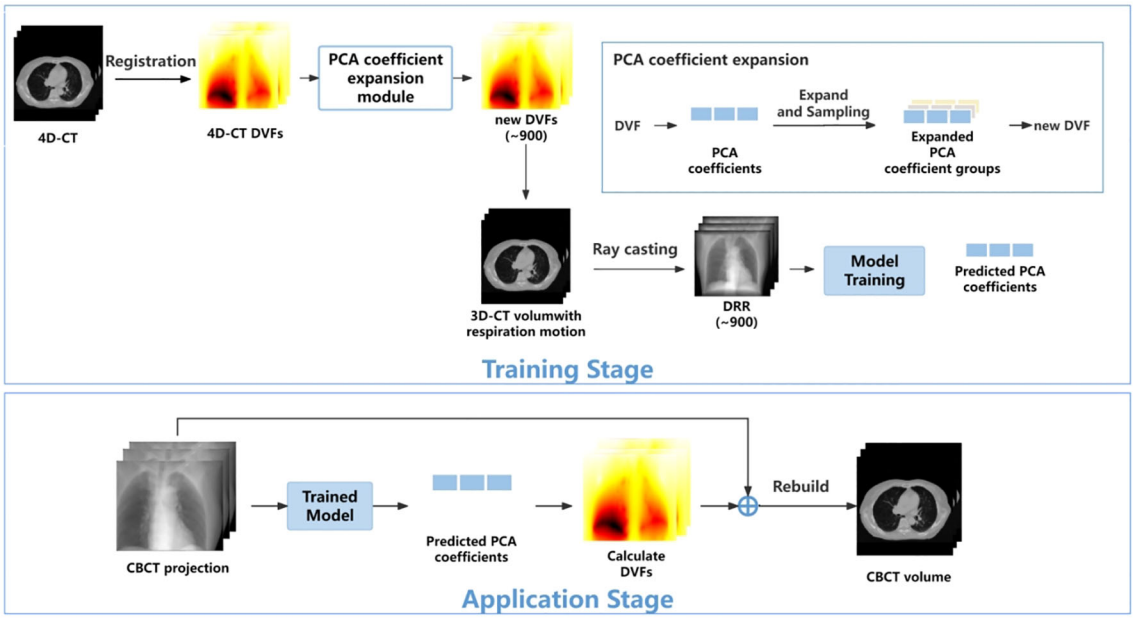


FIGURE 1
The workflow of the proposed method.

In the application stage, a single CBCT online projection that is measured at the same angle will be sent into the trained network. The network predicts three PCA labels to generate a phased 3D-CBCT. Then, more online projections (with different respiration amplitudes) will be continuously measured and sent into the network so that a whole respiration cycle will be covered. In this way, a full-cycle PCA label groups can be achieved and the whole 4D-CBCT. The entire process is performed on time. Below, we summarize our work into five parts: 1) motion modeling, 2) data processing, 3) network architecture, 4) loss function, and 5) experiment design.

2.1 Motion modeling

As mentioned above, the 4D-DVF is initially obtained from 4D-CT via deformable image registration (11–13). The 0% phase was selected as the reference phase to achieve the 4D-DVF. We used PCA, which is a commonly used data decoupling scheme for data dimension reduction (16), to extract DVF's feature label (e.g., the principle components/eigenvectors). For computational efficiency consideration, we select the first three PCA labels for mapping the DVFs. Table 1 illustrates the accuracy of DVF estimation relative to the number of PCA labels used. As expected, DVF accuracy improves with an increasing number of PCA labels. However, this also increases computational complexity. We found that by using the first three principal components, it already achieved 97.22% DVF information. Further increasing the PCA labels will not dramatically increase the information anymore. Therefore, we chose to discard the remaining PCA labels in our experiment.

The mapping relationship between the DVF and the PCA labels is given by Formula 1. Let the DVF size set be $3\times N_{voxelCT}$, where $N_{voxelCT}$ stands for 3D-CT voxel number; 3 stands for the 3D motion. The DVF will be linearly mapped by Equation 1:

$$DVF^{(i)} = \sum_{j=1}^{k} p_j^{(i)} q_j^{(i)} \quad (1)$$

Here, p and q stand for the eigenvectors and their corresponding PCA coefficients. Index i and j represent the respiration phase and eigenvectors, respectively.

2.2 Data processing

Being a regression task, ConvLSTM requires a large number of training data-set samples. In this study, we performed data augmentation and data enhancement. For data augmentation, we enlarged the simulated respiration amplitudes by a 15% interval up and down between two adjacent phases. This is because respiration is a time-

TABLE 1 PCA label versus DVF estimation accuracy.

Number of PCA labels	information (%)	Increment of information (%)
1	71.08	71.02
2	87.37	16.35
3	97.22	9.85
4	98.24	1.02
5	99.20	0.96
6	99.62	0.42
7	99.89	0.27
8	100.00	0.11

continuous physiological motion. The concept of the 4D-CBCT phase is an average reconstruction for projections in one re-binned phase. The lung will move across the re-binned interface between two adjacent phases. Our extended motion amount covers just a bit more than the average motion range (7). This is to make sure all the possible motion amplitude will be modeled for training data generation. We perform PCA label random sampling to generate 900 DRRs as a training data set.

For data enhancement, we considered the influence of quantum noise in the simulated DRRs. Given that quantum noise is typically a combination of Poisson and Gaussian noise (17), we constructed a linear noise combination as follows see Equation 2:

$$N = Poisson(I_0 exp(-p_n)) + Gassian(0, \sigma_e^2) \quad (2)$$

p_n is the noise-free signal line integral; the index N means the noise for each detector; I_0 is the X-ray projection intensity; and σ_e^2 represents background electronic noise. I_0 and σ_e^2 are set to be 10^5 and 10, respectively. DRR was then added to the simulated noise to achieve the real projected image.

We also implemented an intensity correction scheme to minimize the intensity mismatch between the simulated training DRRs versus the measured CBCT projections. The correction is given by Equation 3:

$$I_{DRR}^{\wedge} = (I_{DRR} - I_{Projection}) \times \frac{\sigma_{DRR}}{\sigma_{Projection}} + I_{DRR} \quad (3)$$

where $I_{DRR}^{\wedge}$ represents the corrected DRR intensity. I_{DRR} and σ_{DRR} represent the mean and the standard deviation of the original DRR intensity, and $I_{Projection}$ and $\sigma_{Projection}$ represent the mean and standard deviation of measured CBCT projection.

2.3 Network architecture

We use the ConvLSTM to explore the nonlinear mapping between DRRs and the PCA coefficients. The network architecture is illustrated in Figure 2. It contains a series of ConvLSTM cells and a regression layer. Conventional LSTM (18) contains a memory cell (C_t) and three gate control cells: 1) the forget grate (f_t), 2) the input gate (i_t), and 3) the output gate (o_t). C_t stores the foregone information, and the three gates update the cell. The LSTM sorts the relationships between all of the time flags; meanwhile, it ignores the internal information within each time flag. However, ConvLSTM (19), instead, explores the local features within each time flag via the convolutional operators. For the t^{th} ConvLSTM cell, the internal operations will be represented by (19), see Equations 4–9:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + b_f) \quad (5)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + b_o) \quad (6)$$

$$G_t = tanh(W_{xg} * X_t + W_{hg} * H_{t-1} + b_g) \quad (7)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ G_t \quad (8)$$

$$H_t = o_t \circ tanh(C_t) \quad (9)$$

σ is the sigmoid function, *tanh* stands for the TanHyperbolic function, $*$ and $\circ$ represent the convolutional operator and Hadamard product, respectively. X_t is the input of the current cell, and G_t is a candidate storage unit for information transmission. In addition, W and b denote convolution kernels and the bias terms. W and b have obvious meanings. For instance, W_{xo} is the input–output gate convolution kernel, while b_i is the input gate bias, etc.

Due to the characteristic of the convolutional operator, ConvLSTM can acquire both temporal and spatial information simultaneously (19–22). Our ConvLSTM network contains 40 hidden layers and 20 cell layers. Moreover, it has eight layers, kernel size is 3, padding is set as "valid", and the stride of the convolution kernel is 1.

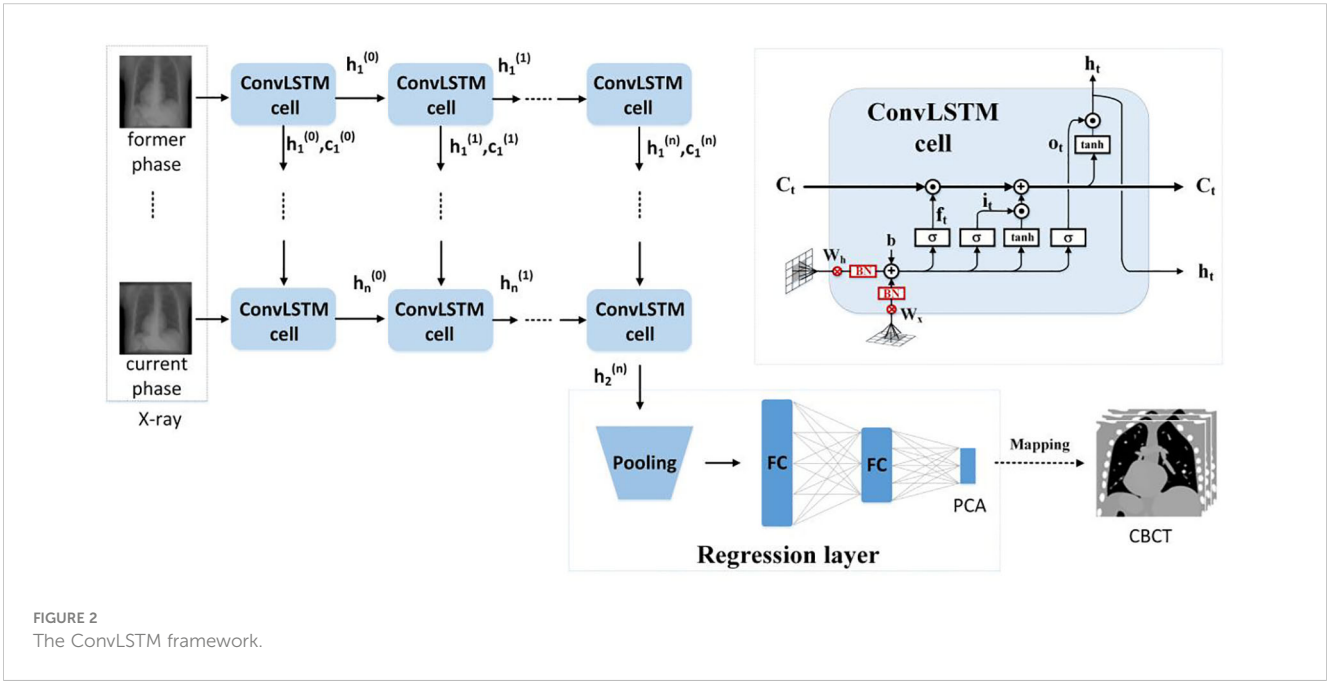


FIGURE 2
The ConvLSTM framework.

The regression layer uses the feature map generated from ConvLSTM to predict PCA coefficients. It contains a pooling layer with two fully connected layers. By using the dominant local information, the pooling layer reduces the computation cost. The pooling was set to twice the down-sampling, and the dimensions of the two completely connected layers are 1,024 and 3.

2.4 Loss function

The normalized mean square error builds the loss function and is given in Formula 5. The PCA coefficients (e.g., output labels in the network) in the loss function (see Equation 10) ensured that the first coefficient has the highest estimation accuracy.

$$Loss = \frac{1}{N}\sum_{i=1}^{N} \| w_{coeff} \circ (y_i - G(x_i, W)) \|_2 \quad (10)$$

N is the training sample number; $\| \|_2$ represents the L_2 norm, and o is the element-wise product. $G(x_i, W)$ is the output of the regression model. x_i is the i^{th} training image, y_i is the PCA coefficient, and W is the network parameters. w_{coeff} is the PCA coefficients weight, which is set to be $[\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}]$.

For model training, the ADAM optimizer was utilized with a dynamic learning rate, initially set at 0.001. The batch size was set to 8, and the training ran for 200 epochs. In an environment configured with Python 3.7 and an NVIDIA GeForce RTX 4080, training the data for 200 epochs took approximately 36 h.

2.5 Experiment design

For network performance evaluation, we use XCAT phantom and patient 4D-CT for the quantification. For testing, we simulated an on-board CBCT projection and then sent it into the pre-trained network to predict PCA coefficients. The quantification evaluation labels that we used are 1) the Mean Absolute Error (MAE), 2) the Normalized Cross Correlation (NCC), 3) the Multi-scale Structural Similarity(SSIM), 4) the Peak Signal-to-Noise Ratio (PSNR), 5) the Root Mean Squared Error (RMSE), and 6) the Absolute Percentage Error (MAPE). MAE is used to quantify the accuracy of regression models. y and $\hat{y}$ represent the label and the predicted value of the model, and i stands for the index of the regression model. We have in Equation 11:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} \left| \hat{y}^{(i)} - y^{(i)} \right| \quad (11)$$

In addition, NCC and SSIM (Multi-scale Structural Similarity Index Measure) are used to evaluate the quality of the reconstructed image. See Equations 12 and 13. S and T represent slice data with size of $H \times W$ of the original image and the reconstructed image, respectively. μ, δ, and δ^2 represent the mean, covariance, and variance of the slice image, respectively.

$$NCC = \frac{\sum_{k=1}^{C}\sum_{i=1}^{H}\sum_{j=1}^{W} |S(i,j) - \mu_s||T(i,j) - \mu_T|}{C\sqrt{\sum_{i=1}^{H}\sum_{j=1}^{W}(S(i,j) - \mu_s)^2 (T(i,j) - \mu_T)^2}} \quad (12)$$

$$SSIM = \frac{(2\mu_s\mu_T + C_1)(2\delta_{ST} + C_2)}{(\mu_s^2 + \mu_T^2 + C_1)(\delta_S^2 + \delta_T^2 + C_2)} = l(s,T) \cdot cs(s,T) \quad (13)$$

PSNR is defined based on MSE (Mean Squared Error). See Equations 14 and 15:

$$MSE = \frac{1}{N}\sum_j \| S(j) - T(j) \|^2 \quad (14)$$

$$PSNR = 10 * log10 \frac{MAX^2}{MSE} \quad (15)$$

N is the image pixel number. MAX is the maximum possible pixel value.

The definition of RMSE is given in Equation 16:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(S_{ij} - T_{ij}^*)^2}{H \times W}} \quad (16)$$

MAPE is the average ratio of the absolute difference between the predicted value and the true value to the true value. The definition of MAPE is given in Equation 17:

$$MAPE = \frac{1}{n}\sum_j \left| \frac{S_j - T_j}{T_j} \right| \quad (17)$$

3 Results

3.1 Network parameter optimization

Being a spatiotemporal sensitive network, the temporal continuous image amount that the network can handle for data training reflects its ability for accurate motion estimation. However, Figure 3 indicates that the model prediction accuracy is not dramatically influenced by the input image number. The MAE values fluctuate between 47 and 57, and the SSIM remains approximately 0.93. We found that the model achieves the best

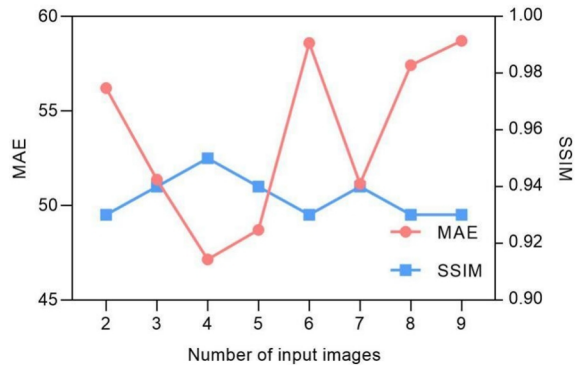


FIGURE 3
Input image quantity vs. MAE/SSIM of model prediction.

performance with four continuous temporal images with the lowest MAE of 47.15 and highest SSIM of 0.95.

The selection of hyper-parameters for the ConvLSTM network was a critical aspect, as these parameters significantly impact the prediction performance of the model. To determine the optimal configuration, we conducted a series of ablation experiments focusing on the number of hidden layers and cell layers within the ConvLSTM network. The experiment results in Figure 4 reveal that increasing the number of hidden layers decreased the MAE without significantly affecting computation time, although it did increase the number of parameters. Conversely, increasing the number of cell layers resulted in a slower decrease in MAE and an increase in computation time, with little change in parameter count. By balancing these factors, we determined that a configuration with 40 hidden layers and two cell layers provided the optimal trade-off, ensuring high prediction accuracy while maintaining computational efficiency.

3.2 Convergence of loss function

The convergence of the loss function is decided by the weightings. Table 2 shows the convergence comparison caused by different weightings. Their MAE and NCC values are also summarized in the table. We found that the second group weighting (e.g., $[2/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}]$) has the smallest first PCA label error. Meanwhile, this group also got the highest NCC.

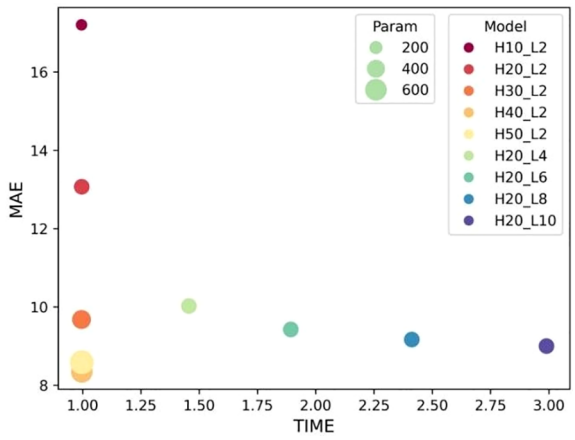


FIGURE 4
Influence of ConvLSTM cells on the model's prediction. "H" stands for the number of hidden layers; "L" denotes the number of cell layers.

TABLE 2 Weighting influence on MAE/NCC.

loss function weighting	MAE			NCC
	1st	2nd	3rd	
$[1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}]$	9.09	9.23	9.36	0.96
$[2/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}]$	6.29	9.81	10.06	0.98
$[\sqrt{3}/\sqrt{6}, \sqrt{2}/\sqrt{6}, 1/\sqrt{6}]$	8.01	9.19	10.01	0.96

Suitable choice of the pooling will also speed up loss function convergence. See Figure 5. The figure compared loss convergence curve with epoch with different pooling scheme such as Maximal pooling, Converlutional pooling, average pooling, and even no pooling at all. The results show that convolutional pooling achieves the best convergence performance. The pooling operation reduces the model's parameters, hence accelerating its convergence.

Suitable choice of pooling will also speed up loss function convergence. Figure 6 compares the loss convergence curve with different pooling schemes such as maximal pooling, convolutional pooling, average pooling, and even no pooling. The results show that convolutional pooling achieves the best convergence performance. The pooling operation reduces the model's parameters, hence accelerating its convergence.

3.3 XCAT simulation results

The XCAT phantom-based digital experiment was first performed. Four state-of-art network structures (e.g., CNN/Unet/ResNet/ConvLSTM) were tested with the phantom to compare their performances. As shown in Table 3, for the two test cases, the ConvLSTM outperforms other models in PCA coefficient prediction, especially for the first coefficient. The bold values provided in Table 3 means that ConvLSTM achieves the best PCA coefficient match compared with that of the ground truth for XCAT phantom. By utilizing PCA to reduce the dimensionality of the DVFs, the ConvLSTM network focuses on the most significant components of respiratory motion. This not only improves computational efficiency but also ensures that the network is learning the most relevant features for accurate motion prediction. Figure 6 presents the reconstructed results based on the PCA coefficients predicted by ConvLSTM versus CNN/UNet/ResNet. The reconstructed coronal plane and sagittal plane images and the different images between each reconstruction and the ground truth image are summarized in Figures 6A, B.

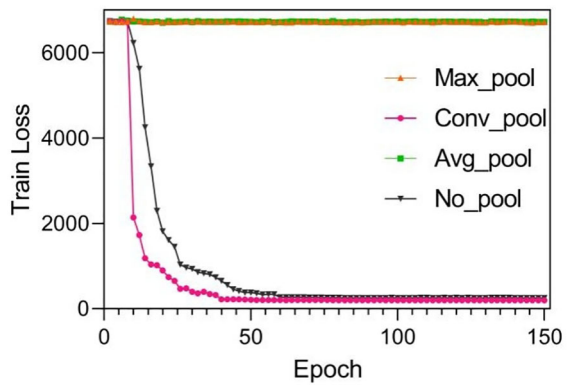


FIGURE 5
Training results by using different pooling optimizations. These pooling operations are set to twice the down-sampling, and the model only performs a single pooling operation.

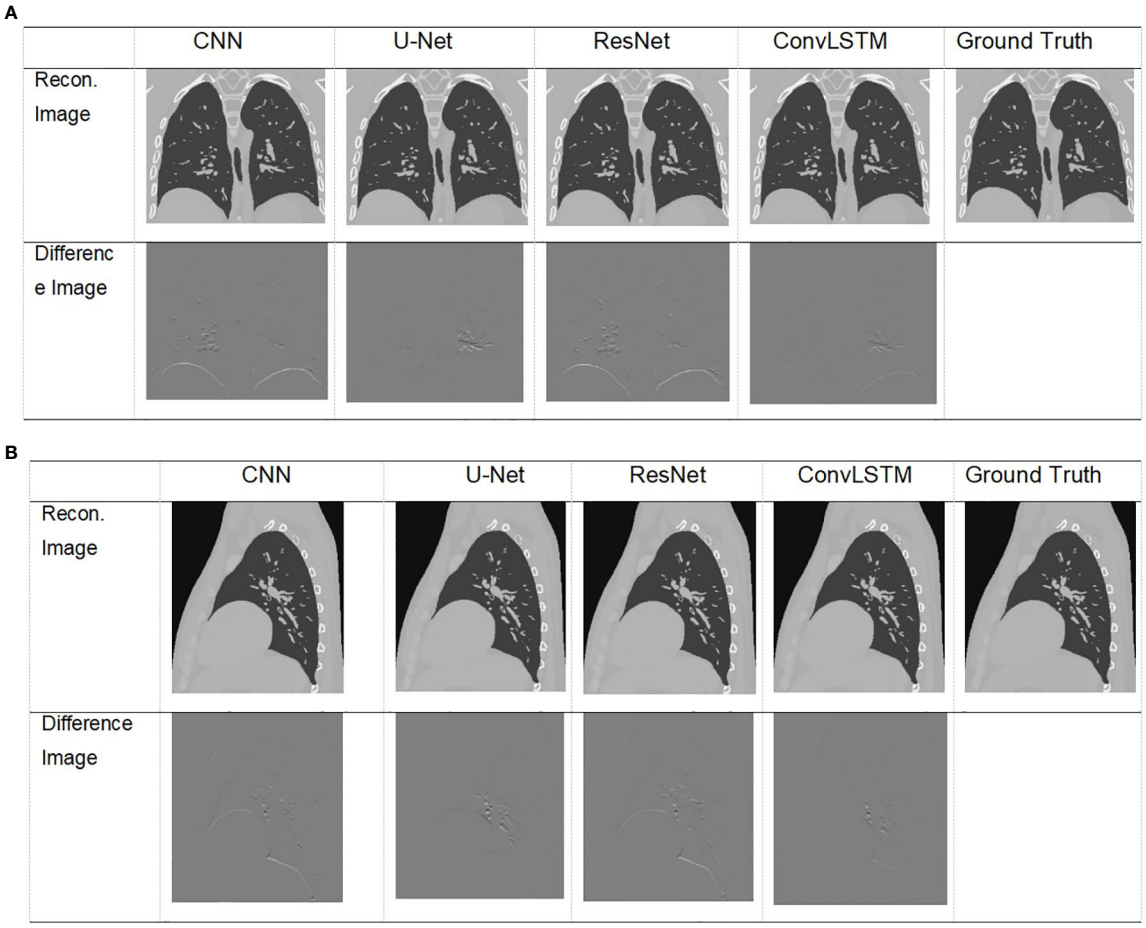


FIGURE 6
Visualization of images result of TestCase1 in different anatomical surfaces for each model with the training data generated from XCAT. **(A)** Coronal plane; **(B)** sagittal plane.

Table 4 summarizes the quantification evaluation comparison between each network. The results indicate that ConvLSTM outperformed other networks for all of the evaluation labels.

TABLE 3 Comparison of prediction results versus ground truth of XCAT data.

Model	PCA coefficients	
	Test Case1	Test Case2
CNN	[−1,121.3269 366.0489 −114.1392]	[5,736.9881 −391.8785 28.2141]
Unet	[−1,201.9354 394.0645 −66.5190]	[5,723.7266 −291.1034 56.5676]
ResNet	[−1,124.9048 378.9768 −60.8043]	[5,610.0141 −292.2644 90.5484]
ConvLSTM	[**−1,173.5433** 407.5900 −53.5265]	[**5,742.6875** −246.5791 181.1309]
Ground Truth	[**−1,163.5334** 454.6699 −78.2698]	[**5,787.5347** −258.0560 186.0697]

Values in bold indicate that our proposed method achieves the best quantification results compared to the ground truth.

3.4 Pilot clinical results

Table 5 shows two cases of the real and predicted first three PCA coefficients of the patient data results. It is well known that the higher the principal component order, the higher the PCA contribution rate. As can be seen from Table 5, the first principal component of the model based on ConvLSTM is closest to the true value, just as the bold values illustrated. Figure 7 shows the reconstructed coronal images based on the PCA coefficients predicted by CNN/UNet/ResNet and ConvLSTM network. We can see that all models have successfully reconstructed the

TABLE 4 Quantification comparison of prediction and reconstruction of each model on the coronal plane in XCAT TestData1.

Model	MAPE	PSNR	RMSE
CNN	0.2092	55.0287	0.0024
UNet	0.0464	62.0018	0.0015
ResNet	0.0628	56.6748	0.0025
ConvLSTM	0.0459	64.6742	0.0011

TABLE 5 Comparison of prediction results versus ground truth of patient data.

Model	PCA coefficients	
	Test Case1	Test Case2
CNN	[−676.5737 −36.4397 −26.6990]	[−87.5940 −117.7669 12.5394]
Unet	[−747.2873 −81.1768 −34.4381]	[−81.5389 −102.9164 11.1525]
ResNet	[−673.7071 −74.0461 −21.3157]	[−107.5652 −120.5355 14.5815]
ConvLSTM	[**−712.0823** −23.8481 −45.7298]	[**−99.6298** −112.0357 20.4654]
Ground Truth	[**−715.3792** −26.7257 −20.5198]	[**−101.5152** −127.8026 19.3956]

Values in bold indicate that our proposed method achieves the best quantification results compared to the ground truth.

anatomical structures, but ConvLSTM achieves the smallest different image to the ground truth. Table 6 summarizes the quantification evaluation comparison between each network on the clinical TestCase1. According to the result, we can see that ConvLSTM supplies a prediction with the minimum error compared with the ground truth, certified that ConvLSTM outperformed other networks. Traditional CNNs and other networks mainly focus on spatial features, which limits their ability to accurately model dynamic processes like respiratory motion. The ConvLSTM's ability to integrate convolutional operations with LSTM's temporal processing allows it to effectively model the temporal evolution of respiratory motion, leading to more accurate 4D-CBCT reconstructions.

TABLE 6 Quantification comparison of prediction and reconstruction of each model on the coronal plane in patient DataTest1.

Model	MAPE	PSNR	RMSE
CNN	0.2206	57.8427	0.0037
UNet	0.3313	53.6098	0.0060
ResNet	0.2706	55.4795	0.0048
ConvLSTM	0.0934	63.7294	0.0019

4 Discussion

In this study, we proposed a spatiotemporal consistent scheme via ConvLSTM and PCA motion modeling to estimate online 4D-CBCT. The network learns the motion features from patient 4D-CT with hundreds of simulated DRRs under a fixed angle. Both digital XCAT phantom experiments and pilot clinical studies were performed to prove the algorithm's efficiency. We compared our proposed method's efficiency with other popular networks such as CNN/Unet/ResNet. Quantification results indicate that ConvLSTM outperforms its competitors. ConvLSTM is an architecture that integrates Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks, enabling the application of convolution operations at each time step to effectively capture spatial information in temporal data. Compared to CNN, U-Net, and ResNet architectures, ConvLSTM can link the feature information of the current projection with that of adjacent projections, providing enhanced temporal and spatial feature connectivity. Hence, it will be able to supply enough information for motion estimation with temporal correlation.

In this work, our goal is to develop a real-time 4D-CBCT imaging model utilizing projection images with high temporal

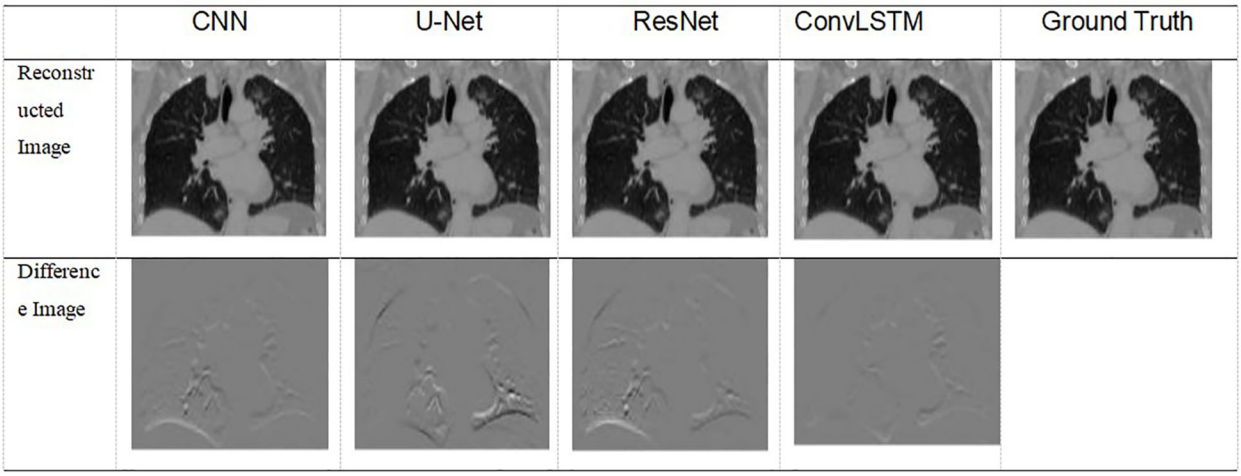


FIGURE 7
Visualization of images result of TestCase1 for each model with the training data generated from 4D-CT.

resolution. The model inference for PCA labels is remarkably fast, taking approximately 0.006 s for one projection. This rapid inference is critical for maintaining real-time processing capabilities, ensuring that the model can handle a continuous stream of projection images without significant latency. However, the reconstruction time for a single volume of 4D-CBCT is approximately 5 s on a personal desktop computer. While this is relatively fast given the complexity of the task, it underscores the computational demands associated with high-resolution 4D imaging. Our ongoing work focuses on optimizing this reconstruction time further, possibly through hardware acceleration or more efficient algorithms, to achieve even faster performance.

Despite the promising results, our study has several limitations that need to be addressed. First, the study relies on simulated data for training the network, including simulated respiratory motion and noise models. While these simulations aim to mimic real-world conditions, they may not fully capture the complexities of actual patient data, potentially affecting the model's performance in clinical settings. Second, the proposed model depends heavily on the consistency of the patient's respiration pattern between the initial 4D-CT scanning and the online treatment stages. Any significant variation in the patient's breathing pattern during treatment could impact the accuracy of the 4D-CBCT reconstruction. Third, the pilot clinical evaluation was conducted with a limited number of patients. Although the results were promising, a larger and more diverse patient cohort is necessary to validate the robustness of the proposed method.

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 authors.

Ethics statement

The studies involving humans were approved by Affiliated Hospital of Jiangsu University Review Board (#2016-034). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

HZ: Writing – original draft, Writing – review & editing, Methodology, Supervision. KC: Investigation, Software, Writing – original draft. XX: Data curation, Software, Writing – original draft. TY: Validation, Writing – review & editing. WS: Funding acquisition, Writing – review & editing. JD: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China under Grants (No. 61871208 and No. 62103366); in part by the Guangdong Basic and Applied Basic Research Foundation (No. 2023A1515011365), in part by the Shenzhen Science and Technology Program (No. JCYJ20220530153801003), in part by the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen (No. E010221003), and by the Zhenjiang City Science and Technology Plan Project (No. SH2021040); and in part by the Shenzhen High-level Hospital Construction Fund.

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.

The reviewer ZD declared a past co-authorship with the author HZ to the handling editor.

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.

References

1. Mah K, Chow B, Swami N, Pope A, Rydall A, Earle C, et al. Early palliative care and quality of dying and death in patients with advanced cancer. *BMJ Supportive Palliative Care*. (2023) 13:e74–7. doi: 10.1136/bmjspcare-2021-002893

2. Bertholet J, Vinogradskiy Y, Hu Y, Carlson DJ. Advances in image-guided adaptive radiation therapy. *Int J Radiat OncologyBiologyPhysics*. (2021) 110:625–8. doi: 10.1016/j.ijrobp.2021.02.047

3. Star-Lack J, Sun M, Oelhafen M, Berkus T, Pavkovich J, Brehm M, et al. A modified McKinnon-Bates (MKB MKB) algorithm for improved 4D cone-beam computed tomography (CBCT CBCT) of the lung. *Med Phys*. (2018) 45:3783–99. doi: 10.1002/mp.13034

4. Dang J, Ying F-F, Dai C, Chen D, Wang J. Simultaneous 4D-CBCT reconstruction with sliding motion constraint. *Med Phys*. (2016) 43(10):5453–63. doi: 10.1118/1.4959998

5. Chapman CH, McGuinness C, Gottschalk AR, Yom SS, Garsa AA, Anwar M, et al. Influence of respiratory motion management technique on radiation pneumonitis risk with robotic stereotactic body radiation therapy. *J Appl Clin Med Phys*. (2018) 19:48–57. doi: 10.1002/acm2.12338

6. Xu Y, Yan H, Ouyang L, Wang J, Zhou L, Cervino L, et al. A method for volumetric imaging in radiotherapy using single x-ray projection. *Med Phys*. (2015) 42:2498–509. doi: 10.1118/1.4918577

7. Li R, Lewis JH, Jia X, Gu X, Folkerts M, Men C, et al. 3D tumor localization through real-time volumetric x-ray imaging for lung cancer radiotherapy. *Med Phys*. (2011) 38:2783–94. doi: 10.1118/1.3582693

8. Zhang YY, Yin F-F, Segars WP, Ren L. A technique for estimating 4D-CBCT using prior knowledge and Limited-angle projections. *Med Phys*. (2013) 40 (12):1217011-1–1217011:16. doi: 10.1118/1.4825097

9. Wei R, Zhou F, Liu B, Bai X, Fu D, Li Y, et al. Convolutional neural network (CNN) based three dimensional tumor localization using single X-ray projection. *IEEE Access*. (2019) 7:37026–38. doi: 10.1109/Access.6287639

10. Wei R, Zhou F, Liu B, Bai X, Fu D, Liang B, et al. Real-time tumor localization with single x-ray projection at arbitrary gantry angles using a convolutional neural network (CNN). *Phys Med Biol*. (2020) 65(6):065012. doi: 10.1088/1361-6560/ab66e4

11. Brown MS, Mcnitt-Gray MF, Goldin JG, Suh RD, Sayre JW, Aberlee DR. Patient-specific models for lung nodule detection and surveillience in CT images. *IEEE Trans On Med Imaging MI*. (2001) 20(12):1242–50. doi: 10.1109/42.974919

12. Chen M, Cao K, Zheng Y, Siochi RAC. Motion-compensated mega-voltage cone beam CT using the deformation derived directly from 2D projection images. *IEEE Trans On Med Imaging MI*. (2013) 32(8):1365–75. doi: 10.1109/TMI.2012.2231694

13. Lin T, Li R, Tang X, Dy JG, Jiang SB. Markerless gating for lung cancer radiotherapy based on machine learning techniques. *Phys Med Biol*. (2009) 54:1555–63. doi: 10.1088/0031-9155/54/6/010

14. Unberath M, Zaech JN, Lee SC, Bier B, Fotouhi J, Armand M, et al. "DeepDRR – A catalyst for machine learning in fluoroscopy-guided procedures". In Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (editors) *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018. Lecture Notes in Computer Science (LNIP, Vol. 11073, pp. 98-106)*. Springer, Cham (2018). doi: 10.1007/978-3-030-00937-3_12

15. Unberath M, Zaech JN, Gao C, Bier B, Navab N. Enabling machine learning in X-ray-based procedures via realistic simulation of image formation. *Int J Comput Assisted Radiol Surg*. (2019) 14(9):1517–28. doi: 10.1007/s11548-019-02011-2

16. Gewers FL, Ferreira GR, De Arruda HF, Silva FN, Comin CH, Amancio DR, et al. Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys (CSUR)*. (2018) 54:1–34. doi: 10.1145/3447755

17. Dang J, Ouyang L, Gu X, Wang J. Deformation vector fields (DVF)-driven image reconstruction for 4D-CBCT. *J Xray Sci Technol*. (2013) 40:457–7. doi: 10.3233/XST-140466

18. Hochreiter S, Schmidhuber J. "Long short-term memory," in *Neural Computation*. (1997) pp. 1735–80. doi: 10.1162/neco.1997.9.8.1735

19. Shi X, Chen Z, Wang H, Yeung DY, Wong W, Woo W. "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting." *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*. MIT Press. (2015) 1. doi: 10.5555/2969239.2969329

20. Wang Q, Huang Y, Jia W, Xiangjian HE, Blumenstein M, Lyu S, et al. FACLSTM: ConvLSTM with focused attention for scene text recognition. *Sci China Inf Sci*. (2020) 63:120103:1–120103:14. doi: 10.1007/s11432-019-2713-1

21. Kong F, Deng J, Fan Z. Gesture recognition system based on ultrasonic FMCW and ConvLSTM model. *Measurement*. (2022) 190:110743. doi: 10.1016/j.measurement.2022.110743

22. Wang W-Y, Li H-C, Deng Y-J, Shao L-Y, Lu X-Q, Du Q. Generative adversarial capsule network with ConvLSTM for hyperspectral image classification. *IEEE Geosci Remote Sens Lett*. (2020) 18:523–7. doi: 10.1109/LGRS.8859

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 06 August 2024
DOI 10.3389/fonc.2024.1374373

Check for updates

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Sarada Prasad Dakua,
Hamad Medical Corporation, Qatar
Emina Talakic,
Medical University of Graz, Austria

*CORRESPONDENCE
Hongzhen Wu
eywuhongzhen@scut.edu.cn
Xinhua Wei
eyxinhuawei@163.com

†These authors have contributed equally to this work

RECEIVED 22 January 2024
ACCEPTED 18 July 2024
PUBLISHED 06 August 2024

CITATION
Liang Y, Wu H and Wei X (2024) Development and validation of a CT-based nomogram for accurate hepatocellular carcinoma detection in high risk patients.
Front. Oncol. 14:1374373.
doi: 10.3389/fonc.2024.1374373

COPYRIGHT
© 2024 Liang, Wu and Wei. 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.

Development and validation of a CT-based nomogram for accurate hepatocellular carcinoma detection in high risk patients

Yingying Liang [1,2], Hongzhen Wu [2*†] and Xinhua Wei [1,2*†]

[1]The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China, [2]Department of Radiology, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China

Purpose: To establish and validate a CT-based nomogram for accurately detecting HCC in patients at high risk for the disease.

Methods: A total of 223 patients were divided into training (n=161) and validation (n=62) cohorts between January of 2017 and May of 2022. Logistic analysis was performed, and clinical model and radiological model were developed separately. Finally, a nomogram was established based on clinical and radiological features. All models were evaluated using the area under the curve (AUC). DeLong's test was used to evaluate the differences among these models.

Results: In the multivariate analysis, gender (p = 0.014), increased Alpha-fetoprotein (AFP) (p = 0.017), non-rim arterial phase hyperenhancement (APHE) (p = 0.011), washout (p = 0.011), and enhancing capsule (p = 0.001) were the independent differential predictors of HCC. A nomogram was formed with well-fitted calibration curves based on these five factors. The area under the curve (AUC) of the nomogram in the training and validation cohorts was 0.961 (95%CI: 0.935~0.986) and 0.979 (95% CI: 0.949~1), respectively. The nomogram outperformed the clinical and the radiological models in training and validation cohorts.

Conclusion: The nomogram incorporating clinical and CT features can be a simple and reliable tool for detecting HCC and achieving risk stratification in patients at high risk for HCC.

KEYWORDS

hepatocellular carcinoma, diagnosis, nomogram, CT, model

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer and the second leading cause of cancer-related mortality worldwide (1, 2). The majority of HCCs occur in patients with hepatitis viruses, alcohol abuse, non-alcoholic fatty liver disease and liver cirrhosis, which are considered as high-risk factors for developing HCC (3, 4). The prognosis of patients with advanced HCC is poor, while the prognosis of patients with early-stage HCC is much better due to effective and curative treatment options, such as resection, percutaneous ablation, or orthotopic liver transplantation (5). Therefore, early detection and accurate diagnosis are important in managing patients with early-stage HCC.

HCC is a unique malignancy that can be diagnosed noninvasively with imaging (6). Once a definitive diagnosis is established, patients may receive priority on the liver transplantation waiting list without mandated pathologic confirmation (7). Although ultrasound (US) is recommended as the main surveillance tool for HCC by most clinical guidelines, its sensitivity for early-stage HCC ranges from 45% to 63%, especially in patients with advanced cirrhosis (8).

Magnetic resonance imaging (MRI) is crucial for the routine diagnosis and evaluation of HCC, however, their wide use may be limited due to the higher cost than that of computed tomography (CT). Thus, CT forms the keystone in the diagnosis of HCC and is recommended as a first-line diagnostic tool for HCC due to its short acquisition time and high spatial resolution. In cirrhotic livers, alterations in hepatic blood flow may cause atypical enhancement patterns (9). Currently, it remains extremely challenging to distinguish HCC from non-HCC lesions because some HCCs show atypical imaging features and some non-HCC lesions may mimic HCC in imaging in patients at high risk of HCC, which may lead to inappropriate treatment (10, 11). In recent years, radiomics, as an emerging methodology in medicine, has shown promising results in detecting HCC in patients at high risk (12, 13). However, the clinical promotion has been limited because it requires special commercial software (14, 15). Thus, there is a need for a simple and precise preoperative approach to detect HCC in patients at high risk.

Our study aimed to identify objective clinical factors and radiological features associated with HCC diagnosis and to develop a new CT-based nomogram that accurately predicts risk in high-HCC risk patients.

Materials and methods

This retrospective study was approved by the Institutional Review Board with waived requirement for informed consent (Ethical Board Approval Number: "K-2022–004-01").

A total of 649 patients with liver lesions were enrolled in the study between January of 2017 and May of 2022. Inclusion criteria included: (a) patients who had CT examinations; (b) patients with hepatitis B virus infection, or liver cirrhosis of any cause confirmed histologically or typical radiologically; and (c) those with a pathological diagnosis confirmed within ten days after CT by biopsy or surgical diagnosis. Exclusion criteria included: (a) inadequate confirmation of pathological findings or biopsy (n = 342); (b) treatment given before imaging or surgery (n = 46); (c) inadequate serological markers (n = 20); and (d) inadequate CT data or poor image quality due to movement during the examination (n = 18). Finally, 223 patients were enrolled and randomly assigned to either the training cohort (n = 161) or the validation cohort (n = 62) at a ratio of 7:3 (Figure 1).

Baseline demographic data, including age, gender, and laboratory parameters, including Alpha-fetoprotein (AFP), serum total bilirubin, total plasma protein, prothrombin time, and blood platelet levels were collected from each patient's medical records.

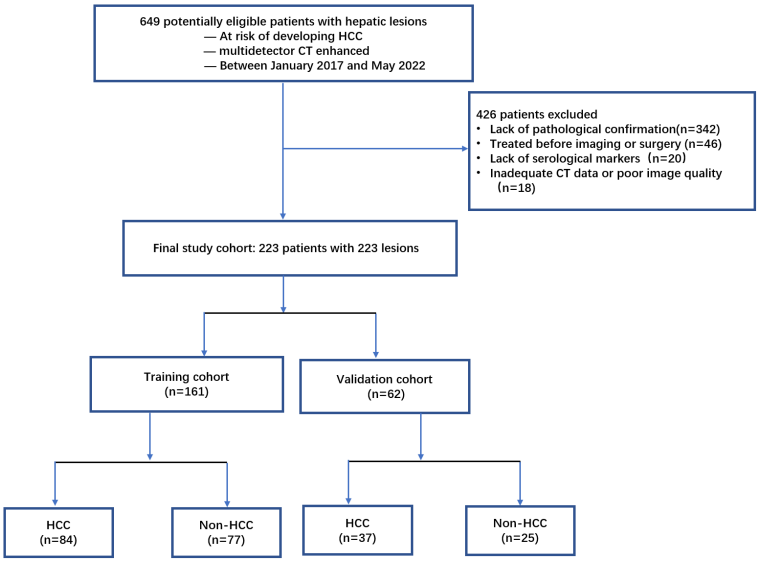


FIGURE 1
Flowchart of the study population.

CT technique

All dynamic acquisition of contrast-enhanced CT exams were performed on a 16-detector CT scanner (Toshiba Aquilion One), 64-detector CT systems (Philips Brilliance), and a 320-detecter CT scanner (Toshiba Aquilion One). The scanning parameters were as follows: 5 mm section thickness, 0.5 s rotation time, 0.9 pitch, 200 mAs tube current, 120 kVP tube voltage, and a 512 × 512 matrix. After an unenhanced CT scan, contrast agent (iodipamide, 370 mg I/mL, Bracco) was injected into the antecubital vein at a rate of 3.5–4.0 mL/s based on their weight (2.0 mL/kg body weight, with a maximal dose of 180ml), followed by 20 mL of saline solution using a power injector. We then obtained the arterial phase (AP, 35–40 seconds), portal venous phase (PVP, 50–60 seconds), and equilibrium phase (EP, 120–250 seconds), respectively.

CT image analysis

Two abdominal radiologists (who had 6 and 15 years of experience, respectively) retrospectively and independently reviewed the CT images. The readers were blinded to the final pathological diagnoses of the lesions. They assessed the presence of major features of HCC, including non-rim arterial phase hyperenhancement (APHE), non-peripheral washout, and enhancing capsule according to the Liver Imaging Reporting and Data System (LI-RADS) version 2018 (16). They also evaluated the number of lesions, size of the largest lesion, and the presence or absence of necrosis, satellite lesions, internal arteries, and non-enhancing capsules. In addition, tumor with pathological result was evaluated if the liver has different or multiple lesions.

Statistical analysis

Continuous variables were expressed as the mean ± standard deviation or median, and compared using the Mann–Whitney U test. Categorical variables were expressed as numbers (percentages) and compared using the chi-square test. Clinical and imaging factors, including age, gender, AFP, serum total bilirubin, total plasma protein, prothrombin time, blood platelet levels, tumor size, non-rim arterial phase hyperenhancement (APHE), non-peripheral washout, enhancing and non-enhancing capsule, necrosis, satellite lesions, and internal arteries, were analyzed using the stepwise regression forward method to select the significant independent predictors in the training cohort. Factors whose P values were less than 0.05 in the univariable analysis were inputted into the multivariable logistic regression analysis to identify the independent predictors of HCC. The clinical model (model 1), radiological model (model 2) and clinical-radiologic nomogram (model 3) were constructed by integrating significant clinical factors, significant radiological factors, combined clinical and radiological factors, respectively. The nomogram is based on proportionally converting each regression coefficient in multivariate logistic regression to a 0- to 100-point scale. The points, according to the b coefficient (absolute value) in different variables, are converted to predicted probabilities. The discrimination power of the nomogram was assessed by calibration curves which were assessed with a 1,000 bootstrap resample to measure the accuracy of the nomogram in the training and validation cohorts. A receive operating characteristic (ROC) curve was conducted to evaluate the performance of the different prediction models using area under the curve (AUC). Sensitivity, specificity and accuracy, were then calculated and the DeLong test was used to compare the models' performances. A decision curve analysis (DCA) was applied to explore clinical usefulness.

The interobserver agreement was analyzed for each feature by using kappa (k) statistics. All statistical analyses were performed using SPSS software (Version 25.0, Chicago, IL, USA) and R software (version 3.6.1). A two-sided p-value of < 0.05 was considered significant in all statistical tests.

Results

Demographic data and laboratory parameters

The baseline demographic characteristics of the training and validation cohorts are summarized in Table 1. Among the 161 patients in the training cohort, the mean age was 55.28 years, with a range of 20–87 years, and 115 patients (71.4%) were men. Pathologic assessment revealed 84 and 77 cases of HCC and non-HCC lesions, respectively. The malignant non-HCC lesions included cholangiocarcinoma (n = 51), metastasis (n = 2), combined hepatocellular-cholangiocarcinoma (n =1), and epithelioid angiomyolipoma (n =1). The benign non-HCC lesions included focal nodular hyperplasia (n = 12), dysplastic nodules (n =7), and hepatocellular adenoma (n =3). Most lesions were surgically diagnosed, although 16.8% (27/161) lesions were diagnosed after a biopsy. Among the 62 patients in the validation cohort, the mean age was 51.10 years, with a range of 21–82 years, and 44 patients (71.0%) were men. Pathologic assessment revealed 37 and 25 cases of HCC and non-HCC lesions, respectively. The malignant non-HCC lesions included cholangiocarcinoma (n = 13), and metastasis (n = 1). The benign non-HCC lesions included focal nodular hyperplasia (n =8), dysplastic nodules (n =2), and hepatocellular adenoma (n =1). There were 16.1% (10/62) lesions confirmed by biopsy, and the remaining lesions were confirmed by surgery.

Interobserver agreement for CT features

The CT features of all patients are shown in Table 2. Patients with HCC were more likely to have non-rim APHE ($p < 0.001$), washout ($p < 0.001$), enhancing capsules ($p < 0.001$), necrosis ($p < 0.001$), and internal arteries ($p < 0.001$) in the training cohort.

The interobserver agreement for CT features was fair for the maximum dimension (k = 0.43), substantial for the nonenhancing capsule (k = 0.75) and almost perfect for non-rim APHE (k = 0.89), washout (k = 0.90), necrosis (k = 0.90), internal arteries (k = 0.90), enhancing capsules (k = 0.95), and satellite lesions (k = 0.95).

10.3389/fonc.2024.1374373

TABLE 1 Patient characteristics in the training and validation cohorts.

	Training dataset (n=161)			Validation dataset (n =62)		
	Non-HCC (n=77)	HCC (n=84)	P	Non-HCC (n=25)	HCC (n=37)	P
Gender(n)						
Male	41(53.2%)	74(88.1%)	<0.001*	11(44.0%)	33(89.2%)	<0.001*
Female	36(46.8%)	10(11.9)		14(56.0%)	4(10.8%)	
Age(years)	55(42,64)	57(50,63.5)	0.509	42(32,60)	57(45,65)	0.045
Serum total bilirubin(g/L)	15.7(11.1,25)	34.5(19.25,67.1)	<0.001*	16.1(11.8,72.2)	24(17.7,35)	0.134
Total plasma protein(g/L)	67.1(60.6,73.4)	52.9(17.7,65)	<0.001*	72.8(40.2,77.7)	64.4(43,71.7)	0.166
Prothrombin time(s)	13.6(13,14.2)	13.6(12.7,14.6)	0.755	13.2(12.8,14)	14(12.8,14.8)	0.175
Blood platelet(10g/L)	245(196,290)	171(135.5,229.5)	<0.001*	258(205,341)	172(117,210)	<0.001*
AFP(ng/ml)						
Negative	62(80.5%)	18(21.4%)	<0.001*	21(84.0%)	7(18.9%)	<0.001*
Positive	15(19.5%)	66(78.6%)		4(16.0%)	30(81.1%)	
Maximum dimension (cm)	5.6(3.4,8.4)	5.15(3.5,7.7)	0.907	6.8(4.7,10.8)	6.8(4.5,8.7)	0.434

AFP, Alpha-fetoprotein; HCC, hepatocellular carcinoma. * P<0.05, significant difference between both groups.

TABLE 2 CT features in the training and validation cohorts.

	Training dataset (n=161)			Validation dataset (n =62)		
	Non-HCC (n=77)	HCC (n=84)	P	Non-HCC (n=25)	HCC (n=37)	P
Non-rim APHE						
Negative	43(55.8%)	7(8.3)	<0.001*	11(44.0%)	5(13.5%)	0.007*
Positive	34(44.2%)	77(91.7%)		14(56.0%)	32(86.5%)	
Washout						
Negative	67(87.0%)	18(21.4%)	<0.001*	19(76.0%)	12(32.4%)	0.001*
Positive	10(13.0%)	66(78.6%)		6(24.0%)	25(67.6%)	
Enhancing capsule						
Negative	69(89.6%)	29(34.5%)	<0.001*	25(100.0%)	7(18.9%)	<0.001*
Positive	8(10.4%)	55(65.5%)		0(0.00)	30(81.1%)	
Necrosis						
Negative	27(35.1%)	5(6.0%)	<0.001*	13(52.0%)	2(5.4%)	<0.001*
Positive	50(64.9%)	79(94.0%)		12(48.0%)	35(94.6%)	
Satellite lesions						
Negative	54(70.1%)	54(64.3%)	0.431	19(76.0%)	24(64.9%)	0.351
Positive	23(29.9%)	30(35.7%)		6(24.0%)	13(35.1%)	
Internal artery						
Negative	37(48.1%)	7(8.3%)	<0.001*	7(28.0%)	4(10.8%)	0.162
Positive	40(51.9%)	77(91.7%)		18(72.0%)	33(89.2%)	

(Continued)

TABLE 2 Continued

	Training dataset (n=161)			Validation dataset (n =62)		
	Non-HCC (n=77)	HCC (n=84)	P	Non-HCC (n=25)	HCC (n=37)	P
Nonenhancing "capsule"						
Negative	71(92.2%)	80(95.2%)	0.426	24(96.0%)	36(97.3%)	1.000
Positive	6(7.8%)	4(4.8%)		1(4.0%)	1(2.7%)	

HCC, hepatocellular carcinoma; APHE, arterial phase hyperenhancement. *P<0.05, significant difference between both groups.

Development of the prediction model

Multiple logistic regression analysis revealed that gender (OR = 0.095, 95% CI = 0.015–0.615, p = 0.014), increased AFP (OR = 5.683, 95% CI = 1.358–23.778, p = 0.017), non-rim APHE (OR = 9.619, 95% CI = 1.683–54.971, p = 0.011), washout (OR = 8.231, 95% CI = 1.614–41.978, p = 0.011), and enhancing capsules (OR = 19.136, 95% CI = 3.478–105.291, p = 0.001) were HCC predictors (Table 3). A nomogram was constructed from these variables to construct a quantitative and predictive tool (Figure 2). Among these significant features, enhancing capsule showed the highest OR and was the best predictor of HCC according to the nomogram.

Predictive performance and validation of the prediction model

The independent risk factors (including gender and increased AFP) for HCC were used to construct the clinical model. The clinical model predicted HCC with an AUC of 0.838 (95% CI, 0.778–0.897) in the training cohort, and 0.873 (95% CI, 0.782–0.964) in the validation cohort (Figure 3). The sensitivity, specificity, and accuracy of the model for the training cohort were 80.5%, 78.6%, and 0.795, respectively, whereas those of the validation cohort were 84.0%, 81.1%, and 0.823, respectively (Table 4).

TABLE 3 Multivariable analysis with logistic regression in the training cohort including clinical and radiological variables.

	B	P value	OR	95% CI	
				Lower	Upper
Serum total bilirubin	0.015	0.185	1.016	0.993	1.039
Total plasma protein	-0.012	0.439	0.988	0.959	1.018
Blood platelet	-0.009	0.065	0.991	0.981	1.001
Gender	-2.357	0.014*	0.095	0.015	0.615
AFP	1.737	0.017*	5.683	1.358	23.778
Non-rim APHE	2.264	0.011*	9.619	1.683	54.971
Washout	2.108	0.011*	8.231	1.614	41.978
Enhancing capsule	2.952	0.001*	19.136	3.478	105.291
Necrosis	2.052	0.051	7.784	0.991	61.176
Internal artery	1.569	0.096	4.801	0.756	30.485

CI, confidence interval; OR, odds ratio.

Three radiological characteristics, including non-rim APHE, washout, and enhancing capsules, were used to construct the radiological model. The model yielded AUC of 0.917 (95% CI, 0.874–0.961) and 0.918 (95% CI, 0.854–0.983) in the training and validation cohort, respectively (Figure 3). The sensitivity, specificity, and accuracy of the model for the training cohort were 87.0%, 86.9%, and 0.870, respectively, whereas those of the validation cohort were 76.0%, 81.1%, and 0.790, respectively (Table 4).

The clinical-radiologic nomogram was constructed which combined the two clinical and three radiological characteristics with different score based on their β coefficient, yielded the AUC values of 0.961 (95% CI: 0.935–0.986) with a sensitivity of 97.4%, a specificity of 81.0%, and 0.979 (95% CI: 0.949–1) with a sensitivity of 100.0%, a specificity of 91.9%, in the training and validation cohorts (Figure 3). The clinical-radiologic nomogram showed the highest AUC than in the clinical and radiological model in the training cohort (p = 0.041 [clinical vs. radiological model], 0.001[clinical vs. clinical-radiologic nomogram], 0.005 [radiological vs. clinical-radiologic nomogram], respectively), and validation cohort (p = 0.421 [clinical vs. radiological model], 0.010 [clinical vs. clinical-radiologic nomogram], 0.012 [radiological model vs. clinical-radiologic nomogram], respectively).

The calibration curve of the prediction model is shown in Figure 4 indicating that the nomogram is in good agreement with the observations of HCC diagnosis in patients at high risk for HCC.

Decision curve analysis

The decision curve analysis (DCA) for the nomogram revealed that our prediction nomogram was better able to predict HCC potential than either the treatment or no treatment schemes with the threshold probability >0.1 in the training cohort and in the range between 0 to 1.0 in the validation cohort (Figure 4).

Discussion

In the present study, it showed that gender, elevated AFP level, positive non-rim APHE, washout, and enhancing capsule were independent, significant parameters predicting HCC. Based on

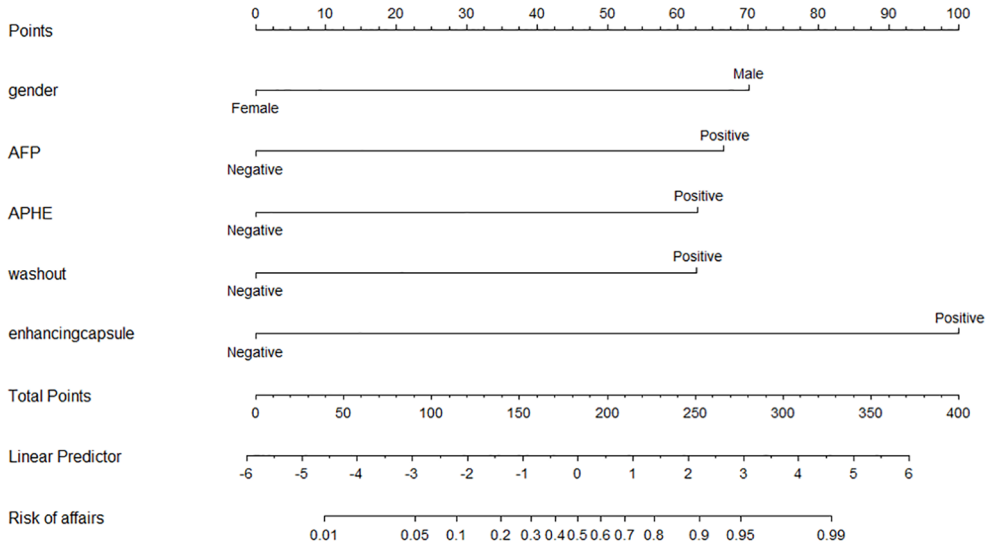


FIGURE 2
Nomogram for estimating the probabilities of HCC.

these significant clinical factors and radiological features, we constructed and validated a noninvasive nomogram to accurately identify HCC from other hepatic lesions in patients at high risk of HCC. This nomogram showed good predictive ability of HCC in both training (AUC = 0.961) and validation cohorts (AUC = 0.979), which outperformed the clinical and radiological model with a good calibration and clinical applicability. This nomogram was easy to use and facilitated the preoperative detection of HCC for clinicians in order to avoid overdiagnosis and overtreatment.

Our study indicated that enhancing capsule was the best single predictor for diagnosing HCC according to the nomogram, consistent with previous reports (17, 18). Enhancing capsule reportedly has the highest specificity (88%–96%) for diagnosing HCC (17–19). Although enhancing capsule is not evaluated according to the American Association for the Study of Liver Diseases (AASLD) (20) and European Association for the Study of the Liver (EASL) (21) guidelines, it is recognized as a major feature in the LI-RADS algorithm, which could increase the sensitivity (18, 22) and specificity (23) of diagnosing HCC.

Among the three CT features, non-rim APHE had the highest sensitivity (91.7%, 77/84) for diagnosing HCC, which reflects the neoangiogenesis and accelerates the carcinogenesis (24). Although non-rim APHE is considered a crucial imaging feature for diagnosing HCC (24), it is non-specific because it could also be detected in other malignant or benign lesions (25). Recent research has reported that non-rim APHE had the highest sensitivity (85%–94%) (17, 18) but lower specificity (58%–64%) (17, 26) for diagnosing HCC, which was consistent with our results. Meanwhile, Granata et al. confirmed that non-rim APHE could be found in most of the dysplastic nodules (70% [17/24]) in their study population (25). Therefore, attention should be taken in using this feature alone which should lead to false HCC diagnosis.

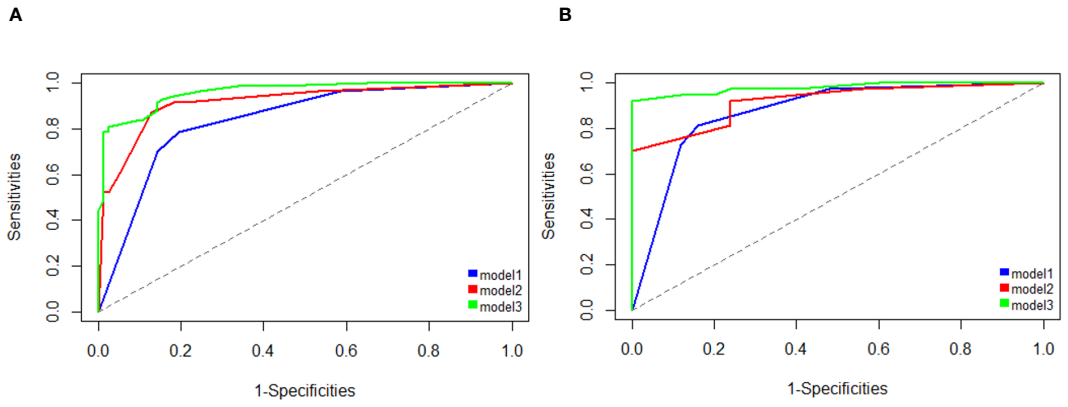


FIGURE 3
Area under the receiver operating characteristic curve (AUC) analysis shows better performance for detecting HCC using the clinical-radiologic nomogram (model 3) compared with the clinical model (model 1) and radiologic model (model 2) on training cohort **(A)** and testing cohorts **(B)**.

TABLE 4 The specific predictive performances of models for HCC.

Model	Cohort	Criterion			
		AUC (95% CI)	Sensitivity	Specificity	Accuracy
Clinical	Training	0.838 (0.778~0.897)	0.805 (0.717~0.894)	0.786 (0.698~0.873)	0.795 (0.793~0.797)
	Testing	0.873 (0.782~0.964)	0.840 (0.696~0.984)	0.811 (0.685~0.937)	0.823 (0.818~0.827)
Radiological	Training	0.917 (0.874~0.961)	0.870 (0.795~0.945)	0.869 (0.797~0.941)	0.870 (0.868~0.871)
	Testing	0.918 (0.854~0.983)	0.760 (0.593~0.927)	0.811 (0.685~0.937)	0.790 (0.785~0.796)
Clinical-radiologic nomogram	Training	0.961 (0.935~0.986)	0.974 (0.938~1)	0.810 (0.726~0.893)	0.888 (0.887~0.889)
	Testing	0.979 (0.949~1)	1.000 (1~1)	0.919 (0.831~1)	0.952 (0.95~0.953)

Washout is the third independent CT feature for diagnosing HCC. There is a discrepancy in washout performance. De Gaetano et al. (17) reported that washout had a high sensitivity (88.2%), but a low specificity (42.3%) for diagnosing HCCs. However, Sangiovanni et al. (26) reported that washout was an effective means of distinguishing HCC from other liver lesions, with a high specificity (100.0%) but a low sensitivity (53.0%). Interestingly, we found that washout was an important independent risk factor of HCC, with a high specificity (87.0%) but a moderate sensitivity (78.6%). Many recent studies have demonstrated that washout combined with non-rim APHE could increase specificity and positive predictability for early HCC diagnosis (18).

AFP is the most widely used tumor marker for diagnosis and evaluation of HCC in clinical practice (27). However, the AASLD does not recommend AFP for the early detection of HCC (20). Previous research has reported that elevated AFP occurred in only 40–65% of HCC patients, while others had normal AFP levels, particularly during the early stages of the disease (28). In addition, many other studies found that an elevated AFP level also occurred in other malignancies or benign liver lesions (29). In our study, gender also had a detrimental effect on HCC diagnosis. Males were more prone to hepatocarcinogenesis, with a prognosis that is worse than in females.

Currently, various imaging models for HCC detection have been described in the literature, especially radiomics models. A study involving 102 patients with liver tumors defined as LR-M based on LI-RADS developed a MRI-based radiomics model to classify HCC and non-HCC tumors with AUC of 0.884 and 0.873, respectively in

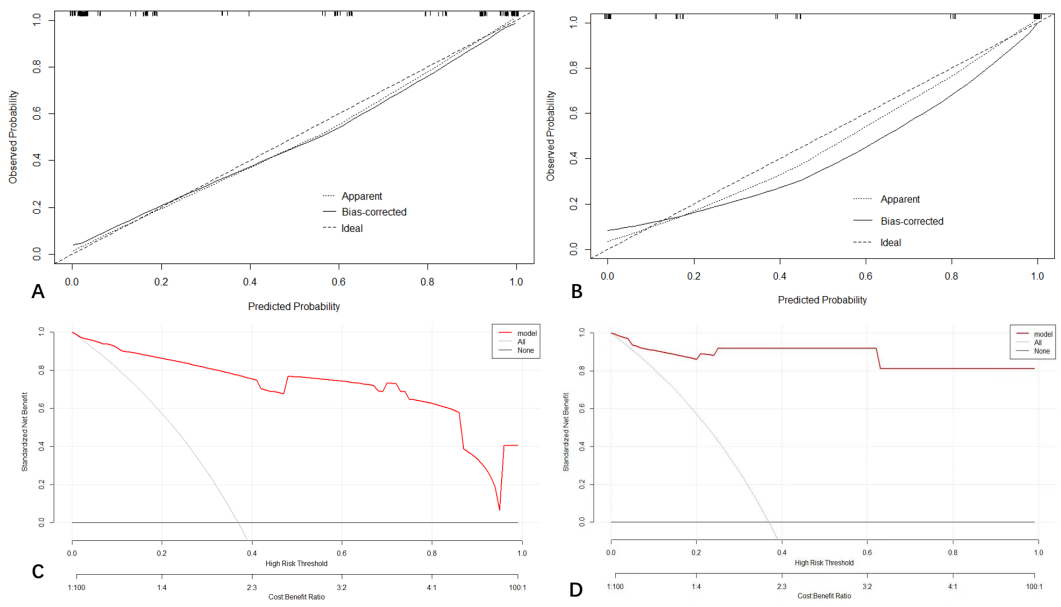


FIGURE 4
Calibration curve demonstrating how predictions from the model to the actual observed probability on training cohort **(A)** and testing cohort **(B)**. Decision curve analysis (DCA) for the nomogram on training **(C)** and validation cohort **(D)**.

the training and validation sets (10). Xu et al. found that the deep learning model based on multiphase CT has the potential in accurately classifying HCC from non-HCC from high-risk liver lesions (LI-4/5/M) with AUC of 0.887 and 0.808 (30). Another previous study developed a deep convolutional neural network-ultrasound (DCNN-US) model to classify HCC from focal hepatic lesions, which exhibited high sensitivity and specificity and outperformed radiologists' visual assessments (31). Undoubtedly, radiomics is important for the diagnosis of HCC with satisfactory model performance. However, it is time-consuming and requires large sample sizes to validate their generalizability. Therefore, to develop a simple and practical model is urgent for radiologists to detect HCC. Our study demonstrated significantly higher performance of the clinical-radiologic nomogram than the clinical or radiological model in detecting HCC without the use of complex software and postprocessing techniques. The AUCs of the training and validation sets were 0.961 and 0.979, respectively, indicating that the nomogram showed good discrimination capability which may aid in the risk stratification and treatment of HCC patients.

There are some limitations to our study. First, it was a retrospective study with a small sample size, especially benign lesions including limited number of regenerative/dysplastic nodules, which may cause potential selection bias. Our work must be validated via prospective studies with larger sample sizes. Second, this was a single-center study, and multi-center validation is required to confirm the nomogram's reproducibility. Third, a selection bias may affect the validity of the study because biopsy may exclude the possibility of combined hepatocellular-cholangiocarcinomas due to sampling error only the cholangiocarcinoma portion was sampled. Fourth, noisy labels exist widely in CT images which may be contributed to statistical noise, structure noise, artifact noise, and various scanned parameters. The suspected HCC was detected and characterized relying on contrast between liver lesion and background seen in different phases of CT (32–34). Some tumors with variable vascular dynamics may be challenging to detect regardless of the phase due to the different noise levels. Thus, it would be of interest to assess the performance of our nomogram using different noisy labels in real practice. Finally, including only patients with a pathologic diagnosis is a design flaw because most HCC patients end up being diagnosed without recourse to a pathological diagnosis. This is very well illustrated by the larger size of the included tumors.

In conclusion, our study presented a nomogram based on gender, increased AFP, positive non-rim APHE, washout, and enhancing capsule to easily and effectively detect HCC at high risk for this disease, allowing clinicians to rapidly evaluate the risk of HCC and reduce unnecessary surgery. Future studies are needed to externally validate the current model.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical Board Approval Number: K-2022-004-01cs committee.

Author contributions

YL: Conceptualization, Data curation, Funding acquisition, Methodology, Software, Writing – original draft. HW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Software, Validation, Writing – review & editing. XW: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by 2021A1515110703, 2022A1515220135, 2021A1515011288, 2021A1515220071 and 2022A1515011470 from Guangdong Basic and Applied Basic Research Foundation (YL, XW, and HW), 202102010031 and 202102010020 from science and Technology Planning Project of Guangzhou (YL, and XW).

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.

References

1. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the L. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol*. (2018) 69:182–236. doi: 10.1016/j.jhep.2018.03.019

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492

3. Yang JD, Roberts LR. Hepatocellular carcinoma: A global view. *Nat Rev Gastroenterol Hepatol*. (2010) 7:448–58. doi: 10.1038/nrgastro.2010.100

4. Kanwal F, Singal AG. Surveillance for hepatocellular carcinoma: current best practice and future direction. *Gastroenterology*. (2019) 157:54–64. doi: 10.1053/j.gastro.2019.02.049

5. Wang Q, Wang A, Wu X, Hu X, Bai G, Fan Y, et al. Radiomics models for preoperative prediction of the histopathological grade of hepatocellular carcinoma: A systematic review and radiomics quality score assessment. *Eur J Radiol*. (2023) 166:111015. doi: 10.1016/j.ejrad.2023.111015

6. Shin J, Lee S, Yoon JK, Roh YH. Diagnostic performance of the 2018 EASL vs. LI-RADS for hepatocellular carcinoma using CT and MRI: A systematic review and meta-analysis of comparative studies. *J Magn Reson Imaging*. (2023) 58:1942–50. doi: 10.1002/jmri.28716

7. Wald C, Russo MW, Heimbach JK, Hussain HK, Pomfret EA, Bruix J. New OPTN/UNOS policy for liver transplant allocation: standardization of liver imaging, diagnosis, classification, and reporting of hepatocellular carcinoma. *Radiology*. (2013) 266:376–82. doi: 10.1148/radiol.12121698

8. Simmons O, Fetzer DT, Yokoo T, Marrero JA, Yopp A, Kono Y, et al. Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis. *Aliment Pharmacol Ther*. (2017) 45:169–77. doi: 10.1111/apt.13841

9. Fraum TJ, Cannella R, Ludwig DR, Tsai R, Naeem M, LeBlanc M, et al. Assessment of primary liver carcinomas other than hepatocellular carcinoma (HCC) with LI-RADS v2018: comparison of the LI-RADS target population to patients without LI-RADS-defined HCC risk factors. *Eur Radiol*. (2019) 30:996–1007. doi: 10.1007/s00330-019-06448-6

10. Zhang H, Guo D, Liu H, He X, Qiao X, Liu X, et al. MRI-based radiomics models to discriminate hepatocellular carcinoma and non-hepatocellular carcinoma in LR-M according to LI-RADS version 2018. *Diagnostics (Basel)*. (2022) 12:1043. doi: 10.3390/diagnostics12051043

11. Hu XX, Bai D, Wang ZL, Zhang Y, Zhao J, Li ML, et al. A model incorporating clinicopathologic and liver imaging reporting and data system-based magnetic resonance imaging features to identify hepatocellular carcinoma in LR-M observations. *Diagn Interv Radiol*. (2023) 29:741–52. doi: 10.4274/dir.2023.232215

12. Kim DW, Lee G, Kim SY, Ahn G, Lee JG, Lee SS, et al. Deep learning-based algorithm to detect primary hepatic Malignancy in multiphase CT of patients at high risk for HCC. *Eur Radiol*. (2021) 31:7047–57. doi: 10.1007/s00330-021-07803-2

13. Hu S, Lyu X, Li W, Cui X, Liu Q, Xu X, et al. Radiomics analysis on noncontrast CT for distinguishing hepatic hemangioma (HH) and hepatocellular carcinoma (HCC). *Contrast Media Mol Imaging*. (2022) 2022:7693631. doi: 10.1155/2022/7693631

14. Wei J, Jiang H, Gu D, Niu M, Fu F, Han Y, et al. Radiomics in liver diseases: Current progress and future opportunities. *Liver Int*. (2020) 40:2050–63. doi: 10.1111/liv.14555

15. Chen C, Chen C, Ma M, Ma X, Lv X, Dong X, et al. Classification of multi-differentiated liver cancer pathological images based on deep learning attention mechanism. *BMC Med Inform Decis Mak*. (2022) 22:176. doi: 10.1186/s12911-022-01919-1

16. Kamath A, Roudenko A, Hecht E, Sirlin C, Chernyak V, Fowler K, et al. CT/MR LI-RADS 2018: clinical implications and management recommendations. *Abdom Radiol (NY)*. (2019) 44:1306–22. doi: 10.1007/s00261-018-1868-6

17. De Gaetano AM, Catalano M, Pompili M, Marini MG, Rodriguez Carnero P, Gulli C, et al. Critical analysis of major and ancillary features of LI-RADS v2018 in the differentiation of small (</= 2 cm) hepatocellular carcinoma from dysplastic nodules with gadobenate dimeglumine-enhanced magnetic resonance imaging. *Eur Rev Med Pharmacol Sci*. (2019) 23:7786–801. doi: 10.26355/eurrev_201909_18988

18. Rimola J, Forner A, Tremosini S, Reig M, Vilana R, Bianchi L, et al. Non-invasive diagnosis of hepatocellular carcinoma </= 2 cm in cirrhosis. Diagnostic accuracy assessing fat, capsule and signal intensity at dynamic MRI. *J Hepatol*. (2012) 56:1317–23. doi: 10.1016/j.jhep.2012.01.004

19. Alhasan A, Cerny M, Olivie D, Billiard JS, Bergeron C, Brown K, et al. LI-RADS for CT diagnosis of hepatocellular carcinoma: performance of major and ancillary features. *Abdom Radiol (NY)*. (2019) 44:517–28. doi: 10.1007/s00261-018-1762-2

20. Bruix J, Sherman M. American Association for the Study of Liver D. Management of hepatocellular carcinoma: an update. *Hepatology*. (2011) 53:1020–2. doi: 10.1002/hep.24199

21. European Association For The Study Of The L and European Organisation For R, Treatment Of C. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. (2012) 56:908–43. doi: 10.1016/j.jhep.2011.12.001

22. Ishigami K, Yoshimitsu K, Nishihara Y, Irie H, Asayama Y, Tajima T, et al. Hepatocellular carcinoma with a pseudocapsule on gadolinium-enhanced MR images: correlation with histopathologic findings. *Radiology*. (2009) 250:435–43. doi: 10.1148/radiol.2501071702

23. Khan AS, Hussain HK, Johnson TD, Weadock WJ, Pelletier SJ, Marrero JA. Value of delayed hypointensity and delayed enhancing rim in magnetic resonance imaging diagnosis of small hepatocellular carcinoma in the cirrhotic liver. *J Magn Reson Imaging*. (2010) 32:360–6. doi: 10.1002/jmri.22271

24. Granata V, Fusco R, Avallone A, Catalano O, Filice F, Leongito M, et al. Major and ancillary magnetic resonance features of LI-RADS to assess HCC: an overview and update. *Infect Agent Cancer*. (2017) 12:23. doi: 10.1186/s13027-017-0132-y

25. Granata V, Fusco R, Avallone A, Filice F, Tatangelo F, Piccirillo M, et al. Critical analysis of the major and ancillary imaging features of LI-RADS on 127 proven HCCs evaluated with functional and morphological MRI: Lights and shadows. *Oncotarget*. (2017) 8:51224–37. doi: 10.18632/oncotarget.17227

26. Sangiovanni A, Manini MA, Iavarone M, Romeo R, Forzenigo LV, Fraquelli M, et al. The diagnostic and economic impact of contrast imaging techniques in the diagnosis of small hepatocellular carcinoma in cirrhosis. *Gut*. (2010) 59:638–44. doi: 10.1136/gut.2009.187286

27. Song P, Tobe RG, Inagaki Y, Kokudo N, Hasegawa K, Sugawara Y, et al. The management of hepatocellular carcinoma around the world: a comparison of guidelines from 2001 to 2011. *Liver Int*. (2012) 32:1053–63. doi: 10.1111/j.1478-3231.2012.02792.x

28. Sherman M, Peltekian KM, Lee C. Screening for hepatocellular carcinoma in chronic carriers of hepatitis B virus: incidence and prevalence of hepatocellular carcinoma in a North American urban population. *Hepatology*. (1995) 22:432–8. doi: 10.1016/0270-9139(95)90562-6

29. Colli A, Fraquelli M, Casazza G, Massironi S, Colucci A, Conte D, et al. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. *Am J Gastroenterol*. (2006) 101:513–23. doi: 10.1111/j.1572-0241.2006.00467.x

30. Xu Y, Zhou C, He X, Song R, Liu Y, Zhang H, et al. Deep learning-assisted LI-RADS grading and distinguishing hepatocellular carcinoma (HCC) from non-HCC based on multiphase CT: a two-center study. *Eur Radiol*. (2023) 33:8879–88. doi: 10.1007/s00330-023-09857-w

31. Yang Q, Wei J, Hao X, Kong D, Yu X, Jiang T, et al. Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: A multicentre study. *EBioMedicine*. (2020) 56:102777. doi: 10.1016/j.ebiom.2020.102777

32. Dakua SP. Use of chaos concept in medical image segmentation. *Comput Methods Biomechanics Biomed Engineering: Imaging Visualization*. (2013) 1:28–36. doi: 10.1080/21681163.2013.765709

33. Wang G, Liu X, Li C, Xu Z, Ruan J, Zhu H, et al. A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Trans Med Imaging*. (2020) 39:2653–63. doi: 10.1109/TMI.2020.3000314

34. Dakua SP, Abi-Nahed J. Patient oriented graph-based image segmentation. *Biomed Signal Process Control*. (2013) 8:325–32. doi: 10.1016/j.bspc.2012.11.009

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 13 August 2024
DOI 10.3389/fonc.2024.1395313

Check for updates

OPEN ACCESS

EDITED BY
Harrison Kim,
University of Alabama at Birmingham,
United States

REVIEWED BY
Radu Hristu,
National University of Science
and Technology POLITEHNICA
Bucharest, Romania
Taihao Jin,
City of Hope National Medical Center,
United States

*CORRESPONDENCE
Youyou Xia
xia.youyou@njmu.edu.cn
Chong Zhou
zhouchongsuda@gmail.com

†These authors have contributed equally to this work

RECEIVED 03 March 2024
ACCEPTED 23 July 2024
PUBLISHED 13 August 2024

CITATION
Sun Y, Liang F, Yang J, Liu Y, Shen Z, Zhou C and Xia Y (2024) Pilot study: radiomic analysis for predicting treatment response to whole-brain radiotherapy combined temozolomide in lung cancer brain metastases.
Front. Oncol. 14:1395313.
doi: 10.3389/fonc.2024.1395313

COPYRIGHT
© 2024 Sun, Liang, Yang, Liu, Shen, Zhou and Xia. 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.

Pilot study: radiomic analysis for predicting treatment response to whole-brain radiotherapy combined temozolomide in lung cancer brain metastases

Yichu Sun[1†], Fei Liang[1†], Jing Yang[2†], Yong Liu[2], Ziqiang Shen[1], Chong Zhou[3*] and Youyou Xia[1,2*]

[1]Department of Radiation Oncology, The First People's Hospital of Lianyungang/Lianyungang Clinical College of Nanjing Medical University, Lianyungang, Jiangsu, China, [2]Department of Radiation Oncology, The Affiliated Lianyungang Hospital of Xuzhou Medical University/The First People's Hospital of Lianyungang, Lianyungang, Jiangsu, China, [3]Department of Radiation Oncology, Xuzhou Central Hospital, Xuzhou, Jiangsu, China

Objective: The objective of this study is to assess the viability of utilizing radiomics for predicting the treatment response of lung cancer brain metastases (LCBM) to whole-brain radiotherapy (WBRT) combined with temozolomide (TMZ).

Methods: Fifty-three patients diagnosed with LCBM and undergoing WBRT combined with TMZ were enrolled. Patients were divided into responsive and non-responsive groups based on the RANO-BM criteria. Radiomic features were extracted from contrast-enhanced the whole brain tissue CT images. Feature selection was performed using t-tests, Pearson correlation coefficients, and Least Absolute Shrinkage And Selection (LASSO) regression. Logistic regression was employed to construct the radiomics model, which was then integrated with clinical data to develop the nomogram model. Model performance was evaluated using receiver operating characteristic (ROC) curves, and clinical utility was assessed using decision curve analysis (DCA).

Results: A total of 1834 radiomic features were extracted from each patient's images, and 3 features with predictive value were selected. Both the radiomics and nomogram models exhibited satisfactory predictive performance and clinical utility, with the nomogram model demonstrating superior predictive value. The ROC analysis revealed that the AUC of the radiomics model in the training and testing sets were 0.776 and 0.767, respectively, while the AUC of the nomogram model were 0.799 and 0.833, respectively. DCA curves demonstrated that both models provided benefits to patients across various thresholds.

Conclusion: Radiomic-defined image biomarkers can effectively predict the treatment response of WBRT combined with TMZ in patients with LCBM, offering potential to optimize treatment decisions for this condition.

KEYWORDS

brain metastases, whole-brain radiation therapy, temozolomide, radiomics, nomogram

Introduction

Lung cancer (LC) ranks among the most prevalent malignant tumors globally, accounting for 11.4% of all cancer incidences and contributing to an 18% mortality rate (1). Brain metastasis (BM), a common form of metastasis in LC, poses a formidable challenge in the treatment and disease management due to its significant threat to patients' functional independence and survival, with a median survival period ranging from 4 to 10 months post-treatment (2). The unique anatomical and physiological characteristics of the brain present challenges in finding effective treatment strategies for BM. Locally, whole-brain radiotherapy (WBRT) remains a pivotal local therapeutic modality for BM, primarily targeting patients with multiple brain metastases who are unsuitable for stereotactic radiosurgery (SRS) or surgical intervention. However, the central nervous system's limited tolerance to radiation restricts the dose of WBRT, making tumor reduction challenging. Systemically, targeted tyrosine kinase inhibitors are applicable only to a minority of patients with specific driver gene mutations. Platinum-based chemotherapy regimens and immune checkpoint inhibitors constitute the mainstay of systemic therapy for advanced LC; however, their efficacy against BM is limited due to the blood-brain barrier (3).

Temozolomide (TMZ), a novel imidazotetrazine alkylating agent, readily crosses the blood-brain barrier to exert intracranial anti-tumor effects, thereby enhancing the local efficacy or survival benefits of radiotherapy for brain metastases (4, 5). A Chinese treatment guideline for lung cancer brain metastases (LCBM) recommends TMZ as a radiosensitizer in combination with WBRT, suggesting it as a promising alternative (6). In a moderately sized real-world study, the addition of TMZ to WBRT increased intracranial objective response rates from 20.2% to 34.9%, disease control rates from 92.7% to 98.4%, progression-free survival from 4.9 to 5.9 months, and overall survival from 5.9 to 8.5 months (7). Nevertheless, a recent meta-analysis incorporating 25 prospective studies indicated that the addition of TMZ to WBRT significantly increased objective response rates while also elevating the risk of hematological and gastrointestinal toxicities (8). This combination therapy is not suitable for every individual. Therefore, developing tools to predict treatment response can facilitate individualized medical decisions, avoid ineffective treatments, and prevent treatment toxicity. However, there is currently a lack of reported predictive model studies for this patient population.

Radiomics, an emerging field that integrates artificial intelligence image recognition with medical imaging technology, extracts high-throughput omics data with biological significance from medical images through a series of standardized quantitative calculation methods. After appropriate feature selection and machine learning methods, these imaging biomarkers can reflect the potential differences and complexities of diseases (9, 10). In recent years, some studies have successfully applied radiomics to predict the microscopic heterogeneity of lung cancer brain metastases, such as pathological subtypes (11), PD-L1 expression (12), and EGFR mutation status (13). Thus, radiomics holds significant potential in distinguishing individual differences in LCBM, perhaps providing a potential pathway for predicting the efficacy of WBRT combined with TMZ.

Therefore, this study aimed to explore the feasibility and application value of radiomics technology in predicting the efficacy of WBRT combined with TMZ for LCBM, expected to optimize personalized treatment decisions.

Materials and methods

Patients recruiting

The medical center's ethics committee approved this retrospective study (Permit No. LW-20240119001-01) with waived written informed consent from patients. Clinicopathological data were retrieved by two radiation oncologists through review of medical records and telephone follow-ups, and then evaluated by a senior radiation oncology expert. Between January 2018 and June 2023, a total of 78 patients with LCBM who underwent WBRT combined with TMZ at our radiotherapy center were identified. After screening based on inclusion and exclusion criteria, 53 patients were included in this study and randomly divided into training and testing sets at an 8:2 ratio. Inclusion criteria were as follows: (1) Definite diagnosis of LCBM through histopathology or typical CT or MRI findings; (2) Available contrast-enhanced whole-brain CT and Eclipse treatment planning system images; (3) Completed WBRT synchronized with temozolomide treatment as planned and evaluated for efficacy. Exclusion criteria were: (1) Presence of intracranial organic diseases or history of intracranial radiotherapy; (2) Occurrence of acute cerebrovascular events during treatment; (3) Incomplete data retrieval despite review of medical records and telephone follow-ups. WBRT was delivered using one of the two American Varian linear accelerators at our radiotherapy center, with a planned target volume (PTV) expansion of 0.3cm beyond the whole brain. The radiotherapy doses were delivered at 30 Gy in 10 fractions, 37.5 Gy in 15 fractions, or 40 Gy in 20 fractions, in accordance with the NCCN guidelines (14). Radiation therapy was administered five times per week, along with concurrent oral or intravenous administration of temozolomide at a dosage of 75-100 mg/m^2/day. Treatment response assessment was conducted at 1.5 to 2 months post-treatment, following the Response Assessment in Neuro-Oncology Brain Metastases (RANO-BM) criteria (15). Patients with complete response (CR) or partial response (PR) were divided into responsive group, while those with stable disease (SD) or disease progression (PD) were divided into non-responsive group.

Image acquisition

All contrast-enhanced CT scans were performed using a Siemens SOMATOM Definition AS+ 64-slice CT simulator at our radiotherapy center, covering from the cranial vertex to the lower boundary of the neck, with a slice thickness of 3 mm. The final WBRT treatment plan was devised and confirmed by the radiation oncology expert team at our institution. DICOM-format whole-brain CT image files and RT-structure files were exported via the Eclipse system and ARIA network workstation. Subsequently, we utilized the OnekeyAI platform (www.medai.icu/) to parse and

convert the DICOM files and RT-structure files into NIFTI-format localization CT images and corresponding radiotherapy target volume. The definition of regions of interest (ROIs) in radiomics primarily relied on lesion-based or anatomy-based criteria (16, 17). Given the scope of local radiotherapy across the entire brain in this study, ROIs were defined in 3D based on the anatomical structures of the whole brain tissue. The final ROIs were manually adjusted by one radiation oncologist using 3dslicer [5.1.0] software (18) and verified by a senior radiation oncology expert.

Radiomics features extraction

Image preprocessing and feature extraction were conducted using the PyRadiomics [3.0.1] Python package. Comprehensive image preprocessing included standardizing the image to a range of 0-1000 Hounsfield Units (HU) and spatially resampling the image to a consistent voxel size of 3mm x 3mm x 3mm. Various filters were applied to enhance the images, including Gaussian Laplacian (LoG) filter, wavelet transformation, Local Binary Patterns in 3D (LBP3D), as well as mathematical transformations such as exponential, square, square root, and logarithmic transformations to highlight different features and textures in the images. Subsequently, shape features, first-order statistical features, and texture features were extracted from the preprocessed images and the transformed images obtained from the aforementioned filters. The texture features included Gray Level Co-occurrence Matrix (GLCM), Gray Level Dependence Matrix (GLDM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM). Further descriptions of the feature extraction process and the YAML configuration file used with pyradiomics were provided in Supplementary File 1. All feature definitions and calculation methods adhered to the standards of the Imaging Biomarker Standardization Initiative (IBSI) (10).

Feature selection

All radiomic features were transformed using the formula $Z = (X - \mu)/\sigma$, and then the feature selection process commenced in the training dataset. Initially, the most significantly distinguishing features between the responsive and non-responsive groups were identified through statistical analysis using the t-test, with only features with $p < 0.01$ retained. Subsequently, redundant features were removed using Pearson correlation coefficients and greedy recursive feature elimination. When any two features had a correlation coefficient exceeding 0.9, one of them was systematically eliminated. In each iteration, the 10 most redundant features with the highest correlation coefficients were discarded. Finally, the Absolute Shrinkage and Selection (LASSO) model was developed, and the lambda parameter was determined through 10-fold cross-validation, aiming to minimize the mean squared error (MSE). The non-zero coefficient features were identified at that lambda.

Radiomics signature

The final features selected by the Lasso model were used to construct the radiomics prediction model. Logistic regression (LR), one of the most commonly used machine learning algorithms in medical data, was chosen for this purpose. The output results of the prediction model were used to construct the radiomic signature. Subsequently, confusion matrices and receiver operating characteristic (ROC) curves were employed for both the training and testing sets to demonstrate predictive performance. Additionally, decision curve analysis (DCA) was conducted to assess clinical utility across different thresholds. Furthermore, the model output results were utilized to generate radiomics signature, serving as biomarkers for subsequent analysis.

Nomogram model

We also explored the potential enhancement of the radiomics model's predictive performance by incorporating clinical variables. Baseline features including age, gender, Karnofsky Performance Status (KPS), serum carcinoembryonic antigen (CEA) levels, and serum neuron-specific enolase (NSE) were used to generate clinical signature through the same LR results, ensuring its comparability to radiomics signature. Subsequently, the nomogram model was developed by integrating clinical signature and radiomics signature. ROC curves were generated for both the training and testing sets to evaluate and compare their performance. DCA curves were employed to evaluate and compare their clinical utility across different thresholds. The comprehensive workflow of the radiomics analysis is depicted in Figure 1.

Statistics and analysis

Normality was assessed using the Shapiro-Wilk test. Continuous variables with normal distribution were presented as mean and standard deviation and analyzed using the t-test. Continuous variables with severe skewed distribution were presented as median and interquartile range and analyzed using the Wilcoxon test. Categorical variables were presented as frequency and percentage and analyzed using the chi-square test or chi-square test with Yates' correction. The significance level for hypothesis testing was set at 0.05. Statistical analysis of baseline data was performed using the stats [4.2.11] R package, nomogram analysis using the rms [6.4.0] R package, ROC analysis using the pROC [1.18.0] R package, DCA analysis using the rmda [1.6] R package, and results were visualized using the ggplot2 [3.3.6] R package. Model building was conducted using the scikit-learn [1.0.2] Python package.

Results

Patient's characteristics

The clinical baseline characteristics presented in Table 1 encompassed a total of 53 patients with LCBM, with 33 cases (62.3%) of non-small cell lung cancer and 20 cases (37.7%) of small

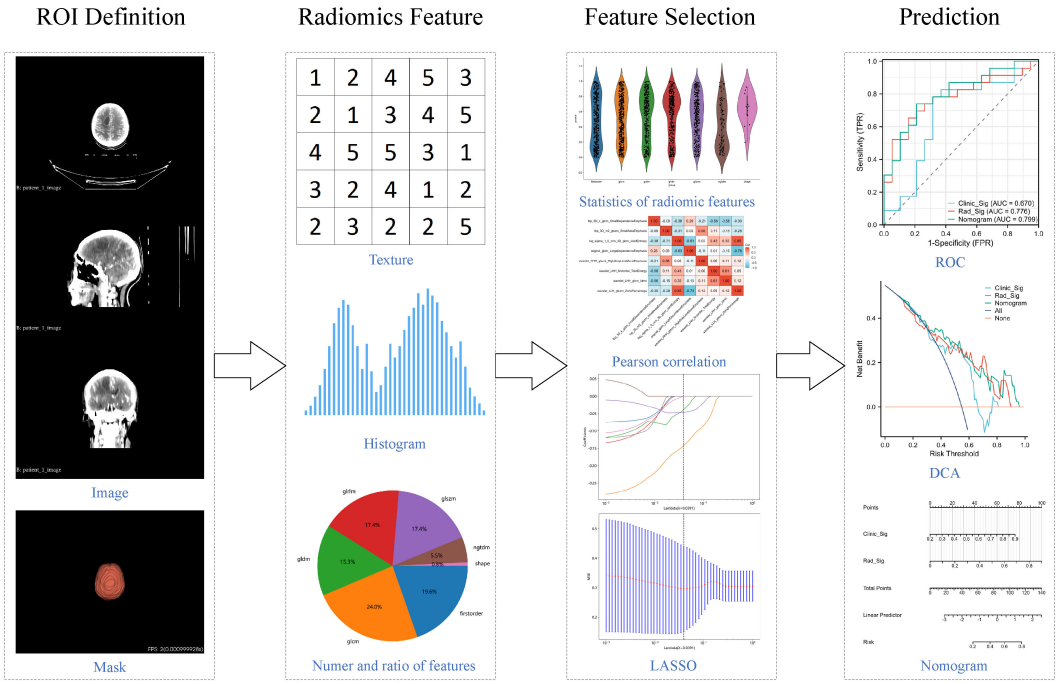


FIGURE 1
Workflow of radiomics analysis in this study.

cell lung cancer. Among these individuals, there were 17 females (32.1%) and 36 males (67.9%), with ages ranging from 36 to 83 years and a mean age of 60.4 years. According to the evaluation results, there were 24 cases in the responsive group and 29 cases in the non-responsive group. Statistical analysis revealed no significant differences in all clinical variables between the training and testing sets, indicating the effectiveness of the random dataset partition.

Radiomics features

Each patient's CT images produced 1834 radiomic features. Detailed results can be found in Supplementary File 1. Initially, 53 features with $p < 0.01$ were identified through t-test analysis, as depicted in Figure 2A. Subsequently, 8 features were selected by the last iteration of greedy recursive elimination. Their correlation coefficients were presented in Figure 2B. Finally, during the 10-fold cross-validation process of the LASSO model, the weight coefficients of the 8 radiomic features varied with lambda, as shown in Figure 2C, while the change in MSE with lambda was illustrated in Figure 2D. The lambda parameter associated with the minimum MSE was found to be 0.0391, corresponding to the retention of 3 non-zero coefficient features in the model: "lbp_3D_m2_glszm_SmallAreaEmphasis", "log_sigma_1_0_mm_3D_glcm_JointEntropy", and "wavelet_HHH_glszm_HighGrayLevelZoneEmphasis".

Predictive performance

The confusion matrices for the radiomics model in the training and testing sets were displayed in Figures 3A, B, respectively, with prediction accuracies of 71.4% and 72.7%. The ROC curves of the radiomics signature, clinical signature, and nomogram models on the training set were illustrated in Figure 4A, while those on the testing set were presented in Figure 4B. The AUC values on the training and testing sets were as follows: clinical signature: 0.670, 0.667; radiomics signature: 0.776, 0.767; nomogram: 0.799, 0.833. There were no indications of overfitting detected in any of the models. Both radiomics signature and nomogram showed AUC values surpassing 0.75, suggesting robust predictive performance. In contrast, clinical signature displayed AUC values below 0.7. The DCA curves on the training set were depicted in Figure 5A, while those on the testing set were shown in Figure 5B. Across various risk thresholds, radiomics signature and nomogram yielded net benefits to patients in most scenarios, outperforming clinical signature. Combining the results of ROC curves and DCA analysis, the nomogram surpassed radiomics signature in performance on both the training and testing sets. The visualization of the nomogram model was portrayed in Figure 6, illustrating the relationships between the nomogram and radiomics signature, as well as clinical signature.

Discussion

This study attempted to utilize radiomic analysis to predict the treatment response of WBRT combined with TMZ for LCBM, demonstrating its feasibility and application value. Furthermore, integrating clinical information into the radiomic model through a nomogram can further improved predictive performance.

TABLE 1 Baseline characteristics of patients in the training and validation sets.

Characteristics	Overall	Training	Test	pvalue
n	53	42	11	
Age, mean ± sd	60.4 ± 9.3	60.8 ± 9.9	58.9 ± 6.3	0.552
CEA, median (IQR)	6.8 (3.0, 64.5)	7.8 (3.4, 65.1)	2.9 (2.2, 34.6)	0.139
NSE, median (IQR)	19.9(14.5, 29.6)	18.9 (14.6, 26.8)	27.0 (18.8, 31.4)	0.303
Gender, n (%)				0.999
Female	17 (32.1%)	13 (24.5%)	4 (7.5%)	
Male	36 (67.9%)	29 (54.7%)	7 (13.2%)	
KPS, n (%)				0.922
≥70	32 (60.4%)	26 (49.1%)	6 (11.3%)	
<70	21 (39.6%)	16 (30.2%)	5 (9.4%)	
Pathology, n (%)				0.649
NSCLC	33 (62.3%)	25 (47.2%)	8 (15.1%)	
SCLC	20 (37.7%)	17 (32.1%)	3 (5.7%)	
BM_Number, n (%)				0.999
<5	24 (45.3%)	19 (35.8%)	5 (9.4%)	
≥5	29 (54.7%)	23 (43.4%)	6 (11.3%)	
CNS_symptom, n (%)				0.999
No	23 (43.4%)	18 (34.0%)	5 (9.4%)	
Yes	30 (56.6%)	24 (45.3%)	6 (11.3%)	
PTV_dose_cGy, n (%)				0.807
3000	35 (66.0%)	27 (50.9%)	8 (15.1%)	
3750	8 (15.1%)	8 (15.1%)	2 (3.8%)	
4000	10 (18.9%)	7 (13.2%)	1 (1.9%)	
TMZ_route, n (%)				0.552
Intravenous	32 (60.4%)	24 (45.3%)	8 (15.1%)	
Oral	21 (39.6%)	18 (34%)	3 (5.7%)	

IQR, interquartile range; sd, standard deviation; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; CEA, carcinoembryonic antigen; NSE, neuron-specific enolase; CNS, central nervous system; TMZ, temozolomide.

Radiomic technology aims to identify valuable biomarkers by extracting quantitative features from images, holding great potential for predicting outcomes in LC treatment (19–21). However, in studies on BM, researchers often focus on using radiomics to predict the efficacy of SRS (22–24), while research involving the prediction of WBRT treatment response is scarce. WBRT, as a crucial local treatment modality for BM, often faces challenges in reducing tumor size due to dose limitations on brain tissue (25). Data from a phase III clinical trial (PCI-P120-9801) suggest a correlation between tumor shrinkage after WBRT and improved survival and neurological function (26). Therefore, the development of tools to predict WBRT treatment response in patients with BM is crucial for making informed clinical decisions. Wang et al.'s study attempted to use radiomic technology to predict the treatment response of BM patients to WBRT, demonstrating excellent predictive performance. The AUC of the radiomic model was 0.928 and 0.837 in the training and testing sets, respectively, while the AUC of the clinical model was 0.650 and 0.598. The nomogram model combining clinical factors and radiomics achieved AUCs of 0.928 and 0.851 (27). Previous studies have also used radiomics to predict the efficacy of TMZ, but in the context of glioblastoma. For example, Li et al.'s study analyzed radiomic features and predicted efficacy in 53 patients with malignant glioblastoma receiving anlotinb and TMZ combination therapy, achieving AUCs above 0.9 in both the training and testing sets. However, this study did not incorporate clinical information into the radiomic model (28). Therefore, it is necessary to investigate the feasibility of using radiomics to predict the response of BM to WBRT combined with TMZ.

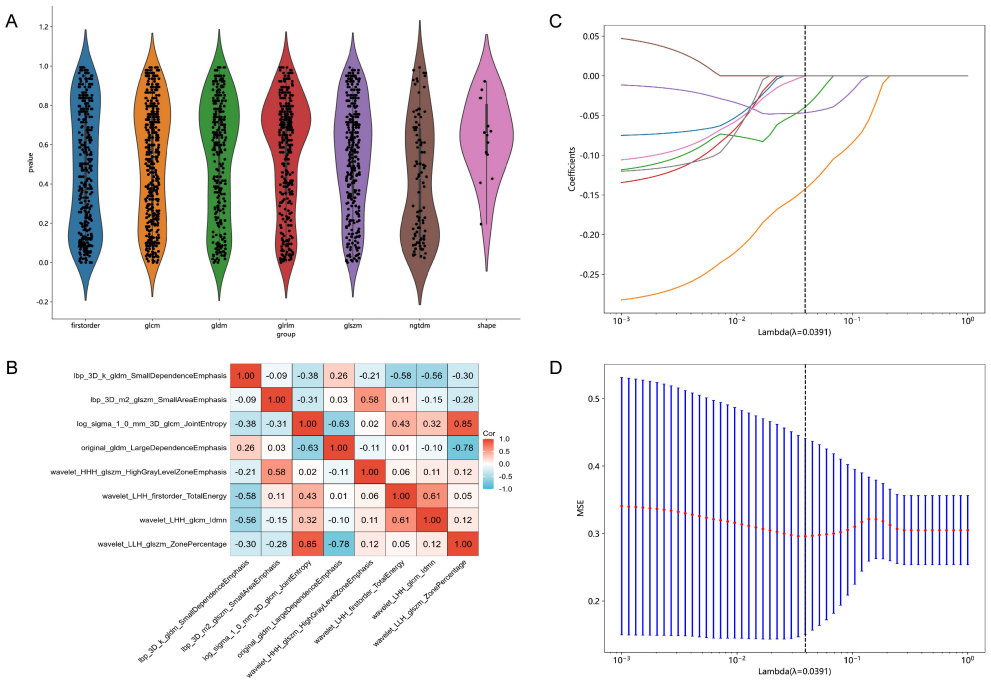


FIGURE 2
Results of feature selection. **(A)** Violin plot illustrating the distribution of p-values from the statistical tests of all radiomic features. **(B)** Heatmap showing the correlation coefficients of the 8 selected features obtained through Pearson correlation coefficients combined with a greedy strategy. **(C)** Variations in the weight coefficients of the 8 radiomic features with lambda during the 10-fold cross-validation process of the least absolute shrinkage and selection operator (LASSO) model. **(D)** The change in mean squared error (MSE) with lambda during the 10-fold cross-validation process of the LASSO model.

In this study, we aimed to reduce the heterogeneity among BM patients by focusing on those with primary LC. We gathered contrast-enhanced CT scans from BM patients who underwent WBRT in combination with TMZ at our academic medical center over the last five years for radiomic analysis. During feature selection, we pinpointed three predictive features using the LASSO model with the minimum mean squared error (MSE), all of which were related to filtered enhancement and transformed grayscale texture features. Notably, these features did not encompass shape or first-order statistical features, likely due to our ROIs definition based on anatomical structures, given the whole-brain scope therapy of our study. This ROIs strategy is advantageous as it incorporates not only intra-tumoral information but also data from the tumor periphery and surrounding normal tissue, all of which hold predictive potential in radiomics (16, 17, 29). Subsequently, we constructed predictive

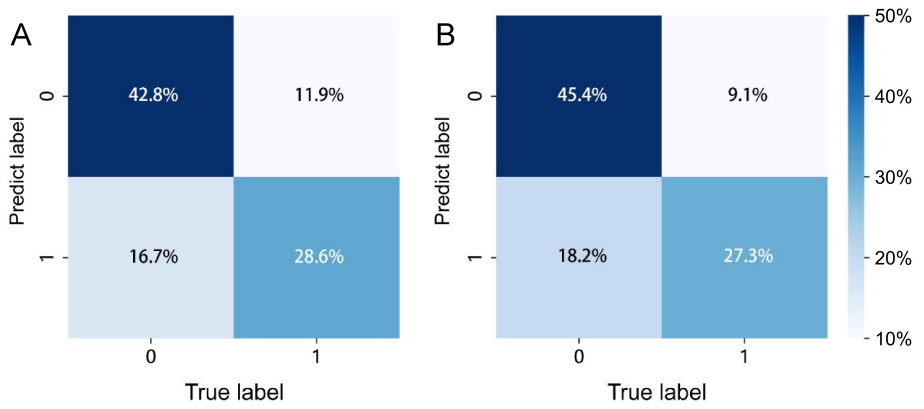


FIGURE 3
The prediction confusion matrices of the radiomics model in the training set **(A)** and testing set **(B)**.

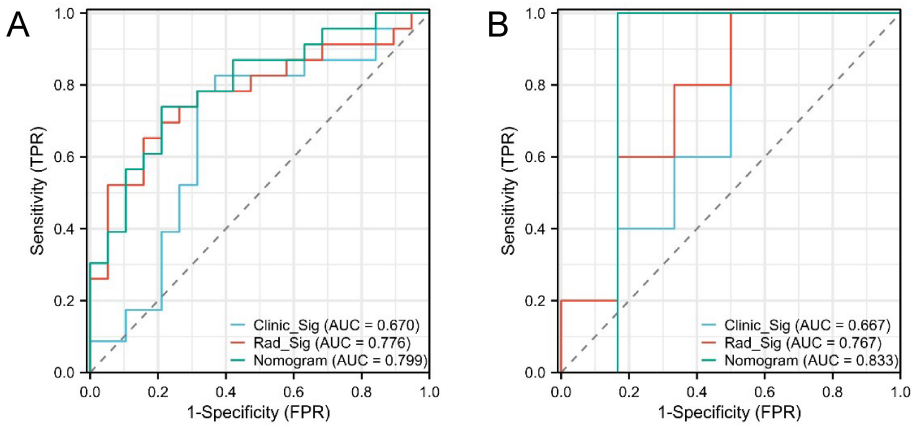


FIGURE 4
ROC curves of the three models in the training set **(A)** and testing set **(B)**. Both radiomics signature (Rad_Sig) and nomogram exhibited AUC values exceeding 0.75, indicating good predictive performance.

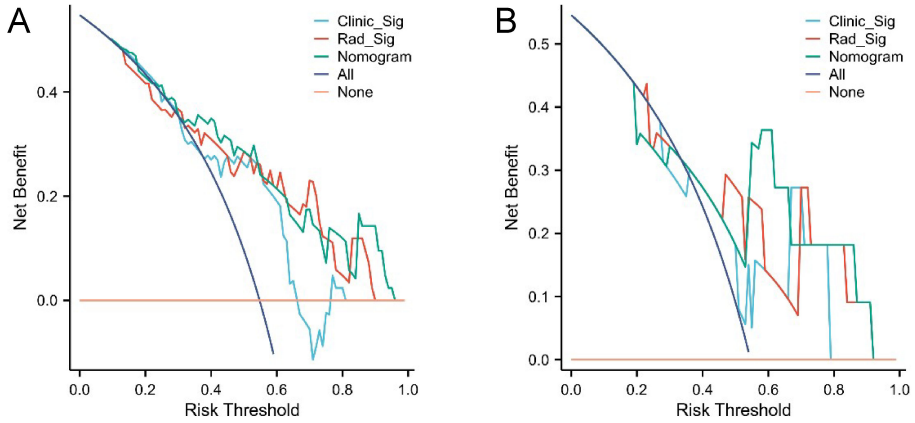


FIGURE 5
DCA curves of the three models in the training set **(A)** and testing set **(B)**. Radiomics signature (Rad_Sig) and nomogram demonstrated net benefits to patients across the majority of risk thresholds.

models and assessed their predictive value using ROC and DCA analysis. In ROC analysis, the AUCs of radiomic model were 0.776 and 0.767 in the training and testing sets, respectively, while those of clinical model were 0.670 and 0.667. The nomogram, combining clinical signature and radiomic signature, achieved AUCs of 0.799 and 0.833. These results suggest the feasibility of utilizing radiomics to predict treatment response of WBRT combined with TMZ in BM patients. Notably, prior research focusing on LC biomarkers has underscored that clinical signature complement radiomic signature by providing valuable insights (30). In our study, while clinical signature alone failed to predict treatment response, it could offer potential information to radiomic signature, thereby enhancing predictive performance, which was consistent with the findings of Wang et al.'s study (27). Moreover, our DCA analysis further affirmed the utility of both the radiomic model and the nomogram for clinical decision-making. In summary, these results underscore the potential of radiomics as a supplementary clinical tool for predicting treatment response of WBRT combined with TMZ in LCBM, thus facilitating personalized treatment strategies.

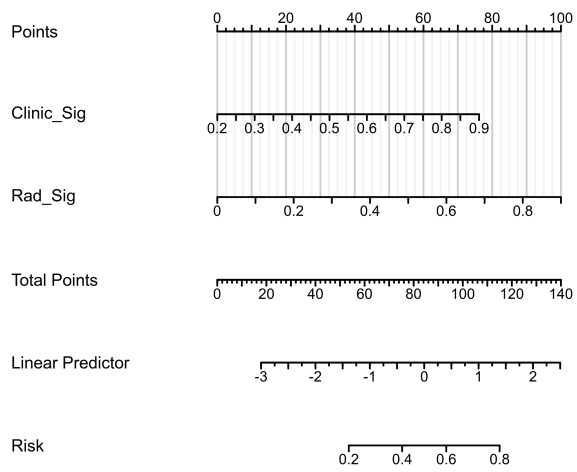


FIGURE 6
The visualization of the nomogram model.

There are several limitations in our study. Firstly, the small sample size and retrospective data collection might introduce bias. Secondly, the clarity of enhanced CT images is not as high as MRI, potentially impacting the accuracy of the radiomic model (31). Thirdly, the single-center division of the testing set may not adequately validate the model's stability. Future research will utilize multi-modal imaging data and acquire multi-center validation datasets to enhance the accuracy and generalizability of predictive models.

Conclusion

Overall, this study successfully predicted the treatment response of LCBM undergoing WBRT combined with TMZ using radiomic-defined image biomarkers. These findings offer potential for optimizing treatment decisions for LCBM in clinical practice and laying the groundwork for further clinical application research.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The original data presented in this article are partially uploaded in Supplementary File 1. Inquiries can be directed to the corresponding author with a reasonable request. Requests to access these datasets should be directed to Xia Youyou, xia.youyou@njmu.edu.cn.

Ethics statement

The studies involving humans were approved by Medical Ethics Committee of the First People's Hospital of Lianyungang. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

YS: Conceptualization, Formal analysis, Software, Visualization, Writing – original draft. FL: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. JY: Data curation, Investigation, Project administration, Supervision, Writing – original draft. YL: Investigation, Writing – original draft. ZS: Resources, Writing – original draft. CZ: Supervision, Visualization, Writing – review & editing. YX: Funding acquisition, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study received funding from the Key Laboratory Open Research Project of Jiangsu Province Higher Education Institutions (Grant No. XZSYSKF2022042), the Clinical Research Special Fund of the First People's Hospital of Lianyungang City (Grant No. LC27) and the Oncology Prevention and Treatment Science and Technology Development Program Project of Lianyungang City (Grant No. MS202305).

Acknowledgments

We sincerely thank the Platform Onekey AI (www.medai.icu/) and the Platform Xiantao (www.xiantaozi.com/) for Code consultation. We also thank Home for Researchers editorial team (www.home-forresearchers.com) for language editing service.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1395313/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

2. Cagney DN, Martin AM, Catalano PJ, Redig AJ, Lin NU, Lee EQ, et al. Incidence and prognosis of patients with brain metastases at diagnosis of systemic Malignancy: a population-based study. *Neuro Oncol*. (2017) 19:1511–21. doi: 10.1093/neuonc/nox077

3. Suh JH, Kotecha R, Chao ST, Ahluwalia MS, Sahgal A, Chang EL. Current approaches to the management of brain metastases. *Nat Rev Clin Oncol*. (2020) 17:279–99. doi: 10.1038/s41571-019-0320-3

4. Zhu Y, Fu L, Jing W, Guo D, Kong L, Yu J. Effectiveness of temozolomide combined with whole brain radiotherapy for non-small cell lung cancer brain metastases. *Thorac Cancer*. (2018) 9:1121–28. doi: 10.1111/1759-7714.12795

5. Bi N, Ma Y, Xiao J, Zhang H, Xu Y, Tian Y, et al. A phase ii trial of concurrent temozolomide and hypofractionated stereotactic radiotherapy for complex brain metastases. *Oncologist*. (2019) 24:e914–20. doi: 10.1634/theoncologist.2018-0702

6. Chinese medical doctor association oncologist branch and China international exchange and promotive association for medical and health care medical oncology branch. Clinical practice guideline for brain metastases of lung cancer in China (2021 version). *Chin J Oncol*. (2021) 43:269–81. doi: 10.3760/cma.j.cn112152-20210104-00009

7. Deng X, Zheng Z, Lin B, Su H, Chen H, Fei S, et al. The efficacy and roles of combining temozolomide with whole brain radiotherapy in protection neurocognitive function and improvement quality of life of non-small-cell lung cancer patients with brain metastases. *BMC Cancer*. (2017) 17:42. doi: 10.1186/s12885-016-3017-3

8. Han J, Qiu M, Su L, Wu C, Cheng S, Zhao Z, et al. Response and safety of whole-brain radiotherapy plus temozolomide for patients with brain metastases of non-small-cell lung cancer: a meta-analysis. *Thorac Cancer*. (2021) 12:3177–83. doi: 10.1111/1759-7714.14183

9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

10. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295:328–38. doi: 10.1148/radiol.2020191145

11. Deng F, Liu Z, Fang W, Niu L, Chu X, Cheng Q, et al. Mri radiomics for brain metastasis sub-pathology classification from non-small cell lung cancer: a machine learning, multicenter study. *Phys Eng Sci Med*. (2023) 46:1309–20. doi: 10.1007/s13246-023-01300-0

12. Meissner AK, Gutsche R, Galldiks N, Kocher M, Junger ST, Eich ML, et al. Radiomics for the non-invasive prediction of pd-l1 expression in patients with brain metastases secondary to non-small cell lung cancer. *J Neurooncol*. (2023) 163:597–605. doi: 10.1007/s11060-023-04367-7

13. Liao CY, Lee CC, Yang HC, Chen CJ, Chung WY, Wu HM, et al. Predicting survival after radiosurgery in patients with lung cancer brain metastases using deep learning of radiomics and egfr status. *Phys Eng Sci Med*. (2023) 46:585–96. doi: 10.1007/s13246-023-01234-7

14. National comprehensive cancer network. *Clinical practice guidelines in oncology for central nervous system cancers*. Available online at: http://www.nccn.org.

15. Lin NU, Lee EQ, Aoyama H, Barani IJ, Barboriak DP, Baumert BG, et al. Response assessment criteria for brain metastases: proposal from the rano group. *Lancet Oncol*. (2015) 16:e270–78. doi: 10.1016/S1470-2045(15)70057-4

16. Kapetanou E, Malamas S, Leventis D, Karantanas AH, Klontzas ME. Developing a radiomics atlas dataset of normal abdominal and pelvic computed tomography (radapt). *J Imaging Inform Med*. (2024). doi: 10.1007/s10278-024-01028-7

17. Chu X, Gong J, Yang X, Ni J, Gu Y, Zhu Z. A "seed-and-soil" radiomics model predicts brain metastasis development in lung cancer: implications for risk-stratified prophylactic cranial irradiation. *Cancers (Basel)*. (2023) 15:307. doi: 10.3390/cancers15010307

18. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001

19. Wei H, Yang F, Liu Z, Sun S, Xu F, Liu P, et al. Application of computed tomography-based radiomics signature analysis in the prediction of the response of small cell lung cancer patients to first-line chemotherapy. *Exp Ther Med*. (2019) 17:3621–29. doi: 10.3892/etm.2019.7357

20. He L, Li Z, Chen X, Huang Y, Yan L, Liang C, et al. A radiomics prognostic scoring system for predicting progression-free survival in patients with stage iv non-small cell lung cancer treated with platinum-based chemotherapy. *Chin J Cancer Res*. (2021) 33:592–605. doi: 10.21147/j.issn.1000-9604.2021.05.06

21. Chen W, Wang L, Hou Y, Li L, Chang L, Li Y, et al. Combined radiomics-clinical model to predict radiotherapy response in inoperable stage iii and iv non-small-cell lung cancer. *Technol Cancer Res Treat*. (2022) 21:2081083008. doi: 10.1177/15330338221142400

22. Mouraviev A, Detsky J, Sahgal A, Ruschin M, Lee YK, Karam I, et al. Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery. *Neuro Oncol*. (2020) 22:797–805. doi: 10.1093/neuonc/noaa007

23. DeVries DA, Lagerwaard F, Zindler J, Yeung T, Rodrigues G, Hajdok G, et al. Performance sensitivity analysis of brain metastasis stereotactic radiosurgery outcome prediction using mri radiomics. *Sci Rep*. (2022) 12:20975. doi: 10.1038/s41598-022-25389-7

24. Jiang Z, Wang B, Han X, Zhao P, Gao M, Zhang Y, et al. Multimodality mri-based radiomics approach to predict the posttreatment response of lung cancer brain metastases to gamma knife radiosurgery. *Eur Radiol*. (2022) 32:2266–76. doi: 10.1007/s00330-021-08368-w

25. Jablonska PA, Bosch-Barrera J, Serrano D, Valiente M, Calvo A, Aristu J. Challenges and novel opportunities of radiation therapy for brain metastases in non-small cell lung cancer. *Cancers (Basel)*. (2021) 13:2141. doi: 10.3390/cancers13092141

26. Li J, Bentzen SM, Renschler M, Mehta MP. Regression after whole-brain radiation therapy for brain metastases correlates with survival and improved neurocognitive function. *J Clin Oncol*. (2007) 25:1260–66. doi: 10.1200/JCO.2006.09.2536

27. Wang Y, Lang J, Zuo JZ, Dong Y, Hu Z, Xu X, et al. The radiomic-clinical model using the shap method for assessing the treatment response of whole-brain radiotherapy: a multicentric study. *Eur Radiol*. (2022) 32:8737–47. doi: 10.1007/s00330-022-08887-0

28. Li Y, Xu W, Fei Y, Wu M, Yuan J, Qiu L, et al. A mri-based radiomics model for predicting the response to anlotinb combined with temozolomide in recurrent Malignant glioma patients. *Discov Oncol*. (2023) 14:154. doi: 10.1007/s12672-023-00751-x

29. Shi J, Dong Y, Jiang W, Qin F, Wang X, Cui L, et al. Mri-based peritumoral radiomics analysis for preoperative prediction of lymph node metastasis in early-stage cervical cancer: a multi-center study. *Magn Reson Imaging*. (2022) 88:1–08. doi: 10.1016/j.mri.2021.12.008

30. Shen J, Du H, Wang Y, Du L, Yang D, Wang L, et al. A novel nomogram model combining ct texture features and urine energy metabolism to differentiate single benign from Malignant pulmonary nodule. *Front Oncol*. (2022) 12:1035307. doi: 10.3389/fonc.2022.1035307

31. Cao G, Zhang J, Lei X, Yu B, Ai Y, Zhang Z, et al. Differentiating primary tumors for brain metastasis with integrated radiomics from multiple imaging modalities. *Dis Markers*. (2022) 2022:5147085. doi: 10.1155/2022/5147085

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 15 August 2024
DOI 10.3389/fonc.2024.1431536

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Silvia Capuani,
National Research Council (CNR), Italy
Muge Karaman,
University of Illinois Chicago, United States

*CORRESPONDENCE
Jianxing Qiu
qjx761225@126.com

RECEIVED 12 May 2024
ACCEPTED 30 July 2024
PUBLISHED 15 August 2024

CITATION
Wang W, Wu J, Shen Q, Li W, Xue K, Yang Y and Qiu J (2024) Assessment of pathological grade and variants of bladder cancer with a continuous-time random-walk diffusion model.
Front. Oncol. 14:1431536.
doi: 10.3389/fonc.2024.1431536

COPYRIGHT
© 2024 Wang, Wu, Shen, Li, Xue, Yang and Qiu. 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.

Assessment of pathological grade and variants of bladder cancer with a continuous-time random-walk diffusion model

Wei Wang[1], Jingyun Wu[1], Qi Shen[2], Wei Li[1], Ke Xue[3], Yuxin Yang[3] and Jianxing Qiu[1]*

[1]Department of Radiology, Peking University First Hospital, Beijing, China, [2]Department of Urology, Peking University First Hospital, Institute of Urology, National Research Center for Genitourinary Oncology, Peking University, Beijing, China, [3]MR Collaboration, United Imaging Research Institute of Intelligent Imaging, Beijing, China

Purpose: To evaluate the efficacy of high b-value diffusion-weighted imaging (DWI) with a continuous-time random-walk (CTRW) diffusion model in determining the pathological grade and variant histology (VH) of bladder cancer (BCa).

Methods: A total of 81 patients (median age, 70 years; range, 35-92 years; 18 females; 66 high grades; 30 with VH) with pathologically confirmed bladder urothelial carcinoma were retrospectively enrolled and underwent bladder MRI on a 3.0T MRI scanner. Multi-b-value DWI was performed using 11 b-values. Three CTRW model parameters were obtained: an anomalous diffusion coefficient (D) and two parameters reflecting temporal (α) and spatial (β) diffusion heterogeneity. The apparent diffusion coefficient (ADC) was calculated using b0 and b800. D, α, β, and ADC were statistically compared between high- and low-grade BCa, and between pure urothelial cancer (pUC) and VH. Comparisons were made using the Mann–Whitney U test between different pathological states. Receiver operating characteristic curve analysis was used to assess performance in differentiating the pathological states of BCa.

Results: ADC, D, and α were significantly lower in high-grade BCa compared to low-grade, and in VH compared to pUC ($p < 0.001$), while β showed no significant differences ($p > 0.05$). The combination of D and α yielded the best performance for determining BCa grade and VH (area under the curves = 0.913, 0.811), significantly outperforming ADC (area under the curves = 0.823, 0.761).

Abbreviations: ADC, Apparent diffusion coefficient; AUC, Area under the curve; BCa, Bladder cancer; CI, Confidence interval; CTRW, Continuous-time random-walk; DWI, Diffusion weighted imaging; FROC, Fractional-order calculus; ICC, Interclass correlation coefficients; IQR, Interquartile range; OPC, Optimal parabolic contour; pUC, Pure urothelial carcinoma; ROC, Receiver operating characteristic; ROI, Region of interest; SI, Signal intensity; TURBT, Transurethral resection of bladder tumor; VH, Variant histology.

Conclusion: The CTRW model effectively discriminated pathological grades and variants in BCa, highlighting its potential as a noninvasive diagnostic tool.

KEYWORDS

urinary bladder neoplasms, pathology, neoplasm grading, tumor microenvironment, diffusion magnetic resonance imaging

Introduction

The pathological status of bladder cancer (BCa) significantly influences its management strategies. BCa is classified as low- or high-grade based on the extent of nuclear anaplasia and architectural abnormalities (1). Transurethral resection of bladder tumor (TURBT) remains the cornerstone of treatment for low-grade, non-muscle invasive BCa. For lesions assessed as high-grade upon resection, a repeat TURBT is recommended to ensure comprehensive removal (2). High-grade BCa carries an elevated risk for muscle invasion or metastatic disease (3). Predominantly, BCa manifests as pure urothelial carcinoma (pUC); however, up to 25% of cases present morphological features that differ from pUC, including different histologic subtypes and/or divergent differentiations (1, 4), which can be indicative of a more aggressive disease trajectory and adverse outcomes (5, 6). Therefore, accurate evaluation of grade and variant is crucial for tailoring therapeutic strategies and forecasting prognosis for BCa patients.

Identifying BCa pathology relies on invasive TURBT (7). However, insufficient TURBT may lead to underestimation and inaccuracy (8). The apparent diffusion coefficient (ADC), a traditional measurement of diffusion in tissue, has been identified as a promising imaging biomarker for detecting BCa pathology (9). Several studies have reported an inverse correlation between ADC and the grade of BCa (10–13). The underlying mechanism for the correlation of ADC values with pathological characteristics involves microstructural changes within malignancy, including larger cell size and density (9). Besides pathological grade, the importance of reporting variant histology (VH) in BCa has been emphasized in recent years (14, 15). The use of DWI and ADC in evaluating variants is lacking but holds significant potential, as the different histological components within UC may create more complex tissue microenvironments than pUC (16). However, water diffusion in complex tumor tissues exhibits a non-Gaussian distribution, which cannot be simply reflected by ADC from the Gaussian diffusion model (17–19). Therefore, non-Gaussian diffusion models are more applicable for characterizing actual microenvironments within complex structures (20).

Recent preliminary studies have reported that high b-value diffusion-weighted imaging (DWI) based on a non-Gaussian continuous-time random-walk (CTRW) model shows promise in diagnosing and evaluating brain, breast, and liver diseases (21–24). Three diffusion parameters derived from the CTRW model—diffusion coefficient (D), temporal diffusion heterogeneity (α), and spatial diffusion heterogeneity (β)—can quantitatively reflect water molecular diffusion and intravoxel structural heterogeneity (22). The fractional order calculus (FROC) model is a simplification of the CTRW model, focusing only on spatial heterogeneity (25). A recent study demonstrated the utility of the FROC model in grading BCa (19), indicating the potential value of non-Gaussian diffusion models in evaluating BCa pathology. However, the performance of the CTRW diffusion model in assessing BCa pathological status remains unknown and warrants further investigation. This study aimed to investigate the efficacy of quantitative diffusion parameters derived from the CTRW model in characterizing the pathological grade and VH of BCa and to compare it with the conventional mono-exponential model.

Materials and methods

This retrospective, single-institution study received approval from the hospital's ethics committee, and the requirement for written informed consent was waived.

Patient enrollment

A total of 107 patients with suspected bladder tumors, who had not received previous treatment and underwent bladder MRI, were consecutively enrolled from August 2022 to November 2023. Patients were excluded if they met any of the following criteria: (1) absence of pathologically confirmed urothelial carcinoma post-MRI (n = 12); (2) poor image quality for diffusion images (n = 1); or (3) lesion diameter less than 5 mm, precluding reliable analysis (n = 4) (10, 26). Notably, if the tumor could not be entirely resected during TURBT, only a biopsy or partial resection was performed for pathological evaluation at our institution. Consequently, 9 patients were excluded from the study due to incomplete pathological information from TURBT, which did not provide comprehensive data for the entire tumor.

MRI acquisition

The MR examinations were performed on a 3.0 T MR scanner (uMR 790, United Imaging Healthcare), with patients in the supine

position, using a 12-channel body phased-array surface coil. A series of axial diffusion-weighted images were obtained for CTRW model analysis using a single-shot spin-echo echo-planar imaging sequence with the following protocol: repetition time/echo time = 2525 ms/56.9 ms, field of view = 240 × 240 mm^2, matrix = 96 × 86, slice thickness = 4 mm, intersection gap = 0.4 mm, acquisition time = 7 min, and b-values = 0_1, 50_1, 100_1, 200_1, 400_1, 800_2, 1000_4, 1500_6, 2000_8, 2500_{14}, and 3000_{16} s/mm^2 (the subscript denotes the number of excitations). To ensure adequate bladder distention, patients were instructed to drink 500-1000 mL of water 30 minutes before the scan (27).

Image analysis

The CTRW model analysis utilized the following formula (23):

$$S = S_0 E_\alpha[-(bD)^\beta] \quad (1)$$

where S is the signal intensity (SI) at a given b-value, and S_0 is the SI without diffusion weighting. E_α is a Mittag-Leffler function (28), b is the b-value, and D is an anomalous diffusion coefficient typically associated with tissue cellularity. The parameters α and β reflect the intravoxel temporal and spatial diffusion heterogeneity, respectively. Based on Equation (1), three CTRW parameter maps (D, α, and β) were generated pixel-by-pixel by fitting the multi-b-value diffusion imaging data via a Levenberg–Marquardt nonlinear fitting algorithm (21).

The nonlinear fitting proceeded with a segmented approach similar to a previous study (22):

a. Estimating D using the DWI with low b-values (b ≤ 1000 s/mm^2) based on a mono-exponential model.
b. Simultaneously estimating α and β using all DWI images while fixing the estimated D for each voxel.

The Mittag-Leffler function was computed using open-source code written by Roberto Garrappa in MATLAB Central (https://www.mathworks.com/matlabcentral/fileexchange/48154-the-mittag-leffler-function), which implemented the optimal parabolic contour (OPC) algorithm (29) and was based on the inversion of the Laplace transform on a parabolic contour suitably chosen in one of the regions of analyticity of the Laplace transform. Its complete form is depicted in the following Equation (2):

$$E_\alpha(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + 1)} \quad (2)$$

The conventional ADC map was calculated using the mono-exponential model formula for comparison:

$$S = S_0 e^{-bADC} \quad (3)$$

In Equation (3) S_0 and S stand for the SI at b values of 0 and 800 s/mm^2, respectively.

The regions of interest (ROIs) for each patient were manually delineated at the slice with the maximum tumor area on DWI images (b = 800 s/mm^2) by two experienced radiologists, with 10 and 13 years of experience in urological radiology, respectively. Necrotic or cystic areas were excluded from the ROIs (25, 30). The mean values of the CTRW parameters (D, α, and β) and ADC were subsequently extracted from the delineated ROIs by the two radiologists through CTRW model fitting. Interobserver agreement was assessed for all diffusion parameters. The average diffusion parameters from the two readers were analyzed and presented in tables and figures.

Reference standard

The histopathological results from surgeries served as the standard reference for evaluating diagnostic performance. The urologists recorded the location of each transurethral resected tumor using a sector map, which was referenced in the histopathological analysis. All pathological specimens were reviewed by an expert urological pathologist with 15 years of experience. The histopathological information included the histological type, grade, T stage, and histological variants (1).

Statistical analysis

The analysis was performed using SPSS software (version 26) and MedCalc software (version 20). The Shapiro–Wilk test was employed to assess the normality of data distributions. Continuous variables are presented as mean ± standard deviation or median with interquartile range, depending on the normality test results. Student's t-test for normally distributed data or a Mann-Whitney U test for non-normally distributed data was used to determine the statistical significance of differences in diffusion parameters between low- and high-grade BCa, and between pUC and VH. The intraclass correlation coefficients (ICCs) of the quantitative diffusion parameters were assessed to evaluate the interobserver agreement between the two radiologists (ICC < 0.20, poor; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, good; 0.81-1.00, perfect).

Univariate and multivariate logistic regression analyses were used to select the optimal diffusion parameters and integrate them to establish a predictive model. Receiver operating characteristic (ROC) analysis was used to evaluate the diagnostic performance of individual diffusion parameters and their combination. The area under the ROC curve (AUC) was compared using the DeLong test. The cut-off value was determined based on the Youden index. The sensitivity, specificity, and accuracy were calculated using the maximum Youden index. P values less than 0.05 indicated statistical significance.

Results

Patient clinicopathological characteristics

A total of 81 patients (18 females; median age, 70 years; interquartile range, 13 years) with 81 lesions (median diameter, 2.4 cm; interquartile range, 2.9 cm) were included in the present

study. Twenty-four patients presented with multiple bladder tumors; the largest tumor was designated as the index tumor for subsequent analysis. The study flowchart is shown in Figure 1. All patients underwent TURBT after MRI. Among them, 25 patients underwent subsequent cystectomies for further treatment after TURBT. Pathology revealed 43 (53.1%) nonmuscle invasive BCa (4 pTa + 39 pT1) and 38 (46.9%) muscle invasive BCa (16 pT2 + 11 pT3 + 11 pT4). Of these 81 BCa lesions, 15 (18.5%) were low grade, and 66 were high grade. Thirty (37.0%) lesions exhibited other histological types, including 12 squamous, 3 sarcomatoid, 2 nested, 2 glandular, 2 small cell, 2 micropapillary, and 7 mixed (3 glandular + squamous, 2 squamous + sarcomatoid, 1 glandular + sarcomatoid, 1 microcystic + sarcomatoid). The patients' clinical and pathological information is listed in Table 1.

Comparison of ADC and CTRW parameters between different pathological groups

Interobserver agreement for mean ADC, D, α, and β values were excellent, with all the ICCs > 0.8 [ICC (95% confidence interval): ADC, 0.936 (0.903-0.959); D, 0.947 (0.919-0.966); α, 0.900 (0.849-0.935); and β, 0.887 (0.830-0.926)]. The average diffusion MRI metrics measured by the two radiologists were used in the analysis. The detailed values of each diffusion parameter and their differences between the different pathological groups are listed in Table 2.

Significant decreases in ADC, D, and α values were observed in high-grade compared to low-grade BCa, as well as in VH relative to pUC. No significant difference in β was found between the low- and high-grade or between the pUC and VH ($p = 0.808$ and $p = 0.059$, respectively). Boxplots illustrating the comparison of ADC and CTRW parameters across different pathological groups are shown in Figure 2. Representative colormaps of ADC and CTRW parameters (D, α, and β) are shown in Figures 3 and 4. Parametric colormaps were smoothed via bilinear interpolation embedded in the 'shading interp' function in MATLAB.

Performance of the diffusion parameters for pathological evaluation

The diagnostic performance of ADC and CTRW parameters for differentiating between low- and high-grade BCa and between pUC and VH are listed in Table 3. The ROC curves for each parameter and their combinations are shown in Figure 5. Among the diffusion parameters for determining BCa grade, α exhibited the highest AUC (0.897), with D showing slightly higher AUC than ADC (0.852 vs. 0.823). However, no statistically significant differences were found in AUCs among α, D, and ADC (all p values > 0.05). In distinguishing VH from pUC, the highest AUC was found for D (0.794), which was significantly higher than that of ADC (0.761, $p = 0.024$). The α had similar AUC to ADC (0.765 vs. 0.761) with no significant difference between them ($p = 0.951$).

Univariate logistic regression analysis revealed that D and α were significantly associated with high grade and VH ($p < 0.05$). Multivariate logistic regression demonstrated that the combination of D and α (D+α) yielded significantly higher AUC than ADC for differentiation between low- and high-grade BCa (0.913 vs. 0.823, $p = 0.043$), and between pUC and VH (0.811 vs. 0.761, $p = 0.026$). For determining the pathological grade, the combination of D and α showed higher sensitivity (92.42% vs. 84.85%), higher specificity (80.00% vs. 73.33%), and higher accuracy (90.12% vs. 82.72%) compared to ADC. For the evaluation of VH, the specificity and accuracy of the combination of CTRW parameters were increased compared to ADC (72.55% vs.

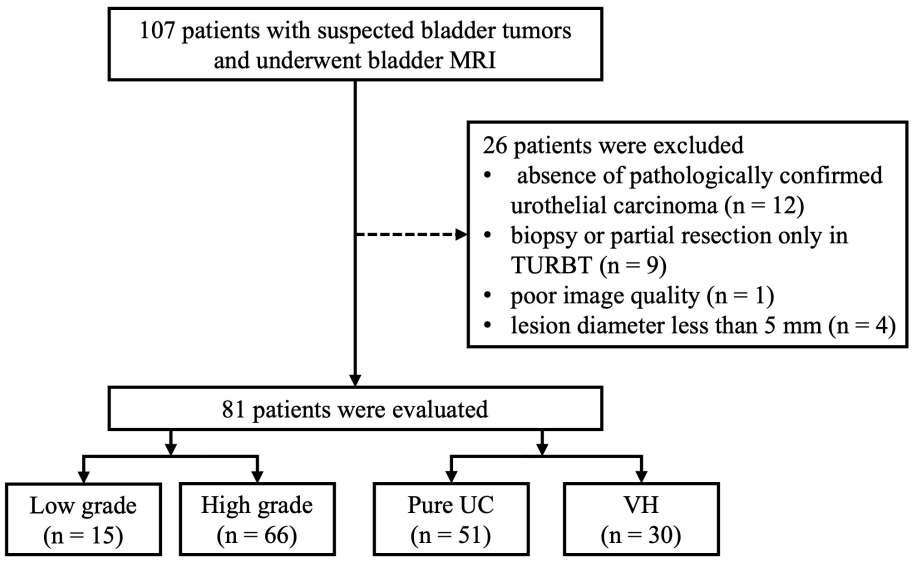


FIGURE 1
Flowchart of the study. TURBT, transurethral resection of bladder tumor; UC, urothelial carcinoma; VH, variant histology.

TABLE 1 Clinicopathological characteristics of the patients.

Variables	Characteristics	
Age (years)*	70 (13)	
Gender	Male	63
	Female	18
No. of lesions	Single	57
	Multiple	24
Tumor size (cm)*	2.4 (2.9)	
Pathological T stage	pTa	4
	pT1	39
	pT2	16
	pT3	11
	pT4	11
Pathological grade	Low	15
	High	66
Variant histology	Absence	51
	Presence	30
Surgery	TURBT	56
	Radical cystectomy	25

*Numbers are medians, with interquartile range in parentheses.
TURBT, Transurethral resection of bladder tumor.

50.98% and 75.31% vs. 66.67%), but the sensitivity was decreased (80.00% vs. 93.33%).

Discussion

Our study demonstrated that both conventional ADC and CTRW parameters were effective in distinguishing between different pathological grades and variants in BCa. These diffusion parameters (ADC, D, and α) significantly differed between low- and high-grade BCa and between pUC and VH. Notably, the α parameter derived from the CTRW model showed the highest performance for BCa grading, while the D parameter from the CTRW model was most effective in differentiating VH from pUC. Furthermore, the combination of CTRW parameters (D+α) achieved a higher AUC than individual parameters, marking a significant improvement over the conventional ADC. These findings suggest that the CTRW model might be a potential tool for noninvasively investigating the pathological characteristics of BCa, offering a supplementary diagnostic approach to the current clinical reliance on TURBT.

Diffusion MRI stands as a powerful tool for probing biological microstructures and has shown potential as an imaging-based marker for predicting the pathological features of BCa (9, 31, 32). The mono-exponential model and its quantitative ADC parameter assume that water molecules follow Gaussian motion in a homogeneous medium. However, tissue structures, especially tumors, present heterogeneous and complex cellular and subcellular microstructures, influencing the diffusion of water molecules in a manner that diverges from a Gaussian distribution. Therefore, ADC may not be accurate enough to evaluate the heterogeneous pathological microstructure of BCa (17). Unlike the mono-exponential diffusion model, non-Gaussian diffusion models can reveal intravoxel tissue cellularity, the extracellular matrix, vascularity, and microstructures of tumors (33). The CTRW model provides insights into the intravoxel heterogeneity of water molecule diffusion both temporally (α) and spatially (β). Our study revealed that the CTRW model could not only aid in evaluating the pathological grade and variant of BCa but also exhibit superior diagnostic capabilities when compared to ADC, reinforcing its potential role in enhancing BCa diagnosis and treatment strategies.

The pathological condition plays a crucial role in guiding the management and forecasting of BCa outcomes (1). The determination of pathologic grade and histologic type typically relies on invasive TURBT procedures, but inadequate resection of the tumor may result in an inaccurate diagnosis (7). BCa grading is based on the organization of cytologic features and architectural abnormalities of the papillae (34). Low-grade BCa exhibits minimal variability in architecture and cytologic features. In contrast, high-grade BCa is characterized by a disorderly appearance resulting from marked architectural and cytologic abnormalities (35). The D parameter from the CTRW model describes the anomalous diffusion process and is analogous to ADC, which is related to tumor cellularity. Our results revealed significantly lower ADC and D values in high-grade compared to low-grade BCa, consistent with previous studies (10–13). This decrease

TABLE 2 Comparison of ADC and CTRW parameters between different pathological groups (median, IQR).

Parameter	ADC ($\times10^{-3}$ mm^2/s)	D ($\times10^{-3}$ mm^2/s)	α	β
Low grade	1.421 (0.348)	1.583 (0.682)	0.898 (0.067)	0.870 (0.148)
High grade	1.024 (0.313)	1.021 (0.306)	0.783 (0.099)	0.868 (0.074)
p value	< 0.001*	< 0.001*	< 0.001*	0.808
pUC	1.206 (0.418)	1.257 (0.556)	0.860 (0.141)	0.852 (0.114)
VH	0.953 (0.225)	0.934 (0.179)	0.774 (0.066)	0.893 (0.064)
p value	< 0.001*	< 0.001*	< 0.001*	0.059

*Significant difference with $p < 0.001$.
ADC, apparent diffusion coefficient; CTRW, continuous-time random-walk; IQR, interquartile range; pUC, pure urothelial carcinoma; VH, variant histology.

10.3389/fonc.2024.1431536

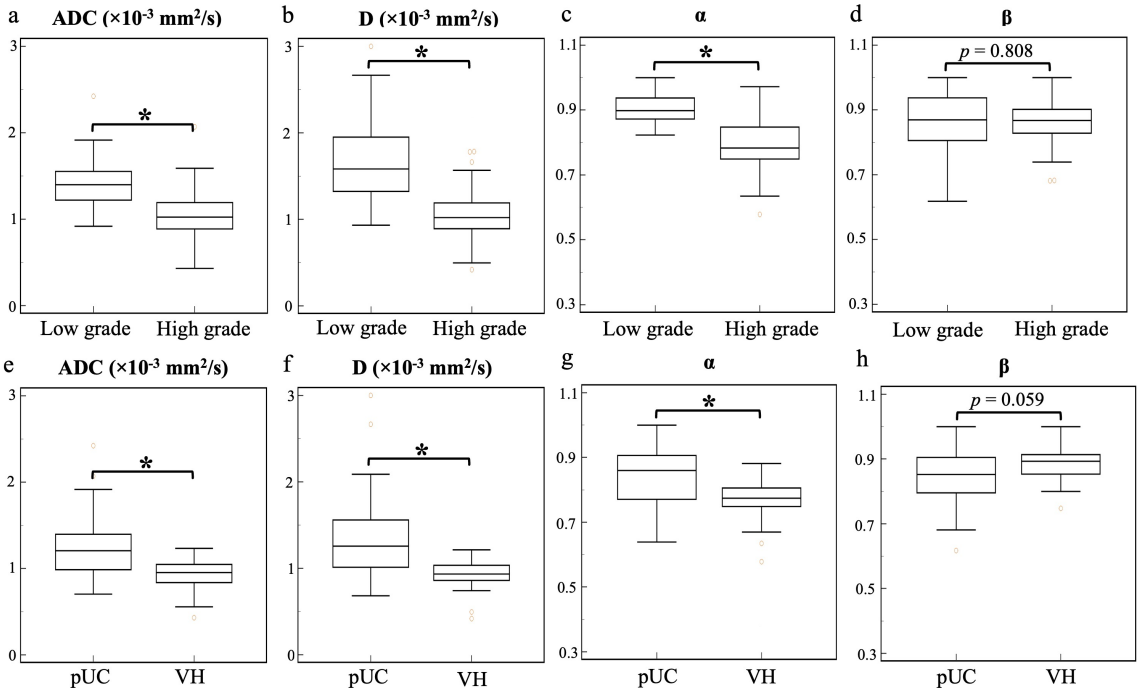


FIGURE 2
Box-and-whisker plots of the mean apparent diffusion coefficient and continuous-time random-walk parameters (D, α, and β) for low- and high-grade bladder cancer (top row, **A-D**) and pure urothelial carcinoma and variant histology (bottom row, **E-H**). The statistically significant level, *$p <$ 0.001. ADC, apparent diffusion coefficient; pUC, pure urothelial carcinoma; VH, variant histology.

can be attributed to the heightened tumor cellularity and decreased extracellular space of high-grade BCa, which consequently impedes water molecule diffusion (36). In addition, high-grade BCa showed increased tumor microstructure heterogeneity, with pronounced architectural and cytologic atypia (37). Our results showed that the temporal diffusion heterogeneity-sensitive parameter α surpassed other diffusion metrics in accurately distinguishing low- from high-grade BCa, indicating the utility of tumor heterogeneity as a grading parameter.

Bladder urothelial carcinoma is recognized for its morphological diversity, encompassing various histological subtypes and divergent differentiations (1). A careful morphological description of the presence of variants is crucial in BCa, as it may affect management and survival expectations (4). Mixed histological components reflect a

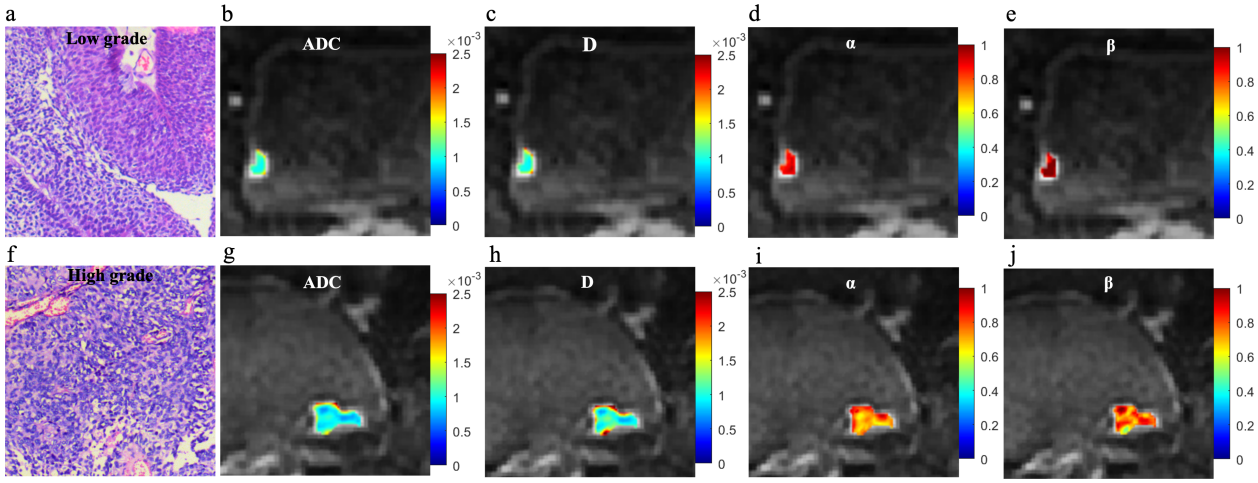


FIGURE 3
Microscopic pathology and colormaps of the apparent diffusion coefficient and continuous-time random-walk parameters of bladder tumors from a 61-year-old male with low-grade urothelial carcinoma (top row, **A-E**) and an 81-year-old female with high-grade urothelial carcinoma (bottom row, **F-J**). The mean values of diffusion parameters were lower in high-grade bladder cancer compared to low-grade bladder cancer.

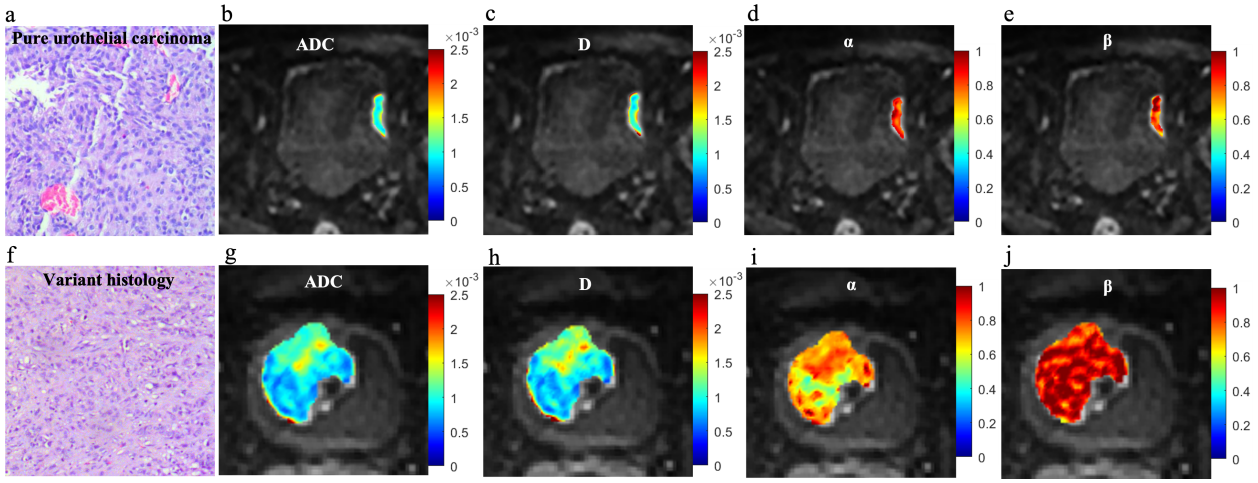


FIGURE 4
Microscopic pathology and colormaps of the apparent diffusion coefficient and continuous-time random-walk parameters of bladder tumors from a 71-year-old male with pure urothelial carcinoma (top row, **A-E**) and a 63-year-old male with urothelial carcinoma and sarcomatoid variant (bottom row, **F-J**). The urothelial carcinoma with variants showed lower ADC, D, and α values than pure urothelial carcinoma.

more heterogeneous tumor microenvironment. Our study observed a lower α value in VH compared to pUC, reflecting the complex and heterogeneous histological components of VH. Additionally, the difference in ADC and D values may be attributed to the varying degrees of cellularity between pUC and VH. The D derived from CTRW presented the highest AUC in the evaluation of VH. The results of this study suggested that the quantitative parameters from the CTRW model could potentially aid pathologists in identifying concurrent histological variants with urothelial carcinoma.

Both α and β are indicative of diffusion heterogeneity within the tumor microstructure. The α parameter is associated with temporal diffusion heterogeneity, revealing the likelihood of water molecules being "retained" or "released" during diffusion, resulting in a variable time during each movement (22). Our findings showed lower α values in higher grades and VH, suggesting a more variable time for diffusing water molecules to travel through them and indicating greater temporal inhomogeneity. The β parameter is related to spatial diffusion heterogeneity and describes the different step lengths of water molecules during diffusion (23). Inconsistency was observed between the α and β values in this study, with no significant differences in the β values between low- and high-grade BCa or between pUC and VH. Water molecules can walk a variable spatial length during each move or spend a variable temporal interval to make a move. The discordant results between α and β in our study could be attributed to the variable time required by water molecules for movement in high-grade or VH BCa, without necessarily resulting in significantly varied displacements. Similarly, previous studies have reported inconsistent changes in α and β in other tumors (24, 38), suggesting that α and β reflect the heterogeneous microstructure in different aspects.

Compared to the single ADC, the CTRW model has the advantage of integrating multiple parameters. The CTRW model encapsulates the attributes of a non-Gaussian distribution and takes into consideration the underlying tumor tissue cellularity and

TABLE 3 Diagnostic performance of diffusion parameters for distinguishing different pathological group.

	ADC	D	α	β	CTRW (D+α)
Low vs. high grade					
AUC (95% CI)	0.823 (0.723-0.899)	0.852 (0.755-0.921)	0.897 (0.809-0.953)	0.520 (0.406-0.633)	0.913 (0.829-0.964)
Sensitivity (%)	84.85	86.36	84.85	77.27	92.42
Specificity (%)	73.33	80.00	93.33	40.00	80.00
Accuracy (%)	82.72	85.19	86.42	70.37	90.12
pUC vs. VH					
AUC (95% CI)	0.761 (0.654-0.849)	0.794 (0.690-0.876)	0.765 (0.657-0.852)	0.626 (0.512-0.731)	0.811 (0.709-0.890)
Sensitivity (%)	93.33	80.00	93.33	76.67	80.00
Specificity (%)	50.98	72.50	64.71	52.94	72.55
Accuracy (%)	66.67	75.31	75.31	61.73	75.31

ADC, apparent diffusion coefficient; AUC, area under the curve; CI, confidence interval; CTRW, continuous-time random-walk; pUC, pure urothelial carcinoma; VH, variant histology.

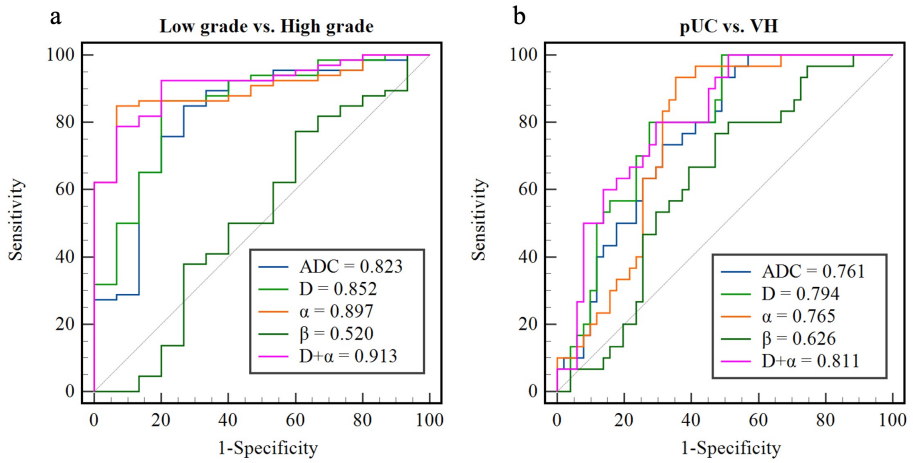


FIGURE 5
Receiver operating characteristic curves of the apparent diffusion coefficient, continuous-time random-walk parameters, and the combination of D and α for differentiating between low- and high-grade bladder cancer **(A)** and between pure urothelial carcinoma and variant histology **(B)**. pUC, pure urothelial carcinoma; VH, variant histology; ADC, apparent diffusion coefficient.

heterogeneity, while ADC does not reveal intravoxel tissue heterogeneity according to the Gaussian distribution. In the present study, the combination of the CTRW parameters (D+α) resulted in the best performance compared with the individual parameters in grade and VH determination. Similarly, Du et al. reported that the combination of CTRW parameters yielded the highest AUC in the differentiation of benign and malignant breast lesions (21). Chang et al. reported that combining multiple CTRW parameters improved the performance of diagnosing molecular subtypes of breast cancer (39). These findings suggest that the combination of the CTRW parameters, which are related to tumor cellularity and heterogeneity, greatly improves the evaluation of pathological changes in tumors compared to individual parameters.

Most clinical MRI scanners allow the DWI signal to be acquired by varying only the diffusion gradient strength (g). Therefore, it is permissible to quantify only D and β by changing gradient strength because in this way D and β have a biophysical meaning. The current α is an estimated parameter extracted from a signal representation model and does not have a biophysical meaning (40). To quantify the true α that characterizes subdiffusion, it is essential to acquire DWI at different diffusion times (Δ). By varying the Δ value while keeping g constant, the decay of the DWI signal depends on Δ and α, where the α quantifies the true subdiffusion (41). In our study, the α does not quantify true subdiffusion because the data were acquired at a fixed diffusion time. Even though, the α parameter was a potential image marker and beneficial in determining pathological states, including grades and variants of BCa. Furthermore, the combination of α and D achieved the greatest discriminating power than individual diffusion parameters.

This study has several limitations. Firstly, the distribution of pathological grades was nonuniform, with more high-grade BCa than low-grade BCa and more pUC than VH, which may introduce bias to the statistical analysis. Secondly, a representative two-dimensional ROI was selected for diffusion parameter measurements, which may not fully reflect whole-tumor characteristics. A three-dimensional whole tumor volume analysis might offer more comprehensive information on BCa but would be more complex and time-consuming. Thirdly, the subgroups of different variants were not further statistically analyzed due to the limited sample size. Future studies with larger sample sizes are needed to validate the role of diffusion models in differentiating different subtypes of variants in BCa.

In conclusion, this study highlighted the distinctiveness of D and α from the non-Gaussian CTRW model and ADC from the mono-exponential model in distinguishing between low- and high-grade BCa as well as between pUC and VH. The CTRW model helped evaluate the grade and variant of BCa. Moreover, the combination of CTRW parameters representing tissue cellularity and heterogeneity outperformed the conventional ADC. Thus, the CTRW model could serve as a promising noninvasive tool, potentially complementing the current pathological evaluation relying on invasive TURBT.

Data availability statement

The datasets presented in this article are not readily available because the clinical and confidential nature of the material but can be made available from the corresponding author on reasonable request. Requests to access the datasets should be directed to JQ, qjx761225@126.com.

Ethics statement

The studies involving humans were approved by the ethics committee of Peking University First Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the retrospective character of this study.

Author contributions

WW: Formal analysis, Investigation, Project administration, Visualization, Writing – original draft. JW: Formal analysis, Investigation, Resources, Writing – original draft, Methodology. QS: Formal analysis, Investigation, Validation, Writing – review & editing, Resources. WL: Resources, Software, Supervision, Writing – original draft, Investigation. KX: Software, Validation, Visualization, Writing – review & editing, Supervision. YY: Supervision, Validation, Visualization, Writing – review & editing, Conceptualization, Resources. JQ: Methodology, Supervision, Writing – review & editing, Conceptualization, Project administration.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author KX and YY work for United Imaging Research Institute of Intelligent Imaging.

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.

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.

References

1. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO classification of tumours of the urinary system and male genital organs—Part B: prostate and bladder tumours. *Eur Urol*. (2016) 70:106–19. doi: 10.1016/j.eururo.2016.02.028

2. National Comprehensive Cancer Network. (NCCN) clinical practice guidelines in oncology. Bladder cancer, version 1. 2024. Available online at: https://www.nccn.org/professionals/physician_gls/pdf/bladder.pdf (Accessed January 2024).

3. Lopez-beltran A, Cookson MS, Guercio BJ, Cheng L. Advances in diagnosis and treatment of bladder cancer. *BMJ*. (2024) 384:e076743. doi: 10.1136/bmj-2023-076743

4. Black AJ, Black PC. Variant histology in bladder cancer: Diagnostic and clinical implications. *Transl Cancer Res*. (2020) 9:6565–75. doi: 10.21037/tcr-20-2169

5. Klaile Y, Schlack K, Boegemann M, Steinestel J, Schrader AJ, Krabbe LM. Variant histology in bladder cancer: How it should change the management in non-muscle invasive and muscle invasive disease? *Transl Androl Urol*. (2016) 5:692–701. doi: 10.21037/tau.2016.06.13

6. Deuker M, Martin T, Stolzenbach F, Rosiello G, Collà Ruvolo C, Nocera L, et al. Bladder cancer: A comparison between non-urothelial variant histology and urothelial carcinoma across all stages and treatment modalities. *Clin Genitourin Cancer*. (2021) 19:60–68.e1. doi: 10.1016/j.clgc.2020.07.011

7. Babjuk M, Burger M, Capoun O, Cohen D, Compérat EM, Dominguez Escrig JL, et al. European association of urology guidelines on non–muscle-invasive bladder cancer (Ta, T1, and carcinoma in situ). *Eur Urol*. (2022) 81:75–94. doi: 10.1016/j.eururo.2021.08.010

8. Hensley PJ, Panebianco V, Pietzak E, Kutikov A, Vikram R, Galsky MD, et al. Contemporary staging for muscle-invasive bladder cancer: accuracy and limitations. *Eur Urol Oncol*. (2022) 5:403–11. doi: 10.1016/j.euo.2022.04.008

9. Yoshida S, Takahara T, Kwee TC, Waseda Y, Kobayashi S, Fujii Y. DWI as an imaging biomarker for bladder cancer. *Am J Roentgenol*. (2017) 208:1218–28. doi: 10.2214/AJR.17.17798

10. Takeuchi M, Sasaki S, Ito M, Okada S, Takahashi S, Kawai T, et al. Urinary bladder cancer: diffusion-weighted MR imaging–accuracy for diagnosing T stage and estimating histologic grade. *Radiology*. (2009) 251:112–21. doi: 10.1148/radiol.2511080873

11. Wang Y, Hu D, Yu H, Shen Y, Tang H, Kamel IR, et al. Comparison of the diagnostic value of monoexponential, biexponential, and stretched exponential diffusion-weighted MRI in differentiating tumor stage and histological grade of bladder cancer. *Acad Radiol*. (2019) 26:239–46. doi: 10.1016/j.acra.2018.04.016

12. Sevcenco S, Ponhold L, Heinz-Peer G, Fajkovic H, Haitel A, Susani M, et al. Prospective evaluation of diffusion-weighted MRI of the bladder as a biomarker for prediction of bladder cancer aggressiveness. *Urol Oncol Semin Orig Investig*. (2014) 32:1166–71. doi: 10.1016/j.urolonc.2014.04.019

13. Yuan L, Li D, Mu D, Zhang X, Kong W, Cheng L, et al. Combined T2 SPAIR, dynamic enhancement and DW imaging reliably detect T staging and grading of bladder cancer with 3.0T MRI. *Front Oncol*. (2020) 10:582532. doi: 10.3389/fonc.2020.582532

14. Netto GJ, Amin MB, Berney DM, Compérat EM, Gill AJ, Hartmann A, et al. The 2022 world health organization classification of tumors of the urinary system and male genital organs—Part B: prostate and urinary tract tumors. *Eur Urol*. (2022) 82:469–82. doi: 10.1016/j.eururo.2022.07.002

15. Lobo N, Shariat SF, Guo CC, Fernandez MI, Kassouf W, Choudhury A, et al. What is the significance of variant histology in urothelial carcinoma? *Eur Urol Focus*. (2020) 6:653–63. doi: 10.1016/j.euf.2019.09.003

16. Alderson M, Grivas P, Milowsky MI, Wobker SE. Histologic variants of urothelial carcinoma: morphology, molecular features and clinical implications. *Bl Cancer*. (2020) 6:107–22. doi: 10.3233/BLC-190257

17. Lin WC, Chen JH. Pitfalls and limitations of diffusion-weighted magnetic resonance imaging in the diagnosis of urinary bladder cancer. *Transl Oncol*. (2015) 8:217–30. doi: 10.1016/j.tranon.2015.04.003

18. Wang F, Wu LM, Hua XL, Zhao ZZ, Chen XX, Xu JR. Intravoxel incoherent motion diffusion-weighted imaging in assessing bladder cancer invasiveness and cell proliferation. *J Magn Reson Imaging*. (2018) 47:1054–60. doi: 10.1002/jmri.25839

19. Feng C, Wang Y, Dan G, Zhong Z, Karaman MM, Li Z, et al. Evaluation of a fractional-order calculus diffusion model and bi-parametric VI-RADS for staging and grading bladder urothelial carcinoma. *Eur Radiol*. (2022) 32:890–900. doi: 10.1007/s00330-021-08203-2

20. Fokkinga E, Hernandez-Tamames JA, Ianus A, Nilsson M, Tax CMW, Perez-Lopez R, et al. Advanced diffusion-weighted MRI for cancer microstructure assessment in body imaging, and its relationship with histology. *J Magn Reson Imaging*. (2023) Published online. doi: 10.1002/jmri.29144

21. Du M, Zou D, Gao P, Yang Z, Hou Y, Zheng L, et al. Evaluation of a continuous-time random-walk diffusion model for the differentiation of Malignant and benign breast lesions and its association with Ki-67 expression. *NMR BioMed*. (2023) 36:e4920. doi: 10.1002/nbm.4920

22. Karaman MM, Sui Y, Wang H, Magin RL, Li Y, Zhou XJ. Differentiating low- and high-grade pediatric brain tumors using a continuous-time random-walk diffusion model at high b-values. *Magn Reson Med*. (2016) 76:1149–57. doi: 10.1002/mrm.26012

23. Karaman MM, Zhang J, Xie KL, Zhu W, Zhou XJ. Quartile histogram assessment of glioma Malignancy using high b-value diffusion MRI with a continuous-time random-walk model. *NMR BioMed*. (2021) 34:1–13. doi: 10.1002/nbm.4485

24. Li C, Wen Y, Xie J, Chen Q, Dang Y, Zhang H, et al. Preoperative prediction of VETC in hepatocellular carcinoma using non-Gaussian diffusion-weighted imaging at high b values: a pilot study. *Front Oncol*. (2023) 13:1167209. doi: 10.3389/fonc.2023.1167209

25. Tang L, Zhou XJ. Diffusion MRI of cancer: From low to high b-values. *J Magn Reson Imaging*. (2019) 49:23–40. doi: 10.1002/jmri.26293

26. Kobayashi S, Koga F, Kajino K, Yoshita S, Ishii C, Tanaka H, et al. Apparent diffusion coefficient value reflects invasive and proliferative potential of bladder cancer. *J Magn Reson Imaging*. (2014) 39:172–8. doi: 10.1002/jmri.24148

27. Panebianco V, Narumi Y, Altun E, Bochner BH, Efstathiou JA, Hafeez S, et al. Multiparametric magnetic resonance imaging for bladder cancer: development of VI-RADS (Vesical imaging-reporting and data system). *Eur Urol*. (2018) 74:294–306. doi: 10.1016/j.eururo.2018.04.029

28. Ingo C, Magin RL, Colon-Perez L, Triplett W, Mareci TH. On random walks and entropy in diffusion-weighted magnetic resonance imaging studies of neural tissue. *Magn Reson Med*. (2014) 71:617–27. doi: 10.1002/mrm.24706

29. Garrappa R. Numerical evaluation of two and three parameter Mittag-Leffler functions. *SIAM J Numer Anal*. (2015) 53:1350–69. doi: 10.1137/140971191

30. Padhani AR, Liu G, Mu-Koh D, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: Consensus and recommendations. *Neoplasia*. (2009) 11:102–25. doi: 10.1593/neo.81328

31. van der Pol CB, Chung A, Lim C, Gandhi N, Tu W, McInnes MDF, et al. Update on multiparametric MRI of urinary bladder cancer. *J Magn Reson Imaging*. (2018) 48:882–96. doi: 10.1002/jmri.26294

32. De Perrot T, Zoua CS, Glessgen CG, Botsikas D, Berchtold L, Salomir R, et al. Diffusion-weighted MRI in the genitourinary system. *J Clin Med*. (2022) 11:1–23. doi: 10.3390/jcm11071921

33. Tang C, Li F, He L, Hu Q, Qin Y, Yan X, et al. Comparison of continuous-time random walk and fractional order calculus models in characterizing breast lesions using histogram analysis. *Magn Reson Imaging*. (2024) 108:47–58. doi: 10.1016/j.mri.2024.01.012

34. Wang G, McKenney JK. Urinary bladder pathology: World Health Organization classification and American joint committee on cancer staging update. *Arch Pathol Lab Med*. (2019) 143:571–7. doi: 10.5858/arpa.2017-0539-RA

35. Kirkali Z, Chan T, Manoharan M, Algaba F, Busch C, Cheng L, et al. Bladder cancer: Epidemiology, staging and grading, and diagnosis. *Urology*. (2005) 66:4–34. doi: 10.1016/j.urology.2005.07.062

36. Hafeez S, Huddart R. Advances in bladder cancer imaging. *BMC Med*. (2013) 11:104. doi: 10.1186/1741-7015-11-104

37. Álvarez KJ, Campos HJP, López BA, Requena TMJ. The 2004 WHO classification of bladder tumors: A summary and commentary. *Actas Urol Esp*. (2007) 31:978–88. doi: 10.1016/s0210-4806(07)73761-0

38. Mao C, Hu L, Jiang W, Qiu Y, Yang Z, Liu Y, et al. Discrimination between human epidermal growth factor receptor 2 (HER2)-low-expressing and HER2-overexpressing breast cancers: a comparative study of four MRI diffusion models. *Eur Radiol*. (2024) 34:2546–59. doi: 10.1007/s00330-023-10198-x

39. Chang H, Wang D, Li Y, Xiang S, Yang YX, Kong P, et al. Evaluation of breast cancer Malignancy, prognostic factors and molecular subtypes using a continuous-time random-walk MR diffusion model. *Eur J Radiol*. (2023) 166:111003. doi: 10.1016/j.ejrad.2023.111003

40. Novikov DS, Kiselev VG, Jespersen SN. On modeling. *Magn Reson Med*. (2018) 79:3172–93. doi: 10.1002/mrm.27101

41. Capuani S, Palombo M. Mini review on anomalous diffusion by MRI: potential advantages, pitfalls, limitations, nomenclature, and correct interpretation of literature. *Front Phys*. (2020) 7:248. doi: 10.3389/fphy.2019.00248

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 August 2024
DOI 10.3389/fonc.2024.1424546

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Kei Shing Ng,
The University of Hong Kong, Hong Kong SAR, China
Zheng Yuan,
China Academy of Chinese Medical Sciences, China
Xiaofeng Li,
Tianjin Medical University Cancer Institute and Hospital, China

*CORRESPONDENCE
Ouwen Li
apocalypse1523@163.com

†These authors have contributed equally to this work

RECEIVED 28 April 2024
ACCEPTED 31 July 2024
PUBLISHED 19 August 2024

CITATION
Xu H, Yu Y, Chang J, Hu X, Tian Z and Li O (2024) Precision lung cancer screening from CT scans using a VGG16-based convolutional neural network. *Front. Oncol.* 14:1424546. doi: 10.3389/fonc.2024.1424546

COPYRIGHT
© 2024 Xu, Yu, Chang, Hu, Tian and Li. 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.

Precision lung cancer screening from CT scans using a VGG16-based convolutional neural network

Hua Xu[1†], Yuanyuan Yu[2,3†], Jie Chang[3], Xifeng Hu[3], Zitong Tian[3] and Ouwen Li[4*]

[1]Department of Infection Control, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Shandong, Jinan, China, [2]Data Science Institute, Shandong University, Jinan, Shandong, China, [3]Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan, China, [4]International Center, Jinan Foreign Language School, Shandong, Jinan, China

Objective: The research aims to develop an advanced and precise lung cancer screening model based on Convolutional Neural Networks (CNN).

Methods: Based on the health medical big data platform of Shandong University, we developed a VGG16-Based CNN lung cancer screening model. This model was trained using the Computed Tomography scans data of patients from Pingyi Traditional Chinese Medicine Hospital in Shandong Province, from January to February 2023. Data augmentation techniques, including random resizing, cropping, horizontal flipping, color jitter, random rotation and normalization, were applied to improve model generalization. We used five-fold cross-validation to robustly assess performance. The model was fine-tuned with an SGD optimizer (learning rate 0.001, momentum 0.9, and L2 regularization) and a learning rate scheduler. Dropout layers were added to prevent the model from relying too heavily on specific neurons, enhancing its ability to generalize. Early stopping was implemented when validation loss did not decrease over 10 epochs. In addition, we evaluated the model's performance with Area Under the Curve (AUC), Classification accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), Sensitivity, Specificity and F1 score. External validation used an independent dataset from the same hospital, covering January to February 2022.

Results: The training and validation loss and accuracy over iterations show that both accuracy metrics peak at over 0.9 by iteration 15, prompting early stopping to prevent overfitting. Based on five-fold cross-validation, the ROC curves for the VGG16-Based CNN model, demonstrate an AUC of 0.963 $\pm$ 0.004, highlighting its excellent diagnostic capability. Confusion matrices provide average metrics with a classification accuracy of 0.917 $\pm$ 0.004, PPV of 0.868 $\pm$ 0.015, NPV of 0.931 $\pm$ 0.003, Sensitivity of 0.776 $\pm$ 0.01, Specificity of 0.962 $\pm$ 0.005 and F1 score of 0.819 $\pm$ 0.008, respectively. External validation confirmed the model's robustness across different patient populations and imaging conditions.

Conclusion: The VGG16-Based CNN lung screening model constructed in this study can effectively identify lung tumors, demonstrating reliability and effectiveness in real-world medical settings, and providing strong theoretical and empirical support for its use in lung cancer screening.

KEYWORDS

lung cancer screening, medical image recognition, computed tomography scans, VGG16 architecture, convolutional neural network

1 Background

Lung cancer is one of the leading causes of cancer-related deaths worldwide, with an estimated 1.8 million deaths, accounting for 18.7% of the total (1–5). Moreover, both the incidence and mortality rates of lung cancer are on the rise, particularly in developing and middle-income countries (6, 7). Given the vast population of China, this implies a large number of new cases and deaths annually, placing immense strain on the healthcare system. In Taiwan, a study assessed low-dose lung computed tomography (CT) screening criteria among Asian ethnic groups, finding that risk-based strategies more effectively identify high-risk non-smokers, emphasizing the need to optimize screening criteria for better early detection (8). Currently, lung CT scans have become a primary tool in modern medicine for early screening of lung diseases, especially lung cancer (9–12). Henschke et al. demonstrated that CT screening can significantly improve the detection rate of early-stage lung cancer compared to traditional screening methods, which is crucial for patient survival rates (13). This further proves the significance of CT screening in the early detection and treatment of lung cancer. However, these place rigorous demands on a physician's interpretive abilities. To effectively interpret these high-resolution images, physicians must possess extensive clinical medical knowledge and expertise in radiology. Subjectivity and variability in interpretation can lead to inconsistent diagnoses and even misdiagnosis. Therefore, there is a pressing need to seek more advanced methods to assist physicians in achieving more accurate and efficient CT screening interpretations, ultimately enhancing diagnostic outcomes and patients' medical experiences.

In recent years, deep learning, especially Convolutional Neural Networks (CNN), has been proven to be an incredibly effective tool for automating and optimizing the analysis process of medical imaging. It has also been widely applied in various medical imaging systems such as Magnetic Resonance Imaging (MRI), CT, and X-rays. Litjens et al. conducted in-depth research on how CNN can improve the accuracy of image classification, segmentation, and disease detection (14). Rajpurkar et al. applied a deep learning model called CheXNeXt to chest X-ray images, demonstrating that CheXNeXt's performance in automatic diagnosis of various chest diseases is comparable to that of practicing radiologists, and in some cases, it even surpasses the performance of human experts (15). These studies provide substantial evidence for the application of deep learning techniques in clinical practice, but still face a series of challenges in areas such as data insufficiency, model overfitting, and interpretability of diagnostic results.

Motivated by the urgent need to improve early lung cancer detection and diagnosis, this study developed a lung cancer screening model based on the CNN architecture. To enhance generalization and reduce overfitting, we employed several strategies. Data augmentation techniques, including random resizing, cropping, horizontal flipping, color jitter, and random rotation, were applied to increase the diversity of the training data. We used five-fold cross-validation to robustly assess performance by ensuring each fold maintained the same ratio of malignant to non-malignant images. The model was fine-tuned with an SGD optimizer (learning rate 0.001, momentum 0.9, and L2 regularization) and a learning rate scheduler. Dropout layers were added to prevent the model from relying too heavily on specific neurons, thus enhancing its ability to generalize. Early stopping was implemented to halt training when validation loss did not decrease over 10 consecutive epochs. In addition, we used a comprehensive set of evaluation metrics, including Area Under the Curve (AUC), Classification accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), Sensitivity, Specificity and F1 score to ensure the model's exceptional performance across diverse criteria. Our aim is to leverage the powerful capabilities of CNNs to provide a cutting-edge, accurate, and effective lung cancer screening method, integrating it into daily medical workflows to improve screening accuracy.

2 Methods

Figure 1 illustrates the comprehensive workflow of our lung cancer screening model using the VGG16 architecture.

2.1 Study population

2.1.1 Training population

2.1.1.1 Construction of the imaging dataset for patients with malignant lung tumors

Based on the health medical big data platform of Shandong University, we gathered a comprehensive dataset from Pingyi

Traditional Chinese Medicine Hospital in Shandong Province, including lung CT scans and EHR data such as age, gender, diagnostic variables, pathological reports, and patient identifiers, collected over January and February 2023 (as shown in Figure 1A). The pathological report clearly indicated whether the patient had a malignant lung tumor. All patient data were anonymized and collected using encrypted identification numbers to protect privacy and comply with ethical standards. As shown in Figure 1B, the inclusion criteria for this study were: 1) Patients with malignant lung tumors who had undergone procedures such as puncture, bronchoscopy, or surgery to confirm the definitive pathological diagnosis. 2) Patients who must have undergone CT scans prior to percutaneous puncture, bronchoscopy biopsy or surgery. 3) Patients who had lung nodules discovered during the CT scan examination. The exclusion criteria were: 1) Patients with concurrent other lung lesions. 2) Patients with a history of lung diseases. 3) Patients with metastatic tumors. Ultimately, a total of 141 patients with malignant lung tumors were included in this study.

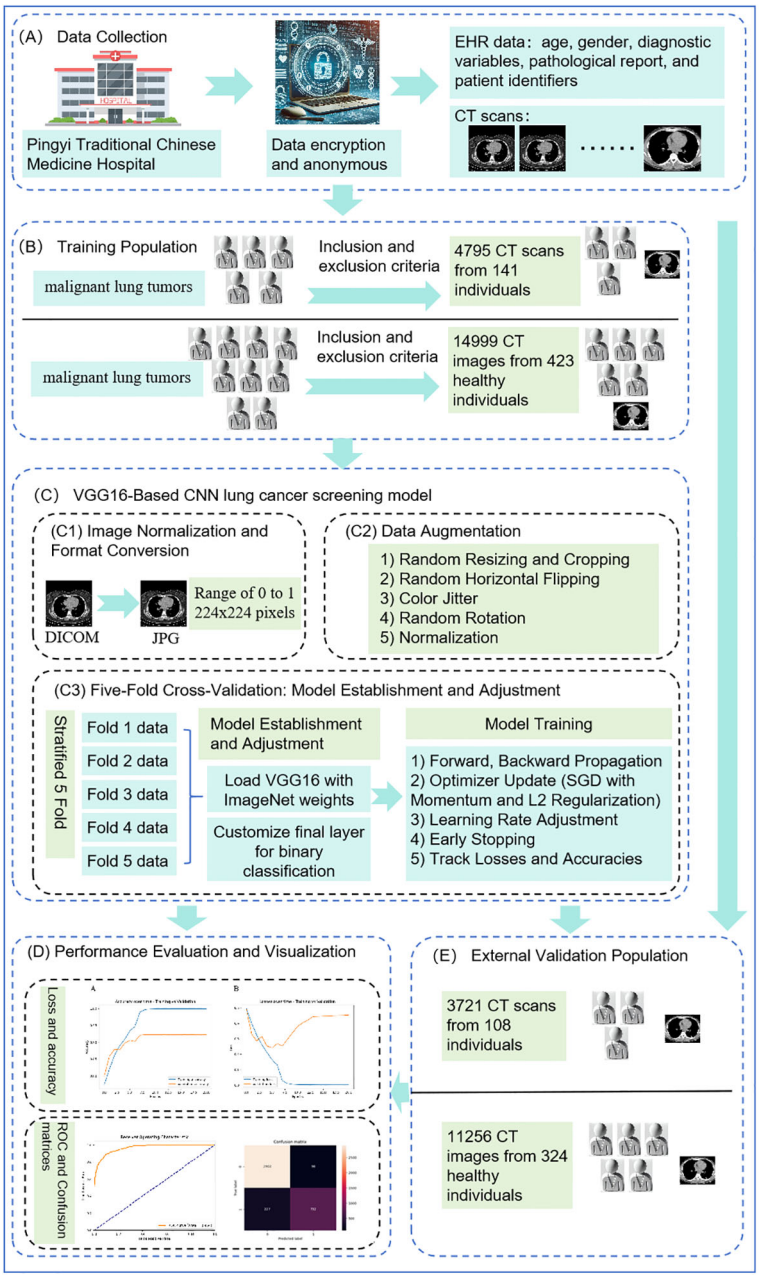


FIGURE 1
Comprehensive workflow of VGG16-based convolutional neural network. **(A)** shows data collection, **(B)** indicates the process of identifying the training population, **(C)** illustrates the construction of the VGG16-Based CNN lung cancer screening model, **(D)** depicts model evaluation and visualization, and **(E)** represents external validation.

2.1.1.2 Construction of the dataset for non-malignant lung tumor patients

For the dataset of non-malignant lung tumor patients, we included same lung CT images and detailed metadata from patients at Pingyi Traditional Chinese Medicine Hospital, collected between January and February 2023. The inclusion criteria required patients to have undergone CT scans. Exclusion criteria (Figure 1B), consistent with those for the malignant lung tumor datasets, included patients with concurrent lung lesions, a history of lung diseases, or metastatic tumors. Using a random match method at a ratio of 1:3, a total of 423 non-malignant lung tumor patients were included.

The final dataset was composed of 4795 CT images from 141 individuals with malignant lung tumors and 14999 CT images from 423 individuals with non-malignant lung tumor.

2.1.2 External validation population

An external validation dataset was obtained from Pingyi Traditional Chinese Medicine Hospital in Shandong Province, comprising imaging data collected between January and February 2022 (Figure 1E). The inclusion and exclusion criteria were consistent with those used for the primary dataset. This external dataset included 108 patients with malignant lung tumors and 324 patients with non-malignant lung tumors, providing a robust basis for validating the performance of the developed model.

In clinical research, especially with smaller sample sizes, sparse data bias can significantly affect the reliability of statistical estimates. By using a 1:3 random matching design, we increased the sample size of the malignant group, thus reducing the impact of sparse data bias and ensuring more stable and reliable estimates of effect sizes and other statistical parameters (16). In addition, random matching helps to minimize the potential biases that can arise from non-random selection processes. By randomly selecting non-malignant cases at a 1:3 ratio, we aimed to ensure a representative sample of the broader population, thus enhancing the generalizability of our results (17).

Our study utilized both plain and contrast-enhanced CT scans. The choice between these two techniques was based on specific clinical indications and the patient's condition at the time of imaging, ensuring that the most appropriate diagnostic approach was utilized for each case. The imaging was performed using three different models of CT scanners, including a Philips 256-slice CT, a 64-slice CT, and a 16-slice CT. This diverse array of scanning equipment enabled a comprehensive assessment of the conditions under study, ensuring a robust analysis through varied imaging capabilities.

This study was approved by the Institutional Review Board of Pingyi Traditional Chinese Medicine Hospital in Shandong Province, China. The ethical approval number for our study is PYX2YYYLLWYh2023030602. It is important to note that no informed consent was required from the participants for this study, as only anonymized lung CT images and detailed metadata were used. The anonymization process was rigorously conducted before the data was accessed for research purposes, ensuring that all personal identifiers were removed to protect patient confidentiality.

2.2 VGG16-based CNN lung cancer screening model

The Convolutional Neural Network (CNN) is a cornerstone algorithm in deep learning, especially adept at image processing tasks. Since the advent of LeNet in 1998, CNN has become mainstream in computer vision tasks (18). This study attempts to use the VGG16 model to construct a lung cancer screening model. The choice of the VGG16 architecture for lung cancer screening is driven by its distinguished capabilities in handling complex image data and its historical success in diverse image recognition tasks (18). VGG16's architectural depth and uniformity are ideal for medical imaging, where precision and reliability are paramount. The architecture features 13 convolutional layers that are exceptionally effective at extracting multi-scale features, a fundamental requirement for identifying subtle and critical anomalies in medical images (19, 20). Moreover, VGG16's robustness and adaptability in processing new and varying datasets make it an exemplary choice for the dynamic requirements of medical diagnostics, as demonstrated by its proven efficacy in diagnosing conditions such as pneumonia from chest X-rays, papillary thyroid carcinomas from cytological images, and brain tumors from MRI scans (15, 21, 22). This combination of deep learning efficiency and versatility underscores why VGG16 is uniquely suited for developing a lung cancer screening model (Figure 1C).

2.2.1 Data preparation

2.2.1.1 Image normalization and format conversion

We use *pydicom* to extract DICOM images and normalize their intensity values to a range of 0 to 1, which is essential for consistent CNN performance. Subsequently, we utilize the Figure module from *matplotlib* to display and manipulate image plots, aiding in converting these images to JPG format (23, 24).

2.2.2 Image pre-processing

2.2.2.1 Resizing and normalization

All images were resized to 224x224 pixels to meet the input requirements of the VGG16 model and were normalized to a range of 0 to 1 to standardize the input (19).

2.2.2.2 Data augmentation

To enhance the model's generalization capabilities and prevent overfitting, data augmentation techniques were applied using the *torchvision.transforms* library. The techniques used included 1) Random Resizing and Cropping: Images were randomly resized and cropped to provide a variety of image sizes and perspectives to the model. 2) Random Horizontal Flipping: Images were randomly flipped horizontally to make the model invariant to left-right orientation. 3) Color Jitter: Random adjustments to brightness, contrast, saturation, and hue to introduce variability. 4) Random Rotation: Images were randomly rotated to make the model invariant to orientation. 5) Normalization: Pixel values were normalized to ensure consistent intensity values across all images (19). By applying these augmentations, we increased the diversity of

our training data, prevented overfitting, and ensured that the model could generalize well to unseen data.

2.2.3 Five-fold cross-validation

We implemented the five-fold cross-validation method to reduce the risk of overfitting to a particular subset of the data. This method involved splitting the dataset into five parts, ensuring each fold maintained the same ratio of malignant to non-malignant images. Each fold served as a validation set once, while the remaining four folds constituted the training set (25).

2.2.4 Model establishment and adjustment

We used the VGG16 architecture, which is a well-known deep CNN model pre-trained on the ImageNet dataset. VGG16 consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. Given our task of binary classification (malignant lung tumors vs. non-malignant lung tumors), we modified the final fully connected layer of the VGG16 model to output two classes (19). Specific layers of the pre-trained model were unfrozen to allow fine-tuning during training (26).

2.2.5 Model training

During each iteration, the model underwent forward and backward propagation on the training data. The SGD optimizer, with a learning rate of 0.001 and momentum of 0.9, and weight decay of 0.0005 (L2 regularization), was used to update the model weights. The binary cross-entropy loss function was employed to calculate the loss and both training and validation losses and accuracies were recorded for performance evaluation (27).

2.2.6 Preventing overfitting

Several strategies were employed to prevent overfitting, including: 1) Dropout layers were added to the model to randomly drop neurons during training, which helps in preventing overfitting. 2) A learning rate scheduler was used to dynamically adjust the learning rate during training, helping to fine-tune the model and avoid overfitting. 3) Early stopping was implemented, terminating training if there was no significant decrease in validation loss over 10 consecutive epochs (28).

2.2.7 Performance evaluation and results visualization

Training and validation loss and accuracy were recorded at the end of each epoch. These metrics were plotted to visualize the learning progress and identify potential overfitting or underfitting issues (27). As shown in Figure 1D, the ROC curve was plotted to evaluate the trade-off between sensitivity and specificity at various threshold settings (29). The area under the ROC curve (AUC) was calculated to quantify the overall ability of the model to discriminate between positive and negative cases (29). A confusion matrix was generated to provide a detailed breakdown of the model's predictions and their alignment with the actual outcomes. This matrix helped in understanding the distribution of true positives, true negatives, false positives, and false negatives, which is crucial for evaluating the performance of the classification model (29, 30). At the end of Area Under the Curve (AUC), classification accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), Sensitivity, Specificity and F1 score were calculated as evaluation metrics (30).

2.2.8 External validation

To ensure the model's generalizability and robustness, the trained model was further evaluated on the external validation dataset obtained from Pingyi Traditional Chinese Medicine Hospital in Shandong Province (Figure 1E). The AUC was computed, and results were visualized using ROC curves to provide a comprehensive evaluation of the model's performance on an independent dataset.

2.3 Software details

The code was written in Python, and executed on the Jupyter server at the Health and Medical Big Data Research Institute of Shandong University.

3 Result

3.1 General characteristics of patients with malignant lung tumors

A total of 141 patients with malignant lung tumors were included, consisting of 97 males and 44 females. Their ages ranged from 44 years old to 89 years, with a median age of 69 years and an interquartile range (IQR) of 63 to 74 years. Additionally, 423 individuals with no-lung tumors were included, comprising 225 males and 198 females. The age of those without lung tumors ranged from 20 to 95 years, with a median age of 67 years and an IQR of 54.5 to 76.6 years. The frequency distribution of ages for both groups is detailed in Supplementary Material Tables 1 and 2.

3.2 Results for pulmonary imaging based on CNN

Figure 2 shows the loss and accuracy of the training and validation sets during each iteration. As can be seen from Figure 2A, the accuracy of both the training and validation sets generally increases with the number of iterations. By the time the number of iterations reaches 10, the accuracy of both the training and validation sets is at its highest, with both achieving an accuracy of over 90%. As shown in Figure 2B, when the number of iterations reaches 15, the program determines that the validation error has not improved for 10 consecutive epochs. Therefore, we chose an iteration count of 15 for early stopping. This early stopping strategy helps prevent overfitting and ensures the model's robustness by terminating training when no significant improvement in validation loss is observed over 10 consecutive epochs.

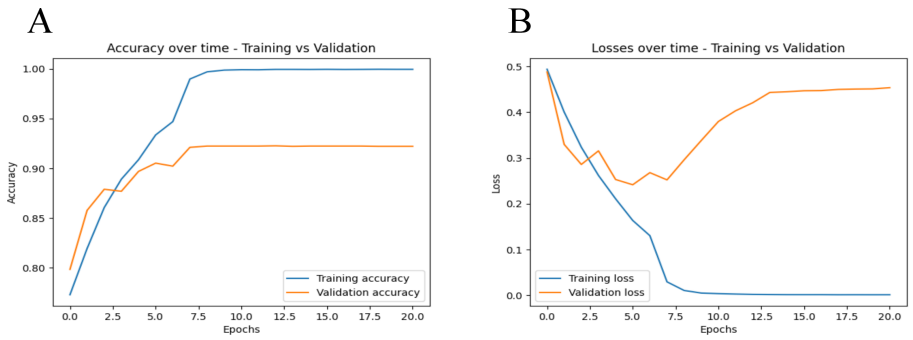


FIGURE 2
Training and validation metrics over iterations. **(A)** displays the accuracy metrics for both training and validation sets across epochs. The blue line represents the training accuracy, while the orange line represents the validation accuracy. **(B)** shows the loss values for both the training and validation datasets over each training epoch. The blue line represents the training loss, while the orange line represents the validation loss.

Figure 3- displays the ROC curves for the five-fold cross-validation and the external dataset, while Table 1 lists the evaluation indices along with their 95% Confidence Intervals (CIs) for five-fold cross-validation. The CNN-based pulmonary imaging diagnostic model consistently demonstrated high diagnostic accuracy, evidenced by an average AUC of 0.963 ± 0.004 across five-fold cross-validation, as depicted in Figures 3A–E. This robust performance is presented in the first row of Table 1, where each fold's AUC score approximates 0.96, with narrow 95% confidence intervals. Such consistently high AUC values across multiple validation folds attest to the model's reliable capability to differentiate between malignant and non-malignant cases.

Additionally, the confusion matrices for all 5 runs are provided in the Supplementary Material Figures 1-5. Based on the confusion matrix, the classification accuracy, PPV and NPV were 0.917 ± 0.004, 0.868 ± 0.015, and 0.931 ± 0.003, respectively, with each fold's accuracy, PPV, and NPV along with their 95% CIs, detailed in lines two to four of Table 1. Classification accuracy measures the overall effectiveness of the model in correctly identifying both positive and negative cases, with a reported accuracy of approximately 91.1%. PPV specifically assesses the accuracy of the model's positive predictions, indicating that approximately 86.6% of the lung cancer diagnoses made by the model are accurate. An NPV of 93.1% ensures a high probability that negative diagnoses are correct, minimizing the risk of missed diagnoses. Together, high PPV and NPV highlight the model's reliability, making it suitable for clinical lung cancer screening. According to the confusion matrix of the five-fold cross-validation, the Sensitivity, Specificity and F1 score were 0.776 ± 0.01, 0.962 ± 0.005 and 0.819 ± 0.008, respectively, with these results for each fold shown in lines five to seven of Table 1. The sensitivity of 77.6% indicates a robust capability to detect true cases of lung cancer, effectively minimizing the risk of missing diagnoses (false negatives). The specificity of 96.2% demonstrates the model's precision in identifying individuals who do not have lung cancer, significantly reducing the occurrence of false positives. Additionally, an F1 score of 81.9% reflects a well-balanced trade-off between precision and sensitivity, ensuring the model's overall accuracy and reliability in medical diagnostics. This comprehensive validation across multiple metrics confirms the model's effectiveness in lung cancer detection.

The external validation showed that our model maintained robust performance (Figure 3F), with an AUC and 95%CI of 0.7564 [0.7563, 0.7564]. The result confirms the model's ability to generalize well across different patient populations and imaging conditions.

4 Discussion

Lung cancer has become a significant public health problem in China, with rising incidence and mortality rates as highlighted in the "Cancer incidence and mortality in China, 2016" report and recent data from the American Cancer Society (ACS) (31, 32). With the advancement of technology and medicine, radiological lung screening has become a pivotal means for early detection and evaluation of pulmonary diseases, especially lung cancer. Despite extensive research and practice, existing methods still have limitations. The survival rate for patients with advanced lung cancer remains low (33, 34), and early diagnosis and treatment are becoming increasingly important (35–37). Aberle et al. have pointed out that compared to traditional chest X-rays, lung cancer screenings using low-dose CT can significantly reduce lung cancer mortality rates (38). This pivotal discovery laid a solid scientific foundation for the promotion and application of low-dose CT screening (39–41). However, traditional imaging analysis methods have certain limitations, such as high missed diagnosis rate, misdiagnosis rate, and limited ability to interpret intricate images (42, 43). Therefore, with the increasing complexity and resolution of medical imaging, there's a growing reliance on advanced computer-aided diagnostic systems to assist in understanding and interpreting these images (44–46). Deep learning techniques, especially CNNs, have revolutionized medical imaging analysis (47).

Our study constructed a precise lung cancer screening model based on CNN. Initial preprocessing and data augmentation ensured the quality and consistency of lung cancer imaging data. Overfitting was mitigated through multiple strategies including data

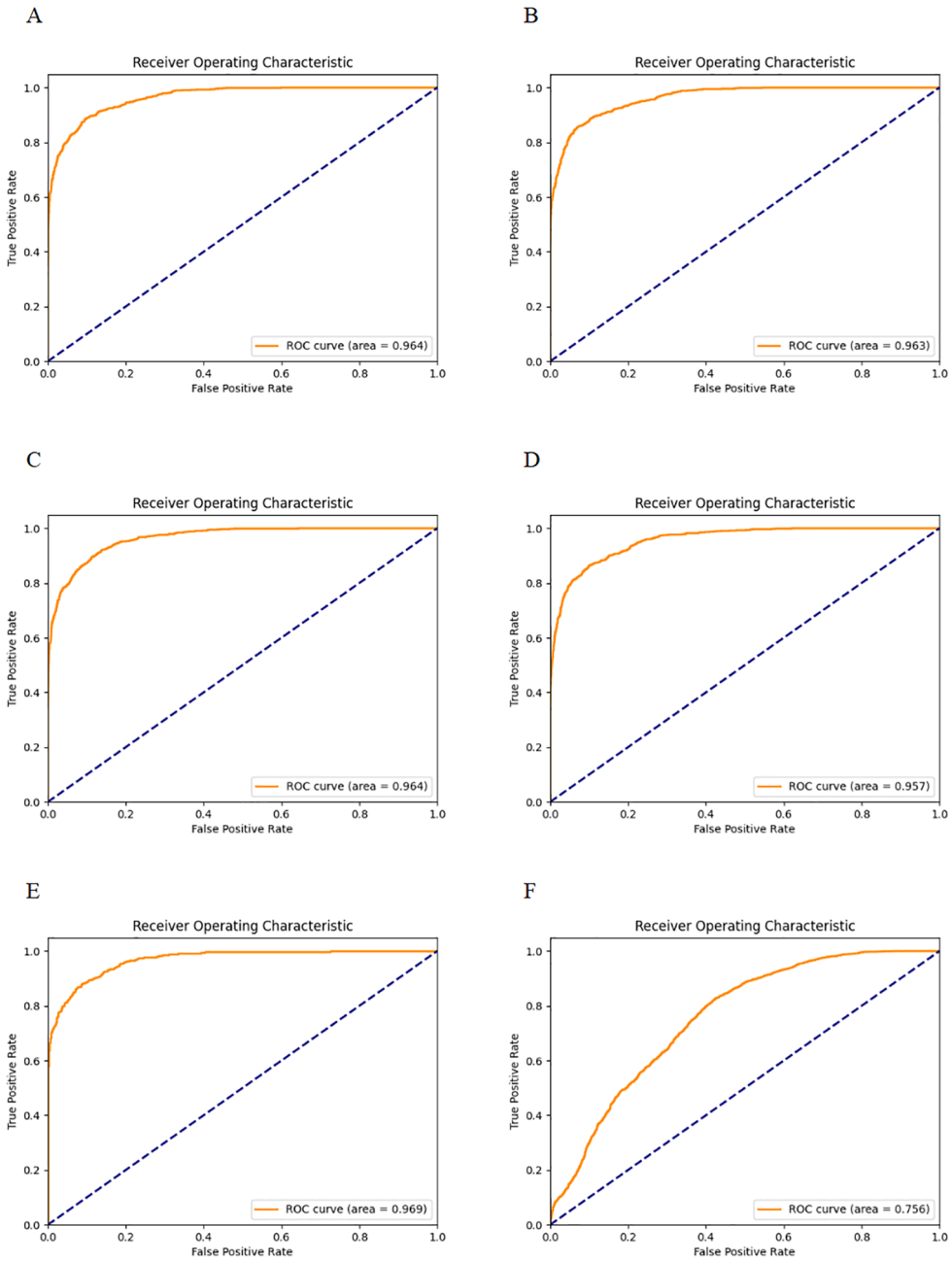


FIGURE 3
ROC curves for five-fold cross-validation and external validation. **(A–E)** display the ROC curves for each fold of the five-fold cross-validation. **(F)** shows the ROC curve for the external validation dataset.

augmentation techniques such as random resizing and cropping, random horizontal flipping, color jitter, and random rotation. L2 regularization was applied via weight decay in the optimizer, and a learning rate scheduler dynamically adjusted the learning rate during training. Dropout layers were added to the model to prevent reliance on specific neurons, and early stopping was implemented to halt training if validation loss did not decrease over 10 consecutive epochs. Five-fold cross-validation demonstrated the model's robustness. Our study achieved an AUC of 0.963 ± 0.004 on the validation set, indicative of excellent diagnostic capability. This high AUC value demonstrated the model's robust capacity to distinguish accurately between lung cancer and those without, minimizing both false positives and false negatives. These performances are consistent with a classification accuracy of 0.917 ± 0.004. Additionally, the PPV, NPV, Sensitivity, Specificity, and F1 score were 0.868 ± 0.015, 0.931 ± 0.003, 0.776 ± 0.01, 0.962 ± 0.005, and 0.819 ± 0.008, respectively. These performance metrics collectively demonstrate the model's

TABLE 1 The evaluation index and it's 95% confidence interval for five-fold cross-validation.

Evaluation index[95%CI]	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
AUC	0.9643 [0.9642, 0.9643]	0.9628 [0.9628, 0.9629]	0.9640[0.9639, 0.9640]	0.9566[0.9565, 0.9567]	0.9690[0.9690, 0.9691]
Classification accuracy	0.9179 [0.9089, 0.9261]	0.9179 [0.9089, 0.9261]	0.9154[0.9063, 0.9237]	0.9113[0.9021, 0.9198]	0.9222[0.9134, 0.9301]
PPV	0.8819[0.8582, 0.9021]	0.8747[0.8507,0.8953]	0.8537[0.8289,0.8755]	0.8486[0.8233,0.8709]	0.8798[0.8563,0.8999]
NPV	0.9275[0.9178,0.936]	0.9296[0.9201,0.9381]	0.9331[0.9237,0.9414]	0.9291[0.9195,0.9376]	0.9339[0.9246,0.9421]
Sensitivity	0.7633[0.7354, 0.7891]	0.7716 [0.7440, 0.7971]	0.7852[0.7581, 0.8100]	0.7716 [0.7440, 0.7971]	0.7862[0.7592, 0.8110]
Specificity	0.9673[0.9604, 0.9731]	0.9647 [0.9574, 0.9707]	0.9570[0.9491, 0.9637]	0.9560[0.9481, 0.9628]	0.9657 [0.9585, 0.9716]
F1 score	0.8183[0.8058, 0.8298]	0.8199 [0.8076, 0.8316]	0.8180 [0.8056, 0.8296]	0.8083[0.7957, 0.8202]	0.8304[0.8182, 0.8416]

high degree of diagnostic accuracy and reliability in identifying lung cancer. The classification accuracy of 91.7% indicates a strong overall ability to correctly classify cases as either having lung cancer or not, supported by a high PPV (86.8%) and NPV (93.1%), which ensure that the positive and negative diagnoses made by the model are likely correct. The sensitivity of 77.6% shows that the model is capable of identifying a substantial majority of true positive cases, critical for early and accurate disease detection. Furthermore, the specificity of 96.2% underscores the model's effectiveness in correctly ruling out disease in healthy individuals, reducing the likelihood of unnecessary treatments. Finally, an F1 score of 81.9% reflects a balanced trade-off between precision and recall, validating the model's utility in clinical settings where both detecting cases and avoiding false alarms are equally important. Together, these metrics not only underline the model's capability but also highlight its potential to significantly enhance patient management and treatment outcomes in clinical practice.

Lu et al. created a CNN model for predicting the long-term incidence of lung cancer, achieving an AUC of 74.9% (48). Ardila et al. reached an AUC of 94.4% using a 3D CNN model for lung nodule detection (49). Cellina et al. reviewed numerous studies on AI applications in lung cancer imaging and diagnosis, reporting AUC values ranging from 87% to 95% (50, 51). This is slightly lower than the AUC of 96.3% achieved by our model, indicating superior performance in distinguishing between lung cancer and non-cancer cases. Additionally, different models reported classification accuracies typically around 90% to 97%, with sensitivity and specificity values ranging from 75% to 95%. Our model's classification accuracy of 91.7% and sensitivity of 77.6% fall within this range. However, our model's sensitivity is slightly lower, suggesting a need for improvement in detecting true positive cases. In contrast, our model's specificity of 96.2% is significantly higher than other studies, indicating its superior ability to correctly identify individuals without lung cancer, minimizing false positives. Comparing positive predictive value (PPV), Hsu et al. achieved 15.0%, whereas our study reported a much higher PPV of 86.8%, reflecting greater accuracy in positive predictions and ensuring fewer false positives. Both studies demonstrated high negative predictive values (NPV), with Hsu et al. at 99.0% and our study at 93.1%, highlighting effectiveness in correctly predicting negative cases (51). Overall, our model's robust performance metrics indicate its potential to significantly enhance lung cancer screening and diagnostic accuracy in clinical settings.

The data collection period from January to February 2023 at Pingyi Traditional Chinese Medicine Hospital was relatively short, which may limit the diversity of our dataset and affect the generalizability of our model. To address this issue, we employed extensive data augmentation techniques, such as random resizing, cropping, horizontal flipping, color jitter, random rotation, and normalization. Despite these efforts, it remains essential to increase the sample size to enhance the model's generalizability. We have planned additional data collection efforts and propose implementing a continuous learning framework to periodically retrain the model with updated data from our hospital and datasets from various regions and medical institutions, ranging from community clinics to large tertiary hospitals. This approach aims to enhance the model's adaptability to variations in clinical practice and its generalizability.

In addition, using data from the same institution collected at different times for external validation may not be ideal. However, existing literature supports the effectiveness of temporal external validation (52, 53), which led us to choose this approach for our study. In addition, the external validation confirmed an AUC of 0.7564 across various patient populations and imaging conditions, which is significantly lower than the training model's AUC of 0.963. The two cohorts, although originating from the same institution, represent different time periods. This temporal difference could potentially affect the model's performance, influenced by changes in patient demographics, CT scanning protocols, or other clinical practices over time. Despite implementing several strategies during model training, such as dropout and L2 regularization, the generalizability of our model across different geographical regions and medical institutions remains a significant challenge. To address this, we outline planned future studies intended to apply our model to datasets collected from different regions, including both urban and rural settings, and from various types of medical institutions

ranging from community clinics to large tertiary hospitals. This will allow us to assess its performance and adaptability in diverse healthcare environments.

We observed an increase in validation loss after the 7th epoch, as shown in Figure 2. To enhance the model's generalization and prevent overfitting, we have implemented various measures, including data augmentation, dynamic adjustment of learning rates, integration of dropout layers and L2 regularization and so on. However, the behavior of machine learning models often remains influenced by inherent characteristics of the training data, such as latent noise and complex nonlinear relationships. These factors may cause the model to quickly adapt to these characteristics in the early stages of training, which could pose challenges as training progresses. Despite extensive efforts to prevent overfitting, the model may still exhibit heightened sensitivity to certain specific features of the training data, especially after prolonged training periods. Looking forward, we plan to further enhance our model's performance by expanding our dataset and exploring the use of updated and more sophisticated model architectures.

In practical clinical applications, deep learning models, including ours, have shown remarkable potential for improving the accuracy and speed of medical diagnoses, as supported by research from Erickson et al. and others (54). Our CNN-based model demonstrates high accuracy in the early detection and diagnosis of lung cancer, offering the potential for continuous improvement through updates and retraining with new data. In addition, challenges such as data privacy, model interpretability, and acceptance within the medical community remain. Therefore, therefore, future work will also need to focus on strengthening data privacy measures and enhancing the interpretability of the model. Addressing these challenges comprehensively will be key to fully integrating advanced diagnostic tools like ours into clinical practice, ultimately improving early lung cancer diagnosis and patient outcomes.

Data availability statement

The datasets presented in this article are not readily available because the data is individual-level data restricted for sharing due to ethical requirements. Requests to access the datasets should be directed to apocalypse1523@163.com.

Ethics statement

The studies involving humans were approved by Ethics review committee of Pingyi county Hospital of Traditional Chinese Medicine (PYX2YYYLLWYh2023030602). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because Data for this study were collected and anonymized by Pingyi County Hospital of Traditional Chinese Medicine. It was retrospectively acquired from all patients who underwent imaging examinations during a specified period. Due to the retrospective nature of the study and the anonymization of the data, obtaining informed consent from specific subjects was not feasible.

Author contributions

HX: Funding acquisition, Validation, Writing – original draft, Writing – review & editing. YY: Methodology, Validation, Writing – original draft, Writing – review & editing. OL: Methodology, Software, Writing – original draft, Writing – review & editing. JC: Data curation, Validation, Writing – review & editing. XH: Data curation, Writing – review & editing. ZT: Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Research on optimizing the informatization and standardization of hospital infection monitoring systems in medical institutions.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1424546/full#supplementary-material.

References

1. Oliver AL. Lung cancer: epidemiology and screening. *Surg Clin North Am*. (2022) 102:335–44. doi: 10.1016/j.suc.2021.12.001

2. Wéber A, Morgan E, Vignat J, Laversanne M, Pizzato M, Rumgay H, et al. Lung cancer mortality in the wake of the changing smoking epidemic: A descriptive study of the global burden in 2020 and 2040. *BMJ Open*. (2023) 13:e065303. doi: 10.1136/bmjopen-2022-065303

3. Lu S, Cheng Y, Huang D, Sun Y, Wu L, Zhou C, et al. Efficacy and safety of selpercatinib in chinese patients with advanced RET fusion-positive non-small-cell lung cancer: A phase II clinical trial (LIBRETTO-321). *Ther Adv Med Oncol*. (2022) 14:17588359221105020. doi: 10.1177/17588359221105020

4. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J Clin*. (2024) 74:229–63. doi: 10.3322/caac.21834

5. Yuan Z, Yu X, Wu S, Wu X, Wang Q, Cheng W, et al. Instability mechanism of osimertinib in plasma and a solving strategy in the pharmacokinetics study. *Front Pharmacol*. (2022) 13:928983. doi: 10.3389/fphar.2022.928983

6. Huang J, Deng Y, Tin MS, Lok V, Ngai CH, Zhang L, et al. Distribution, risk factors, and temporal trends for lung cancer incidence and mortality: A global analysis. *Chest*. (2022) 161:1101–11. doi: 10.1016/j.chest.2021.12.655

7. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. *Nat Rev Clin Oncol*. (2023) 20:624–39. doi: 10.1038/s41571-023-00798-3

8. Wu FZ, Huang YL, Wu CC, Tang EK, Chen CS, Mar GY, et al. Assessment of selection criteria for low-dose lung screening CT among asian ethnic groups in Taiwan: from mass screening to specific risk-based screening for non-smoker lung cancer. *Clin Lung Cancer*. (2016) 17:e45–56. doi: 10.1016/j.cllc.2016.03.004

9. Li C, Wang H, Jiang Y, Fu W, Liu X, Zhong R, et al. Advances in lung cancer screening and early detection. *Cancer Biol Med*. (2022) 19:591–608. doi: 10.20892/j.issn.2095-3941.2021.0690

10. Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: current application and future perspective. *Semin Cancer Biol*. (2023) 89:30–7. doi: 10.1016/j.semcancer.2023.01.006

11. Thong LT, Chou HS, Chew HSJ, Lau Y. Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: A systematic review and meta-analysis. *Lung Cancer*. (2023) 176:4–13. doi: 10.1016/j.lungcan.2022.12.002

12. Feng J, Jiang JJ. Deep learning-based chest CT image features in diagnosis of lung cancer. *Comput Math Methods Med*. (2022) 2022:4153211. doi: 10.1155/2022/4153211

13. Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med*. (2006) 355:1763–71. doi: 10.1056/NEJMoa060476

14. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005

15. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the cheXNeXt algorithm to practicing radiologists. *PloS Med*. (2018) 15:e1002686. doi: 10.1371/journal.pmed.1002686

16. Peduzzi P, Concato J, Kemper E, Kemper E, Holford TR, Feinstein AR, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. (1996) 49:1373–9. doi: 10.1016/S0895-4356(96)00236-3

17. Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci*. (2010) 25:1–21. doi: 10.1214/09-STS313

18. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. (1998) 86:2278–324. doi: 10.1109/5.726791

19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. (2014) 1409:1556. arXiv:1409.1556.

20. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev BioMed Eng*. (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442

21. Guan Q, Wang Y, Ping B, Li D, Du J, Qin Y, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. *J Cancer*. (2019) 10:4876–82. doi: 10.7150/jca.28769

22. Rohini A, Praveen C, Mathivanan SK, Muthukumaran V, Mallik S, Alqahtani MS, et al. Multimodal hybrid convolutional neural network based brain tumor grade classification. *BMC Bioinf*. (2023) 24:382. doi: 10.1186/s12859-023-05518-3

23. pydicom. pydicom documentation. Available online at: https://pydicom.github.io/ (Accessed 2024-07-16).

24. matplotlib. matplotlib documentation. Available online at: https://matplotlib.org/stable/contents.html (Accessed 2024-07-16).

25. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. (1995) Volume 2 (IJCAI'95). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1137–1143.

26. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst*. (2014) 27:3320–8.

27. Goodfellow I, Bengio Y, Courville A. *Deep learning [M]*. MIT Press (2016) Cambridge, Massachusetts, USA. https://www.deeplearningbook.orgv.

28. Prechelt L. Early stopping - but when? *Neural Networks: Tricks Trade*. (1998) 1524:55–69. doi: 10.1007/3-540-49430-8_3

29. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Lett*. (2006) 27:861–74. doi: 10.1016/j.patrec.2005.10.010

30. Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol*. (2011) 2:37–63.

31. Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2016. *J Natl Cancer Center*. (2022) 2:1–9. doi: 10.1016/j.jncc.2022.02.002

32. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA: A Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

33. Yang D, Woodard G, Zhou C, Wang X, Liu Z, Ye , et al. Significance of different response evaluation criteria in predicting progression-free survival of lung cancer with certain imaging characteristics. *Thorac Cancer*. (2016) 7:535–42. doi: 10.1111/1759-7714.12363

34. Wei N, Zheng B, Que W, Zhang J, Liu M. The association between proton pump inhibitor use and systemic anti-tumour therapy on survival outcomes in patients with advanced non-small cell lung cancer: A systematic review and meta-analysis. *Br J Clin Pharmacol*. (2022) 88:3052–63. doi: 10.1111/bcp.15276

35. Saha P, Nyarko RO, Lokare P, Lokare P, Boateng A, Kahwa I, et al. Effect of covid-19 in management of lung cancer disease: A review. *Asian J Pharm Res Dev*. (2022) 10(3):58–64. doi: 10.22270/ajprd.v10i3.1131

36. Kerpel-Fronius A, Tammemägi M, Cavic M, Henschke C, Jiang L, Kazerooni E, et al. Screening for lung cancer in individuals who never smoked: an international association for the study of lung cancer early detection and screening committee report. *J Thorac Oncol*. (2022) 17:56–66. doi: 10.1016/j.jtho.2021.07.031

37. Herath S, Sadeghi Rad H, Radfar P, Ladwa R, Warkiani M, O'Byrne K, et al. The role of circulating biomarkers in lung cancer. *Front Oncol*. (2022) 11:801269. doi: 10.3389/fonc.2021.801269

38. Aberle DR, Adams AM, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. (2011) 365:395–409. doi: 10.1056/NEJMoa1102873

39. Grenier PA, Brun AL, Mellot FJ. The potential role of artificial intelligence in lung cancer screening using low-dose computed tomography. *Diagnostics (Basel)*. (2022) 12:2435. doi: 10.3390/diagnostics12102435

40. Reck M, Dettmer S, Kauczor H-U, Kaaks R, Reinmuth N, Vogel-Claussen J, et al. Lung cancer screening with low-dose computed tomography: current status in Germany. *Dtsch Arztebl Int*. (2023) 120(23):387. doi: 10.3238/arztebl.m2023.0099

41. Nekolla EA, Brix G, Griebel J. Lung cancer screening with low-dose CT: radiation risk and benefit-risk assessment for different screening scenarios. *Diagnostics (Basel)*. (2022) 12:364. doi: 10.3390/diagnostics12020364

42. Donovan K, Tustison NJ, Linn KA, Shinohara RT. Multivariate residualization in medical imaging analysis. . *bioRxiv: Preprint Server Biol*. (2023). doi: 10.1101/2023.02.15.528657

43. Reznikov N, Buss DJ, Provencher B, McKee MD, Piché N. Deep learning for 3D imaging and image analysis in biomineralization research. *J Struct Biol*. (2020) 212:107598. doi: 10.1016/j.jsb.2020.107598

44. Hyun Y, Kim D. Development of deep-learning-based single-molecule localization image analysis. *Int J Mol Sci*. (2022) 23:6896. doi: 10.3390/ijms23136896

45. Zhang H, Ma R, Zhao Y, Zhang Q, Sun Q, Ma Y, et al. Optimized convolutional neural network recognition for athletes' Pneumonia image based on attention mechanism. *Entropy (Basel Switzerland)*. (2022) 24:1434. doi: 10.3390/e24101434

46. Efimenko M, Ignatev A, Koshechkin K. Review of medical image recognition technologies to detect melanomas using neural networks. *BMC Bioinf*. (2020) 21:270. doi: 10.1186/s12859-020-03615-1

47. Miotto R, Wang F, Wang S, Jiang X, Dudley JT, et al. Deep learning for healthcare: review, opportunities and challenges. *Briefings Bioinf*. (2017) 19:1236–46. doi: 10.1093/bib/bbx044

48. Lu MT, Raghu VK, Mayrhofer T, Aerts HJWL, Hoffmann U. Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening computed tomography: development and validation of a prediction model. *Ann Intern Med*. (2020) 173:704–13. doi: 10.7326/M20-1868

49. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x

50. Cellina M, Cè M, Irmici G, Ascenti V, Khenkina N, TotoBrocchi M, et al. Artificial intelligence in lung cancer imaging: unfolding the future. *Diagnostics (Basel)*. (2022) 12:2644. doi: 10.3390/diagnostics12112644

51. Hsu YC, Tsai YH, Weng HH, Hsu LS, Tsai YH, Lin YC, et al. Artificial neural networks improve LDCT lung cancer screening: a comparative validation study. *BMC Cancer*. (2020) 20:1023. doi: 10.1186/s12885-020-07465-1

52. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. (2009) 338:b605. doi: 10.1136/bmj.b605

53. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol*. (2022) 22:316. doi: 10.1186/s12874-022-01801-8

54. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. (2017) 37:505–15. doi: 10.1148/rg.2017160130

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 29 August 2024
DOI 10.3389/fonc.2024.1362850

Check for updates

OPEN ACCESS

EDITED BY
Harrison Kim,
University of Alabama at Birmingham, United States

REVIEWED BY
Shanglong Liu,
Affiliated Hospital of Qingdao University, China
Jakub Nalepa,
Silesian University of Technology, Poland
Lukas Vrba,
University of Arizona, United States
Man Lu,
Sichuan Cancer Hospital, China

*CORRESPONDENCE
Anurag S. Rathore
asrathore@biotechcmz.com

RECEIVED 29 December 2023
ACCEPTED 01 August 2024
PUBLISHED 29 August 2024

CITATION
Mandal S, Balraj K, Kodamana H, Arora C, Clark JM, Kwon DS and Rathore AS (2024) Weakly supervised large-scale pancreatic cancer detection using multi-instance learning.
Front. Oncol. 14:1362850.
doi: 10.3389/fonc.2024.1362850

COPYRIGHT
© 2024 Mandal, Balraj, Kodamana, Arora, Clark, Kwon and Rathore. 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.

Weakly supervised large-scale pancreatic cancer detection using multi-instance learning

Shyamapada Mandal[1], Keerthiveena Balraj[2], Hariprasad Kodamana[1,2], Chetan Arora[2,3], Julie M. Clark[4], David S. Kwon[4,5] and Anurag S. Rathore[1,2]*

[1]Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi, India, [2]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi, India, [3]Department of Computer Science and Engineering, Indian Institute of Technology Delhi, New Delhi, India, [4]Henry Ford Pancreatic Cancer Center, Henry Ford Health, Detroit, MI, United States, [5]Department of Surgery, Henry Ford Health, Detroit, MI, United States

Introduction: Early detection of pancreatic cancer continues to be a challenge due to the difficulty in accurately identifying specific signs or symptoms that might correlate with the onset of pancreatic cancer. Unlike breast or colon or prostate cancer where screening tests are often useful in identifying cancerous development, there are no tests to diagnose pancreatic cancers. As a result, most pancreatic cancers are diagnosed at an advanced stage, where treatment options, whether systemic therapy, radiation, or surgical interventions, offer limited efficacy.

Methods: A two-stage weakly supervised deep learning-based model has been proposed to identify pancreatic tumors using computed tomography (CT) images from Henry Ford Health (HFH) and publicly available Memorial Sloan Kettering Cancer Center (MSKCC) data sets. In the first stage, the nnU-Net supervised segmentation model was used to crop an area in the location of the pancreas, which was trained on the MSKCC repository of 281 patient image sets with established pancreatic tumors. In the second stage, a multi-instance learning-based weakly supervised classification model was applied on the cropped pancreas region to segregate pancreatic tumors from normal appearing pancreas. The model was trained, tested, and validated on images obtained from an HFH repository with 463 cases and 2,882 controls.

Results: The proposed deep learning model, the two-stage architecture, offers an accuracy of 0.907 $\pm$ 0.01, sensitivity of 0.905 $\pm$ 0.01, specificity of 0.908 $\pm$ 0.02, and AUC (ROC) 0.903 $\pm$ 0.01. The two-stage framework can automatically differentiate pancreatic tumor from non-tumor pancreas with improved accuracy on the HFH dataset.

Discussion: The proposed two-stage deep learning architecture shows significantly enhanced performance for predicting the presence of a tumor in the pancreas using CT images compared with other reported studies in the literature.

KEYWORDS

pancreatic cancer, multi-instance learning, image segmentation, feature extraction, medical image analysis

1 Introduction

Pancreatic adenocarcinoma is currently one of the deadliest cancers, with an overall 5-year survival rate of approximately 11% (1). Signs and symptoms of pancreatic cancer are non-specific and thus have limited utility in early detection. Moreover, efficient screening tests for the early detection of pancreatic tumors do not currently exist. It is clear that the pancreatic cancer survival rate is likely to significantly improve if the cancer can be detected at an early stage where definitive treatment with surgery and systemic therapy can be offered (2). Computed tomography (CT) and magnetic resonance imaging (MRI) are two common screening modalities that can be better utilized for diagnosing pancreatic cancer. Recent advancements in artificial intelligence and radiographic imaging provide hope that there may be the opportunity to use the aforementioned modalities as early screening detection tests, especially for early pancreatic cancer detection (3–5). Automated medical image segmentation and classification have been extensively investigated in the image analysis community due to the fact that manual, dense labeling of large amounts of medical images is tedious and error-prone. Accurate and reliable solutions are desired to increase clinical workflow efficiency and support decision-making through fast and automatic extraction of quantitative measurements (6, 7).

In the field of biomedical image analysis, the analysis of pancreatic images has significant importance in clinical diagnosis and research, including a range of tasks: (1) segmentation of tumor region, (2) diagnosing the presence of cancer, and (3) clustering the region. This research primarily focuses on an integrated framework that has the potential to perform classification, segmentation, and clustering. Traditional fully supervised techniques require accurately annotated data, which is laborious, uncertain, and time-consuming. However, unsupervised methods extract features from unlabeled data and have limited application to high-level tasks such as image segmentation and classification. In this scenario, the proposed algorithm strikes a balance by leveraging the benefits of both supervised and unsupervised approaches. This paper addresses the effectiveness and efficiency of accomplishing high-level tasks with a minimum of manual annotation and to automatically extract fine-grained information from coarse-grained labels.

Convolutional neural networks (CNNs) (8) have made incredible strides in the medical imaging industry over the past 10 years, particularly in the areas of CT, MRI, and ultrasound image analysis. Since their advent, near-radiologist level performance has been achieved in automated medical image analysis tasks, including detection or prediction of hypertrophic cardiomyopathy (9, 10), future cardiovascular event (11), cancerous lung nodules (12, 13), liver tumors (14), and hepatocellular carcinoma (15). Figure 1 demonstrates the difficulty in traditionally segmenting the pancreas compared with other organs due to various factors such as relatively small size, complicated anatomical structure due to adjacent structures, and uncertain boundaries of the organ in the limited slices in which the pancreas appears on traditional CTs.

High representation power and fast inference properties have made CNNs the *de-facto* standard for image segmentation and classification. Fully convolutional networks (FCNs) (12, 16) in general and U-Net (6) in particular are some of the commonly used architectures for automated medical image segmentation. The architectures are typically cascaded with multistage CNN models when the target organs show large inter-patient variation in terms of shape and size (17). Although deep learning has been studied for the detection of pancreatic cancer (18–24), pancreatic neuroendocrine tumors (pNETs) (25), and intraductal papillary mucinous neoplasms (IPMNs) in pancreas (26), it is yet to be incorporated as a part of the routine workup for patients diagnosed with pancreatic cancers. The U-Net model and its extended versions have been used in the literature for organ segmentation and have been demonstrated to deliver good accuracy (27–29).

Pancreatic cancer detection from CT images by applying deep learning is a challenging task because the pancreas is a small organ and is located in a complicated position in the retroperitoneum [Nakao et al., (30)]. A typical CT scan of a patient contains 131 slices in the full axial view, with approximately 20 to 60 slices containing the image of the pancreas. In early stages, the pancreatic tumors are too small and irregularly shaped for easy identification. Following that, several researchers proposed cutting-edge CNN techniques based on segmentation of the pancreas using either cascaded or coarse-to-fine segmentation networks. However, prior investigations of pancreatic segmentation were conducted on extremely limited populations (4, 18, 31, 32) and the results have been unsatisfactory (maximum dice coefficient 0.58) (33). This is primarily due to the minor differences between a singular image that contains the tumor vs. an image that does not. To the best of our knowledge, there are very few DL studies that have been conducted on big CT datasets that encompass a variety of pancreatic volumes. Figure 2 illustrates the private and publicly available pancreatic dataset with normal and abnormal patients used for training the proposed framework. Therefore, the objective of this investigation is to carry out an effectiveness study using a weakly supervised algorithm with a total of 463 patients suffering from pancreatic tumors and a total of 2,882 controls.

Furthermore, as the tumor size is significantly small compared with the overall size of the CT image, the task of identification is further complicated when all slices are analyzed together. To circumvent these issues, we propose a weakly supervised two-stage architecture with a cascade of segmentation and classification. As we do not have annotated mask of the pancreas in our local dataset (Henry Ford Health (HFH)), the segmentation model was trained using MSKCC data to segment the pancreas. It should be noted that the Memorial Sloan Kettering Cancer Center (MSKCC) data set is small and does not capture all possible clinical variations of tumors. On the other hand, HFH is a much larger dataset but contains only patient-label information (cancer versus non-cancer). Our proposed approach intelligently combines "the best of both the worlds" by using a weakly supervised classification that utilizes patient label information of the HFH data and applies it to the cropped pancreas images, after initial processing by the segmentation model trained on the MSKCC data.

The goal of semantic segmentation (34) is to identify common features in an input image by learning and then to label each pixel in an image with a class. In this technique, raw image data are converted to quantitative, spatially structured information and

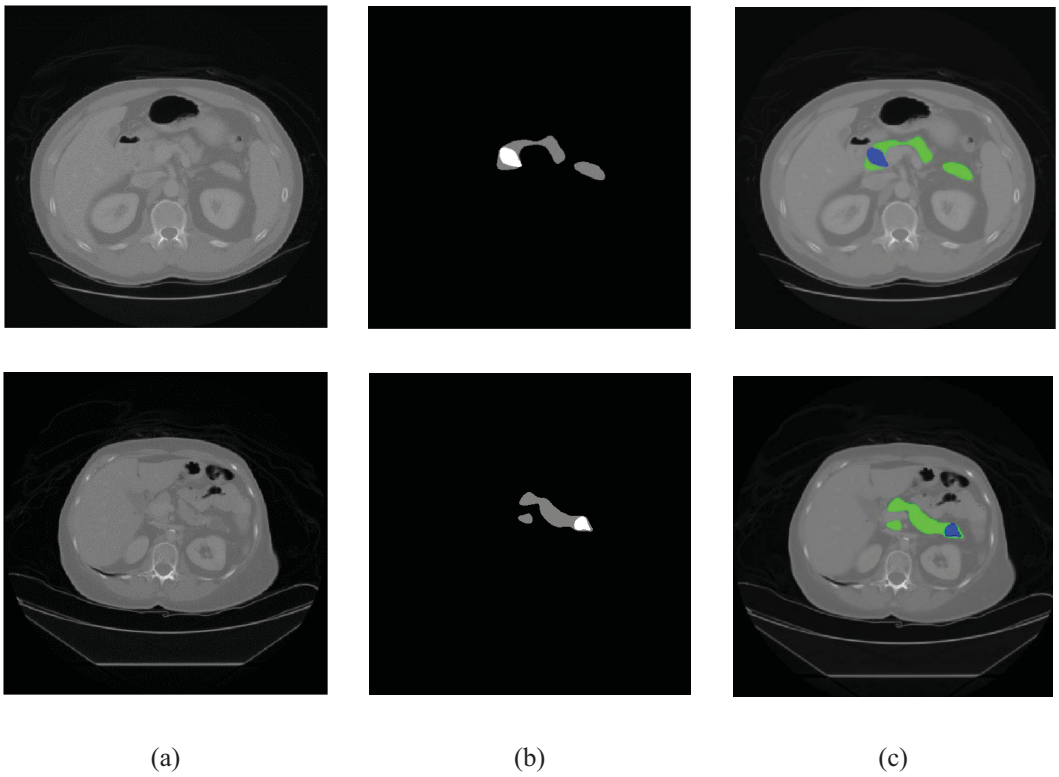


FIGURE 1
Sample CT images. **(A)** Original image, **(B)** ground truth, and **(C)** color-coded depiction of overlay on an axial CT image (pancreas: green, cancer: blue).

can then be used for further processing. The segmentation method is an essential component in finding features in several clinical applications, such as applications of artificial intelligence in a diagnostic support system, tumor growth monitoring, therapy planning, and intraoperative assistance.

Based on FCN models, researchers have proposed a variety of strategies recently, including Hierarchical 3D FCN (35), DeepLab (36), SegNet (37), PSPNet (38), and RefineNet (39). The majority of these techniques automatically fit into the category of fully supervised learning methods, hence requiring a sufficient number

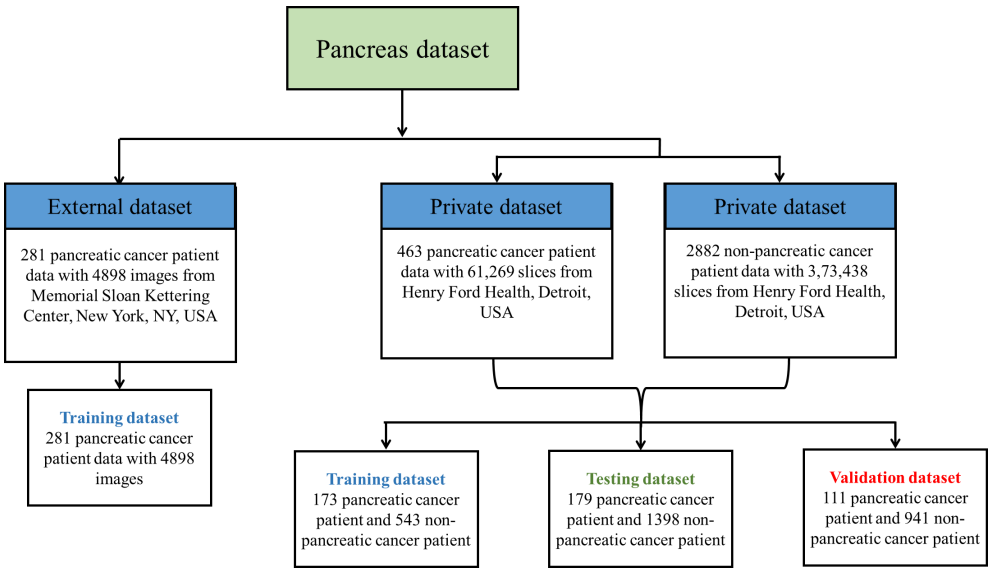


FIGURE 2
Private and external pancreas datasets used for model training.

of annotated data to train. Overall, the existing state-of-the-art methods for performing segmentation and classification either involve careful generation of handcrafted features or heavily rely on the extensive delineation of pancreatic tumor areas to give the annotation masks, which really place a tremendous load on the oncologists and researchers, respectively. Fully supervised methods have achieved remarkable performance in segmentation tasks such as brain tumor segmentation, lesion segmentation, and multiorgan segmentation (40). However, when a fully supervised algorithm is applied to pancreatic cancer detection and segmentation, these models have not achieved satisfactory results (3).

When employing advanced machine learning to the diagnosis of pancreatic tumors, the following major challenges arise: (1) Over 70% of pancreatic tumors have irregular shapes and ambiguous margins, resulting in imperceptible boundaries with the surrounding tissues. This characteristic increases the complexity of the segmentation process and may result in oversights when segmenting tumors. (2) The pancreas region is surrounded by many organs and tissues, and cancers affect a small area of the organ. Due to this, training a CNN architecture becomes difficult and the model gets distracted by irrelevant regions of the image, potentially leading to misclassification. (3) Training a deep learning model requires a substantial quantity of precisely annotated images for training. However, owing to the anatomical intricacy of the organ and differences in tumor appearance, physically identifying the pancreas and tumor locations is a labor-intensive and time-consuming task.

To overcome these issues, we herein propose a new weakly supervised algorithm that has a two-stage architecture, namely, segmentation and classification. In the first stage, the pancreas is segmented by the supervised segmentation model, and in the second stage, the multi-instance learning-based weakly supervised classification method is applied to the cropped pancreas images, which were obtained from segmentation, to classify pancreatic tumor images and normal-appearing pancreas, as schematically illustrated in Figure 3. There are three major impactful contributions from this work:

1. An end-to-end model for high-accurate pancreas segmentation along with classification has been proposed. The segmentation model is built upon an nn-Unet architecture, which segregates the pancreas, pancreatic cancers, and residual background organ and intraperitoneal space.
2. Furthermore, end-to-end multiple instance learning has been performed by multiple-instance neural networks, which accept a bag containing different numbers of instances as input and output the bag label right away.
3. Finally, comprehensive experiments on the unannotated HFH dataset have been conducted to demonstrate that the proposed approach outperforms other state-of-the-art techniques. To validate the effectiveness of the overall approach, the proposed architecture has been tested on a large volume of data obtained from the local data repository at Henry Ford Health (HFH) in Detroit, Michigan.

In this paper, Section 2 highlights the dataset used, network architectures, its components, and methodology. Sections 3 and 4

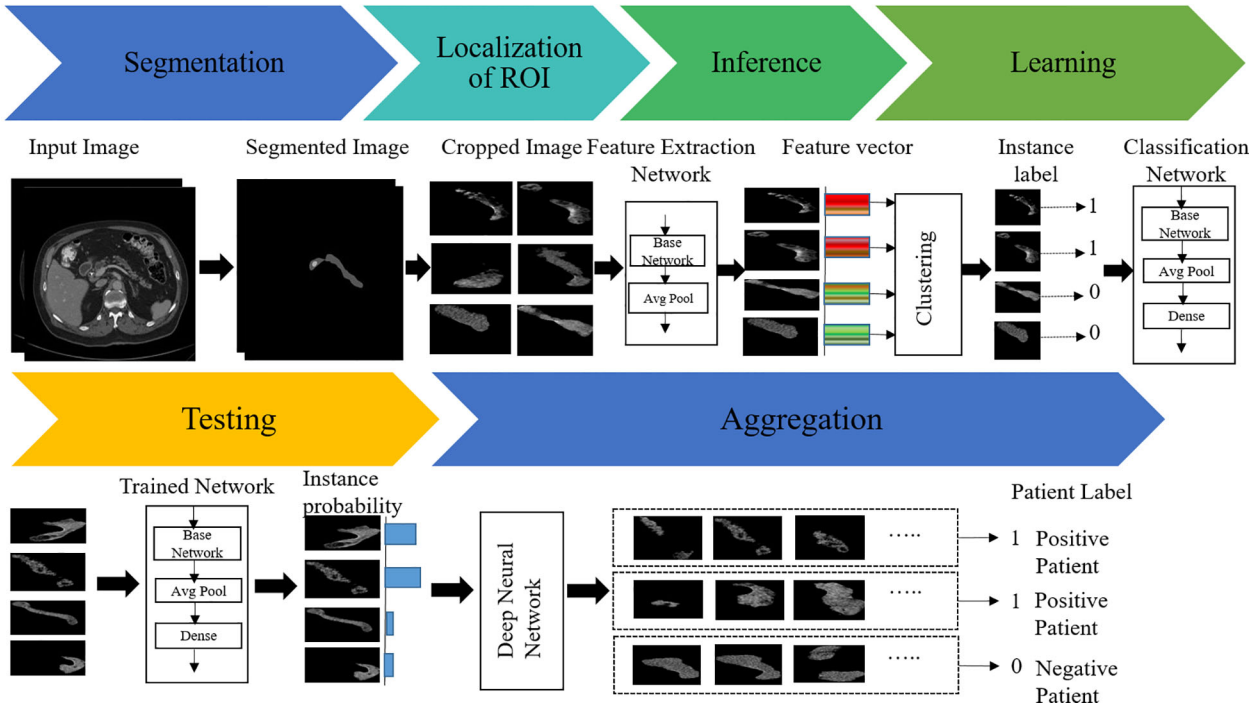


FIGURE 3
Illustration of the proposed two-stage architecture.

present the results and discussion in terms of ablation and comparison study. Finally, the conclusions of the study are presented in Section 5.

2 Materials and methods

CT images from HFH and Memorial Sloan Kettering Cancer Centre (MSKCC) were used to develop an end-to-end algorithm to detect the presence of pancreatic tumors in case versus control images. MSKCC is an open-source dataset that was used to train the segmentation model. Following this, we focused on the larger HFH dataset to further fine-tune and validate the model. A method has been proposed utilizing the axial view CT images and segmentation and classification techniques.

2.1 Dataset

In the proposed algorithm, the publicly available MSKCC dataset has been used to develop the segmentation model, followed by a separate and more robust HFH dataset used to train, test, and validate the MIL classification model. The MSKCC dataset comprised patients undergoing resection of pancreatic masses (31). These data consist of portal venous phase CT scans of 281 patients. Each patient has a single nifty (.nii) file that contains a full series of axial view images at the volumetric level. The classification model train, test, and validation data included CT images of 3,453 adult patients from HFH. Cases were images from patients diagnosed with pancreatic ductal adenocarcinoma, and controls were those where there was no suspicion of pancreas disease. Patients who had pancreatitis and women who were pregnant were excluded from the dataset.

2.2 Data preprocessing

In this study, retrospective imaging data were collected at Henry Ford Health (HFH), Detroit, Michigan, United States, from 2013 to 2020. Each patient on the HFH dataset had axial, sagittal, and coronal view images. In this investigation, the axial view is preferred because it provides a higher resolution and more detailed cross-sectional images of the pancreas. In addition, the axial perspective is consistent with the conventional clinical methods, which guarantees that our analysis and findings are reliable and consistent every time. For patients with known pancreatic tumors, image acquisition was performed with a pancreas protocol, high-resolution imaging cut at 2.5 mm with dedicated arterial and portal venous phases to identify vascular abnormalities, characteristics of hypoattenuating tumors, and to recognize hepatic metastases. The axial raw images were in DICOM format with several images (60 to 350) per study.

A Python function was used to convert a 2D slice-level multiple images to a single 3D volumetric level image in nifty (.nii) format. To account for scanner and acquisition variability, a third-order spline interpolation was used for image data and nearest-neighbor interpolation for the corresponding segmentation mask to convert heterogeneous spacing to homogeneous (3, 41). During training and inference, each image was normalized using a global normalization scheme. While preprocessing HFH data, a total of 108 patients' CT images were excluded due to incomplete axial view series (38), inconsistent pixel spacing (33), inconsistent image orientation (13), and segmentation error (29) due to poor quality of the image. This is represented in the flowchart (Figure 4). After this preprocessing step, the train, test and validation dataset sizes were 463 patients with pancreatic tumors and 2,882 controls.

2.3 Model architectures and training

In this study, the objectives of the segmentation method were to identify the slices containing the pancreas from the full-volume image and thereby identify the location of the pancreas in each slice. The segmentation method was applied to the volumetric axial view images, and it produced a segmented portion of the pancreas. In the first stage, the pancreas was segmented by the nnU-Net model (3). The segmented pancreas image was then used to find the location of the pancreas in the input image. The supervised classification model provides satisfactory results if the training data set captures a wide variation of the clinical samples. Therefore, we applied a weakly supervised classification model known as multi-instance learning (MIL). The MIL-based classification model was applied to distinguish pancreatic tumor and non-tumor pancreas on the cropped pancreas images. MIL is a weakly supervised classification wherein a label is only assigned to a collection of observations or a bag of instances (42–45).

In this approach, a three-dimensional image is converted into a bag of two-dimensional slices. Initially, each image is divided into patches as instance, with bags representing a subset of these patches. The instances are assigned with a label based on the presence or absence of the cancer region. If at least one instance within the bag consists of the tumor region, the bag is labeled as positive and none of the instances in the negative bag are positive. One of the challenges in applying MIL is finding the positive instances and negative instances from the positive bag. If all instances within the bad are classified as non-tumors, then the bag is labeled as negative. For the given coarse-grained (bag) label, MIL aims to predict the fine-grained (instance) labels for each patch within the image.

To guarantee that each positive bag consists of more than one positive-labeled instance, the K-means clustering algorithm is employed to partition the data into distinct groups based on the features extracted from the images. The architecture of the proposed MIL approach is given in Figure 5. The feature of the instances was extracted by the average pool layer, where the base network is taken as ResNet50, a popular CNN framework used for image processing (46–48). This ensures that each positive bag consists of multiple instances representing tumors regions within the image. The following are the steps involved to obtain the predicted instance probability:

Step 1: We label the instances in a positive bag by clustering the image features into two groups: one is a positive group that contains all positive instances, and another is a negative group that contains

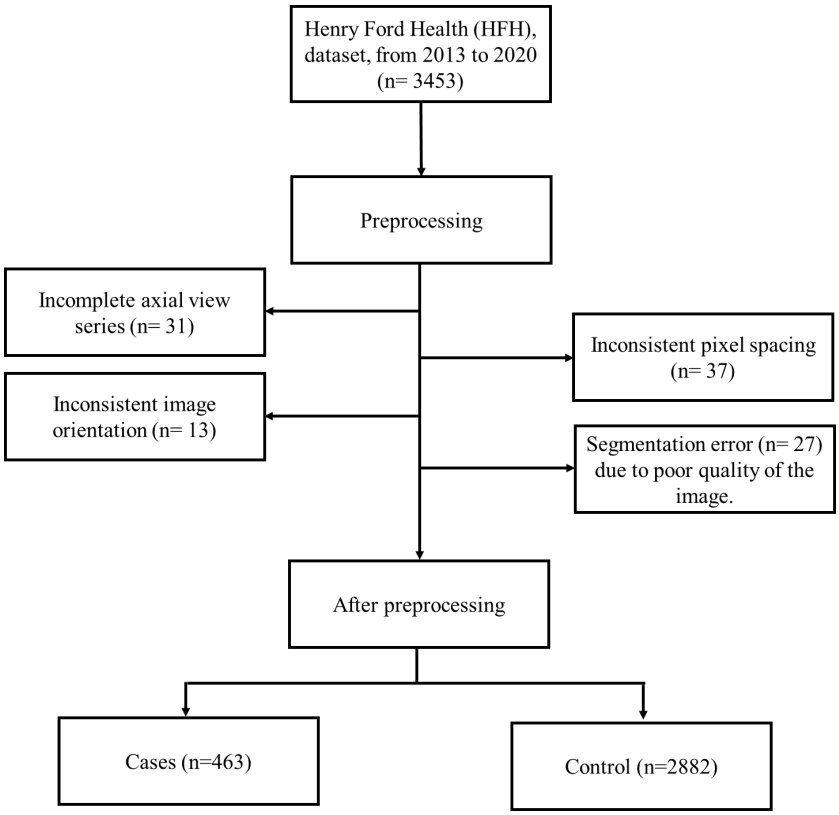


FIGURE 4
Patients' enrollment and exclusion process in the HFH database.

negative instances. Given a dataset, $\{X^{(i)}, Y^{(i)}\}_{i=1}^{N}$ containing N volumetric images, $X^{(i)} \in \mathbb{R}^{h_i \times w_i \times z_i}$, and $Y^{(i)} \in \{0,1\}$ is the patient label, where 1 indicates positive patient and 0 indicates a normal patient. Each volumetric image $X^{(i)}$ is a bag of instances $\left\{x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\right\} \in X^{(i)}$.

Step 2: For the negative bag, all instances are negative. In the MIL pipeline, the features of the images are captured by ResNet50 and classified by ResNet50 (46).

Step 3: After extracting features of the images in ResNet50, K-means clustering was employed to get the label of the instance. In the feature extraction network, instance features are extracted at average pooling layers and returned as $\left\{\left(x_1^{(i)}, f_1^{(i)}\right), \left(x_2^{(i)}, f_2^{(i)}\right), \ldots, \left(x_n^{(i)}, f_n^{(i)}\right)\right\}$, where $f_1^{(i)}$ is the feature of instance $x_1^{(i)}$.

Step 4: The instance labels are assigned by clustering method as $\left\{\left(x_1^{(i)}, y_1^{(i)}\right), \left(y_2^{(i)}, f_2^{(i)}\right), \ldots, \left(x_n^{(i)}, y_n^{(i)}\right)\right\}$, where $y_j \in \{0,1\}$.

Step 5: The label instances are fed into the classification model that produces the predicted instance probability.

In this case, the ResNet50 model was used as a classification model. ResNet50 is a CNN-based classification method available in the Keras environment with pretrained weights in the TensorFlow backend. The original model was trained on the ImageNet dataset and was slightly modified in the proposed method. The fully connected output layers that were used for the prediction in the original model were not used. Instead, an average pooling layer with a pool size of (4 × 4) was added, followed by a dense layer with the number of neurons as 256. As the problem is that of binary classification, the output layer is a dense layer with dimension 2, and the softmax activation function is used for this layer. A dropout rate of 0.3 was applied in between the output layer and its previous layer. The final layer of the network, called the probability layer, calculates the probability of the input (cropped) image being of that class. Furthermore, a multilayer perceptron neural network (NN) was applied to aggregate the instance probability to patient probability and schematically illustrated in Figure 5. The NN structure is optimized, and the best-performing network is found to have one hidden with the number of neurons 18. The MIL and NN were trained by 173 cases and 543 control patients of the HFH dataset.

3 Results

To validate the effectiveness of the proposed architecture, a series of ablation studies with different baseline models were conducted in this section. The model was trained, tested, and validated on images obtained from an HFH repository with 463 cases and 2,882 controls.

3.1 Quantitative evaluation

We created a two-stage weakly supervised deep learning-based model to identify pancreatic tumors using CT images from publicly

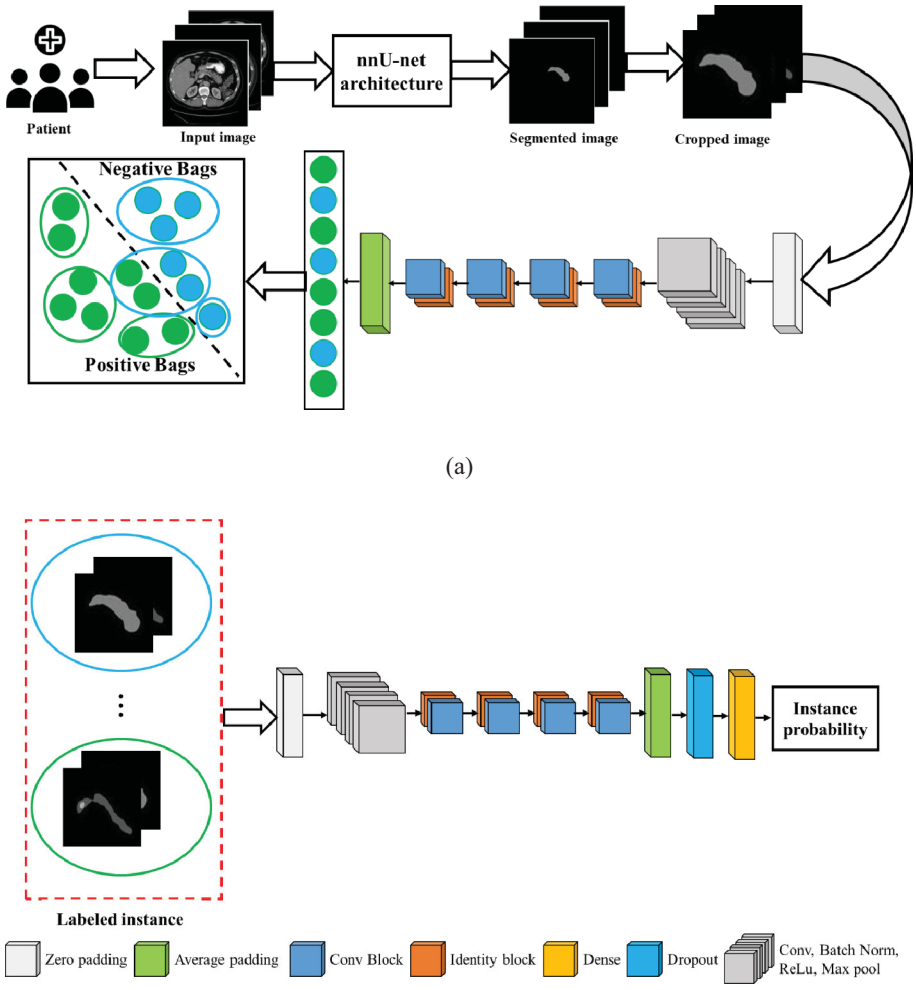


FIGURE 5
Architecture of the proposed model: **(A)** multi-instance learning, and **(B)** classification model.

available Memorial Sloan Kettering Cancer Center (MSKCC) and Henry Ford Health (HFH) data sets. In the first stage, the nnU-Net supervised segmentation model was used to crop an area in the location of the pancreas, which was trained on the MSKCC repository of 281 patient image sets with established pancreatic tumors. In the second stage, a multi-instance learning-based weakly supervised classification model was applied on the cropped pancreas region to segregate pancreatic tumors from normal-appearing pancreas. The performance of the proposed two-stage architecture was then compared with the existing models that have been recently published.

The performance of the proposed method was tested on the HFH dataset comprising 179 patients with known pancreatic tumors (23715 slices) as well as 1,398 patients without pancreatic cancer (182,757 slices), whereas the model parameters were fixed by the validation dataset including 111 patients with known pancreatic tumors (14,612 slices) as well as 941 patients without pancreatic cancer (122,214 slices), both selected randomly, yielding an average number of slices per patient of 131. A training dataset of HFH patients was created with 173 randomly selected patients with pancreatic cancer (22,942 slices) and 543 patients without pancreatic cancer (68,467 slices). The training dataset was used to train the MIL classification and NN aggregation models. The segmentation model was fed with the input CT image. The train and test data of HFH were segmented by the trained nnU-Net model. The nnU-Net model was trained by 281 patients of MSKCC dataset. Figure 6 depicts the segmentation results of samples with both normal and abnormal by using the nnU-Net model applied on the HFH dataset. The Supplementary Material (Supplementary Figure S1) contains a box plot of the dice score for pancreas and pancreatic cancer regions.

The cropped pancreas images from the HFH train dataset were used to train the MIL classification model. In this case, only the list of positive patients and negative patients is known. In the MIL

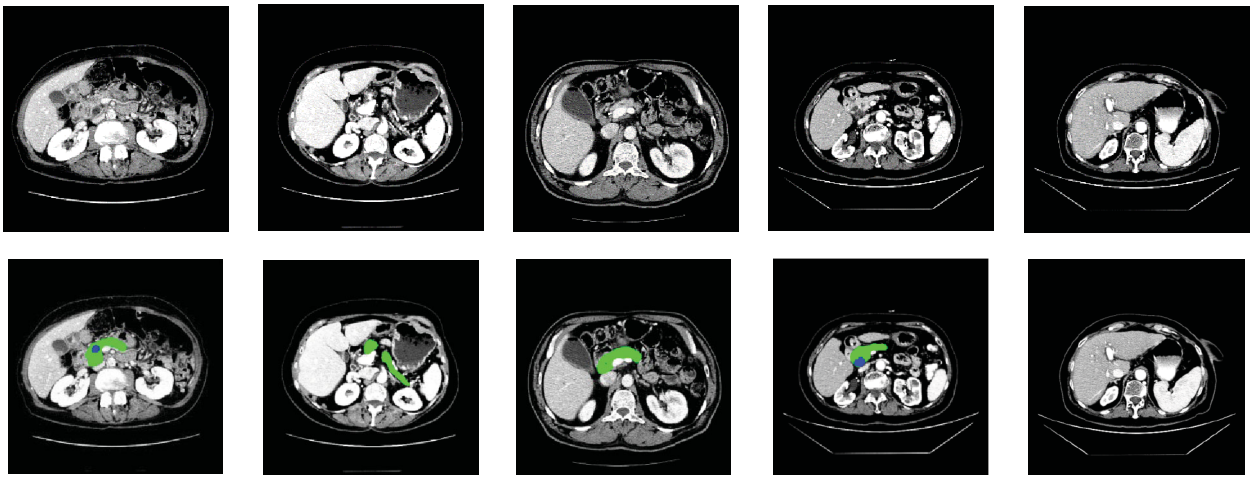


FIGURE 6
Segmentation results on HFH data: the first row represents the raw input image, and the second row represents the results of segmentation (pancreas: green, cancer: blue).

approach, each patient is considered as a bag, a cancerous patient is called a positive bag, and a control patient is called a negative bag. All instances in the negative bag are negative, and we directly fed the instances with label 0 to the classification model without clustering. In the positive bag, positive and negative instances are mixed. The instances in the positive bag are fed to the feature extraction model to extract the feature. The clustering method segregates the positive and negative instances by utilizing their features. The positive instances are labelled as 1 and fed to the classification model. The classification model was trained by the labeled instances from the training dataset. For test data, instance probability for the risk of cancer was predicted by the trained classification model. In the aggregation model, the neural network produces patients' cancer probability by combining all instances of the probability of the patients. The neural network model parameters were fixed by the validation dataset. The results of validation dataset were as follows: sensitivity 0.847 ± 0.015, specificity 0.880 ± 0.025, 0.876 ± 0.021, and AUC 0.863 ± 0.01.

3.2 Ablation study of the proposed model

The proposed pancreatic tumor detection method was then compared with the existing nnU-Net-based segmentation method shown in Table 1. As the nnU-Net is a segmentation method for pancreas and pancreatic tumor segmentation (3), its efficacy toward identification of pancreatic cancer in the HFH dataset was characterized by measuring sensitivity, specificity, accuracy, and area under curve (AUC) based on the tumor segmentation results. The corresponding 95% confidence intervals (CI) were obtained using the Delong technique (49). With a 95% CI of (0.88, 0.92), the sensitivity was 0.905, demonstrating a high level of reliability in detecting true positives. The proposed framework's ability to detect true negatives is reflected by specificity, which is 0.908 (95% CI: 0.86, 0.94). The model demonstrated outstanding performance across different assessment measures, with an overall accuracy of 0.907 (95% CI: 0.887, 0.927) and an AUC of 0.903 (95% CI: 0.883, 0.923).

As is evident, nnU-Net + MIL with the NN aggregation model (the proposed method) yielded the best performance with sensitivity 0.905 ± 0.01, specificity 0.908 ± 0.02 accuracy 0.907 ± 0.01, and AUC(ROC) 0.903 ± 0.01. In comparison, the nnU-Net + MIL and nnU-Net alone significantly underperformed, as shown in Table 1. The AUC(ROC) for the three methods are illustrated in Figure 7. As the nnU-Net + MIL + NN method provides each patient label as probability, the ROC is a smooth curve unlike other two methods which involve detection by setting threshold yielding binary patient label (0 or 1).

Several classification methods such as ResNet50, Xception, VGG16, and InceptionV3 were implemented in the second-phase test dataset on the cropped pancreas image. The pancreatic tumor detection results of these methods with the proposed architecture are depicted in Figure 8. It is evident that the proposed algorithm yields better performance in terms of accuracy and sensitivity. We have compared the proposed (nnU-Net + MIL with the NN aggregation model) architecture with various models such as nnU-Net. As can be deduced from Figure 8, the InceptionV3 classification model yielded the best performance (accuracy 0.83, sensitivity 0.79, and specificity 0.84) compared with the nnU-Net (0.79, sensitivity 0.78, specificity 0.8), nnU-Net + ResNet50 (accuracy 0.82, sensitivity 0.78, specificity 0.82), nnU-Net + Xception (accuracy 0.82, sensitivity 0.78, specificity, 0.81), and nnU-Net + VGG16 (accuracy 0.82, sensitivity 0.78, specificity, 0.82) models.

P-values are computed for both proposed and state-of-the-art techniques and are represented in Supplementary Figure S2.

TABLE 1 Comparison of the proposed model with different classifications on the HFH test dataset.

Model	Sensitivity	Specificity	Accuracy	AUC
nnU-Net	0.780 ± 0.03	0.801 ± 0.04	0.790 ± 0.03	0.791 ± 0.02
nnU-Net + MIL	0.831 ± 0.02	0.917 ± 0.01	0.908 ± 0.01	0.874 ± 0.02
nnU-Net + MIL + NN (Proposed approach)	0.905 ± 0.01	0.908 ± 0.02	0.907 ± 0.01	0.903 ± 0.01

The above mentioned results are tested on the HFH test dataset, which includes 179 cases (23,715 slices) and 1,398 control (182,757 slices). Data are sensitivity (95% CI).

4 Discussion

We have developed and validated an image analysis model that can identify pancreatic tumors on CT images by a combination of segmentation and classification methods, both based on deep learning techniques. The two-stage framework was developed utilizing a previously reported pancreas dataset from MSKCC and validated on the HFH dataset of control patients and pancreatic tumor patients. It has been demonstrated that the proposed method results in satisfactory performance. As is evident from Table 1, the proposed method yields superior performance when compared with other existing models and offers an accuracy of 0.907. Equally importantly, the performance of the proposed two-stage approach is demonstrated to be stable in a large dataset. The robust performance of the proposed method is an outcome of the verification and optimization that are performed at each layer. The information output from the pancreas is high in the segmented region. Hence, in the proposed method, the MIL classification method was applied only on the cropped pancreas image. As a result, the chances of identifying false positives are minimized. Overall, our results indicate that two-stage image analysis can distinguish between presence and absence of pancreatic tumors on CT images when given blinded images.

The nnU-Net used in this study has been applied toward medical segmentation decathlon (MSD) competition for multiple-organ segmentation and tumor detection tasks such as liver, spleen, kidney, pancreas, gallbladder, colon, and prostate (3). However, the detection efficacy of the pancreatic tumor by employing nnU-Net is limited, exemplified by a low dice coefficient of 0.53. In our study, we identified that a single-stage approach using nnU-Net produces more false positives, primarily because the contour of the pancreatic tumor is irregular and there are ill-defined margins on the CT image, leading to false detection of the normal pancreas portion as cancer. The CNN patch-based classification method was also attempted in literature for pancreatic cancers detection (18) but suffered from the following lacunae. First, CNN was trained to classify pancreas patches and pancreatic cancer patches. The patches were generated by the sliding windows method in a region of interest determined by the presence of pancreas fed in as the input. While this may be feasible in training, for a test image, such masks are unlikely to be available. Secondly, in this patch-based analysis, a patch is labeled positive, even if only a single pixel is predicted as cancerous. Apart from the above method, whole slide image (WSI) classification has been studied for pancreatic cancer detection (50, 51). Since the pancreas is a small organ, the tumor size is also very small in the early stages and thus a classification model alone is unlikely to be sufficient for locating the features of the small tumor portion with respect to the whole slide. This study also had limitations. Manual labeling of pancreatic images are labor-intensive, so we used publicly available datasets and private datasets for training and validation; the testing dataset included only American participants from a single institution. In response to this limitation, we used the MIL technique to balance both supervised and unsupervised approaches and verified the generalizability of the model. The findings demonstrate the effectiveness of a proposed two-stage weakly supervised deep learning system for detecting pancreatic cancer. By employing the proposed prediction model to aid in the radiographic diagnosis of tumors, therapeutic intervention may be accelerated, leading to better clinical results for the patient.

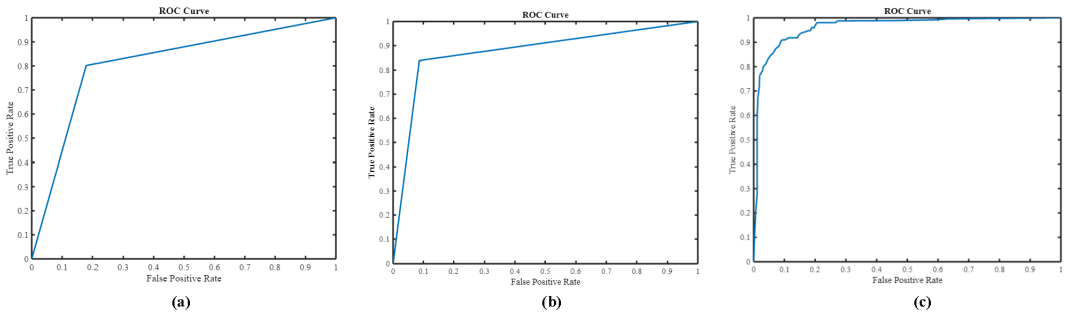


FIGURE 7
ROC curve on HFH test data: **(A)** nnU-Net model, **(B)** nnU-Net with MIL, **(C)** nnU-Net +MIL+NN.

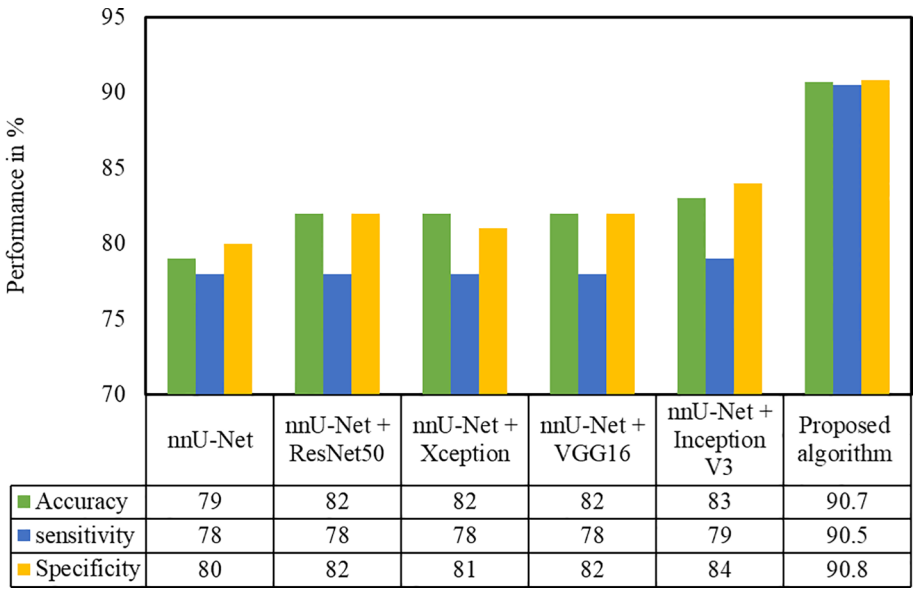


FIGURE 8
Performance evaluation of the proposed approach with different classification techniques on the HFH test dataset.

TABLE 2 Summarization of the existing techniques for pancreatic cancer detection.

Author	Dataset	Segmentation	Classification	Algorithm	Performance measure
Zhang et al., 2021 (52)	FLARE 2021	✓	✗	Efficient context aware network	DSC: 75.3
					NSD: 60.5
Wang et al., 2020 (53)	FLARE	✓	✗	Enhancement of pancreatic cancer using local and global multi-scale feature fusion	DSC: 79.5
					Jaccard: 66.6
Zhang et al., 2021 (47)	NIH and MSD	✓	✗	Lightweight deep convolutional neural network	Mean DSC: 84.90
					Min DSC: 61.82
					Max DSC: 91.46
Yu et al., 2018 (8)	NIH	✓	✗	Recurrent saliency transformation network	Mean DSC: 84.50
					Min DSC: 62.81
					Max DSC: 91.02
Oktay et al., 2018 (7)	NIH	✓	✗	Attention u-net	DSC: 82.2
Chen et al., 2022 (54)	NIH and MSD	✓	✗	Attention mechanism-based feature propagation and fusion	Precision: 85.6
					Recall: 85.9
					IoU: 74.8
Kim et al., (55)	MSD	✓	✗	Scalable gradient-based optimization	DSC 1: 80.61
					DSC 2: 51. 75
					NSD 1: 95.83
					NSD 2: 73.09
Liu et al., 2020 (18)	Private, MSD and TCIA	✗	✓	CNN with modified VGG network	Accuracy: 87.4
					Specificity: 86.7
					Sensitivity: 91.5

(Continued)

TABLE 2 Continued

Author	Dataset	Segmentation	Classification	Algorithm	Performance measure
Li et al., 2023 (24)	Private	✓	✓	Adaptive-metric graph neural network and causal contrastive mechanism	Accuracy: 88.9
					Sensitivity: 88.7
					Specificity: 89.1
					AUC: 94.9
Qu et al., 2023 (23)	Private and MSD	✓	✓	Multiple instance learning and anatomically guided shape normalization	Accuracy: 89
					Sensitivity: 88
					Specificity: 89
					AUC: 94
Proposed algorithm	**MSD and HFH**	✓	✓	**nn-Unet and multi-instance learning**	**DSC 1: 81.64**
					DSC 2: 52.78
					Sensitivity: 90.5
					Specificity: 90.8
					Accuracy: 90.7
					AUC: 90.3

MSD, medical segmentation decathlon; NIH, National Institutes of Health; HFH, Henry Ford Health; DSC, dice similarity coefficients; IoU, intersection over union; NSD, normalized surface distance; AUC, area under the ROC curve.
The symbol ✓ indicates that the reported paper used the method or result, whereas the symbol ✗ signifies that the reported paper does not include method, or result.

Table 2 summarizes the state-of-the-art techniques validated on different datasets such as FLARE, NIH, and MSD. As indicated by the references in the table, the codes for these studies are not available, making it impossible to reproduce their investigations. Therefore, a direct comparison of these approaches is not feasible. However, all these studies used pancreatic images either for classification or segmentation tasks. A modified CNN model was trained to classify patches as cancerous or non-cancerous (18). The model was trained on the local dataset and externally tested on 281 patients with pancreatic cancer, and 82 individuals with normal pancreas. Researchers have developed a generalized pancreatic cancer diagnosis, and the method consists of anatomically guided shape normalization, instance-level contrastive learning, and a balance-adjustment strategy (23) on two unseen datasets (a private test set with 316 and a publicly available test set with 281). The effectiveness of the adaptive-metric graph neural network and causal contrastive mechanism has been developed to enhance the discriminability of the target features and improve early diagnosis stability (24). The training dataset with cross-validation consists of 953 subjects including 554 pancreatic cancer and 399 non-tumor pancreas. Despite the superior performance reported in the majority of the studies, a limited dataset with just a few hundred samples was utilized. We evaluated our pipeline on a significantly larger dataset with 463 cases and 2,882 controls CT images, whereby > 90 % accuracy and sensitivity values were achieved. The significant diversity in the HFH dataset ensured strong generalization capabilities, demonstrating its superior performance. Moreover, our deployment of the segmentation model enabled both accurate normal pancreas and pancreatic cancer detection. As a result, our approach may enable an early detection modality that affords comprehensive options for clinicians to assess earlier onset of pancreatic cancers as well as offer curative intention options for pancreatic cancers that would not otherwise be feasible.

5 Conclusion

In this paper, we propose an end-to-end model for accurate pancreatic tumor prediction. The model incorporates segmentation using the nnU-Net architecture and multi-instance classification using weakly supervised learning. The pancreatic tumor samples are processed by localizing the area of interest from the segmented image. A bag is then formed for each region, which is labeled based on the grade. Finally, the multi-instance learning model is trained for classification. The proposed MIL classification technique achieves an optimal performance by utilizing patient label information on the cropped image, not on the whole pixel patches. Our experimental findings demonstrate that the proposed framework outperforms nnU-Net with Inception V3 by a large margin (7.0%) using the HFH test dataset. From the results, it is evident that the two-stage deep learning architecture of patient radiographic imaging has the potential to be of great assistance in the pursuit of early pancreatic tumor detection. Furthermore, it has the potential to reduce the number of incorrect diagnoses of pancreatic cancer, which would ultimately result in much-required improvements in patient care. We will investigate the possibility of employing auto-encoding DNN rather than K-means in the future.

Data availability statement

The datasets presented in this article are not readily available because CT images from HFH and Memorial Sloan Kettering Cancer Centre (MSKCC) were used to develop a novel algorithm to detect the presence of pancreas tumors in case versus control images. This data cannot be shared without patient approval. MSKCC is an open-source dataset that was used to train the segmentation model. Requests to access the datasets should be directed to DK, System Director, Surgical Oncology, Department of Surgery, Henry Ford Health, Detroit, MI, USA, dkwon1@hfhs.org.

Ethics statement

The studies involving humans were approved by Henry Ford Health, Detroit MI, USA. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because data has been anonymized for this study. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because data has been anonymized for this study.

Author contributions

SM: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. KB: Formal analysis, Software, Validation, Writing – review & editing. HK: Formal analysis, Validation, Writing – review & editing, Methodology. CA: Formal analysis, Methodology, Validation, Writing – review & editing. JC: Data curation, Investigation, Validation, Visualization, Writing – review & editing. DK: Data curation, Investigation, Supervision, Validation, Visualization, Writing – review & editing. AR: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Henry Ford Pancreatic Cancer Centre (Grant no. FT/03/2032/2019).

Acknowledgments

The Department of Radiology at Henry Ford Health assisted in obtaining images for this project. The authors would like to express their gratitude to the Yardi School of Artificial Intelligence for their valuable assistance with this project. This study used the high-performance computing capabilities of the IIT Delhi.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1362850/full#supplementary-material

References

1. Pancreatic cancer prognosis, Aug. 08, 2021. Available online at: https://www.hopkinsmedicine.org/health/conditions-and-diseases/pancreatic-cancer/pancreatic-cancer-prognosis (Accessed Oct. 04, 2022).

2. Kenner B, Chari ST, Kelsen D, Klimstra DS, Pandol SJ, Rosenthal M, et al. Artificial intelligence and early detection of pancreatic cancer: 2020 summative review. *Pancreas*. (2021) 50:251. doi: 10.1097/MPA.0000000000001762

3. Isensee F. From manual to automated design of biomedical semantic segmentation methods. University of Heidelberg, Heidelberg, Germany (2020). doi: 10.11588/HEIDOK.00029345

4. Wang T, Lei Y, Schreibmann E, Roper J, Liu T, Schuster D, et al. Lesion segmentation on 18 F-fluciclovine (18 F-FACBC) PET/CT images using deep learning. *Front Oncol*. (2023) 13:1274803. doi: 10.3389/fonc.2023.1274803

5. Zhou Y, LALANDE A, Chevalier C, Baude J, Aubignac L, Boudet J, et al. Deep learning application for abdominal organs segmentation on 0.35 T MR-Linac images. *Front Oncol*. (2024) 13:1285924. doi: 10.3389/fonc.2023.1285924

6. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. *Lecture Notes in Computer Science* Cham: Springer (2015) , pp. 234–41. doi: 10.1007/978-3-319-24574-4_28

7. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. (2018). doi: 10.48550/arXiv.1804.03999

8. Yu Q, Xie L, Wang Y, Zhou Y, Fishman EK, Yuille AL. (2018). Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT, USA. pp. 8280–9. doi: 10.1109/CVPR.2018.00864

9. Augusto JB, Davies RH, Bhuva AN, Knott KD, Seraphim A, Alfarih M, et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *Lancet Digital Health*. (2021) 3:e20–8. doi: 10.1016/S2589-7500(20)30267-3

10. Zhou H, Li L, Liu Z, Zhao K, Chen X, Lu M, et al. Deep learning algorithm to improve hypertrophic cardiomyopathy mutation prediction using cardiac cine images. *Eur Radiol*. (2021) 31:3931–40. doi: 10.1007/s00330-020-07454-9

11. Wu JR, Leu HB, Yin WH, Tseng WK, Wu YW, Lin TH, et al. The benefit of secondary prevention with oat fiber in reducing future cardiovascular event among CAD patients after coronary intervention. *Sci Rep*. (2019) 9:1–6. doi: 10.1038/s41598-019-39310-2

12. Long J, Shelhamer E, Darrell T. (2015). Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–40. doi: 10.48550/arXiv.1411.4038

13. Jiang C, Liu X, Qu Q, Jiang Z, Wang Y. Prediction of adenocarcinoma and squamous carcinoma based on CT perfusion parameters of brain metastases from lung cancer: a pilot study. *Front Oncol*. (2023) 13:1225170. doi: 10.3389/fonc.2023.1225170

14. Wang CJ, Hamm CA, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. *Eur Radiol*. (2019) 29:3348–57. doi: 10.1007/s00330-019-06214-8

15. Yao W, He JC, Yang Y, Wang JM, Qian YW, Yang T, et al. The prognostic value of tumor-infiltrating lymphocytes in hepatocellular carcinoma: a systematic review and meta-analysis. *Sci Rep*. (2017) 7:1–11. doi: 10.1038/s41598-017-08128-1

16. Wang G, Van Stappen G, De Baets B. Automated detection and counting of Artemia using U-shaped fully convolutional networks and deep convolutional networks. *Expert Syst Appl*. (2021) 171:114562. doi: 10.1016/j.eswa.2021.114562

17. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. *Artif Intell healthcare*. (2020), 25–60. doi: 10.1016/B978-0-12-818438-7.00002-2

18. Liu KL, Wu T, Chen PT, Tsai YM, Roth H, Wu MS, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. *Lancet Digital Health*. (2020) 2:e303–13. doi: 10.1016/S2589-7500(20)30078-9

19. Sekaran K, Chandana P, Krishna NM, Kadry S. Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer. *Multimedia Tools Appl*. (2020) 79:10233–47. doi: 10.1007/s11042-019-7419-5

20. Xuan W, You G. Detection and diagnosis of pancreatic tumor using deep learning-based hierarchical convolutional neural network on the internet of medical things platform. *Future Generation Comput Syst*. (2020) 111:132–42. doi: 10.1016/j.future.2020.04.037

21. Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U. Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. *IEEE Trans Med Imaging*. (2019) 38:1777–87. doi: 10.1109/TMI.42

22. Liu Z, Su J, Wang R, Jiang R, Song YQ, Zhang D, et al. Pancreas Co-segmentation based on dynamic ROI extraction and VGGU-Net. *Expert Syst Appl*. (2022) 192:116444. doi: 10.1016/j.eswa.2021.116444

23. Qu J, Wei X, Qian X. Generalized pancreatic cancer diagnosis via multiple instance learning and anatomically-guided shape normalization. *Med Image Anal*. (2023) 86:102774. doi: 10.1016/j.media.2023.102774

24. Li X, Guo R, Lu J, Chen T, Qian X. Causality-driven graph neural network for early diagnosis of pancreatic cancer in non-contrast computerized tomography. *IEEE Trans Med Imaging*. (2023) 42:1656–67. doi: 10.1109/TMI.2023.3236162

25. Luo Y, Chen X, Chen J, Song C, Shen J, Xiao H, et al. Preoperative prediction of pancreatic neuroendocrine neoplasms grading based on enhanced computed tomography imaging: validation of deep learning with a convolutional neural network. *Neuroendocrinology*. (2020) 110:338–50. doi: 10.1159/000503291

26. Song C, Wang M, Luo Y, Chen J, Peng Z, Wang Y, et al. Predicting the recurrence risk of pancreatic neuroendocrine neoplasms after radical resection using deep learning radiomics with preoperative computed tomography images. *Ann Trans Med*. (2021) 9:833. doi: 10.21037/atm

27. Fu M, Wu W, Hong X, Liu Q, Jiang J, Ou Y, et al. Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images. *BMC Syst Biol*. (2018) 12:119–27. doi: 10.1186/s12918-018-0572-z

28. Qiao Z, Du C. RAD-UNet: a residual, attention-based, dense UNet for CT sparse reconstruction. *J Digital Imaging*. (2022) 35:1748–58. doi: 10.1007/s10278-022-00685-w

29. Qu T, Wang X, Fang C, Mao L, Li J, Li P, et al. M3Net: A multi-scale multi-view framework for multi-phase pancreas segmentation based on cross-phase non-local attention. *Med Image Anal*. (2022) 75:102232. doi: 10.1016/j.media.2021.102232

30. Nakao M, Nakamura M, Mizowaki T, Matsuda T. Statistical deformation reconstruction using multi-organ shape features for pancreatic cancer localization. *Med image Anal*. (2021) 67:101829. doi: 10.1016/j.media.2020.101829

31. Simpson AL, Antonelli M, Bakas S, Bilello M, Farahani K, Van Ginneken B, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*. (2019). doi: 10.48550/arXiv.1902.09063

32. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J digital Imaging*. (2013) 26:1045–57. doi: 10.1007/s10278-013-9622-7

33. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. (2019) 25:1301–9. doi: 10.1038/s41591-019-0508-1

34. Dong D, Fu G, Li J, Pei Y, Chen Y. An unsupervised domain adaptation brain CT segmentation method across image modalities and diseases. *Expert Syst Appl*. (2022) 207:118016. doi: 10.1016/j.eswa.2022.118016

35. Roth HR, Oda H, Hayashi Y, Oda M, Shimizu N, Fujiwara M, et al. Hierarchical 3D fully convolutional networks for multi-organ segmentation. *arXiv preprint arXiv:1704.06382*. (2017). doi: 10.48550/arXiv.1704.06382

36. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell*. (2017) 40:834–48. doi: 10.1109/TPAMI.2017.2699184

37. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. (2017) 39:2481–95. doi: 10.1109/TPAMI.34

38. Zhao H, Shi J, Qi X, Wang X, Jia J. (2017). Pyramid scene parsing network, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, USA., pp. 2881–90. doi: 10.1109/CVPR.2017.660

39. Lin G, Milan A, Shen C, Reid I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, USA, pp. 1925–34. doi: 10.1109/CVPR.2017.549

40. Tahri S, Texier B, Nunes JC, Hemon C, Lekieffre P, Collot E, et al. A deep learning model to generate synthetic CT for prostate MR-only radiotherapy dose planning: a multicenter study. *Front Oncol*. (2023) 13:1279750. doi: 10.3389/fonc.2023.1279750

41. Yao J, Shi Y, Lu L, Xiao J, Zhang L. (2019). Deepprognosis: Preoperative prediction of pancreatic cancer survival and surgical margin via contrast-enhanced CT imaging, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2020. Lecture Notes in Computer Science*, Lima, Peru, October 4–8, 2020, vol 12262, pp. 272–82. Springer, Cham. doi: 10.1007/978-3-030-59713-9_27

42. Tennakoon R, Bortsova G, Ørting S, Gostar AK, Wille MM, Saghir Z, et al. Classification of volumetric images using multi-instance learning and extreme value theorem. *IEEE Trans Med Imaging*. (2019) 39:854–65. doi: 10.1109/TMI.42

43. Tibo A, Jaeger M, Frasconi P. Learning and interpreting multi-multi-instance learning networks. *J Mach Learn Res*. (2020) 21:1–60.

44. He K, Zhang X, Ren S, Sun J. (2016). Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, pp. 770–8. doi: 10.1109/CVPR.2016.90

45. Waqas M, Tahir MA, Khan SA. Robust bag classification approach for multi-instance learning via subspace fuzzy clustering. *Expert Syst Appl*. (2023) 214:119113. doi: 10.1016/j.eswa.2022.119113

46. Li Y, Wu H. A clustering method based on K-means algorithm. *Phys Proc*. (2012) 25:1104–9. doi: 10.1016/j.phpro.2012.03.206

47. Zhang D, Zhang J, Zhang Q, Han J, Zhang S, Han J. Automatic pancreas segmentation based on lightweight DCNN modules and spatial prior propagation. *Pattern Recognition*. (2021) 114:107762. doi: 10.1016/j.patcog.2020.107762

48. Ibragimov B, Toesca DA, Chang DT, Yuan Y, Koong AC, Xing L. Automated hepatobiliary toxicity prediction after liver stereotactic body radiation therapy with deep learning-based portal vein segmentation. *Neurocomputing*. (2020) 392:181–8. doi: 10.1016/j.neucom.2018.11.112

49. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44(3):837–45. doi: 10.2307/2531595

50. Luo Y, Chen X, Chen J, Song C, Shen J, Xiao H, et al. Preoperative prediction of pancreatic neuroendocrine neoplasms grading based on enhanced computed tomography imaging: validation of deep learning with a convolutional neural network. *Neuroendocrinology*. (2020) 110:338–50. doi: 10.1159/000503291

51. Ma H, Liu ZX, Zhang JJ, Wu FT, Xu CF, Shen Z, et al. Construction of a convolutional neural network classifier developed by computed tomography images for pancreatic cancer diagnosis. *World J Gastroenterol*. (2020) 26:5156. doi: 10.3748/wjg.v26.i34.5156

52. Zhang F, Wang Y, Yang H. Efficient context-aware network for abdominal multi-organ segmentation. *arXiv preprint arXiv:2109.10601*. (2021). doi: 10.48550/arXiv.2109.10601

53. Wang H, Wang G, Liu Z, Zhang S. Global and local multi-scale feature fusion enhancement for brain tumor segmentation and pancreas segmentation. In: Crimi A, Bakas S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes. Lecture Notes in Computer Science, Shenzhen, China, October 17, 2019, Springer, Cham. (2020) 11992:80–88. doi: 10.1007/978-3-030-46640-4_8

54. Chen H, Liu Y, Shi Z. FPF-Net: feature propagation and fusion based on attention mechanism for pancreas segmentation. *Multimedia Syst*. (2022) 29(2):525–38. doi: 10.1007/s00530-022-00963-1

55. Kim S, Kim I, Lim S, Baek W, Kim C, Cho H, et al. Scalable neural architecture search for 3d medical image segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham (2019), pp. 220–228. doi: 10.48550/arXiv.1906.05956

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 16 September 2024
DOI 10.3389/fonc.2024.1401095

Check for updates

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Edoardo Poletto,
University of Verona, Italy
Ji-Qiao Zhu,
Capital Medical University, China

*CORRESPONDENCE
Xiang Zhang
seva1101@163.com
Shangeng Weng
shangeng@sina.com

†These authors share first authorship

RECEIVED 14 March 2024
ACCEPTED 22 August 2024
PUBLISHED 16 September 2024

CITATION
Zhong Y, Chen L, Ding F, Ou W, Zhang X and Weng S (2024) Assessing microvascular invasion in HBV-related hepatocellular carcinoma: an online interactive nomogram integrating inflammatory markers, radiomics, and convolutional neural networks. *Front. Oncol.* 14:1401095. doi: 10.3389/fonc.2024.1401095

COPYRIGHT
© 2024 Zhong, Chen, Ding, Ou, Zhang and Weng. 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.

Assessing microvascular invasion in HBV-related hepatocellular carcinoma: an online interactive nomogram integrating inflammatory markers, radiomics, and convolutional neural networks

Yun Zhong[1,2,3,4†], Lingfeng Chen[1,2,3,4†], Fadian Ding[1,2,3,4], Wenshi Ou[1,2,3,4], Xiang Zhang[1,2,3,4*] and Shangeng Weng[1,2,3,4*]

[1]Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital, Fujian Medical University, Fuzhou, China, [2]Fujian Abdominal Surgery Research Institute, The First Affiliated Hospital, Fujian Medical University, Fuzhou, China, [3]Department of Hepatobiliary and Pancreatic Surgery, National Regional Medical Center, Binhai Campus of the First Affiliated Hospital, Fujian Medical University, Fuzhou, China, [4]Fujian Provincial Key Laboratory of Precision Medicine for Cancer, The First Affiliated Hospital, Fujian Medical University, Fuzhou, China

Objective: The early recurrence of hepatocellular carcinoma (HCC) correlates with decreased overall survival. Microvascular invasion (MVI) stands out as a prominent hazard influencing post-resection survival status and metastasis in patients with HBV-related HCC. The study focused on developing a web-based nomogram for preoperative prediction of MVI in HBV-HCC.

Materials and methods: 173 HBV-HCC patients from 2017 to 2022 with complete preoperative clinical data and Gadopentetate dimeglumine-enhanced magnetic resonance images were randomly divided into two groups for the purpose of model training and validation, using a ratio of 7:3. MRI signatures were extracted by pyradiomics and the deep neural network, 3D ResNet. Clinical factors, blood-cell-inflammation markers, and MRI signatures selected by LASSO were incorporated into the predictive nomogram. The evaluation of the predictive accuracy involved assessing the area under the receiver operating characteristic (ROC) curve (AUC), the concordance index (C-index), along with analyses of calibration and decision curves.

Results: Inflammation marker, neutrophil-to-lymphocyte ratio (NLR), was positively correlated with independent MRI radiomics risk factors for MVI. The performance of prediction model combined serum AFP, AST, NLR, 15 radiomics features and 7 deep features was better than clinical and radiomics models. The combined model achieved C-index values of 0.926 and 0.917, with AUCs of 0.911 and 0.907, respectively.

Conclusion: NLR showed a positive correlation with MRI radiomics and deep learning features. The nomogram, incorporating NLR and MRI features, accurately predicted individualized MVI risk preoperatively.

KEYWORDS

hepatocellular carcinoma, microvascular invasion, radiomics, convolutional neural network, inflammation marker

1 Introduction

Liver cancer, the fifth most common malignant tumor, ranks fourth in mortality of cancer (1). Hepatocellular carcinoma (HCC) accounts for approximately 90% of the cases of liver cancer worldwide (2). At least 50% cases of HCC worldwide were caused by hepatitis B virus (HBV). In China, chronic HBV infection is also the main cause of HCC (3, 4).

At present, the main treatment options for HCC encompass liver resection, liver transplant, and transcatheter arterial chemoembolization (TACE) (5). Post-surgical resection, the annual recurrence rate of hepatocellular carcinoma (HCC) is at least 10%, escalating to 70–80% within a five-year period. The recurrent HCC tumors may probably progress into incurable, advanced-stage disease in most patients (6). So accurately identifying high-risk patients, estimating the prognosis of those with HCC, and extending survival time are of critical importance in clinical practice.

Microvascular invasion (MVI) serves as a significant independent prognostic factor for patients with hepatocellular carcinoma (HCC) following curative treatments, including surgical resection, liver transplantation, or alternative therapeutic interventions (7). Under microscopic examination, microvascular invasion (MVI) is characterized by the presence of neoplastic cell clusters within the lumina of endothelial-lined vascular channels, including those of the portal and hepatic venous systems (8). Recent evidence suggested that MVI might be the first step in the occurrence of intra-hepatic or systemic metastasis of HCC (9). Several retrospective studies suggest an association between inflammation and MVI in HBV-HCC patients (10). Inflammatory environment may increase hepatic microvascular permeability so that cancer cells can invade through the blood vessel wall (11). The occurrence of microvascular invasion (MVI) signifies the infiltration of cancer cells into the vasculature, heralding the potential onset of metastasis.

There has been an increasing interest in predicting MVI through preoperative data. Recent studies have highlighted the potential utility of blood-cell-inflammatory markers as non-invasive predictors of MVI (12, 13). These markers, including neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII), reflect the patient's systemic inflammatory response, which is closely associated with tumor progression and metastasis. It has been previously observed that an elevated NLR is correlated with an increased incidence of MVI in HCC patients (14). Furthermore, the SII, which combines neutrophil, lymphocyte, and platelet counts, has shown superior predictive accuracy for MVI compared to individual markers (15). The integration of these inflammatory markers into clinical practice could enhance the preoperative assessment of MVI, allowing for more tailored and effective treatment plans for HCC patients.

As known, preoperative images are also important preoperative data. Radiomics, defined as the transformation of digital medical images into high-dimensional, analyzable data through advanced computational techniques, facilitates the exhaustive extraction and quantification of data from standard radiological imagery, thereby yielding critical insights into the cancer phenotype and the tumor microenvironment (16). Recent studies suggest an association between some radiographic features and local inflammatory status and vascular response in tumor (17). Li Yang reported that a nomogram incorporating radiomic features extracted from hepatobiliary phase (HBP) imaging demonstrated efficacy in the preoperative prognostication of MVI risk in HCC patients (18). PENG LIU has substantiated that radiomic analysis of computed tomography images exhibits a definitive predictive value for MVI in solitary HCC with a dimension less than or equal to 5 cm (19). Gadobenate dimeglumine-enhanced MRI imaging carries additional information on tumors than computed tomography, which can also reflect changes in the tumor micro-environment (20).

Deep learning with convolutional neural networks (CNNs) was applied to extract the inherent features of input data automatically (21). Recently, Residual Neural Network, a classical deep learning model, has been widely used for 3D imaging data analysis in medical field including MRI (22). Li et al. utilized a six-layer CNN to extract features from MR images to classify low grade gliomas and found an improvement on the traditional radiomics (23). Unlike traditional computed features, deep features retain a large amount of the global spatial information. So far, very little attention has been paid to the role of blood-cell-inflammatory markers combined with radiomics and deep learning features from preoperative MRI in predicting MVI.

In this study, we aimed to integrate preoperative MRI characteristics with inflammatory markers to create and confirm a new predictive nomogram for the preoperative estimation of MVI in HBV- related HCC. This nomogram facilitates the preoperative

determination of the individualized risk of MVI in HCC patients, which is especially beneficial for categorizing patients into appropriate treatment groups.

2 Materials and methods

2.1 Patients and follow-up

Ethical approval was obtained for this retrospective study, and the requirement for informed consent was waived. Patients were retrospectively collected from January 1, 2017 to November 31, 2022 (The First Affiliated Hospital of Fujian Medical University). [MTCA, ECFAH of FMU[2015] No.084-1] The inclusion criteria were: (a) Pathologically diagnosed hepatocellular carcinoma with MVI evaluation; (b) HBV related HCC; (c) Without history of prior intervention therapy; (d) Gadobenate dimeglumine-enhanced MRI was performed before surgery within 1 week; (e) No portal or hepatic vein invasion; (f) No lymph node or distant metastasis. The exclusion criteria were: (a) Complicated with other malignant tumors, and multiple primary or recurrent liver cancer; (b) Pathology-confirmed malignancies were not HCC; (c) Combined with other infectious diseases, immune diseases, hematologic diseases or allergic diseases; (d) Patients with emergency surgery for heparorrhexis; (e) Incomplete medical information. The ultimate composition of the patient cohort encompassed a total of 173 individuals (150 men and 23 women). The training group included 120 patients, and 52 patients were allocated to the validation group. The process for patient inclusion was detailed in the flowchart provided in Figure 1. Recurrence-free survival (RFS) time referred to the time interval from surgery to the date of recurrence, death or the last follow-up. The RFS was analyzed using the Kaplan-Meier method to estimate the survival distribution in this population. The Log-rank test was employed to compare the survival differences between MVI positive and MVI negative groups.

2.2 Laboratory and pathology data acquisition

Clinical-pathological baseline data were systematically extracted from our institution's medical archives. Clinical features included age, gender, BMI, neutrophil, lymphocytes, hemoglobin, platelets, serum albumin, alkaline phosphatase (ALP), aspartate aminotransferase (AST), alanine aminotransferase (ALT), γ-glutamyl transpeptidase (γ-GT), AFP, the Systemic Immune-inflammation Index (SII), the Neutro-phil-to-Lymphocyte Ratio (NLR), the Platelet-to-Lymphocyte Ratio (PLR). MR image characteristics including enhancement pattern, peritumoral enhancement on artery phase, radiologic capsule on delay phase, peritumoral hypointensity were analyzed by 2 experienced radiologists. The pathological attributes, including tumor count, MVI presence, tumor size, Edmondson-Steiner grading, and cirrhosis status in non-tumorous liver tissue, were assessed by 2 experienced pathologists. Laboratory analyses consisted of common hematology tests were performed within one week prior to the surgical intervention.

2.3 Magnetic resonance imaging data acquisition

all study patients underwent gadobenate dimeglumine-enhanced MR imaging using 3.0-Tesla MR scanners (Magnetom

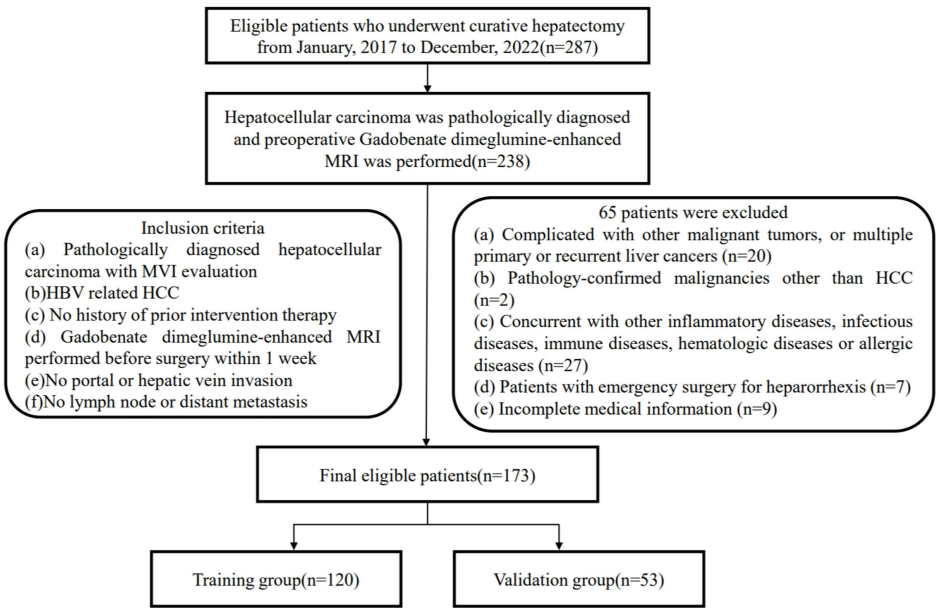


FIGURE 1
Flow chart of the patient enrollment process.

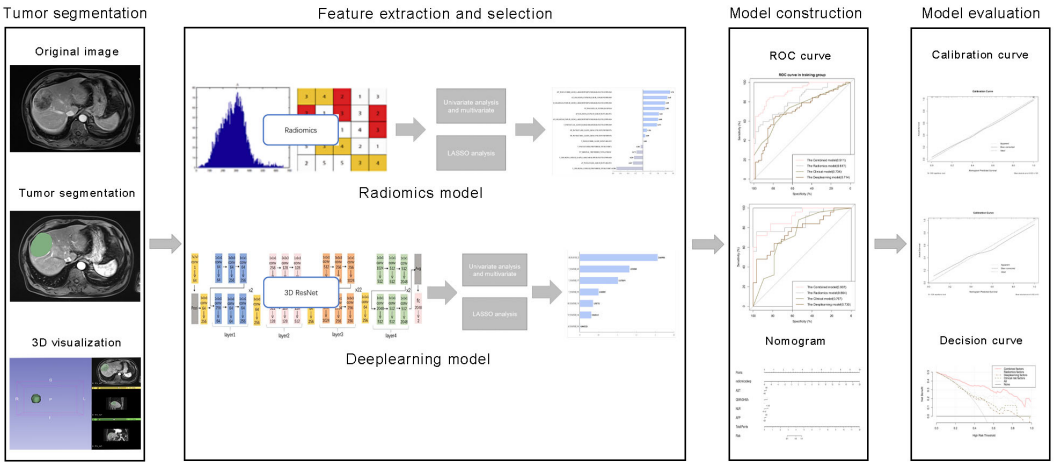


FIGURE 2
Comprehensive workflow diagram of the prediction model. Tumor segmentation in MR images is the first step. After that, MRI feature extraction was conducted separately via radiomics and neural convolutional networks. Student's t-test, Mann-Whitney U test and least absolute shrinkage and selection operator (LASSO) were used to feature selection, sequentially. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) were calculated to evaluate the prediction efficiency of the radiomic features. Finally, a nomogram was developed and evaluated.

Skyra Siemens Healthcare). Conventional MRI parameters were as follows: repetition time (TR)/echo time (TE), 6000/125 ms; number of excitation (NEX), 1; field of view (FOV), 20 mm; slices, 22; slice thickness, 3 mm. Gadopentetate dimeglumine was administrated with a dose of 0.1 mmol/kg, followed by a 20-mL continuous saline flush. Imaging sequences included T2-weighted imaging with fat suppression, T1-weighted imaging, and contrast-enhanced T1-weighted im-aging obtained at 20–30 s (by monitoring, the scan is triggered when the contrast agent reaches the ascending aorta), 70–90 s, 100–120 s, and 160–180 s respectively after contrast medium injection. That included a transverse arterial phase, transverse portal venous phase and transverse delayed phase. Image characteristics including enhancement pattern, peritumoral enhancement on artery phase, radiologic capsule on delay phase and peritumoral hypointensity were analyzed by 2 experienced radiologists (Figure 2).

2.4 Workflow of radiomics analysis

The radiomics analysis workflow encompassed the segmentation of the tumor, the extraction and selection of features, followed by the construction and evaluation of the predictive model (Figure 2). 1223 candidate texture parameters were extracted, including shape_LeastAxisLength, shape_MeshVolume,shape_MinorAxisLength, glrlm_RunEntropy,

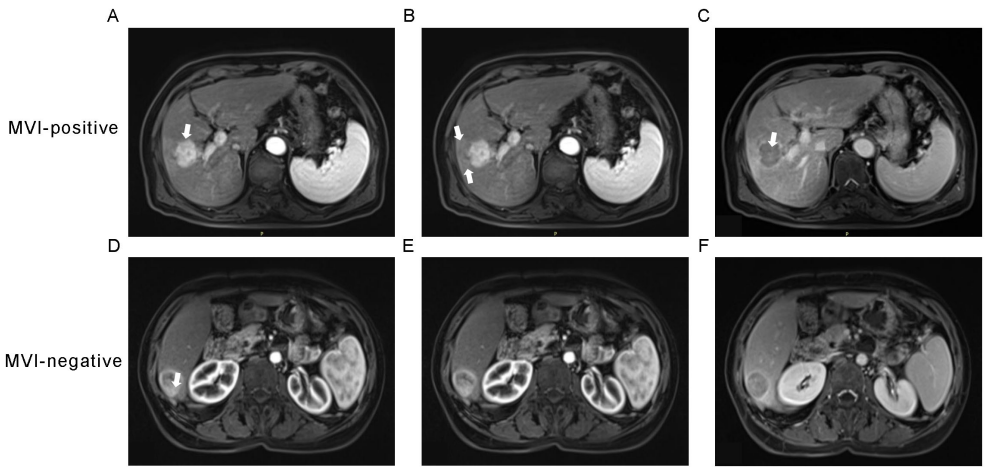


FIGURE 3
Characteristic MRI images of microvascular invasion (MVI)-positive and MVI-negative hepatocellular carcinoma. **(A, B)** The axial arterial phase image shows a non-smooth hype vascular tumor (arrow) with peritumoral enhancement (arrows). **(C)** The delayed phase image exhibits a rapid low signal intensity (arrow) within the tumor. **(D, E)** The arterial phase image demonstrates a smooth tumor, accompanied by **(F)** a radiologic capsule in the delayed phase.

ngtdm_Coarseness, gldm_DependenceNonUniformity and so on. All feature extraction was implemented by Python (Version 3.10.5).

2.4.1 Tumor segmentation

The MRI images of the patients were exported in DICOM format. Blind to the pathological findings, the radiologist, boasting three years of expertise in abdominal imaging, employed the 3D-Slice software (www.slice.org) to demarcate the region of interest (ROI) on each slice encompassing the tumor. Regions of interest (ROIs) were drawn on all arterial phase (AP), venous phase (VP), delayed phase (DP) and T2-weighted (T2) images slice-by-slice for each patient (Figure 3).

2.4.2 Feature extraction

Features, including radiomics features and deep learning features, were extracted through Python package Pyradiomics and trained 3D ResNet, respectively. 1223 radiomics texture parameters were extracted, including first-order features and texture features in every phase. In another way, 3D ResNet, a pre-trained deep learning architecture, was utilized to harvest 512 deep features from the ROIs of the images for each patient in every MRI sequence. All feature extraction was implemented by Python (Version 3.10.5).

2.4.3 Model construction and evaluation

The clinical model was constructed using features that demonstrated a statistical significance with p-values less than 0.05 in the multivariate logistic regression analysis. We built a radiomics model and a deep learning model using features that were extracted by the above method. A combined model was constructed by integrating elements from the clinical model with the most effective signatures from radiomics and deep learning models. The efficiency of the radiomic feature-based predictions was assessed by computing the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC). The area under the curve (AUC) was ascertained along with its 95% confidence interval (CI), in addition to quantifying the sensitivity, specificity, and precision of the model. A nomogram was constructed to visually represent the combined model. The discriminative efficacy of the nomogram was quantified employing Harrell's concordance index (C-index). Calibration curves were generated and analyzed to assess the diagnostic concordance of the nomogram within the training and validation cohorts. Decision curve analysis was employed to ascertain the clinical utility of the nomogram. A web-based calculator for the dynamic prediction of MVI was created utilizing the "shiny" and "DynNom" packages (https://www.shinyapps.io/).

2.5 Evaluation of NLR and radiomics and deep features

The correlation between the NLR and extracted features were represented by "ggpubr" and "ggExtra" R packages. Besides, we explored the relationship between the NLR and risk score of every model by Pearson correlation analysis.

2.6 Statistical analysis

All statistical analyses were conducted utilizing IBM SPSS Statistics (Version 24.0), Python (Version 3.10.5), and R (Version 4.3.1) software packages. A two-tailed p value less than 0.05 was considered statistically significant.

3 Results

3.1 Baseline characteristics and survival curve

The baseline characteristics of the study participants were presented in Supplementary Table S1. These data provided a comprehensive overview of the demographic, clinical, and biochemical characteristics at the onset of the study. There was no significant difference in the incidence of MVI between the training and validation groups. Meanwhile, we conducted a follow-up study involving 111 patients from this population and plotted the survival curve (Supplementary Figure S1). This survival curve demonstrated that MVI was a major risk factor for HCC recurrence.

3.2 Clinical feature analysis

The clinical and MR image characteristics collected from MVI-negative and MVI-positive groups were presented in Table 1. Youden index was used to find the best cut-off value of inflammation markers (Table 2). NLR, AST, serum AFP levels, and peritumoral enhancement were significantly different between groups through univariate and multivariate analyses ($P < 0.05$).

3.3 Traditional radiomics feature analysis

For all MRI radiomics features, 982 stable features were retained through the evaluation of consistency (ICCs > 0.75). 4 features in artery phase, 2 features in venous phase, 3 features in delayed phase and 6 features in T2-weighted images were selected by Mann-Whitney U-test, univariate logistic, and LASSO analysis, sequentially ($P < 0.05$). The lasso dimension reduction analysis of radiomics features was presented in Supplementary Figure S2. Supplementary Figure S3 shows the predictive performance of each feature.

3.4 Deep learning feature analysis

A total of 951 deep learning features were preserved following the consistency check (ICCs > 0.75). Following the same method, 2 features in artery phase, 1 feature in venous phase, 1 feature in delayed phase, and 3 features in T2-weighted images were selected. The lasso dimension reduction analysis of deep learning features was shown in Supplementary Figure S4. The predictive performance of each characteristic was illustrated in Supplementary Figure S3.

10.3389/fonc.2024.1401095

TABLE 1 Comparisons of patients' characteristics in training and validation datasets.

Characteristics	Training dataset (n = 120)			Validation dataset (n = 53)		
	MVI (+)	MVI (-)	*P*	MVI (+)	MVI (-)	*P*
Age, years			0.583			0.838
< 35	2	3		1	1	
35-65	28	36		21	22	
≥ 65	27	24		3	5	
Gender			0.283			0.614
Male	49	58		21	22	
Female	8	5		4	6	
BMI			0.652			0.994
< 18.5	5	3		1	1	
18.5-25	30	33		18	20	
≥ 25	22	27		6	7	
α-Fetoprotein			**<0.001**			**<0.001**
< 20 ng/mL	16	38		9	16	
≥20ng/mL	41	25		16	12	
Edmondson-Steiner Grade			0.015			0.478
I	7	11		1	1	
II	30	44		16	22	
III	20	8		8	5	
Cirrhosis of background liver			0.866			0.053
Absent	19	18		5	13	
Present	44	39		19	15	
Serum albumin			0.629			0.028
< 35 g/L	8	7		6	1	
≥35 g/L	49	56		19	27	
Alanine transaminase			0.034			0.224
< 40 U/L	29	44		11	17	
≥ 40 U/L	28	19		14	11	
Aspartate transaminase			**0.018**			**0.043**
< 40 U/L	24	46		11	20	
≥ 40 U/L	33	17		14	8	
Total billrubin			0.522			0.346
< 21 μmol/L	47	49		20	25	
≥ 21 μmol/L	10	14		5	3	
γ-Glutamyltransferase			0.022			0.200
< 60 U/L	6	17		9	15	
≥ 60 U/L	51	46		16	13	
Neutrophils, 10^9/L*	3.50 ± 2.12	3.28 ± 1.25	0.483	3.62 ± 1.18	3.36 ± 0.74	0.142
Lymphocyte, 10^9/L*	1.74 ± 0.60	1.64 ± 0.52	0.337	1.80 ± 0.74	1.52 ± 0.43	<0.001
Hemoglobin, g/L*	141 ± 16	142 ± 13	0.837	147 ± 10	132 ± 10	0.006
Platelet,10^9/L*	182 ± 72	166 ± 66	0.251	208 ± 83	169 ± 24	0.034
NLR			**0.019**			**0.004**
< 1.45	24	40		9	21	
≥ 1.45	33	23		16	7	
Tumor size			0.247			0.001
< 5cm	39	49		5	18	
≥ 5cm	18	14		20	10	
Tumor margin			0.005			0.100
Smooth margin	8	23		2	7	
Non-smooth margin	49	40		23	21	
Enhancement pattern			0.398			0.224
Typical	21	28		14	11	
Atypical	36	35		11	17	
Peritumoral enhancement on artery phase			**0.004**			**0.040**
Absent	27	46		9	18	
Present	30	17		16	10	
Radiologic capsule on delay phase			0.698			0.884
Absent	37	43		12	14	
Present	20	20		13	14	
Peritumoral hypointensity			0.269			0.487
Absent	25	34		11	15	
Present	32	29		14	13	

Unless otherwise noted, data are shown as number of patients, with the percentage in parentheses.
*Data are medians, with interquartile ranges in parentheses.
Bold indicate values below 0.05, which are statistically significant.

3.5 Correlation between the NLR and extracted features

The correlation between NLR and radiomics and deep features was tested using Pearson analysis. As shown in Figure 4, the result

TABLE 2 Predictive efficacy of the inflammation markers.

Inflammation marker	Threshold	Youden index	Sensitivity	Specificity	*P*
Neutrophil to lymphocyte ratio,NLR	1.45	0.277	0.585	0.692	**0.002**
Systemic Immune Inflammation Index, SII	482.04	0.104	0.280	0.824	0.341
Platelet to lymphocyteratio, PLR	115.89	0.118	0.415	0.703	0.376

Bold indicate values below 0.05, which are statistically significant.

showed that NLR was positively correlated with features like wavelet.LLL.glszm.LargeAreaHighGrayLevelEmphasis radiomics feature, deep feature 292, deep feature 157 in T2 phase and deep feature 22 in artery phase. Furthermore, we discovered the positive relationship between the NLR and 3 predictive model including deep learning model, radiomics model, and combined model (r = 0.244; *P* = 0.001; r = 0.161; *P* = 0.038; r = 0.209; *P* = 0.007). Reasonably, the results indicated that the inflammatory signature could be a good indicator for reflecting inflammation in MRI images and a new biomarker for MVI prediction with radiomics features in HBV-HCC patients.

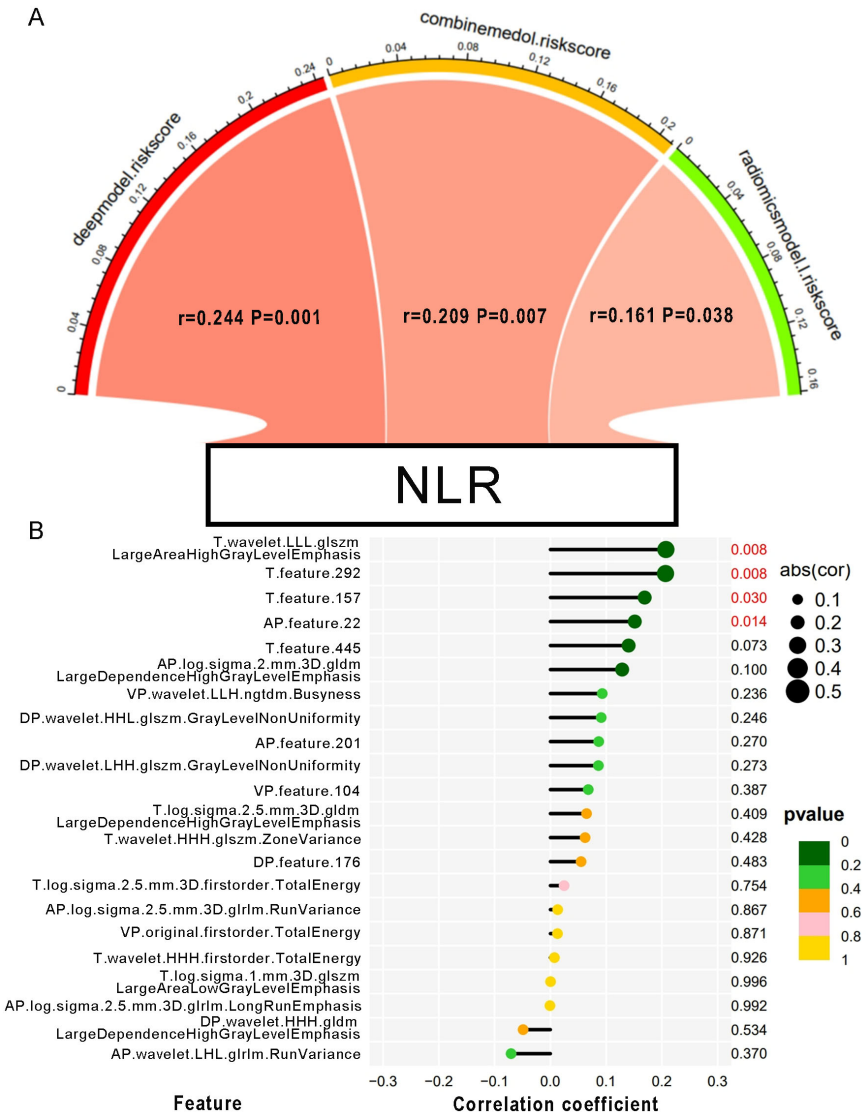


FIGURE 4
The correlation between the NLR and MRI features. **(A)** Pearson correlation between NLR and the risk scores of each predictive model. The color of the band represented the R-value. **(B)** Correlation of the NLR with every single MVI-related MRI feature. The size of the circle represents the magnitude of the correlation coefficient. The color of the circle indicates the magnitude of the p-value.

TABLE 3 Logistics regression analysis for MVI.

Characteristic	Hazard ratio	95%CI	*P*
AFP	3.95	1.936-8.061	**<0.001**
NLR	2.837	1.412-5.703	**0.003**
Peritumoral enhancement	2.605	1.285-5.283	**0.008**
AST	2.916	1.433-5.935	**0.003**

Bold indicate values below 0.05, which are statistically significant.

3.6 Model construction and evaluation

3.6.1 Clinical model

The univariate analysis indicated a significant association between MVI and several independent variables, including NLR, serum AFP, AST, and the presence of peritumoral enhancement. Upon multivariate logistic regression analysis, it was ascertained that NLR [odds ratio (OR) 2.837; 95% confidence interval (CI) 1.412-5.703; P = 0.003], serum AFP [OR 3.95; 95% CI 1.936-8.061; P < 0.001], peritumoral enhancement [OR 2.605; 95% CI 1.285-5.283; P = 0.008], and AST [OR 2.916; 95% CI 1.433-5.935; P = 0.003] emerged as independent predictors of MVI, as shown in Table 3. These were also effective components in the construction of the clinical model. In conclusion, the AUCs for the clinical model yielded values of 0.734 [95% CI: 0.7024-0.8685] for the training group and 0.767 [95% CI: 0.6778-0.9194] for the validation cohort, respectively (Figure 5).

3.6.2 Single radiomics model

Multivariate logistic regression was used to build the model for each MRI phase. The quartet of radiomics models were designated as the arterial, venous+delay, arterial+venous+delay, and T2 model, each predicated on features derived from the arterial phase, portal venous phase, delayed phases, and T2-weighted imaging sequences, respectively. The performance of each model is shown in Table 4. Within the training group, the AUCs for the artery, venous+delay, artery+venous+delay, and T2 models were 0.696 [95% CI: 0.6014-0.7910], 0.774 [95% CI: 0.6917-0.8566], 0.779 [95% CI: 0.6978-0.8611], and 0.698 [95% CI: 0.6025-0.7927]. In the validation cohort, the diagnostic efficacy of the radiomics models was quantitatively evaluated. The arterial model demonstrated a notable area under the curve (AUC) of 0.820, with a 95% confidence interval (CI) ranging from 0.7018 to 0.9382. The venous+delay model exhibited an AUC of 0.769 (95% CI: 0.6343-0.9029), reflecting substantial discriminative capability. The arterial +venous+delay model achieved an AUC of 0.803 (95% CI: 0.6844-0.9213), indicating a robust performance in prognostic evaluation. Lastly, the T2 model presented an AUC of 0.724 (95% CI: 0.5731-0.8755), suggesting a respectable level of diagnostic accuracy within the context of the studied parameters. In the training and validation cohorts, the artery model and artery+venous+delay model predicted more accurately than the other radiomic signature models, respectively.

3.6.3 Radiomics and deep learning model

The Radiomics model, utilizing features extracted from four distinct MRI phases, yielded an AUC of 0.817, with a 95% CI spanning from 0.742 to 0.892 in the training cohort. And the AUC value of the Radiomics model is 0.864(95% CI: 0.761-0.967) in validation group. Then we applied residual networks (ResNet) to construct a predictive model. This prediction model, which used deep MR image features to forecast MVI, gave back the following performance data. In the training cohort, the model demonstrated a specificity of 0.778 and a sensitivity of 0.649. The AUC for this model was determined to be 0.714, with a 95% CI ranging from 0.620 to 0.808. The AUC for the test cohort was 0.767 (95% CI: 0.637-0.896), while the corresponding specificity and sensitivity

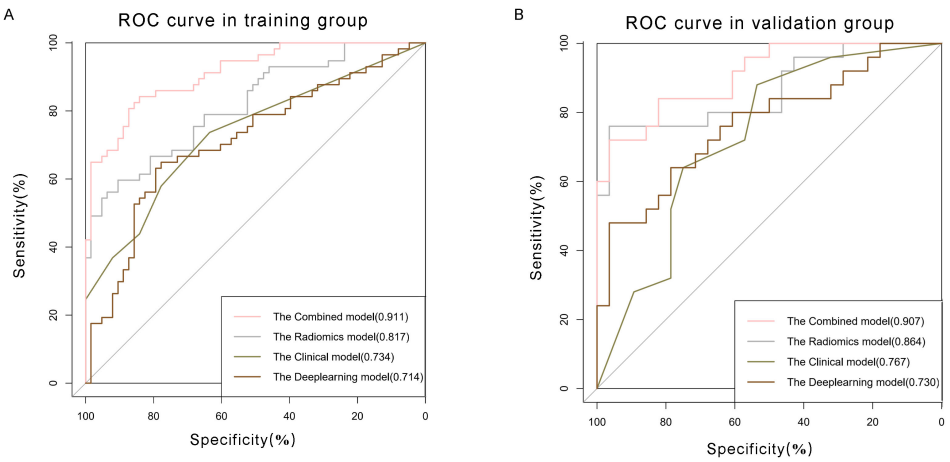


FIGURE 5
Comparison of receiver operating characteristic (ROC) curves for prediction of microvascular invasion. ROC curves of clinical model (pink curve), radiomics model (grey curve), deep learning model (olive green curve), and combined model (brown curve) in the **(A)** training and **(B)** validation datasets. The X-axis represents the specificity predicted by the model, the Y-axis represents the model's sensitivity, and the AUC indicates the predictive performance of the predictive model.

TABLE 4 Predictive efficacy of different radiomics models.

Different models	Training group (n = 120)				Validation group (n = 53)			
	sensitivity	specificity	AUC	AUC (95% CI)	sensitivity	specificity	AUC	AUC (95% CI)
The radiomics model	0.807	0.73	0.858	0.7928-0.9226	0.84	0.786	0.851	0.7401-0.9628
The artery+venous+ delayed model	0.807	0.635	0.779	0.6978-0.8611	0.64	0.857	0.803	0.6844-0.9213
The artery model	0.439	0.905	0.696	0.6014-0.7910	0.857	0.76	0.82	0.7018-0.9382
The venous+delayed model	0.86	0.587	0.774	0.6917-0.8566	0.76	0.75	0.769	0.6343-0.9029
The T2 model	0.386	0.937	0.698	0.6025-0.7927	0.6	0.929	0.724	0.5731-0.8755

were 0.964 and 0.480. The ROC curves of radiomics and deep learning models were presented in Figure 5.

3.6.4 Development and validation of the combined model

Clinical characteristics, radiomics features and deep learning features were incorporated to construct a Combined model. This MVI predictive model combining all significant independent predictors outperformed other models with an AUC of 0.900(95% CI: 0.8474-0.9532). The combined model was able to predict MVI more accurately. The performance of the model in the training set was characterized by an AUC of 0.911, with a 95% CI ranging from 0.862 to 0.960. It demonstrated a sensitivity of 84.2% and a specificity of 84.1%, as detailed in Figure 5. When utilized in the validation group, the model produced an AUC of 0.907, with a 95% confidence interval (CI) from 0.831 to 0.984, with specificity values of 96.4%, sensitivity values of 72.0% respectively. The performance of the combined model, the radiomics model, the deep learning model, and the clinical model is shown in Table 5. Figure 6 depicts the nomogram based on the merged model. The nomogram demonstrated satisfactory predictive accuracy, as reflected by a C-index of 0.926, with a 95% CI ranging between 0.881 and 0.969, in the training group. Similarly, in the validation group, the C-index was an impressive 0.917, with a 95% CI of 0.846 to 0.988. The calibration curves (Figure 6) illustrated that the nomogram's predicted probabilities exhibited a high degree of correlation with the actual occurrences of MVI in both the training and validation groups. In the training group, the calibration was statistically significant with a P-value of 0.022, while in the validation group, the result nearly reached statistical significance with a P-value of 0.052. Decision curve for the nomogram was demonstrated in Figure 6. Considering the findings, a dynamic online tool (https://zhongyun.shinyapps.io/HBV-HCC_MVI_NOMOGRAM/) has been created for forecasting the likelihood of microvascular invasion in individuals diagnosed with HBV-related HCC (Figure 7).

4 Discussion

The role of systemic inflammation in the pathogenesis of hepatocellular carcinoma (HCC) in the context of hepatitis B virus (HBV) infection has been extensively examined (18). The design of this retrospective study aimed to construct and confirm a novel nomogram that incorporates preoperative indicators, including markers of blood-cell-mediated inflammation and radiomics-derived inflammation, for the prognostication of MVI in individuals with HBV-HCC. In contrast to earlier findings, we introduced MRI deep radiomics features to construct a new predictive model called the deep radiomics model.

In the systemic milieu induced by the tumor, a myriad of inflammatory cytokines contribute to the invasion of HCC cells and the progression of metastases (20). Inflammation markers derived from blood cells, such as the neutrophil-to-lymphocyte ratio (NLR), systemic immune-inflammation index (SII), and aspartate aminotransferase-to-platelet ratio index (APRI), represent ratios that correlate various cellular proportions implicated in the inflammatory response. In our investigation, patients with HBV-HCC exhibited elevated levels of these inflammatory indices. Hernandez-Ainsa et al. have reported an association between

TABLE 5 Predictive efficacy of predictive models.

Different models	Training group (n = 120)				Validation group (n = 53)			
	sensitivity	specificity	AUC	(95% CI)	sensitivity	specificity	AUC	(95% CI)
The combined model	0.841	0.842	0.911	0.862-0.961	0.720	0.964	0.907	0.831-0.984
The radiomics model	0.807	0.73	0.858	0.793-0.923	0.84	0.786	0.851	0.740-0.963
The deep learning model	0.649	0.778	0.714	0.620-0.808	0.480	0.964	0.767	0.638-0.897
The clinical model	0.737	0.635	0.734	0.645-0.823	0.88	0.536	0.730	0.593-0.867

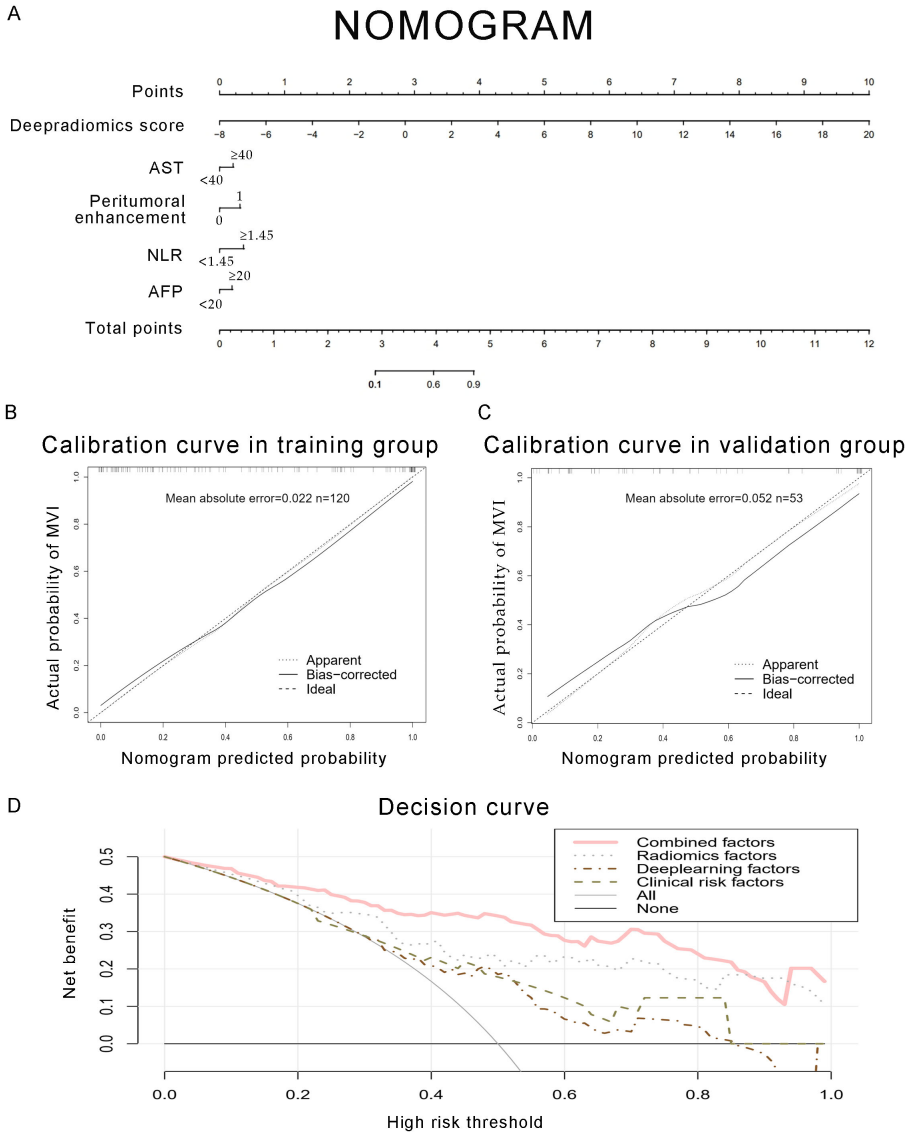


FIGURE 6
Development of a predictive nomogram for assessing the probabilities of MVI, along with calibration and decision curves. **(A)** A nomogram combining fusion radiomics-deep risk score, inflammation markers (NLR, AST), serum AFP, and radiological factor (peritumoral enhancement). **(B, C)** Calibration curves of the nomogram in the training and validation groups. The X-axis is the nomogram-predicted probability of MVI. The Y-axis is the actual probability of MVI. **(D)** Decision curve of the nomogram for predicting MVI. The pink line represents the expected net benefit per patient derived from the predictive nomogram.

elevated values of NLR, PLR, and SII with increased tumor invasiveness, lower differentiation grades, and diminished overall survival rates (24). In the cascade of cancer cell metastasis, neoplastic cells first infiltrate the adjacent microvasculature, encompassing the intratumoral vascular compartments, before propagating through these microvascular structures. A strong relationship between MVI and tumor recurrence and survival after transplant and surgical resection has been reported in recent retrospective analysis (25, 26). Consequently, this gives rise to the hypothesis that the presence of extratumoral MVI constitutes a significant risk factor for the recurrence of the tumor (27).

The NLR is an inflammatory marker that has been studied as a predictive indication of recurrence and survival in HCC patients (28). Higher NLR correlates with a heightened infiltration of inflammatory cells and an enhanced secretion of inflammatory cytokines, leading to a proliferation of neutrophil populations (29). According to the findings of our study, NLR was significantly associated with MVI, showing an increasing value at a higher incidence of MVI. This phenomenon could be attributed to the substantial release of neutrophils, which amplifies the potential for tumor progression and vascular invasion through the upregulation of vascular endothelial growth factor and additional pro-inflammatory cytokines (30).

The AST is an important enzyme in the liver, which is mainly present in the mitochondria of liver cells. In highly proliferative cancer tissues, the level of AST is more frequently activated. Ellen

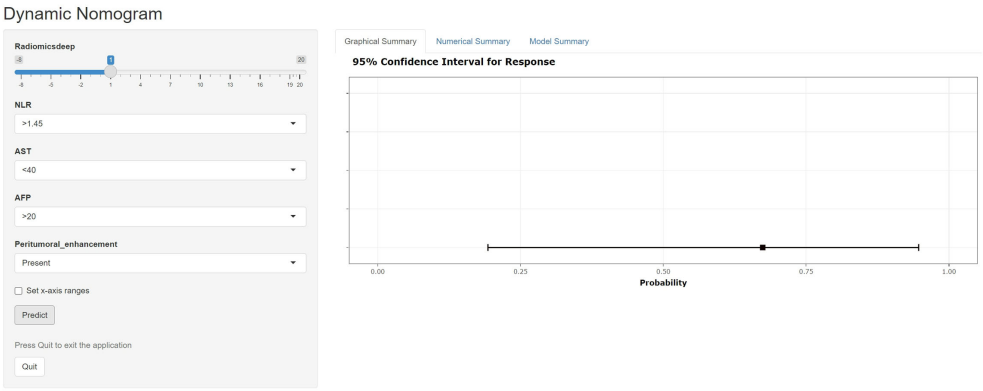


FIGURE 7
A web-dynamic nomogram for MVI prediction. The MVI probability and 95% confidence interval of an HBV-HCC patient with a radiomics score of 1, NLR over 1.45, AST below 40, serum AFP over 20, and signs of peritumoral enhancement on preoperative MRI were evaluated.

Friday has found that the increase in AST might be related to the tumor growth and progression. Our research has also demonstrated that the higher level of AST was associated with MVI in HCC patients.

As a conventional clinical biomarker in HCC, our observations corresponded with an established trend, wherein serum AFP paralleled the likelihood of MVI. These inflammatory indices would probably identify high-risk populations and improve our screening methods. Using these biomarkers, we developed a clinical model that combined NLR, AFP, and AST levels to identify HCC with MVI, achieving an accuracy with an AUC of 0.767. Our results are in line with those reported by Hidetoshi Nitta, indicating that the nomogram offers a reliable method for predicting extra tumoral MVI in patients undergoing hepatic resection or liver transplantation (27).

Prior studies have noted the importance of preoperative tumor images in tumor immune biology and immunotherapy response. Traditional radiomics leverages advanced computing tools to extract deeper and more granular data from imaging (31). As Marius E. Mayerhoefer mentioned, radiomics models based on large high-quality and well-curated data sets have a better performance, so we tried to use radiomics features to construct a predictive model (32). In the context of this retrospective analysis, we identified 15 radiomic features correlated with MVI in HCC from T2-weighted phase images and standard triphasic phase images. These features were subsequently utilized for the construction of a radiomics score. In these 15 texture parameters, GLRLM_Long RunEmphasis, NGTDM_Busyness, GLDM_LargeDependenceHighGrayLevelEmphasis and Firstorder_TotalEnergy were the most significant difference between the two groups in 4 phases, respectively. LargeDependenceHighGrayLevel Emphasis measures the joint distribution of large dependence with higher gray-level values, with a higher value indicating more homogeneity (33). This parameter in venous phase has strongest predictive performance. The increased heterogeneity observed in the MVI-positive group within MRI scans can be ascribed to the presence of a greater diversity of atypical vessels, a higher incidence of necrotic vessels stemming from rapid tumor growth, and a more heterogeneous internal structure of the tumors (34). The inflammatory reaction within tumor microenvironment causes the proliferation of abnormal blood vessels and the necrosis of the tumor tissue, resulting in the appearance of uneven tumor internal structure in images.

With the development of high-throughput computing technology and artificial intelligence, deep image features are obtained by neural convolutional network algorithm such as CNN. In contrast to conventional radiomics methods, the image features are directly extracted from the deep neural network, hence ensuring that the deep learning radiomics extraction technique remains error-free (35). After extracting MRI deep features with 3D ResNet and LASSO analysis, 7 deep features were selected as MVI risk factors to construct a deep radiomics prediction model. We developed models based on radiomics features and deep learning features for predicting MVI in training dataset, with AUCs of 0.817 and 0.714, respectively. And validated in validation dataset, the AUCs of the two models were 0.864 and 0.730 respectively. Two models performed well in terms of MVI prediction.

The literature has documented that radiomics features might serve as indicators of the tumor microenvironment in patients (36). The occurrence of MVI might be related to the local inflammation in the tumor microenvironment (37). The radiomics model has a good prediction effect on MVI, and we speculate that there is a correlation between radiomics and inflammatory indicators. Finally, we also proved that radiomics features were positively correlated with NLR, which reflects inflammation. Simultaneously, it was confirmed that the risk scores of the deep and radiomics models were positively connected with NLR. As both preoperative examinations, we infer that the clinical and MRI indices could reflect the pro/antitumorigenic inflammatory status in two different ways. Hence, it could conceivably be hypothesized that clinical characteristics plus MRI signatures can improve the predictive value for MVI. We constructed a combined model, the result is just as what we have supposed. Our study supported the findings of Wenjun Yao, which indicated that a combined model leveraging both clinical and radiomic signatures delivered superior predictive performance, as evidenced by a high AUC, and more effectively differentiated MVI when compared to models based on clinical or radiomic markers alone (38).

Our combined model achieved an AUC of 0.907, sensitivity of 72.0%, and specificity of 96.4%. These metrics were comparable to

those reported by Yang et al., who reported an AUC of 0.861, using a similar MRI-based approach (13). Notably, the AUCs of the clinical model and the radiomics model were similar to those obtained in their study, but the combined model yielded a higher AUC. This underscores the effectiveness of deep learning features in accurately identifying MVI. When compared to the study by Zhou et al., which employed 3D convolutional neural networks on contrast-enhanced MRI to predictive MVI in HCC, our study also utilized advanced feature extraction method (39). This led to an AUC improvement, highlighting the superiority of our approach in capturing significant features linked to MVI. Additionally, the interpretability of our model was enhanced through the analysis of the correlation between radiomic features and NLR, providing insights into the contribution of individual radiomic features. This transparency is a significant advancement over earlier studies, such as that by Mu He, where the model incorporating neutrophils lacks interpretability (40).

Our analysis extended to examining the performance of radiomics models in different imaging phases, revealing that the model integrating venous and delayed phase images yielded better predictive accuracy than the model based on arterial phase imaging. Hypervascularity during the arterial phase of enhancement and wash-out during the portal phase correspond to the Barcelona criteria for HCC (41). The high AUC of the model may be due to the enhancement patterns and "wash-out" characteristics typical of HCC seen in imaging studies (42). Even though deep learning model's prediction performance was lower than radiomics model, there was still a high positive correlation between some deep learning features and NLR. In the radiomics analysis of HBV-related HCC, which was closely related to inflammation, we speculated deep learning features were a good complement to traditional radiomics analysis. The AUCs of the combined model also corresponded to our thoughts, and the C-index and the decision curve both verified the good predictive performance of the combined model.

This study is subject to certain limitations, including its single-center design and retrospective nature. Furthermore, the sample size is relatively small, which may affect the generalizability of the findings. To develop and validate an accurate prediction model for microvascular invasion (MVI) grading, additional research involving larger populations is imperative.

5 Conclusion

In conclusion, the combined model achieves satisfactory preoperative prediction of MVI in HBV-related HCC. In other models, the radiomics model has good prediction performance, and deep learning features are a better complement for MVI prediction. NLR is positively correlated with MRI features. The nomogram based on clinical risk factors and MRI characteristics would help clinicians and patients make an individualized risk assessment of MVI.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Branch for Medical Research and Clinical Technology Application, Ethics Committee of the First Affiliated Hospital of Fujian Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YZ: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LC: Conceptualization, Data curation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. FD: Conceptualization, Methodology, Software, Writing – review & editing. WO: Conceptualization, Data curation, Methodology, Writing – review & editing. XZ: Conceptualization, Writing – review & editing. SW: Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Backbone Talents Training Project of Fujian Health Department (Grant No. 2020GGB033) and the Natural Science Foundation of Fujian Province (Grant No. 2020J01980).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1401095/full#supplementary-material

References

1. Mattiuzzi C, Lippi G. Current cancer epidemiology. *J Epidemiol Glob Health*. (2019) 9:217–22. doi: 10.2991/jegh.k.191008.001

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

3. Wu Q, Qin SK. Features and treatment options of Chinese hepatocellular carcinoma. *Chin Clin Oncol*. (2013) 2:38. doi: 10.3978/j.issn.2304-3865.2013.09.07

4. Tanaka M, Katayama F, Kato H, Tanaka H, Wang J, Qiao YL, et al. Hepatitis B and C virus infection and hepatocellular carcinoma in China: a review of epidemiology and control measures. *J Epidemiol*. (2011) 21:401–16. doi: 10.2188/jea.JE20100190

5. Couri T, Pillai A. Goals and targets for personalized therapy for HCC. *Hepatol Int*. (2019) 13:125–37. doi: 10.1007/s12072-018-9919-1

6. Bhatti ABH, Sheikh AAE, Mahmud US, Zeeshan S, Khan NY, Zia HH, et al. Clinical profile and treatment of hepatocellular carcinoma: A single-center experience. *South Asian J Cancer*. (2021) 10:76–80. doi: 10.1055/s-0041-1729451

7. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*. (2007) 132:2557–76. doi: 10.1053/j.gastro.2007.04.061

8. Weber A, Boege Y, Reisinger F, Heikenwalder M. Chronic liver inflammation and hepatocellular carcinoma: persistence matters. *Swiss Med Wkly*. (2011) 141:w13197. doi: 10.4414/smw

9. Zheng J, Seier K, Gonen M, Balachandran VP, Kingham TP, D'Angelica MI, et al. Utility of serum inflammatory markers for predicting microvascular invasion and survival for patients with hepatocellular carcinoma. *Ann Surg Oncol*. (2017) 24:3706–14. doi: 10.1245/s10434-017-6060-7

10. Li S, Zeng Q, Liang R, Long J, Liu Y, Xiao H, et al. Using systemic inflammatory markers to predict microvascular invasion before surgery in patients with hepatocellular carcinoma. *Front Surg*. (2022) 9:833779. doi: 10.3389/fsurg.2022.833779

11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

12. Ronot M, Kerbaol A, Rautou PE, Brancatelli G, Bedossa P, Cazals-Hatem D, et al. Acute extrahepatic infectious or inflammatory diseases are a cause of transient mosaic pattern on CT and MR imaging related to sinusoidal dilatation of the liver. *Eur Radiol*. (2016) 26:3094–101. doi: 10.1007/s00330-015-4124-2

13. Yang L, Gu D, Wei J, Yang C, Rao S, Wang W, et al. A radiomics nomogram for preoperative prediction of microvascular invasion in hepatocellular carcinoma. *Liver Cancer*. (2019) 8:373–86. doi: 10.1159/000494099

14. Yu Y, Song J, Zhang R, Liu Z, Li Q, Shi Y, et al. Preoperative neutrophil-to-lymphocyte ratio and tumor-related factors to predict microvascular invasion in patients with hepatocellular carcinoma. *Oncotarget*. (2017) 8:79722–30. doi: 10.18632/oncotarget.v8i45

15. Wu Y, Tu C, Shao C. The value of preoperative systemic immune-inflammation index in predicting vascular invasion of hepatocellular carcinoma: a meta-analysis. *Braz J Med Biol Res*. (2021) 54:e10273. doi: 10.1590/1414-431x202010273

16. Liu P, Tan XZ, Zhang T, Gu QB, Mao XH, Li YC, et al. Prediction of microvascular invasion in solitary hepatocellular carcinoma </= 5 cm based on computed tomography radiomics. *World J Gastroenterol*. (2021) 27:2015–24. doi: 10.3748/wjg.v27.i17.2015

17. Marin D, Di Martino M, Guerrisi A, De Filippis G, Rossi M, Ginanni Corradini S, et al. Hepatocellular carcinoma in patients with cirrhosis: qualitative comparison of gadobenate dimeglumine-enhanced MR imaging and multiphasic 64-section CT. *Radiology*. (2009) 251:85–95. doi: 10.1148/radiol.2511080400

18. Nouri-Vaskeh M, Mirza-Aghazadeh-Attari M, Pashazadeh F, Azami-Aghdash S, Alizadeh H, Pouya P, et al. Prognostic impact of monocyte to lymphocyte ratio in clinical outcome of patients with hepatocellular carcinoma: A systematic review and meta-analysis. *Galen Med J*. (2020) 9:e1948. doi: 10.31661/gmj.v9i0.1948

19. Timperi E, Barnaba V. Viral hepatitides, inflammation and tumour microenvironment. *Adv Exp Med Biol*. (2020) 1263:25–43. doi: 10.1007/978-3-030-44518-8_3

20. Beudeker BJB, Groothuismink ZMA, van der Eijk AA, Debes JD, Boonstra A. Circulating cytokines reflect the etiology-specific immune environment in cirrhosis and HCC. *Cancers (Basel)*. (2022) 14:4900. doi: 10.3390/cancers14194900

21. Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, et al. Recent advances and applications of deep learning methods in materials science. *NPJ Comput Materials*. (2022) 8. doi: 10.1038/s41524-022-00734-6

22. Kim M, Yun J, Cho Y, Shin K, Jang R, Bae HJ, et al. Deep learning in medical imaging. *Neurospine*. (2019) 16:657–68. doi: 10.14245/ns.1938396.198

23. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep*. (2017) 7:5467. doi: 10.1038/s41598-017-05848-2

24. Hernandez-Ainsa M, Velamazan R, Lanas A, Carrera-Lasfuentes P, Piazuelo E. Blood-cell-based inflammatory markers as a useful tool for early diagnosis in colorectal cancer. *Front Med (Lausanne)*. (2022) 9:843074. doi: 10.3389/fmed.2022.843074

25. Famularo S, Piardi T, Molfino S, Di Martino M, Ferrari C, Ielpo B, et al. Factors affecting local and intra hepatic distant recurrence after surgery for hcc: an alternative perspective on microvascular invasion and satellitosis - A western european multicentre study. *J Gastrointest Surg*. (2021) 25:104–11. doi: 10.1007/s11605-019-04503-7

26. Rodriguez-Peralvarez M, Luong TV, Andreana L, Meyer T, Dhillon AP, Burroughs AK. A systematic review of microvascular invasion in hepatocellular carcinoma: diagnostic and prognostic variability. *Ann Surg Oncol*. (2013) 20:325–39. doi: 10.1245/s10434-012-2513-1

27. Nitta H, Allard MA, Sebagh M, Ciacio O, Pittau G, Vibert E, et al. Prognostic value and prediction of extratumoral microvascular invasion for hepatocellular carcinoma. *Ann Surg Oncol*. (2019) 26:2568–76. doi: 10.1245/s10434-019-07365-0

28. Lin S, Hu S, Ran Y, Wu F. Neutrophil-to-lymphocyte ratio predicts prognosis of patients with hepatocellular carcinoma: a systematic review and meta-analysis. *Trans Cancer Res*. (2021) 10:1667–78. doi: 10.21037/tcr

29. Cedrés S, Torrejon D, Martínez A, Martinez P, Navarro A, Zamora E, et al. Neutrophil to lymphocyte ratio (NLR) as an indicator of poor prognosis in stage IV non-small cell lung cancer. *Clin Trans Oncol*. (2012) 14:864–9. doi: 10.1007/s12094-012-0872-5

30. Wu L, Saxena S, Singh RK. Neutrophils in the tumor microenvironment. *Adv Exp Med Biol*. (2020) 1224:1–20. doi: 10.1007/978-3-030-35723-8_1

31. Harding-Theobald E, Louissaint J, Maraj B, Cuaresma E, Townsend W, Mendiratta-Lala M, et al. Systematic review: radiomics for the diagnosis and prognosis of hepatocellular carcinoma. *Alimentary Pharmacol Ther*. (2021) 54:890–901. doi: 10.1111/apt.16563

32. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med*. (2020) 61:488–95. doi: 10.2967/jnumed.118.222893

33. Mayerhoefer ME, Szomolanyi P, Jirak D, Berg A, Materka A, Dirisamer A, et al. Effects of magnetic resonance image interpolation on the results of texture-based pattern classification: a phantom study. *Invest Radiol*. (2009) 44:405–11. doi: 10.1097/RLI.0b013e3181a50a66

34. Li Y, Xu X, Weng S, Yan C, Chen J, Ye R. CT image-based texture analysis to predict microvascular invasion in primary hepatocellular carcinoma. *J Digit Imaging*. (2020) 33:1365–75. doi: 10.1007/s10278-020-00386-2

35. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and deep learning methods for radiomics. *Med Phys*. (2020) 47:e185–202. doi: 10.1002/mp.13678

36. Kang W, Qiu X, Luo Y, Luo J, Liu Y, Xi J, et al. Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. *J Trans Med*. (2023) 21:598. doi: 10.1186/s12967-023-04437-4

37. Critelli R, Milosa F, Faillaci F, Condello R, Turola E, Marzi L, et al. Microenvironment inflammatory infiltrate drives growth speed and outcome of hepatocellular carcinoma: a prospective clinical study. *Cell Death Disease*. (2017) 8:e3017–e. doi: 10.1038/cddis.2017.395

38. Yao W, Yang S, Ge Y, Fan W, Xiang L, Wan Y, et al. Computed tomography radiomics-based prediction of microvascular invasion in hepatocellular carcinoma. *Front Med (Lausanne)*. (2022) 9:819670. doi: 10.3389/fmed.2022.819670

39. Zhou W, Jian W, Cen X, Zhang L, Guo H, Liu Z, et al. Prediction of microvascular invasion of hepatocellular carcinoma based on contrast-enhanced MR and 3D convolutional neural networks. *Front Oncol*. (2021) 11:588010. doi: 10.3389/fonc.2021.588010

40. He M, Zhang P, Ma X, He B, Fang C, Jia F. Radiomic feature-based predictive model for microvascular invasion in patients with hepatocellular carcinoma. *Front Oncol*. (2020) 10:574228. doi: 10.3389/fonc.2020.574228

41. Willatt J, Ruma JA, Azar SF, Dasika NL, Syed F. Imaging of hepatocellular carcinoma and image guided therapies - how we do it. *Cancer Imaging*. (2017) 17:9. doi: 10.1186/s40644-017-0110-z

42. Fronda M, Doriguzzi Breatta A, Gatti M, Calandri M, Maglia C, Bergamasco L, et al. Quantitative assessment of HCC wash-out on CT is a predictor of early complete response to TACE. *Eur Radiol*. (2021) 31:6578–88. doi: 10.1007/s00330-021-07792-2

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 04 October 2024
DOI 10.3389/fonc.2024.1446736

Check for updates

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Lorenzo Faggioni,
University of Pisa, Italy
Murat Beyhan,
Tokat Gaziosmanpaşa University, Türkiye

*CORRESPONDENCE
Chuanhai Jia
csjiachuanhai@163.com
Shengyu Wang
rjhnwsy@163.com

†These authors have contributed equally to this work

RECEIVED 10 June 2024
ACCEPTED 18 September 2024
PUBLISHED 04 October 2024

CITATION
Sun J, Kuai X, Huang D, Ji X, Jia C and Wang S (2024) Assessment of synthetic MRI to distinguish Warthin's tumor from pleomorphic adenoma in the parotid gland: comparison of two methods of positioning the region of interest for synthetic relaxometry measurement. *Front. Oncol.* 14:1446736. doi: 10.3389/fonc.2024.1446736

COPYRIGHT
© 2024 Sun, Kuai, Huang, Ji, Jia 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.

Assessment of synthetic MRI to distinguish Warthin's tumor from pleomorphic adenoma in the parotid gland: comparison of two methods of positioning the region of interest for synthetic relaxometry measurement

Jiabin Sun[1†], Xinping Kuai[2†], Dawei Huang[3], Xinghua Ji[2], Chuanhai Jia[1*] and Shengyu Wang[4*]

[1]Department of Radiology, Changshu No.2 People's Hospital, the Fifth Affiliated Clinical Medical College of Yangzhou University, Changshu, Jiangsu, China, [2]Department of Radiology, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [3]Department of Stomatology, Changshu No.2 People's Hospital, the Fifth Affiliated Clinical Medical College of Yangzhou University, Changshu, Jiangsu, China, [4]Department of Radiology, Ruijin Hospital, shanghai Jiao Tong University School of Medicine, Jiading, Shanghai, China

Purpose: To assess the diagnostic potential of the synthetic MRI (SyMRI) for differentiating Warthin's tumors (WT) from pleomorphic adenomas (PA).

Materials and methods: Forty-nine individuals with parotid gland tumors (PA, n = 23; WT, n = 26) were recruited. Using two distinct regions of interest (ROI), SyMRI quantitative parameters of lesions were calculated, including mean and standard deviation (T1, T2, PD, T1sd, T2sd, and PDsd). Meanwhile, T1ratio, T2ratio, and PDratio (lesion/masseter muscle) were calculated based on the mean SyMRI quantitative parameters of masseter muscle (T1, T2, PD). Using the independent samples *t* test, we compared PA and WT parameters, while comparing the areas under the curve (AUC) using the DeLong's test. A multi-parameter SyMRI model was constructed using logistic regression analysis.

Results: In PA, the T1, T1sd, T2, PD, T1ratio, T2ratio, and PDratio derived from full and partial lesion ROIs were significantly higher than in WT. According to the receiver operating curve analysis, the AUC of the quantitative parameters derived from full-lesion and partial-lesion ROIs ranged from 0.722 to 0.983 for differentiating PA from WT. T1 values derived from partial-lesion ROI delineation demonstrated the best diagnostic performance among all single parameters, achieving an AUC of 0.983. Using 1322 ms as a cutoff value, the sensitivity, specificity, and accuracy were 88.46%, 100% and 93.88%, respectively.

Conclusion: The SyMRI-derived quantitative parameters demonstrated excellent performance for discriminating PA from WT in the parotid gland.

KEYWORDS

parotid neoplasms, pleomorphic adenomas, Warthin's tumor, magnetic resonance imaging, synthetic MRI, magnetic resonance image compilation

Introduction

Pleomorphic adenomas (PA) and Warthin's tumors (WT) are the most common benign tumors, accounting for more than 70% of the neoplasms in the parotid gland (1, 2). Although PA and WT are benign tumors, they differ in biological behavior and prognosis (2, 3). Unlike WT, PA presents a 1.8% to 6.2% risk of malignant transformation (4, 5). For WT, due to the intact fibrous capsule, enucleation of the tumor alone is sufficient, while for PA, extensive field removal of the tissue in the parotid bed is recommended due to the risk of recurrence or disruption to the capsule (6, 7). Therefore, an accurate preoperative diagnosis of PA and WT is critically necessary for determining the optimal treatment strategy.

Various imaging modalities play essential roles in the diagnosis and management of parotid gland tumors (3, 7). Among these methods, conventional magnetic resonance imaging (MRI) provides information regarding tumor size, position, and relationship to adjacent structures. However, due to variability in the histological component of PA, conventional MRI is insufficient to distinguish PA from WT (8). Quantitative MRI techniques such as diffusion-weighted imaging and dynamic contrast-enhanced MRI (DCE-MRI) can complement physiological and functional information to distinguish subtypes of parotid neoplasms (9, 10). However, the apparent diffusion coefficient (ADC) value of WT overlaps with that of malignant tumors, and ADC alone cannot accurately distinguish benign and malignant parotid tumors (4, 11). In addition, in the case of DCE-MRI, there is a disadvantage due to the risk of serious adverse events and overlapping time-intensity curve pattern between PA and WT (12). Recently, in order to overcome the limited discriminative performance of conventional MRI and other imaging techniques, radiomics has been applied in differentiating various types of parotid tumors (13). Radiomics-based techniques, however, have a number of drawbacks, including limited generalizability of quantitative features due to low sample sizes, single-center studies, and different imaging equipment and radiomics models. Accordingly, novel approaches are required to improve the ability of differential diagnosis of PA from WT.

Synthetic MR imaging, also called magnetic resonance image compilation (MAGiC), can quantitatively measure multiple physical properties based on longitudinal relaxation time (T1), transverse relaxation time (T2) and proton density (PD) parameters (14). Unlike conventional and functional MRI techniques, which are influenced by scan parameters and biophysical model for data postprocessing, the quantitative SyMRI technique is inherently independent of MRI methods and parameters (15). In previous studies, the SyMRI has been applied in the differential diagnosis of breast tumors, prostate lesions and retropharyngeal lymph node (16–18). To our knowledge, there have been no studies exploring SyMRI's clinical utility for distinguishing PA from WT. Meanwhile, various regions of interest (ROI) delineations have been used in previous studies, but a standardized approach for ROI placement methods is still lacking (19–21). Whole-volume ROI analysis rather than single-slice ROI delineation may accurately reflect tumor heterogeneity (22). In clinical practice, however, it is time-consuming and difficult to perform such analysis. Furthermore, no studies have evaluated the diagnostic performance of SyMRI in distinguishing between PA and WT using different ROI delineations. Hence, this study aimed to assess the clinical feasibility and diagnostic potential of SyMRI for differentiating PA from WT using two different ROI delineation methods.

Materials and methods

Participants

All the subjects who participated in the study completed the informed consent form and completed it uniformly. The project was approved by the Ethics Committee of Changshu No.2 People's Hospital, the Fifth Affiliated Clinical Medical College of Yangzhou University. From January 2021 to December 2023, the SyMRI examination was performed on 71 consecutive individuals with suspected parotid gland tumors using a 3.0T MRI scanner. Exclusion criteria were as follows (4): (1) imaging artifacts interfering with interpretation, (2) prior history of head or neck disease, (3) small parotid gland tumors less than 10 mm in size, (4) the parotid tumor received biopsy, surgery, chemotherapy or radiotherapy and (5) postoperative histopathologically confirmed non-PA or non-WT. Subsequently, three patients with tiny parotid gland tumors under 10 mm in diameter, four patients with severe dental metal artefacts, and 15 patients with histopathologically proven non-PA or non-WT (one carcinoma ex pleomorphic adenoma, two lymphomas, three mucoepidermoid carcinomas, one adenoid cystic carcinomas, one acinar cell carcinomas, one squamous cell carcinomas, two granulomatous lymphadenitis, two myoepithelioma, one basal cell adenoma, and one benign lymphoepithelial lesion) were excluded from this study. Finally, there were 49 patients enrolled in this study, including 23 patients with PAs and 26 patients with WTs. All patients underwent surgical resection, and complete tumor specimens were histopathologically examined within 1 month after their MR examinations.

MRI data acquisition

All examinations were performed using a GE Signa Architect 3.0T MRI scanner with a 28-channel head-neck coil. Before contrast was administered, SyMRI sequence in the axial direction was performed. The parameters of the SyMRI were as follows: repetition time (TR) of 4000 ms; echo time (TE) of 19.2/96.3 ms; field of view (FOV) of 24 × 24 cm; number of excitations (NEX): 1; Echo Train Length (ETL): 16; Bandwidth:31.25 kHz; matrix: 320 × 256; number of slices:20; slice thickness: 3 mm; slice gap: 1 mm; acquisition time: 4 min.

Image analysis

The SyMRI quantitative parameters were post-processed by two radiologists blinded to pathology (rater 1 and rater 2 with more than five and ten years of experience in head and neck imaging, respectively). The ROIs were manually delineated from synthetic T2-weighted images by referencing contrast-enhanced T1WI images.

Two ROI delineation methods were used: (1) solid portion within the largest section of the lesion, avoiding necrotic, cystic, hemorrhagic, or apparent vessel components (partial lesion); (2) the largest area of the largest section (full lesion) (4, 18). The mean and standard deviation (sd) for longitudinal relaxation time (T1, T1sd), transverse relaxation time (T2, T2sd), and proton density (PD, PDsd) were calculated for WT and PA. Meanwhile, we delineated a ROI on the unilateral masseter muscle at the same level of the lesion and obtained T1, T2, and PD values (23). T1ratio, T2ratio, and PDratio were calculated as follows: T1ratio = T1 value of lesion/T1 value of masseter muscle; T2ratio = T2 value of lesion/T2 value of masseter muscle; PDratio = PD value of lesion/PD value of masseter muscle (23). Tumor volumes were calculated using the following Eq: $V=\pi\times L\times W^2/6$, where L is length and W is width of the largest section (24).

Statistical analysis

Two-way mixed, absolute agreement, single measure intraclass correlation coefficients (ICC) were calculated to assess the reliability of SyMRI quantitative parameters between two raters (25). Normal distribution and homogeneity of variance were assessed using the Kolmogorov–Smirnov test and Levene's test, respectively.PA and WT quantitative parameters were compared through independent sample *t*-tests and categorical variables through chi-squared tests. We plotted receiver operating characteristic curves (ROCs), calculated area under the curve (AUC), optimal threshold values, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value. Multiparameter diagnostic models were constructed using binary logistic regression analysis (forward stepwise selection model) that included significant parameters with $P < 0.05$. To compare the differences between the ROC curves, DeLong's test was performed. All statistical analyses were performed with R 4.1.3 and MedCalc software (version 15.2.2) using *P*-values below 0.05.

Results

As shown in Table 1, 49 patients were enrolled in this study, including 23 patients with PA and 26 with WT. No statistical difference was found in tumor volume between the two entities ($P > 0.05$). However, there were statistically significant differences in age and gender ratio between PA and WT ($P < 0.05$). In this study, WT showed a predilection for older males.

Assessment of SyMRI inter-rater reliability

The ICCs for the quantitative parameters of SyMRI are all above 0.85, indicating excellent inter-rater agreement (Table 2).

SyMRI quantitative parameters for PA and WT

Table 3 summarizes SyMRI's quantitative parameters for PA and WT. The T1, T1sd, T2, PD, T1ratio, T2ratio, and PDratio of PA derived from full-lesion ROI were significantly higher than those of WT ($P < 0.05$). The T1, T1sd, T2, T2sd, PD, PDsd, T1ratio, T2ratio, and PDratio of PA derived from partial-lesion ROI were significantly higher than those of WT ($P < 0.05$) (Figures 1, 2). In terms of T2sd and PDsd derived from full-lesion ROIs, there were no significant differences between PA and WT ($P > 0.05$). The T1, T2, and PD values measured on the masseter muscle did not differ significantly between the two types of lesions ($P > 0.05$).

Differentiation between PA and WT patients using SyMRI

For differentiating PA from WT, the AUCs for these quantitative parameters range from 0.722 to 0.983 based on the ROC analysis. The T1 values obtained from the partial-lesion ROI delineation showed the highest AUC of 0.983 among all single parameters (Table 4; Figures 3A, B). Using a cutoff value of 1322 ms to discriminate between PA and WT subjects, the T1 values obtained from partial-lesion ROI delineation achieved 88.46% sensitivity, 100% specificity, and 93.88% accuracy, respectively.

The AUCs for T1, T2, PD, T1ratio, T2ratio and PDratio values did not differ between the partial and full lesion ROI delineation methods (all $P > 0.05$). Combined SyMRI-derived multiple parameters resulted in similar AUCs when delineating ROIs with these two different approaches ($P > 0.05$) (Figure 3C). The AUCs for T1sd, T2sd, and PDsd values differed significantly between partial and full lesion ROI delineations (all $P < 0.05$).

Discussion

In this study, based on full and partial lesion ROIs, we observed that T1, T1sd, T2, PD, T1ratio, T2ratio, and PDratio were higher in PA than in WT. Moreover, the T2sd and PDsd of PA derived from partial-lesion ROI were significantly higher than those of WT. No significant differences, however, were found in the T2sd and PDsd derived from full-lesion ROI between the PA and WT groups. We also demonstrated the ability of SyMRI parameters to discriminate between subjects with PA and those with WT. As far as we know, this study is the first to investigate the quantitative parameters of SyMRI and assess its diagnostic power for PA and WT individuals.

TABLE 1 Characteristics of patients with PA and WT.

Variable	PA	WT	*P* value
Age (years)	48.74 ± 13.88	67.54 ± 7.44	0.005
Gender			
male	12	23	0.005
female	11	3	
Tumor volume (cm^3)	5.72 ± 8.39	6.89 ± 4.93	0.294

Please note that PA stands for pleomorphic adenoma, and WT stands for Warthin's tumor. Data are expressed as mean ± SD, or number.

TABLE 2 Assessment of SyMRI inter-rater reliability.

Parameter	Inter-rater	
	ICC	95%CI
Full-lesion ROI		
T1	0.993	0.988-0.996
T1sd	0.987	0.977-0.993
T2	0.992	0.986-0.996
T2sd	0.997	0.994-0.998
PD	0.988	0.980-0.993
PDsd	0.980	0.965-0.989
T1ratio	0.980	0.964-0.989
T2ratio	0.990	0.982-0.994
PDratio	0.969	0.937-0.984
Partial-lesion ROI		
T1	0.996	0.992-0.998
T1sd	0.988	0.978-0.993
T2	0.998	0.997-0.999
T2sd	0.993	0.988-0.996
PD	0.980	0.964-0.988
PDsd	0.957	0.925-0.975
T1ratio	0.982	0.968-0.990
T2ratio	0.976	0.959-0.987
PDratio	0.960	0.928-0.977
Masseter muscle		
T1	0.914	0.853-0.950
T2	0.887	0.807-0.935
PD	0.962	0.918-0.981

Please note that T1 stands for longitudinal relaxation time, T2 stands for transverse relaxation time, PD stands for proton density, T1sd stands for standard deviation of the T1 measurements, T2sd stands for standard deviation of the T2 measurements, PDsd stands for standard deviation of the PD measurements, ICC stands for intraclass correlation, and 95%CI stands for 95% Confidence Interval.

TABLE 3 Comparison of quantitative parameters between PA and WT based on full and partial lesion ROIs.

Parameters	PA	WT	*P* value
Full-lesion ROI			
T1	2032.70 ± 486.06	1126.46 ± 229.11	**0.000***
T1sd	573.83 ± 230.98	339.62 ± 244.51	**0.001***
T2	109.87 ± 43.51	72.19 ± 12.44	**0.000***
T2sd	23.48 ± 31.00	11.88 ± 9.13	0.075
PD	88.66 ± 7.62	80.48 ± 7.38	**0.000***
PDsd	10.73 ± 5.55	8.88 ± 3.58	0.166
T1ratio	2.04 ± 0.49	1.05 ± 0.23	**0.000***
T2ratio	2.34 ± 0.87	1.45 ± 0.27	**0.000***
PDratio	1.27 ± 0.22	1.11 ± 0.30	**0.031***
Partial-lesion ROI			
T1	2011.91 ± 421.23	1135.61 ± 191.36	**0.000***
T1sd	521.65 ± 173.69	231.96 ± 168.48	**0.000***
T2	102.74 ± 29.42	70.23 ± 7.89	**0.000***
T2sd	16.65 ± 15.47	7.04 ± 2.58	**0.003***
PD	87.85 ± 7.35	78.67 ± 4.61	**0.000***
PDsd	10.54 ± 5.56	7.23 ± 2.40	**0.013***
T1ratio	2.02 ± 0.47	1.07 ± 0.26	**0.000***
T2ratio	2.20 ± 0.64	1.42 ± 0.22	**0.000***
PDratio	1.26 ± 0.21	1.07 ± 0.24	**0.006***
Masseter muscle			
T1	1003.65 ± 112.81	1088.12 ± 198.82	0.071
T2	46.87 ± 2.90	50.35 ± 7.77	0.050
PD	71.13 ± 12.21	75.89 ± 13.66	0.207

ROI, region of interest; T1, longitudinal relaxation time; T1sd, standard deviation of the T1 measurements; T2, transverse relaxation time; T2sd, standard deviation of the T2 measurements; PD, proton density; PDsd, standard deviation of the PD measurements; T1ratio = T1 value of lesion/T1 value of masseter muscle; T2ratio = T2 value of lesion/T2 value of masseter muscle; PDratio = PD value of lesion/PD value of masseter muscle. *Statistical difference. Bold value indicates statistical significance.

A previous study using T1 mapping in parotid gland tumors has demonstrated that the mean T1 relaxation time is higher in PA than in WT (26). Similarly, we revealed that the T1 value of PA was significantly higher than that of WT using SyMRI, with either partial or full lesion ROI delineation approaches. The T1 values obtained from the partial-lesion ROI delineation showed the highest AUC of 0.983 among all single parameters. Using a cutoff value of 1322 ms to discriminate between PA and WT subjects, the T1 values obtained from partial-lesion ROI delineation achieved 88.46% sensitivity, 100% specificity, and 93.88% accuracy, respectively. T1 value is influenced by water content, binding with macromolecules (water mobility), and cell content (27, 28). The higher T1 values of PA were attributed to the tiny nuclei of the epithelial and mesenchymal cells, low cell and neovascular densities, abundant stroma, large extracellular spaces, and excessive free water content in these tumors (29). The lower T1 values of WT may be attributed to the abundance of lymphocytes and lymph stromal cells, high microvascular density and cellular-stromal grade, low free water content, relatively small extracellular space, and limited diffusion of water molecules (29). This suggests that T1 values may be valuable to distinguish PA from WT. To our knowledge, there are no studies published in the literature comparing PA and WT based on the standard deviation of T1 measurement (T1sd) value and T1 value of lesion/T1 value of masseter muscle (T1ratio). In this study, PA had a significantly higher T1sd than WT, suggesting that PA has more T1 variability over image pixels compared to WT. At the cellular level, PA is characterized by morphological diversity that includes both epithelial and mesenchymal components, which

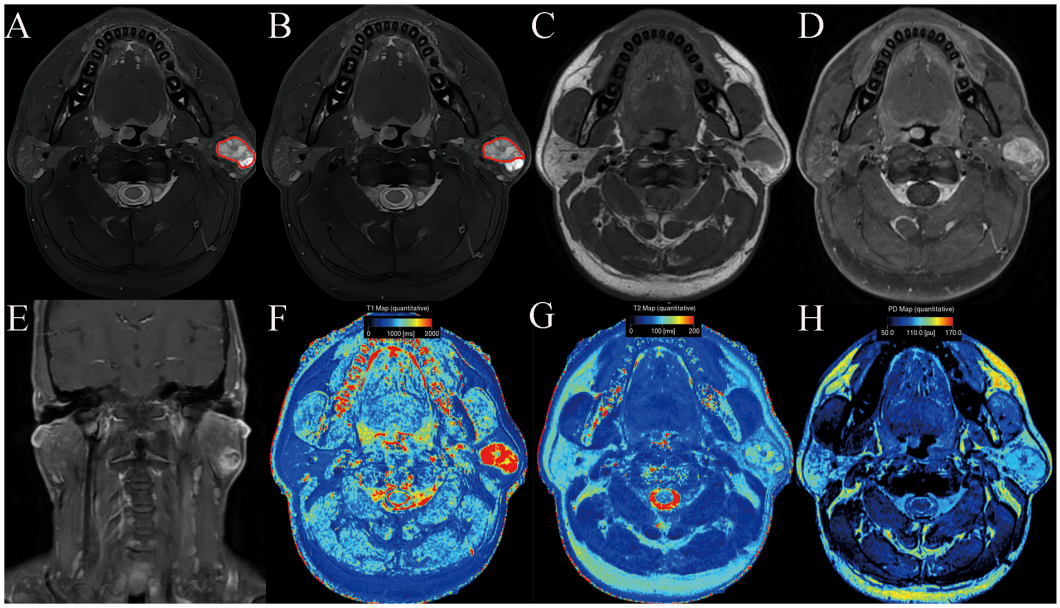


FIGURE 1
Pleomorphic adenoma of the parotid gland in a 42-year-old male subject. The ROI delineation methods included the largest area of the largest section (full lesion, **A**) and the solid portion within the largest section (partial lesion, **B**). On conventional MRI, this mass showed hyperintense on T2WI **(A, B)**, a well-defined border **(C)**, and marked enhancement on contrast enhanced T1WI **(D, E)**. In comparison with WT, it demonstrated significantly higher SyMRI-derived quantitative parameters for partial-lesion ROI (T1: 2077 ms; T1sd: 529 ms; T2: 90 ms; T2sd: 13 ms; PD:99 pu; PDsd: 8.4 pu) on axial T1 mapping **(F)**, T2 mapping **(G)**, and PD mapping **(H)**.

may lead to higher T1 variability within ROIs. Lastly, we found that T1ratios were significantly higher in PA than in WT, yielding an AUC of 0.973 for full-lesion ROI, and 0.970 for partial-lesion ROI. Therefore, the T1ratio value may also be a valuable quantitative parameter for differentiating PA from WT.

In different tissues, T2 values (quantification of transverse relaxation time) have been identified as excellent and reproducible biomarkers of water content (especially free water molecules). A previous study using T2 mapping in parotid gland tumors has demonstrated that the mean T2 relaxation time is higher in PA than in WT (30). Our previous study found that the solid components of PA mostly showed hyperintense on T2-weighted imaging compared with WT (4). Using SyMRI, either partial or full lesion ROI delineation, we found that PA had a higher T2 value than WT. The similarities in these findings may be explained by the pathological features of PA and WT. This suggests that T2 values may be valuable to distinguish PA from WT. To our knowledge, there are no studies published in the literature comparing PA and WT based on the standard deviation of T2 measurement (T2sd) value and T2 value of lesion/T2 value of masseter muscle (T2ratio). In this study, we found that the T2sd of PA derived from partial-lesion ROI was significantly higher than that of WT. However, no significant difference between the PA and WT groups was confirmed in the T2sd derived from the full lesion ROI. Thus, the T2sd derived from partial-lesion ROI and T2ratio value may be valuable quantitative parameters for differentiating PA from WT.

The PD value, as another magnetic property of the tissue, primarily reflects the water content of tissue (31). A previous study found that the PD value could distinguish benign from malignant retropharyngeal lymph node (18). In our study, we found that PA had higher PD and PDratio values compared to WT. In addition, we found that the PDsd of PA derived from partial-lesion ROI was significantly higher than that of WT. However, no significant difference between the PA and WT groups was confirmed in the PDsd of PA derived from full-lesion ROI. Thus, the PD value, PDsd derived from partial-lesion ROI and PDratio value may be also valuable quantitative parameters for differentiating PA from WT.

In our study, two different ROI approaches, partial or full-lesion ROI, were conducted in parameter measures. No significant differences between partial- and full-lesion ROI delineation approaches were observed in the AUCs of T1, T2, PD, T1ratio, T2ratio, and PDratio values. Meanwhile, combined SyMRI-derived multiple parameters showed similar AUCs for two different ROI delineation approaches. Thus, different types of ROI selection schemes have insignificant effect on the quantitative parameters of SyMRI for distinguishing PA from WT.

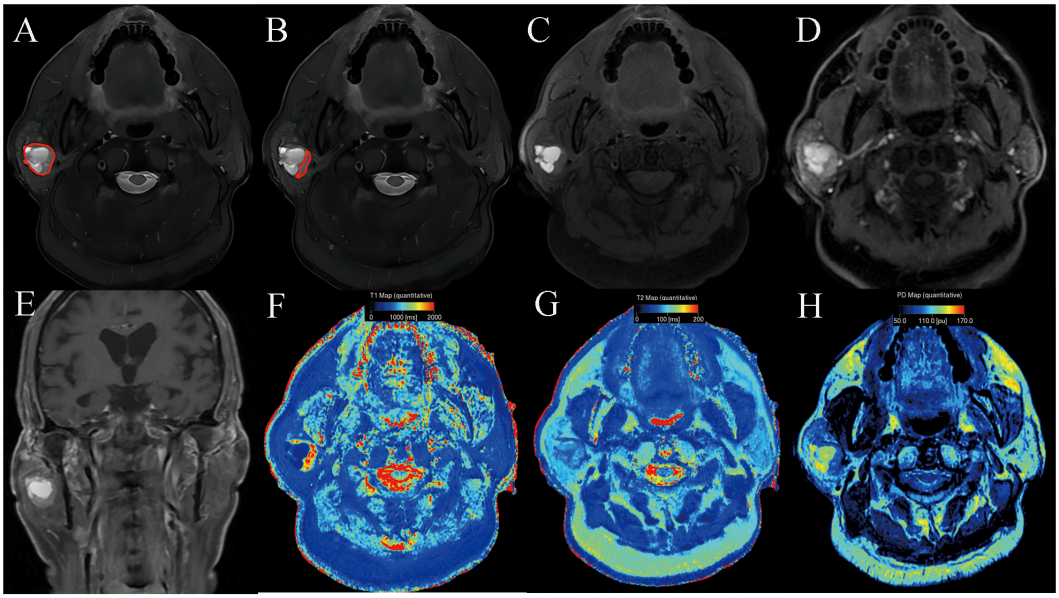


FIGURE 2
Warthin's tumor of the parotid gland in a 77-year-old male subject. The ROI delineation methods included the largest area of the largest section (full lesion, **A**) and the solid portion within the largest section (partial lesion, **B**). On conventional MRI, the solid component shows moderate signal and cyst/necrosis shows hyperintense on T2WI **(A, B)**, a well-defined border **(C)**, and moderate enhancement on contrast enhanced T1WI **(D, E)**. In comparison with PA, it demonstrated significantly lower SyMRI-derived quantitative parameters for partial-lesion ROI (T1: 1723 ms; T1sd: 269 ms; T2: 79 ms; T2sd: 7 ms; PD: 86.4 pu; PDsd: 6.1 pu) on axial T1 mapping **(F)**, T2 mapping **(G)**, and PD mapping **(H)**.

TABLE 4 Diagnostic capability of quantitative parameters calculated from full-lesion and partial-lesion ROIs for differentiating PA from WT.

	AUC (95% CI)	Cutoff value	Youden index	Sen (%)	Spe (%)	PPV (%)	NPV (%)	Acc (%)
Full-lesion ROI								
T1	0.975 (0.935-1.000)	1378.5	0.923	92.31	100	100	92.00	95.92
T1sd	0.804 (0.677-0.930)	250	0.538	53.84	100	100	65.71	75.51
T2	0.911 (0.830-0.991)	78.5	0.725	76.92	95.65	95.24	78.57	85.71
PD	0.828 (0.697-0.959)	84.1	0.672	84.61	82.61	84.61	82.61	83.67
T1ratio	0.973 (0.926-1.000)	1.38	0.918	96.15	95.65	96.15	95.65	95.92
T2ratio	0.932 (0.866-0.999)	1.80	0.749	92.31	82.61	85.71	90.47	87.75
PDratio	0.753 (0.611-0.894)	1.07	0.523	65.38	86.96	85.00	68.96	75.51
Combined (SyMRI)	0.977 (0.930-1.000)	0.24	0.961	96.15	100	100	95.83	97.96
Partial-lesion ROI								
T1	0.983 (0.958-1.000)	1322	0.885	88.46	100	100	88.46	93.88
T1sd	0.943 (0.864-1.000)	301.5	0.841	88.46	95.65	95.83	88.00	91.84

(Continued)

TABLE 4 Continued

	AUC (95% CI)	Cutoff value	Youden index	Sen (%)	Spe (%)	PPV (%)	NPV (%)	Acc (%)
Partial-lesion ROI								
T2	0.958 (0.909-1.000)	80.5	0.831	96.15	86.96	89.28	95.24	91.84
T2sd	0.902 (0.807-0.998)	8.5	0.759	84.61	91.30	91.67	84.00	87.75
PD	0.865 (0.747-0.982)	83.75	0.749	92.31	82.61	85.71	90.47	87.75
PDsd	0.722 (0.574-0.871)	8.25	0.416	80.77	60.87	70.00	73.68	71.43
T1ratio	0.970 (0.919-1.000)	1.42	0.918	96.15	95.65	96.15	95.65	95.92
T2ratio	0.963 (0.921-1.000)	1.73	0.793	92.31	86.96	88.89	90.91	89.80
PDratio	0.769 (0.632-0.906)	1.08	0.518	69.23	82.61	81.82	70.37	75.51
Combined (SyMRI)	0.990 (0.971-1.000)	0.26	0.923	92.31	100	100	92.00	95.92

ROI, region of interest; Sen, Sensitivity; Spe, Specificity; PPV, positive predictive value; NPC, negative predictive value; Acc, accuracy; T1, longitudinal relaxation time; T1sd, standard deviation of the T1 measurements; T2, transverse relaxation time; T2sd, standard deviation of the T2 measurements; PD, proton density; PDsd, standard deviation of the PD measurements; T1ratio = T1 value of lesion/T1 value of masseter muscle; T2ratio = T2 value of lesion/T2 value of masseter muscle; PDratio = PD value of lesion/PD value of masseter muscle. Combined (SyMRI) derived from full-lesion ROI: Logit (p = PA) = 6.74×T1ratio+2.37×T2ratio-13.89; Combined (SyMRI) derived from partial-lesion ROI: Logit (p = PA) = 7.68× T1ratio+0.40×T2sd-15.38.

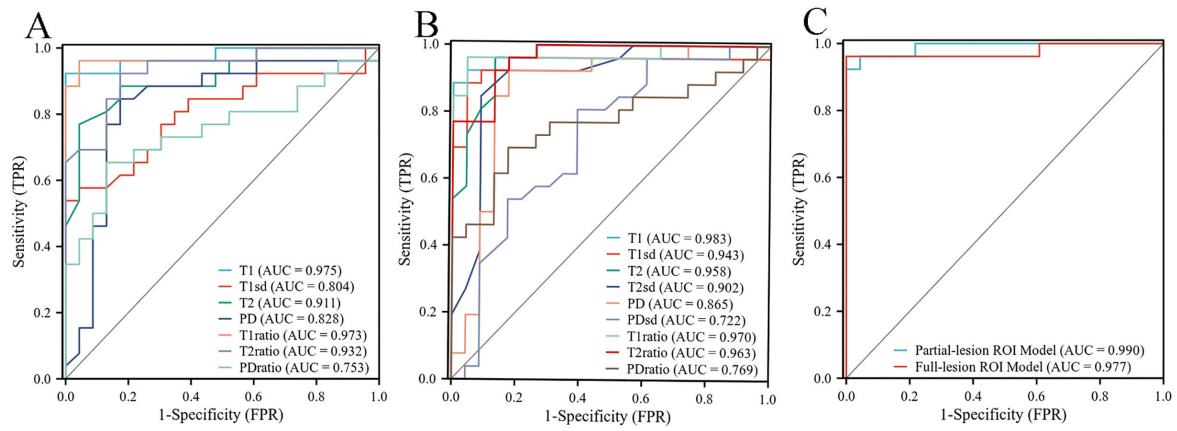


FIGURE 3
(A) ROC curves of quantitative parameters derived from full-lesion ROI delineation for differentiating between PA and WT. **(B)** ROC curves of quantitative parameters derived from partial-lesion ROI delineation for differentiating between PA and WT. **(C)** ROC curves of the combined SyMRI-derived multiple parameters for differentiating between PA and WT. Combined SyMRI-derived multiple parameters showed similar AUCs with these two different ROI delineation methods.

Our study had several limitations. Firstly, we revolved only around PA and WT due to low incidence and low sample size of the other parotid tumors. Secondly, this was a single-center study, which limits the generalizability of our findings. So future multicenter studies are needed to confirm our findings. However, the T1, T2, and PD values of MRI are fundamental properties independent of the MRI scanners or scanning parameters at a given field strength, suggesting our findings could be generalized to other hospitals. Finally, future studies should use multiple imaging modalities to detect more sensitive and robust biomarkers for distinguishing PA from WT.

In conclusion, the quantitative parameters of SyMRI were shown to be efficient in distinguishing PA from WT, with T1 values derived from partial lesion ROI delineation demonstrating the best diagnostic performance among all individual parameters.

Our findings may have important implications for preoperatively identifying optimal treatment strategies.

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.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Changshu No.2 People's Hospital, the Fifth Affiliated Clinical Medical College of Yangzhou University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JS: Data curation, Investigation, Software, Writing – original draft. XK: Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. DH: Data curation, Writing – original draft. XJ: Data curation, Writing – original draft. CJ: Funding acquisition, Investigation, Supervision, Writing – original draft. SW: Conceptualization, Data curation, Software, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the China Postdoctoral Science Foundation (2021M702170) and Changshu science and technology bureau (CSEY2021018).

Acknowledgments

We thank all the subjects who participated in this study. The authors have declared no conflict of interest with respect to the subject of this study.

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.

References

1. Sentani K, Ogawa I, Ozasa K, Sadakane A, Utada M, Tsuya T, et al. Characteristics of 5015 salivary gland neoplasms registered in the hiroshima tumor tissue registry over a period of 39 years. *J Clin Med*. (2019) 8:566. doi: 10.3390/jcm8050566

2. Andreasen S, Therkildsen MH, Bjørndal K, Homøe P. Pleomorphic adenoma of the parotid gland 1985-2010: A Danish nationwide study of incidence, recurrence rate, and Malignant transformation. *Head neck*. (2016) 38 Suppl 1:E1364–9. doi: 10.1002/hed.v38.S1

3. Espinoza S, Felter A, Malinvaud D, Badoual C, Chatellier G, Siauve N, et al. Warthin's tumor of parotid gland: Surgery or follow-up? Diagnostic value of a decisional algorithm with functional MRI. *Diagn interventional imaging*. (2016) 97:37–43. doi: 10.1016/j.diii.2014.11.024

4. Jia CH, Wang SY, Li Q, Qiu JM, Kuai XP. Conventional, diffusion, and dynamic contrast-enhanced MRI findings for differentiating metaplastic Warthin's tumor of the parotid gland. *Sci progress*. (2021) 104:368504211018583. doi: 10.1177/00368504211018583

5. Antony J, Gopalan V, Smith RA, Lam AK. Carcinoma ex pleomorphic adenoma: a comprehensive review of clinical, pathological and molecular data. *Head Neck pathol*. (2012) 6:1–9. doi: 10.1007/s12105-011-0281-z

6. Quer M, Hernandez-Prera JC, Silver CE, Casasayas M, Simo R, Vander Poorten V, et al. Current trends and controversies in the management of warthin tumor of the parotid gland. *Diagnostics (Basel Switzerland)*. (2021) 11:1467. doi: 10.3390/diagnostics11081467

7. Sood S, McGurk M, Vaz F. Management of salivary gland tumours: United Kingdom national multidisciplinary guidelines. *J laryngol otol*. (2016) 130:S142–s149. doi: 10.1017/S0022215116000566

8. Zaghi S, Hendizadeh L, Hung T, Farahvar S, Abemayor E, Sepahdari AR. MRI criteria for the diagnosis of pleomorphic adenoma: a validation study. *Am J otolaryngol*. (2014) 35:713–8. doi: 10.1016/j.amjoto.2014.07.013

9. Yabuuchi H, Matsuo Y, Kamitani T, Setoguchi T, Okafuji T, Soeda H, et al. Parotid gland tumors: can addition of diffusion-weighted MR imaging to dynamic contrast-enhanced MR imaging improve diagnostic accuracy in characterization? *Radiology*. (2008) 249:909–16. doi: 10.1148/radiol.2493072045

10. Coudert H, Mirafzal S, Dissard A, Boyer L, Montoriol PF. Multiparametric magnetic resonance imaging of parotid tumors: A systematic review. *Diagn interventional imaging*. (2021) 102:121–30. doi: 10.1016/j.diii.2020.08.002

11. Xu Z, Chen M, Zheng S, Chen S, Xiao J, Hu Z, et al. Differential diagnosis of parotid gland tumours: Application of SWI combined with DWI and DCE-MRI. *Eur J radiol*. (2022) 146:110094. doi: 10.1016/j.ejrad.2021.110094

12. Yabuuchi H, Fukuya T, Tajima T, Hachitanda Y, Tomita K, Koga M. Salivary gland tumors: diagnostic value of gadolinium-enhanced dynamic MR imaging with histopathologic correlation. *Radiology*. (2003) 226:345–54. doi: 10.1148/radiol.2262011486

13. Gabelloni M, Faggioni L, Attanasio S, Vani V, Goddi A, Colantonio S, et al. Can magnetic resonance radiomics analysis discriminate parotid gland tumors? A pilot study. *Diagnostics (Basel Switzerland)*. (2020) 10:900. doi: 10.3390/diagnostics10110900

14. Tanenbaum LN, Tsiouris AJ, Johnson AN, Naidich TP, DeLano MC, Melhem ER, et al. Synthetic MRI for clinical neuroimaging: results of the magnetic resonance image compilation (MAGiC) prospective, multicenter, multireader trial. *AJNR Am J neuroradiol*. (2017) 38:1103–10. doi: 10.3174/ajnr.A5227

15. Hagiwara A, Hori M, Cohen-Adad J, Nakazawa M, Suzuki Y, Kasahara A, et al. Linearity, bias, intrascanner repeatability, and interscanner reproducibility of quantitative multidynamic multiecho sequence for rapid simultaneous relaxometry at 3 T: A validation study with a standardized phantom and healthy controls. *Invest radiol*. (2019) 54:39–47. doi: 10.1097/RLI.0000000000000510

16. Gao W, Zhang S, Guo J, Wei X, Li X, Diao Y, et al. Investigation of synthetic relaxometry and diffusion measures in the differentiation of benign and Malignant

breast lesions as compared to BI-RADS. *J magnetic resonance imaging: JMRI*. (2021) 53:1118–27. doi: 10.1002/jmri.27435

17. Cui Y, Han S, Liu M, Wu PY, Zhang W, Zhang J, et al. Diagnosis and grading of prostate cancer by relaxation maps from synthetic MRI. *J magnetic resonance imaging: JMRI*. (2020) 52:552–64. doi: 10.1002/jmri.27075

18. Wang P, Hu S, Wang X, Ge Y, Zhao J, Qiao H, et al. Synthetic MRI in differentiating benign from metastatic retropharyngeal lymph node: combination with diffusion-weighted imaging. *Eur radiol*. (2023) 33:152–61. doi: 10.1007/s00330-022-09027-4

19. Xu X, Su G, Hu H, Wang Y, Hong X, Shi H, et al. Effects of regions of interest methods on apparent coefficient measurement of the parotid gland in early Sjögren's syndrome at 3T MRI. *Acta radiologica (Stockholm Sweden: 1987)*. (2017) 58:27–33. doi: 10.1177/0284185116637245

20. Bruvo M, Mahmood F. Apparent diffusion coefficient measurement of the parotid gland parenchyma. *Quantitative Imaging Med surgery*. (2021) 11:3812–29. doi: 10.21037/qims-20-1178

21. Yabuuchi H, Kamitani T, Sagiyama K, Yamasaki Y, Hida T, Matsuura Y, et al. Characterization of parotid gland tumors: added value of permeability MR imaging to DWI and DCE-MRI. *Eur radiol*. (2020) 30:6402–12. doi: 10.1007/s00330-020-07004-3

22. Han X, Suo S, Sun Y, Zu J, Qu J, Zhou Y, et al. Apparent diffusion coefficient measurement in glioma: Influence of region-of-interest determination methods on apparent diffusion coefficient values, interobserver variability, time efficiency, and diagnostic ability. *J magnetic resonance imaging: JMRI*. (2017) 45:722–30. doi: 10.1002/jmri.25405

23. Zhang H, Zhao J, Dai J, Chang J, Hu S, Wang P. Synthetic MRI quantitative parameters in discriminating stage T1 nasopharyngeal carcinoma and benign hyperplasia: Combination with morphological features. *Eur J radiol*. (2024) 170:111264. doi: 10.1016/j.ejrad.2023.111264

24. Kuai X, Zhu Y, Yuan Z, Wang S, Lin L, Ye X, et al. Perfluorooctyl bromide nanoemulsions holding MnO(2) nanoparticles with dual-modality imaging and glutathione depletion enhanced HIFU-eliciting tumor immunogenic cell death. *Acta Pharm Sinica B*. (2022) 12:967–81. doi: 10.1016/j.apsb.2021.07.025

25. Kuai X, Shao D, Wang S, Wu PY, Wu Y, Wang X. Neuromelanin-sensitive MRI of the substantia nigra distinguishes bipolar from unipolar depression. *Cereb Cortex*. (2024) 34:bhad423. doi: 10.1093/cercor/bhad423

26. Wen B, Zhang Z, Fu K, Zhu J, Liu L, Gao E, et al. Value of pre-/post-contrast-enhanced T1 mapping and readout segmentation of long variable echo-train diffusion-weighted imaging in differentiating parotid gland tumors. *Eur J radiol*. (2023) 162:110748. doi: 10.1016/j.ejrad.2023.110748

27. Cameron IL, Ord VA, Fullerton GD. Characterization of proton NMR relaxation times in normal and pathological tissues by correlation with other tissue parameters. *Magnetic resonance imaging*. (1984) 2:97–106. doi: 10.1016/0730-725X(84)90063-8

28. Meng T, He H, Liu H, Lv X, Huang C, Zhong L, et al. Investigation of the feasibility of synthetic MRI in the differential diagnosis of non-keratinising nasopharyngeal carcinoma and benign hyperplasia using different contoured methods for delineation of the region of interest. *Clin radiol*. (2021) 76:238.e9–238.e15. doi: 10.1016/j.crad.2020.10.010

29. Gökçe E, Beyhan M. Advanced magnetic resonance imaging findings in salivary gland tumors. *World J radiol*. (2022) 14:256–71. doi: 10.4329/wjr.v14.i8.256

30. Baohong W, Jing Z, Zanxia Z, Kun F, Liang L, Eryuan G, et al. T2 mapping and readout segmentation of long variable echo-train diffusion-weighted imaging for the differentiation of parotid gland tumors. *Eur J radiol*. (2022) 151:110265. doi: 10.1016/j.ejrad.2022.110265

31. Gracien RM, Reitz SC, Hof SM, Fleischer V, Zimmermann H, Droby A, et al. Changes and variability of proton density and T1 relaxation times in early multiple sclerosis: MRI markers of neuronal damage in the cerebral cortex. *Eur radiol*. (2016) 26:2578–86. doi: 10.1007/s00330-015-4072-x

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 October 2024
DOI 10.3389/fonc.2024.1436189

Check for updates

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Hsien-Chi Liao,
National Taiwan University Hospital, Taiwan
Dilara Akbulut,
National Institutes of Health (NIH),
United States

*CORRESPONDENCE
Pengliang Xu
xupliang@163.com

†These authors have contributed equally to this work

RECEIVED 21 May 2024
ACCEPTED 12 September 2024
PUBLISHED 11 October 2024

CITATION
Peng X, Zhao H, Wu S, Jia D, Hu M, Guo B, Hu J and Xu P (2024) Habitat-based CT radiomics enhances the ability to predict spread through air spaces in stage T1 invasive lung adenocarcinoma. *Front. Oncol.* 14:1436189. doi: 10.3389/fonc.2024.1436189

COPYRIGHT
© 2024 Peng, Zhao, Wu, Jia, Hu, Guo, Hu 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.

Habitat-based CT radiomics enhances the ability to predict spread through air spaces in stage T1 invasive lung adenocarcinoma

Xiuhua Peng[1†], Hongxing Zhao[1†], Shiyong Wu[1], Dan Jia[2], Miaomiao Hu[1], Biping Guo[3], Jinliang Hu[1] and Pengliang Xu[4*]

[1]Department of Radiology, The First People's Hospital of Huzhou, Huzhou, China, [2]Department of Respiratory Medicine, The First People's Hospital of Huzhou, Huzhou, China, [3]Department of Ultrasound, The First People's Hospital of Huzhou, Huzhou, China, [4]Department of Thoracic Surgery, The First People's Hospital of Huzhou, Huzhou, China

Introduction: Spread through air spaces (STAS) represents a novel invasive pattern in lung adenocarcinoma (LUAD) and is a risk factor for poor prognosis in stage T1 LUAD. This study aims to develop and validate a CT habitat imaging analysis model for predicting STAS in stage T1 invasive LUAD.

Methods: We retrospectively analyzed 217 patients with preoperative stage T1 invasive LUAD (115 STAS-positive and 102 STAS-negative cases, including 151 in the train set and 66 in the test set). Semi-automatic segmentation was performed on the regions of interest (ROIs) in all CT images, with an automatic 3mm expansion around the tumor, considering the intratumoral and peritumoral 3mm area. This area was divided into three sub-regions via K-means clustering, and 1197 radiomic features were extracted from each sub-region and the overall combined region. After dimension reduction through the Mann-Whitney U test, Pearson correlation analysis, and least absolute shrinkage and selection operator (LASSO), the best features for each sub-region and overall were selected. Models were then built using the selected radiomic features through the Adaptive Boosting (AdaBoost) and Multilayer Perceptron (MLP) classifiers. Four different models were established based on different sub-regions and the overall features. The performance of these models was evaluated through receiver operating characteristic curves (AUC) under the DeLong test, calibration curves via the Hosmer-Lemeshow test, and decision curve analysis to assess the performance of these features.

Results: In this study, we evaluated the predictive performance of AdaBoost and MLP classifiers on rad feature models across various subregions and the overall dataset. In the test set, the AdaBoost classifier achieved a maximum AUC of 0.871 in Habitat 3, whereas the MLP classifier demonstrated slightly superior performance with an AUC of 0.879. Both classifiers exhibited high efficiency in habitat 3, with the MLP algorithm showing enhanced model performance.

Conclusions: CT habitat imaging analysis for the preoperative prediction of STAS in stage T1 invasive LUAD shows satisfactory diagnostic performance, with the habitat3 model exhibiting the highest efficacy, reflecting tumor heterogeneity.

KEYWORDS

invasive lung adenocarcinoma, radiomics, habitat imaging, tumor microenvironment, air space spread

1 Introduction

Lung adenocarcinoma (LUAD) is one of the malignancies with high incidence and mortality rates globally, and early-stage diagnosis and treatment significantly impact patient survival rates (1). In recent years, with the advancement of low-dose helical computed tomography and increased public health screening awareness, the detection rate of early-stage lung cancer has been on the rise; according to guidelines by the National Comprehensive Cancer Network(NCCN)in the United States, the current standard treatment for stage T1 LUAD is radical pulmonary resection, specifically lobectomy (2). Numerous domestic and international studies have also suggested that sublobar resection for early-stage lung nodules not only preserves lung function but also reduces the incidence of postoperative complications (3, 4). Despite surgical treatment significantly enhancing the cure rate and survival of patients with stage T1 lung adenocarcinoma, the recurrence rate in these patients still stands at 20-30% (5). Therefore, identifying patients with stage T1 LUAD at high risk of recurrence to provide precise surgical strategies has become a focal point of current clinical research.

Spread through air spaces (STAS) is a dissemination mode specific to lung cancer, involving cancer cells detaching from the primary site and spreading to surrounding areas via respiratory movements, reimplanting in the respiratory tract and alveolar walls, and further growing, thus facilitating lung cancer metastasis. This concept was first introduced in 2015 and was quickly recognized as an independent risk factor for lung adenocarcinoma recurrence and poor prognosis (6). Current studies have explored the relationship between STAS and postoperative prognosis of lung adenocarcinoma, where STAS-positive patients exhibit higher local and distant recurrence rates and shorter recurrence-free survival (RFS) (7–9). Some scholars have compared lobectomy and sublobar resection in stage T1 LUAD STAS-positive patients, finding significantly increased recurrence and metastasis rates in sublobar resections, suggesting that STAS positivity is an independent risk factor for stage T1 LUAD (10). Therefore, preoperative prediction of STAS status is crucial for choosing the surgical approach for stage T1 LUAD. If STAS can be diagnosed preoperatively or intraoperatively, it would guide the selection of clinical surgical treatment methods. Due to the limited inflation of lung tissues in intraoperative frozen sections, their ability to predict STAS status is limited. Currently, choosing a preoperative surgical approach for stage T1 LUAD remains a challenge, and accurate preoperative prediction of STAS is a hot research topic and an essential means to improve lung cancer survival rates, offering significant guidance for planning clinical surgical strategies.

Currently, radiomics technology has been extensively validated for predicting air space spread in LUAD but is limited to traditional radiomic analysis, i.e., of the tumor and surrounding areas (11–15). Each tumor is not just a homogenous entity but a mosaic of unique microenvironments, or habitats, composed of clusters of voxels with similar characteristics, consisting of tumor cells with identical genotypes and phenotypes. Habitat imaging distinctly segments the tumor into these sub-regions, which can reflect the spatial heterogeneity within the tumor to a certain extent, providing a new perspective for understanding and predicting the invasive behavior of tumors (16, 17). Habitat analysis has been applied to other types of tumors, such as breast cancer (18, 19), glioma (20, 21), cervical cancer (22, 23), liver cancer (24), colorectal cancer (25), recurrence of non-small cell lung cancer (26), pulmonary metastases (27), etc., but reports on its application in predicting STAS in lung adenocarcinoma are yet to be seen. The purpose of this study is to use habitat imaging to segment tumor sub-regions and incorporate peritumoral imaging to characterize spatial heterogeneity more accurately, aiming to predict the STAS status in stage T1 invasive LUAD more precisely, thereby providing a new scientific basis for preoperative assessment and treatment decision-making in stage T1 invasive LUAD.

2 Materials and methods

2.1 Patient selection and clinicopathological information

This study follows the Declaration of Helsinki, has been approved by the Ethics Committee of Huzhou First People's Hospital, and has waived the patients' informed consent. The data of patients admitted to T1 invasive lung adenocarcinoma from January 2019 to December 2023. Inclusion criteria: (1) patients with maximum CT tumor diameter less than 3CM; (2) patients with CT imaging data within one month before surgery; (3) patients diagnosed with invasive lung adenocarcinoma; and (4)

patients without distant metastasis before surgery. Exclusion criteria: (1) patients who had received neoadjuvant therapy; (2) patients with multiple pulmonary nodules reported on preoperative CT images; (3) patients with current or previous history of other malignancy; (4) patients with incomplete clinical data collection; and (5) patients whose images were not identified by ITK-SNAP. Ultimately, 455 eligible patients were continuously enrolled, of whom 115 were positive for STAS and 340 were negative for STAS. To overcome the possible imbalance in the data, we randomly grouped the STAS negative cases by 3:7 and matched them to an almost 1:1 ratio of the STAS positive group. This data balancing method has been demonstrated in previous studies (15), divided into a training group (n=151) and a validation group (n=66) (Figure 1).

Clinical and pathological variables included age, sex, serum tumor markers (carcinoembryonic antigen (CEA) and sugar antigen CA125 (CA125)), smoking status, tumor location, and emphysema; KI67, vascular infiltration, nerve infiltration, and invasion of pleural membrane.

2.2 Histopathological evaluation

According to the WHO, STAS is defined as the spread of micropapillary clusters, solid nests, or individual tumor cells into the alveolar spaces beyond the main tumor edge. The main morphological features include: (1) alveolar spaces filled with ring-shaped micropapillary structures without or with occasional central fibrovascular cores; (2) alveolar spaces filled with solid nests or tumor islands composed of tumor cells; and (3) alveolar spaces filled with discontinuous individual tumor cells. In this study, pathological examinations were reassessed and diagnosed by an

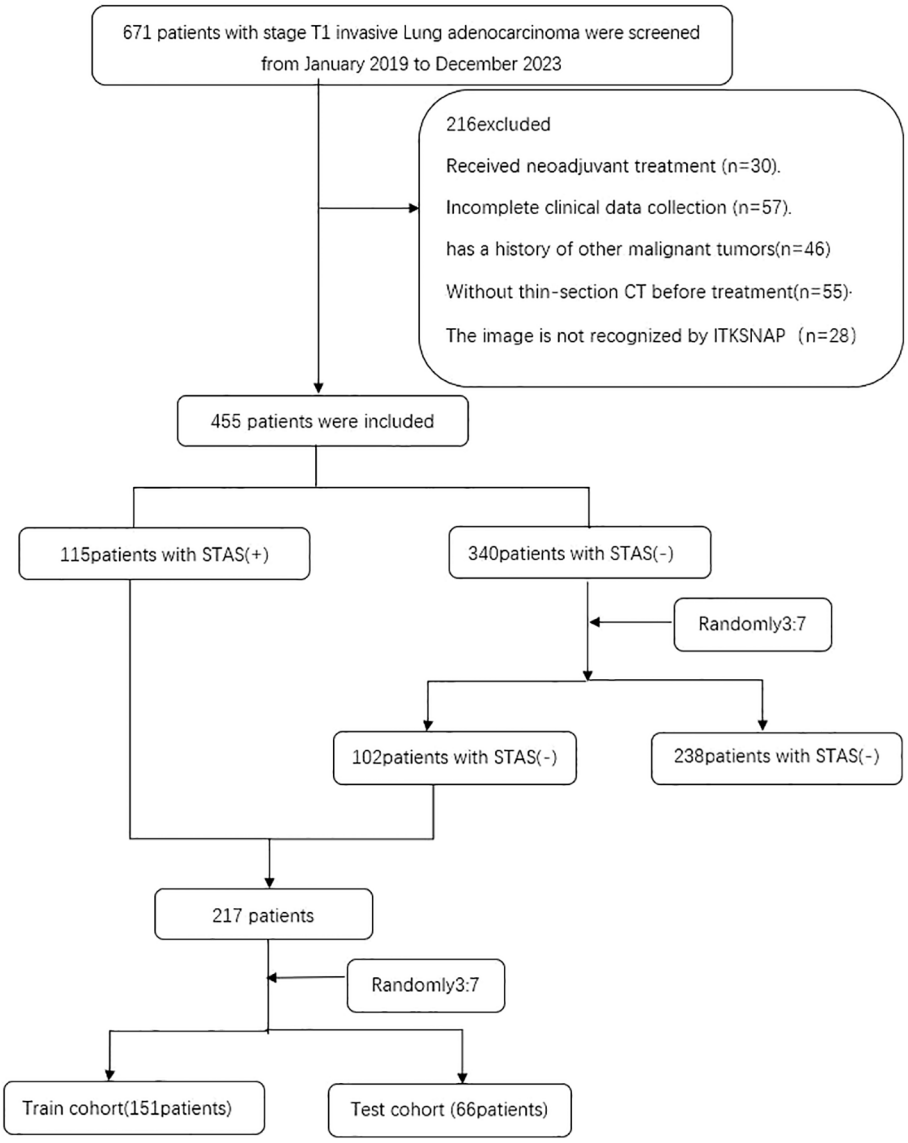


FIGURE 1
Flow diagram of the enrolment patients. STAS spread through air spaces; STAS(+), presence of STAS; STAS(-), absence of STAS.

attending physician with 5 years of experience and a chief physician with 15 years of experience, based on the WHO definition of STAS.

2.3 CT examination protocol

The chest scan was performed with German Siemens Definition AS 64-row 128-slice spiral CT. Scan from thoracic entrance to diaphragm level. The subjects were placed in the supine position and held their breath after deep inhalation. Scanning parameters: tube voltage 120kv, tube current 120mA, window width 1300-1500, window position: -600~-700, pitch 1.0, frame rotation time 0.33S/360 degrees. Lung window reconstruction was performed using the lung method with a reconstruction thickness of 1.25mm and layer spacing of 1.25mm. Mediastinal window reconstruction thickness and layer spacing were 5mm.

2.4 Image segmentation and data preprocessing

Image segmentation was performed independently by two radiologists with extensive experience. They were blinded to the patients' histopathology. One of the radiologists (radiologist A, with 5 years of experience) manually drew the ROI slice by slice using the open-source software ITK-SNAP (version3.8.0, http://www.itksnap.org). Another radiologist (radiologist B, with 10 years of experience) reviewed all ROIs manually segmented by radiologist A. The software automatically expands its boundaries by 3mm to get a gross peritumor ROI. The 3mm area surrounding the tumor was manually removed where the soft tissue, bone, and mediastinum overlapped in the chest wall. Intra-class correlation coefficients (ICCs) are used to evaluate feature extraction. In terms of intra-observer and inter-observer consistency, ICCs≥0.75 indicates good consistency.

The dataset was randomly assigned in a 7:3 ratio to either the training dataset or the test dataset. All cases in the training dataset were used to train the predictive model, while cases in the test dataset were used to evaluate the model's performance independently. Medical volumes are common with heterogeneous voxel spacing because of different scanners or acquisition protocols. Such spacing refers to the physical distance between two pixels in an image. Spatial normalization is often employed to reduce the effect of voxel spacing variation. The fixed-resolution resampling method was used in our experiment to handle the problems mentioned above. All images were resampled to a voxel size of 1*1*1 mm to standardize the voxel spacing. Finally, the data were standardized using z-score standardization (zero-mean normalization).

2.5 Sub-region clustering and feature extraction

Habitat utilizes voxel and entropy values from CT images to cluster VOIs into sub-regions (28–30). The voxel counts for each tumor VOI were determined using a traditional method, whereas the entropy values were computed for each layer of the CT images using the following formula:

$$V_{voxel} = \sum_{k=1}^{N_V} V_k$$

$$\text{entropy} = -\sum_{i=1}^{N_g} p(i)\log_2(p(i)+\varepsilon) =$$

The k-means method was employed to cluster the VOI regions at the patient level, forming multiple habitats, and the distance correlation between samples was calculated using the Euclidean distance (voxel values and entropy values). The number of habitats was tested from 2 to 10 to determine the optimized number of habitats with the highest evaluation metric, the Calinski–Harabasz index (31). The optimal k-value was the criterion for selecting the optimal number of clusters at the patient population level. The optimal k-value was found to be 3. Using Python software, we imported the volume of interest (VOI) for each patient into the system. The T1 aggressive lung adenocarcinomas were classified into three distinct categories: habitat 1, habitat 2, and habitat 3.

The handcrafted features can be divided into three groups: (I) geometry, (II) intensity, and (III) texture. The geometry features describe the three-dimensional shape characteristics of the tumor. The intensity features describe the first-order statistical distribution of the voxel intensities within the tumor. The texture features describe the patterns or the second and high-order spatial distributions of the intensities. Here the texture features are extracted using several different methods, including the gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighborhood gray-tone difference matrix (NGTDM) methods. Based on habitat imaging, image omics features were extracted from three subregions and the peri-tumor region, respectively. All features were extracted using the Pyradiomics package in Python version 3.9. Eight wavelet transform algorithms are used to obtain high-throughput features for first-order statistics and texture features, namely LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH. Do z-score normalization for all features and change the feature value to 0 mean 1 variance.

2.6 Feature selection and model design

Statistics: We performed a Mann-Whitney U test and feature screening for all radiomic features. Only the radiomic features with a p-value< 0.05 were kept. Correlation: For features with high repeatability, Spearman's rank correlation coefficient was also used to calculate the correlation between features, and one of the features with a correlation coefficient greater than 0.9 between any two features is retained. We use a greedy recursive deletion strategy for feature filtering to maintain the ability to depict features to the greatest extent. That is, the feature with the greatest redundancy in the current set is deleted each time. Lasso: LASSO regression model was used on the discovery data set for signature construction. Depending on the regulation weight λ, LASSO shrinks all regression coefficients toward zero and sets the coefficients of

many irrelevant features exactly to zero. To find an optimal λ, 10-fold cross-validation with minimum criteria was employed, where the final value of λ yielded minimum cross-validation error. The retained features with nonzero coefficients were used for regression model fitting and combined into a radiomics signature. Subsequently, we obtained a radiomics score for each patient by a linear combination of retained features weighed by their model coefficients. The Python scikit-learn package was used for LASSO regression modeling.

After Lasso feature screening, we input the final features into the AdaBoost and MLP classifiers for risk model construction. Here, we adopt 5 5-fold cross-verification to obtain the final Rad Signature. Receiver operating characteristic (ROC) curves were plotted to assess the diagnostic performance of the predictive models, and the corresponding area under the curve (AUC), diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were analyzed.

2.7 Statistical analysis

The Python state models (version 0.13.2) package was used to perform statistical analysis, and a p-value< 0.05 was considered statistically significant. We analyzed the differences between groups using Student's t-test or Mann–Whitney U tests for continuous variables; the chi-square test or Fisher's exact test was applied for categorical variables.

3 Results

3.1 Patient characteristics

A total of 217 patients, including STAS positive 115and STAS negative 102patients, were included in our study. Patients were divided into a training set (151 patients) and an independent test set (66 patients) based on treatment duration. A pathologist reviewed the pathological data. All patients underwent surgical treatment; there were 77 (51%) patients with STAS positive and 74 (49%) patients with STAS negative in the training group and37 (56%) patients with STAS positive and 29(44%) patients with STAS negative in the test group.

The characteristics of the patients in the cohort are shown in Table 1. The comparison of age, gender, CA125, smoking status, lobular locations, Vascular infiltration, perineural invasion, and Pleural infiltration showed no significant difference between the two groups and within each group (p>0.05), ensuring a reasonable classification. Significant differences between the cohorts were found in CEA (p<0.05).

3.2 Workflow of radiomics analysis

The radiomics analysis consisted of a series of steps: image segmentation, feature extraction, feature selection, signature construction, and evaluation (Figure 2).

3.3 Feature selection and radiomics signature development

Features Statistics: Optimal CH value emerged when tumors were clustered into three sub-regions in the entire cohort (Figure 3). A total of 1197 manual features were extracted for each sub-region and population, among which 234 were the first feature, 14 were shape features, and the last were texture features. All handcrafted features are extracted with an in-house feature analysis program implemented in Pyradiomic (http://pyradiomics.readthedocs.io).

Feature selection: Through the Mann-Whitney U test, Pearson correlation analysis, and LASSO dimension reduction processing, we screened out the best features of each sub-region and the whole. Fifteen features were obtained in habitat 1. 23 features were obtained in habitat 2. Fifteen features were obtained in habitat 3. In the whole region, 20 features are obtained.

We use AdaBoost and MLP classifiers, respectively, to predict rad feature models for each habitat. Among them, the test group habitat 3 of the AdaBoost classifier has the highest AUC, 0.871. The test group of the MLP classifier also has the highest AUC, which is 0.879. Therefore, among the two classifiers, habitat 3 has higher efficiency, and the model efficiency of the MLP classifier is higher. (Tables 2, 3). Figure 4 shows the receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA) for all features in the training and testing sets of the MLP model. Barplots depicting the classification performance of the habitat3 signature in the training and validation cohorts are shown in Figure 5. The distribution of three different habitats in a three-dimensional space is shown in Figure 6.

CT images and histopathological photographs of STAS-positive nodule is shown in Figure 7.

4 Discussion

In this study, we utilized CT images to extract radiomic features from the tumor and an additional 3mm peritumoral sub-regions, establishing a multi-regional radiomics model for predicting the status of STAS. Results indicated that Habitat3 models exhibited robust predictive performance, with AUC values of 0.881 in the training set and 0.879 in the testing set using the MLP classifier. This approach, employing radiomics from the tumor and 3mm peritumoral sub-regions divided into different areas, is a novel method proposed for predicting STAS status, providing a robust basis for preoperative precise surgical planning for patients with stage T1 invasive LUAD.

Numerous researchers have explored radiomics to predict LUAD STAS. Chen et al. (32) segmented tumors in a cohort of 233 stage I LUAD patients, extracted radiomic features, and constructed a model to predict the presence of STAS. The model demonstrated AUC values of 0.63 and 0.69 during internal and external validation, respectively, underscoring CT-based radiomics's utility in the preoperative STAS prediction in stage I LUAD. Han et al. (15) employed a similar radiomic approach on preoperative stage IA LUAD patients, achieving AUC values of

10.3389/fonc.2024.1436189

TABLE 1 Baseline characteristics of patients in the training cohort and test cohort.

Characteristics	Total (N=217)	Train(N=151)	Test(N=66)	P-value
Age (years)	64.96 ± 9.58	65.44 ± 9.21	63.88 ± 10.37	0.424
Maximum-diametere(cm) () ()(cm)	1.68 ± 0.64	1.67 ± 0.65	1.70 ± 0.61	0.574
Gender				0.07
Male	104(47.93)	79(52.32)	25(37.88)	
Female	113(52.07)	72(47.68)	41(62.12)	
Smoke				0.414
Non-smoker	164(75.58)	117(77.48)	47(71.21)	
Smoker	53(24.42)	34(22.52)	19(28.79)	
CEA				0.007
Negative	171(78.80)	111(73.51)	60(90.91)	
Positive	46(21.20)	40(26.49)	6(9.09)	
CA125				0.056
Negative	206(94.93)	140(92.72)	66(100.00)	
Positive	11(5.07)	11(7.28)	0	
Lobular locations				0.533
RUL	67(30.88)	48(31.79)	19(28.79)	
RML	27(12.44)	15(9.93)	12(18.18)	
RLL	31(14.29)	21(13.91)	10(15.15)	
LUL	59(27.19)	43(28.48)	16(24.24)	
LLL	33(15.21)	24(15.89)	9(13.64)	
Vascular_infiltration				1.0
No	142(65.44)	99(65.56)	43(65.15)	
Yes	75(34.56)	52(34.44)	23(34.85)	
perineural_invasion				0.984
No	212(97.70)	147(97.35)	65(98.48)	
Yes	5(2.30)	4(2.65)	1(1.52)	
Pleural_infiltration				0.235
No	179(82.49)	121(80.13)	58(87.88)	
Yes	38(17.51)	30(19.87)	8(12.12)	
KI67				0.827
<20%	181(83.41)	127(84.11)	54(81.82)	
≥20%	36(16.59)	24(15.89)	12(18.18)	
lymphatic_metastasis				0.707
No	193(88.94)	133(88.08)	60(90.91)	
Yes	24(11.06)	18(11.92)	6(9.09)	
Emphysema				0.214
No	143(65.90)	104(68.87)	39(59.09)	
Yes	74(34.10)	47(31.13)	27(40.91)	

LLL, left lower lobe; LUL, left upper lobe; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.

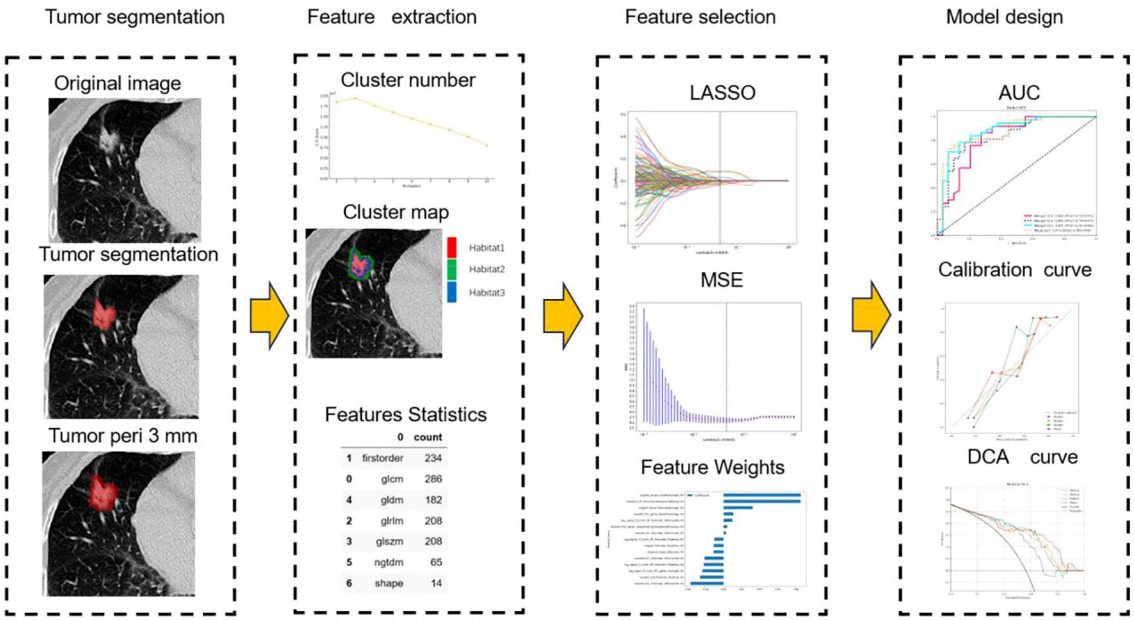


FIGURE 2
Workflow of radiomics analysis.

0.812 and 0.850 in the training and testing sets, respectively, using logistic regression. Jiang et al. (11) utilized the same segmentation technique on a dataset of 462 LUAD patients and developed a model using the random forest classifier, which achieved an AUC of 0.754. These studies focused solely on the tumor without analyzing the peritumoral area's impact on STAS. Zhuo and Qi et al. (12, 33) studied various regions surrounding lung adenocarcinoma tumors and determined that an integrated model incorporating peritumoral areas along with other clinical parameters demonstrates higher efficacy than a radiomic model based solely on the tumor itself. These studies highlight that STAS occurrence is mainly related to the intrinsic aggressiveness of the tumor, with the peritumoral area also reflecting the tumor's invasive behavior to some extent.

Our study combined features from the tumor and its surroundings, avoiding the oversight of additional value from the tumor microenvironment. We discovered that representing the spatial heterogeneity within different habitats of the tumor and its surroundings can better predict the status of STAS. Tumors themselves, as complex ecosystems akin to natural habitats, consist of heterogeneous sub regions that follow the principle of

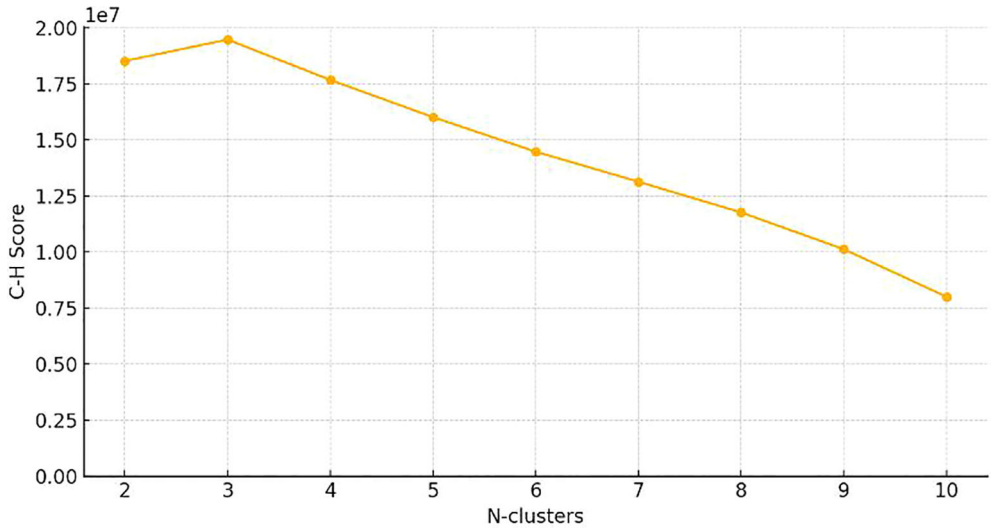


FIGURE 3
Calinski–Harabasz score plot. The red dotted line represented the optimal value beyond which the scores started to decrease in the radiomics features from CT images.

TABLE 2 Performance of each sub-region and the whole imaging rad model of the AdaBoost classifier in predicting STAS.

Signature	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
Habitat1	0.815	0.905	0.8608 - 0.9499	0.885	0.740	0.784	0.857	Train
Habitat2	0.881	0.957	0.9307 - 0.9838	0.885	0.877	0.885	0.877	Train
Habitat3	0.861	0.952	0.9230 - 0.9801	0.859	0.863	0.870	0.851	Train
Whole	0.854	0.940	0.9057 - 0.9745	0.795	0.918	0.912	0.807	Train
Habitat1	0.773	0.852	0.7553 - 0.9484	0.730	0.828	0.844	0.706	Test
Habitat2	0.727	0.802	0.6975 - 0.9064	0.757	0.690	0.757	0.690	Test
Habitat3	0.788	0.871	0.7845 - 0.9583	0.784	0.793	0.829	0.742	Test
Whole	0.727	0.816	0.7061 - 0.9258	0.676	0.793	0.806	0.657	Test

TABLE 3 Performance of each subregion and the whole imaging rad model of the MLP classifier in predicting STAS.

Signature	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
Habitat1	0.795	0.844	0.7824 - 0.9057	0.872	0.712	0.764	0.839	Train
Habitat2	0.815	0.906	0.8615 - 0.9509	0.833	0.795	0.812	0.817	Train
Habitat3	0.808	0.881	0.8295 - 0.9334	0.885	0.726	0.775	0.855	Train
Whole	0.828	0.907	0.8616 - 0.9516	0.756	0.904	0.894	0.776	Train
Habitat1	0.788	0.822	0.7119 - 0.9311	0.838	0.724	0.795	0.778	Test
Habitat2	0.788	0.838	0.7386 - 0.9371	0.757	0.828	0.848	0.727	Test
Habitat3	0.803	0.879	0.7900 - 0.9677	0.757	0.862	0.875	0.735	Test
Whole	0.803	0.871	0.7838 - 0.9590	0.703	0.931	0.929	0.711	Test

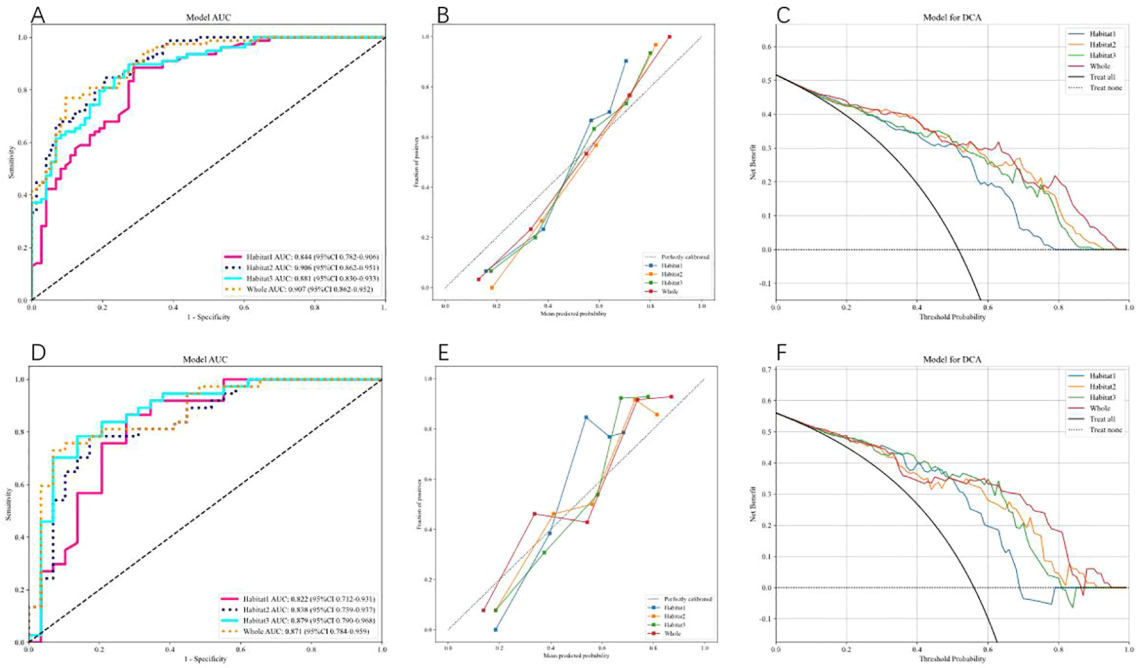


FIGURE 4
shows the Receiver Operating Characteristic (ROC) curve, correction curve, and decision curve analysis (DCA) for all features in the training cohort **(A–C)** and test cohort **(D–F)** of the MLP model.

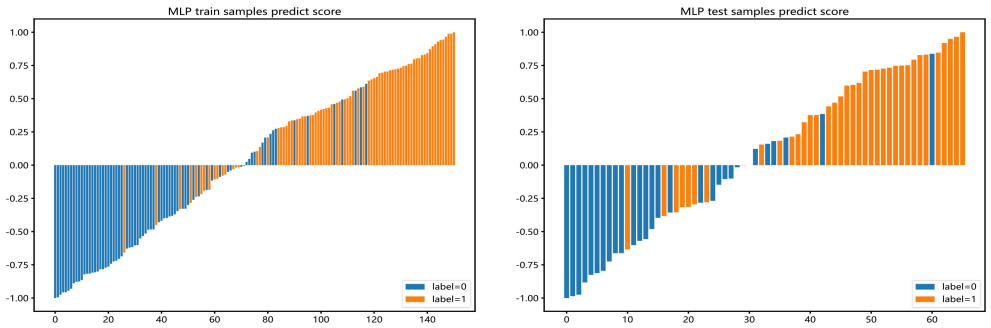


FIGURE 5
Barplots depicting the classification performance of radiomic features of habitat 3 in the MLP model. The yellow bar with a prediction value > 0 indicates that the signature successfully classifies the STAS patients; the red bar with a prediction value< 0 indicates that the signature fails to classify the STAS patients. For the blue bar, the contrary applies.

survival of the fittest, growing and increasing under various pressures, closely related to the progression and prognosis of the tumor. Based on radiomics principles and differences in pathology and biology, Habitat imaging uses quantitative imaging markers to comprehensively and non-invasively characterize the tumor environment, visualizing and quantifying the tumor's internal heterogeneity. We obtained 1197 high-dimensional features from each sub region for STAS based on CT images. Therefore, feature selection is a critical step before constructing a radiomics model. Our study employed a three-step method for feature selection from different aspects. To describe tumor heterogeneity, we identified the best features from each sub region and the overall combined regions. We obtained 15 features in habitat1, 23 in habitat2, and 15 in habitat3, with 20 features in the whole area. The obtained features primarily focused on filtered image intensity and texture characteristics, exploring deeper correlations with tumor biological changes through filtering transformations.

We used the AdaBoost and MLP classifiers to predict rad feature models for each habitat and the whole. Among these, the test set for habitat3 had the highest AUC with the AdaBoost classifier (0.871) and the MLP classifier (0.879). Thus, habitat3 demonstrated high efficacy with both classifiers, with the MLP classifier showing higher model performance. AdaBoost's main advantage lies in its ability to enhance difficult-to-classify samples, improving the overall accuracy of the model, and compared to some complex classifiers, it is more efficient in implementation and operation, making it a powerful and practical tool for various classification challenges. MLP, a type of feedforward artificial neural network, consists of multiple layers, including input, one or more hidden layers, and an output layer, with each layer composed of numerous neurons interconnected by learnable weights. Its main advantages are its strong non-linear learning capabilities and complex pattern representation ability, making it highly effective in many modern AI applications such as image recognition, speech processing, and complex classification tasks. The results obtained using these two classifiers established a more robust and resilient model.

This study concludes that the model constructed by habitat3 exhibits the highest efficacy, surpassing that of the whole model, indicating that intra-tumoral subregions are more indicative of STAS status than the tumor as a whole. This also suggests that habitat analysis reflects tumor heterogeneity more than traditional

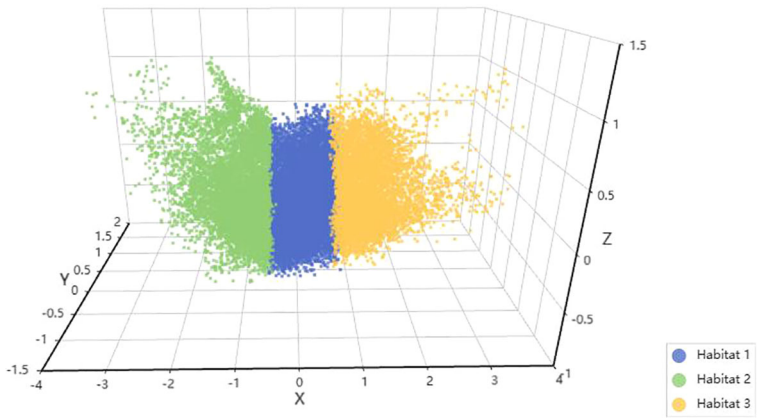


FIGURE 6
Habitat 3D visualization. Each habitat is represented by a different color: green for habitat 1, blue for habitat 2, and yellow for habitat 3.

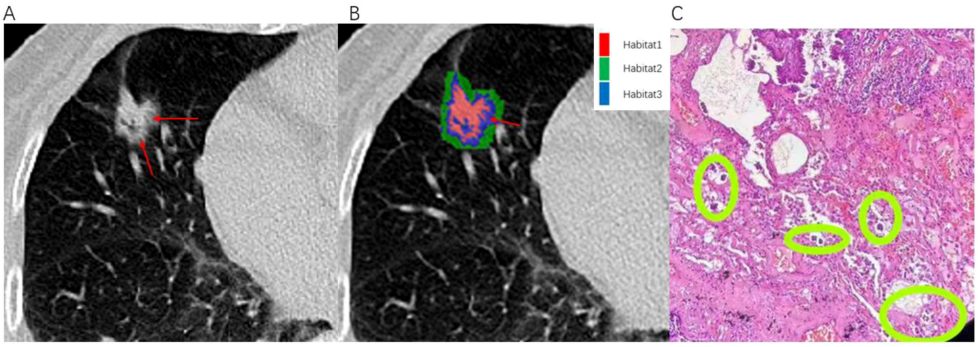


FIGURE 7
Illustrates an example of a patient with infiltrative pulmonary adenocarcinoma. An 80-year-old female presented with positive STAS associated with lung adenocarcinoma. **(A)** The axial CT image (width, 1300 HU; level, -300 HU) shows a mixed nodule in the right middle lobe, with ground-glass components indicated by the red arrow. **(B)** The habitat analysis image divides the lung nodule into three habitats, with the blue area representing Habitat 3 (indicated by the red arrow). **(C)** Microscopic photograph (Hematoxylin-Eosin staining, magnification 40x) reveals clusters of isolated small papillary tumor cells (within the green circle) present in the alveoli beyond the margins of the primary tumor.

radiomics (22). Habitat3 represents the peripheral ground-glass opacity (GGO) components of the tumor, possibly due to tumor cells in the GGO area enhancing their migration and invasion capabilities by altering cell adhesion mechanisms and signaling pathways, leading to the occurrence of STAS (34). The GGO area typically reflects changes in the tumor microenvironment, including inflammatory responses and extracellular matrix remodeling, which may also promote the occurrence of STAS (35, 36). Habitat3 manifests radiologically as the range inward or outward from the tumor edge, corresponding to the radiological tumor margin, which aligns with the pathological tumor margin. It is also possible that tumor growth gradually expands outward, with central tumor cells gradually hypoxic and necrotic due to aggregation, resulting in the richest blood supply at the tumor's edge, thus reflecting tumor heterogeneity most prominently. However, habitat2, representing the area surrounding the tumor, exhibits lower model efficacy than Habitat3, possibly due to interference from normal lung tissue in the peritumoral region. Previous studies (33) extracting peritumoral areas at 5mm, 10mm, and 15mm have yielded good predictive model performance but with deviations from the ideal curve in calibration. This suggests that predictive models based on radiomic parameters within lung tumors outperform those based on radiomic parameters from the peritumoral region, possibly due to greater interference from normal lung tissue surrounding the tumor. Some studies separately extracted radiomic features from the tumor and peritumoral regions at 3mm, 6mm, and 9mm, constructed models, and found that radiomic features from the tumor and peritumoral regions at 3mm can enhance the impact on the overall survival rate of patients with non-small cell lung cancer after surgery, and radiomic features from the peritumoral region at 3mm are associated with STAS (37). This study only extracted 3mm around the tumor for this reason. Previous methods of automatic expansion of peritumoral areas have been rigid and prone to interference, unable to reflect the true characteristics of the tumor surroundings accurately. This study utilizes unsupervised K-means clustering to categorize identical components within the tumor and peritumoral regions into the same subregion, resulting in a higher efficacy in predicting STAS status. Research indicates (38) that a 5mm expansion or contraction within the tumor edge yields higher predictive efficacy for STAS, similar to the results of this study, although with a lower AUC of 0.79 compared to our findings. This may be due to the automatic expansion of the previous study's contraction area, whereas our study is based on clustering of identical components, resulting in slightly more precise results.

This study has several limitations: firstly, as a retrospective study with a small sample size from a single center, there may be potential bias. It is recommended that future studies be multi-centered to increase the number of cases and validate the model across different patient populations, thereby enhancing its generalizability and practicality. Secondly, this study utilized manual segmentation for ROI delineation, which may introduce subjective variability. Future research should employ automated segmentation techniques to mitigate these limitations and enhance segmentation consistency and accuracy. Additionally, this study only used conventional CT scans for subregion segmentation, not accurately representing the tumor's information. Energy spectrum CT enhancement, containing more material energy and blood supply information, will be included in future research. Besides machine learning, deep learning is also becoming increasingly popular, and this aspect can also be explored in LUAD STAS habitat analysis.

In conclusion, this study constructed a CT habitat subregion radiomics model that non-invasively predicts the STAS status of T1-stage invasive lung adenocarcinoma preoperatively. The model has improved the robustness of predictive performance, providing a basis for distinguishing the benign or malignant nature of lung nodules and has auxiliary diagnostic value in assessing malignancy severity. The results offer a quantitative reference for preoperative surgical planning and postoperative chemotherapy selection in patients with T1-stage invasive lung adenocarcinoma, suggesting its potential as a non-invasive biomarker for preoperative STAS in lung adenocarcinoma patients.

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.

Ethics statement

The studies involving humans were approved by The Medical Ethical Committee of The First People's Hospital of Huzhou. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

XP: Conceptualization, Formal analysis, Methodology, Writing – original draft. HZ: Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing – original draft. SW: Data curation, Investigation, Writing – review & editing. DJ: Software, Validation, Writing – review & editing. MH: Software, Validation, Writing – review & editing. BG: Data curation, Formal analysis, Writing – review & editing. JH: Software, Validation, Visualization, Writing – review & editing. PX: Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Project funded by the Municipal and Bureau-level Scientific Research Program of 2023 (Grant No. 2023GY33).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1436189/full#supplementary-material

References

1. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin.* (2019) 69:184–210. doi: 10.3322/caac.21557

2. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, et al. Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* (2017) 15:504–35. doi: 10.6004/jnccn.2017.0050

3. Hu J, Chen J, Chen C, Zhong W, Geng Q. Wedge resection of pulmonary nodules (</=2 cm): A consensus statement by specialists of thoracic surgery (2023 edition). *Zhongguo Fei Ai Za Zhi.* (2023) 26:338–47. doi: 10.3779/j.issn.1009-3419.2023.102.17

4. Altorki N, Wang X, Kozono D, Watt C, Landrenau R, Wigle D, et al. Lobar or sublobar resection for peripheral stage IA non-small-cell lung cancer. *N Engl J Med.* (2023) 388:489–98. doi: 10.1056/NEJMoa2212083

5. Lv G, Feng S, Zhang Y, Li Y, Cui Y. Clinical therapeutic effects and prognosis of video-assisted thoracoscopic surgery-guided pulmonary lobectomy combined with mediastinal lymph node dissection in lung carcinoma. *Am J Cancer Res.* (2023) 13:5138–50.

6. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 world health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol.* (2015) 10:1243–60. doi: 10.1097/JTO.0000000000000630

7. Chen Z, Wu X, Fang T, Ge Z, Liu J, Wu Q, et al. Prognostic impact of tumor spread through air spaces for T2aN0 stage IB non-small cell lung cancer. *Cancer Med.* (2023) 12:15246–55. doi: 10.1002/cam4.v12.14

8. Shiono S, Endo M, Suzuki K, Yanagawa N. Spread through air spaces affects survival and recurrence of patients with clinical stage IA non-small cell lung cancer after wedge resection. *J Thorac Dis.* (2020) 12:2247–60. doi: 10.21037/jtd.2020.04.47

9. Liu H, Yin Q, Yang G, Qie P. Prognostic impact of tumor spread through air spaces in non-small cell lung cancers: a meta-analysis including 3564 patients. *Pathol Oncol Res.* (2019) 25:1303–10. doi: 10.1007/s12253-019-00616-1

10. Ren Y, Xie H, Dai C, She Y, Su H, Xie D, et al. Prognostic impact of tumor spread through air spaces in sublobar resection for 1A lung adenocarcinoma patients. *Ann Surg Oncol.* (2019) 26:1901–8. doi: 10.1245/s10434-019-07296-w

11. Jiang C, Luo Y, Yuan J, You S, Chen Z, Wu M, et al. CT-based radiomics and machine learning to predict spread through air space in lung adenocarcinoma. *Eur Radiol.* (2020) 30:4050–7. doi: 10.1007/s00330-020-06694-z

12. Qi L, Li X, He L, Cheng G, Cai Y, Xue K, et al. Comparison of diagnostic performance of spread through airspaces of lung adenocarcinoma based on morphological analysis and perinodular and intranodular radiomic features on chest CT images. *Front Oncol.* (2021) 11:654413. doi: 10.3389/fonc.2021.654413

13. Wang Y, Lyu D, Hu L, Wu J, Duan S, Zhou T, et al. CT-based intratumoral and peritumoral radiomics nomograms for the preoperative prediction of spread through air spaces in clinical stage IA non-small cell lung cancer. *J Imaging Inform Med.* (2024) 37:520–35. doi: 10.1007/s10278-023-00939-1

14. Lin MW, Chen LW, Yang SM, Hsieh MS, Ou DX, Lee YH, et al. CT-based deep-learning model for spread-through-air-spaces prediction in ground glass-predominant lung adenocarcinoma. *Ann Surg Oncol.* (2024) 31:1536–45. doi: 10.1245/s10434-023-14565-2

15. Han X, Fan J, Zheng Y, Ding C, Zhang X, Zhang K, et al. The value of CT-based radiomics for predicting spread through air spaces in stage IA lung adenocarcinoma. *Front Oncol.* (2022) 12:757389. doi: 10.3389/fonc.2022.757389

16. Napel S, Mu W, Jardim-Perassi BV, Aerts HJWL, Gillies RJ. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. *Cancer*. (2018) 124:4633–49. doi: 10.1002/cncr.31630

17. Li J, Qiu Z, Zhang C, Chen S, Wang M, Meng Q, et al. ITHscore: comprehensive quantification of intra-tumor heterogeneity in NSCLC by multi-scale radiomic features. *Eur Radiol.* (2023) 33:893–903. doi: 10.1007/s00330-022-09055-0

18. Weinfurtner RJ, Abdalah M, Stringfield O, Ataya D, Williams A, Mooney B, et al. Quantitative changes in intratumoral habitats on MRI correlate with pathologic response in early-stage ER/PR+ HER2- breast cancer treated with preoperative stereotactic ablative body radiotherapy. *J Breast Imaging*. (2022) 4:273–84. doi: 10.1093/jbi/wbac013

19. Xu R, Yu D, Luo P, Li X, Jiang L, Chang S, et al. Do habitat MRI and fractal analysis help distinguish triple-negative breast cancer from non-triple-negative breast carcinoma. *Can Assoc Radiol J.* (2024) 75:584–92. doi: 10.1177/08465371241231573

20. Qiao J, Kang H, Ran Q, Tong H, Ma Q, Wang S, et al. Metabolic habitat imaging with hemodynamic heterogeneity predicts individual progression-free survival in high-grade glioma. *Clin Radiol.* (2024) 79:e842–53. doi: 10.1016/j.crad.2024.02.011

21. Wang H, Zeng L, Wu H, Tian J, Xie H, Zhang L, et al. Preoperative vascular heterogeneity based on dynamic susceptibility contrast MRI in predicting spatial pattern of locally recurrent high-grade gliomas. *Eur Radiol.* (2024) 34:1982–93. doi: 10.1007/s00330-023-10149-6

22. Wang S, Liu X, Wu Y, Jiang C, Luo Y, Tang X, et al. Habitat-based radiomics enhances the ability to predict lymphovascular space invasion in cervical cancer: a multi-center study. *Front Oncol.* (2023) 13:1252074. doi: 10.3389/fonc.2023.1252074

23. Mu W, Liang Y, Hall LO, Tan Y, Balagurunathan Y, Wenham R, et al. (18)F-FDG PET/CT habitat radiomics predicts outcome of patients with cervical cancer treated with chemoradiotherapy. *Radiol Artif Intell.* (2020) 2:e190218. doi: 10.1148/ryai.2020190218

24. Liu HF, Wang M, Lu YJ, Wang Q, Lu Y, Xing F, et al. CEMRI-based quantification of intratumoral heterogeneity for predicting aggressive characteristics of hepatocellular carcinoma using habitat analysis: comparison and combination of deep learning. *Acad Radiol.* (2023) 31:2346–55. doi: 10.1016/j.acra.2023.11.024

25. Zhao H, Su Y, Wang Y, Lyu Z, Xu P, Gu W, et al. Using tumor habitat-derived radiomic analysis during pretreatment (18)F-FDG PET for predicting KRAS/NRAS/BRAF mutations in colorectal cancer. *Cancer Imaging.* (2024) 24:26. doi: 10.1186/s40644-024-00670-2

26. Sujit SJ, Aminu M, Karpinets TV, Chen P, Saad MB, Salehjahromi M, et al. Enhancing NSCLC recurrence prediction with PET/CT habitat imaging, ctDNA, and integrative radiogenomics-blood insights. *Nat Commun.* (2024) 15:3152. doi: 10.1038/s41467-024-47512-0

27. Huang H, Chen H, Zheng D, Chen C, Wang Y, Xu L, Wang Y, et al. Habitat-based radiomics analysis for evaluating immediate response in colorectal cancer lung metastases treated by radiofrequency ablation. *Cancer Imaging*. (2024) 24:44. doi: 10.1186/s40644-024-00692-w

28. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

29. Xie C, Yang P, Zhang X, Xu L, Wang X, Li X, et al. Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. *EBioMedicine*. (2019) 44:289–97. doi: 10.1016/j.ebiom.2019.05.023

30. Chen L, Liu K, Zhao X, Shen H, Zhao K, Zhu W. Habitat imaging-based (18)F-FDG PET/CT radiomics for the preoperative discrimination of non-small cell lung cancer and benign inflammatory diseases. *Front Oncol.* (2021) 11:759897. doi: 10.3389/fonc.2021.759897

31. Calinkski T, Harabasz J. A dendrite method for cluster analysis. *Commun Stat.* (1974) 3:1–27.

32. Chen D, She Y, Wang T, Xie H, Li J, Jiang G, et al. Radiomics-based prediction for tumour spread through air spaces in stage I lung adenocarcinoma using machine learning. *Eur J Cardio-thoracic Surg.* (2020) 1:51–8. doi: 10.1093/ejcts/ezaa011

33. Zhuo Y, Feng M, Yang S, Zhou L, Ge D, Lu S, et al. Radiomics nomograms of tumors and peritumoral regions for the preoperative prediction of spread through air spaces in lung adenocarcinoma. *Transl Oncol.* (2020) 13:100820. doi: 10.1016/j.tranon.2020.100820

34. Lee JS, Kim EK, Kim M, Shim HS. Genetic and clinicopathologic characteristics of lung adenocarcinoma with tumor spread through air spaces. *Lung Cancer*. (2018) 123:121–6. doi: 10.1016/j.lungcan.2018.07.020

35. Koezuka S, Mikami T, Tochigi N, Sano A, Azuma Y, Makino T, et al. Toward improving prognosis prediction in patients undergoing small lung adenocarcinoma resection: Radiological and pathological assessment of diversity and intratumor heterogeneity. *Lung Cancer*. (2019) 135:40–6. doi: 10.1016/j.lungcan.2019.06.023

36. Niu YY, Han XH, Zeng Y, Nanding A, Bai Q, Guo S, et al. The significance of spread through air spaces in the prognostic assessment model of stage I lung adenocarcinoma and the exploration of its invasion mechanism. *J Cancer Res Clin Oncol.* (2023) 10:7125–38. doi: 10.1007/s00432-023-04619-z

37. Tominaga M, Yamazaki M, Umezu H, Sugino H, Fuzawa Y, Yagi T, et al. Prognostic value and pathological correlation of peritumoral radiomics in surgically resected non-small cell lung cancer. *Acad Radiol.* (2024) 23:S1076-6332(24)00055-2. doi: 10.1016/j.acra.2024.01.033

38. Takehana K, Sakamoto R, Fujimoto K, Matsuo Y, Nakajima N, Yoshizawa A, et al. Peritumoral radiomics features on preoperative thin-slice CT images can predict the spread through air spaces of lung adenocarcinoma. *Sci Rep*. (2022) 12:10323. doi: 10.1038/s41598-022-14400-w

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 October 2024
DOI 10.3389/fonc.2024.1427743

Check for updates

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Liyuan Zhang,
Second Affiliated Hospital of Soochow University, China
Shuheng Bai,
The First Affiliated Hospital of Xi'an Jiaotong University, China

*CORRESPONDENCE
Linyou Zhang
lyzhang@hrbmu.edu.cn

RECEIVED 04 May 2024
ACCEPTED 18 September 2024
PUBLISHED 11 October 2024

CITATION
Xu R, Wang K, Peng B, Zhou X, Wang C, Lu T, Shi J, Zhao J and Zhang L (2024) Evaluating peritumoral and intratumoral radiomics signatures for predicting lymph node metastasis in surgically resectable non-small cell lung cancer.
Front. Oncol. 14:1427743.
doi: 10.3389/fonc.2024.1427743

COPYRIGHT
© 2024 Xu, Wang, Peng, Zhou, Wang, Lu, Shi, Zhao 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.

Evaluating peritumoral and intratumoral radiomics signatures for predicting lymph node metastasis in surgically resectable non-small cell lung cancer

Ran Xu[1,2], Kaiyu Wang[1,2], Bo Peng[1,2], Xiang Zhou[1,2], Chenghao Wang[1,2], Tong Lu[3], Jiaxin Shi[1,2], Jiaying Zhao[1,2] and Linyou Zhang[1]*

[1]Department of Thoracic Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]The Second Clinical Medical College, Harbin Medical University, Harbin, China, [3]Department of Thoracic Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: Whether lymph node metastasis in non-small cell lung cancer is critical to clinical decision-making. This study was to develop a non-invasive predictive model for preoperative assessing lymph node metastasis in patients with non-small cell lung cancer (NSCLC) using radiomic features from chest CT images.

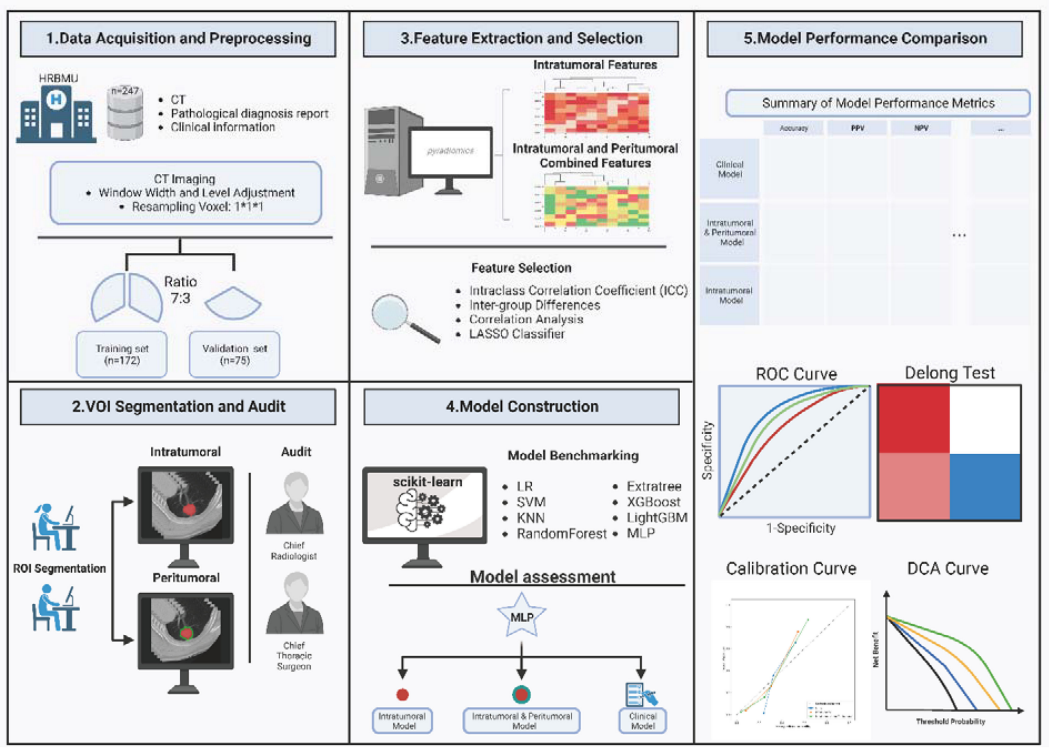
Materials & methods: In this retrospective study, 247 patients with resectable non-small cell lung cancer (NSCLC) were enrolled. These individuals underwent preoperative chest CT scans that identified lung nodules, followed by lobectomies and either lymph node sampling or dissection. We extracted both intratumoral and peritumoral radiomic features from the CT images, which were used as covariates to predict the lymph node metastasis status. By using ROC curves, Delong tests, Calibration curve, and DCA curves, intra-tumoral-peri-tumoral model performance were compared with models using only intratumoral features or clinical information. Finally, we constructed a model that combined clinical information and radiomic features to increase clinical applicability.

Results: This study enrolled 247 patients (117 male and 130 females). In terms of predicting lymph node metastasis, the intra-tumoral-peri-tumoral model (0.953, 95%CI 0.9272-0.9792) has a higher AUC compared to the intratumoral radiomics model (0.898, 95%CI 0.8553-0.9402) and the clinical model (0.818, 95%CI 0.7653-0.8709). The DeLong test shows that the performance of the Intratumoral and Peritumoral radiomics models is superior to that of the Intratumoral or clinical feature model ($p < 0.001$). In addition, to increase the clinical applicability of the model, we combined the intratumoral-peritumoral model and clinical information to construct a nomogram. Nomograms still have good predictive performance.

Conclusion: The radiomics-based model incorporating both peritumoral and intratumoral features from CT images can more accurately predict lymph node metastasis in NSCLC than traditional methods.

KEYWORDS

peritumoral, intratumoral, radiomics, lymph node metastasis, non-small cell lung cancer



GRAPHICAL ABSTRACT

Introduction

Lung cancer is a prominent cause of cancer-related deaths, with non-small cell lung cancer (NSCLC) accounting for a significant proportion (1). For stages IA to IIIA localized lymph node metastasis, surgery is the primary treatment approach (2); however, for advanced stage IIIB-IV NSCLC that is not resectable through surgery, chemoradiotherapy, immune therapy, or targeted therapy are necessary. Hence, accurate staging is critical for effective NSCLC treatment. Clinical staging typically begins with chest CT scans, which may suggest an indication for lymph node metastasis. Recently, mediastinoscopic lymph node biopsy or endobronchial ultrasound-guided fine needle aspiration have been proposed as essential methods for definitive staging (3). However, these methods are associated with high costs and potential complications. Therefore, there is an urgent need to develop non-invasive methods for preoperative staging of surgically resectable NSCLC and guide the treatment of NSCLC patients. In addition, PET/CT is also considered an important basis for determining staging, but CT is currently the preferred imaging examination for patients with newly treated pulmonary nodules (4, 5), so the model developed using CT images is more universal.

Radiomics is a powerful method that can transform CT images into high-throughput quantifiable data (6). The combination of these radiomic features with machine learning algorithms to construct clinical prediction models is a promising research direction (7). Prior studies have suggested that radiomic features are useful in predicting patient overall survival, pathological response to adjuvant therapy, or lymph node metastasis status in various cancers (8–12), including NSCLC. However, most studies focused only on intratumoral radiomic features and neglected peritumoral areas despite their importance in encompassing information related to tumor progression and evolution

Abbreviations: NSCLC, non-small cell lung cancer; CT, computed tomography; KNN, K-Nearest Neighbor; SVM, Support Vector Machine; MLP, Multilayer Perceptron.

influenced by interactions between intra-tumoral cells and elements in the peritumoral region, such as lymphatic or vascular invasion and angiogenesis (10, 13–15). Previous radiomics models of lymph node metastasis in non-small cell lung cancer usually only focused on the characteristics of the tumor itself without introducing the characteristics of the surrounding tissue, which may lead to limitations in model performance (16–20). The metastatic status of lymph nodes is not only affected by the characteristics of the primary tumor, such as tumor size, shape, density, etc., but is also related to changes in the tumor microenvironment, including surrounding tissues. Surrounding tissue characteristics can provide important information about the interaction of the tumor with its microenvironment. For example, changes in texture of surrounding tissue, abnormalities in local blood vessel density, irregularities in tumor boundaries, and the degree of tumor infiltration may indicate tumor invasiveness and metastatic potential. Therefore, a comprehensive consideration of radiomic features of surrounding tumor tissues may better characterize tumor invasiveness and metastatic potential. Constructing a comprehensive, non-invasive preoperative radiomic prediction model for lymph node metastasis status in surgically resectable NSCLC is essential.

In this study, we collected CT images and clinical data from 247 patients with surgically resectable NSCLC. We established three preoperative lymph node metastasis prediction models using intratumoral radiomic features, a combination of peritumoral and intratumoral radiomic features, and clinical characteristics. We compared the predictive performance of these models, highlighting the importance of peritumoral radiomic features in predicting lymph node metastasis status. Finally, we constructed a nomogram with wider applicability by fusing the best-performing intratumoral-peritumoral radiomics model and clinical indicators. Our results may provide a new perspective for non-invasive preoperative lymph node diagnosis.

Materials and methods

Data acquisitions

The chest CT images, and clinical information of patients used in this study were obtained from our center, and the study was approved by the Ethics Committee (Approval No.: KY2022-144). The acquisition of written informed consent from patients was waived because of the retrospective design. All data usage was performed in accordance with the supervision of the Ethics Committee. The cases included in the study were diagnosed as surgically resectable NSCLC at our hospital between 2019 and 2020. Lymph node metastasis status and pathological diagnosis were determined through postoperative pathological examination. Lymph node dissection strategies for all patients were consistent with current recommendations from The American Association for Thoracic Surgery (21). The exact pathological classification was determined by multiple pathological experts in our center after surgery and a formal medical diagnosis was issued. The detailed inclusion criteria were as follows (1): Patients with pulmonary nodules diagnosed by chest CT; (2) Patients with chest CT images available in Digital Imaging and Communications in Medicine (DICOM) format; (3) Patients who underwent pulmonary lobectomy and systematic lymph node sampling or dissection; (4) Patients with confirmed postoperative histopathology of primary NSCLC with clear lymph node metastasis status for each station; (5) Patients who did not receive neoadjuvant treatment or chemoradiotherapy in the past; (6) Patients who did not have previous thoracic surgery. All CT images were acquired using a 64-channel CT scanner (Discovery 750, GE Healthcare, Milwaukee, USA) with the following scanning parameters: tube voltage of 120kV, tube current of 100-250mAs, layer thickness of 0.625-5 mm, field of view (FOV) of 350-400 mm, 512 x 512 matrix, and reconstructed layer thickness of 0.625-5 mm. The CT images were reconstructed using filtered back-projection (FBP) and adaptive statistical iterative reconstruction (ASIR) at a level of 40% ASIR. The standard kernel was used for the reconstruction. In total, 247 cases were included in the study, with 87 positive and 160 negatives for lymph node metastasis. In addition, we extracted the CT imaging data of lung adenocarcinoma patients (n=219) from two datasets in The Cancer Imaging Archive (TCIA) database for external validation of the radiomics model (https://www.cancerimagingarchive.net/browse-collections/, Collection label: NSCLC-Radiomics-Genomics and NSCLC Radiogenomics).

CT image data preprocessing and ROI segmentation

To eliminate batch differences between CT image data, all CT data were first adjusted to a window width of -150 and a window level of 1700. The voxel spacing was then adjusted to $1 \times 1 \times 1$ using the nearest interpolation algorithm to account for different scanning parameters and image resolutions (22). Grayscale discretization was performed using a fixed bin width of 25 Hounsfield Units (HU All chest CTs were loaded into ITK-SNAP software (23) (version: 3.8.0) for tumor region segmentation by two thoracic surgeons with over 5 years of clinical experience. The segmented tumor area was examined layer by layer and revised by a chief thoracic surgeon and a radiologist.

Extraction of radiomic features from VOI

The volume of interest (VOI) comprised intratumoral and extratumoral regions. The intratumoral region was manually segmented layer by layer, while the extratumoral region was obtained by extending 3 voxels around the intratumoral VOI. Radiomic features were extracted using the 'pyradiomics' package (24) in Python. To capture a comprehensive set of features, we extracted features under multiple filters, including Original, Wavelet, Square, SquareRoot, Logarithm, Gradient, Exponential, LBP3D, and Laplacian of Gaussian filter (LoG). For the LoG filter,

we used sigma values of 1.0, 2.0, and 3.0 to enhance texture recognition for fine and rough textures. Under these filters, we extracted radiomic features from the intratumoral and peritumoral regions, including First Order Features, Shape Features, Gray Level Co-occurrence Matrix (GLCM) Features, Gray Level Zone Size Matrix (GLSZM) Features, Gray Level Run Matrix (GLRLM) Features, Neighbungay Gray Tone Difference Matrix (NGTDM) Features, and Gray Level Features (GLDM) Features. The detailed definition methods of all features can be found at https://pyradiomics.readthedocs.io/en/latest/features.html#, and all defined features comply with the Imaging Biomarker Standardization Initiative (IBSI) (25).

Feature preprocessing and screening

The radiomic features of VOI delineated by different physicians were assessed for consistency using the Interclass correlation coefficients method (ICC) (26). Variables with an agreement greater than 0.75 between groups were considered reliable imaging features. Subsequently, the features were normalized using the z-score method. To identify radiomic features associated with lymph node metastasis, we performed a t-test between the negative and positive lymph node metastasis groups. Variables with a p-value< 0.05 were retained for further screening (27). Furthermore, to address collinearity among variables, Pearson correlation analysis was conducted to examine variable correlations. In cases where correlations exceeded 0.9, one of the paired variables was randomly chosen for model development. Additionally, the LASSO classifier was employed to mitigate collinearity and screen variables relevant to lymph node metastasis. Variables with non-zero coefficients in the LASSO model were ultimately used in subsequent machine learning model construction (28).

Model construction

We utilized the presence or absence of lymph node metastasis as the target for prediction, with the characteristics obtained through the aforementioned screening strategy serving as covariates. The model was developed using the Scikit-learn (29) framework. For the intratumoral radiomics feature model, we benchmarked several algorithms, including Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Random Forests, Extremely Randomized Trees (ExtraTrees), XGBoost, Light Gradient Boosting Machine (LightGBM), Multi-Layer and Perceptron (MLP), to select the best algorithm. Performance indicators such as Accuracy, area under the curve (AUC), Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) were calculated based on the prediction results of each model. After evaluating the combined performance of the models on the training and test sets, the MLP model was determined to be the most suitable. To ensure model comparability, MLP models were also constructed for subsequent intratumoral-peritumoral combined models and clinical feature models.

Model performance comparison

The DeLong test was employed to compare the AUC values between different models. Additionally, the model accuracy, sensitivity, specificity, precision, recall, NPV (Negative Predictive Value), PPV (positive predictive value) and F1 score were also used for evaluating. The Decision Curve Analysis was conducted to assess the potential clinical benefit performance of the model.

Nomogram construction

The nomogram was constructed using the R-based rms package. We integrated the radiomics model scores and clinical risk factors for lymph node metastasis and used logistic regression formulas to construct the final nomogram. The DeLong test was used to evaluate the performance of the AUC of the nomogram compared to other models. Calibration curve and DCA curve are used to evaluate the clinical benefit performance of the nomogram.

Results

Clinical characteristics of patients

This study enrolled 247 patients with surgically resectable NSCLC, and their chest CT images, and clinical information were collected. All patients underwent lobectomy or segmentectomy with systematic lymph node dissection. Among them, 87 patients had lymph node metastasis confirmed by postoperative pathology, while 160 patients had no lymph node metastasis. Table 1 presents the clinical information of all included patients. Significant differences were observed between the lymph node metastasis group and the group without lymph node metastasis in terms of maximum tumor diameter (p<0.001) and smoking history (p<0.001). The total of 247 patients were randomly divided into training and validation cohorts, consisting of 172 and 75 patients respectively, following a 7:3 ratio. There was no statistical difference in various clinical information between the training group and the validation group (Table 2).

Feature extraction of intratumoral radiomics and feature selection

After performing tumor region segmentation on the collected chest CT images of all patients (details provided in the Methods section), a total of 1874 intratumoral radiomic features were extracted using different filters. These features included 360 First Order Features, 14 Shape Features, 480 GLCM Features, 320 GLSZM Features, 320 GLRLM Features, 100 NGTDM Features, and 280 GLDM Features (Figures 1A, B). Subsequently, we performed statistical analysis between the groups with and without lymph node metastasis to screen for relevant radiomics features. We retained 1383 features with a p-value < 0.05 for further analysis (Figure 1C). To address collinearity among variables,

TABLE 1 Clinical characteristics of patients.

Covariates		All (n=247)	Without lymph node metastasis (n=160)	Lymph node metastasis (n=87)	*p*-value
Age		59.99 ± 9.06	59.9 4± 9.31	60.08 ± 8.64	0.91
Gender (%)	Female	130 (52.6)	89 (55.6)	41 (47.1)	0.252
	Male	117 (47.4)	71 (44.4)	46 (52.9)	
Location (%)	LLL*	43 (17.4)	27 (16.9)	16 (18.4)	0.525
	LUL*	65 (26.3)	41 (25.6)	24 (27.6)	
	RLL*	55 (22.3)	33 (20.6)	22 (25.3)	
	RML*	13 (5.3)	11 (6.9)	2 (2.3)	
	RUL*	71 (28.7)	48 (30.0)	23 (26.4)	
Histology (%)	LUAD†	203 (82.2)	134 (83.8)	69 (79.3)	0.486
	LUSC†	44 (17.8)	26 (16.2)	18 (20.7)	
Maximum Tumor Diameter (mean ± SD)		25.11 ± 15.43	20.16 ± 14.30	34.20 ± 13.18	<0.001
Smoking History (%)	No	130 (52.6)	103 (64.4)	27 (31.0)	<0.001
	Yes	117 (47.4)	57 (35.6)	60 (69.0)	

*LLL, Left Lower Lobe; LUL, Left Upper Lobe; RLL, Right Lower Lobe; RML, Right Middle Lobe; RUL, Right Upper Lobe. †LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinoma.

TABLE 2 Clinical characteristics of training and testing sets.

Covariates		All (n=247)	Testing set (n=75)	Training set (n=172)	*p*-value
Age		59.99 ± 9.06	60.93 ± 8.83	59.58 ± 9.16	0.282
Gender (%)	Female	130 (52.6)	39 (52.0)	91 (52.9)	1
	Male	117 (47.4)	36 (48.0)	81 (47.1)	
Location (%)	LLL*	43 (17.4)	10 (13.3)	33 (19.2)	0.752
	LUL*	65 (26.3)	20 (26.7)	45 (26.2)	
	RLL*	55 (22.3)	19 (25.3)	36 (20.9)	
	RML*	13 (5.3)	5 (6.7)	8 (4.7)	
	RUL*	71 (28.7)	21 (28.0)	50 (29.1)	
Histology (%)	LUAD†	203 (82.2)	62 (82.7)	141 (82.0)	1
	LUSC†	44 (17.8)	13 (17.3)	31 (18.0)	
Maximum Tumor Diameter (mean ± SD)		25.11 ± 15.43	24.80 ± 15.38	25.24 ± 15.50	0.838
Smoking History (%)	No	130 (52.6)	37 (49.3)	93 (54.1)	0.584
	Yes	117 (47.4)	38 (50.7)	79 (45.9)	
Lymph node metastasis (%)	No	160 (64.8)	44 (58.7)	116 (67.4)	0.237
	Yes	87 (35.2)	31 (41.3)	56 (32.6)	

*LLL, Left Lower Lobe; LUL, Left Upper Lobe; RLL, Right Lower Lobe; RML, Right Middle Lobe; RUL, Right Upper Lobe. †LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinoma.

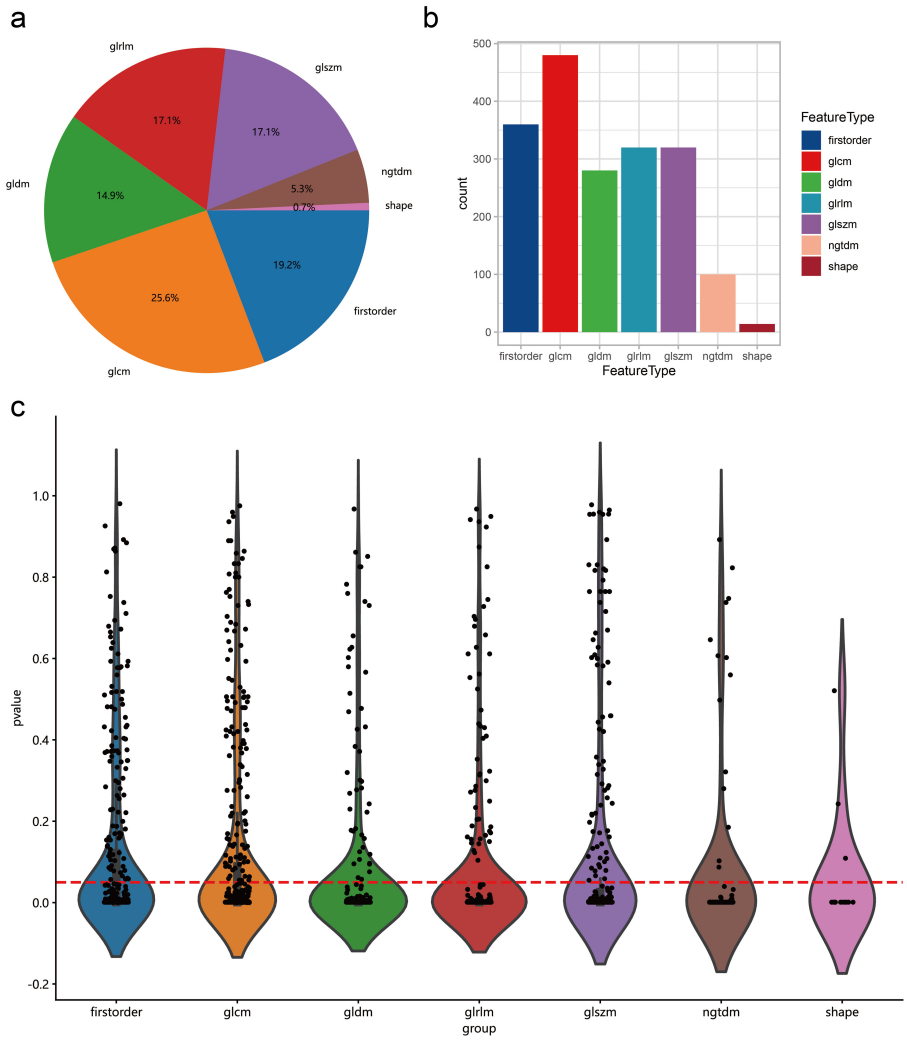


FIGURE 1
Distribution of intratumoral radiomics features data. **(A)** Proportions of different types of radiomic features. **(B)** Number of different types of radiomics features. **(C)** T-test results comparing different types of radiomic features between lymph node metastasis and non-metastasis groups. The horizontal axis represents different feature types, and the vertical axis represents p-values. Features below the red dashed line indicate a p-value <0.05.

Pearson correlation analysis was conducted on these 1383 variables. Based on the experience of previously published articles (30), for variables with correlations greater than 0.9, one variable was randomly retained, resulting in 271 variables being finally retained.

Model construction of intratumoral radiomic features

In the training group, we employed the LASSO classifier to further identify variables that significantly influenced lymph node metastasis for constructing subsequent models. When λ reaches the optimal value of 0.0339, fourteen variables with non-zero coefficients were retained for subsequent modeling (Figures 2A–C). Various machine learning algorithms including SVM, KNN, Random Forests, ExtraTrees, LightGBM, and MLP were benchmarked to select the best algorithm. All models underwent 5-fold cross-validation in entire dataset. In 5-fold cross-validation, the MLP model showed the highest median AUC (Supplementary Figure 1A). Therefore, the MLP model may be a potentially optimal model. In subsequent modeling, we use the training set for training and the testing set to evaluate the final metrics of the model. (Table 3). The accuracy, AUC values and F1 scores of the models were visualized for testing set (Figures 2D–F). It is worth noting that among all models, although SVM has the highest AUC, the MLP model has higher accuracy and F1 Score. Since there is a certain imbalance between the two categories in our data, F1 scores is more robust to evaluating the performance of the model than AUC. Therefore, we chose the MLP model as the final model (Figure 2G). The MLP model constructed using intratumoral radiomic features achieved AUC values of 0.912 (95% CI: 0.871-0.953) and 0.877 (95% CI: 0.779-0.975) in the training and testing sets, respectively, for predicting lymph node metastasis.

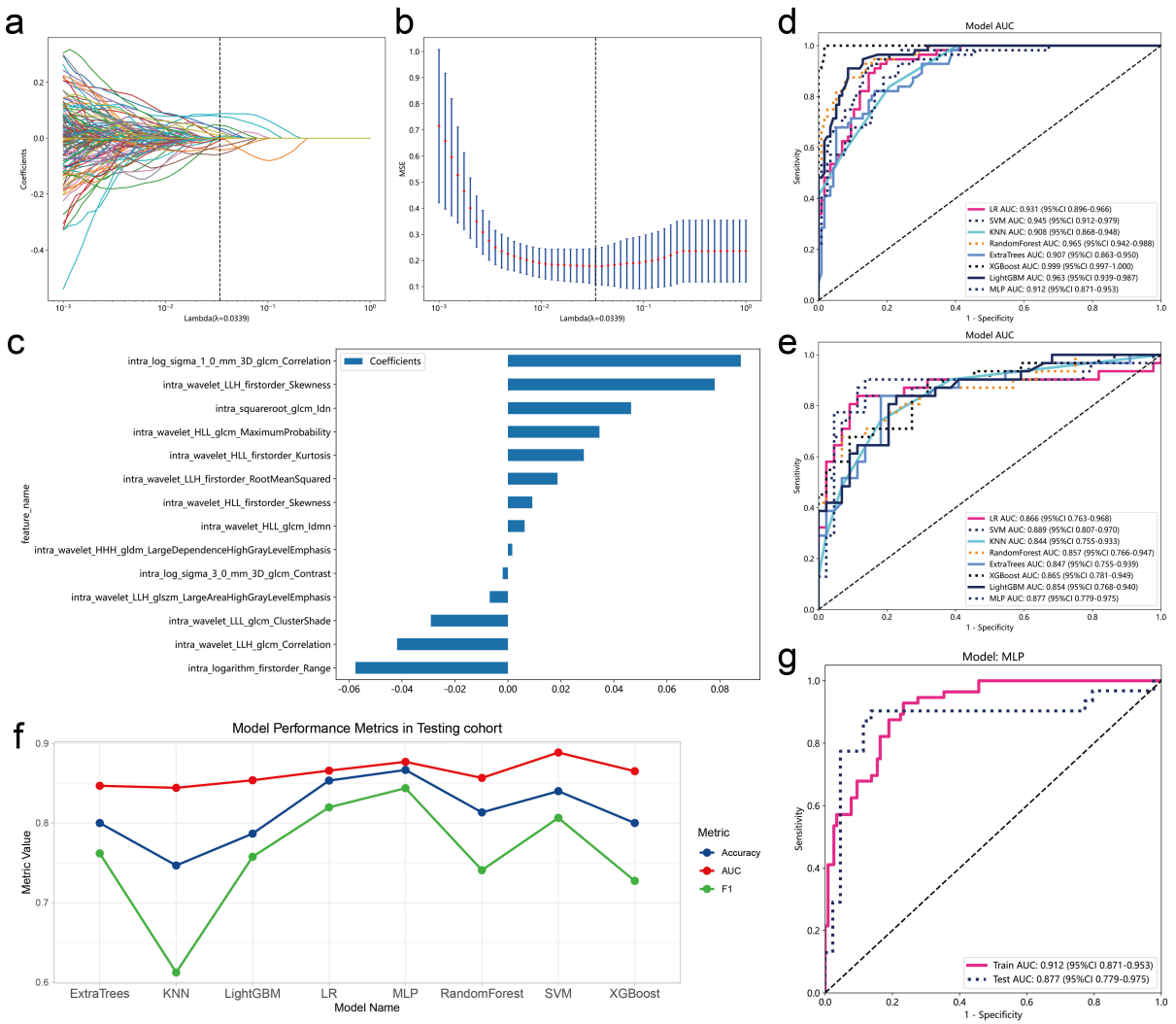


FIGURE 2
Screening and model construction of intratumoral radiomics features. The LASSO classifier is used for feature selection. **(A)** Coefficient trajectory plot for various Lambda values. **(B)** Mean squared error (MSE) of the model at different Lambda values. The black dashed line indicates the optimal Lambda value determined by minimizing the MSE. **(C)** Coefficient plot for variables in the model at the optimal Lambda value. The X-axis represents model coefficients. Model metrics **(D)** and ROC curves **(E, F)** of the model on the training and testing sets of different machine learning models. **(G)** ROC curve of the MLP model, determined as the optimal model, on the training and testing sets.

Radiomics feature extraction and feature selection for intratumoral and peritumoral combination

To investigate the impact of peritumoral radiomics features on the predictive ability of lymph node metastasis, we extended manually delineated tumor ROI regions outwards by 3 individual voxels to obtain peritumoral regions by referring to methods in previously published articles (8, 31). Peritumoral features were extracted using the same method as intratumoral feature extraction and merged with intratumoral features for subsequent feature selection. After combination, we obtained a total of 3748 intratumoral and peritumoral radiomics features. Similarly, we screened and obtained 2572 features with p-value < 0.05 in lymph node metastasis and non-lymph node metastasis samples (Figures 3A–C). As in the modeling of intratumoral features only, we performed Pearson correlation analysis on these combined features and retained one variable from paired variables whose correlation between variables was greater than 0.9. Eventually, we selected 539 variables for subsequent construction of intratumoral and peritumoral prediction models.

Model construction of intratumoral and peritumoral combination radiomic features

Consistent with the modeling strategy using only intratumoral features, we employed the LASSO classifier to screen combined intratumoral and peritumoral features and used the MLP algorithm to test the impact of peritumoral features on the model for variables with absolute values of coefficients greater than zero. After LASSO classifier screening, we finally obtained 64 variables with non-zero

TABLE 3 Model performance of intratumoral model.

Model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV*	NPV*	Precision	Recall	F1	Cohort
LR	0.843	0.931	0.8956 - 0.9661	0.929	0.802	0.693	0.959	0.693	0.929	0.794	Train
	0.853	0.866	0.7633 - 0.9684	0.806	0.886	0.833	0.867	0.833	0.806	0.820	Test
SVM	0.860	0.945	0.9118 - 0.9789	0.929	0.828	0.722	0.960	0.722	0.929	0.812	Train
	0.840	0.889	0.8071 - 0.9700	0.806	0.864	0.806	0.864	0.806	0.806	0.806	Test
KNN	0.814	0.908	0.8684 - 0.9482	0.571	0.931	0.800	0.818	0.800	0.571	0.667	Train
	0.747	0.844	0.7551 - 0.9333	0.484	0.932	0.833	0.719	0.833	0.484	0.612	Test
RandomForest	0.872	0.965	0.9418 - 0.9880	0.929	0.845	0.743	0.961	0.743	0.929	0.825	Train
	0.813	0.857	0.7662 - 0.9471	0.645	0.932	0.870	0.788	0.870	0.645	0.741	Test
ExtraTrees	0.826	0.907	0.8635 - 0.9498	0.804	0.836	0.703	0.898	0.703	0.804	0.750	Train
	0.800	0.847	0.7547 - 0.9388	0.774	0.818	0.750	0.837	0.750	0.774	0.762	Test
XGBoost	0.983	0.999	0.9965 - 1.0000	0.982	0.983	0.965	0.991	0.965	0.982	0.973	Train
	0.800	0.865	0.7810 - 0.9492	0.645	0.909	0.833	0.784	0.833	0.645	0.727	Test
LightGBM	0.907	0.963	0.9394 - 0.9870	0.893	0.914	0.833	0.946	0.833	0.893	0.862	Train
	0.787	0.854	0.7675 - 0.9399	0.806	0.773	0.714	0.850	0.714	0.806	0.758	Test
MLP	0.814	0.912	0.8710 - 0.9532	0.911	0.767	0.654	0.947	0.654	0.911	0.761	Train
	0.867	0.877	0.7790 - 0.9746	0.871	0.864	0.818	0.905	0.818	0.871	0.844	Test

*NPV, Negative Predictive Value; PPV, positive predictive value.

10.3389/fonc.2024.1427743

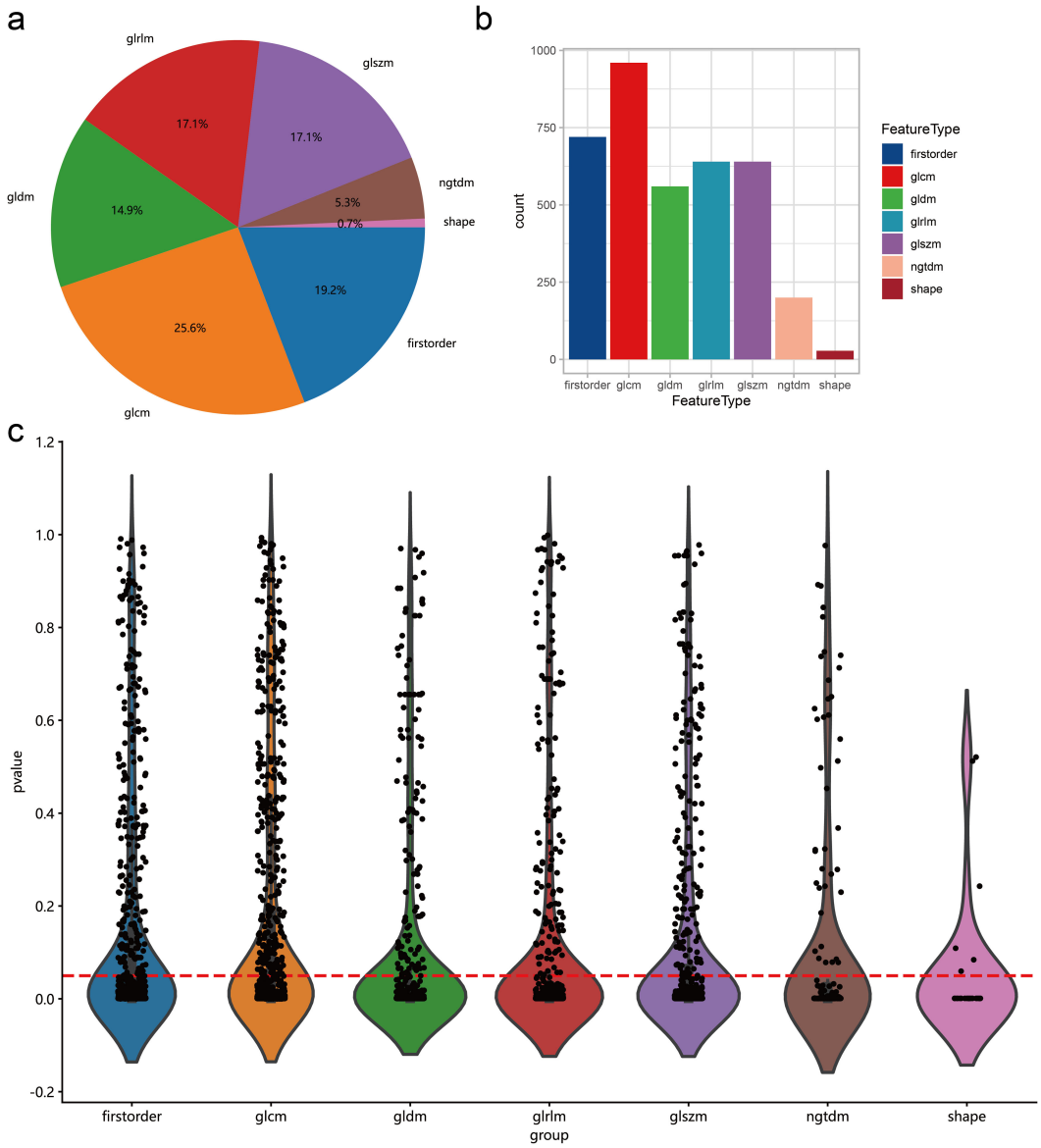


FIGURE 3
Distribution of intratumoral and peritumoral radiomics features data. **(A)** Proportions of different types of radiomic features. **(B)** Number of different types of radiomics features. **(C)** T-test results comparing different types of radiomic features between lymph node metastasis and non-metastasis groups. The horizontal axis represents different feature types, and the vertical axis represents p-values. Features below the red dashed line indicate a p-value <0.05.

coefficients (Figures 4A–C). Using the MLP algorithm to predict lymph node metastasis status, AUC values of the model reached 0.977 (95%CI: 0.960-0.994) in the training set and 0.905 (95%CI: 0.833-0.977) in the testing set (Figure 4D). In addition, since the SVM model of intratumoral radiomics features showed the highest accuracy (Figure 2F), we also tried the SVM modeling after combining intratumoral and peritumoral features. In the test group, the SVM model of intratumoral and peritumoral features had an AUC of 0.876 (95% CI 0.792 – 0.960) (Supplementary Figure 2A), which was lower than the MLP model. In addition, the F1 score of the SVM model in testing set is also lower than that of the MLP model (Table 4). Finally, the MLP model was identified as the suitable model of intratumoral and peritumoral features. Overall, compared to the AUC in the MLP model when using only intratumoral features (Figure 2G), the predictive performance of the MLP model improved to a certain extent upon inclusion of peritumoral features. In addition, we validated the intratumor-peritumoral feature MLP model in an external dataset. The AUC of the external dataset was 0.812 (95% CI, 0.741-0.884), which showed that the model had good robustness (Supplementary Figure 3A).

Clinical model construction

Clinical information as an important indicator of current clinical assessment, we collected patients' clinical information,

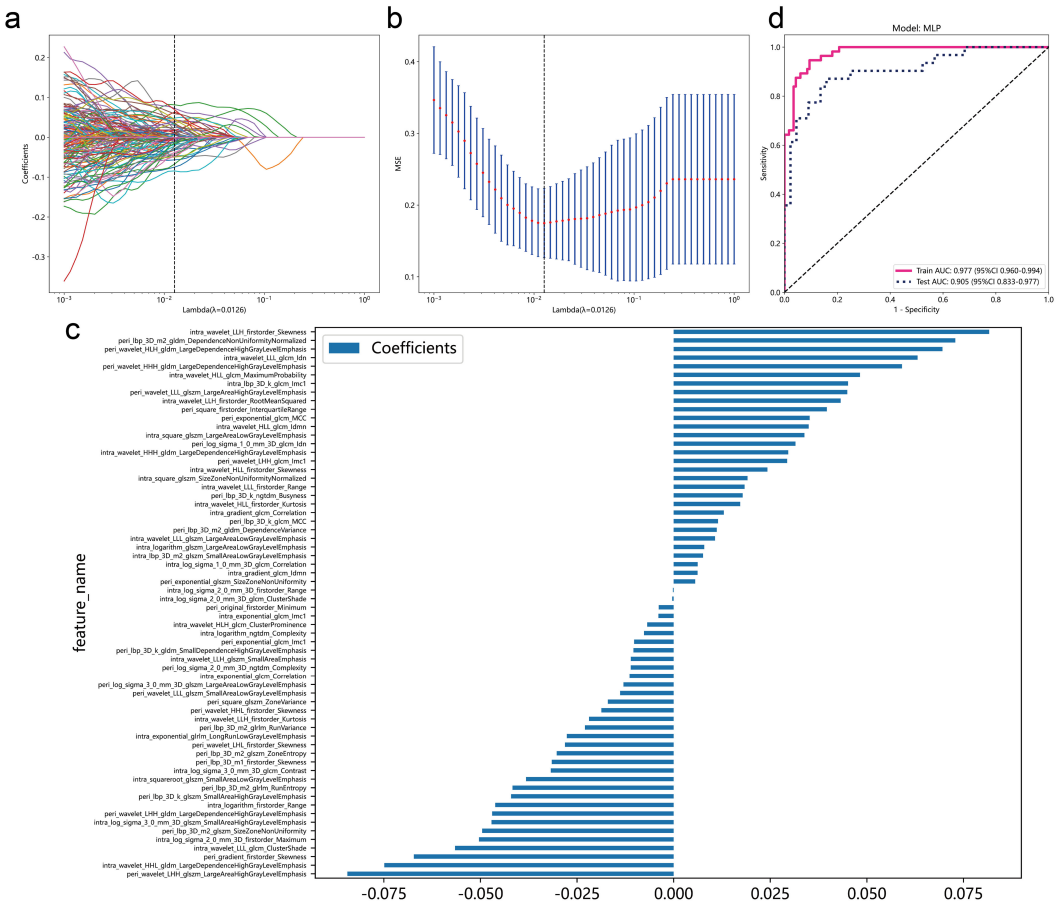


FIGURE 4
Screening and model construction of intratumoral and peritumoral radiomics features. The LASSO classifier is used for feature selection. **(A)** Coefficient trajectory plot for various Lambda values. **(B)** Mean squared error (MSE) of the model at different Lambda values. The black dashed line indicates the optimal Lambda value determined by minimizing the MSE. **(C)** Coefficient plot for variables in the model at the optimal Lambda value. The X-axis represents model coefficients. **(D)** ROC curve of the MLP model on the training and testing sets.

including age, gender, preoperative CT diagnosis of tumor location, smoking history, maximum tumor diameter, pathological types of tumors diagnosed by intraoperative frozen sections and a series of laboratory indicators. Univariate logistic regression was applied to these clinical details to assess their risks for lymph node metastasis. Results indicated that maximum tumor diameter, smoking history, CA153, CA125 and CEA were risk factors for lymph node metastasis (Figure 5A). Additionally, we conducted multivariate logistic regression on these risk factors. In the multivariate analysis, smoking history and maximum tumor diameter were considered independent risk factors for lymph node metastasis (Figure 5B, Supplementary Table 1). To ensure the comparability of the model, we also used the MLP method to construct a lymph node metastasis prediction model using these independent risk factors, and the results showed that the clinical information MLP model had an AUC of 0.810 (95%CI: 0.747-0.872) and 0.846 (95%CI: 0.749-0.943) in the training and validation groups, respectively (Figure 5C).

Model performance evaluation

To evaluate the performance of the three models we constructed above (intratumoral peritumoral radiomics model, intratumoral radiomics model and clinical information model), we compared various model indicators in the entire data set. Radiomic signatures of intratumoral and peritumoral regional features still exhibit higher area under the ROC curve than intratumoral features or clinical feature models alone and DeLong test shows that p values are all <0.05 (Figures 6A, B). Furthermore, when we summarized other evaluation metrics for the three models, we found that the intratumoral and peritumoral feature models were overall higher than other models in terms of accuracy, sensitivity, specificity, precision, recall, NPV (Negative Predictive Value), PPV (positive predictive value) and F1 score (Figure 6C, Table 5). Clinical decision curves further demonstrated that the model combining intratumoral and peritumoral radiomics features potentially offers

10.3389/fonc.2024.1427743

TABLE 4 Model performance of intratumoral-peritumoral model.

Model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV*	NPV*	Precision	Recall	F1	Cohort
SVM	0.948	0.993	0.9851 - 1.0000	0.946	0.948	0.898	0.973	0.898	0.946	0.922	Train
	0.827	0.876	0.7921 - 0.9601	0.774	0.864	0.800	0.844	0.800	0.774	0.787	Test
MLP	0.913	0.977	0.9595 - 0.9937	0.929	0.905	0.825	0.963	0.825	0.929	0.874	Train
	0.840	0.905	0.8327 - 0.9766	0.839	0.841	0.788	0.881	0.788	0.839	0.812	Test

*NPV, Negative Predictive Value; PPV, positive predictive value.

greater clinical benefits (Figure 6D). In order to further evaluate the stability of the predictive performance of the three models across the entire dataset, we adopted 5-fold cross-validation. The results demonstrate that the model incorporating both intratumoral and peritumoral features still exhibits the highest average AUC value. (Supplementary Figures 4A–C) In summary, incorporating peritumoral radiomic features into the prediction model improves the performance of the model in predicting lymph node metastasis and may provide potential clinical benefits. Therefore, in this study the results show that including peritumoral features in radiological models is necessary for predicting lymph node metastasis.

Construction of lymph node metastasis nomogram of lung adenocarcinoma

Because of the importance and usefulness of clinical information in clinical assessment, we should consider it. Therefore, we further constructed a nomogram by integrating intratumoral peritumoral radiomics models and clinical risk indicators of lymph node metastasis to increase clinical applicability (Figure 7A). The AUC of the nomogram in the entire cohort was 0.947 (95% CI 0.920-0.974) (Figure 7B). DeLong test showed that the AUC of the nomogram was higher than the clinical features model ($p<0.001$) and comparable to the intratumoral peritumoral radiomics model ($p=0.212$) (Figure 7C). The calibration curves and Decision Curve Analysis still show that the nomogram has comparable performance to the intratumoral peritumoral radiomics model and may bring potential clinical benefit (Figures 7D, E).

Discussion

Accurate clinical staging is crucial for determining treatment strategies in newly diagnosed NSCLC patients, particularly for surgically resectable patients. Accurate N staging can guide decisions regarding preoperative neoadjuvant therapy (32) or intraoperative lymph node dissection strategies (33). Radiomics is a promising non-invasive diagnostic approach for lymph node N staging compared to mediastinoscopic lymph node biopsy or endobronchial ultrasound-guided fine needle aspiration. At present, lymph node staging often relies on intraoperative sampling. However, the diagnosis based on preoperative radiomics is promising. In this study, we established a MLP model combining the intratumoral and peritumoral radiomics characteristics of NSCLC on CT images, which was superior to traditional intratumoral feature models and clinical information models in predicting lymph node metastasis status preoperatively. In addition, the Nomogram integrating clinical information and intratumoral-peritumoral radiomics still have a good predictive ability.

Although machine learning models based on intratumoral radiomics profiles have proved effective in predicting lymph node metastasis status (16, 34, 35), peritumoral profiles have received limited attention in research. The peritumoral area provides

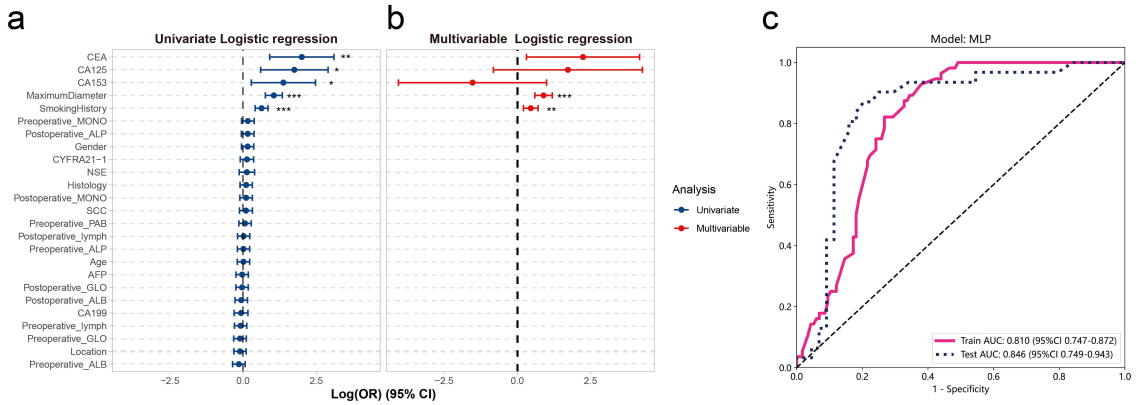


FIGURE 5
Clinical feature selection and model construction. Univariate **(A)** and multivariate **(B)** logistic regression of clinical features, with the x-axis representing log2-transformed odds ratios. **(C)** ROC curve of the MLP model for clinical features. In **(A, B)** "*" indicates *P*-value < 0.05, "**" indicates *P*-value < 0.01, and "***" indicates *P*-value < 0.001.

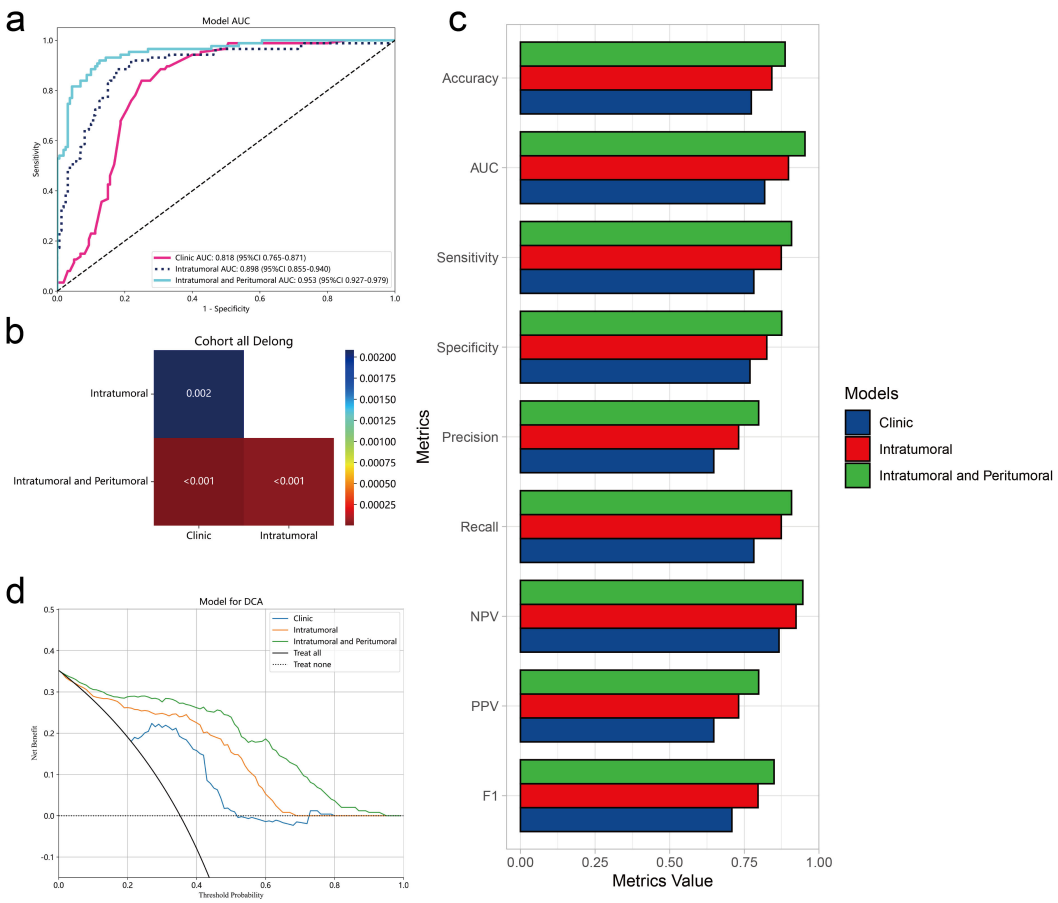


FIGURE 6
Comparison of multiple model performance. **(A)** ROC curves of the clinical feature model, intratumoral radiomic feature model, and combined intratumoral and peritumoral radiomic feature model in the entire dataset. **(B)** DeLong test for AUC values of the ROC curves of the three models. **(C)** The comparison of all performance metrics of the three models. **(D)** Decision Curve Analysis (DCA) curves of the three models in the entire dataset.

10.3389/fonc.2024.1427743

TABLE 5 Model metrics comparison.

Model name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV*	NPV*	Precision	Recall	F1	Cohort
Clinic	0.773	0.818	0.7653 - 0.8709	0.782	0.769	0.648	0.866	0.648	0.782	0.708	Entire
Intratumoral	0.842	0.898	0.8553 - 0.9402	0.874	0.825	0.731	0.923	0.731	0.874	0.796	Entire
Intratumoral and Peritumoral	0.887	0.953	0.9272 - 0.9792	0.908	0.875	0.798	0.946	0.798	0.908	0.849	Entire

*NPV, Negative Predictive Value; PPV, positive predictive value.

insights into tumor infiltration and the invasion of microvessels and lymphatic vessels, making the radiomic characteristics of this region crucial (36, 37). Additionally, manual segmentation and measurement of the tumor area often result in unstable features at the tumor edge. Due to the irregular shape of the tumor edge area, parts of the tumor margin area will inevitably be missing due to manual division. Therefore, expanding the artificially segmented tumor area to include the peritumoral region partially addresses this issue to enhance model generalization ability. In addition, the texture changes, abnormal local blood vessel density, and information on the degree of tumor infiltration provided by the surrounding tissue characteristics also enable the model to evaluate the invasiveness and metastasis potential of the tumor from a more comprehensive perspective, thus enhancing the performance of the model. We found an article on intratumoral and peritumoral radiomics profiles for predicting lymph node metastasis status, it focused exclusively on patients with stage IA NSCLC (38). However, it is necessary to include surgically resectable cases ranging from stage IA to stage IIIA, encompassing only localized lymph node metastases. Compared with previously published models for predicting lymph nodes based on PET/CT (39, 40), CT is the first choice examination for patients with initial treatment of pulmonary nodules, may make our model more universal. In addition, PET/CT relies on FDG (fluorodeoxyglucose) uptake imaging, which may be affected by infectious or non-infectious diseases (such as tuberculosis, pneumoconiosis, or chronic obstructive pulmonary disease. In our study, we developed three prediction models for lymph node metastasis status in all included patients with surgically resectable NSCLC: an intratumoral radiomics profile model, a combined peritumoral and intratumoral radiomics profile model, and a clinical information prediction model. Upon comparing their performance, we observed that models incorporating peritumoral radiomics profiles was better in each metric (accuracy:0.887, sensitivity:0.908, specificity:0.875, precision:0.798, recall:0.908, negative predictive value:0.946, positive predictive value:0.798 and F1 score:0.849). Fusion radiomics features and clinical information making our model more applicable. Our constructed nomogram can predict lymph node metastasis of lung cancer in a non-invasive manner, which is of great significance for clinical treatment planning of lung cancer patients. Our results found that incorporating radiomic features of the peritumoral region is necessary in the prediction of lymph node metastasis in non-small cell lung cancer. The information it provides on texture changes, abnormal local blood vessel density, and the extent of tumor infiltration is indispensable for clinical assessment of the invasiveness and metastatic potential of primary tumors. Comprehensive consideration of the characteristics of the tumor itself and the characteristics of these surrounding areas can enhance the accuracy of lymph node metastasis prediction, bringing potential clinical benefits to patients.

This study also has some limitations. Even though we used a publicly available dataset for external validation, there is still a lack of multicenter data to further verify the generalization ability of the model. In addition, the dataset in this study also has the problem of imbalance in the number of outcome categories. To eliminate the impact of class imbalance on model evaluation as much as possible,

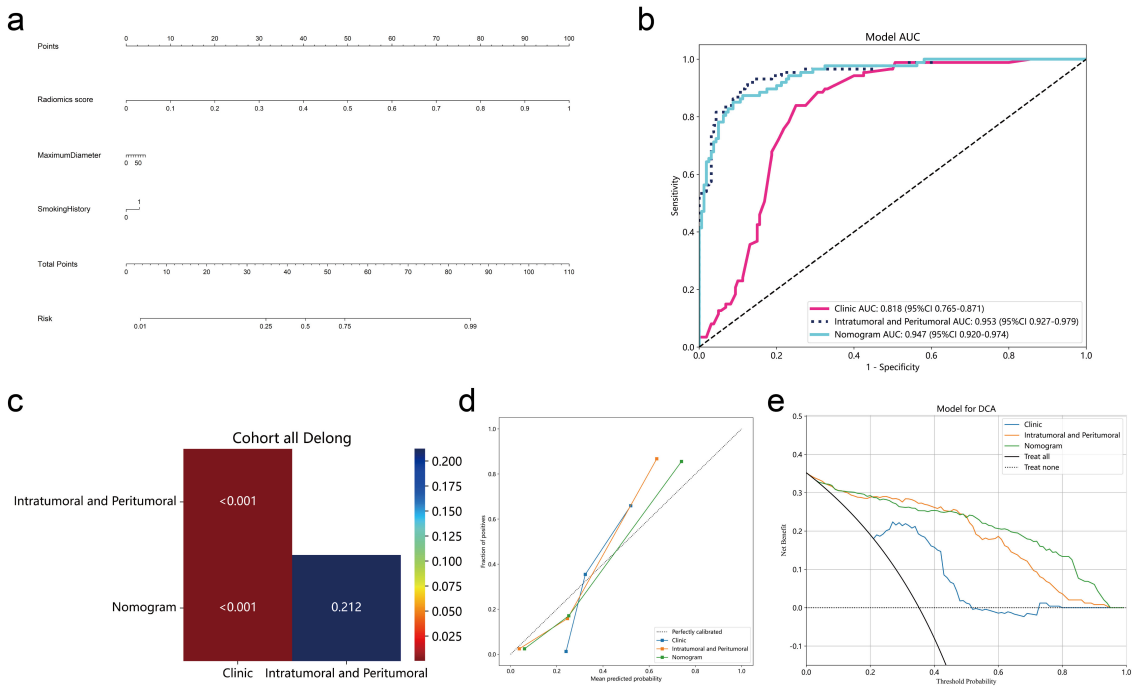


FIGURE 7
Nomogram integrating intratumoral-peritumoral radiomics model and clinical indicators. **(A)** Nomogram for predicting lymph node metastasis in non-small cell lung cancer. **(B)** The ROC curve of nomogram, intratumoral-peritumoral radiomics model and clinical model. **(C)** DeLong test of AUC values between different models. **(D, E)** Calibration curves and Decision Curve Analysis of different models.

we not only focus on the accuracy of the model, but also compare the AUC value under binary classification between models and the F1 score calculated by precision and recall. In addition, the error caused by manual segmentation of ROI areas should also be considered. Although we performed Intraclass Correlation Coefficient analysis after segmentation by multiple doctors to screen for robust radiomics features, this issue has not been completely resolved. In future studies, we would collect multicenter data for validation and establish better deep learning algorithm models to solve feature biases caused by manual segmentation of tumor regions. Furthermore, in the practical application of radiomics models, how to ensure the standardization and automation of the preprocessing process of images from different hospitals or different CT scanners is also a problem that needs to be solved urgently.

Conclusion

Our study proposed a predictive model for lymph node metastasis status based on preoperative CT radiomics profiles in patients with operable resected NSCLC. This model included the characteristics of peritumoral region and improved the performance of traditional intratumoral feature models and clinical feature models. Additionally, our constructed nomogram may provide new insights into the development of clinical treatment strategies for patients with surgically resectable NSCLC. Our results show that radiomics models considering the characteristics of the tumor region alone are insufficient and future studies should pay more attention to the region surrounding the tumor.

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.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Second Affiliated Hospital of Harbin Medical University (Approval No.: KY2022-144). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because This study is a retrospective study.

Author contributions

RX: Data curation, Software, Visualization, Writing – original draft, Writing – review & editing. KW: Data curation, Formal analysis, Writing – original draft. BP: Data curation, Software, Writing – original draft. XZ: Data curation, Writing – original

draft. CW: Writing – original draft. TL: Writing – original draft. JS: Data curation, Software, Writing – original draft. JZ: Data curation, Software, Writing – original draft. LZ: Conceptualization, Data curation, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to thank the Department of Radiology and the Department of Informatics at the Second Affiliated Hospital of Harbin Medical University for their assistance with this study. The graphical abstract was created using BioRender.com. Part of the analysis conducted in this study was performed on the OnekeyAI platform (http://www.medai.icu/). We acknowledge the contributions of its developers in this field.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1427743/full#supplementary-material

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

2. Santana-Davila R, Martins R. Treatment of stage iiia non-small-cell lung cancer: A concise review for the practicing oncologist. *J Oncol Pract*. (2016) 12:601–6. doi: 10.1200/JOP.2016.013052

3. Dunne EG, Fick CN, Jones DR. Mediastinal staging in non-small-cell lung cancer: saying goodbye to mediastinoscopy. *J Clin Oncol*. (2023) 41:3785–90. doi: 10.1200/JCO.23.00867

4. Liam CK, Liam YS, Poh ME, Wong CK. Accuracy of lung cancer staging in the multidisciplinary team setting. *Transl Lung Cancer Res*. (2020) 9:1654–66. doi: 10.21037/tlcr.2019.11.28

5. Erasmus LT, Strange TA, Agrawal R, Strange CD, Ahuja J, Shroff GS, et al. Lung cancer staging: imaging and potential pitfalls. *Diagnostics (Basel)*. (2023) 13. doi: 10.3390/diagnostics13213359

6. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141

7. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med*. (2020) 61:488–95. doi: 10.2967/jnumed.118.222893

8. Vaidya P, Bera K, Gupta A, Wang X, Corredor G, Fu P, et al. Ct derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, ii resectable non-small cell lung cancer: A retrospective multi-cohort study for outcome prediction. *Lancet Digit Health*. (2020) 2:e116–e28. doi: 10.1016/s2589-7500(20)30002-9

9. Tang X, Huang H, Du P, Wang L, Yin H, Xu X. Intratumoral and peritumoral ct-based radiomics strategy reveals distinct subtypes of non-small-cell lung cancer. *J Cancer Res Clin Oncol*. (2022) 148:2247–60. doi: 10.1007/s00432-022-04015-z

10. Mao N, Shi Y, Lian C, Wang Z, Zhang K, Xie H, et al. Intratumoral and peritumoral radiomics for preoperative prediction of neoadjuvant chemotherapy effect in breast cancer based on contrast-enhanced spectral mammography. *Eur Radiol*. (2022) 32:3207–19. doi: 10.1007/s00330-021-08414-7

11. Khorrami M, Jain P, Bera K, Alilou M, Thawani R, Patil P, et al. Predicting pathologic response to neoadjuvant chemoradiation in resectable stage iii non-small cell lung cancer patients using computed tomography radiomic features. *Lung Cancer*. (2019) 135:1–9. doi: 10.1016/j.lungcan.2019.06.020

12. Chen Q, Shao J, Xue T, Peng H, Li M, Duan S, et al. Intratumoral and peritumoral radiomics nomograms for the preoperative prediction of lymphovascular invasion and overall survival in non-small cell lung cancer. *Eur Radiol*. (2023) 33:947–58. doi: 10.1007/s00330-022-09109-3

13. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep learning vs. Radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: don't forget the peritumoral region. *Front Oncol*. (2020) 10:53. doi: 10.3389/fonc.2020.00053

14. Ding J, Chen S, Serrano Sosa M, Cattell R, Lei L, Sun J, et al. Optimizing the peritumoral region size in radiomics analysis for sentinel lymph node status prediction in breast cancer. *Acad Radiol*. (2022) 29 Suppl 1:S223–S8. doi: 10.1016/j.acra.2020.10.015

15. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast dce-mri. *Breast Cancer Res*. (2017) 19:57. doi: 10.1186/s13058-017-0846-1

16. Zhang H, Liao M, Guo Q, Chen J, Wang S, Liu S, et al. Predicting N2 lymph node metastasis in presurgical stage I-ii non-small cell lung cancer using multiview radiomics and deep learning method. *Med Phys*. (2023) 50:2049–60. doi: 10.1002/mp.16177

17. Thomas PA. Development of radiomics models to predict lymph node metastasis and de-escalated non-small-cell lung cancer surgery: A word of caution. *Eur J Cardiothorac Surg*. (2021) 60:72–3. doi: 10.1093/ejcts/ezab021

18. Qiao J, Zhang X, Du M, Wang P, Xin J. (18)F-fdg pet/ct radiomics nomogram for predicting occult lymph node metastasis of non-small cell lung cancer. *Front Oncol*. (2022) 12:974934. doi: 10.3389/fonc.2022.974934

19. Laros SSA, Dieckens D, Blazis SP, van der Heide JA. Machine learning classification of mediastinal lymph node metastasis in nsclc: A multicentre study in a western european patient population. *EJNMMI Phys*. (2022) 9:66. doi: 10.1186/s40658-022-00494-8

20. He L, Huang Y, Yan L, Zheng J, Liang C, Liu Z. Radiomics-based predictive risk score: A scoring system for preoperatively predicting risk of lymph node metastasis in patients with resectable non-small cell lung cancer. *Chin J Cancer Res*. (2019) 31:641–52. doi: 10.21147/j.issn.1000-9604.2019.04.08

21. Expert Consensus P, Kidane B, Bott M, Spicer J, Backhus L, Chaft J, et al. The american association for thoracic surgery (Aats) 2023 expert consensus document: staging and multidisciplinary management of patients with early-stage non-small cell lung cancer. *J Thorac Cardiovasc Surg*. (2023) 166:637–54. doi: 10.1016/j.jtcvs.2023.04.039

22. Ligero M, Jordi-Ollero O, Bernatowicz K, Garcia-Ruiz A, Delgado-Munoz E, Leiva D, et al. Minimizing acquisition-related radiomics variability by image

resampling and batch effect correction to allow for large-scale data analysis. *Eur Radiol*. (2021) 31:1460–70. doi: 10.1007/s00330-020-07174-0

23. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015

24. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–e7. doi: 10.1158/0008-5472.CAN-17-0339

25. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295:328–38. doi: 10.1148/radiol.2020191145

26. Meng F, Guo Y, Li M, Lu X, Wang S, Zhang L, et al. Radiomics nomogram: A noninvasive tool for preoperative evaluation of the invasiveness of pulmonary adenocarcinomas manifesting as ground-glass nodules. *Transl Oncol*. (2021) 14:100936. doi: 10.1016/j.tranon.2020.100936

27. Wu M, Xu W, Fei Y, Li Y, Yuan J, Qiu L, et al. Mri-based clinical radiomics nomogram may predict the early response after concurrent chemoradiotherapy in locally advanced nasopharyngeal carcinoma. *Front Oncol*. (2023) 13:1192953. doi: 10.3389/fonc.2023.1192953

28. Mao B, Ma J, Duan S, Xia Y, Tao Y, Zhang L. Preoperative classification of primary and metastatic liver cancer via machine learning-based ultrasound radiomics. *Eur Radiol*. (2021) 31:4576–86. doi: 10.1007/s00330-020-07562-6

29. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30. doi: 10.48550/arXiv.1201.0490

30. Gao C, Xiang P, Ye J, Pang P, Wang S, Xu M. Can texture features improve the differentiation of infiltrative lung adenocarcinoma appearing as ground glass nodules in contrast-enhanced ct? *Eur J Radiol*. (2019) 117:126–31. doi: 10.1016/j.ejrad.2019.06.010

31. Chang R, Qi S, Zuo Y, Yue Y, Zhang X, Guan Y, et al. Predicting chemotherapy response in non-small-cell lung cancer via computed tomography radiomic features: peritumoral, intratumoral, or combined? *Front Oncol*. (2022) 12:915835. doi: 10.3389/fonc.2022.915835

32. Saw SPL, Ong BH, Chua KLM, Takano A, Tan DSW. Revisiting neoadjuvant therapy in non-small-cell lung cancer. *Lancet Oncol*. (2021) 22:e501–e16. doi: 10.1016/S1470-2045(21)00383-1

33. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science*. (2017) 357. doi: 10.1126/science.aan2507

34. Zeng C, Zhang W, Liu M, Liu J, Zheng Q, Li J, et al. Efficacy of radiomics model based on the concept of gross tumor volume and clinical target volume in predicting occult lymph node metastasis in non-small cell lung cancer. *Front Oncol*. (2023) 13:1096364. doi: 10.3389/fonc.2023.1096364

35. Laqua FC, Woznicki P, Bley TA, Schoneck M, Rinneburger M, Weisthoff M, et al. Transfer-learning deep radiomics and hand-crafted radiomics for classifying lymph nodes from contrast-enhanced computed tomography in lung cancer. *Cancers (Basel)*. (2023) 15. doi: 10.3390/cancers15102850

36. Tanaka Y, Nakai T, Suzuki A, Kagawa Y, Noritake O, Taki T, et al. Clinicopathological significance of peritumoral alveolar macrophages in patients with resected early-stage lung squamous cell carcinoma. *Cancer Immunol Immunother*. (2023) 72:2205–15. doi: 10.1007/s00262-023-03393-8

37. Libling WA, Korn R, Weiss GJ. Review of the use of radiomics to assess the risk of recurrence in early-stage non-small cell lung cancer. *Transl Lung Cancer Res*. (2023) 12:1575–89. doi: 10.21037/tlcr-23-5

38. Chen QL, Li MM, Xue T, Peng H, Shi J, Li YY, et al. Radiomics nomogram integrating intratumoural and peritumoural features to predict lymph node metastasis and prognosis in clinical stage ia non-small cell lung cancer: A two-centre study. *Clin Radiol*. (2023) 78:e359–e67. doi: 10.1016/j.crad.2023.02.004

39. Zheng K, Wang X, Jiang C, Tang Y, Fang Z, Hou J, et al. Pre-operative prediction of mediastinal node metastasis using radiomics model based on (18)F-fdg pet/ct of the primary tumor in non-small cell lung cancer patients. *Front Med (Lausanne)*. (2021) 8:673876. doi: 10.3389/fmed.2021.673876

40. Wang M, Liu L, Dai Q, Jin M, Huang G. Developing a primary tumor and lymph node 18f-fdg pet/ct-clinical (Tlpc) model to predict lymph node metastasis of resectable T2-4 nsclc. *J Cancer Res Clin Oncol*. (2023) 149:247–61. doi: 10.1007/s00432-022-04545-6

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 14 November 2024
DOI 10.3389/fonc.2024.1441064

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Chunyin Zhang,
Southwest Medical University, China
Giorgia Garganese,
Mater Olbia Hospital, Italy

*CORRESPONDENCE
Biyun Zhang
bianka0830@126.com

RECEIVED 30 May 2024
ACCEPTED 25 October 2024
PUBLISHED 14 November 2024

CITATION
Song Q, Zhang B and Gu Y (2024) ^{18}F-FDG PET/CT detection of primary mammary-like adenocarcinoma of the vulva with multiple metastases: a case report.
Front. Oncol. 14:1441064.
doi: 10.3389/fonc.2024.1441064

COPYRIGHT
© 2024 Song, Zhang and Gu. 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.

^{18}F-FDG PET/CT detection of primary mammary-like adenocarcinoma of the vulva with multiple metastases: a case report

Qing Song[1], Biyun Zhang[2*] and Yu Gu[3]

[1]The First School of Clinical Medicine, Nanjing University of Chinese Medicine, Nanjing, China, [2]Department of Nuclear Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China, [3]Department of Oncology, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China

Mammary-like adenocarcinoma of the vulva is a malignancy with a low incidence rate compared with the squamous cell carcinoma occurring at the same site. We present a rare case of mammary-like adenocarcinoma of the vulva with multiple-organ involvement using ^{18}F-FDG PET/CT. This study indicates that ^{18}F-FDG PET/CT can not only detect the primary lesion but also distinguish the stage of the mammary-like adenocarcinoma of the vulva.

KEYWORDS

mammary-like adenocarcinoma, vulvar adenocarcinoma, FDG, PET/CT, vulvar cancer

Introduction

The mammary-like adenocarcinoma (MLA) of the vulva is a rare primary vulvar tumor occurring in ectopic vulvar breast tissue (1), and it shows a low incidence rate, and has immunohistochemical profiles and aggressive behaviors similar to those of breast cancers, which make it challenging for clinical diagnosis and differential diagnosis. PET/CT plays an important role in the management of vulvar cancer (2–4). However, at present, there is little literature on the detection of MLA of the vulva with ^{18}F-FDG PET/CT imaging.

In this case, the ^{18}F-FDG PET/CT imaging indicated a malignant tumor according to morphological manifestations and metabolic characteristics, which was definitely diagnosed as primary MLA of the vulva by the pathological examination.

Case description

A 68-year-old woman presented with low back pain for 1 month, and a non-contrast-enhanced abdominal CT revealed several lesions on her spine bones, suggesting multiple

metastatic lesions. Furthermore, laboratory examinations showed increased tumor markers: CA125: 50.7 U/ml (normal value < 35 U/ml), Ferritin: 539 ng/ml (normal value 13-150 ng/ml). This patient had no previous history of malignancy. For evaluating the general condition, this patient underwent ^{18}F-FDG PET/CT imaging, which demonstrated multiple fluorodeoxyglucose (FDG) uptake in the liver, bones and lymph nodes in the PET/CT MIP image (Figure 1A). Multiple foci of abnormal activity (broken arrows) were seen in the liver (Figures 1B, D), although there was no clear evidence of anatomical abnormality in the corresponding CT image (Figure 1C). Multiple hypermetabolic foci (solid arrows) were also seen in the bones (Figures 1E–P), with mixed bone destruction on CT (Figure 1L). Enlarged retroperitoneal and bilateral inguinal lymph nodes (dashed arrows) with increased activities were seen on PET/CT (Figures 1E–J). An additional focal activity (Figures 1N, P, curved arrows) with SUV_{max} of 6.2 was also observed in the

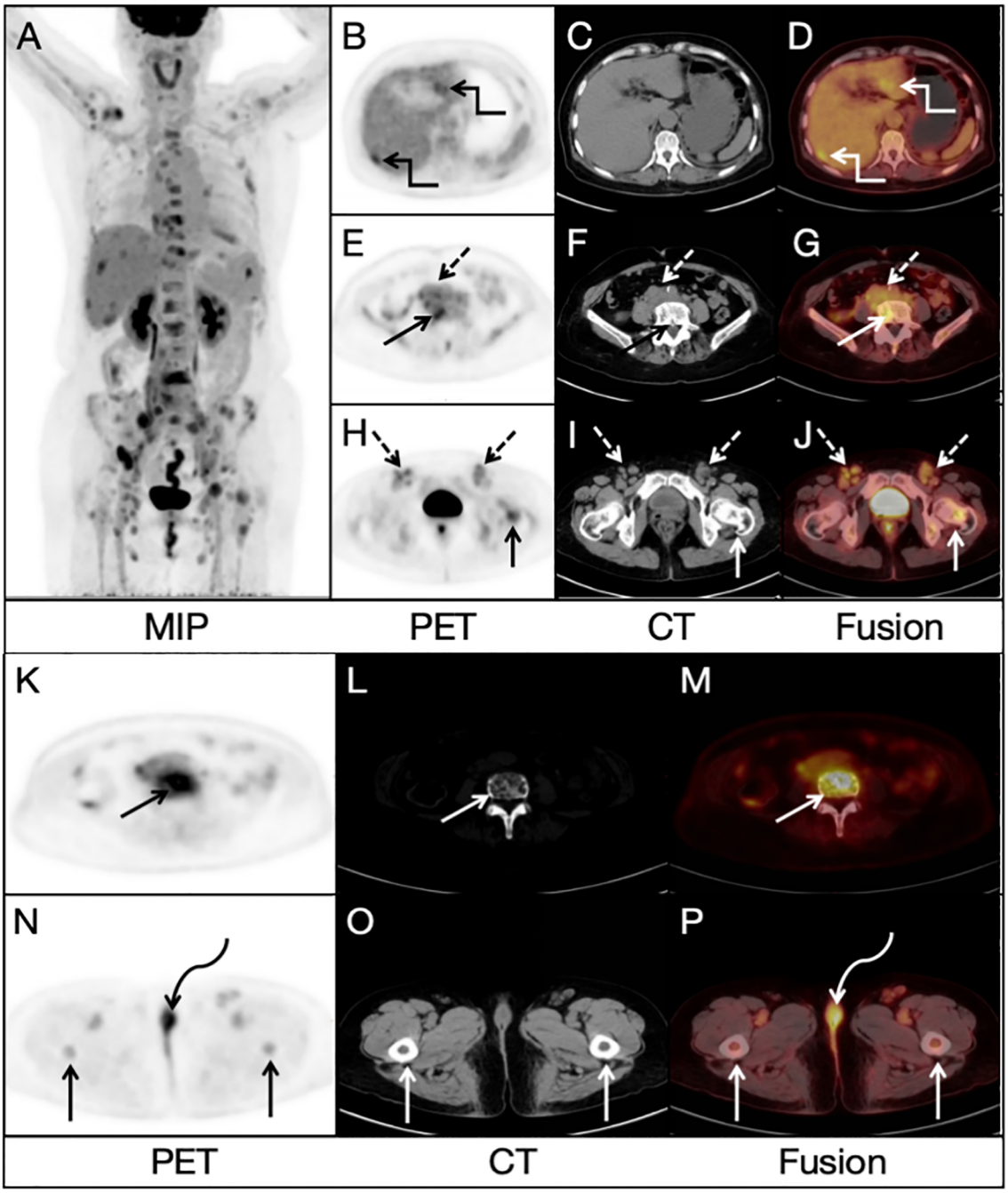


FIGURE 1
Radiological findings on ^{18}FDG PET/CT image: **(A)** Multiple lesions in the bones, liver and lymph nodes. **(B, D)** Multiple foci of abnormal activity (broken arrows) in the liver. **(C)** No clear evidence of anatomical abnormality in the corresponding area on CT image. **(E–J)** Enlarged retroperitoneal and bilateral inguinal lymph nodes (dashed arrows). **(K–M)** Multiple hypermetabolic foci (solid arrows) in the bones **(N, P)** An additional focal activity (curved arrows) with a SUV_{max} of 6.2 in the vulva. **(O)** No clear evidence of anatomical abnormality in the corresponding area on CT image.

vulva, although CT indicated no clear evidence of anatomical abnormality in the corresponding area (Figure 1O).

FDG uptake in many lesions in the liver, bones and lymph nodes might suggest multiple metastases, but the location of primary lesion was unclear. Due to the existence of bilateral inguinal lymph node metastases, the vulvar lesion was suspected to be the primary lesion. To further clarify the diagnosis, this patient underwent a delayed pelvic FDG PET/CT imaging at 2 h after FDG administration, and the results demonstrated multiple possible malignant lesions in the vulva (Figures 2A, C, curved arrows), with a furtherly increased SUV^{max} of 10.2 while no evidence of anatomical abnormality in the corresponding area on CT image (Figure 2B). A more detailed medical history was then taken. The woman had been complaining of vulvar inflammation and pain for years. The physical examination had revealed the presence of stiffness and swelling in the mons pubis, accompanied by two

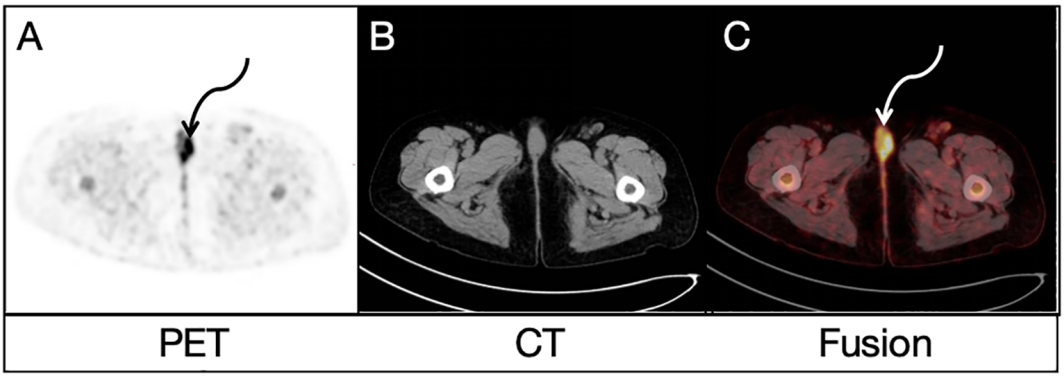


FIGURE 2
Delayed pelvic FDG PET/CT findings: Both PET image **(A)** and PET/CT fusion image **(C)** demonstrated similar possible malignant lesions (curved arrows) in the vulva, with a SUVmax of 10.2 while no evidence of anatomical abnormality in the corresponding area on CT image **(B)**.

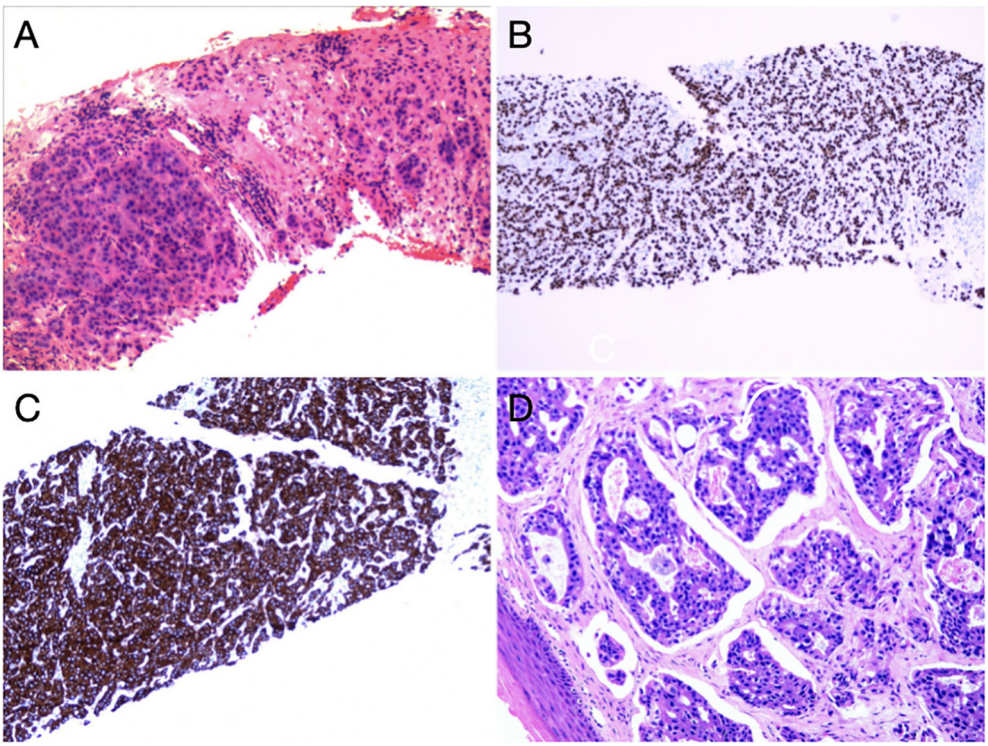


FIGURE 3
Pathohistological findings of inguinal lymph nodes and vulva: **(A)** HE staining of right inguinal lymph nodes. In fibrous connective tissue and lymphoid tissue, there were nests and cords of heterogeneous cells with enlarged nuclei and an increased nucleo-plasmatic ratio, and these features were consistent with those of malignant epithelial tumors. **(B)** Positive expression of GATA3. **(C)** Strongly positive expression of CK7. **(D)** HE staining of the vulva. There were cribriform hyperplasia and nest-shaped clusters of cells, local necrosis and vascular cancer thrombus in the tissues.

erythematous lesions of 1.0 cm in diameter. According to the medical history and findings of dual-time-point ^{18}F-FDG PET/CT, vulvar carcinoma with liver, bone and lymph node metastasis was suggested.

Subsequently, biopsy and pathological examination of affected vulva (Figure 3D) and right inguinal lymph nodes (Figures 3A–C) were performed, indicating that the vulvar lesion was an infiltrative, moderately/poorly differentiated adenocarcinoma arising from mammary gland-like epithelium of the vulva. In addition, the immunohistochemical results showed that the estrogen receptor/progesterone receptor and positive common breast cancer markers such as epithelial membrane antigen (EMA), carcinoembryonic antigen (CEA), cytokeratin 7(CK7), GATA-binding protein 3 (GATA3) and Ki-67 were increasingly expressed by more than 40%. The morphology and immunoprofile of the tumor cells suggested the incidence of MLA.

This patient had no history of breast malignancy, and no abnormal ^{18}F-FDG uptake was observed in bilateral breasts. Finally, this patient was diagnosed with a primary MLA of the vulva, with multiple metastases in liver, bones and lymph nodes. At present, the patient is receiving palliative radiotherapy in local lesion of vulvar MLA (with a prescription dose of 50 Gy in 25 fractions) after 6 cycles of chemotherapy with paclitaxel and carboplatin. The outcome after chemotherapy was assessed as stable.

Discussion

It has been reported that the squamous cell carcinoma is the most common malignant tumor in the vulva, and accounts for 76% of vulvar cancers, while the adenocarcinoma only accounts for less than 2% of vulvar cancers (5). Van der Linden et al. (6) have observed that the majority of glandular malignancies in the vulva are primary tumors, which account for 55%. Primary adenocarcinomas in the vulva predominantly include extramammary Paget's disease and sweat gland carcinomas. In addition, a rare form of adenocarcinoma is known as MLA of the vulva, which originates from putative mammary-like glands of the vulva, and exhibits a spectrum of pathological reactions that are similar to those in their mammary counterparts (7–9). Notably, the extramammary Paget's disease can also present as a manifestation of underlying breast carcinoma, thus contributing to the differential diagnosis (10). The diagnosis of vulvar malignancies is obviously dependent on pathological evaluation (11). Since vulvar MLA and breast cancer have morphological similarities and similar aggressive behaviors, the criteria for diagnosing vulvar malignancy were defined by referring to breast carcinoma (12): in the absence of concurrent breast carcinoma, a primary adenocarcinoma of mammary-like glands of the vulva can be diagnosed when its morphological pattern is consistent with that of the breast carcinoma, with the presence of estrogen/progesterone receptors and positive markers such as CK 7 and GATA 3, which are common in breast cancers. In this case, the immunohistochemical results were measured against the above-mentioned criteria.

It is worth noting that HER-2, as one of the molecular typing markers of breast cancer, theoretically has an important reference value in the diagnosis and treatment of vulvar MLA (13). Unfortunately, the HER-2 expression was not detected in this case. Among previous reports (14), HER-2 expression was detected in a few cases, indicating a potential value of HER-2 in the diagnosis and treatment of vulvar MLA.

Ki67 is widely recognized as the best indicator of cell proliferation activity. Rolfe et al. (15) demonstrated that the increased expression of Ki67 reflects the changes in the proliferation during vulvar carcinogenesis, suggesting that Ki67 can serve as an indicator of tumor invasiveness. This patient is still under treatment, and attention will be paid to the correlation between tumor progression and ki67 expression. However, a study (16) has shown that Ki67 has no prognostic value in patients with vulvar Paget's disease (PDV) or mammary Paget's disease (PDB). Therefore, further study is needed to explore the correlation between Ki67 expression and this disease in the future.

The MLA originating from the vulva is exceedingly uncommon, and fewer than 40 cases have been reported since 1935 (14, 17). Only two patients in these reports underwent FDG PET/CT to assess the stage of the adenocarcinoma, and the results revealed that one patient had no metastatic disease, while the other patient exhibited distant metastases in lymph nodes and bones (18, 19). PET/CT plays an important role in the management of vulvar cancer. In the case reported by Patel D et al. (20), PET/CT exerted an important effect in staging of an identified vulvar cancer. Moreover, Triumbari EKA et al. (21) demonstrated that a negative preoperative PET/CT imaging may indicate no groin metastases in early-stage vulvar cancer patients, thus it's not necessary for them to undergo sentinel node biopsy. And Treglia G et al. (22) reported a case of vulvar extramammary Paget's disease (EMPD) restaged by PET/CT. In this case, PET/CT scan was helpful in locating the primary tumor. The manifestations of some inflammatory lesions and tumors on PET/CT are similar. Zhuang H et al. (23) have proved that the SUVs of delayed images of malignant lesions increase over time compared with those of earlier images, while the SUVs of the inflammatory lesions and benign lesions remain stable or decrease slightly over time. Another clinical trial (24) also showed that dual-time-point FDG-PET imaging has a potential to improve the accuracy of distinguishing between inflammation and tumor lesions. In this case, the manifestations of vulvar lesions on dual-time-point FDG PET/CT clearly showed this feature, and the SUV_{max} increased from 6.2 on the initial image to 10.2 on the delayed image. Furthermore, some studies (25, 26) have shown that the radiomics features of PET/CT images have an important value in lymph node assessment and prognosis prediction in patients with vulvar cancers.

Because of the rarity of this disease and the lack of definitive treatment guidelines, this type of cancers is currently staged and treated according to the treatment method for primary breast cancers (27, 28). Surgical excision with adjuvant therapies such as radiation, anthracycline-based chemotherapy and hormonal therapy is a common option (29). Benito V et al. (30) presented a case of an elderly patient with metastatic vulvar adenocarcinoma arising from mammary-like glands successfully treated with a combination of surgery and hormonal therapy. And Tanaka H et al. (31) reported a case of MLA successfully treated by paclitaxel

weekly without excision. In addition, Butler B et al. (32) explored the application of sentinel lymph node mapping in patients with MLA, this technique has been almost exclusively applied in patients with breast carcinoma. It is very suitable to use sentinel lymph node mapping in the patients with vulvar cancers, which can predict the lymphatic drainage of the vulva, with a lower false negative rate and a substantially reduced adverse reaction rate (e.g. lymphedema) compared with the total groin lymph node dissection (33, 34). Tessier-Cloutier B et al. (35) observed that the breast carcinoma and MLA of the vulva had similar intrinsic luminal molecular subtypes, and thus proposed a new treatment strategy, that is, molecular subtyping of breast cancer can be performed to optimize individual treatment.

To sum up, we report a rare case of MLA of the vulva with multiple-organ involvement detected by ^{18}F-FDG PET/CT. In this case, according to the distribution characteristics of lesions on PET/CT, the primary tumor was found, indicating the important role of PET/CT. Furthermore, dual-time-point PET/CT imaging demonstrates a great diagnostic value and potential significance in differential diagnosis. The effects of PET/CT imaging in prognostic prediction of MLA remain to be confirmed in future studies. This case is described based on the real world, and there are some deficiencies in the diagnosis of MLA of the vulva, such as insufficient collection of detailed medical histories and a lack of physical examination before PET/CT imaging, which have increased the difficulty of diagnosis. The study of this case reveals that optimized diagnostic strategy and appropriate imaging technique are especially important in the clinical diagnosis of rare diseases.

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.

Ethics statement

Written informed consent was obtained from the patient for the publication of this case report.

Author contributions

QS: Writing – original draft. BZ: Writing – review & editing. YG: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Natural Science Foundation of Nanjing University of Chinese Medicine (GrantNo.XZR2020006) and Special Project for Director Promotion of Jiangsu Province Hospital of Chinese Medicine (GrantNo.Y2021ZR26).

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.

References

1. McMaster J, Dua A, Dowdy SC. Primary breast adenocarcinoma in ectopic breast tissue in the vulva. *Case Rep Obstet Gynecol.* (2013) 2013:721696. doi: 10.1155/2013/721696

2. Peiró V, Chiva L, González A, Bratos R, Alonso S, Márquez R, et al. Utility of the PET/CT in vulvar cancer management. *Rev espanola medicina Nucl e imagen Mol.* (2014) 33:87–92. doi: 10.1016/j.remn.2013.07.009

3. Dolanbay M, Ozcelik B, Abdulrezzak U, Serin IS, Kutuk MS, Uludag S. F-18 fluoro-D-glucose (FDG)-positron emission tomography (PET)/computed tomography (CT) in planning of surgery and sentinel lymph node screening in vulvar cancers. *Arch gynecology obstetrics.* (2016) 293:1319–24. doi: 10.1007/s00404-015-3927-3

4. Cohn DE, Dehdashti F, Gibb RK, Mutch DG, Rader JS, Siegel BA, et al. Prospective evaluation of positron emission tomography for the detection of groin node metastases from vulvar cancer. *Gynecologic Oncol.* (2002) 85:179–84. doi: 10.1006/gyno.2002.6605

5. Wohlmuth C, Wohlmuth-Wieser I. Vulvar Malignancies: an interdisciplinary perspective. *J Dtsch Dermatol Ges.* (2019) 17:1257–76. doi: 10.1111/ddg.13995

6. van der Linden M, Schuurman M, Bulten J, van der Aa M, Massuger L, de Hullu J. Incidence and survival of glandular vulvar Malignancies in the Netherlands. *Gynecol Oncol.* (2017) 144:553–7. doi: 10.1016/j.ygyno.2017.01.020

7. van der Putte SC, van Gorp LH. Adenocarcinoma of the mammary-like glands of the vulva: a concept unifying sweat gland carcinoma of the vulva, carcinoma of supernumerary mammary glands and extramammary Paget's disease. *J Cutan Pathol.* (1994) 21:157–63. doi: 10.1111/j.1600-0560.1994.tb00251.x

8. Van der Putte SC. Mammary-like glands of the vulva and their disorders. *Int J Gynecol Pathol.* (1994) 13:150–60. doi: 10.1097/00004347-199404000-00009

9. Desouki MM, Fadare O. Primary adenocarcinomas of the vulva and related structures: an enigmatic and diverse group of tumors. *Semin Diagn Pathol.* (2021) 38:71–84. doi: 10.1053/j.semdp.2020.09.011

10. Ohira S, Itoh K, Osada K, Oka K, Suzuki A, Osada R, et al. Vulvar Paget's disease with underlying adenocarcinoma simulating breast carcinoma: case report and review of the literature. *Int J Gynecol Cancer.* (2004) 14:1012–7. doi: 10.1111/j.1048-891X.2004.14544.x

11. Michalski BM, Pfeifer JD, Mutch D, Council ML. Cancer of the vulva: a review. *Dermatol Surg.* (2021) 47:174–83. doi: 10.1097/DSS.0000000000002584

12. Irvin WP, Cathro HP, Grosh WW, Rice LW, Andersen WA. Primary breast carcinoma of the vulva: a case report and literature review. *Gynecol Oncol.* (1999) 73:155–9. doi: 10.1006/gyno.1998.5269

13. Lob S, Linsmeier E, Herbert SL, Schlaiss T, Kiesel M, Wischhusen J, et al. Prognostic effect of HER2 evolution from primary breast cancer to breast cancer

metastases. *J Cancer Res Clin Oncol*. (2023) 149:5417–28. doi: 10.1007/s00432-022-04486-0

14. Niakan S, Love H, Cao Q, Kawar N. Primary invasive lobular carcinoma arising in mammary-like glands of the vulva managed with neoadjuvant trastuzumab-based chemotherapy, excision, and sentinel lymph node biopsy. *Clin Case Rep*. (2021) 9:118–22. doi: 10.1002/ccr3.3475

15. Rolfe KJ, Eva LJ, MacLean AB, Crow JC, Perrett CW, Reid WM. Cell cycle proteins as molecular markers of Malignant change in vulvar lichen sclerosus. *Int J Gynecol Cancer*. (2001) 11:113–8. doi: 10.1046/j.1525-1438.2001.011002113.x

16. Ellis PE, Fong LF, Rolfe KJ, Crow JC, Reid WM, Davidson T, et al. The role of P53 and Ki67 in Paget's disease of the vulva and the breast. *Gynecol Oncol*. (2002) 86:150–6. doi: 10.1006/gyno.2002.6629

17. Abbott JJ, Ahmed I. Adenocarcinoma of mammary-like glands of the vulva: report of a case and review of the literature. *Am J Dermatopathol*. (2006) 28:127–33. doi: 10.1097/01.dad.0000171601.25315.2b

18. Deshmukh AA, Greenwalt J, Whitworth JM, Fox MD, Crozier JA. Mammary-like carcinoma of the vulva: A diagnostic challenge. *Breast J*. (2020) 26:1242–4. doi: 10.1111/tbj.13831

19. Ananthula A, Lockwood B, Savage J, Malak S, Chen C, Makhoul I, et al. Primary breast carcinoma of the vulva metastatic to lymph nodes and bones: a case report and literature review. *Perm J*. (2020) 24:19.084. doi: 10.7812/TPP/19.084

20. Patel D, Anderson TM, Lu Y. FDG-PET/CT of vulvar adenocarcinoma with diffuse metastases. *Clin Nucl Med*. (2016) 41:710–1. doi: 10.1097/RLU.0000000000001257

21. Triumbari EKA, de Koster EJ, Rufini V, Fragomeni SM, Garganese G, Collarino A. ^{18}F-FDG PET and ^{18}F-FDG PET/CT in vulvar cancer: a systematic review and meta-analysis. *Clin Nucl Med*. (2021) 46:125–32. doi: 10.1097/RLU.0000000000003411

22. Treglia G, Giovannini E, Bertagna F, Giovanella L, Malaggese M. An unusual case of metastatic extramammary Paget's disease of the vulva identified by ^{18}F-FDG PET/CT. *Rev Esp Med Nucl Imagen Mol*. (2013) 32:402–3. doi: 10.1016/j.remn.2013.05.001

23. Zhuang H, Pourdehnad M, Lambright ES, Yamamoto AJ, Lanuti M, Li P, et al. Dual time point ^{18}F-FDG PET imaging for differentiating Malignant from inflammatory processes. *J Nucl Med*. (2001) 42:1412–7.

24. Xiu Y, Bhutani C, Dhurairaj T, Yu JQ, Dadparvar S, Reddy S, et al. Dual-time point FDG PET imaging in the evaluation of pulmonary nodules with minimally increased metabolic activity. *Clin Nucl Med*. (2007) 32:101–5. doi: 10.1097/01.rlu.0000252457.54929.b7

25. Collarino A, Garganese G, Fragomeni SM, Pereira Arias-Bouda LM, Ieria FP, Boellaard R, et al. Radiomics in vulvar cancer: first clinical experience using (18)F-FDG PET/CT Images. *J Nucl Med*. (2018) 60:199–206. doi: 10.2967/jnumed.118.215889

26. Collarino A, Garganese G, Valdes Olmos RA, Stefanelli A, Perotti G, Mirk P, et al. Evaluation of dual-timepoint (18)F-FDG PET/CT imaging for lymph node staging in vulvar cancer. *J Nucl Med*. (2017) 58:1913–8. doi: 10.2967/jnumed.117.194332

27. Meddeb S, Rhim MS, Mestiri S, Kouira M, Bibi M, Khairi H, et al. Mammary-like adenocarcinoma of the vulva associated to Paget's disease: a case report. *Pan Afr Med J*. (2014) 19:188. doi: 10.11604/pamj.2014.19.188.4521

28. Morais M, Vaz Silva J, Tavares MV. Diagnosis and management of primary vulvar adenocarcinoma of mammary gland type: report of two distinct cases. *BMJ Case Rep*. (2022) 15:6. doi: 10.1136/bcr-2021-245580

29. Lopes G, DeCesare T, Ghurani G, Vincek V, Jorda M, Glück S, et al. Primary ectopic breast cancer presenting as a vulvar mass. *Clin Breast Cancer*. (2006) 7:278–9. doi: 10.3816/CBC.2006.n.041

30. Benito V, Arribas S, Martínez D, Medina N, Lubrano A, Arencibia O. Metastatic adenocarcinoma of mammary-like glands of the vulva successfully treated with surgery and hormonal therapy. *J Obstet Gynaecol Res*. (2013) 39:450–4. doi: 10.1111/j.1447-0756.2012.01937.x

31. Tanaka H, Umekawa T, Nagao K, Ishihara A, Toyoda N. Adenocarcinoma of mammary-like glands in the vulva successfully treated by weekly paclitaxel. *Int J Gynecol Cancer*. (2005) 15:568–71. doi: 10.1111/j.1525-1438.2005.15328.x

32. Butler B, Leath CA 3rd, Barnett JC. Primary invasive breast carcinoma arising in mammary-like glands of the vulva managed with excision and sentinel lymph node biopsy. *Gynecol Oncol Case Rep*. (2013) 7:7–9. doi: 10.1016/j.gynor.2013.09.001

33. Collarino A, Fuoco V, Garganese G, Pereira Arias-Bouda LM, Perotti G, Manca G, et al. Lymphoscintigraphy and sentinel lymph node biopsy in vulvar carcinoma: update from a European expert panel. *Eur J Nucl Med Mol Imaging*. (2020) 47:1261–74. doi: 10.1007/s00259-019-04650-8

34. Levenback CF, Ali S, Coleman RL, Gold MA, Fowler JM, Judson PL, et al. Lymphatic mapping and sentinel lymph node biopsy in women with squamous cell carcinoma of the vulva: a gynecologic oncology group study. *J Clin Oncol*. (2012) 30:3786–91. doi: 10.1200/JCO.2011.41.2528

35. Tessier-Cloutier B, Asleh-Aburaya K, Shah V, McCluggage WG, Tinker A, Gilks CB, et al. Molecular subtyping of mammary-like adenocarcinoma of the vulva shows molecular similarity to breast carcinomas. *Histopathology*. (2017) 71:446–52. doi: 10.1111/his.13239

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 17 February 2025
DOI 10.3389/fonc.2025.1475133

Check for updates

OPEN ACCESS

EDITED BY
Harrison Kim,
University of Alabama at Birmingham, United States

REVIEWED BY
Jianing Xi,
Guangzhou Medical University, China
Ge Ren,
Hong Kong Polytechnic University, Hong Kong SAR, China

*CORRESPONDENCE
Edward Castillo
edward.castillo@utexas.edu

†These authors have contributed equally to this work and share first authorship

RECEIVED 02 August 2024
ACCEPTED 27 January 2025
PUBLISHED 17 February 2025

CITATION
Nomura GRO, Luong AT, Prakash A, Alemand A, Bhowmick T, Ali A, Ren J, Rehani B, Nair G, Castillo R, Vinogradskiy Y and Castillo E (2025) TriSwinUNETR lobe segmentation model for computing DIR-free CT-ventilation. *Front. Oncol.* 15:1475133. doi: 10.3389/fonc.2025.1475133

COPYRIGHT
© 2025 Nomura, Luong, Prakash, Alemand, Bhowmick, Ali, Ren, Rehani, Nair, Castillo, Vinogradskiy and Castillo. 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.

TriSwinUNETR lobe segmentation model for computing DIR-free CT-ventilation

Gabriela Roque Oliveira Nomura[1†], Aaron T. Luong[1†], Ananya Prakash[1], Annabelle Alemand[1], Tanish Bhowmick[1], Alisa Ali[1], Jaimie Ren[1], Basil Rehani[1], Girish Nair[2], Richard Castillo[3], Yevgeniy Vinogradskiy[4] and Edward Castillo[1*]

[1]Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, United States, [2]Division of Pulmonary and Critical Care, William Beaumont University Hospital, Royal Oak, MI, United States, [3]Department of Radiation Oncology, Emory University School of Medicine, Atlanta, GA, United States, [4]Department of Radiation Oncology, Thomas Jefferson University, Philadelphia, PA, United States

Purpose: Functional radiotherapy avoids the delivery of high-radiation dosages to high-ventilated lung areas. Methods to determine CT-ventilation imaging (CTVI) typically rely on deformable image registration (DIR) to calculate volume changes within inhale/exhale CT image pairs. Since DIR is a non-trivial task that can bias CTVI, we hypothesize that lung volume changes needed to calculate CTVI can be computed from AI-driven lobe segmentations in inhale/exhale phases, without DIR. We utilize a novel lobe segmentation pipeline (TriSwinUNETR), and the resulting inhale/exhale lobe volumes are used to calculate CTVI.

Methods: Our pipeline involves three SwinUNETR networks, each trained on 6,501 CT image pairs from the COPDGene study. An initial network provides right/left lung segmentations used to define bounding boxes for each lung. Bounding boxes are resized to focus on lung volumes and then lobes are segmented with dedicated right and left SwinUNETR networks. Fine-tuning was conducted on CTs from 11 patients treated with radiotherapy for non-small cell lung cancer. Five-fold cross-validation was then performed on 51 LUNA16 cases with manually delineated ground truth. Breathing-induced volume change was calculated for each lobe using AI-defined lobe volumes from inhale/exhale phases, without DIR. Resulting lobar CTVI values were validated with 4DCT and positron emission tomography (PET)-Galligas ventilation imaging for 19 lung cancer patients. Spatial Spearman correlation between TriSwinUNETR lobe ventilation and ground-truth PET-Galligas ventilation was calculated for each patient.

Results: TriSwinUNETR achieved a state-of-the-art mean Dice score of 93.72% (RUL: 93.49%, RML: 85.78%, RLL: 95.65%, LUL: 97.12%, LLL: 96.58%), outperforming best-reported accuracy of 92.81% for the lobe segmentation task. CTVI calculations yielded a median Spearman correlation coefficient of 0.9 across 19 cases, with 13 cases exhibiting correlations of at least 0.5, indicating strong agreement with PET-Galligas ventilation.

Conclusion: Our TriSwinUNETR pipeline demonstrated superior performance in the lobe segmentation task, while our segmentation-based CTVI exhibited strong agreement with PET-Galligas ventilation. Moreover, as our approach leverages deep-learning for segmentation, it provides interpretable ventilation results and facilitates quality assurance, thereby reducing reliance on DIR.

KEYWORDS

lobe segmentation, CT-ventilation, artificial intelligence, transformer networks, functional radiotherapy, deformable image registration, medical image segmentation

1 Introduction

The lungs are situated within the thoracic cavity, enclosed by the ribcage. Although the right and left lungs appear similar, they exhibit notable anatomical differences. The right lung consists of three lobes: right upper lobe (RUL), right middle lobe (RML), and right lower lobe (RLL). The left lung consists of only two: left upper lobe (LUL) and left lower lobe (LLL). Lobar boundaries are characterized by fissures that appear as thin white lines on computed tomography (CT) scans. There is variability in lobe fissures, which may be complete when the lobes are connected solely at the hilum by the bronchi and pulmonary vessels, incomplete when there are areas of parenchymal fusion between the lobes, or entirely absent (1). In the right lung, the oblique fissure divides the lower and middle lobes, and the horizontal fissure divides the upper and middle lobes (2). The left lobe contains only one oblique fissure separating the upper and lower lobes (2). Radiologists and machine learning models alike identify the fissure locations in the lungs to determine the boundaries between lobes and subsequently perform the task of lobe segmentation.

Lobe segmentation is crucial for various medical applications, including disease diagnosis, severity assessment, and treatment planning (3). Previous research has shown that lung regions receiving high doses of radiation (> 20 Gy) experience a decrease in post-treatment lung function, as measured by decreased CTVI values (4). To improve lung cancer treatment, functional radiotherapy has been proposed. This approach aims to avoid delivering high radiation doses to high-functioning, or high-ventilated, lung areas during radiotherapy treatment planning. Methods to calculate CTVI typically use deformable image registration (DIR), in which lung voxels are registered from the inhale to the exhale phase of the breathing cycle, and the displacement of each voxel is measured (4). However, iterative DIR is a non-trivial task that can potentially bias CTVI due to its long computation time, potential inaccuracies in alignment, and inherent uncertainty (5). Therefore, the use of a DIR-free automated lung lobe segmentation method as a means for calculating CTVI based on volume changes is a potential avenue to be explored.

While healthy patients pose little challenge to existing lobe segmentation methods, the accuracy of the segmentations can be significantly worsened by the presence of disease states. It has been demonstrated that conditions such as COVID-19 can affect the appearance of lung tissue and cause existing methods to fail (6). Similarly, parenchymal fibrosis associated with chronic obstructive pulmonary disorder (COPD) can pose similar challenges to existing methods since the density of the fibrotic tissue is similar to that of tissue outside the lung, which can obscure fissures and affect thresholding-based algorithms (6).

Previous works on deep learning-based segmentation have been faced with the issue of downsampling images without losing substantial contextual information in order to fit the memory capacity of graphics processing units (GPUs) needed for model training. As shown in Figure 1, fissure identification is challenging at lower resolutions, but due to memory requirements, it is a common procedure to downsample original CT scans from a full-resolution to a lower-resolution space. To preserve contextual information, previous works have employed a random sampling tactic, resizing the original 512x512x256 CT scan to a lower resolution of 256x256x128 and then random sampling in 128x128x64 patches (7). The issue with such approaches, however, is in the first downsampling step, which diminishes the resolution of the images by a factor of 2, thus possibly leading to a loss of essential information that could have guided the model to a better result.

While convolutional neural networks (CNNs), proposed in the 1980s, have been widely used for medical imaging segmentation tasks, newer machine learning models have improved upon the basic CNN. Ronneberger et al. took the basic CNN structure to create the U-NET, widely used for medical imaging segmentation tasks. This "U-shaped" network consists of an encoder, a contracting path to capture contextual information, and a decoder, an expanding path that enables localization (8). In contrast to CNNs which are notably constrained to the local features captured by the kernel, the innovative transformer architecture provides the capability to capture more global features. Recent advancements in large language models are in large part thanks to the transformer architecture, introduced in the famous "Attention is All You Need" paper (9). Expanding upon the basic transformer network, the vision transformer (ViT) provided the precedent and foundation for implementing transformers in computer vision tasks (10). The UNETR architecture developed by Hatamizadeh et al. replaces the convolutional encoding arm of a

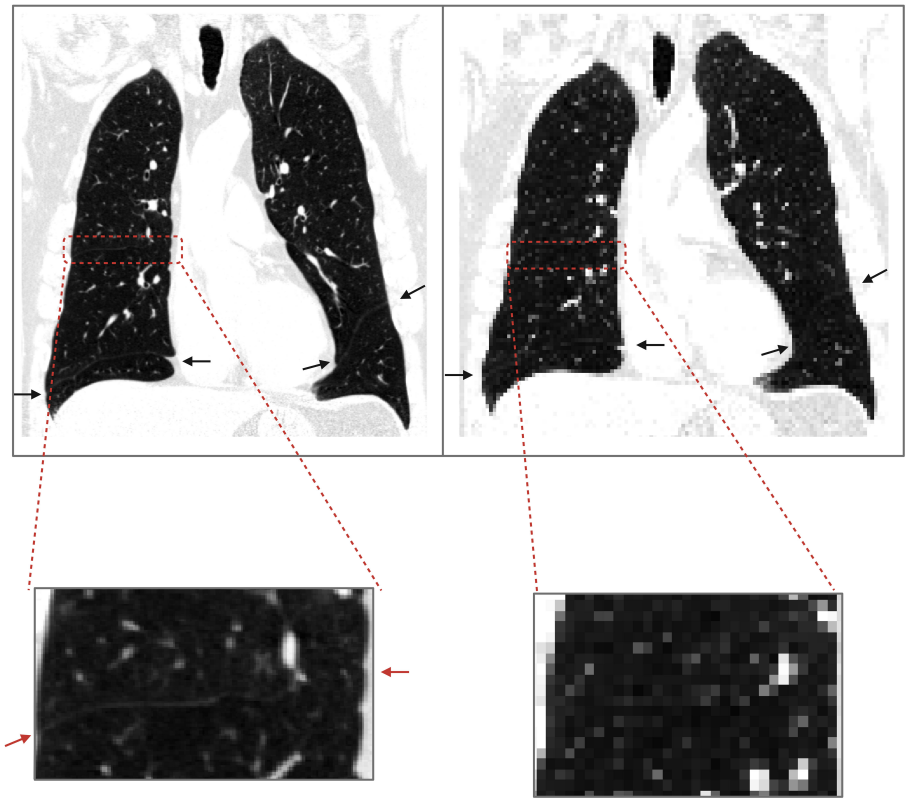


FIGURE 1
Loss of contextual information upon downsampling. The image on the left is a COPD Gold 1 case at resolution 512x512x649; the endpoints of fissures are shown with arrows. On the right, the same image was downsampled with nearest neighbor interpolation to 128x128x128, and upon visual inspection, the horizontal fissure on the right lung is not clearly identified.

traditional UNET with a ViT encoder, achieving state-of-the-art performance at the time in the Beyond the Cranial Vault (BTCV) abdominal CT multi-organ segmentation challenge (11). In order to translate images into a format compatible with the ViT, 16x16 pixel large patches of the image are taken and linearly projected. However, images often vary in scale and require higher resolution than the 16x16 patches can capture. The Swin Transformer resolves this limitation by dividing the image into a variety of patches ranging from 4x4 to 16x16 in order to capture the different scales and resolution of details in an image (12). This approach generalizes into the third dimension and is utilized in the SwinUNETR, replacing the ViT encoder with a Swin Transformer. This improved architecture is demonstrated to provide superior performance compared to its ViT-based predecessor in the BTCV challenge (13).

In this study, we propose the novel three SwinUNETR (TriSwinUNETR) ensemble pipeline network for CT lobe segmentation, using the proven state-of-the-art transformer-based machine learning architecture. Furthermore, we break down the lobe segmentation task into three steps to prevent loss of contextual information caused by image downsampling. Then, we train the machine learning pipeline in multiple datasets and disease states to ensure that the model is generalizable to multiple patients. Lastly, we test the model on a clinical task for calculating CTVI of lung cancer patients prior to undergoing radiotherapy.

2 Materials and methods

2.1 COPDGene and preprocessing

CT images from the COPDGene dataset were used to train the three SwinUNETR networks comprising our TriSwinUNETR model. COPDGene data was acquired from an observational study conducted to identify genetic factors that contributed to COPD (14). Images were acquired using multi-detector CT scanners, with 3D volumetric scans acquired on both full inspiration (200 mAs) and end-of-normal expiration (50 mAs) (14). For each patient, the scan is composed of sub-millimeter spaced (0.625 - 0.9 mm) 512x512 slices with a pixel spacing of 0.5 mm (14). The severity of COPD is categorized by a value for the Global Initiative for Obstructive Lung Disease, or GOLD score. GOLD scores range from 0 to 4, where an increasing score denotes increased severity. Our model was trained on the full range of GOLD scores, as shown in Table 1.

A total of 13,002 unique breath-hold CT scans, or 6,501 inhale-exhale CT image pairs, from all GOLD scores at the initial time point of the COPDGene imaging study were preprocessed prior to the training phase. Scans were first converted from Hounsfield units, which are calculated based on the attenuation coefficient of the X-ray beam, to density values ranging from 0 to 1 (15). All lung tissues fall within this density range while excluding denser tissue

10.3389/fonc.2025.1475133

TABLE 1 CODPGene training data.

Gold Score	Qty
0	4387
1	787
2	1926
3	1164
4	606
Uncategorized	1782

Number of CT images per GOLD score.

and bone. In order to satisfy computational and memory constraints, the initial full-resolution training CT scans and their corresponding masks were then downsampled with nearest-neighbor interpolation to a resolution of 128x128x128.

2.2 SwinUNETR training

This paper proposes a lobe segmentation pipeline, TriSwinUNETR, composed of three SwinUNETR networks. The overview of the SwinUNETR architecture is shown in Figure 2. The SwinUNETR is a "U-shaped" network composed of a Swin-transformer encoder and a CNN-based decoder with skip connections at each resolution. The encoder begins with a patch partition layer and then proceeds to contain 4 stages comprising 2 transformer blocks each. Each stage contains a window multi-head self-attention transformer mechanism, applied individually within each partition, and a sliding window multi-head self-attention transformer mechanism, applied across different local windows. Patch merging occurs at the end of each stage. Encoded feature representations are concatenated to the decoder input via skip connections at each resolution along the path. In each of the four stages, output features are reshaped and sent to a convolutional residual block. The final segmentation is outputted using a 1x1x1 convolutional layer and softmax activation function. Hyperparameter and optimizer details are included on the - Supplementary Material page.

As shown in Figure 3, an initial network denoted as Lung SwinUNETR provides right and left lung segmentations on a resolution space of 128x128x128, which are then subsequently utilized to determine bounding boxes for each lung. The bounding boxes acquired from the left and right lung segmentations from the first network are then upsampled and localized back onto the original CT scan, which is at a full resolution of 512x512x512. Each individual lung is then cropped from the original scan, downsampled back to 128x128x128, and provided as input to a dedicated SwinUNETR network that is trained to output the corresponding amount of segmentation classes for lobes in that lung. This strategy only downsamples the lung region instead of the entire CT scan for the task of lobe segmentation, thus preserving a substantial amount of contextual information.

2.3 Post-processing

After each model forward pass, the result was dusted with an implementation of a block-based union-find algorithm. All connected components were evaluated with a connectivity of 26 and a threshold of 5000 voxels for the initial segmentation and 3000 voxels for the subsequent segmentations was used for dusting.

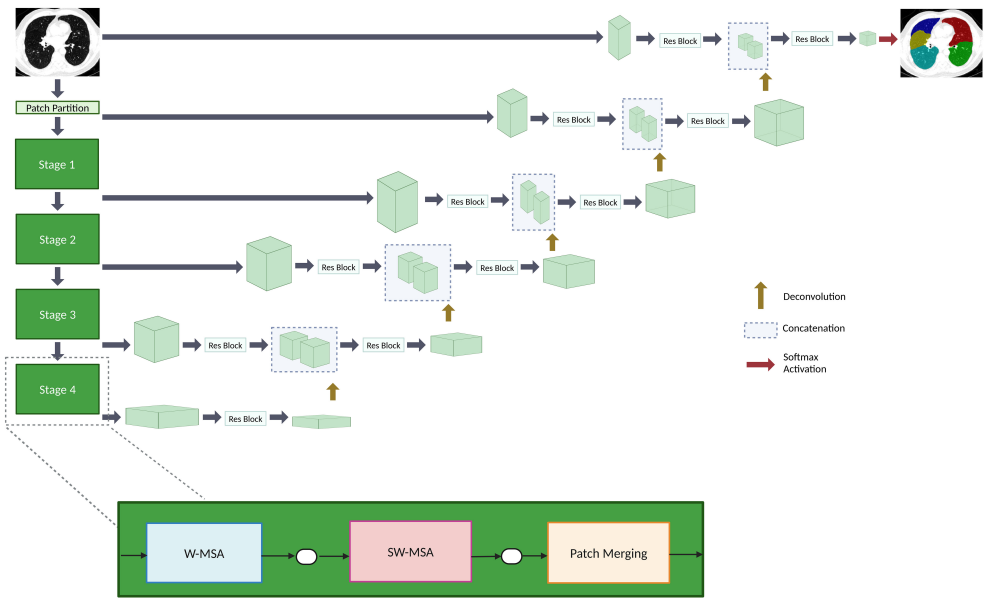


FIGURE 2
An overview of the SwinUNETR architecture. The encoder contains four stages. Each stage is composed of a window multi-head self-attention (W-MSA) transformer block, a sliding window multi-head self-attention (SW-MSA) transformer block, and a patch merging mechanism. The decoder reshapes output features that are sent to a convolutional block up on the path. The encoder and decoder are connected via skip connections. A softmax activation function outputs the final segmentations.

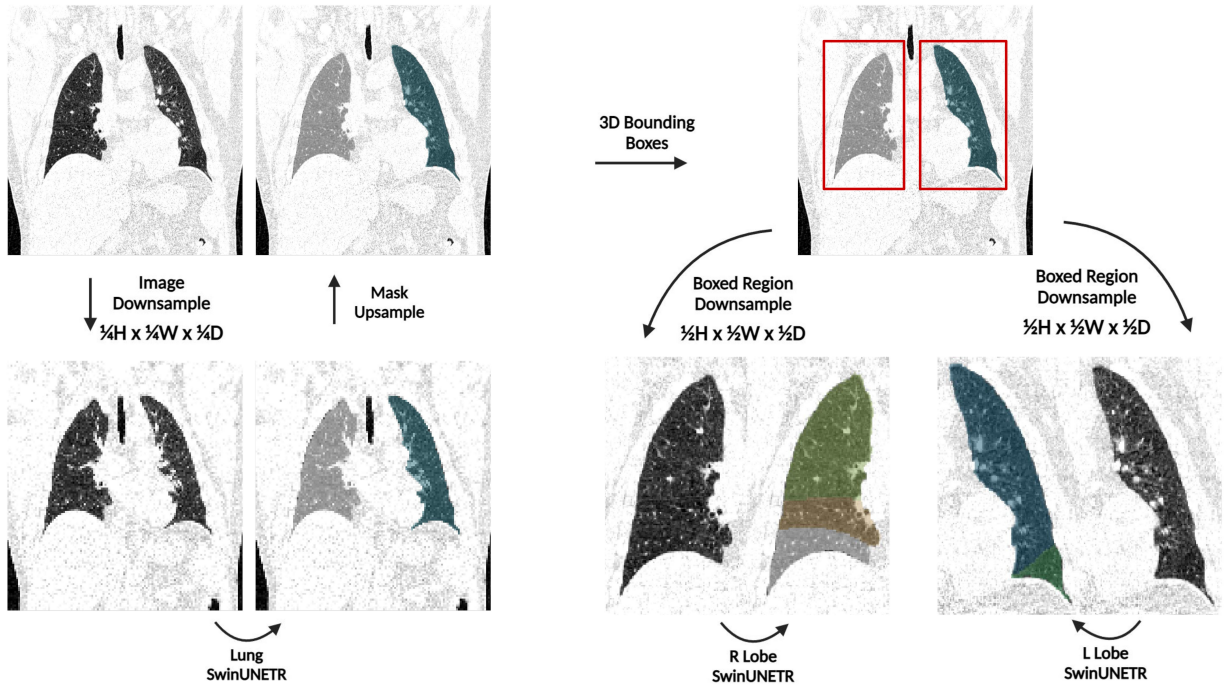


FIGURE 3
An overview of the TriSwinUNETR pipeline. A full-resolution (512x512x512) CT image is first downsampled to 128x128x128, left and right lung segmentations are acquired then upsampled back to the original image space, bounding boxes for each lung are determined based on the initial segmentations, regions delineated by the bounding boxes are downsampled, then corresponding Right Lobe and Left Lobe SwinUNETR networks output the corresponding lobe segmentations for each lung.

2.4 Fine-tuning and testing datasets

As shown in Table 2, two datasets were used to finetune the L Lobe SwinUNETR and R Lobe SwinUNETR networks. The first fine-tuning step was done on 22 4DCT scans of non-small cell lung cancer patients, with lobe segmentations manually delineated by experts. These scans were obtained as part of a study to incorporate lung function imaging into radiation therapy to preserve function after treatment. Of the 202 inhale and exhale phases from the 101 participants in the study, 22 were randomly selected after validation of image quality (16).

The second dataset used is a subset of the LUNA-16 dataset with lobe segmentations manually created by radiologists. The LUNA-16 dataset was originally a dataset of 888 scans selected from the LIDC-IDRI dataset of lung CT scans with nodules as part of a lung nodule segmentation Grand Challenge in 2016. From this dataset, 51 scans were selected, segmented by radiologists, and presented in (17).

2.5 Fine-tuning with 4DCT lung cancer dataset

4DCT non-small cell lung cancer images used in fine-tuning were preprocessed similarly to the COPDGene cases described previously. The process of acquiring bounding boxes for the lungs and using them to crop to each lung on the high-resolution image is the same. However, the two new images and their corresponding lobe masks were not resized to 128^3; instead, they were resized to the largest possible resolution divisible by 32 with a volume less than $6*128^3$ to fit into memory. This fine-tuning process yielded the best results and was conducted on the right and left networks separately. Hyperparameter and optimizer details are included on the - Supplementary Material page.

2.6 K-fold cross validation with LUNA16

Following fine-tuning on the cancer dataset, K-fold cross-validation was performed on a random subset of 51 LUNA16 lobe segmentation cases, as previously described. The original LUNA16 scans were preprocessed in the same manner as the COPDGene dataset. The 51 cases were divided into k = 5 folds of 10 or 11 cases each. The optimizer and hyperparameters used were identical to the previous fine-tuning step. The accuracy of segmentations was calculated using Dice percent score.

2.7 CT-ventilation calculation

Using the trained and fine-tuned AI model previously described, lobe segmentations were acquired from the inhale and exhale phases of the 4DCT of 19 lung cancer patients from a publicly available dataset (18). As shown in Figure 4, CTVI calculation for each lobe

was acquired by determining the breathing-induced lobar volume change using AI-defined lobe segmentation volumes, without DIR. The CTVI was calculated for each individual patient as the percent change in lobar volumetric segmentations between the full inhale and full exhale phases of the breathing cycle. Each patient had five CTVI values, one for each main lobar region.

Resulting lobar CTVI values were validated with positron emission tomography (PET)-Galligas ventilation imaging for each lung cancer patient, which was acquired from the same publicly available dataset (18). As shown in Figure 4, the PET-Galligas ventilation was acquired for each patient's lobe (LUL, LLL, RUL, RML, RLL) based on the number of counts or detected photon events recorded by the PET scanner. Each of the five PET-Galligas ventilation values was compared against the CTVI value calculated for that specific lobe. The spatial Spearman correlation between TriSwinUNETR lobe ventilation and ground-truth PET-Galligas ventilation was calculated for each patient. A Spearman correlation value of at least 0.5 suggests a moderately strong correlation between AI-based ventilation and the ground truth. It is important to note that the PET imaging itself, as shown in Figure 4, could have been a source of error since the radioactive material can be seen going beyond the extremities of the lungs. Since the calculations were isolated to lobes only, the radioactive material outside of the lungs was not considered.

To determine whether the proposed volume-change approach may improve CTVI calculations compared to DIR-based methods, the CTVIs for the 19 lung cancer cases were generated using both methods. The chosen DIR-based method to use was the integral formulation of the Jacobian (IJF), which aims to estimate the apparent voxel volume changes within an inhale/exhale CT image pair (19). The Spatial Spearman correlation between IJF ventilation and ground-truth PET-Galligas was generated for each patient.

TABLE 2 Summary of datasets used.

Name and Source	Number of CT images	Purpose
COPDGene (14)	13,002	Training Lung, Right Lobe, and Left Lobe SwinUNETR models
4DCT Lung Cancer (18)	22	Fine-tuning Right Lobe and Left Lobe SwinUNETR models
LUNA16 (17)	51	K-fold cross-validation of Right Lobe and Left Lobe SwinUNETR models

The name, source, number of CT images used, and purpose of each dataset in this study are included in this table.

3 Results

3.1 Dice comparison for lobe segmentation

Table 3 shows the lobe segmentation results of previous high-performing model architectures. Our method attained a mean Dice percent score of 93.75 ± 1.81% on the LUNA16 cases, with RUL at 93.49 ± 2.76%, RML at 85.78 ± 5.61%, RLL at 95.65 ± 0.69%, LUL at 97.12 ± 0.17%, and LLL at 96.58 ± 0.42%. 4.08M working parameters were used in the Lung SwinUNETR, and 15.7M working parameters were used in the L Lobe SwinUNETR and R Lobe SwinUNETR each. TriSwinUNETR contains 35.48M working parameters in total.

TriSwinUNETR Dice percent scores included on Table 3 are the average across the five folds from our K-fold cross-validation.

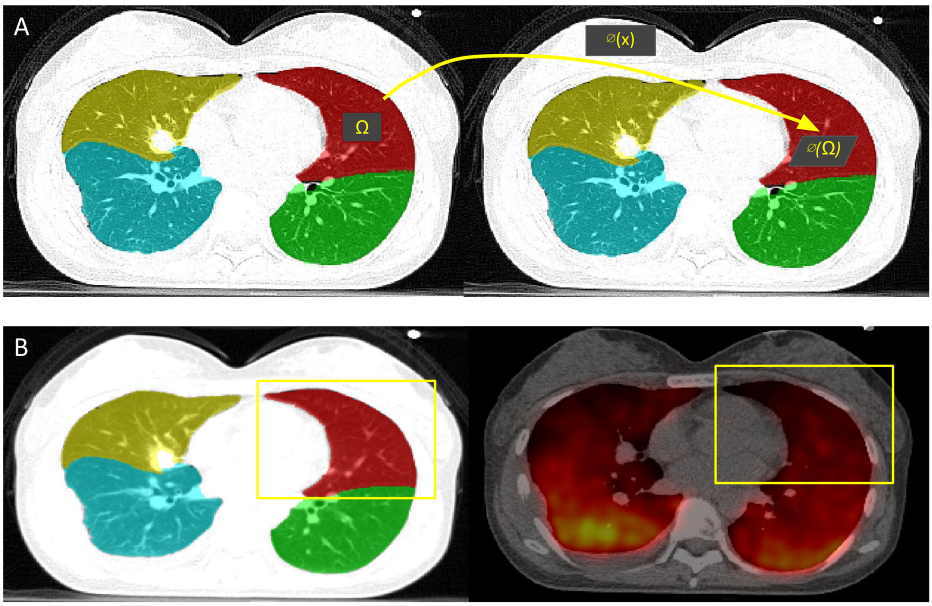


FIGURE 4
CTVI calculation and validation. **(A)** CTVI is calculated based on the change in volume between the inhale (right) and exhale (left) segmentations, where ventilation(Ω) = |1 - vol(∅(Ω))/vol(Ω)|. **(B)** The ground-truth PET-Galligas (right) is acquired by isolating the number of counts per lobe based on the TriSwinUNETR lobe segmentation from the PET-CT scan (left).

Table 4 shows the results for the K-fold cross validation of 51 LUNA16 cases using TriSwinUNETR network.

Representative test cases from the three datasets used to train and finetune the model were selected, and their corresponding lobe segmentations and ground truth are shown in Figure 5.

3.2 Spearman correlation for CT-ventilation

Table 5 shows the Spearman correlation coefficients between the CTVI and the PET-Galligas for each of the 19 lung cancer patients. The correlation was calculated between the CTVI and PET-Galligas count for all five lobes (LUL, LLL, RUL, RML, RLL) per patient. The median Spearman correlation coefficient was 0.9 across 19 cases, with 13 cases exhibiting correlations of at least 0.5, indicating moderately strong agreement between CTVI and PET-Galligas ventilation.

Figure 6 compares the proposed method's CTVI values and the number of PET-counts per lobar region for four patients. As shown, higher percent ventilation values should correspond to higher numbers of counts detected by the PET-scanner in order to result in a strong Spearman correlation. Refer to the - Supplementary Material page for all patients' data.

4 Discussion

4.1 Lobe segmentation task

Lung cancer is the leading cause of cancer death in the United States, causing more deaths in 2020 than breast, colorectal, and prostate cancers combined (23). However, advancements in the understanding of tumor biology, development of targeted therapies, and introduction of low-dose computed tomography (LDCT) for lung cancer screening have increased survival rates (23). Annual cancer screening using LDCT is an integral step in detecting cancer at its earlier stages, and lobe segmentation is a necessary part of the process. Computer-assisted diagnosis (CAD) methods aid radiologists in early lung nodule detection, but to do so, automatic segmentation of pulmonary lobes must be completed to eliminate other confounding structures such as the heart, the thoracic wall, abdominal organs, and the vertebrae (24). Although current automatic lobe segmentation models exist, previous methods may fail to train on different lung conditions and disease states in addition to struggling to downsample images without losing substantial contextual information, as previously discussed (6, 7). For these reasons we propose a novel CT lobe segmentation pipeline (TriSwinUNETR), which employs transfer learning to ensure generalizability to multiple datasets and disease states and breaks down segmentation tasks to prevent loss of contextual information caused by image downsampling.

In the first part of this study, we used K-fold cross-validation applied to the LUNA16 dataset to compare against previously published results. Although previous works do not mention the process of K-fold cross-validation nor a clear method for calculating standard deviation of Dice scores, we have chosen to perform 5-fold validation to ensure that LUNA16 test results are robust and reproducible. The proposed TriSwinUNETR achieves a mean Dice score of 93.75%, surpassing the mean Dice percent accuracy reported by the current state-of-the-art model, SCLMnet (22). In particular, TriSwinUNETR outperforms pre-existing architectures on the segmentation of lobes in the right lung.

As shown in Figure 5, our pipeline performs well in comparison to ground truth manual segmentations. Since we downsample a smaller portion of the image instead of the entire CT scan, a significant amount of information is preserved on the lower-resolution image, thus improving the accuracy of our results. In addition, fine-tuning on higher-resolution lung cancer images preserved higher-quality contextual information that could have been useful in the network's learning process. Due to their vast number of parameters as well as the complexity of architecture, transformer-based models require a significant amount of images to learn (25). Therefore, in addition to preserving as much of the image resolution as possible, transfer learning also increased the accuracy of our results. Training the models first on the COPDGene dataset allowed the architecture to familiarize itself with lung anatomy and the segmentation task. However, fine-tuning on the lung cancer dataset prior to doing K-fold cross-validation on LUNA16 allowed us to ensure that the model would be exposed to different types of scanners and lung conditions.

A limitation of our study is the quality of the training data itself. As previously stated, transformer-based models require a substantial amount of images to be optimized, but manual lobe segmentation is a time-consuming task. Therefore, we trained TriSwinUNETR on the larger COPDGene dataset, despite its automated segmentations that are prone to errors. This bias could have influenced the training of our model. However, in order to

TABLE 3 Dice percent score comparison.

Model	RUL Dice %	RML Dice %	RLL Dice %	LUL Dice %	LLL Dice %	Mean Dice %	Params (M)
UNETR (11)	92.47	83.56	93.15	93.51	92.84	91.11	101.96
UNet++ (20)	92.20	82.26	94.08	95.89	95.53	92.00	N/A
AttentionUnet (21)	93.01	82.91	94.24	95.32	94.89	92.07	23.63
SCLMnet (22)	92.81	78.28	95.02	97.68	97.05	92.17	87.31
TriSwinUNETR	**93.50**	**85.82**	**95.65**	97.16	96.61	**93.75**	35.48

Dice percent scores for RUL, RML, RLL, LUL, LLL, and mean Dice percent score are given for each model. Model parameters (in millions) are also included for each model. Bolded values indicate tasks where TriSwinUNETR outperforms existing architectures.

TABLE 4 Dice percent scores per fold.

Fold	RUL Dice %	RML Dice %	RLL Dice %	LUL Dice %	LLL Dice %	Mean Dice %
1	93.38	84.80	94.73	97.09	96.32	93.27
2	96.22	92.09	95.91	97.50	97.33	95.81
3	93.09	83.30	96.15	97.13	96.10	93.15
4	88.65	77.14	94.97	96.98	96.75	90.90
5	96.14	91.75	96.51	97.13	96.54	95.61

Dice percent scores for RUL, RML, RLL, LUL, LLL, and mean Dice percent score are given for each fold.

minimize errors caused by such bias, we have fine-tuned the Left and Right SwinUNETR networks with 62/63 images (22 lung cancer and 40/41 lung nodule CTs). In addition, our pipeline, along with other existing methods, struggles to segment cases where the patient has undergone a lobectomy due to the lack of available lobe segmentations of cases involving lobectomies. Therefore, future directions for this work include expanding the datasets and disease states that we train our models on, in addition to developing an even more effective method to preserve image resolution prior to the segmentation task.

4.2 CT-ventilation task

Given the high mortality of lung cancer in the United States, radiotherapy (RT) has been undergoing several technological innovations in recent years (26). One specific type of RT, functional RT, allows irradiation of tumors with high doses, while sparring healthy lung tissue. In order to make use of this advancement, however, functional image information is necessary to determine high-functioning lung areas in addition to applying personalized dose prescriptions for patients (26). CTVI is a functional metric that has been positively correlated with high-functioning areas of the lung, as previously discussed (4). Deformable image registration (DIR) is a core process in radiotherapy treatment planning and in calculating CTVI. However, DIR is a process that requires caution, as previously stated, as it is a process that is highly subject to variation in algorithm and user input (27). It has also been observed that DIR may return results that are physically implausible (27). Therefore, in the second part of this study, we propose the use of our TriSwinUNETR network to calculate CTVI without the use of DIR.

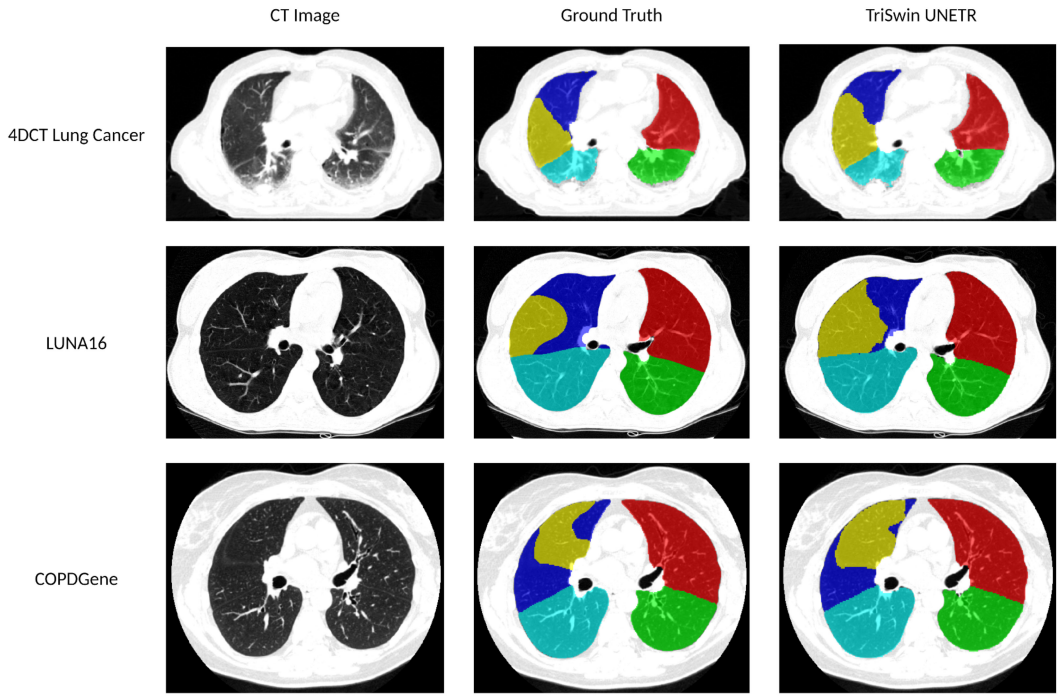


FIGURE 5
Lobar segmentations comparison. The 4DCT Lung Cancer test case has an average percent Dice score of 92.61 (LUL: 96.77 LLL: 89.69 RUL: 96.26 RML: 93.99 RLL: 86.36). The LUNA16 test case has an average percent Dice score of 96.05 (LUL: 97.45 LLL: 97.69 RUL: 95.16 RML: 92.35 RLL: 97.61). The COPDGene test case has an average percent Dice score of 96.62 (LUL: 97.83 LLL: 97.28 RUL: 97.20 RML: 94.09 RLL: 96.69).

TABLE 5 Spearman correlation coefficients of proposed method's CTVI vs. IJF CTVI.

Patient ID	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	18	19	20
Proposed Method	**0.9**	0.2	-1	**1**	**0.6**	**0.9**	**0.9**	**0.9**	**0.6**	**0.9**	**0.9**	0	**0.9**	0.3	0.2	**0.5**	**0.9**	-1	**1**
DIR-Based IJF	**0.54**	**0.52**	0.39	0.45	**0.56**	0.27	0.48	**0.59**	**0.64**	**0.71**	0.27	0.17	**0.54**	0.42	0.33	0.35	**0.56**	0.41	0.41

Patient ID is the number describing each anonymous patient. Spearman coefficient was calculated from the correlations of the CTVI and PET-Galligas for all five lobes per patient. Bolded numbers represent successful cases (correlation ≥ 0.5). Patient 7's PET-Galligas ventilation could not be calculated as it lacks a necessary CT image.

The Spearman correlation coefficients reveal that when using the proposed lobe volume change method, 13 out of 19 lung cancer cases (~68%) yielded successful CTVI results when compared to the ground-truth PET-Galligas per lobe. On the other hand, when using the DIR-based IJF method, only 8 out of 19 lung cancer cases (~42%) yielded successful CTVI results compared to ground truth. These results indicate the potential benefit of implementing a lobe volume change approach to calculate CTVI instead of solely relying on DIR. Since the lobe segmentation pipeline has been proven over 93% accurate on LUNA16 cases and has been fine-tuned on lung cancer cases, the model is well-prepared to segment lobes from the lung cancer 4DCT test set. Possible reasons for the proposed model's Spearman correlation coefficients < 0.5 include patients who have had a lobectomy, which as previously discussed could be challenging for the model to segment. In addition, the 4DCT lung cancer cases have an image resolution of 512 x 512 x ~170; the lower image resolution on the z-axis could make it more challenging for the AI model to identify fissures in certain patients. Lastly, most of the unsuccessful cases have abnormal PET-Galligas, which are either incorrectly cropped or missing to fill a sublobar region. Refer to the - Supplementary Material page to see figures of patients' with Spearman correlation < 0.5.

Possible future directions for this work include combining the proposed CTVI method with an iterative DIR method. While CTVI calculations per lobe provide interpretable ventilation results and can be verified by referencing lobe segmentation outputs, iterable DIR-methods allows for every voxel to be registered from inhale to exhale phase instead of comparing the volume change of a lobar region. Therefore, a combination of iterative DIR limited to a specific lobar region outputted by the AI model, or DIR-ventilation results cross-referenced to the AI-outputted lobe ventilation results could yield a mathematically stable and accurate CTVI method.

4.3 Ethical considerations

Our proposed method of calculating CTVI directly from lobe segmentations has the potential to be implemented in the clinic. In the United States, 4DCT imaging is a standard aspect of radiotherapy treatment planning for patients with lung cancer

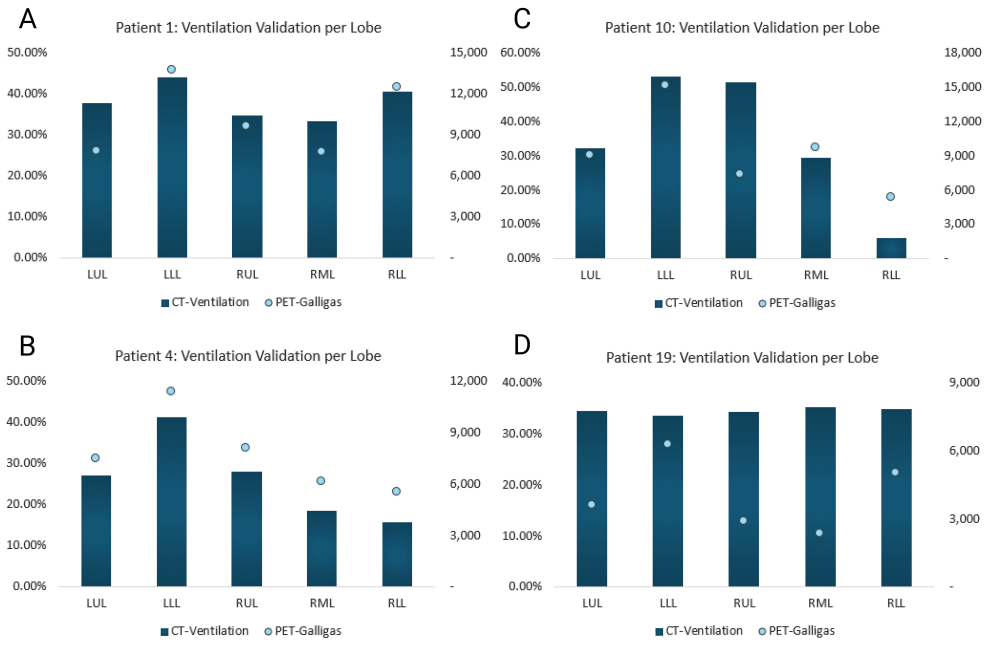


FIGURE 6
Proposed method's CTVI vs. PET-Galligas counts per lobe. CTVI is calculated in percent and PET-Galligas ventilation is calculated by the number of counts. The graphs included serve to visualize strong vs. weak Spearman correlations. **(A)** Patient 1's calculated Spearman correlation = 0.9, **(B)** Patient 4's calculated Spearman correlation = 1, **(C)** Patient 10's calculated Spearman correlation = 0.6, **(D)** Patient 19's calculated Spearman correlation = -1.

(28). Therefore, acquiring lobe segmentations directly from patients' CT images and performing CTVI calculations in local hospital machines would not disrupt the clinical workflow or risk patient data leaking. With the appropriate quality assurance procedures already in place at radiation oncology clinics, AI-defined lobe segmentations and their corresponding ventilation values may be inspected by medical physicists and radiation oncologists prior to implementation. The benefits of CTVI implementation in the clinic have been shown in a 2022 study. It was proven that CTVI as a functional imaging metric for functional avoidance radiotherapy planning reduced pneumonitis rates in lung cancer patients, thus proving CTVI's potential for clinical implementation (28).

5 Conclusion

In this work, we proposed a novel implementation of state-of-the-art segmentation architecture for automated CT lobe segmentation and made it publicly available to the scientific community. We utilized a TriSwinUNETR composed of three SwinUNETR networks for three distinct segmentation tasks: left and right lung segmentation, right lobes segmentation, and left lobes segmentation. Our proposed method, trained on a section of the COPDGene dataset and fine-tuned on manual lobe segmentations, includes minimal preprocessing and postprocessing. Dice score comparison on a subsection of the LUNA16 dataset showed that our proposed method outperforms currently proposed state-of-the-art methods. Using the proposed TriSwinUNETR AI-defined lobe volumes from a 4DCT lung cancer dataset, we have calculated the per-patient CTVI value for each lobe. Spatial Spearman correlation between TriSwinUNETR lobe ventilation and ground-truth PET-Galligas ventilation indicates strong agreement, thus possibly revealing a DIR-free alternative for calculating CTVI with the use of an AI-based lobe segmentation model. Future directions for this work include developing a more effective method to preserve image resolution prior to the segmentation task, expanding training datasets to include more disease states, and possibly stabilizing the mathematical uncertainties of DIR calculations with the proposed AI-based CTVI method.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://luna16.grand-challenge.org/, https://copdgene.org/, https://www.cancerimagingarchive.net/collection/ct-vs-pet-ventilation-imaging/. Trained model weights for this study can be found in the TriSwinUNETR repository at https://github.com/DMIC-Lab/segmentation-pipeline.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

GRON: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AL: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AP: Data curation, Software, Writing – review & editing. AnA: Data curation, Software, Writing – review & editing. TB: Data curation, Software, Writing – review & editing. AlA: Data curation, Software, Writing – review & editing. JR: Data curation, Software, Writing – review & editing. BR: Data curation, Software, Writing – review & editing. GN: Conceptualization, Writing – review & editing. RC: Conceptualization, Writing – review & editing. YV: Conceptualization, Writing – review & editing. EC: Conceptualization, Methodology, Software, Funding acquisition, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was partially funded by grant R01CA236857. The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1475133/full#supplementary-material

References

1. Mpolokeng KS, Madolo MY, Louw GJ, Gunston G. Anatomical variations in lung fissures leading to supernumerary lobes in the lungs. *Trans Res Anat*. (2022) 28:100209. doi: 10.1016/j.tria.2022.100209

2. Chaudhry R, Omole AE, Bordoni B. *Anatomy, thorax, lungs*. StatPearls Publishing (2024).

3. Carmo D, Ribeiro J, Dertkigil S, Appenzeller S, Roberto L, Rittner L. A systematic review of automated segmentation methods and public datasets for the lung and its lobes and findings on computed tomography images. *Yearbook Med Inf*. (2022) 31:277–95. doi: 10.1055/s-0042-1742517

4. Vinogradskiy Y. CT-based ventilation imaging in radiation oncology. *Br Institute Radiol*. (2019) 1. doi: 10.1259/bjro.20180035

5. Dowling JA, O'Connor LM. Deformable image registration in radiation therapy. *J Med Radiat Sci*. (2020) 67:257–595. doi: 10.1002/jmrs.446

6. G Sarah E, Herrmann J, Xin Y, Martin K, Rezoagli E, Ippolito D, et al. CT image segmentation for inflamed and fibrotic lungs using a multi-resolution convolutional neural network. *Sci Rep*. (2021) 11. doi: 10.1038/s41598-020-80936-4

7. Ferreira FT, Sousa P, Galdran A, Sousa M, Campilho A. *End-to-end supervised lung lobe segmentation*. Portuguese National Funding Agency for Science, Research and Technology (RCAAP Project by FCT) (2018). doi: 10.1109/ijcnn.2018.8489677

8. Ronneberger O, Fischer P, Brox T. *U-net: convolutional networks for biomedical image segmentation*. ArXiv (Cornell University (2015). doi: 10.48550/arxiv.1505.04597

9. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. *Attention is all you need*. ArXiv.org (2017). doi: 10.48550/arXiv.1706.03762

10. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. *An image is worth 16x16 words: transformers for image recognition at scale*. ArXiv.org (2021). doi: 10.48550/arXiv.2010.11929

11. Hatamizadeh A, Yang DH, Roth HR, Xu D. *UNETR: transformers for 3D medical image segmentation*. ArXiv (2021). doi: 10.48550/arxiv.2103.10504

12. Liu Z, Lin Y, Cao Y, Qiu J, Wei Y, Zhang ZG, et al. *Swin transformer: hierarchical vision transformer using shifted windows*. ArXiv (Cornell University (2021). doi: 10.48550/arxiv.2103.14030

13. Hatamizadeh A, Nath V, Tang Y, Yang DH, Roth H, Xu D. *Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images*. ArXiv (2022). doi: 10.48550/arxiv.2201.01266

14. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD: J Chronic Obstructive Pulmonary Dis*. (2010) 7:32–435. doi: 10.3109/15412550903499522

15. DenOtter TD, Schubert J. *Hounsfield unit*. StatPearls Publishing (2023).

16. Yevgeniy V. *Feasibility study incorporating lung function imaging into radiation therapy for lung cancer patients*. ClinicalTrials.gov (2016).

17. Zheng S, Nie W, Pan L, Zheng B, Shen Z, Huang L, et al. A dual-attention V-network for pulmonary lobe segmentation in CT scans. *IET Image Process*. (2021) 15:1644–54. doi: 10.1049/ipr2.12133

18. Eslick EM, Kipritidis J, Gradinscak D, Stevens MJ, Bailey DL, Harris B, et al. *CT Ventilation as a functional imaging modality for lung cancer radiotherapy (CT-vs-PET-Ventilation-Imaging) (Version 1) [Data set]*. The Cancer Imaging Archive (2022). doi: 10.7937/3ppx-7s22

19. Castillo E, Castillo R, Vinogradskiy Y, Dougherty M, Solis D, Myziuk N, et al. Robust CT ventilation from the integral formulation of the jacobian. *Med Phys*. (2019) 46:2115–25. doi: 10.1002/mp.13453

20. Zhou Z, Siddiquee M, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging*. (2020) 39:1856–67. doi: 10.1109/tmi.2019.2959609

21. Oktay O. *Attention u-net: Learning where to look for the pancreas*. arXiv preprint (2018). doi: 10.48550/arXiv.1804.03999

22. Zhang S, Yuan H, Cao H, Yang M, Zhang C. SCLMnet: A dual-branch guided network for lung and lung lobe segmentation. *Biomed Signal Process Control*. (2023) 86:105211–11. doi: 10.1016/j.bspc.2023.105211

23. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: A Cancer J Clin*. (2020) 70:7–305. doi: 10.3322/caac.21590

24. Zhang Z, Ren J, Tao X, Tang W, Zhao S, Zhou L, et al. Automatic segmentation of pulmonary lobes on low-dose computed tomography using deep learning. *Ann Trans Med*. (2021) 9:291–91. doi: 10.21037/atm-20-5060

25. Lin T, Wang Y, Liu X, Qiu X. *A survey of transformers*. AI Open (2022). doi: 10.1016/j.aiopen.2022.10.001

26. Thorwarth D. Functional imaging for radiotherapy treatment planning: current status and future directions-a review. *Br J Radiol*. (2015) 88. doi: 10.1259/bjr.20150056

27. Barber J, Yuen J, Jameson M, Schmidt L, Sykes J, Gray A, et al. Deforming to best practice: key considerations for deformable image registration in radiotherapy. *J Med Radiat Sci*. (2020) 67:318–32. doi: 10.1002/jmrs.417

28. Vinogradskiy Y, Castillo R, Castillo E, Schubert L, Jones BL, Faught A, et al. Results of a multi-institutional phase 2 clinical trial for 4DCT-ventilation functional avoidance thoracic radiation therapy. *Int J Radiat OncologyBiologyPhysics*. (2022) 112:986–95. doi: 10.1016/j.ijrobp.2021.10.147

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 February 2025
DOI 10.3389/fonc.2025.1520972

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Yao Luo,
Sichuan University, China
Chita Ranjan Sahoo,
Indian Council of Medical Research (ICMR), India

*CORRESPONDENCE
Anusha Klett
anusha.klett@intelligent-oncology.org

†These authors have contributed equally to this work and share first authorship

RECEIVED 01 November 2024
ACCEPTED 14 January 2025
PUBLISHED 19 February 2025

CITATION
Klett A, Raith D, Silvestrini P, Stingl M, Bermeitinger J, Sapre A, Condor M, Melachrinos R, Kusterer M, Brand A, Pisani G, Ullrich E, Follo M, Duque-Afonso J and Mertelsmann R (2025) Leveraging automated time-lapse microscopy coupled with deep learning to automate colony forming assay. *Front. Oncol.* 15:1520972. doi: 10.3389/fonc.2025.1520972

COPYRIGHT
© 2025 Klett, Raith, Silvestrini, Stingl, Bermeitinger, Sapre, Condor, Melachrinos, Kusterer, Brand, Pisani, Ullrich, Follo, Duque-Afonso and Mertelsmann. 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.

Leveraging automated time-lapse microscopy coupled with deep learning to automate colony forming assay

Anusha Klett[1,2*†], Dennis Raith[3,4†], Paula Silvestrini[1,2,5†], Matías Stingl[4], Jonas Bermeitinger[4], Avani Sapre[1,2,4], Martin Condor[4], Roman Melachrinos[4], Mira Kusterer[6], Alexandra Brand[6], Guido Pisani[1,2], Evelyn Ullrich[1,7,8,9], Marie Follo[6,10], Jesús Duque-Afonso[6] and Roland Mertelsmann[1,2]

[1]Collaborative Research Institute Intelligent Oncology (CRIION), Freiburg, Germany, [2]Mertelsmann Foundation, Freiburg, Germany, [3]Neurorobotics Lab, Department of Computer Science, University of Freiburg, Freiburg, Germany, [4]Department of Applied Computer Sciences, LABMaiTE GmbH, Freiburg, Germany, [5]Laboratory of Applied Cellular and Molecular Biology, Institute of Veterinary Sciences of the Litoral (ICIVET), Universidad Nacional del Litoral (UNL) - National Scientific and Technical Research Council (CONICET), Esperanza, Argentina, [6]Department of Medicine I, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany, [7]Department of Pediatrics, Experimental Immunology and Cell Therapy, Goethe University, Frankfurt am Main, Frankfurt, Germany, [8]German Cancer Consortium Deutsches Konsortium für Translationale Krebsforschung (DKTK) and German Cancer Research Center (DKFZ), Partner Site Frankfurt, Frankfurt, Germany, [9]Frankfurt Cancer Center (FCI), Frankfurt, Germany, [10]Lighthouse Core Facility, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Introduction: The colony forming assay (CFA) stands as a cornerstone technique for evaluating the clonal expansion ability of single cancer cells and is crucial for assessing drug efficacy. However, traditional CFAs rely on labor-intensive, endpoint manual counting, offering limited insights into the dynamic effects of treatment. To overcome these limitations, we developed an Artificial Intelligence (AI)-assisted automated CFA combining time-lapse microscopy for real-time tracking of colony formation.

Methods: Using B-acute lymphoblastic leukemia (B-ALL) cells from an E2A-PBX1 mouse model, we cultured them in a collagen-based 3D matrix with cytokines under static conditions in a low volume (60 µl) culture vessel and validated its comparability to methylcellulose-based media. No significant differences in final colony count or plating efficiency were observed. Our automated platform utilizes a deep learning and multi-object tracking approach for colony counting. Brightfield images were used to train a YOLOv8 object detection network, achieving a mAP50 score of 86% for identifying single cells, clusters, and colonies, and 97% accuracy for Z-stack colony identification with a multi-object tracking algorithm. The detection model accurately identified the majority of objects in the dataset.

Results: This AI-assisted CFA was successfully applied for density optimization, enabling the determination of seeding densities that maximize plating efficiency (PE), and for IC50 determination, offering an efficient, less labor-intensive method for testing drug concentrations. In conclusion, our novel AI-assisted automated colony counting platform enables automated, high-throughput analysis of colony dynamics, significantly reducing labor and increasing accuracy. Furthermore, it allows detailed, long-term studies of cell-cell interactions and treatment responses using live-cell imaging and AI-assisted cell tracking.

Discussion: Future integration with a perfusion-based drug screening system promises to enhance personalized cancer therapy by optimizing broad drug screening approaches and enabling real-time evaluation of therapeutic efficacy.

KEYWORDS

automated colony forming assay, time-lapse microscopy, primary B-ALL cells, artificial intelligence, personalized cancer therapy, live cell imaging

Highlights

- The study introduces an AI-assisted, automated CFA that integrates time-lapse microscopy-based drug screening for dynamic, real-time insights into treatment effects.
- The AI model, trained on brightfield images, achieved high accuracy in colony identification, with promising results for real-time analysis of B-ALL cells.
- This innovative approach optimizes drug screening processes and supports personalized cancer therapy development.

1 Introduction

A colony-forming assay (CFA) is a type of cell survival assay that measures the ability of a single cancer cell to grow into a colony and have an unlimited ability to expand. This proliferative capacity of cancer cells can be used to identify the potential of the cell to form cancer and relapse (1, 2). Hematopoietic cancer cells, such as leukemia cells, exhibit diverse growth properties *in vitro*. These cells can grow in either liquid culture or semi-solid media depending on their characteristics. Healthy hematopoietic progenitor cells, in contrast, can grow in semi-solid media enriched with cytokines but do not proliferate in liquid culture. Similarly, primary leukemia cells from patients typically grow in semi-solid media, although in rare cases, they may also grow in liquid culture (3). Creating a microenvironment that promotes self-renewal without inducing differentiation is also difficult. The precise combination and concentrations of cytokines are critical but hard to optimize. Moreover, the natural bone marrow microenvironment is complex, involving stromal cells, extracellular matrix components, and signaling molecules. Reproducing this *in vitro* is challenging. Therefore, the choice of culture medium and supplements is crucial in the expansion of these cells *ex-vivo* (4, 5). Traditional CFA is performed in a very labor-intensive manual way where the cells are seeded in a 6-well plate embedded in a semi-solid medium such as methylcellulose supplemented with necessary cytokines and nutrients. One of the main applications of CFA is to perform screening of potential drugs with anti-tumor activity (6–8). In drug screenings, the antitumor agents are applied to the cells at the time of seeding and the cells are allowed to form colonies over 7-10 days (9). Therefore, it is also an important technique in estimating drug inhibitory effect (IC_{50}) which is defined as the concentration of a drug required to inhibit the growth of colonies by half (8, 9). This dose-response curve allows us to determine the lowest concentration which has inhibitory effects and therefore less toxicity when administered in patients. Moreover, CFA is also used to measure biological damage to the cells after ionizing radiation treatment (10).

An arbitrary threshold is set for the definition of a colony (more than 15 cells). At the end point, the number of colonies is counted manually under a microscope, and parameters such as plating efficiency (PE) or survival fraction (SF) are evaluated to assess the efficiency of the assay (11). PE is calculated as the number of colonies formed at the final timepoint divided by the number of cells initially seeded, while SF is determined by dividing the number of colonies formed after treatment by the number of cells seeded and the PE.

Although a powerful tool to measure cell proliferative ability and to screen for the effectiveness of chemotherapy drugs, this method has many limitations. The availability of large amounts of samples is needed to seed 6-well plates and many plates to screen multiple drugs. Furthermore, it is extremely cumbersome to count the colonies under the microscope manually, and often individual biases make it complicated to obtain consistent results. Evaluating the colonies at the endpoint does not take into account the response of the cells to the drugs over time. Therefore, tracking not only colonies but also single cells and clusters over time provides a comprehensive understanding of proliferation rates of different cells, clonogenic potential, and resistance mechanisms (9, 11).

Automated CFA using high-throughput methods, such as conducting the assay in a 96-well microplate with fluorescence microscopy, presents a significant improvement over the traditional approach (12). The Agilent BioTek Cytation 5 cell imaging multi-mode reader, with its wide field of view, is particularly advantageous as it allows for efficient screening in a 96-well format, saving considerable time and resources. However, this method relies on staining cells with crystal violet or fluorescent dyes to visualize colonies, which introduces some limitations (13). Specifically, it does not allow for the tracking of live cells over time, thereby failing to capture the dynamic nature of colony formation. This lack of real-time monitoring could overlook crucial aspects of cellular behavior, particularly how colonies evolve and respond to treatments across different time points. Several AI-driven colony counting systems exist, such as Axion Biosystems' Omni platform, which integrates live-cell analysis with an AI-powered clonogenic

assay module. This system enables the automated evaluation of parameters such as PE and SF, reducing manual labor while increasing accuracy and reproducibility (7, 8, 11). However, current AI-based systems often have limitations, including high costs associated with proprietary microscopes and restricted plate formats, as is the case with Omni, which only supports 6-well plates. Additionally, many existing solutions are incompatible with pre-existing lab equipment and with microfluidic platform integration adding further cost. In a broader context, AI tools such as convolutional neural networks (CNNs) could be trained on colony images to improve colony detection and classification across various experimental conditions (14, 15). Moreover, open-source platforms like CellProfiler and ImageJ can be integrated with machine learning models for more flexible, cost-effective solutions. Additionally, transfer learning techniques, where pre-trained models can be fine-tuned to new datasets, also hold promise for improving colony recognition without requiring extensive computational resources or custom hardware (16–19). Although manual counting is still considered the gold standard, it is labor-intensive and can introduce variability and bias due to operator influence. To address these challenges, recent efforts have focused on automating cell identification and quantification (20). An automated colony counting software ideally should reduce the manual labor when it comes to colony counting, incorporate automated time-lapse microscopy to acquire as much information as possible to follow the formation of colonies over time and enable low-volume culture vessels to address low sample availability.

CFA is routinely used for drug screening and testing combination therapies in leukemia cells. In these assays, the leukemia cells are cultured in methylcellulose and exposed to increasing concentrations of various drugs. Colonies are counted 7-10 days after treatment to assess the effectiveness of the drugs (21, 22). Moreover, the colonies are counted at the endpoint which ignores the dynamics of colony formation over time. Acute Lymphoblastic Leukemia (ALL) is a hematologic malignancy of the lymphoid lineage that accounts for 25% of all childhood cancers. With multi-agent chemotherapy, the fatality of pediatric ALL has been reduced by ~90%. In adults, the response has not been as superior as in pediatric patients. Nevertheless, immunotherapies like CAR-T cell therapies, and CD-19 targeted T cell engagers have emerged as new treatment options specifically in B-cell ALL or B-ALL (23–25). ALL is classified based on the cell type (immature precursor of lymphoid lineage), immunophenotyping, and genetic features of the leukemic cells. When the malignancy arises from a precursor B-cell, it is termed a B-cell ALL and T-cell progenitor cells derived ALL are termed a T-cell ALL or T-ALL (26, 27). The E2A-PBX1 fusion gene plays a crucial role in the development of a specific subtype of B-ALL. This gene is formed by a translocation between parts of chromosomes 1 and 19 and is found in approximately 3-5% of pediatric B-ALL cases. In this study, we aimed to utilize the E2A-PBX1 pre-B-ALL cells from an established murine model to develop an Artificial Intelligence (AI)-assisted automated colony forming assay, integrating time-lapse microscopy and microfluidics-based drug screening to facilitate a time efficient method to accurately evaluate the effects of combination therapies. To ensure compatibility with microfluidic systems, we intend to use collagen as the matrix which can better mimic the natural *in-vivo* microenvironment (3, 22). Our main goal was to reduce labor-intensive manual counting and to get more accurate insights into the dynamics of colony formation throughout the experiment with the help of automated imaging and finally integrating a microfluidics-based approach to automate drug applications.

2 Materials and methods

2.1 Cell culture

The m159 primary cells used are derived from a population of B-ALL cells isolated from a mouse model expressing the E2A-PBX-1 fusion gene (22, 22). The E2A-PBX-1 positive mouse B-ALL cells, hereafter referred to as m159, were kindly provided by Prof. Dr. Jesús Duque-Afonso.

2.2 Reagents

For methylcellulose-based CFA, m159 cell suspension in complete IMDM [cIMDM, Iscove's Modified Dulbecco's Medium (Stem Cell) supplemented with 10% fetal bovine serum (FBS; Gibco), 1x L-glutamine (Gibco) and 1% Pen Strep (10,000 Units/ml Penicillin, 10,000 µg/ml Streptomycin; Gibco)] containing IL-7 (10 ng/ml, Stem Cell) was mixed with MethoCult™ (Stem Cell). For collagen-based CFA, m159 cell suspension in cIMDM containing IL-7 was mixed with a collagen solution (1.09 mg/ml). This collagen solution was prepared using Collagen Type I Rat Tail (5 mg/ml, Ibidi) and 1 M NaCl, 7,5% $NaHCO_3$, and millipore H_2O. For IC_{50} determination, JQ1 (S7110, Selleckchem), Prednisolone (PRDL, S1737, Selleckchem), and Daunorubicin (DNR, S3035, Selleckchem) were tested.

2.3 Transitional methodology from traditional to automated collagen-based CFA

A density of 50,000 cells/ml of m159 cells were cultured in both collagen-based and methylcellulose-based media. 60 µl of each cell suspension was seeded in triplicate in a µ-Slide 15-well 3D (Ibidi) and incubated in the BioTek Lionheart FX Automated Microscope. A stage-top incubator system (Ibidi) was coupled to the microscope to maintain controlled conditions of temperature (37°C), CO_2 (5%), and humidity (95%). Time-lapse images were captured every 12 h for 4 days. Six regions of interest (ROIs) were analyzed per well, and brightfield images were acquired at 10x magnification. The absolute cell number was determined by manually counting single cells, clusters, and colonies at both the start and end time points. Colony-forming efficiency (% PE) was calculated as the number of colonies formed at the end time point divided by the number of single cells at the start point.

2.4 Collagen-based CFA for cell density optimization

Different cell densities of m159 cells (25,000, 50,000, 100,000, and 150,000 cells/ml) were cultured in a collagen-based medium. In parallel, 25,000 cells/ml were cultured in a methylcellulose-based medium as a positive control. Each cell density was seeded in triplicates at 60 µl per well in a 15-well slide and incubated in the UC2 Investigator Automated Microscope (UC2i) at 37°C, 5% CO_2, and 95% humidity controlled by the stage top incubator system. The UC2i microscope was developed in-house in cooperation with LABMaiTE GmbH & OpenUC2 GmbH. Image acquisition was performed every 8 h for 5 days and consisted of 25 Z-slices with 30 µm separation, resulting in a total height of 750 µm. Two ROIs were considered per well and brightfield images were captured at 10x magnification. Z-stacks of time-lapse images generated were annotated using Roboflow (28), an annotation tool for ground truth labeling, with the labels single-cell, cluster, and, compact- and dispersed-colony. Additionally, any out-of-focus clump of cells that could become a cluster or colony was labeled as cluster-candidate or colony-candidate, respectively (Supplementary Figure S1). Once annotated, the dataset was exported and used for AI model training.

2.5 Automated colony counting using deep learning & multi-object tracking

The analysis process for counting colony formation on microscopy Z-stacks consists of two major stages. The first stage is the localization and classification of objects on a single 2D image or slice of the Z-stack which results in a list of objects as well as their class, coordinates, and size of the enclosing rectangle around the object, per image. We apply the multi-object tracking (MOT) BOTSort algorithm to merge the slice information along the Z-stack to form a 2.5D representation of unique objects and their location (29).

The approach was chosen as colony formation in a 3D culture usually involves scanning along a large vertical range to capture all colonies growing in different layers of the material. This eliminates the possibility of merging objects strictly by their x and y location as this could combine multiple colonies into a single detection. Multi-object tracking is usually applied in the temporal dimension where the spatial movement of objects is tracked and objects are likely to disappear or be occluded occasionally and therefore have to be re-identified but not confused with other objects traveling across the same location in the meantime. This domain can be easily reformulated to fit the need for tracking objects along a Z-stack as similar events such as positional drift, out-of-focus, or occlusion by another object provide an equal challenge here.

Given the substantial volume of data generated per imaged position and the large number of unique positions captured in each experiment, we decided on a real-time object detection model, specifically the YOLOv8-m model provided by the *ultralytics* library (30). YOLOv8 is a popular, efficient, and powerful Deep Convolutional Neural Network that can perform real-time bounding box detection and instance segmentation tasks. We attempt to keep the computational cost of the analysis process low to keep this solution accessible for consumer-level hardware and preserve reasonable analysis runtime for a full experimental analysis of the 11520 image slices per chip in these experiments. Real-time models usually sacrifice accuracy for faster inference speed which is why we expect a lower performance compared to larger but slower models. This work however investigates the feasibility of the method in general which can be easily adapted to incorporate a different model once its applicability has been shown.

Individual object detections on each independent slice of a stack are combined into trajectories using the BOTSort tracking algorithm. Classification of a complete track is done by taking the maximum of all classes present in the trajectory as class IDs are ordered hierarchical in terms of importance or relevance, i.e. candidates (not relevant unless actual class detected), cells (a subset of following classes), cluster (consisting of cells, a subset of a colony), colony (extended definition of a cluster). A typical trajectory starts with an object being recognized as a candidate first since by definition these are objects that are out of focus (and therefore not reliably classifiable) but their shape and size allow the hypothesis that a larger, relevant object will come into focus. It is followed by one or multiple cluster or colony detections after again moving out of focus and becoming a candidate again. This approach provides reliable classification as multiple focal planes are considered in the classification procedure as long as the true class is at least detected in one slice of the stack.

The detector model was trained on a dataset of 82 microscopy images (4024x3036 pixels) containing 2.732 annotated objects across 4 distinct classes: candidate objects, individual cells, cell clusters, and colonies. Due to ambiguity and under-representation of the cluster-candidate, colony-candidate as well as dispersed- and compact colony classes in the original dataset, we decided to combine these into the *candidate* and *colony* classes respectively. Each training image was augmented to produce 2 additional images to enhance the robustness and generalization of the model. The choice of augmentation transformations included 90-degree rotations, random cropping (0-25% zoom), random rotations (-45° to +45°), brightness adjustments (± 15%), exposure modifications (± 5%), and Gaussian blur up to 3.6 pixels, resulting in 246 training images. In addition, 29 images have been held out from the training data and augmentation and are used to evaluate the detector performance after the hyperparameters for the detection and tracking have been jointly optimized using the Optuna framework (31). We optimize the parameters at the same time as the tracking behavior heavily depends on what kind of detections are generated by the upstream model. Note that the used hyperparameters of the detector (confidence, NMS threshold, and runtime augmentations) only alter the inference behavior but the underlying trained model will remain the same.

Separately, five Z-stacks (19 slices per stack, time points 8 and 10) taken from existing experiments have been annotated using CVAT with MOT annotations to perform and evaluate the tracking procedure (32). Each track was assigned the proper class using the maximum operator according to the previous definition. To reduce the annotation effort, only cluster and colony classes have been

annotated as these are the primary classes considered in this work and following experiments. The python library *pymotmetrics* was used to evaluate the tracking output and compute the necessary MOT scores (33). Optuna was used to jointly optimize the YOLOv8 detection and BOTSort hyperparameters for 100 iterations on one of the five annotated Z-stacks. The product of the IDF1 score and percentage of objects tracked (mostly-tracked and partially tracked, relative to total object count) was used as the optimization objective. We aim to maximize this objective and include the tracked percentage metric to emphasize recall of the model over precision as consistent annotation of in-focus vs. out-of-focus in this setting turned out to be very challenging. Hyperparameters and details for the training of the YOLOv8 model can be found in Supplementary Figure S2. The final optimized parameters for both YOLOv8 inference and BOTSort which were used for the analysis of the biological results can be found in Supplementary Figure S3.

2.6 Collagen-based CFA for IC_{50} determination

A 5-point drug titration was set up as follows. For JQ1, PRDL, and DNR, a 10x dilution was prepared from the stock in cIMDM, followed by 1:5 serial dilutions. The tested concentrations were 500 nM, 100 nM, 20 nM, 4 nM, and 0.8 nM for JQ1, 125 nM, 25 nM, 5 nM, 1 nM, and 0.2 nM for PRDL, and 25 nM, 5 nM, 1 nM, 0.2 nM, and 0.04 nM for DNR. m159 cells (100.000 cells/ml) were resuspended in Megacult™(Stem cell)with IL-7 and then mixed with collagen solution and the corresponding 10x dilution of each drug to achieve a 1x final concentration. Controls include m159 cells mixed with dimethyl sulfoxide (DMSO, 1:1000, Sigma) as negative control and with cIMDM as an untreated (UT) control. JQ1, PRDL, DNR, and control groups were all tested in the same experiment using a 4-slide adapter (Ibidi) coupled to the UC2i microscope. Each slide contained triplicates of each dilution/condition per drug. Image acquisition was performed every 12 h for 5 days and consisted of 20 Z-slices with 30 µm separation, resulting in a total height of 600 µm. Two ROIs were considered per well and brightfield images were captured at 10x magnification. Data were normalized to the DMSO control. The IC50 was calculated by fitting a sigmoidal dose-response curve to the normalized total number of colonies and clusters.

2.7 Statistical analysis

Raw data from the AI-detection model was further processed and statistical analyses were performed using GraphPad Prism 8 software (version 8.4.3). Each experiment was performed in triplicate. For the comparison between methylcellulose and collagen-based media, data of cell, and cluster counts were analyzed using a two-way ANOVA Results 3.1, Figures 1A–C, while a Student's t-test was used to analyze colony counts and PE data Figure 1C. In the cell density experiments shown in Results 3.3, Figures 2A–F, comparisons were made using a two-way ANOVA to assess the effects of seeding cell density and time, while one-way ANOVA was used to analyze differences in PE Figure 2G. P values < 0.05 were considered significant. For the IC_{50} determination experiments (Results 3.4, Figure 3), the dose-response curve was made by fitting a nonlinear curve (sigmoidal, 4-parameter model) of the normalized colony and cluster count on a logarithmic scale for JQ1 and PRDL, and a simple linear regression curve of the normalized colony and cluster count on a logarithmic scale was used to estimate IC_{50} for DNR. Results are displayed as each individual experiment or as the mean of the three experiments with the corresponding standard error of the mean (SEM).

3 Results

3.1 Establishing a collagen-based CFA

To develop a robust, user-friendly, and microfluidics-compatible CFA, the traditional methodology was modified by specifically replacing the methylcellulose medium with collagen for cell growth and significantly reducing the volume used. The m159 cells were cultured in collagen-based media supplemented with Megacult and IL-7, seeded in a 15-well slide, and incubated under controlled conditions in the microscope at 37°C with 5% CO_2 and 95% humidity. Images were acquired every 12 h, consisting of 10 Z-slices with 10 µm separation, covering a total height of 100 µm. Single cells, clusters, and colonies within 6 ROIs were manually counted at the start and after a 4-day incubation. Only in-focus Z-stacks were included to avoid overcounting. While images were taken at 12-h intervals, data analysis was focused on the initial and final time points.

To assess the efficiency of colony formation in a collagen-based medium in a low-volume culture vessel and to confirm its comparability to a methylcellulose-based medium, the absolute cell counts and their capability to proliferate in both media were evaluated. For single cells, no significant effect of the matrix, time, or interaction between these factors was detected (p_{matrix} = 0.1375, p_{time} = 0.2740, and $p_{interaction}$ = 0.5176). Accordingly, no differences were observed in the total number of single cells grown in collagen compared to methylcellulose medium at any time point Figure 1A. Clusters, defined as groups of 2-14 cells, showed a significant effect of time (p_{time} = 0.0081), but neither the matrix nor interaction had a significant effect (p_{matrix} = 0.0570, and $p_{interaction}$ = 0.0962) Figure 1B. Colonies, defined as groups of more than 15 cells, were only analyzed at the final time point since only single cells or small clusters were present initially. No significant differences were found in the number of colonies formed in the collagen-based medium compared to the methylcellulose-based medium (p = 0.4304) Figure 1C. Lastly, the colony-forming efficiency (PE) after 4 days was similar in both matrices (p = 0.5640) Figure 1D. These results demonstrate the efficiency of scaling down from the traditionally used CFA volume of 3 ml down to the 60 µl volume used in these experiments. Indeed, the ability of B-ALL cells to grow and form colonies in a collagen matrix was not affected, making it an excellent option for use in combination with microfluidics.

10.3389/fonc.2025.1520972

3.2 Automated colony counting using deep learning & multi-object tracking

We evaluated both, colonies and clusters together as well as colony and cluster tracking individually to assess the class-based performance (only ground truth and detections for specific class considered). We averaged the results per evaluation across the four tracking datasets which have an average unique object count of 96.5 clusters and 28 colonies. Averaged and individual dataset results can be seen in Table 1. Note that the dataset with id 126 is not present as this dataset was used for the hyperparameter optimization. We focused on the IDF1 score, which is the F1-Score adapted to the MOT scenario, as well as the mostly- and partially-tracked metric, which measures how many of the unique objects have been detected by the pipeline. Similarly, the mostly-lost metric reflects how many objects have not been tracked or less than 20% of their trajectory.

We found an IDF1 score of 0.752 with 61.2% of objects tracked more than 80% of their trajectory (mostly-tracked), 28.2% between 80% and 20% (partially-tracked) and 10.6% failed to be recognized by our system (mostly-lost). The majority of the missed objects appeared to be smaller clusters caused by inconsistencies in the annotations. Out of the total of 112 colonies present in the dataset, only 3 are missed by the pipeline (2.7%). These misses can be explained by large colonies that span across a wide range of focal planes, exposing different portions of the colony in focus at each step and therefore the detection model recognizing only portions of the colony in contrast to the complete colony including the out-of-focus parts (Figure 4). These large changes in coordinates cause the tracker to not be able to associate the objects, therefore failing to combine them into a single trajectory. Similarly, the partial detection of colonies can also lead to a recognition of a large colony as multiple smaller sub-colonies, resulting in an overestimation of colony count. Both issues should be addressable by revisiting the annotation criteria of colonies and including out-of-focus parts to get more consistent large colony detection. Furthermore, we observed that the general definition of cluster (15 cells or less) vs. colony (more than 15 cells) generally seems to be picked up well by the model but occasionally sees misclassifications towards the decision boundary.

We also noted that our experiment shows no ID-switching errors, indicating that the tracker is capable of re-associating objects very reliably which is important as ID switches would cause a single true object to be counted multiple times. Overall, a qualitative inspection yielded a reliable tracking and detection mechanism with

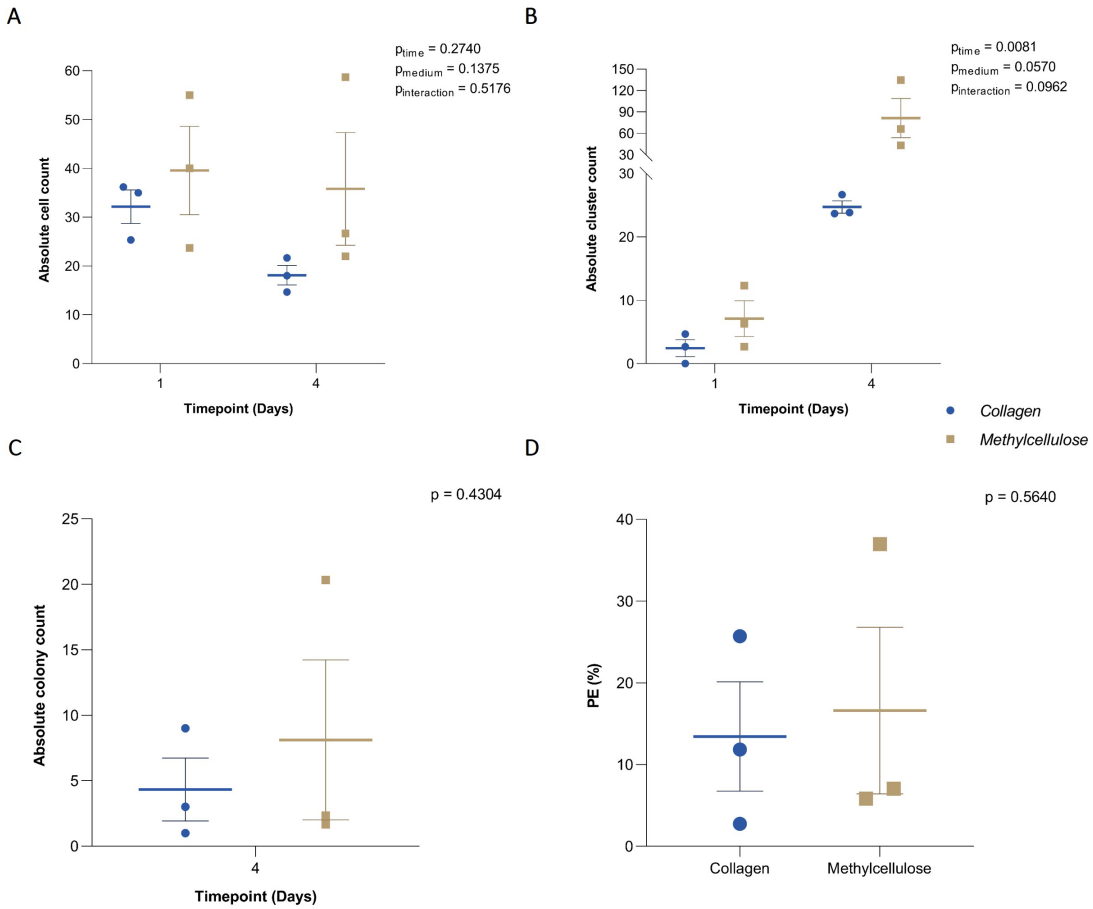


FIGURE 1
Efficiency of collagen-based CFA. **(A)** Absolute cell count, **(B)** clusters, and **(C)** colonies grown on collagen- and methylcellulose-based media and manually counted at baseline and after 4 days of incubation. **(D)** Plating efficiency (PE) in collagen and methylcellulose-based media after 4-day incubation. Individual results from each independent experiment (n = 3) are plotted alongside the mean ± SEM.

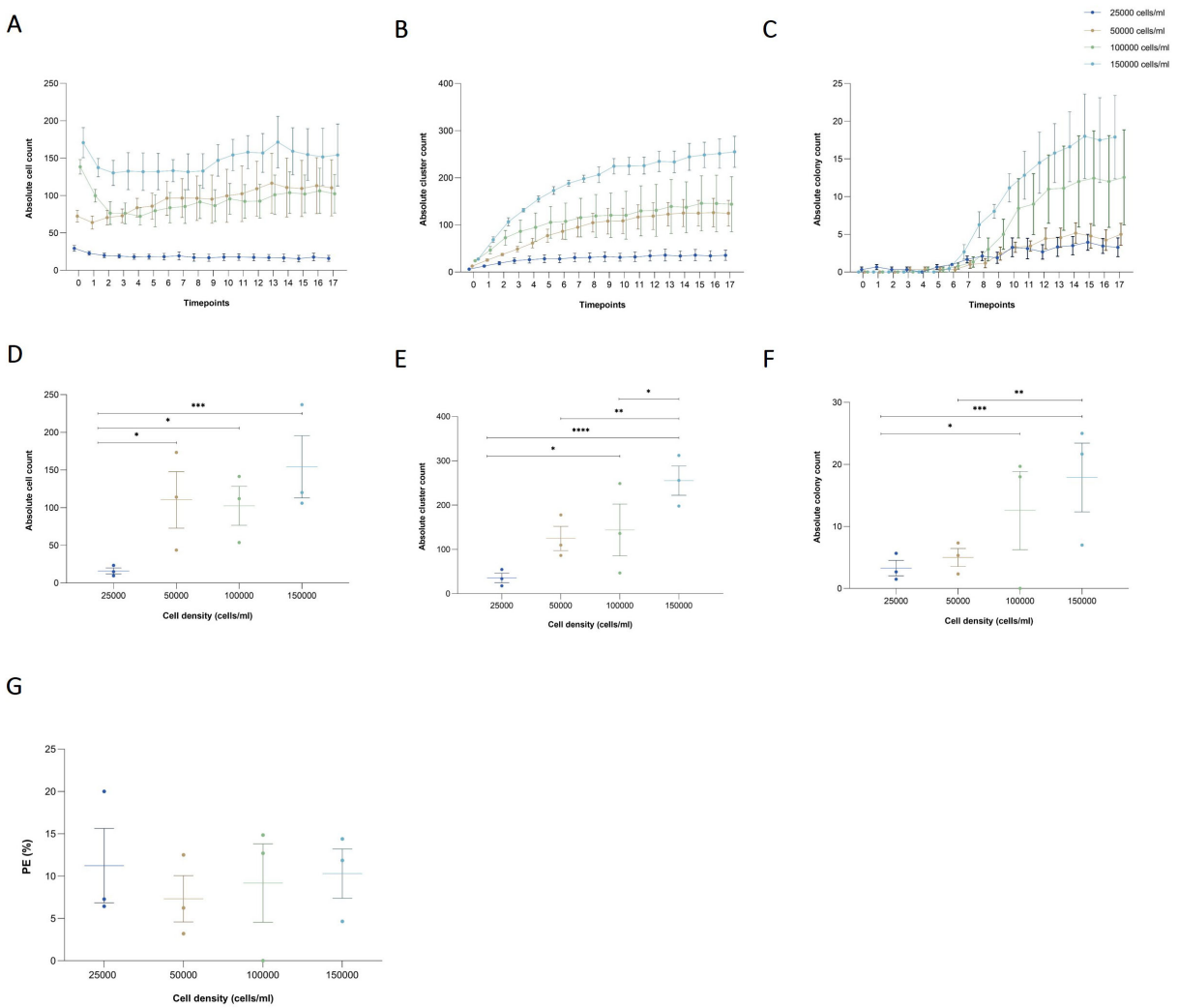


FIGURE 2
Effects of seeding cell density on cell, cluster, and colony counts in a collagen-based CFA. **(A)** Absolute cell count, **(B)** cluster, and **(C)** colony over time for varying initial seeding densities (25,000, 50,000, 100,000, and 150,000 cells/ml) enumerated by the AI-detection model. Results are presented as the mean ± SEM from three independent experiments. **(D)** Absolute counts of single cells, **(E)** clusters, and **(F)** colonies at the final time point (day 4) across the different cell densities. Individual results from independent experiments are presented, along with mean ± SEM. Statistically significant differences between groups are indicated by asterisks (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). **(G)** PE across different seeding cell densities after 4 days of incubation. Individual results from independent experiments are presented, along with mean ± SEM.

occasional errors in edge cases. An example of the pipeline output can be found in Figure 5.

With optimized parameters for the detector and the tracking algorithm, we can additionally evaluate the detector on the initial manual annotations to assess the frame-by-frame classification performance. In contrast to the tracking datasets, annotations for candidates and single-cell classes are also present and the classification performance for these objects can also be assessed.

We found that the evaluation results on the 29 ground truth images confirm the results seen in the tracking evaluation. The detector achieved a mean Average Precision (mAP) of 0.668 at an Intersection over Union (IoU) threshold of 50% across all classes. Table 2 shows that performance varied considerably between classes, with colonies showing the highest mAP50 of 0.861, followed by clusters (0.722), cells (0.697), and candidates (0.392). The lower scores for clusters and cells can most likely be explained by inconsistencies in the ground truth due to the different focal planes and subjective decisions on which potential object to annotate and which one is too far out of focus. The confusion matrix shown in Figure 6 supports this by revealing a large number of supposedly false positives for single cells and clusters, which upon inspection turn out to be unlabeled objects, either due to the focus conditions or due to the object being enclosed by a larger object (e.g. cells within a cluster annotation). Only in rare cases actual false positives, e.g. detecting debris are seen which can be explained by limited examples of debris in the training dataset. Overall, the results show that the detection model can correctly recognize the majority of objects in the dataset although with confusion in the classification caused by overlapping or inconsistent class definitions.

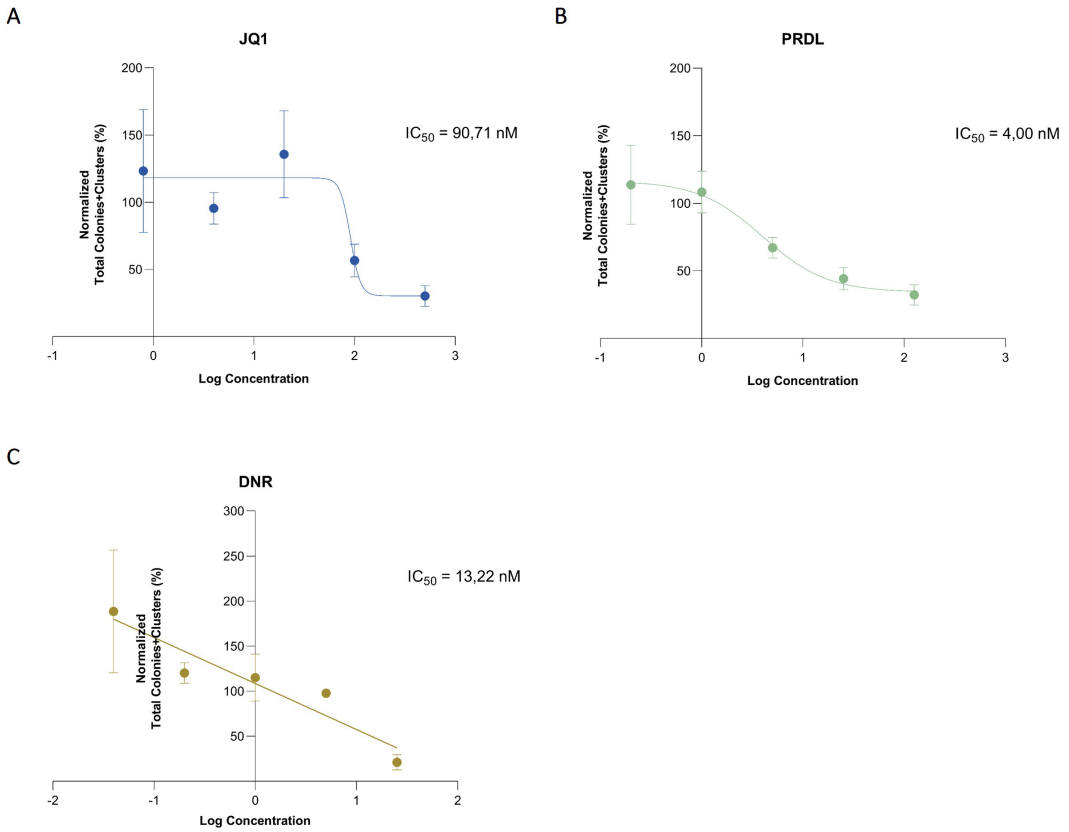


FIGURE 3
IC_{50} determination using data generated by the AI software. Sigmoidal curves using a 4-parameter nonlinear model for **(A)** JQ1 and **(B)** PRDL and linear regression for **(C)** The total number of colonies and clusters was normalized to the DMSO control and the concentrations were represented in logarithmic scale. Results are presented as the mean ± SEM from three independent experiments.

3.3 AI-assisted CFA for cell density optimization

After establishing the collagen-based CFA, we aimed to utilize this setup to optimize and select the most suitable cell seeding density that leads to the most optimal PE, ensuring consistent and reliable colony formation results. Furthermore, the initial seeding density mustn't lead to overcrowded colonies which makes it difficult to track and count manually as well as for the AI detection model.

The detection model recognized, counted, and tracked single cells, clusters, and colonies throughout the time. For single cells, the absolute count remained consistent over time across all tested seeding densities. A significant effect of cell density on cell number was observed, but there was no effect of time or the interaction between both factors ($p_{cell\ density} < 0.0001$, $p_{time} = 0.8253$, and $p_{interaction} > 0.9999$) Figure 2A. By the final time point, after 8 days of incubation, the number of single cells in the lowest density group (25,000 cells/ml) was significantly lower than in the 50,000, 100,000, and 150,000 cells/ml groups ($p = 0.0181$, $p = 0.0366$, and $p < 0.0001$, respectively) Figure 2D. For clusters, an increase in the absolute count was observed starting at 12 h of incubation, after which the number remained constant until the last time point in all evaluated groups. A significant effect of both cell density and time on the number of clusters was detected, but not of their interaction ($p_{cell\ density} < 0.0001$, $p_{time} < 0.0001$, and $p_{interaction} > 0.7832$) Figure 2B. At the final timepoint, although the total number of clusters formed in the 25,000 cells/ml group was lower than in the other groups, it was only significantly different from the 100,000 and 150,000 cells/ml groups ($p = 0.0237$, and $p < 0.0001$, respectively). While the total number of clusters in the 150,000 cells/ml group was higher than in the other groups, it was only significantly different from the 50,000 and 100,000 cells/ml groups ($p = 0.0039$, and $p = 0.0191$, respectively) Figure 2E. For colonies, an increase in the absolute count was observed after 3 days of incubation across all evaluated groups. A significant effect of both cell density and time on colony number was detected, though there was no significant interaction between these factors ($p_{cell\ density} < 0.0001$, $p_{time} < 0.0001$, and $p_{interaction} = 0.0547$) Figure 2C. At the final time point, although the mean number of colonies in the 25,000 cells/ml group was lower than in the other groups, it was only significantly different from the 100,000 and 150,000 cells/ml groups ($p = 0.0419$, and $p = 0.0003$, respectively). Similarly, while the mean number of colonies in the 150,000 cells/ml group was higher than in the other groups, it was only significantly different from the 50,000 cells/ml group ($p = 0.0017$) Figure 2F. Finally, no significant differences were observed in PE among the groups tested ($p = 0.8931$) Figure 2G. As a result, the key criterion for selecting the

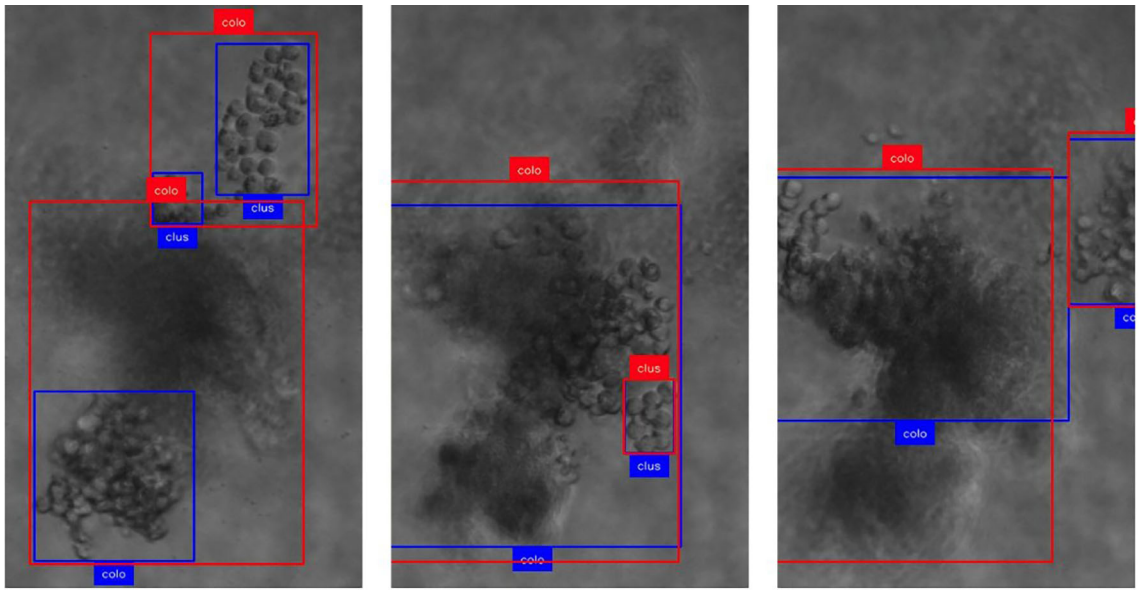


FIGURE 4
Demonstration of focus-related challenges for the detection and tracking mechanism. Boxes in red show ground truth annotations, and blue display model predictions. The text represents the class provided by ground truth and model, respectively. (Col, colony; clus, cluster). As the focal plane progresses along z (left to right), specific parts of the colony are detected and tracked which leads to rapid changes in the (center-) location of the object.

optimal seeding density was based on colony visualization and ensuring adequate space for cells to grow. Based on the findings, we determined that 100,000 cells per well is the optimal seeding density for future experiments. Manual colony counting would yield only a final count for each density, making the process both time-consuming and labor-intensive. The implementation of AI-assisted CFA reduced the time and effort required to analyze a large number of samples, while also providing deeper insights into colony formation dynamics, including single-cell and cluster counts over time.

3.4 AI-assisted CFA for IC_{50} determination

The applicability of the AI-assisted CFA developed was further evaluated by assessing the ability of B-ALL cells to form colonies and clusters after treatment and determining the IC_{50} for different drugs. Furthermore, the effects of the individual drugs on the number of cells, clusters, and colonies over time provide insights into the effectiveness of the treatments. To begin, IC_{50} values for JQ1, DNR, and PRDL were established using the traditional CFA with manual colony counting as a reference (data not shown). Based on these values, five concentrations within the upper and lower IC_{50} range were selected to evaluate the performance of the detection model to detect and track single cells, clusters, and colonies over time. The AI software accurately detected the three categories generating raw data that included the absolute number of single cells, clusters, and colonies for each ROI, time point, and Z-plane across all wells. Consequently, single-cell, cluster, and colony counts were analyzed over time for all tested drug concentrations and control groups. The absolute cell count followed a consistent pattern, remaining stable over time across all evaluated groups, including those treated with JQ1, PRDL, and DNR, as well as the control groups (DMSO and UT) (Figure 7). Regarding cluster formation, the absolute cluster count remained constant over time in cells treated with higher concentrations of JQ1 (500 nM and 100 nM), PRDL (125 nM, 25 nM, and 5 nM), and DNR (25 nM), whereas it increased over time in cells treated with lower concentrations of JQ1 (20 nM, 4 nM, and 0.8 nM), PRDL (1 nM and 0.2 nM), and DNR (5 nM, 1 nM, 0.2 nM, and 0.04 nM) (Figures 8A–C, respectively). In the control groups, both UT and DMSO-treated cells showed a steady increase in the absolute cluster count, following a similar pattern Figure 8D. In terms of colony formation, it was completely inhibited in cells treated with higher concentrations of JQ1 (500 nM and 100 nM), PRDL (125 nM, 25 nM, and 5 nM), and DNR (25 nM), as the absolute colony count remained at zero over time. In contrast, the absolute colony count increased after approximately 60 h of incubation in cells treated with lower concentrations of JQ1 (20 nM, 4 nM, and 0.8 nM), PRDL (1 nM and 0.2 nM), and DNR (5 nM, 1 nM, 0.2 nM, and 0.04 nM) (Figures 9A–C, respectively). Similarly, for the control groups (both UT and DMSO-treated cells), colonies began to form after approximately 60 h of incubation, with the absolute colony count steadily increasing over time Figure 9D.

For the determination of IC_{50}, absolute counts for each ROI and Z-plane at the final time point across all conditions were estimated using the presented AI-detection model. In this context, the total number of colonies and clusters was considered to optimize the model's fit and accurately estimate the IC_{50}, thereby reducing the standard deviation and improving the r^2 value. The estimated IC_{50} values were 90.71 nM for JQ1, 4 nM for PRDL, and for DNR 13,22 nM (Figure 3).

10.3389/fonc.2025.1520972

TABLE 1 Detection & tracking results.

Dataset	Class	# Objects	Mostly tracked	Partially tracked	Mostly lost	idf1	idp	idr
Average	all	124	76	35	13.25	0.752	0.739	0.770
Average	cluster	97	61	23	13.00	0.728	0.692	0.768
Average	colony	28	15	12	1.50	0.805	0.850	0.772
123	cluster	91	49	24	18	0.682	0.670	0.695
123	colony	28	15	12	1	0.831	0.927	0.752
123	all	119	64	36	19	0.730	0.746	0.715
124	cluster	103	78	14	11	0.764	0.696	0.847
124	colony	14	14	0	0	0.887	0.839	0.940
124	all	117	93	14	10	0.788	0.719	0.872
125	cluster	80	51	17	12	0.749	0.713	0.789
125	colony	31	10	17	4	0.704	0.800	0.629
125	all	111	62	36	13	0.744	0.751	0.738
127	cluster	112	64	37	11	0.715	0.690	0.739
127	colony	39	21	17	1	0.799	0.835	0.766
127	all	150	85	54	11	0.748	0.739	0.753

MOT scores of the analysis pipeline were evaluated on four MOT datasets after being optimized on a separate dataset using Optuna.
Objects represent the number of unique objects and their trajectories present in the dataset. The "Average" dataset represents the averaged scores across the four datasets. Mostly and Partially tracked are considered a successful recognition while Mostly lost indicates untracked objects.

4 Discussion

Transitioning from a traditional CFA approach to an AI-assisted automated assay offers numerous advantages, including the reduction of labor-intensive manual counting tasks, elimination of human error, increased throughput, real-time monitoring, and automated analysis. CFA has long been a foundational technique for evaluating the effects of various drugs in ALL. In this study, together with our collaborators, we aimed to transform the traditional approach of CFA for a certain type of murine E2A-PBX1 B- ALL cells to a more efficient AI-assisted CFA approach coupled with time-lapse microscopy.

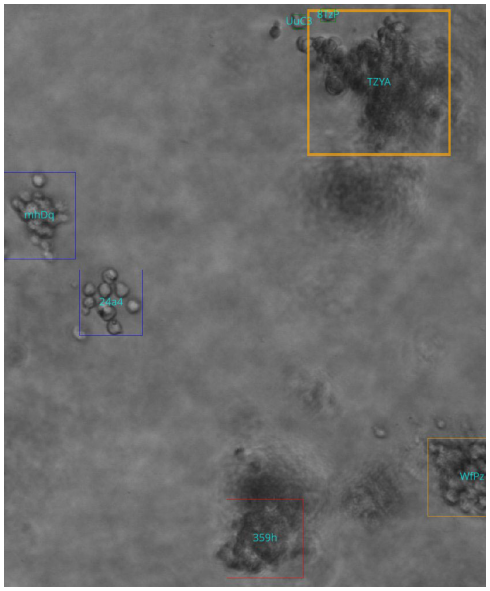


FIGURE 5
Example output of the analysis pipeline. Box colors represent classification output at the current z-index, and text inside the box represents the unique track ID assigned. Colors represent class (red: candidate, green: single cell, blue: cluster, yellow: colony).

Our first approach was to scale down from a 6-well plate format to a 15-well slide format, thereby enhancing throughput even with limited sample quantities. This miniaturization, reducing the traditional 3 ml requirement in CFA to just 60 µl, not only preserves the ability of B-ALL cells to grow and form colonies effectively but also offers a major economic advantage. While not the primary focus of this study, the reduced volume requirement could greatly benefit research groups enabling them to conduct many more tests with the same amount of resources. Moreover, the small volumes required make it particularly advantageous when working with limited patient samples, maximizing the use of precious biological material. This scalability and compatibility with microfluidic platforms make this method an attractive option for wider adoption in various experimental setups.

Secondly, we aimed to integrate microfluidics-based approaches for automated drug application. Therefore, we focused on finding an alternative to methylcellulose as a semi-solid medium as we found it to be unsuitable for perfusion-based assays. We tested and found rat tail collagen to be a good alternative for methylcellulose as it did not dissolve during perfusion-based assays. Subsequently, we validated that the scaling down approach and switching to collagen did not show any significant differences in colony formation. The

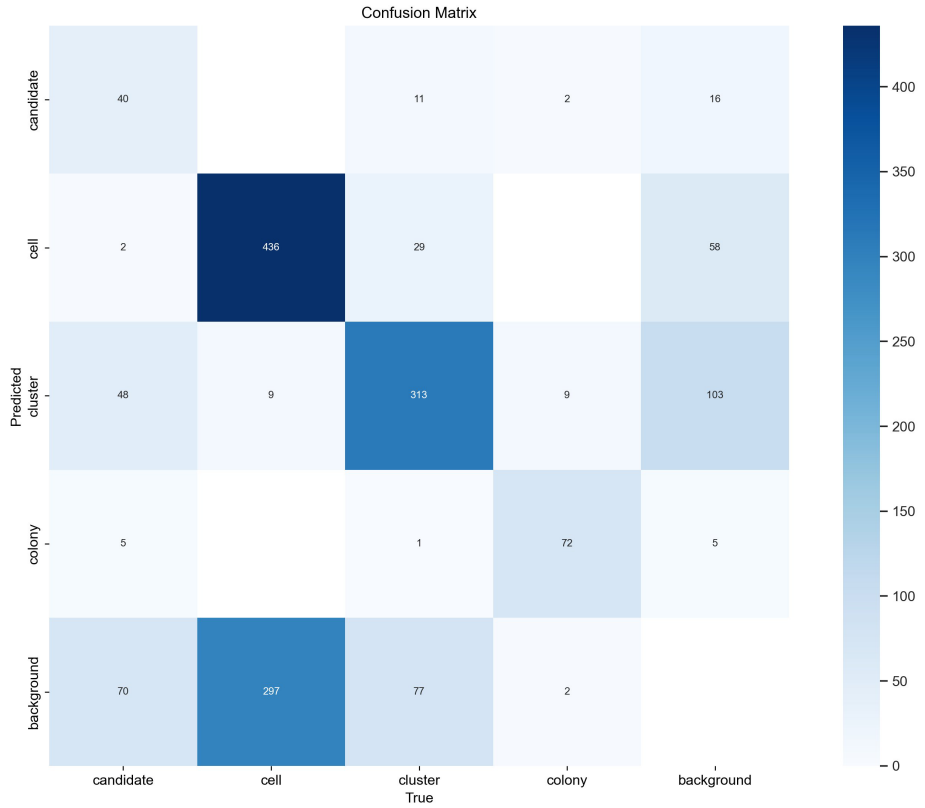


FIGURE 6
YOLOv8 confusion matrix. The confusion matrix of the final object detector was evaluated on 29 test images after hyperparameter optimization was performed on the tracking datasets. Rows represent class predictions while columns represent true ground truth annotations. Diagonal elements represent correct predictions while off-diagonal elements represent misclassifications. The number in each entry corresponds to the number of times this combination of prediction vs. ground truth occurred.

strategy behind developing an automated CFA involved leveraging automated time-lapse microscopy to continuously record cell colony formation over time. These images were then used to train an AI detection model capable of accurately identifying single cells, clusters, and colonies, and providing comprehensive analysis of critical parameters such as PE, which is essential for effective drug screening. Furthermore, tracking single-cell and cluster counts at multiple time points, rather than relying solely on endpoint colony counts as in traditional CFAs, provides a deeper understanding of cell behavior dynamics, such as differential proliferation rates and resistance mechanisms. Investigating single cells that persist after drug exposure and evade apoptosis could also offer critical insights for evaluating the efficacy of different therapeutic agents. Given that time-lapse microscopy enables prolonged observation of cells, it naturally generates an enormous amount of data. To manage this, we strategically limited imaging to 2 to 4 ROIs within each well. Additionally, because the cells are embedded in collagen and grow in a three-dimensional manner, Z-stacks were employed to capture the Z-plane, ensuring that the full depth and structure of the colonies were accurately recorded and analyzed. Roboflow and CVAT software were used to annotate the images using various classes including colony and cluster candidates. To ensure accurate

TABLE 2 Object detection evaluation results.

Class	Images	Instances	Precision	Recall	mAP50	mAP50-95
all	29	1423	0.722	0.594	0.668	0.451
candidate	18	165	0.565	0.236	0.392	0.27
cell	27	742	0.819	0.58	0.697	0.356
cluster	28	431	0.647	0.724	0.722	0.492
colony	22	85	0.855	0.835	0.861	0.685

Evaluation results of the YOLOv8 object detector on the 29 manually annotated test images, separated by class. mAP is the mean Average Precision and represents the average precision between ground truth and prediction boxes at different overlap thresholds (50%, 50-95%).

AI model training, it was essential to label all visible objects in the images, including blurry colonies or clusters that were only in focus on a different plane.

Our proposed pipeline combines a YOLOv8 object detector with the BOTSort tracking algorithm to analyze colony formation in 3D cell cultures. The detector achieves good performance for the main classes of interest - colonies (mAP50 0.861) and clusters (mAP50 0.722). Lower performance on candidate objects and individual cells can be attributed to annotation inconsistencies or ambiguities rather than model limitations. Our tracker successfully maintains object identities across focal planes with IDF1 scores of 0.805 for colonies and 0.728 for clusters, with notably zero identity switches. This indicates that our adaptation of temporal tracking algorithms to the spatial domain of Z-stacks is effective. The system shows particular strength in tracking colonies, with on average 95% of colonies being either mostly or partially tracked.

Colony formation is not purely determined by the number of cells seeded; it also depends on the proximity and interactions between cells. This means that simply increasing the number of cells seeded does not necessarily lead to a proportionate increase in the number of colonies. Cooperative behavior disrupts the expected linear relationship between cell seeding and colony formation, making the clonogenic assay less predictable (34). To determine the optimal seeding density, various initial densities must be tested, which increases manual labor. Therefore, we tested the AI detection model to assess colony formation across different seeding densities to streamline the process and help identify the ideal conditions for accurate results. The preliminary results generated by the AI software served as a baseline for further downstream analysis.

To further validate the AI model, we performed IC_{50} as a part of a drug screening process. Traditionally, this is a cumbersome process, where the cells are treated with varying concentrations of drugs and monitored for colony formation and the colonies are counted manually at the endpoint to provide a relative colony count between treated and non-treated control samples. We utilized our automated microscopy and multiplexing approach to screen five different concentrations of three drugs, alongside a control chip. This innovative method allowed us to conduct a drug screening that differs from traditional techniques by enabling the tracking of individual cells to observe their dynamic responses to the drugs. Through this approach, we were able to capture the formation of colonies over time allowing us to capture crucial insights such as volume, size, and shape of the colonies, that are often overlooked when using conventional methods that only count colonies at the endpoint.

Our evaluation of the colony counting software against manual counts demonstrated Mean Absolute Errors (MAE) of 24 for single cells, 12 for clusters, and 1 for colonies. These findings align with the higher performance of the AI detection model in identifying colonies and lower performance in identifying single cells. The detection model occasionally misclassified larger single cells as clusters, leading to a slight overestimation in cluster counts. Additionally, colonies that were correctly classified at earlier time points were sometimes reclassified as clusters at later time points.

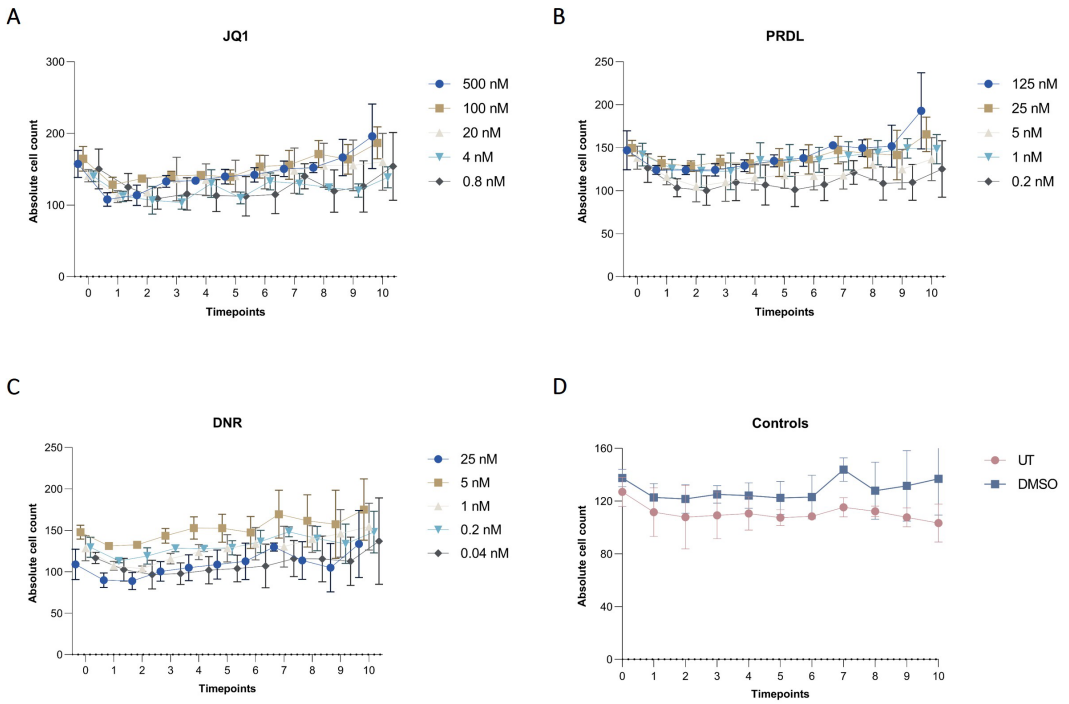


FIGURE 7
Cell dynamics for different drugs and concentrations in collagen-based CFA over time. Absolute cell counts for m159 cells treated with different concentrations of **(A)** JQ1 (500 nM, 100 nM, 20 nM, 4 nM, and 0.8 nM), **(B)** PRDL (125 nM, 25 nM, 5 nM, 1 nM, and 0.2 nM), **(C)** DNR (25 nM, 5 nM, 1 nM, 0.2 nM, and 0.04 nM), and **(D)** control groups, including untreated (UT) and DMSO-treated cells enumerated by the AI-software. Results are presented as the mean ± SEM from three independent experiments.

10.3389/fonc.2025.1520972

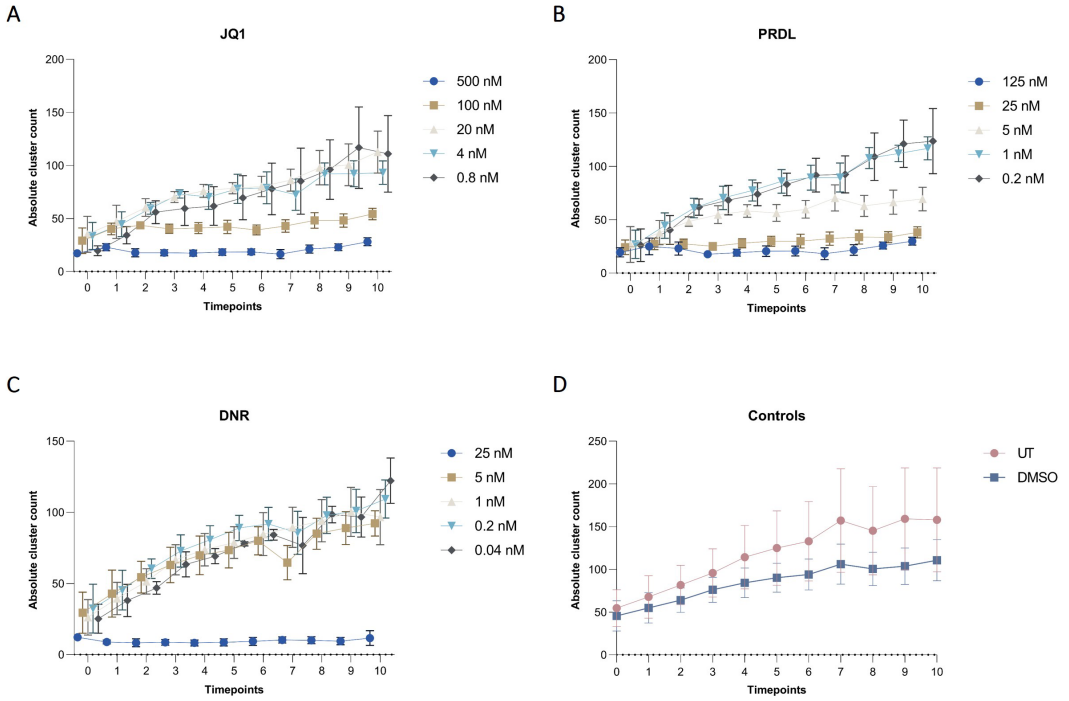


FIGURE 8
Cluster dynamics for different drugs and concentrations in collagen-based CFA over time. Absolute cluster counts for m159 cells treated with different concentrations of **(A)** JQ1 (500 nM, 100 nM, 20 nM, 4 nM, and 0.8 nM), **(B)** PRDL (125 nM, 25 nM, 5 nM, 1 nM, and 0.2 nM), **(C)** DNR (25 nM, 5 nM, 1 nM, 0.2 nM, and 0.04 nM), and **(D)** control groups, including untreated (UT) and DMSO-treated cells enumerated by the AI-software. Results are presented as the mean ± SEM from three independent experiments.

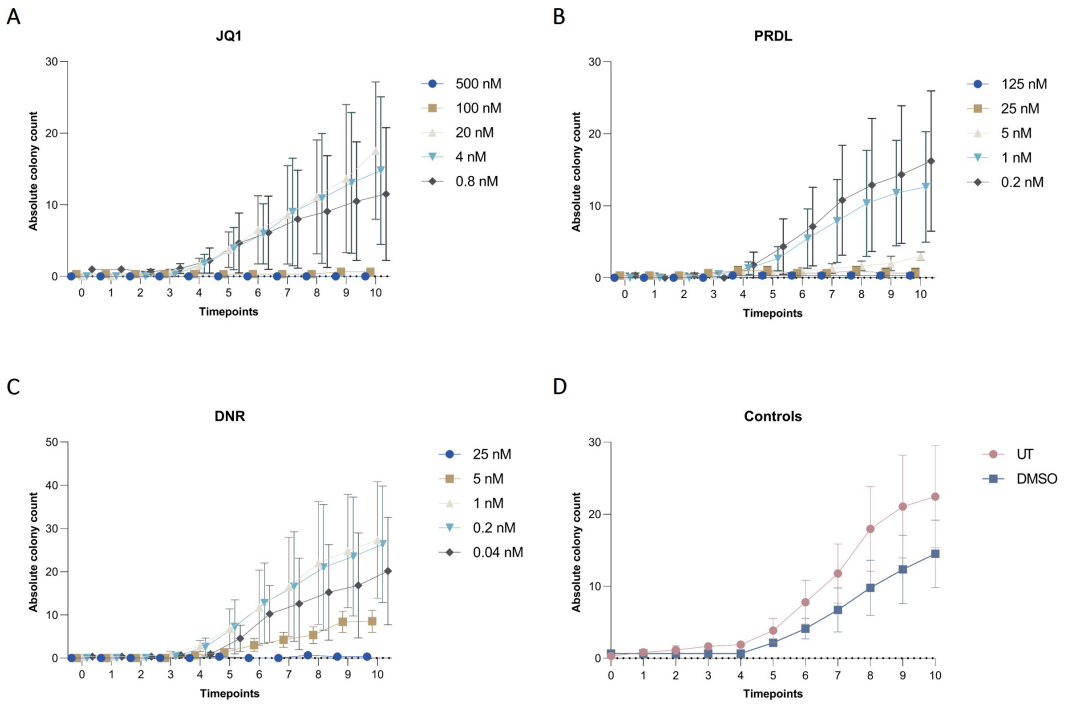


FIGURE 9
Colony dynamics for different drugs and concentrations in collagen-based CFA over time. Absolute colony counts for m159 cells treated with different concentrations of **(A)** JQ1 (500 nM, 100 nM, 20 nM, 4 nM, and 0.8 nM), **(B)** PRDL (125 nM, 25 nM, 5 nM, 1 nM, and 0.2 nM), **(C)** DNR (25 nM, 5 nM, 1 nM, 0.2 nM, and 0.04 nM), and **(D)** control groups, including untreated (UT) and DMSO-treated cells enumerated by the AI-software. Results are presented as the mean ± SEM from three independent experiments.

The detection model also missed several single cells, leading to undercounts in this category. These discrepancies may reflect limitations in our training annotations and variations in manual counting. By minimizing these sources of human error such as creating a curated dataset that has been validated by multiple experts, the AI pipeline holds promise for standardized reproducible, and unbiased CFA.

Microfluidic devices are advanced platforms that are used as a way to deliver nutrients or compounds to cells closely mimicking physiological conditions (35–37). Although they have not fully replaced animal models, advancements in organoid or 3D cell culture are paving the way for more sophisticated *in-vitro* models to mimic drug treatments *in vivo*. Microfluidic devices enable continuous or intermittent perfusion of media containing nutrients, cytokines, or drugs at a constant flow rate.

In our system, we integrated a microfluidic device with automated nutrient and drug delivery, using Ibidi perfusion-based chips that allow cells to be embedded in collagen. Although we were able to successfully integrate the system, there are still parameters that need to be optimized to ensure the appropriate microenvironment for colony growth and formation. Our goal is to validate this approach in future studies. We expect to demonstrate that, once fully optimized and combined with AI-assisted colony-forming assays, *in vitro* microfluidics-based screening has the potential to revolutionize therapeutic approaches and advance personalized medicine.

The transition from traditional static assays to microfluidic systems presents significant opportunities but also several challenges that need to be addressed. One critical aspect is the establishment of microfluidic conditions, particularly in determining the appropriate drug concentrations and application frequency. Unlike traditional assays where the drug remains in a static environment, microfluidic systems continuously flush out drugs, which can lead to wastage and non-circulatory dynamics. This necessitates precise calibration to ensure that drug concentrations are effective while minimizing wastage.

Moreover, the AI detection model we currently employ is capable of generating raw data, but there is untapped potential in automating the subsequent analysis and data visualization steps. This advancement would not only streamline workflows but also enable more comprehensive, real-time insights into cell behavior during drug screening. Another advantage of software could be to predict the colony behavior and suggest an earlier or later endpoint to evaluate the effect of drugs or provide cell seeding density suggestions.

To date, our efforts have been focused on murine E2A-PBX1 B-ALL cells and cell lines, but an important future direction involves expanding the application to patient-derived samples. This would provide a more clinically relevant understanding of how therapies function in human cancer cells. Additionally, exploring combination therapies using this platform could uncover synergies between different drugs, which are often critical in complex diseases like cancer. Another area interesting for exploration is the biological underpinnings of why certain cells fail to form colonies. Understanding the heterogeneity in colony formation and why some cells behave differently could provide important insights into mechanisms of cancer relapse and resistance, offering new targets for therapeutic intervention.

5 Conclusions

In conclusion, transitioning from traditional colony formation assays to an AI-assisted, automated approach offers substantial advantages in terms of reducing manual labor, increasing accuracy, and enabling real-time monitoring. This study successfully demonstrated the feasibility of integrating automated time-lapse microscopy with AI-driven colony analysis for murine E2A-PBX1 B-ALL cells. By scaling down to a chip format and switching to collagen as a medium, we achieved comparable results to conventional methods while enhancing throughput and efficiency. The AI-assisted system not only automates colony counting but also provides dynamic insights into cell behavior during drug screening, capturing critical events like colony splitting. Overall, this innovative approach has the potential to significantly advance therapeutic development and personalized medicine, offering a more efficient, precise, and scalable method for evaluating treatment efficacy. Some limitations remain, particularly regarding annotation consistency and edge cases involving overlapping objects or ambiguous class definitions. In addition, cluster recognition and tracking could be further improved using additional subclasses or a more consistent single-cell detection, alleviating the need to combine smaller groups of individual cells into clusters. Nevertheless, the overall system demonstrates promising results for automated analysis of colony formation experiments.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

All experiments using mice were performed with the approval of and in accordance with the guidelines of the Administrative Panel on Laboratory Animal Care of Stanford University. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

AK: Conceptualization, Formal analysis, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing, Methodology, Project administration, Visualization. DR: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Data curation. PS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing –

review & editing, Validation. MS: Software, Writing – review & editing. JB: Conceptualization, Investigation, Project administration, Writing – review & editing. AS: Data curation, Methodology, Writing – review & editing. MC: Software, Writing – review & editing. RMel: Project administration, Resources, Writing – review & editing. MK: Methodology, Resources, Writing – review & editing. AB: Methodology, Resources, Writing – review & editing. GP: Data curation, Writing – review & editing. EU: Conceptualization, Supervision, Writing – review & editing. MF: Conceptualization, Methodology, Supervision, Writing – review & editing. JD: Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – review & editing. RMer: Conceptualization, Investigation, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We would like to thank JD-A and his team for providing the B-ALL cells, reagents, and knowledge transfer. The Mertelsmann Foundation provided financial support for the project. DR is funded by the Lab-On-A-Chip AI (LOCAI) grant of the Federal Ministry of Education and Research (BMBF) Grant ID: 01IS21098A. The responsibility for the content of this publication lies with the author. Institutional support and project executing agency: Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), DLR-Projektträger, Softwaresysteme und Wissenstechnologien.

Conflict of interest

Author DR, MS, JB, AS, MC and RM were employed by company LABMaiTE 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1520972/full#supplementary-material

References

1. Shoemaker RH, Wolpert-DeFilippes MK, Kern DH, Lieber MM, Makuch RW, Melnick NR, et al. Application of a human tumor colony-forming assay to new drug screening. *Cancer Res*. (1985) 45:2145–53.

2. Franken NAP, Rodermond HM, Stap J, Haveman J, Van Bree C. Clonogenic assay of cells *in vitro*. *Nat Protoc*. (2006) 1:2315–9. doi: 10.1038/nprot.2006.339

3. Svozilova H, Vojtova L, Matulova J, Bruknerova J, Polakova V, Radova L, et al. *In vitro* culture of leukemic cells in collagen scaffolds and carboxymethyl cellulose-polyethylene glycol gel. *PeerJ*. (2024) 12:e18637. doi: 10.7717/peerj.18637

4. Astori G, Malangone W, Adami V, Risso A, Dorotea L, Falasca E, et al. A novel protocol that allows short-term stem cell expansion of both committed and pluripotent hematopoietic progenitor cells suitable for clinical use. *Blood Cells Mol Dis*. (2001) 27:715–24. doi: 10.1006/bcmd.2001.0439

5. Mertelsmann R, Lindemann A, Oster W, Gamm H, Kolbe K, Herrmann F. Hematopoietic growth factors in oncology. *Cancer Detect Prev*. (1990) 14:613–6.

6. Li B, Liu J, Qu S, Gale RP, Song Z, Xing R, et al. Colony-forming unit cell (CFU-C) assays at diagnosis: CFU-G/M cluster predicts overall survival in myelodysplastic syndrome patients independently of IPSS-R. *Oncotarget*. (2016) 7:68023–32. doi: 10.18632/oncotarget.12105

7. Kaufman DS, Hanson ET, Lewis RL, Auerbach R, Thomson JA. Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc Natl Acad Sci*. (2001) 98:10716–21. doi: 10.1073/pnas.191362598

8. Rajendran V, Jain MV. *In vitro* tumorigenic assay: colony forming assay for cancer stem cells. In: Papaccio G, Desiderio V, editors. *Cancer Stem Cells*. Springer New York, New York, NY (2018). p. 89–95. doi: 10.1007/978-1-4939-7401-6_8

9. Katz D, Ito E, Lau KS, Mocanu JD, Bastianutto C, Schimmer AD, et al. Increased efficiency for performing colony formation assays in 96-well plates: novel applications to combination therapies and high-throughput screening. *BioTechniques*. (2008) 44:ix–xiv. doi: 10.2144/000112757

10. Buch K, Peters T, Nawroth T, Sänger M, Schmidberger H, Langguth P. Determination of cell survival after irradiation via clonogenic assay versus multiple MTT Assay–a comparative study. *Radiat Oncol Lond Engl*. (2012) 7:1. doi: 10.1186/1748-717X-7-1

11. Available online at: http://www.axionbiosystems.com/resources/application-note/quantifying-chemotoxicity-cancer-cell-colony-formation-using-live-cell? (Accessed September 17, 2024).

12. Available online at: www.agilent.com/about/tektalk/en/newsletter-cancer-research.html (Accessed October 19, 2024).

13. Available online at: https://www.agilent.com/cs/library/applications/automated-colony-formation-assay-5994-3403EN-agilent.pdf? (Accessed October 19, 2024).

14. Zhang L. Machine learning for enumeration of cell colony forming units. *Vis Comput Ind BioMed Art*. (2022) 5:26. doi: 10.1186/s42492-022-00122-3

15. Carl SH, Duempelmann L, Shimada Y, Bühler M. A fully automated deep learning pipeline for high-throughput colony segmentation and classification. *Biol Open*. (2020) 9:bio052936. doi: 10.1242/bio.052936

16. Siragusa M, Dall'Olio S, Fredericia PM, Jensen M, Groesser T. Cell colony counter called CoCoNut. Montazeri Aliabadi H, editor. *PloS One*. (2018) 13:e0205823. doi: 10.1371/journal.pone.0205823

17. Bray MA, Vokes MS, Carpenter AE. Using cellProfiler for automatic identification and measurement of biological objects in images. *Curr Protoc Mol Biol*. (2015) 109:14.17.1–14.17.13. doi: 10.1002/0471142727.mb1417s109

18. Guzmán C, Bagga M, Kaur A, Westermarck J, Abankwa D. ColonyArea: an imageJ plugin to automatically quantify colony formation in clonogenic assays. *PloS One*. (2014) 9:e92444. doi: 10.1371/journal.pone.0092444

19. Brzozowska B, Gałecki M, Tartas A, Ginter J, Kaźmierczak U, Lundholm L. Freeware tool for analysing numbers and sizes of cell colonies. *Radiat Environ Biophys*. (2019) 58:109–17. doi: 10.1007/s00411-018-00772-z

20. Shao Z, Buchanan LB, Zuanazzi D, Khan YN, Khan AR, Prodger JL. Comparison between a deep-learning and a pixel-based approach for the automated quantification of HIV target cells in foreskin tissue. *Sci Rep*. (2024) 14:1985. doi: 10.1038/s41598-024-52613-3

21. Duque-Afonso J, Lin CH, Han K, Wei MC, Feng J, Kurzer JH, et al. E2A-PBX1 remodels oncogenic signaling networks in B-cell precursor acute lymphoid leukemia. *Cancer Res*. (2016) 76:6937–49. doi: 10.1158/0008-5472.CAN-16-1899

22. Duque-Afonso J, Feng J, Scherer F, Lin CH, Wong SHK, Wang Z, et al. Comparative genomics reveals multistep pathogenesis of E2A-PBX1 acute lymphoblastic leukemia. *J Clin Invest*. (2015) 125:3667–80. doi: 10.1172/JCI81158

23. Fry TJ, Aplan PD. A robust *in vivo* model for B cell precursor acute lymphoblastic leukemia. *J Clin Invest*. (2015) 125:3427–9. doi: 10.1172/JCI83799

24. Tucker ER, George S, Angelini P, Bruna A, Chesler L. The promise of patient-derived preclinical models to accelerate the implementation of personalised medicine for children with neuroblastoma. *J Pers Med*. (2021) 11:248. doi: 10.3390/jpm11040248

25. Pasvolsky O, Kebriaei P, Shah BD, Jabbour E, Jain N. Chimeric antigen receptor T-cell therapy for adult B-cell acute lymphoblastic leukemia: state-of-the-(C)ART and the road ahead. *Blood Adv*. (2023) 7:3350–60. doi: 10.1182/bloodadvances.2022009462

26. Tang TCY, Xu N, Nordon R, Haber M, Micklethwaite K, Dolnikov A. Donor T cells for CAR T cell therapy. *biomark Res*. (2022) 10:14. doi: 10.1186/s40364-022-00359-3

27. Raetz EA, Teachey DT. T-cell acute lymphoblastic leukemia. *Hematol Am Soc Hematol Educ Program*. (2016) 1:580–8. doi: 10.1182/asheducation-2016.1.580

28. Dwyer B, Nelson J, Hansen T, et al. *Roboflow (Version 1.0) [Software]*. (2024). Available at: https://roboflow.com/research#cite

29. Aharon N, Orfaig R, Bobrovsky BZ. *BoT-SORT: Robust Associations Multi-Pedestrian Tracking [Internet]. arXiv* (2022). Available online at: http://arxiv.org/abs/2206.14651 (Accessed October 31, 2024).

30. Jocher G, Qiu J, Chaurasia A. *Ultralytics YOLO [Internet]* (2023). Available online at: https://github.com/ultralytics/ultralytics (Accessed October 31, 2024).

31. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. (2019). Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, . pp. 2623–31. New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3292500.3330701

32. CVAT.ai Corporation. *Computer Vision Annotation Tool (CVAT)* (2023). Available online at: https://github.com/cvat-ai/cvat.

33. Heindl C. *Pymotmetrics*. Available online at: https://github.com/cheind/py-motmetrics (Accessed October 31, 2024).

34. Brix N, Samaga D, Hennel R, Gehr K, Zitzelsberger H, Lauber K. The clonogenic assay: robustness of plating efficiency-based analysis is strongly compromised by cellular cooperation. *Radiat Oncol*. (2020) 15:248. doi: 10.1186/s13014-020-01697-y

35. Petreus T, Cadogan E, Hughes G, Smith A, Pilla Reddy V, Lau A, et al. Tumour-on-chip microfluidic platform for assessment of drug pharmacokinetics and treatment response. *Commun Biol*. (2021) 4:1001. doi: 10.1038/s42003-021-02526-y

36. Zhai J, Liu Y, Ji W, Huang X, Wang P, Li Y, et al. Drug screening on digital microfluidics for cancer precision medicine. *Nat Commun*. (2024) 15:4363. doi: 10.1038/s41467-024-48616-3

37. Sun J, Warden AR, Ding X. Recent advances in microfluidics for drug screening. *Biomicrofluidics*. (2019) 13:061503. doi: 10.1063/1.5121200

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 February 2025
DOI 10.3389/fonc.2025.1489450

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Ji-Qiao Zhu,
Capital Medical University, China
Sarada Prasad Dakua,
Hamad Medical Corporation, Qatar

*CORRESPONDENCE
Chuansheng Zheng
hqzcsxh@sina.com
Guofeng Zhou
zhouguofeng69@126.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 01 September 2024
ACCEPTED 24 January 2025
PUBLISHED 19 February 2025

CITATION
Chai B, Xiang D, Zhou G and Zheng C (2025) Pilot study of an arterial enhancement fraction-based model for progression prediction in HCC undergoing chemoembolization.
Front. Oncol. 15:1489450.
doi: 10.3389/fonc.2025.1489450

COPYRIGHT
© 2025 Chai, Xiang, Zhou and Zheng. 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.

Pilot study of an arterial enhancement fraction-based model for progression prediction in HCC undergoing chemoembolization

Bin Chai[1,2†], Dongqiao Xiang[1,2†], Guofeng Zhou[1,2*] and Chuansheng Zheng[1,2*]

[1]Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Hubei Province Key Laboratory of Molecular Imaging, Wuhan, China

Objective: To develop a prognostic model including arterial enhancement fraction of residual tumor (AEF-RT) for predicting progression-free survival (PFS) in hepatocellular carcinoma (HCC) patients after drug-eluting beads transarterial chemoembolization (DEB-TACE).

Materials and methods: Between March 2019 and March 2024, 111 HCC patients undergoing DEB-TACE were randomly allocated to a training cohort and a validation cohort in a 7:3 ratio. LASSO regression was applied in the training cohort to identify risk factors for recurrence, which were subsequently used to construct the Cox model. Model performance was assessed using the concordance index (C-index, where 0.5 indicates non-informative discrimination and 1 represents perfect discrimination) and Brier score (ranging from 0 to 1, 0 indicating higher calibration) and was compared with five existing prognostic models.

Results: The final model, termed ADMAN model, incorporated AEF-RT, Diameter, Margin appearance, Aspartate transaminase, and Neutrophil-to-lymphocyte ratio. High-risk patients defined by ADMAN had 4.69 times greater progression risk than low-risk ones in the training cohort ($p < 0.001$) and 3.52 times greater in the validation cohort ($p = 0.005$). The C-index of ADMAN (0.75) was significantly higher than that of other models in the training cohort ($p < 0.05$ for all) and remained significantly higher than three of them in the validation cohort [0.71 vs. 0.55 ($p = 0.041$), 0.54 ($p = 0.033$), 0.53 ($p = 0.004$)]. The ADMAN model showed a significantly lower Brier score than that of other models at 6 months and 12 months in the training cohort ($p < 0.05$ for all). In the validation cohort, the ADMAN model remained to have significantly lower Brier score than the four models ($p < 0.05$) at 6 months, while it had significantly lower score than one model at 12 months.

Conclusions: The AEF-based model may be a promising tool for progression risk stratification in HCC patients undergoing DEB-TACE. Further external validation in independent cohorts with larger sample sizes is necessary to confirm the robustness of the ADMAN model.

KEYWORDS

hepatocellular carcinoma, transarterial chemoembolization, prognosis, multidetector computed tomography, quantitative evaluation

1 Introduction

Transarterial chemoembolization (TACE) is a well-established treatment strategy for patients with hepatocellular carcinoma (HCC) who are not eligible for curative treatments (1). Nevertheless, the objective response rates 6 months after TACE range from 27% to 76% and 70% to 80% in patients who eventually die due to tumor progression (2). Transitioning from TACE to systemic therapies is advisable for patients unlikely to benefit from repeated embolization before liver function deteriorates. Therefore, it is imperative to develop methods for estimating individualized treatment efficacy. Several studies have suggested that 18F-fluorodeoxyglucose uptake, clinical characteristics, tumor radiological features, and certain serum biomarkers may serve as promising indicators for tumor progression after TACE (3–7). However, few studies have developed multivariate algorithms combining clinical and imaging findings to predict progression for individual patients.

A key element that influences the aggressiveness of HCC is tumor neo-angiogenesis, the process of developing new capillary blood vessels that results in tumor vascularization. The presence of high vascularity typically indicates aggressive tumor behavior and is linked to poorer clinical outcomes. CT perfusion may reveal tumor aggressiveness and predict prognosis based on tumor vascularity (8). However, the application of such a technique is limited by the high radiation exposure. Arterial enhancement fraction (AEF) is defined as the ratio of the absolute increment of attenuation in the arterial phase to that of the portal venous phase: $AEF = [(HU_A - HU_U)/(HU_P - HU_U)] \times 100\%$, where HU, A, P, and U represent attenuation, arterial phase, portal phase, and unenhanced, respectively. The strong correlation ($r = 0.91$, $p < 0.001$) between hepatic perfusion and AEF was observed in 10 rabbits with VX2 liver tumor by Kim et al. (9). Thus, AEF is an ideal biomarker, which can be readily derived from routine triphasic liver CT examinations, to indirectly reflect the ratio of hepatic arterial perfusion to total perfusion (10). A recent study showed that elevated AEF of residual tumor after embolization was strongly associated with poor prognosis in drug-eluting beads (DEB) TACE-treated HCC patients (11). Thus, the purpose of the current study was to establish a prognostic model for progression in HCC patients after incomplete DEB-TACE by integrating AEF and clinical–radiological characteristics.

2 Materials and methods

The present analysis adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guideline (12). This retrospective study was approved by the institutional ethics committee and was conducted following the 1975 Declaration of Helsinki. The requirement for informed consent was waived, as all data were anonymized and collected from the electronic medical system.

2.1 Study population

From October 2018 to March 2024, clinical and image information of 253 consecutive treatment-naive patients with unresectable HCC undergoing DEB-TACE were retrieved from the electronic medical database of our tertiary medical center. The follow-up was completed on 30 June 2024. Of these, 142 patients were excluded for the following reasons: (a) extrahepatic metastases (confounding factor for disease progression caused by the primary lesion, n = 6); (b) conditions not eligible for AEF measurement, comprising hemorrhage in HCC lesion (non-contrast agent caused high density, n = 1), HCC involving major branch of portal vein (low background enhancement caused by impaired portal venous blood supply, n = 1), arterioportal shunt (atypical lesion enhancement, n = 7), and complete tumor necrosis after first DEB-TACE treatment (n = 14); (c) missing triphasic CT scan data at baseline or at follow-up (n = 63); (d) disease progression after initial TACE treatment (n = 33); and (e) receiving any locoregional treatments other than TACE at follow-up (n = 17) (Figure 1). The diagnosis of HCC was either biopsy-proven or met the European Association for the Study of the Liver imaging criteria (13).

2.2 Procedure

DEB-TACE technique. All DEB-TACE procedures were performed by a team of interventional radiologists with over 10 years of experience. CalliSpheres (Jiangsu Hengrui Medicine Co. Ltd.) beads were loaded with 60 mg or 80 mg of epirubicin per vial

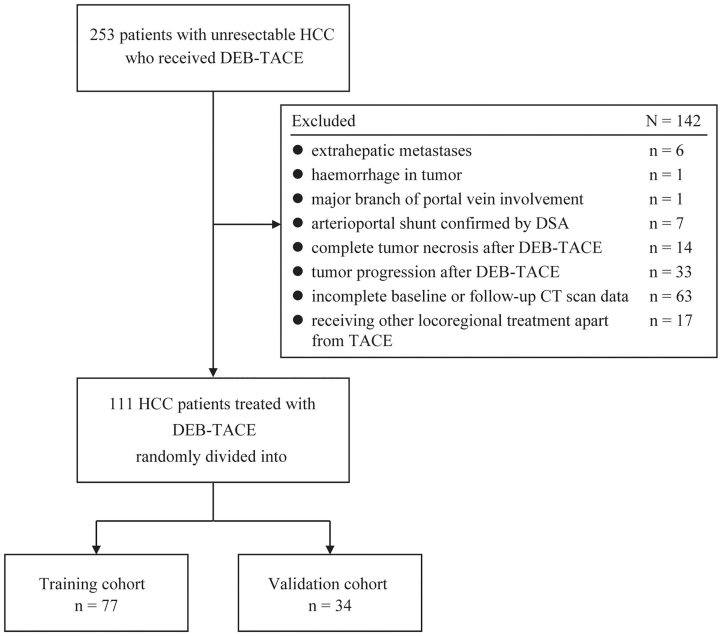


FIGURE 1
Flowchart showing the patient selection and grouping strategy for model derivation and validation. DEB-TACE, drug-eluting bead transarterial chemoembolization; HCC, hepatocellular carcinoma; DSA, digital subtraction angiography.

and mixed with a non-ionic contrast medium to obtain the final injectable beads. The details of the DEB-TACE procedure has been described previously (11).

Follow-up protocol. The "on-demand" TACE procedure was scheduled at 6–12-week intervals based on CT or MR evidence of viable tumor or intrahepatic recurrence, unless contraindications were present. The median number of TACE sessions was 3 (range, 1–4), with one session in six patients, two in 39 patients, three in 51 patients, and four in 15 patients.

Image and AEF acquisition. CT acquisitions in this study were performed using three Siemens CT scanners: Somatom Definition, Somatom Definition AS, or Somatom Force (Siemens Healthcare, Erlangen, Germany). After unenhanced scanning, triphasic contrast-enhanced scans were performed following intravenous administration of 80–100 mL non-ionic contrast agent (Iopamidol, 370 mg I/mL, Bracco) at a rate of 2.5–3.0 mL/s through the antecubital vein using an automatic power injector. The arterial phase, portal venous phase, and equilibrium phase images were acquired at 30 s, 50 s, and 180 s after initiation of the contrast medium injection, respectively. The unenhanced, arterial phase, and portal venous phase datasets were then sent to a syngo.via workstation (Siemens Healthcare, Erlangen, Germany) for generating the quantitative color mapping of AEF by using the dedicated software in the MM Oncology mode. Typically, AEF was calculated based on the ratio of the absolute increment of attenuation during the arterial phase to the absolute increment of attenuation during the portal venous phase per pixel ($AEF = \frac{HU_A - HU_U}{HU_P - HU_U} \times 100\%$, where HU stand for attenuation, A for arterial phase, P for portal phase, and U for unenhanced). The resulting data are mapped to a spectral color scale that displays from black (0%) to red (100%) (10) (Supplementary Figure S1).

2.3 Outcome and predictors

Outcome. The outcome of interest was progression-free survival (PFS) (months), defined as the period between the TACE initiation and radiological detection of tumor progression (n = 109) or death (n = 2). Radiological progression was determined according to the modified response evaluation criteria in solid tumors (mRECIST) (14).

Clinical predictors. On the basis of the literature, the following clinical parameters were collected: sex, age, cause of HCC (hepatitis B infection or other disease), alanine transaminase (ALT), aspartate transaminase (AST), peripheric platelet count, albumin, total bilirubin, Child–Pugh classification, neutrophil-to-lymphocyte ratio (NLR), platelet–to–lymphocyte ratio (PLR), and alpha-fetoprotein. All the laboratory tests and physical examinations were performed within 3 days before the DEB-TACE procedure.

Radiological predictors. Radiological features of each patient were reviewed through pre-therapeutic liver CT examinations scheduled within 1 week before the DEB-TACE procedure. The first follow-up CT scan for treatment response assessment was conducted in 35.7 ± 4.1 days (range, 22–46) after initial treatment. Selected predictors of radiological characteristics included the diameter of the dominant HCC lesion (largest diameter of viable tumor on the axial section of arterial-phase images), margin appearance (smooth or non-smooth), enhancing capsule appearance in portal venous phase or equilibrium phase (absence or presence), lesion number (solitary or multifocal), tumor extent (unilobar or bilobar), and vascular invasion (presence of portal vein tumor thrombosis). The dominant tumor was determined as the largest measurable target lesion per patient. All image analyses, incorporating dominant lesion determination, radiological

characteristics identification, and treatment response assessment, were conducted by two nonauthor abdominal radiologists (14 years and 10 years of experience in abdominal radiology) who were unfamiliar with the study design, and any discrepancy during analysis was resolved by consensus.

AEF of residual tumor. The AEF of residual tumor (hereafter, AEF-RT) was measured in the AEF color map derived from the first follow-up CT scan. The measuring procedure was consistent with the previous study (11). In short, two authors (B.C. and D.Q.X.), blinded to patient outcomes, manually delineated regions of interest on three transverse planes. The mean AEF-RT, calculated across all three planes, was used for analysis.

2.4 Statistical analysis

Continuous variables with normal distribution are expressed as the mean ± standard deviation, and those with non-normal distribution are expressed as the median (interquartile range). Categorical variables are expressed as frequencies (percentages). Missing data in predictors were assumed to be random, and five imputations using chained equations were performed to correct for bias. Unless otherwise specified, statistical analyses were performed using GraphPad Prism (Version 8.3.4) and packages of "mice," "survival," "survminer," "rms," and "compareC" in R software (Version 4.4.1) (http://cran.r-project.org).

Model building procedure. The Least Absolute Shrinkage and Selection Operator (LASSO) regression was applied to select the most significant predictors for PFS from the 19 candidate predictors mentioned above in the training cohort. Tenfold cross-validation of LASSO regression was performed using the R package "glmnet." The final predictors were determined by a stepwise Cox proportional hazards regression algorithm from all significant predictors selected by the LASSO regression at the lambda value corresponding to the minimum error. The proportional hazards assumption of the models was tested by examining the plots of scaled Schoenfeld residuals against time for each variable in the final model.

Generating different progression risk categories. To generate different risk categories, the weighted sum of regression coefficients from the Cox model, or the linear predictor, was computed. Survival curves according to the risk categories were plotted using Kaplan–Meier (KM) method, allowing a visual comparison of discrimination. The more widely separated are the curves, the better is the discrimination. The median PFS of each risk category, *p*-values of the log-rank test, hazard ratio (HR), and *p*-values of the Wald coefficient test were also reported. The calibration, or prediction accuracy of PFS, in the different risk groups was assessed graphically following the procedure proposed by Royston and Altman (15). Briefly, the model-predicted mean survival curves were created by applying fractional polynomial regression to approximate the log baseline cumulative hazard function as a smooth function of time. Then, the KM estimate and model-predicted survival curves were superimposed on the same plot to examine the model calibration visually.

Model performance assessment. Harrell's concordance index (C-index, where 0.5 indicates non-informative discrimination and 1 represents perfect discrimination) was computed to assess the model performance of discrimination. The plots of the time-dependent concordance index were also created to visualize the discrepancy of the C-index between models at each time point (months).

The calibration performance was assessed by the Brier score (from 0 to 1, where 0 indicates better calibration) at 6 months and 12 months using the R package "riskRegression." Additionally, integrated Brier score (IBS) was calculated with inverse probability of censoring weights to adjust for right-censored data using the R package "SurvMetrics" (16).

3 Results

3.1 Clinical characteristics and PFS

In total, 111 patients were included in the study, with 77 and 34 randomly allocated to the training and validation cohort, respectively. Table 1 reports the clinical and disease related information. The median follow-up time was 25 months for the whole cohort, 24 months for the training cohort, and 25 months for the validation cohort. Progression was recorded in 91 patients, with 66 in training cohort and 25 in the validation cohorts. The median PFS was 7 months for both the whole and training cohorts and 8 months for validation cohort.

3.2 Construction of multivariate Cox proportional hazards model for PFS

In this study, the optimal λ value was chosen as 0.11623 (λ.min), leading to the selection of six variables with nonzero coefficients using LASSO regression (Figure 2, Table 2), five of which were retained as significant predictors in the final Cox model. These variables included AEF-RT, tumor diameter, margin appearance (smooth or non-smooth), AST, NLR, and all treated as continuous variables except margin appearance. The parameters of the final Cox model, referred to as the ADMAN model, are depicted in Figure 3A. A nomogram based on the ADMAN model was constructed to estimate the probability of PFS at 6 months, 9 months, and 12 months after TACE for individual patients (Figure 3B).

A nested DMAN model was developed by deleting AEF-RT from the complete model to determine the contribution of AEF-RT in predicting tumor progression. For the model performance, the C-index of the DMAN model was marginally significantly lower than that of the ADMAN model (0.706 vs. 0.754, $p = 0.051$). However, as comparing C-indexes of two nested models is a low-power procedure, the likelihood ratio χ^2 test was used instead to compare the ADMAN and DMAN models (17). A significant loss of predictive power was observed in the likelihood ratio test when AEF-RT was removed from the ADMAN model ($\chi^2 = 9.685$, $p = 0.002$).

TABLE 1 Clinical and disease-related information of 111 patients with HCC.

Characteristics	Whole cohort (n = 111)	Training cohort (n = 77)	Validation cohort (n = 34)	*p*
Median age (years)	57 (24, 84)[†]	60 (24, 84)[†]	52.5 (34, 81)[†]	0.051
Sex (male/female)	97 (87.4%)/14 (12.6%)	67 (87.0%)/10 (13.0%)	30 (88.2%)/4 (11.8%)	0.896
Cause of HCC (Hepatitis B/Other)	97 (87.4%)/14 (12.6%)	68 (88.3%)/9 (11.70%)	29 (85.3%)/5 (14.7%)	0.758
ALT (U/L)[‡]	42.0 (28, 47)	41.0 (26.0, 46.8)	45.4 (30.5, 50.0)	0.324
AST (U/L)[‡]	53.0 (36, 69)	51.0 (36.0, 69.0)	55.0 (37.8, 69.0)	0.797
NLR[‡]	2.6 (2.0, 4.0)	2.5 (2.0, 4.0)	2.9 (1.9, 4.3)	0.537
PLR[‡]	133.0 (100.0, 184.4)	133.0 (108.2, 187.8)	110.4 (84.4, 165.8)	0.089
PLT ($\times10^9$/L)[‡]	165.0 (111.0, 199.0)	166.2 (111.5, 189.5)	151.5 (100.3, 236.8)	0.945
ALB (g/L)[‡]	36.6 (34.8, 39.0)	36.6 (34.7, 39.1)	36.8 (34.8, 39.2)	0.790
TBIL (μmol/L)[‡]	17.6 (13.9, 20.3)	17.6 (14.0, 20.9)	17.7 (13.0, 19.5)	0.936
Child-Pugh class (A/B)	91 (82.0%)/20 (18.0%)	65 (84.4%)/12 (15.6%)	26 (76.5%)/8 (23.5%)	0.315
AFP (ng/mL) (≤400/>400)	63 (56.8%)/48 (43.2%)	40 (51.9%)/37 (48.1%)	23 (67.7%)/11 (32.3%)	0.124
No. of lesions (solitary/multifocal)	56 (50.5%)/55 (49.5%)	39 (50.7%)/38 (49.3%)	17 (50.0%)/17 (50.0%)	0.950
Diameter (cm)	9.2 ± 4.0	9.4 ± 4.2	8.8 ± 3.5	0.507
Vascular invasion (absent/present)	76 (68.5%)/35 (31.5%)	51 (66.2%)/26 (33.8%)	25 (73.5%)/9 (26.5%)	0.446
Tumor extent (unilobar/bilobar)	79 (71.2%)/32 (28.8%)	53 (68.8%)/24 (31.2%)	26 (76.5%)/8 (23.5%)	0.413
Enhancing capsule (absent/present)	63 (56.8%)/48 (43.2%)	43 (55.8%)/34 (44.2%)	20 (58.8%)/14 (41.2%)	0.770
Margin appearance (smooth/non-smooth)	62 (55.9%)/49 (44.1%)	43 (55.8%)/34 (44.2%)	19 (55.9%)/15 (44.1%)	0.997
AEF of residual tumor (%)	44.3 ± 15.9	43.8 ± 15.9	45.7 ± 16.0	0.558
Median PFS (months)	7 (6, 10)[§]	7 (6, 9)[§]	8 (5, 17)[§]	–

Unless otherwise indicated, data in parentheses are percentages. AEF, arterial enhancement fraction; AFP, α-fetoprotein; ALT, alanine transaminase; AST, aspartate transaminase; HCC, hepatocellular carcinoma; NLR, neutrophil-to-lymphocyte ratio; PFS, progression-free survival; PLT, peripheric platelet count; PLR, platelet-to-lymphocyte ratio; PR, partial response; SD, stable disease; TBIL, total bilirubin.
[†]Data in parentheses are range.
[‡]Data are expressed as median (interquartile range).
[§]Data in parentheses are 95% confidence interval.

3.3 ADMAN-based risk categories

The ADMAN linear predictor for observation in the training cohort was computed based on the following formula:

$$\begin{aligned}\text{ADMAN linear predictor} = {} & 0.032 \times \text{AEF (value without percent sign)} \\ & + 0.092 \times \text{Diameter of dominant tumor (cm)} \\ & + 0.723 \times \text{Margin appearance (0 = smooth, 1 = non-smooth)} \\ & + 0.010 \times \text{AST (U/L)} \\ & + 0.199 \times \text{NLR}\end{aligned}$$

The median ADMAN linear predictor (3.55) was used as the cutoff to categorize patients into high- and low-risk groups. This cutoff was then applied to the validation cohort for subsequent grouping.

In the training cohort, good separation of survival curves was achieved for the high- and low-risk groups according to ADMAN-based risk categories [median PFS: 4.5 months vs. 12 months, $p < 0.001$; HR = 4.69 (95% CI, 2.68–8.19), $p < 0.001$] (Figure 4A). In the validation cohort, there remained a significant difference in PFS between the high- and low-risk groups [median PFS: 4.5 months vs. 15 months, $p = 0.003$; HR = 3.52 (95% CI, 1.47–8.43), $p = 0.005$) (Figure 4B).

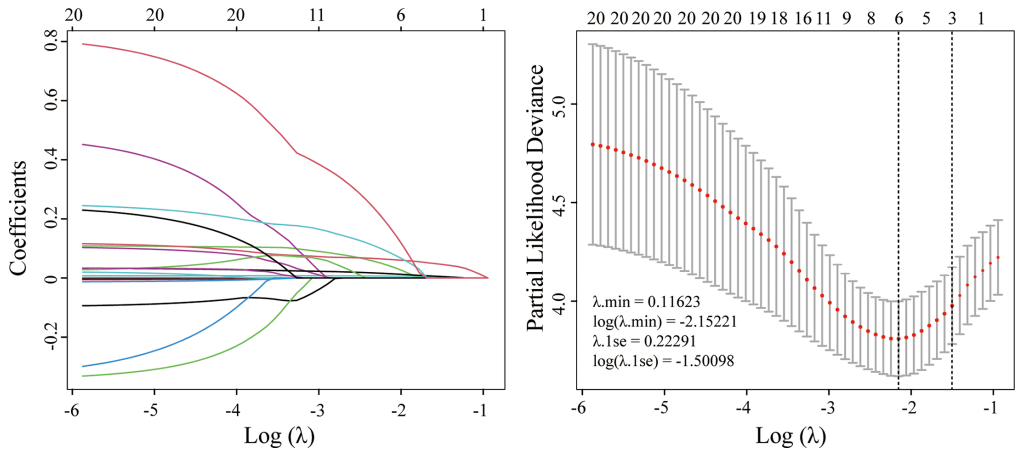


FIGURE 2
LASSO regression for candidate predictors selection.

TABLE 2 Features entered the LASSO for PFS in the training cohort.

Variables	Coefficients (λ.min)	Coefficients (λ.1se)
Age (y)	–	–
Sex (male/female)	–	–
Cause of HCC (hepatitis B/other)	–	–
ALT (U/L)	–	–
AST (U/L)	0.006412359	0.003018912
NLR	0.042218270	–
PLR	–	–
PLT ($\times 10^9$/L)	–	–
ALB (g/L)	–	–
TBIL (μmol/L)	–	–
Child-Pugh class (A/B)	–	–
AFP (ng/mL) (≤400/>400)	–	–
No. of lesions (solitary/multifocal)	0.099686063	–
Diameter (cm)	0.063100231	0.046008561
Vascular invasion (absent/present)	–	–
Tumor extent (unilobar/bilobar)	–	–
Enhancing capsule (absent/present)	–	–
Margin appearance (smooth/non-smooth)	0.177587833	–
AEF of residual tumor (%)	0.014771557	0.004969961

The "–" means zero coefficient of the variable at respective λ value. AEF-RT, arterial enhancement fraction of residual tumor; AFP, α-fetoprotein; ALT, alanine transaminase; AST, aspartate transaminase; HR, hazard ratio; LASSO, least absolute shrinkage and selection operator; NLR, neutrophil-to-lymphocyte ratio; PFS, progression-free survival; PLR, platelet –to–lymphocyte ratio; PLT, peripheric platelet count; TBIL, total bilirubin.

The calibration of the ADMAN-based risk categorization was visualized in Figures 4C and D. Good agreement was observed between the KM estimate and the ADMAN-predicted survival curves in the training cohort (Figure 4C) and was maintained in the validation cohort (Figure 4D).

3.4 Comparison of PFS between models

Given the lack of an established strategy for predicting PFS in HCC treated with TACE, we compared the performance of the ADMAN model with other widely recognized prognostic scoring systems designed for overall survival prediction, including the hepatoma arterial-embolization prognostic (HAP) score (18), modified HAP score (3), mHAP-II score (19), Six-and-Twelve score (20), and Up-to-11 criteria (21). The scores assignment and classification strategies are detailed in Supplementary Table S1. To avoid the training bias, the comparison was also performed in the validation set.

In the training cohort, the ADMAN model (0.75) demonstrated a significantly higher C-index than the other five prognostic score systems ($p < 0.05$ for all) (Table 3). In the validation cohort, the ADMAN model maintained sufficient discriminatory performance, with a significantly higher C-index (0.71) compared to the HAP Score (0.55, $p = 0.041$), mHAP Score (0.54, $p = 0.033$), and the Up-to-11 criteria (0.53, $p = 0.004$) while demonstrating a marginally higher C-index than the mHAP-II Score (0.61, $p = 0.080$) and Six-and-Twelve Score (0.64, $p = 0.231$) (Table 3). Additionally, the plots of time-dependent C-index were created to visualize the differences in the performance of discrimination between models over time. The ADMAN model demonstrated the consistently highest C-index during follow-up either in the training (Figure 5A) or validation cohort (Figure 5B).

The calibration of each model was assessed by Brier score of 6 months and 12 months PFS. In the training cohort, the Brier score of the ADMAN model was significantly lower than those of the other existing models, indicating better calibration (Table 4). In the validation cohort, the ADMAN model remained to have significantly lower 6-month Brier score than HAP Score, mHAP Score, mHAP-II Score, and

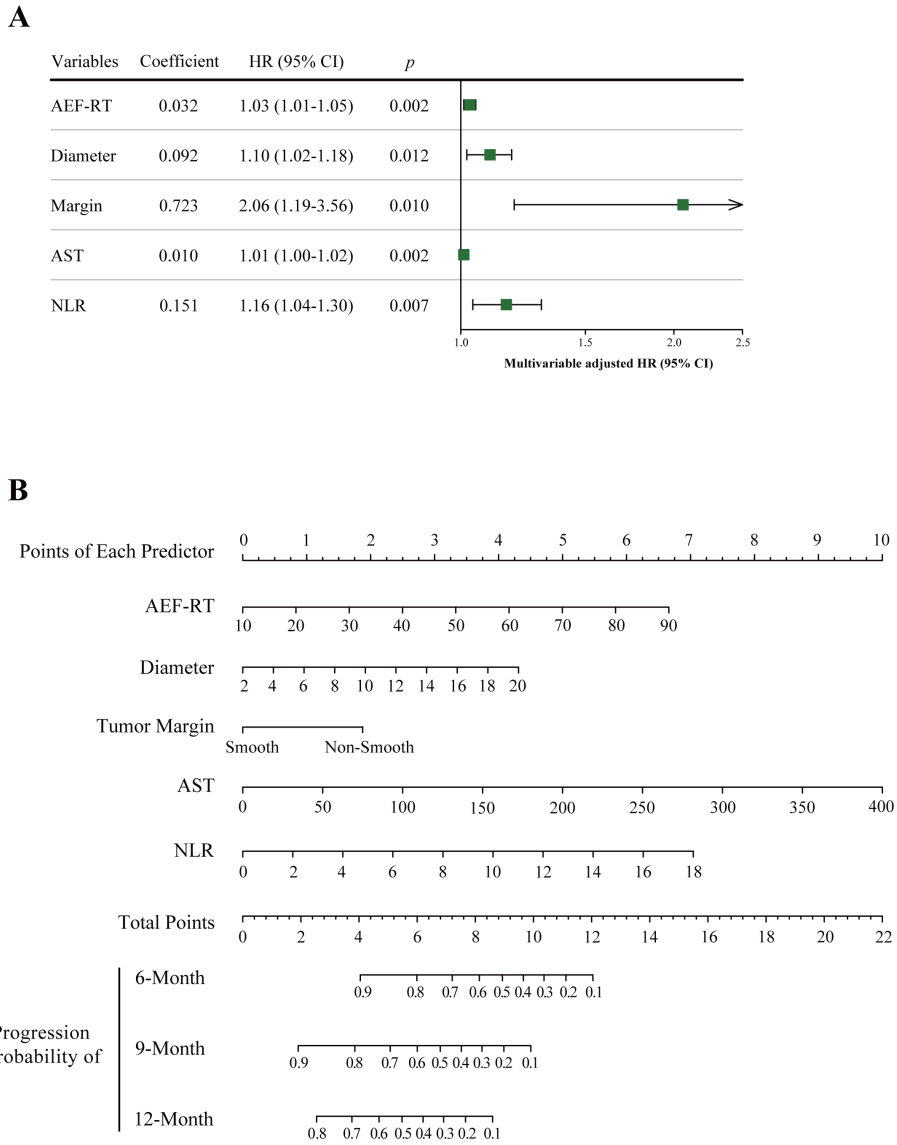


FIGURE 3
Visualization of the ADMAN model parameters **(A)** and ADMAN model-based nomogram **(B)**. Instruction for ADMAN model-based nomogram: Locate an individual patient's value on each independent variable axis, and then draw a line upward to obtain the points for each variable. Next, locate the sum of these points on the total points axis, and draw a line downward to the progression axis to obtain the probability of 6-, 9-, and 12-month PFS. AEF-RT = arterial enhancement fraction of residual tumor, AST, aspartate transaminase; DEB-TACE, drug-eluting beads transarterial chemoembolization; HCC, hepatocellular carcinoma; NLR, neutrophil-to-lymphocyte ratio; PFS, progression-free survival.

Up-to-11 criteria, while it had significantly lower Brier score than mHAP Score and Up-to-11 criteria at 12 months (Table 4). The integrated Brier score (IBS) extends the Brier score over time, evaluating the overall performance of probability forecasts across multiple time points. It is calculated by integrating the Brier score over time, offering a comprehensive view of a model's performance in time-series forecasting. In the training cohort, the ADMAN model (0.086) had a lower IBS than HAP Score (0.100), mHAP Score (0.099), mHAP-II Score (0.098), Six-and-Twelve Score (0.103), and Up-to-11 criteria (0.104). In the validation cohort, the IBS of the ADMAN model showed moderate calibration (0.170), while the IBS of other models indicated poor calibration [HAP Score (0.285), mHAP Score (0.291), mHAP-II Score (0.261), Six-and-Twelve Score (0.246), Up-to-11 criteria (0.358)].

4 Discussion

Existing scoring systems for TACE commonly incorporate pre-treatment characteristics, serum tumor markers, liver function tests, and tumor burden as prognostic indicators (1, 3, 18–21). However, despite these systems, TACE remains the preferred palliative treatment for patients ineligible for surgery due to the absence of a widely accepted pre-treatment scoring system, as noted in various

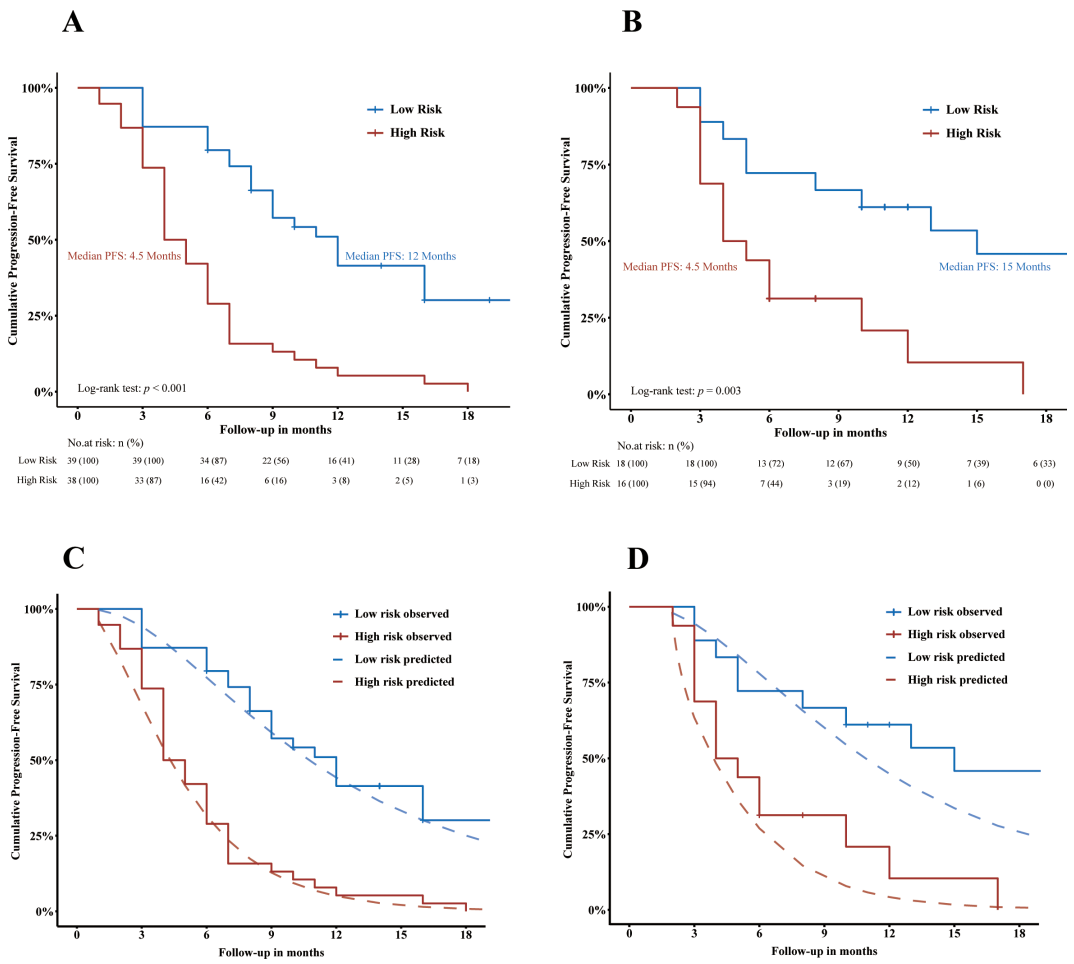


FIGURE 4
Visualization of the discrimination **(A, B)** and calibration **(C, D)** of ADMAN model. Kaplan–Meier curves showing the PFS of HCC patients stratified by progression risk in training cohort **(A)** and in validation cohort **(B)**. Plots depicting Kaplan–Meier estimate survival curves (jagged line) against ADMAN model-predicted mean survival curves (smooth dash line) in the training cohort **(C)** and validation cohort **(D)**. PFS, progression-free survival.

international guidelines (1, 22, 23). Unfortunately, 70%–80% of patients would eventually die due to tumor progression following repeat TACE procedures (2). This highlights the importance of transitioning from TACE to systemic therapies before liver function is compromised by repeated and ineffective treatments. Our purpose in developing this post-treatment scoring system is grounded in clinical practice, aiming to predict the suitability of TACE for patients based on certain post-treatment disease characteristics, such as the blood flow perfusion status of residual lesions.

TABLE 3 C-index of different models for PFS in the training and validation cohorts.

Models	Training cohort	$p^{\dagger}$	Validation cohort	$p^{\dagger}$
ADMAN Model	0.75 (0.69–0.82)	–	0.71 (0.60–0.83)	–
HAP Score	0.62 (0.55–0.71)	**0.004**	0.55 (0.47–0.70)	**0.041**
mHAP Score	0.65 (0.58–0.73)	**0.014**	0.54 (0.40–0.69)	**0.033**
mHAP-II Score	0.63 (0.55–0.70)	**0.002**	0.61 (0.48–0.74)	0.080
Six-and-Twelve Score	0.63 (0.56–0.70)	**<0.001**	0.64 (0.53–0.75)	0.231
Up-to-11 criteria	0.60 (0.53–0.67)	**<0.001**	0.53 (0.44–0.65)	**0.004**

HAP, hepatoma arterial-embolization prognostic score; mHAP, modified hepatoma arterial-embolization prognostic score; PFS, progression-free survival.
$^{\dagger}p$-value for comparing the C-index of ADMAN model and other prognostic model.
Bold values represent p-value < 0.05.

In the present study, we integrated a perfusion-like parameter of residual tumor with other clinical and radiological predictors to preliminarily develop and validate a prognostic model for PFS in patients undergoing DEB-TACE. The ADMAN model allows for the categorization of patients into distinct prognostic risk groups, enabling tailored follow-up schedules or alternative treatments for high-risk individuals. A comparison of five existing prognostic scores demonstrated that the ADMAN model exhibited promising performance in predicting PFS, which was further confirmed in the validation cohort. However, slightly declined performance of the ADMAN model in the validation cohort warrants further discussion. In the validation cohort, C-index of the ADMAN model consistently outperformed that of the HAP Score (p = 0.041), mHAP Score

TABLE 4 Brier score of different models for PFS at 6 months and 12 months.

Time point / Models	Training cohort	$p^{\dagger}$	Validation cohort	$p^{\dagger}$
6-Month				
ADMAN	0.161	–	0.183	
HAP Score	0.219	**0.009**	0.247	**0.041**
mHAP Score	0.210	**0.025**	0.257	**0.012**
mHAP-II Score	0.220	**0.004**	0.232	**0.049**
Six-and-Twelve Score	0.225	**<0.001**	0.214	0.238
Up-to-11 criteria	0.233	**<0.001**	0.244	**0.047**
12-Month				
ADMAN	0.140	–	0.191	–
HAP Score	0.203	**0.007**	0.248	0.054
mHAP Score	0.202	**0.014**	0.240	**0.049**
mHAP-II Score	0.209	**0.006**	0.234	0.158
Six-and-Twelve Score	0.211	**0.001**	0.196	0.888
Up-to-11 criteria	0.215	**<0.001**	0.243	**0.041**

HAP, hepatoma arterial-embolization prognostic score; mHAP, modified hepatoma arterial-embolization prognostic score; PFS, progression-free survival.
$^{\dagger}p$-value for comparing the brier score of ADMAN model and other prognostic model.
Bold values represent p-value < 0.05.

(p= 0.033), and Up-to-11 criteria (p = 0.004). These statistically significant advantages may be attributable to differences in the modeling cohorts, as each of these three models was developed in Western populations (3, 18, 21), where the etiological factors for HCC, such as hepatitis C virus, alcohol, obesity, and metabolic syndrome, differ notably from those in Asian populations, where hepatitis B virus predominates. In contrast, the mHAP-II score, developed in a Korean population, aligns more closely with the etiological and demographic characteristics of Chinese patients (19). As a result, the ADMAN model's C-index was only slightly higher than the mHAP-II score. The Six-and-Twelve score, developed in an entirely Chinese cohort, demonstrated the most comparable performance to the ADMAN model (20). Regarding the statistically insignificant Brier score of the ADMAN model, we hypothesize that this result may stem not only from demographic factors but also from the increasing proportion of censored observations over time, which impacts the dispersion of the empirical Brier score. In our cohort, the median PFS was 7 months [95% CI (6, 10)], leading to higher Brier score dispersion at time points with concentrated censoring. This may explain why the ADMAN model's Brier score showed no significant difference from that of other models in the validation cohort. To address this limitation, we included the IBS, which evaluates prediction inaccuracy over an interval rather than at a single time point by integrating loss functions. The IBS provides a more comprehensive measure of model performance over time and further supports the robustness of the ADMAN model.

Furthermore, the applicability of the ADMAN model being limited to patients with incomplete embolization may be considered a limitation. However, from our perspective, patients achieving complete response after initial TACE tend to maintain a relatively stable condition, whereas those with incomplete embolization are more likely to experience disease progression in the short term, despite repeated TACE attempts. Previous studies have also identified complete response as a significant protective factor for survival (24).The underlying mechanism may be attributed to intratumoral hypoxia induced by incomplete embolization. Hypoxia within the tumor microenvironment has been shown to promote epithelial–mesenchymal transition in cancer cells, thereby enhancing tumor aggressiveness and resistance to treatment (25, 26). Moreover, hypoxia triggers neo-angiogenesis and vascularization, processes that can be reflected by the AEF of tumor tissue (9).

Unlike chemoembolization, radioembolization delivers radioactive microspheres directly to tumors via super-selective catheterization, providing localized radiation to kill tumor cells without significant embolic effects. Although this technique has minimal impact on arterial blood flow, its potential influence on tumor angiogenesis or perfusion remains uncertain. A recent study suggests that hypoperfused primary liver tumors treated with Y-90 may have worse clinical outcomes compared to hyperperfused tumors (27). In that study, tumor perfusion was visually assessed by researchers using preoperative CT or MRI. Hypoperfused lesions were defined as those with less enhancement than the surrounding liver parenchyma, while hyperperfused lesions exhibited similar or greater enhancement. This raises an intriguing question: could AEF assessment serve as a complementary method to visual evaluation?

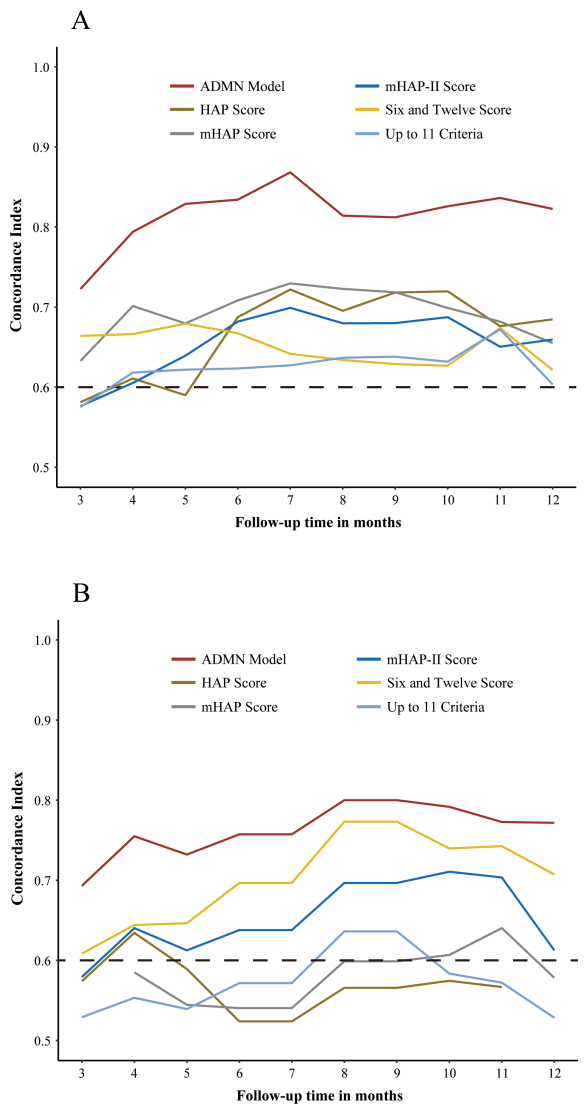


FIGURE 5
Plots of time-dependent C-index showing the C-index of the ADMAN model and five other prognostic models for PFS over time in the training **(A)** and validation cohorts **(B)**.

Given that AEF can be easily integrated into routine liver CT scans with less than 2 min of additional postprocessing time, future TARE studies could explore its feasibility for preoperative perfusion evaluation and for assessing changes in postoperative blood flow.

Although not directly applied in the current study, artificial intelligence (AI)-based techniques hold significant potential for overcoming limitations in AEF-based lesion analysis. AI-driven methods can provide a more comprehensive assessment of lesion perfusion dynamics, addressing challenges such as tumor heterogeneity, post-embolization alterations in blood supply, and high-perfusion artifacts caused by inflammation in necrotic region. For instance, stochastic resonance, as demonstrated by Dakua et al. in aneurysm segmentation, enhances contrast in low-signal environments, improving lesion-background differentiation and offering potential applications in AEF analysis (28, 29). Similarly, multi-modality registration methods, such as diffeomorphic mapping, ensure consistent segmentation across imaging protocols. Level-set methods integrated with denoising techniques, including maximal overlap discrete wavelet transforms, further refine lesion contours in noisy datasets (30, 31). Advanced preprocessing and regularization techniques further improve model robustness. For example, Dense-PSP-UNet employs contrast limited adaptive histogram equalization (CLAHE) to enhance boundary visualization and reduce noise in low-contrast images, preserving critical features and improving segmentation accuracy (32). Res-PAC-UNet incorporates Pyramid Atrous Convolution modules to capture multi-scale features, enriching contextual information and resolving ambiguous boundaries (33). Additionally, synthetic oversampling methods like SMOTE address class imbalance by generating synthetic samples for underrepresented categories, thereby improving model generalization and representation (34). Collectively, these techniques—contrast enhancement, multi-scale feature extraction, and data balancing—help mitigate overfitting and improve the generalizability of AEF-based models, establishing a foundation for robust, accurate predictive tools in complex imaging environments.

Apart from the inherent limitations of its retrospective design, our study has several other constraints. First, the relatively small sample size, especially in the validation cohort, limits the generalizability of our findings to broader patient populations. External validation with larger, independent cohorts is necessary to confirm the robustness of the ADMAN model in different clinical settings. Second, this study excluded patients who underwent chemoembolization with Lipiodol due to high-density artifacts caused by iodized oil deposits. These artifacts significantly hinder the identification of residual tumors and distort the AEF color map, making it unsuitable for evaluating postoperative imaging in these patients. This exclusion restricts the applicability of our findings to this subgroup of patients. Third, the model was developed and validated within an Asian population predominantly affected by hepatitis B virus, which limits its applicability to populations with different etiological profiles, such as those characterized by hepatitis C virus, alcohol-related liver disease, or metabolic syndrome. Variations in underlying etiologies may influence the performance and generalizability of the ADMAN model. Lastly, technical limitations in AEF measurement remain a challenge. Post-embolization changes, such as necrotic inflammation and altered blood perfusion, can introduce variability in measurements. Additionally, manual delineation of residual tumors on the AEF color map may lead to interobserver variability. These issues might be mitigated by integrating AI-based lesion segmentation techniques, which could standardize measurements and enhance the consistency and reliability of the results.

In conclusion, this pilot study underscores the potential utility of AEF-RT in predicting progression after DEB-TACE. The ADMAN model enabled the progression risk stratification and individualized estimation of PFS in patients with HCC undergoing DEB-TACE. Further external validation in independent cohorts with larger sample sizes is necessary to confirm the robustness of the ADMAN model.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics committee of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

BC: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. DX: Conceptualization, Data curation, Investigation, Software, Validation, Writing – review & editing. GZ: Project administration, Resources, Validation, Writing – review & editing. CZ: Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by the National Key Research and Development Program of China (2023YFC2413500) and the National Natural Science Foundation of China (U22A20352). The funding body had no role on the design, data collection, analysis, and manuscript writing of this study.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1489450/full#supplementary-material

References

1. Reig M, Forner A, Rimola J, Ferrer-Fabrega J, Burrel M, Garcia-Criado A, et al. Bclc strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol*. (2022) 76:681–93. doi: 10.1016/j.jhep.2021.11.018

2. Shin SW. The current practice of transarterial chemoembolization for the treatment of hepatocellular carcinoma. *Korean J Radiol*. (2009) 10:425–34. doi: 10.3348/kjr.2009.10.5.425

3. Pinato DJ, Arizumi T, Allara E, Jang JW, Smirne C, Kim YW, et al. Validation of the hepatoma arterial embolization prognostic score in european and asian populations and proposed modification. *Clin Gastroenterol Hepatol*. (2015) 13:1204–8 e2. doi: 10.1016/j.cgh.2014.11.037

4. Park C, Kim JH, Kim PH, Kim SY, Gwon DI, Chu HH, et al. Imaging predictors of survival in patients with single small hepatocellular carcinoma treated with transarterial chemoembolization. *Korean J Radiol*. (2021) 22:213–24. doi: 10.3348/kjr.2020.0325

5. Lee S, Kim KA, Park MS, Choi SY. Mri findings and prediction of time to progression of patients with hepatocellular carcinoma treated with drug-eluting bead transcatheter arterial chemoembolization. *J Korean Med Sci*. (2015) 30:965–73. doi: 10.3346/jkms.2015.30.7.965

6. Cho E, Kim HC, Lee JH, Jeong-Ju Y, Choi WM, Cho YY, et al. Serum insulin-like growth factor-1 predicts disease progression and survival in patients with hepatocellular carcinoma who undergo transarterial chemoembolization. *PloS One*. (2014) 9:e90862. doi: 10.1371/journal.pone.0090862

7. Wang B, Li F, Cheng L, Zhan Y, Wu B, Chen P, et al. The pretreatment platelet count is an independent predictor of tumor progression in patients undergoing transcatheter arterial chemoembolization with hepatitis B virus-related hepatocellular carcinoma. *Future Oncol*. (2019) 15:827–39. doi: 10.2217/fon-2018-0591

8. Kim SH, Kamaya A, Willmann JK. Ct perfusion of the liver: principles and applications in oncology. *Radiology*. (2014) 272:322–44. doi: 10.1148/radiol.14130091

9. Kim KW, Lee JM, Kim JH, Klotz E, Kim HC, Han JK, et al. Ct color mapping of the arterial enhancement fraction of vx2 carcinoma implanted in rabbit liver: comparison with perfusion ct. *AJR Am J Roentgenol*. (2011) 196:102–8. doi: 10.2214/AJR.09.3971

10. Kim KW, Lee JM, Klotz E, Park HS, Lee DH, Kim JY, et al. Quantitative ct color mapping of the arterial enhancement fraction of the liver to detect hepatocellular carcinoma. *Radiology*. (2009) 250:425–34. doi: 10.1148/radiol.2501072196

11. Chai B, Xiang D, Wang W, Ren Y, Wang F, Wang J, et al. Arterial enhancement fraction in evaluating the therapeutic effect and survival for hepatocellular carcinoma patients treated with deb-tace. *Cancer Imaging*. (2022) 22:38. doi: 10.1186/s40644-022-00477-z

12. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (Tripod): the tripod statement. *BMJ*. (2015) 350:g7594. doi: 10.1136/bmj.g7594

13. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the L. Easl Clinical Practice Guidelines: Management of Hepatocellular Carcinoma. *J Hepatol*. (2018) 69:182–236. doi: 10.1016/j.jhep.2018.03.019

14. Llovet JM, Lencioni R. Mrecist for hcc: performance and novel refinements. *J Hepatol*. (2020) 72:288–306. doi: 10.1016/j.jhep.2019.09.026

15. Royston P, Altman DG. External validation of a cox prognostic model: principles and methods. *BMC Med Res Methodol*. (2013) 13:33. doi: 10.1186/1471-2288-13-33

16. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. (1999) 18:2529–45. doi: 10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5

17. Harrell FE. Statistically Efficient Ways to Quantify Added Predictive Value of New Measurements . Available online at: https://hbiostat.org/blog/post/addvalue/ (Accessed July 16, 2024).

18. Kadalayil L, Benini R, Pallan L, O'Beirne J, Marelli L, Yu D, et al. A simple prognostic scoring system for patients receiving transarterial embolisation for hepatocellular cancer. *Ann Oncol*. (2013) 24:2565–70. doi: 10.1093/annonc/mdt247

19. Park Y, Kim SU, Kim BK, Park JY, Kim DY, Ahn SH, et al. Addition of tumor multiplicity improves the prognostic performance of the hepatoma arterial-embolization prognostic score. *Liver Int*. (2016) 36:100–7. doi: 10.1111/liv.12878

20. Wang Q, Xia D, Bai W, Wang E, Sun J, Huang M, et al. Development of a prognostic score for recommended tace candidates with hepatocellular carcinoma: A multicentre observational study. *J Hepatol*. (2019) 70:893–903. doi: 10.1016/j.jhep.2019.01.013

21. Kim JH, Shim JH, Lee HC, Sung KB, Ko HK, Ko GY, et al. New intermediate-stage subclassification for patients with hepatocellular carcinoma treated with transarterial chemoembolization. *Liver Int*. (2017) 37:1861–8. doi: 10.1111/liv.13487

22. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the american association for the study of liver diseases. *Hepatology*. (2018) 68:723–50. doi: 10.1002/hep.29913

23. Omata M, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, et al. Asia-pacific clinical practice guidelines on the management of hepatocellular carcinoma: A 2017 update. *Hepatol Int*. (2017) 11:317–70. doi: 10.1007/s12072-017-9799-9

24. Prajapati HJ, Spivey JR, Hanish SI, El-Rayes BF, Kauh JS, Chen Z, et al. Mrecist and easl responses at early time point by contrast-enhanced dynamic mri predict survival in patients with unresectable hepatocellular carcinoma (Hcc) treated by doxorubicin drug-eluting beads transarterial chemoembolization (Deb tace). *Ann Oncol*. (2013) 24:965–73. doi: 10.1093/annonc/mds605

25. Liu Z, Wang Y, Dou C, Xu M, Sun L, Wang L, et al. Hypoxia-induced up-regulation of vasp promotes invasiveness and metastasis of hepatocellular carcinoma. *Theranostics*. (2018) 8:4649–63. doi: 10.7150/thno.26789

26. Li F, Mei H, Gao Y, Xie X, Nie H, Li T, et al. Co-delivery of oxygen and erlotinib by aptamer-modified liposomal complexes to reverse hypoxia-induced drug resistance in lung cancer. *Biomaterials*. (2017) 145:56–71. doi: 10.1016/j.biomaterials.2017.08.030

27. Kalaghchi B, Ince S, Barnes J, Kiser K, Chin RI, Mikell J, et al. Arterial hypoperfusion as a negative predictive marker for primary hepatic Malignancies treated with Y-90 glass microsphere transarterial radioembolization. *Front Oncol*. (2024) 14:1433480. doi: 10.3389/fonc.2024.1433480

28. Dakua SP, Abinahed J, Al-Ansari A. A pca-based approach for brain aneurysm segmentation. *Multidimensional Syst Signal Process*. (2018) 29:257–77. doi: 10.1007/s11045-016-0464-6

29. Regaya Y, Amira A, Dakua SP. Development of a cerebral aneurysm segmentation method to prevent sentinel hemorrhage. *Network Modeling Anal Health Inf Bioinf*. (2023) 12:18. doi: 10.1007/s13721-023-00412-7

30. Mohanty S, Dakua SP. Toward computing cross-modality symmetric non-rigid medical image registration. *IEEE Access*. (2022) 10:24528–39. doi: 10.1109/ACCESS.2022.3154771

31. Dakua SP, Abinahed J, Zakaria A, Balakrishnan S, Younes G, Navkar N, et al. Moving object tracking in clinical scenarios: application to cardiac surgery and cerebral aneurysm clipping. *Int J Comput Assisted Radiol Surg*. (2019) 14:2165–76. doi: 10.1007/s11548-019-02030-z

32. Ansari MY, Yang Y, Meher PK, Dakua SP. Dense-Psp-Unet: a neural network for fast inference liver ultrasound segmentation. *Comput Biol Med*. (2023). 153:106478. doi: 10.1016/j.compbiomed.2022.106478

33. Ansari MY, Yang Y, Balakrishnan S, Abinahed J, Al-Ansari A, Warfa M, et al. A Lightweight neural network with multiscale feature enhancement for liver ct segmentation. *Sci Rep*. (2022) 12:14153. doi: 10.1038/s41598-022-16828-6

34. Ansari MY, Chandrasekar V, Singh AV, Dakua SP. Re-routing drugs to blood brain barrier: A comprehensive analysis of machine learning approaches with fingerprint amalgamation and data balancing. *IEEE Access*. (2023) 11:9890–906. doi: 10.1109/ACCESS.2022.3233110

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 24 February 2025
DOI 10.3389/fonc.2025.1524634

Check for updates

OPEN ACCESS

EDITED BY
Frontiers in Oncology Editorial Office,
Frontiers Media SA, Switzerland

REVIEWED BY
Zheng Han,
University of Oklahoma Health Sciences Center, United States
Alexandros Vamvakas,
Amsterdam University Medical Center, Netherlands

*CORRESPONDENCE
Dibash Basukala
dibash.basukala@nyulangone.org

RECEIVED 07 November 2024
ACCEPTED 28 January 2025
PUBLISHED 24 February 2025

CITATION
Basukala D, Mikheev A, Li X, Goldberg JD, Gilani N, Moy L, Pinker K, Partridge SC, Biswas D, Kataoka M, Honda M, Iima M, Thakur SB and Sigmund EE (2025) Retrospective BReast Intravoxel Incoherent Motion Multisite (BRIMM) multisoftware study. *Front. Oncol.* 15:1524634.
doi: 10.3389/fonc.2025.1524634

COPYRIGHT
© 2025 Basukala, Mikheev, Li, Goldberg, Gilani, Moy, Pinker, Partridge, Biswas, Kataoka, Honda, Iima, Thakur and Sigmund. 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.

Retrospective BReast Intravoxel Incoherent Motion Multisite (BRIMM) multisoftware study

Dibash Basukala[1,2]*, Artem Mikheev[1], Xiaochun Li[3], Judith D. Goldberg[3], Nima Gilani[1], Linda Moy[1], Katja Pinker[4,5], Savannah C. Partridge[6,7], Debosmita Biswas[6,7], Masako Kataoka[8], Maya Honda[8,9], Mami Iima[8,10], Sunitha B. Thakur[2,5] and Eric E. Sigmund[1,11]

[1]Department of Radiology, Grossman School of Medicine, New York University, New York, NY, United States, [2]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, United States, [3]Department of Population Health, Grossman School of Medicine, New York University, New York, NY, United States, [4]Department of Radiology, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY, United States, [5]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, United States, [6]Department of Bioengineering, University of Washington, Seattle, CA, United States, [7]Department of Radiology, School of Medicine, University of Washington, Seattle, WA, United States, [8]Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, [9]Department of Diagnostic Radiology, Kansai Electric Power Hospital, Osaka, Japan, [10]Department of Fundamental Development for Advanced Low Invasive Diagnostic Imaging, Graduate School of Medicine, Nagoya University, Nagoya, Japan, [11]Center for Advanced Imaging Innovation and Research, New York University, New York, NY, United States

Introduction: The intravoxel incoherent motion (IVIM) model of diffusion weighted imaging (DWI) provides imaging biomarkers for breast tumor characterization. It has been extensively applied for both diagnostic and prognostic goals in breast cancer, with increasing evidence supporting its clinical relevance. However, variable performance exists in literature owing to the heterogeneity in datasets and quantification methods.

Methods: This work used retrospective anonymized breast MRI data (302 patients) from three sites employing three different software utilizing least-squares segmented algorithms and Bayesian fit to estimate 1_{st} order radiomics of IVIM parameters perfusion fraction (f_p), pseudo-diffusion (D_p) and tissue diffusivity (D_t). Pearson correlation (r) coefficients between software pairs were computed while logistic regression model was implemented to test malignancy detection and assess robustness of the IVIM metrics.

Results: D_t and f_p maps generated from different software showed consistency across platforms while D_p maps were variable. The average correlation between the three software pairs at three different sites for 1_{st} order radiomics of IVIM parameters were $D_t min/D_t max/D_t mean/D_t variance/D_t skew/D_t kurt$: 0.791/0.891/0.98/0.815/0.697/0.584; $f_p max/f_p mean/f_p variance/f_p skew/f_p kurt$: 0.615/0.871/0.679/0.541/0.433; $D_p max/D_p mean/D_p variance/D_p skew/D_p kurt$: 0.616/0.56/0.587/0.454/0.51. Correlation between least-squares algorithms were the highest. $D_t mean$ showed highest area under the ROC curve (AUC) with 0.85 and lowest coefficient of variation (CV) with 0.18% for benign and malignant differentiation using logistic regression. D_t metrics were highly diagnostic as well as consistent along with f_p metrics.

Discussion: Multiple 1_{st} order radiomic features of D_t and f_p obtained from a heterogeneous multi-site breast lesion dataset showed strong software robustness and/or diagnostic utility, supporting their potential consideration in controlled prospective clinical trials.

KEYWORDS

IVIM, DWI, breast cancer, diagnosis, multisite, multisoftware, radiomics, robust

1 Introduction

Breast cancer remains a leading cause of cancer-related deaths in women in the U.S (1). Diffusion weighted MRI (DW-MRI or DWI) provides biomarkers for cancer diagnosis and characterization (2–6), and has been demonstrated to distinguish benign and malignant breast lesions (7–10) without using contrast agents.

Intravoxel incoherent motion (IVIM) (11–13), an advanced DWI technique allows simultaneous quantification of diffusion and perfusion properties of the tissue. IVIM is sensitive to cellularity and microvascular flow and there is a growing evidence base of its clinical utility for both diagnostic and prognostic goals in the setting of breast cancer (14–18). IVIM uses a biexponential function (see Equation 1.1) to describe the diffusion signal decay over different b-values to estimate tissue diffusivity (D_t), pseudo-diffusion (D_p), and perfusion fraction (f_p). These IVIM coefficients serve as biomarkers for the identification of different tumor biologic characteristics. Specifically, D_t is sensitive to restrictions to Brownian water motion such as cell membranes, fibrosis, or macromolecules. f_p reflects the volume fraction of faster microcirculation, often originating from the microvascular space. Finally, D_p reflects the apparent diffusion process in the microcirculatory space which is impacted by both fluid flow speed and vascular architecture. For the specific case of breast cancer, malignant tumors often exhibit lower D_t values due to higher cellularity, higher f_p due to higher vascularity and lower D_p due to slower blood velocity compared to benign lesions.

One obstacle to clinical implementation is the variability of algorithms and tools used to determine the IVIM metrics, which can introduce corresponding variability in clinical performance. For example, most IVIM parameters' estimation is based on nonlinear least squares (19, 20), segmented least squares fitting (17, 21–23), or the Bayesian (24–27) approach. More recently, deep learning (DL) based approaches have gained significant attention for their mitigation of acquisition (28) and noise-induced variability compared to traditional methods, especially for f_p and D_p (29–32). Furthermore, most prior studies calculated IVIM coefficients based on the mean values within the region of interest (ROI), whereas radiomic features of IVIM maps may potentially provide more information and capture tumor heterogeneity (33–35).

Nevertheless, differences in patient cohorts, scanners, acquisition protocols, and analysis algorithms (36–39) contribute to variable diagnostic performance between studies and can dilute the potential of the IVIM biomarkers for more widespread adoption in clinical trials or daily practice (8, 10). A retrospective cross-sectional view of a large subset of available clinical data from patients presenting with suspicious lesions, acquired at different sites analyzed with widely used software platforms, may be illuminating to highlight the software dependency of IVIM parameters as well as the most robust and diagnostic 1^{st} order radiomic features in the IVIM dataset and guide future harmonization efforts in multi-center trials.

2 Materials and methods

This study evaluated retrospective anonymized breast MR imaging data from three different sites. The patients were scanned using 1.5 T or 3 T scanners at each site (Site A: GE Healthcare, Waukesha, Wisconsin, USA, 1.5 T and 3.0 T; Site B: Philips Healthcare, Best, the Netherlands, 3.0 T; and Site C: Siemens trio, Siemens Healthcare, Erlangen, Germany, 3.0 T). Details of each acquisition protocol and studied cohort are listed in Tables 1, 2. Criteria for evaluation varied between sites. Site A included the patients who underwent breast MRI screening when they were suspicious of Breast Imaging Reporting and Data System (BI-RADS) 4, 5 and/or cancer-proven BI-RADS 6 lesions. Site B included patients underwent breast MRI screening and had BI-RADS 4, 5 lesions detected. Site C included the patients who underwent breast MRI when they were known to have or were suspected of having breast carcinoma. This included patients with BI-RADS categories 2-5 lesions.

Lesion conspicuity was assessed by radiologists on either b0 or b>0 DWI images, in comparison with dynamic contrast enhanced (DCE) MRI at Site A, Site B and Site C. Referencing the accompanying DCE MRI, ROIs were drawn on either b0 or b>0 DWI images in consultation with the respective team radiologist. At Site A ROIs were prescribed on all lesion slices, while for Sites B and C only the central slice of largest cross section was prescribed. The ROI contains at least 3 voxels, and no obvious artifacts were included at all sites as per the guidelines from the European Society of Breast Radiology (EUSOBI) (8). Single lesion per patient was used at all sites. In addition, lesions were histologically confirmed (Site A and Site B), or sometimes based on radiologist reports, and based on stability on imaging for more than 18 months for benign lesions at Site C.

TABLE 1 MRI system and acquisition parameters used at each site in the multicenter study. Resolutions are given in acquired and reconstructed voxel sizes.

		Site A		Site B	Site C
MRI system	Vendor	GE 1.5 T/3 T		Philips 3 T	Siemens Trio 3 T
Resolution (mm)		2.2 - 2.8/2.2 - 2.8/4 Recon. 1.1 - 1.4/1.1 - 1.4/4		1.8/1.8/4 Recon. 1.3/1.3/4	2.0/2.0/3
Echo time (ms)		96.2		59.2	88.4
Repetition time (ms)		6000		4987.5	4700
b-values (s/mm^2)		9	0,30,60,90,120,250, 400,600,800	0,100,600,800,1000	0,5,10,20,30,50,70,100,200,400,600,800,1000
		24	0,30,60,90,120,400, 600,800,1000		
		10	0,30,60,90,120,250, 450,600,800,1000		
		14	0,30,60,90,120,250, 400,600,800,1000		
		1	0,10,30,60,90,120, 200,400,600,800, 1000		

2.1 Data analysis

IVIM data from all sites were independently analyzed using three software packages: a shareware tool with least-squares segmented fitting (Firevoxel, https://firevoxel.org/ (Software a)), an MR vendor research software package with least-squares segmented fitting (Siemens MR Body Diffusion Toolbox from Siemens Healthineers (Software b)) and a commercial software package with Bayesian fit algorithm Olea Sphere (Software c).

IVIM parameters were estimated from a fit of all acquired b-values (see Table 1) to a biexponential decay:

$$\frac{S}{S_0} = f_p \exp(-b \cdot D_p) + (1 - f_p) \exp(-b \cdot D_t) \quad 1.1$$

IVIM parameters f_p, D_p and D_t were calculated from the voxels in the lesion ROI using each software tool. Histogram analysis of parametric maps generated by each software was also performed within a separate module for histogram generation in Firevoxel (100 bins, f_p: 0 – 1, D_p: 0 – 0.1 mm^2/s and D_t: 0 – 0.003 mm^2/s) to estimate 1st order radiomic features from each parameter: mean/minimum/maximum/variance/skewness/kurtosis. This single histogram module was used to limit the software differences to that in IVIM estimation alone.

2.2 Statistical analysis

The Pearson correlation (r) coefficient of IVIM parameters for the 1st order radiomic features was computed between each software pair at each site separately. The average correlation coefficient and coefficient of variation (CV) over all software pairs and sites was computed for each metric and ranked in numerical order to assess the consistency of performance of a clinical task. The intraclass correlation coefficient (ICC) was also computed for the agreement among three software for the IVIM metrics at each site. Additionally, Bland-Altman analysis (40) of IVIM parameters for the 1st order radiomic features was also carried out between each software pair at each site separately. Measures of absolute difference mean, absolute difference standard deviation, and CV (%) were derived from each software pair comparison.

TABLE 2 Number of patients with breast lesions from multiple centers along with average age at each site.

		Site A			Site B			Site C		
		N	ROI Size		N	ROI Size		N	ROI Size	
			Voxels	cm^3		Voxels	cm^3		Voxels	cm^3
Lesions	**Benign**	12	201 ± 143	0.97 ± 0.69	70	51 ± 101	0.32 ± 0.63	38	33 ± 40	0.39 ± 0.47
	Malignant	46	1363 ± 1644	6.6 ± 7.96	19	34 ± 28	0.21 ± 0.18	117	56 ± 69	0.67 ± 0.83
	Total	58	1123 ± 1537	5.43 ± 7.44	89	47 ± 91	0.29 ± 0.57	155	50 ± 64	0.6 ± 0.77
Age (yrs)		48.26 ± 9.61			46.12 ± 11.34			57.03 ± 15.25		

Voxel count, size values (in cubic centimeters) and age are given in mean ± standard deviation.

Within the context of each software, each IVIM metric was tested for benign/malignant differentiation, separately for each software, using logistic regression for all three sites' data together, with each variable adjusted by site (coefficient and intercept). In addition, we also performed leave-one-patient-out (LOU) cross validation for each IVIM metric for the logistic regressions adjusted by site for each software. The area under the ROC curve (AUC) and standard error (SE) were quantified for each software separately. An average of AUCs (separately for original and LOU analysis) across software was computed for each IVIM metric. CVs of the three AUCs from each software were computed for benign and malignant differentiation. These average metrics were then ranked in numerical order for assessment of consistency of performance of a clinical task. Additionally, AUCs from all pairs of software were separately compared with DeLong's test. Statistical analysis was performed using MATLAB software for Bland-Altman analysis and R 4.2 software for ICC and logistic regression.

3 Results

The study included 58, 89 and 155 patients from Site A, Site B, and Site C respectively. Site A, Site B and Site C included 79.3%, 21.4%, and 75.5% of patients with malignant lesions respectively, with each patient contributing one lesion. Table 2 shows the distribution of the patients including the number of biopsy-confirmed benign/malignant lesions across sites in this retrospective multicenter study along with ROI size. The number of voxels per ROI ranged from 47 ± 91 (Site B) to 50 ± 64 (Site C), and up to 1123 ± 1537 (Site A). In addition, average age across sites is also reported.

Example IVIM parameter maps obtained from each software for malignant lesions from Site A, Site B and Site C are shown in Figure 1, Figures 2, 3 respectively. Example benign breast lesions are shown in Supplementary Figure S1-S3. Overall D_t maps and f_p maps show consistency while D_p maps exhibit the most variability across

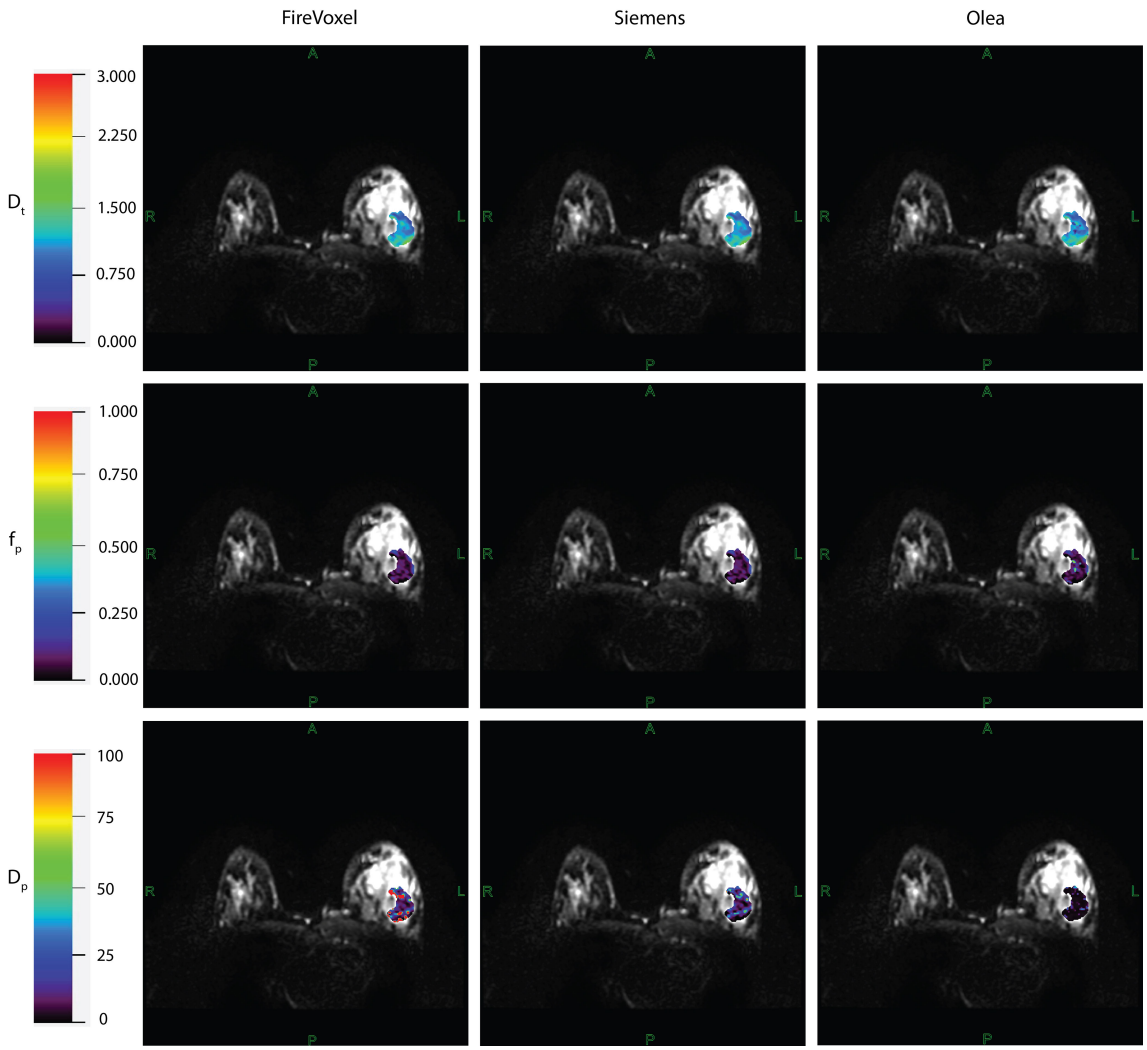


FIGURE 1
IVIM parametric maps overlaid on raw DWI images in a patient with malignant breast lesion for Site A. IVIM parameters tissue diffusivity (D_t), perfusion fraction (f_p) and pseudodiffusivity (D_p) obtained from Firevoxel, Siemens and Olea software in the breast lesion. D_t maps and f_p maps are the most consistent across software platforms, while D_p maps show the most variability with fit algorithms. D_t and D_p are given in units of 10^{-3} mm^2/s.

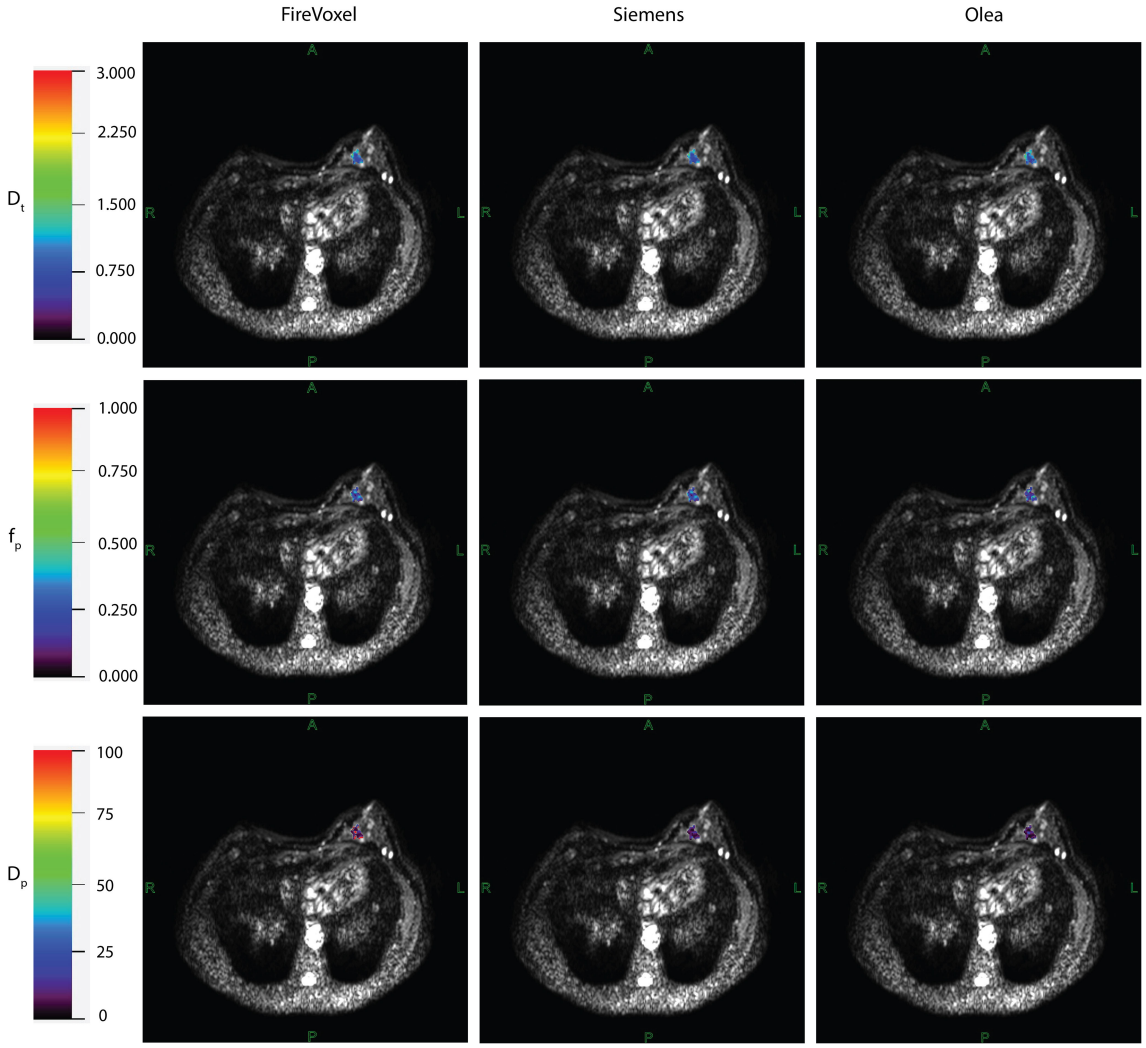


FIGURE 2
IVIM parametric maps overlaid on raw DWI images in a patient with malignant breast lesion for Site B. IVIM parameters tissue diffusivity (D_t), perfusion fraction (f_p) and pseudodiffusivity (D_p) obtained from Firevoxel, Siemens and Olea software in the breast lesion. D_t maps and f_p maps are the most consistent across software platforms, while D_p maps show the most variability with fit algorithms. D_t and D_p are given in units of 10^{-3} mm^2/s.

the software platforms. The average fractions of utilized voxels per lesion (i.e. having values within the prescribed histogram ranges) were as follows. D_t utilized 99.72%/99.88%/99.87% of lesion voxels at Site A; 100%/100%/100% at Site B; 99.99%/99.95%/100% at Site C using Software a/b/c. D_p utilized 77.51%/99.77%/100% of lesion voxels at Site A; 80.58%/100%/100% at Site B; 59.78%/99.23%/100% at Site C using Software a/b/c while f_p utilized 100% of lesion voxels at all sites using Software a, b, c. The mean IVIM parameter values for benign and malignant lesions at Site A, Site B and Site C are shown in Table 3 which clearly indicates the consistency of D_t and f_p values across all software platforms except for f_p at Site B. Mean f_p values were found to be somewhat variable between least squares segmented fitting and Bayesian fitting at Site B.

The correlation coefficient of IVIM parameters between each software pair for 1st order radiomic features at each site is shown in Supplementary Table S1 along with ICC values. Correlations between least-squares segmented fitting algorithms are generally higher than those between least squares and Bayesian algorithms. The average correlation between the three software at three different sites for 1st order radiomic features mean/maximum/variance/skewness/kurtosis were f_p (r = 0.871/0.615/0.679/0.541/0.433), D_p (r = 0.56/0.616/0.587/0.454/0.51) and D_t (r = 0.98/0.891/0.815/0.697/0.584) respectively while that for $D_t min$ was 0.791. The correlations between the three software for mean D_t at Site A, Site B and Site C are shown in Figure 4; excellent correlation observed between least-squares segmented algorithms (Firevoxel and Siemens) and Bayesian algorithms (Olea) at each site. Similarly, the correlations between the three software for mean f_p at Site A, Site B and Site C are shown in Figure 5; strongest correlation observed between least-squares segmented algorithms (Firevoxel and Siemens) at each site. Figure 6 shows the average of correlation coefficients of 1st order radiomic features of f_p, D_t and D_p across all software and sites along with CV of correlation coefficients. In general, D_t radiomics showed the highest average software correlation along with mean f_p while $D_t mean$ showed the lowest CV. Additionally, Bland-Altman analysis of IVIM parameters between each Software (a, b, c) pair for 1st order

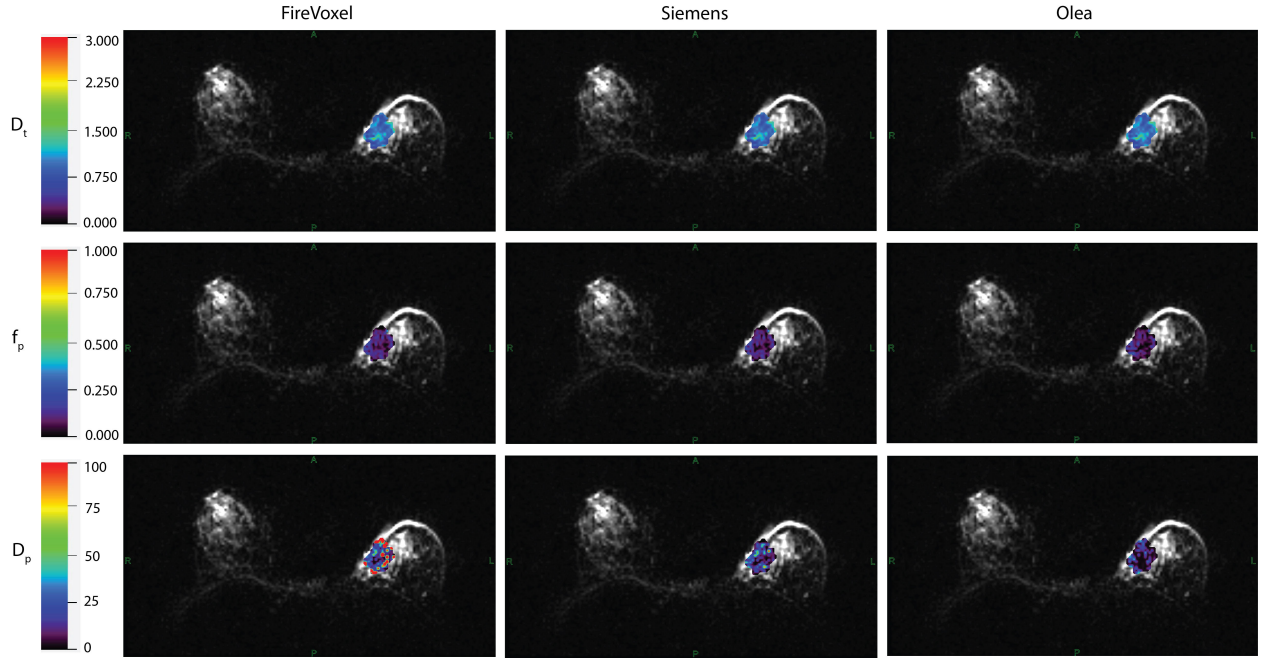


FIGURE 3
IVIM parametric maps overlaid on raw DWI images in a patient with malignant breast lesion for Site C. IVIM parameters tissue diffusivity (D_t), perfusion fraction (f_p) and pseudodiffusivity (D_p) obtained from Firevoxel, Siemens and Olea software in the breast lesion. D_t maps and f_p maps are the most consistent across software platforms, while D_p maps show the most variability with fit algorithms. D_t and D_p are given in units of 10^{-3} mm^2/s.

radiomics at each site is shown in Supplementary Table S2. Bland-Altman plots between the three software for mean D_t and mean f_p at Site A, Site B and Site C are shown in Supplementary Figures S4, S5.

No pair of parameter AUCs from different software were significantly different ($p>0.05$). Regarding the pooled site analyses, the AUC with SE for benign and malignant differentiation for different IVIM metrics employing different software from three sites using logistic regression and LOU cross validation is shown in Table 4 while the average of AUC as well as CV of AUC (%) for benign and malignant differentiation is shown in Figure 7. For both AUC analyses, mean, minimum, maximum and skewness of D_t showed the highest average AUC followed by D_p metrics for the benign/malignant task while mean and variance of f_p along with several D_t radiomics showed high consistency among software. LOU AUCs showed a similar ranking of performance to logistic regression AUC with a few exceptions (such as higher ranking of f_p mean), with slightly lower and more spread values of average AUC, and higher and more spread values of CV of AUC.

4 Discussion

Our study evaluated variability across software tools for IVIM measurements of breast tumors in a heterogeneous multicenter multivendor dataset to test the robustness and diagnostic utility of IVIM biomarkers in a worst-case scenario paradigm. Broadly speaking, D_t metrics present markers of tissue microstructure (especially tumor cellularity) and f_p metrics report on microvascularity. Both of these

TABLE 3 Mean IVIM parameter values for benign and malignant lesion employing Software (a, b, c) at Site A, Site B and Site C.

		Site A			Site B			Site C		
		a	b	c	a	b	c	a	b	c
D_t	Benign	1.26 ± 0.29	1.26 ± 0.27	1.25 ± 0.24	1.04 ± 0.33	1.05 ± 0.31	1.12 ± 0.31	1.36 ± 0.4	1.31 ± 0.4	1.31 ± 0.35
	Malignant	1.2 ± 0.39	1.19 ± 0.38	1.18 ± 0.38	0.88 ± 0.29	0.88 ± 0.28	0.96 ± 0.31	0.93 ± 0.29	0.9 ± 0.28	0.9 ± 0.29
f_p	Benign	15.79 ± 9.87	14.42 ± 8.87	16.34 ± 8.36	18.62 ± 7.49	16.47 ± 7.41	11.06 ± 7.46	12.77 ± 6.95	11.78 ± 7	11.08 ± 6.77
	Malignant	13.85 ± 5.02	13.03 ± 4.51	14.34 ± 4.85	20.66 ± 7.15	19.32 ± 7.85	14.03 ± 6.79	11.41 ± 4.91	10.21 ± 3.9	9.4 ± 4.61
D_p	Benign	10.57 ± 3.83	9 ± 3.54	5.46 ± 2.63	6.07 ± 2.34	5.25 ± 2.99	6.03 ± 3.51	18.02 ± 8.47	10.87 ± 4.81	6.77 ± 4.12
	Malignant	12.02 ± 3.12	9.85 ± 3.17	7.85 ± 2.91	7.06 ± 2.02	7.27 ± 3.24	8.62 ± 4.05	16.23 ± 5.94	9.84 ± 4.03	11.43 ± 4.64

a: Firevoxel; b: Siemens; c: Olea.
Perfusion fraction (f_p) is given in %, while pseudo-diffusion (D_p) and tissue diffusivity (D_t) are given in units of 10^{-3} mm^2/s.
Data are given in mean ± standard deviation.

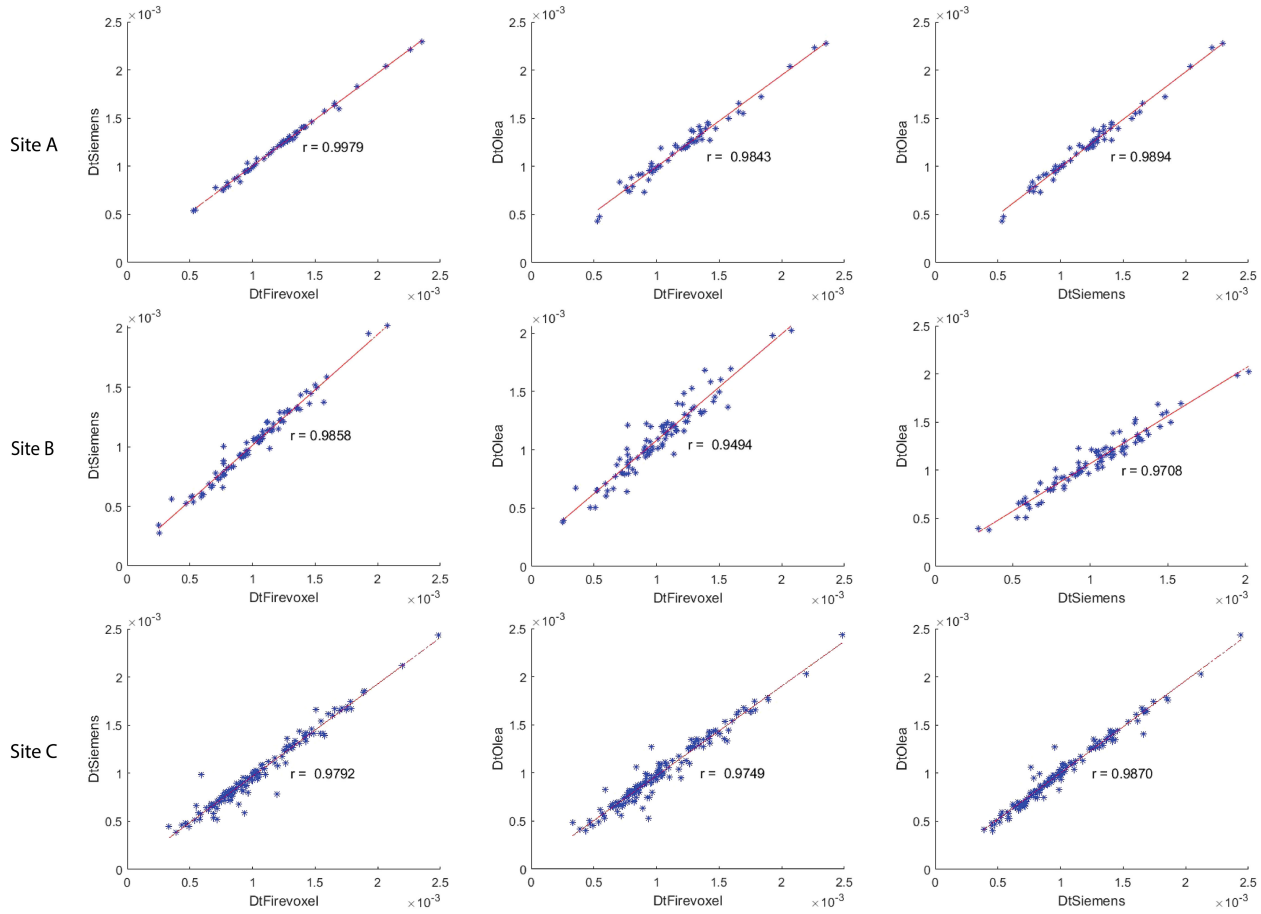


FIGURE 4
Correlation coefficient between Firevoxel, Siemens and Olea for mean of tissue diffusivity (D_t) at Site A, Site B and Site C. Comparisons shown left to right: Firevoxel *vs*. Siemens, Firevoxel *vs*. Olea, and Siemens *vs*. Olea. Least-squares segmented algorithms (Firevoxel, Siemens) and Bayesian algorithms (Olea) show excellent agreement. D_t is given in unit of mm^2/s.

features are known to be biologically important in determining malignancy and monitoring or predicting response to differently targeted treatment (such as cytotoxic or anti-angiogenic agents). In order to maximize the potential of their separate biologic sensitivities, their numerical robustness must be scrutinized as in the present study.

IVIM parametric maps obtained from different software employing least-squares segmented fitting and Bayesian fitting generated similar D_t and f_p maps. D_t maps were the most consistent across the software platforms at all sites while some differences in f_p maps could be observed particularly at Site B between Software a/b and c. The lower correlations at Site B between f_p values obtained from Bayesian and least-squares packages may have been affected by that site's low number of b-values sampled in the pseudodiffusion regime ($b<200$ s/mm^2); with fewer data constraints Bayesian approaches may regress to their prior. D_p maps were the most variable between the software platforms.

Several D_t radiomic features as well as mean f_p demonstrated high correlations between software pairs. Software correlations were highest between the least squares segmented algorithms (a/b) and mean values are the most consistent across contexts. Multiple D_t radiomic features were highly diagnostic for benign and malignant differentiation as well as consistent across software platforms. However, for f_p metrics, mean and variance, moderately diagnostic on average, were highly consistent among software.

Results of this study indicate some variability in software robustness and benign/malignant differentiation among multi-site data. Some site variability (lesion size, b-value distribution, cohort size, selection criteria) may limit consistency; therefore, a logistic regression model with site adjustment factors was employed to obtain AUCs to account for such heterogeneity in the dataset. LOU AUCs was also derived as a more stringent test of the data, which revealed slight reduction in performance but analogous ranking of parameters. Several D_t metrics showed both software robustness and consistently high diagnostic performance. The robust performance of D_t metrics across different software platforms and sites, particularly for benign/malignant differentiation, supports the potential for widespread implementation of IVIM-DWI beyond its current limited clinical use and research applications (5). On the other hand, several D_p metrics although showing consistency across software platforms were moderately diagnostic on average for benign and malignant differentiation. Several f_p metrics showed only slightly lower diagnostic performance in the logistic regression and were highly consistent across software platforms. These results, obtained in the challenging context of a retrospective analysis of

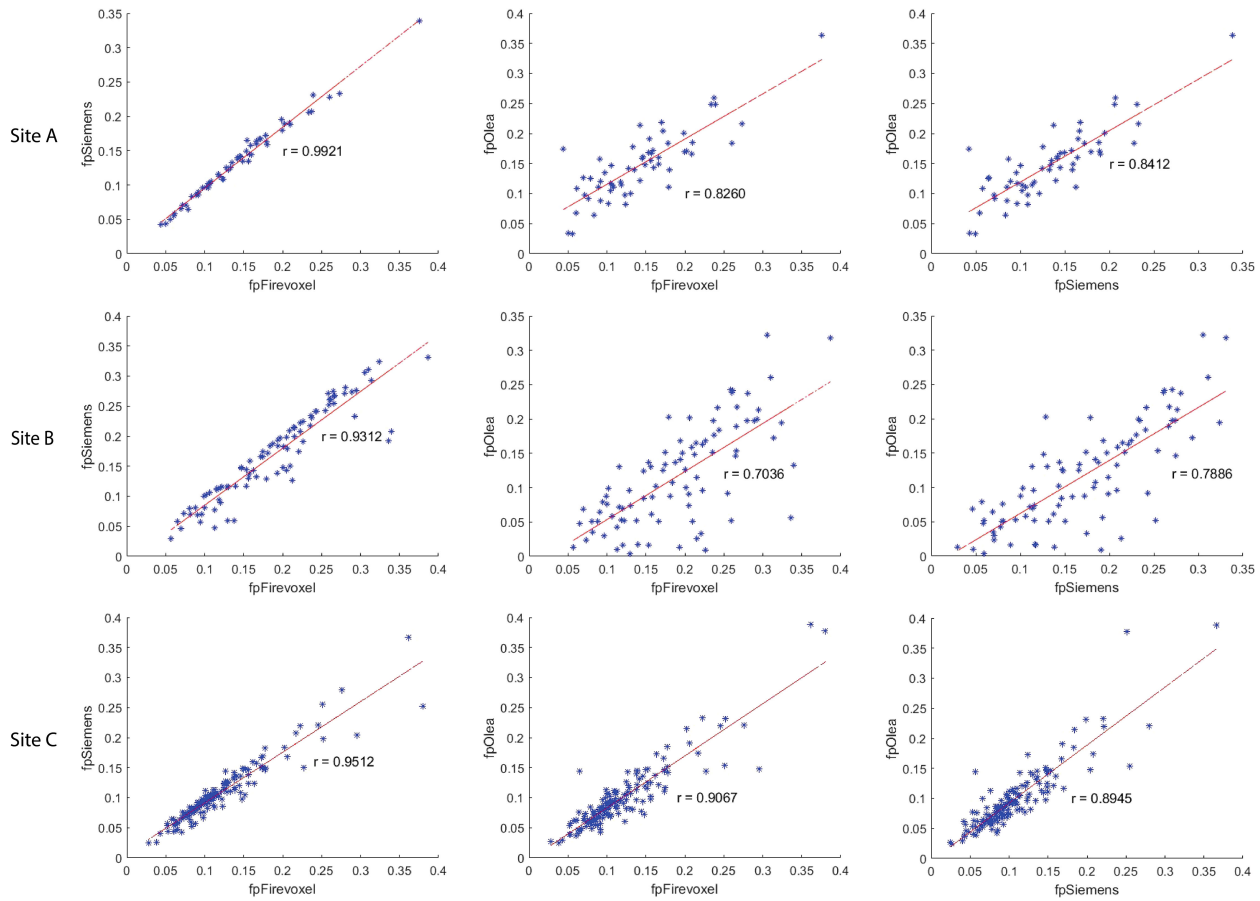


FIGURE 5
Correlation coefficient between Firevoxel, Siemens and Olea for mean of perfusion fraction (f_p) at Site A, Site B and Site C. Comparisons shown left to right: Firevoxel *vs*. Siemens, Firevoxel *vs*. Olea, and Siemens *vs*. Olea. Least-squares segmented algorithms (Firevoxel, Siemens) show the highest agreement while correlation between least-squares and Bayesian algorithms (Olea) is somewhat less.

heterogeneous multi-site data, underline the potential additive value of f_p in future prospective multi-site studies.

In general, consistency of D_t radiomic features from least-squares segmented algorithms and Bayesian algorithms agrees with the study conducted by Scalco et al. (35) in that the choice of the quantification method can be neglected for the extraction of 1st order histogram features from D_t maps in case of retrospective multi-center analyses. However, our study also validated that D_t

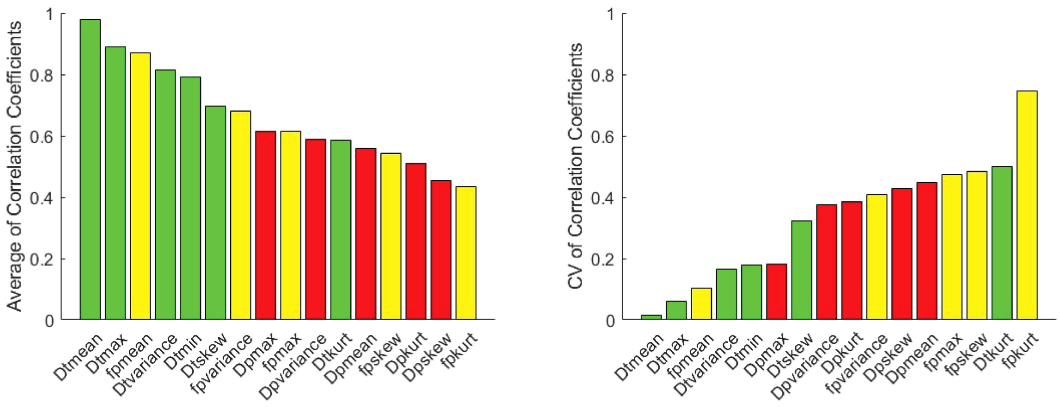


FIGURE 6
Average Pearson correlation coefficients of 1st order radiomic features of f_p (yellow), D_t (green) and D_p (red) between software pairs at Site A, Site B and Site C along with coefficient of variation (CV) of correlation coefficients. Highest correlations are observed for mean D_t (lowest CV) and f_p metrics as well as other D_t radiomics.

TABLE 4 Area under the ROC curve (AUC) with standard error (SE) using logistic regression and leave-one-patient-out (LOU) cross validation AUC with SE for benign and malignant differentiation for different IVIM metrics using Software (a, b, c) from Site A, Site B and Site C.

	Software	AUC (SE)	LOU AUC (SE)		Software	AUC (SE)	LOU AUC (SE)		Software	AUC (SE)	LOU AUC (SE)
$D_t min$	a	0.82 (0.02)	0.8 (0.03)								
	b	0.82 (0.02)	0.8 (0.03)								
	c	0.83 (0.02)	0.8 (0.03)								
$D_t max$	a	0.8 (0.03)	0.78 (0.03)	$f_p max$	a	0.76 (0.03)	0.7 (0.03)	$D_p max$	a	0.81 (0.02)	0.78 (0.03)
	b	0.81 (0.02)	0.79 (0.03)		b	0.77 (0.03)	0.73 (0.03)		b	0.79 (0.03)	0.76 (0.03)
	c	0.8 (0.03)	0.78 (0.03)		c	0.76 (0.03)	0.69 (0.03)		c	0.83 (0.02)	0.81 (0.02)
$D_t mean$	a	0.85 (0.02)	0.83 (0.02)	$f_p mean$	a	0.77 (0.03)	0.73 (0.03)	$D_p mean$	a	0.77 (0.03)	0.74 (0.03)
	b	0.85 (0.02)	0.83 (0.02)		b	0.76 (0.03)	0.72 (0.03)		b	0.78 (0.03)	0.73 (0.03)
	c	0.86 (0.02)	0.84 (0.02)		c	0.77 (0.03)	0.73 (0.03)		c	0.85 (0.02)	0.84 (0.02)
$D_t variance$	a	0.75 (0.03)	0.62 (0.03)	$f_p variance$	a	0.77 (0.03)	0.72 (0.03)	$D_p variance$	a	0.78 (0.03)	0.72 (0.03)
	b	0.77 (0.03)	0.68 (0.03)		b	0.77 (0.03)	0.71 (0.03)		b	0.77 (0.03)	0.71 (0.03)
	c	0.76 (0.03)	0.64 (0.03)		c	0.77 (0.03)	0.7 (0.03)		c	0.81 (0.02)	0.79 (0.03)
$D_t skew$	a	0.82 (0.02)	0.78 (0.03)	$f_p skew$	a	0.77 (0.03)	0.66 (0.03)	$D_p skew$	a	0.8 (0.03)	0.76 (0.03)
	b	0.82 (0.02)	0.8 (0.03)		b	0.77 (0.03)	0.64 (0.03)		b	0.78 (0.03)	0.72 (0.03)
	c	0.85 (0.02)	0.83 (0.02)		c	0.77 (0.03)	0.71 (0.03)		c	0.82 (0.02)	0.79 (0.03)
$D_t kurt$	a	0.78 (0.03)	0.74 (0.03)	$f_p kurt$	a	0.79 (0.03)	0.65 (0.03)	$D_p kurt$	a	0.8 (0.03)	0.75 (0.03)
	b	0.78 (0.03)	0.73 (0.03)		b	0.75 (0.03)	0.68 (0.03)		b	0.78 (0.03)	0.73 (0.03)
	c	0.78 (0.03)	0.74 (0.03)		c	0.77 (0.03)	0.66 (0.03)		c	0.79 (0.03)	0.7 (0.03)

a: Firevoxel; b: Siemens; c: Olea.

10.3389/fonc.2025.1524634

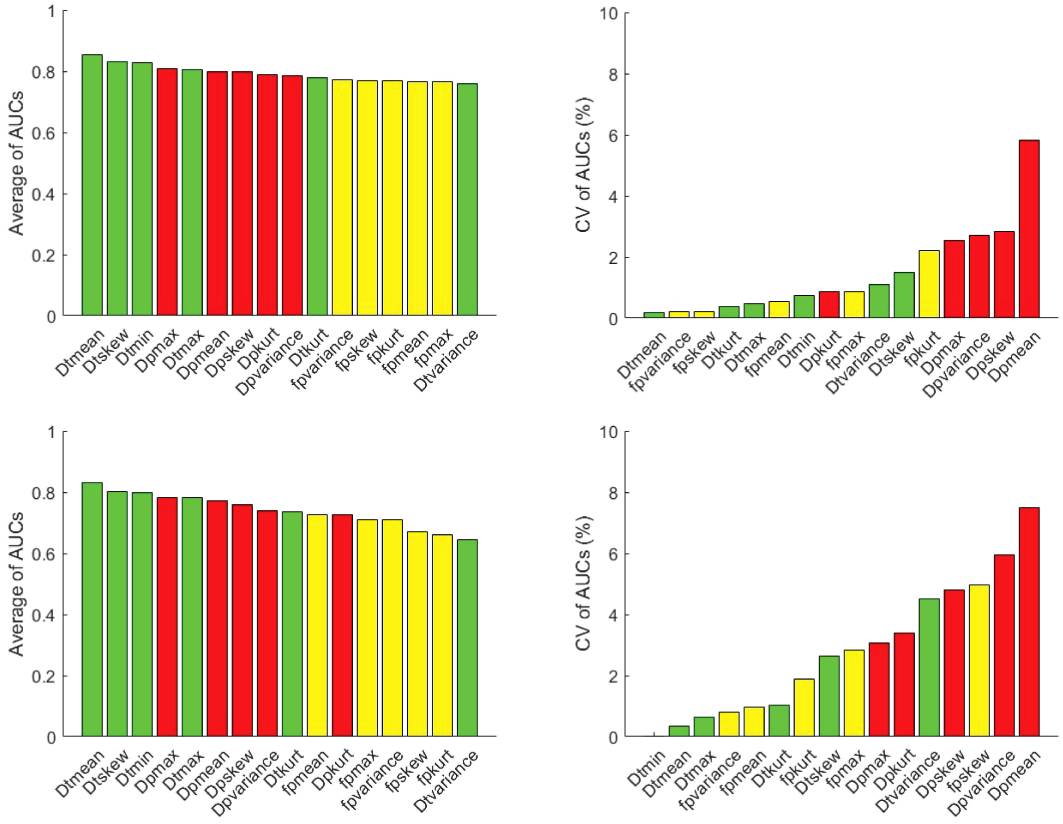


FIGURE 7
Average area under the ROC curve (AUC) and coefficient of variation (CV) of AUC for benign and malignant differentiation via metrics of f_p (yellow), D_t (green) and D_p (red) using logistic regression (top row) and leave-one-patient-out (LOU) cross validation (bottom row). D_t metrics generally show the highest average and most consistent performance for the benign/malignant task, and several f_p metrics (e.g. mean and variance) show high consistency among software.

radiomic features obtained from least-squares segmented fitting could be consistent with the Bayesian fitting and therefore the fitting methods for the estimation of D_t maps could be completely neglected. Their study also revealed that D_p is the most sensitive to quantification method and therefore is less robust across software platforms as demonstrated in this study.

Meeus et al. (41) reported that the constrained IVIM fitting method provides robust and reproducible IVIM parameters particularly D_t and f_p in low-perfused brain tissue similar to our study. D_t consistency across software tools reported in the current study was good and in agreement with the reproducibility studies conducted for phantom (42) and kidney (43–46). In addition, we also observed good f_p reproducibility in most contexts.

Our present study had some limitations. Since the study was retrospective there was no control over the differences in acquisition protocols or hardware platforms at different sites; this might be one of many reasons for inconsistency in IVIM parameter maps particularly D_p. There is a possibility that robustness and consistency of D_p maps among software packages was impacted by the different amount of outlier rejection fractions particularly in the case of D_p maps. D_p maps from Firevoxel generated considerably more lesion voxels outside the histogram range (0 – 0.1 mm^2/s) than did Software b and Software C, which included almost all the lesion voxels. Moreover, the harmonization in b-values would be beneficial for future prospective studies to maintain robustness. Site A in particular may have been affected by heterogeneous sets of b-values and resolution levels within its cohort. The non-Gaussian effect/noise floor was not accounted for in the software used in this study, potentially leading to overestimations of f_p values. While the lesion size among the recruited population in the study cannot be foreseen, however the difference in ROI size in patient population in this retrospective study is also because of the multi-slice segmentation (Site A) or single slice segmentation (Site B and Site C) employed, which could also be the reason for some inconsistency in results. Therefore, uniformity in delineating the lesion must be maintained in addition to recruiting a similar cohort size and consistent recruitment criteria for prospective multicenter studies. Finally, there was some heterogeneity in lesion validation standard (biopsy confirmation at Sites A, B *vs.* radiologic assessment at Site C for benign lesions) in the studied cohorts.

5 Conclusion

Even in a heterogeneous multisite cohort with varying acquisition and analysis settings, certain 1st order IVIM radiomic features (specifically mean, minimum and maximum of D_t) show potential for robustness and diagnostic applicability. Pseudodiffusion features (f_p and D_p) are more sensitive to fit algorithms and clinical cohorts,

but the mean and variance of f_p still demonstrates potential for consistent behavior among site/software contexts that controlled prospective studies might leverage.

Data availability statement

The datasets presented in this article are not readily available because Data sharing is restricted. The current data use agreement doesn't allow it to be shared and post publicly. However, upon request, under appropriate institutional data-use agreements' sharing might be possible. Requests to access the datasets should be directed to DiB, dibash.basukala@nyulangone.org.

Ethics statement

The data collection for studies involving humans was approved by each participating Site's local institutional review board (IRB). Data analysis of all Sites's data in the current work was done under a retrospective IRB approved at NYU, with which all participating Sites partnered with approved Agreements covering the sharing of anonymized data. All studies were conducted in accordance with the local legislation and institutional requirements.

Author contributions

DiB: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Validation. AM: Software, Writing – review & editing. XL: Methodology, Writing – review & editing, Formal analysis, Validation, Visualization. JDG: Methodology, Writing – review & editing, Formal analysis, Validation, Visualization. NG: Writing – review & editing. LM: Funding acquisition, Writing – review & editing. KP: Methodology, Funding acquisition, Writing – review & editing, Data curation, Investigation. SCP: Methodology, Funding acquisition, Writing – review & editing, Data curation, Investigation. DeB: Writing – review & editing, Data curation. MK: Data curation, Investigation, Methodology, Writing – review & editing. MH: Data curation, Writing – review & editing, Investigation, Methodology. MI: Data curation, Writing – review & editing, Investigation, Methodology. SBT: Software, Supervision, Validation, Visualization, Data curation, Investigation, Writing – review & editing, Conceptualization, Funding acquisition, Methodology, Project administration, Resources. EES: Data curation, Investigation, Writing – review & editing, Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Institutes of Health (NIH) (UG3CA239861 (EES, SBT, AM, KP, SCP, DiB, DeB, MK, MI), UH3CA239861 (EES, SBT, DiB, JDG, XL, AM, LM)). EES: This work was performed under the rubric of the Center for Advanced Imaging Innovation and Research (CAI2R, www.cai2r.net), an NIBIB National Center for Biomedical Imaging and Bioengineering (NIH P41 EB017183). JDG and XL: This study was supported by National Institutes of Health/National Cancer Institute Cancer Center Support Grant P30CA016087. SBT: This study was supported by National Institutes of Health/National Cancer Institute Cancer Center Support Grant P30 CA008748. SCP and DeB: This study was supported by Q16 funding from the National Institutes of Health/National Cancer Institute R01CA207290, R01CA248192 and R01CA190299.

Acknowledgments

We acknowledge Mahesh Keerthivasan and Robert Grimm at Siemens Healthineers and Astrid Saulnier at Olea Medical for assistance and useful discussions.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1524634/full#supplementary-material

References

1. DeSantis CE, Ma JM, Gaudet MM, Newman LA, Miller KD, Sauer AG, et al. Breast cancer statistics, 2019. *Ca-Cancer J Clin*. (2019) 69:438–51. doi: 10.3322/caac.21583

2. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. *Neoplasia*. (2009) 11:102–25. doi: 10.1593/neo.81328

3. Partridge SC, Nissan N, Rahbar H, Ae K, Sigmund EE. Diffusion-weighted breast MRI: Clinical applications and emerging techniques. *J Magn Reson Imaging*. (2017) 45:337–55. doi: 10.1002/jmri.25479

4. Gullo RL, Partridge SC, Shin HJ, Thakur SB, Pinker K. Update on DWI for breast cancer diagnosis and treatment monitoring. *AJR Am J Roentgenol*. (2024) 222:e2329933. doi: 10.2214/AJR.23.29933

5. Honda M, Sigmund EE, Le Bihan D, Pinker K, Clauser P, Karampinos D, et al. Advanced breast diffusion-weighted imaging: what are the next steps? A proposal from the EUSOBI International Breast Diffusion-weighted Imaging working group. *Eur Radiol*. (2024). doi: 10.1007/s00330-024-11010-0

6. Iima M, Honda M, Sigmund EE, Ohno Kishimoto A, Kataoka M, Togashi K. Diffusion MRI of the breast: Current status and future directions. *J Magn Reson Imaging*. (2020) 52:70–90. doi: 10.1002/jmri.26908

7. Amornsiripanitch N, Bickelhaupt S, Shin HJ, Dang M, Rahbar H, Pinker K, et al. Diffusion-weighted MRI for unenhanced breast cancer screening. *Radiology*. (2019) 293:504–20. doi: 10.1148/radiol.2019182789

8. Baltzer P, Mann RM, Iima M, Sigmund EE, Clauser P, Gilbert FJ, et al. Diffusion-weighted imaging of the breast-a consensus and mission statement from the EUSOBI International Breast Diffusion-Weighted Imaging working group. *Eur Radiol*. (2020) 30:1436–50. doi: 10.1007/s00330-019-06510-3

9. Chen BY, Xie Z, Nie P, Yang D, Hu YC, Liu ST, et al. Multiple b-value diffusion-weighted imaging in differentiating benign from Malignant breast lesions: comparison of conventional mono-, bi- and stretched exponential models. *Clin Radiol*. (2020) 75:642.e1–.e8. doi: 10.1016/j.crad.2020.03.039

10. Lo Gullo R, Sevilimedu V, Baltzer P, Le Bihan D, Camps-Herrero J, Clauser P, et al. A survey by the European Society of Breast Imaging on the implementation of breast diffusion-weighted imaging in clinical practice. *Eur Radiol*. (2022) 32:6588–97. doi: 10.1007/s00330-022-08833-0

11. Le Bihan D. What can we see with IVIM MRI? *NeuroImage*. (2019) 187:56–67. doi: 10.1016/j.neuroimage.2017.12.062

12. Lebihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Lavaljeantet M. Mr imaging of intravoxel incoherent motions - application to diffusion and perfusion in neurologic disorders. *Radiology*. (1986) 161:401–7. doi: 10.1148/radiology.161.2.3763909

13. Iima M, Yano K, Kataoka M, Umehana M, Murata K, Kanao S, et al. Quantitative non-Gaussian diffusion and intravoxel incoherent motion magnetic resonance imaging: differentiation of Malignant and benign breast lesions. *Invest radiol*. (2015) 50:205–11. doi: 10.1097/RLI.0000000000000094

14. Bokacheva L, Kaplan JB, Giri DD, Patil S, Gnanasigamani M, Nyman CG, et al. Intravoxel incoherent motion diffusion-weighted MRI at 3.0 T differentiates Malignant breast lesions from benign lesions and breast parenchyma. *J Magn Reson Imaging*. (2014) 40:813–23. doi: 10.1002/jmri.24462

15. Liang J, Zeng S, Li Z, Kong Y, Meng T, Zhou C, et al. Intravoxel incoherent motion diffusion-weighted imaging for quantitative differentiation of breast tumors: A meta-analysis. *Front Oncol*. (2020) 10:585486. doi: 10.3389/fonc.2020.585486

16. Ma WL, Mao JW, Wang T, Huang YN, Zhao ZH. Distinguishing between benign and Malignant breast lesions using diffusion weighted imaging and intravoxel incoherent motion: A systematic review and meta-analysis. *Eur J Radiol*. (2021) 141:109809. doi: 10.1016/j.ejrad.2021.109809

17. Sigmund EE, Cho GY, Kim S, Finn M, Moccaldi M, Jensen JH, et al. Intravoxel incoherent motion imaging of tumor microenvironment in locally advanced breast cancer. *Magn Reson Med*. (2011) 65:1437–47. doi: 10.1002/mrm.22740

18. Yao FF, Zhang Y. A review of quantitative diffusion-weighted MR imaging for breast cancer: Towards noninvasive biomarker. *Clin Imaging*. (2023) 98:36–58. doi: 10.1016/j.clinimag.2023.03.016

19. Lu YG, Jansen JFA, Stambuk HE, Gupta G, Lee N, Gonen M, et al. Comparing primary tumors and metastatic nodes in head and neck cancer using intravoxel incoherent motion imaging: A preliminary experience. *J Comput Assist Tomo*. (2013) 37:346–52. doi: 10.1097/RCT.0b013e318282d935

20. Suo ST, Lin N, Wang H, Zhang LB, Wang R, Zhang S, et al. Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: Comparison of different curve-fitting methods. *J Magn Reson Imaging*. (2015) 42:362–70. doi: 10.1002/jmri.24799

21. Callot V, Bennett E, Decking UKM, Balaban RS, Wen H. *In vivo* study of microcirculation in canine myocardium using the IVIM method. *Magn Reson Med*. (2003) 50:531–40. doi: 10.1002/mrm.10568

22. Cho GY, Moy L, Zhang JL, Baete S, Lattanzi R, Moccaldi M, et al. Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer. *Magn Reson Med*. (2015) 74:1077–85. doi: 10.1002/mrm.25484

23. Luciani A, Vignaud A, Cavet M, Van Nhieu JT, Mallat A, Ruel L, et al. Liver cirrhosis: intravoxel incoherent motion MR imaging-pilot study. *Radiology*. (2008) 249:891–9. doi: 10.1148/radiol.2493080080

24. Freiman M, Perez-Rossello JM, Callahan MJ, Voss SD, Ecklund K, Mulkern RV, et al. Reliable estimation of incoherent motion parametric maps from diffusion-weighted MRI using fusion bootstrap moves. *Med Image Anal*. (2013) 17:325–36. doi: 10.1016/j.media.2012.12.001

25. Neil JJ, Bretthorst GL. On the use of bayesian probability-theory for analysis of exponential decay data - an example taken from intravoxel incoherent motion experiments. *Magn Reson Med*. (1993) 29:642–7. doi: 10.1002/mrm.1910290510

26. Orton MR, Collins DJ, Koh DM, Leach MO. Improved intravoxel incoherent motion analysis of diffusion weighted imaging by data driven bayesian modeling. *Magn Reson Med*. (2014) 71:411–20. doi: 10.1002/mrm.24649

27. While PT. A comparative simulation study of bayesian fitting approaches to intravoxel incoherent motion modeling in diffusion-weighted MRI. *Magn Reson Med*. (2017) 78:2373–87. doi: 10.1002/mrm.26598

28. Wada A, Akashi T, Hagiwara A, Nishizawa M, Shimoji K, Kikuta J, et al. Deep learning-driven transformation: A novel approach for mitigating batch effects in diffusion MRI beyond traditional harmonization. *J Magn Reson Imaging*. (2024) 60:510–22. doi: 10.1002/jmri.29088

29. Barbieri S, Gurney-Champion OJ, Klaassen R, Thoeny HC. Deep learning how to fit an intravoxel incoherent motion model to diffusion-weighted MRI. *Magn Reson Med*. (2020) 83:312–21. doi: 10.1002/mrm.27910

30. Bertleff M, Domsch S, Weingartner S, Zapp J, O'Brien K, Barth M, et al. Diffusion parameter mapping with the combined intravoxel incoherent motion and kurtosis model using artificial neural networks at 3 T. *Nmr BioMed*. (2017) 30:e3833. doi: 10.1002/nbm.v30.12

31. Kaandorp MPT, Barbieri S, Klaassen R, van Laarhoven HWM, Crezee H, While PT, et al. Improved unsupervised physics-informed deep learning for intravoxel incoherent motion modeling and evaluation in pancreatic cancer patients. *Magn Reson Med*. (2021) 86:2250–65. doi: 10.1002/mrm.28852

32. Mastropietro A, Procissi D, Scalco E, Rizzo G, Bertolino N. A supervised deep neural network approach with standardized targets for enhanced accuracy of IVIM parameter estimation from multi-SNR images. *Nmr Biomed*. (2022) 35:e4774. doi: 10.1002/nbm.v35.10

33. Baidya Kayal E, Kandasamy D, Khare K, Bakhshi S, Sharma R, Mehndiratta A. Texture analysis for chemotherapy response evaluation in osteosarcoma using MR imaging. *Nmr Biomed*. (2021) 34:e4426. doi: 10.1002/nbm.4426

34. Li HH, Sun B, Tan C, Li R, Fu CX, Grimm R, et al. The value of whole-tumor histogram and texture analysis using intravoxel incoherent motion in differentiating pathologic subtypes of locally advanced gastric cancer. *Front Oncol*. (2022) 12:821586. doi: 10.3389/fonc.2022.821586

35. Scalco E, Rizzo G, Mastropietro A. The quantification of IntraVoxel incoherent motion - MRI maps cannot preserve texture information: An evaluation based on simulated and in-vivo images. *Comput Biol Med*. (2023) 154:106495. doi: 10.1016/j.compbiomed.2022.106495

36. Vidić I, Jerome NP, Bathen TF, Goa PE, While PT. Accuracy of breast cancer lesion classification using intravoxel incoherent motion diffusion-weighted imaging is improved by the inclusion of global or local prior knowledge with bayesian methods. *J Magnet Resonance Imaging*. (2019) 50:1478–88. doi: 10.1002/jmri.26772

37. Gurney-Champion OJ, Klaassen R, Froeling M, Barbieri S, Stoker J, Engelbrecht MRW, et al. Comparison of six fit algorithms for the intravoxel incoherent motion model of diffusionweighted magnetic resonance imaging data of pancreatic cancer patients. *PloS One*. (2018) 13:1–18. doi: 10.1371/journal.pone.0194590

38. Taimouri V, Afacan O, Perez-Rossello JM, Callahan MJ, Mulkern RV, Warfield SK, et al. Spatially constrained incoherent motion method improves diffusion-weighted MRI signal decay analysis in the liver and spleen. *Med Phys*. (2015) 42:1895–903. doi: 10.1118/1.4915495

39. Barbieri S, Donati OF, Froehlich JM, Thoeny HC. Impact of the calculation algorithm on biexponential fitting of diffusion-weighted MRI in upper abdominal organs. *Magnet Resonance Med*. (2016) 75:2175–84. doi: 10.1002/mrm.25765

40. Klein R. Bland-Altman and Correlation Plot (2025). Available online at: https://www.mathworks.com/matlabcentral/fileexchange/45049-bland-altman-and-correlation-plot (Accessed December 24, 2024).

41. Meeus EM, Novak J, Withey SB, Zarinabad N, Dehghani H, Peet AC. Evaluation of intravoxel incoherent motion fitting methods in low-perfused tissue. *J Magn Reson Imaging*. (2017) 45:1325–34. doi: 10.1002/jmri.25411

42. Basukala D, Mikheev A, Sevilimedu V, Gilani N, Moy L, Pinker K, et al. Multisite MRI intravoxel incoherent motion repeatability and reproducibility across 3 T scanners in a breast diffusion phantom: A BReast intravoxel incoherent motion multisite (BRIMM) study. *J Magn Reson Imaging*. (2024) 59:2226–37. doi: 10.1002/jmri.29008

43. de Boer A, Harteveld AA, Stemkens B, Blankestijn PJ, Bos C, Franklin SL, et al. Multiparametric renal MRI: an intrasubject test-retest repeatability study. *J Magn Reson Imaging*. (2021) 53:859–73. doi: 10.1002/jmri.27167

44. Makino Y, Ohno N, Miyati T, Hori N, Matsuura Y, Kobayashi S, et al. Tri- and bi-exponential diffusion analyses of the kidney: effect of respiratory-controlled acquisition on diffusion parameters. *Radiol Phys Technol*. (2023) 16:478–87. doi: 10.1007/s12194-023-00734-1

45. Sigmund EE, Vivier PH, Sui D, Lamparello NA, Tantillo K, Mikheev A, et al. Intravoxel incoherent motion and diffusion-tensor imaging in renal tissue under hydration and furosemide flow challenges. *Radiology*. (2012) 263:758–69. doi: 10.1148/radiol.12111327

46. Thoeny HC, De Keyzer F, Oyen RH, Peeters RR. Diffusion-weighted MR imaging of kidneys in healthy volunteers and patients with parenchymal diseases: initial experience. *Radiology*. (2005) 235:911–7. doi: 10.1148/radiol.2353040554

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 May 2025
DOI 10.3389/fonc.2025.1498832

Check for updates

OPEN ACCESS

EDITED BY
Sharon R. Pine,
University of Colorado Anschutz Medical Campus, United States

REVIEWED BY
Gusheng Tang,
Naval Medical University, China
Sibaji Gaj,
Cleveland Clinic, United States
Patrick McNeely,
George Washington University, United States

*CORRESPONDENCE
Humera Tariq
humeratariq883@gmail.com
Thomas L. Chenevert
tlchenev@med.umich.edu

RECEIVED 19 September 2024
ACCEPTED 28 April 2025
PUBLISHED 23 May 2025

CITATION
Tariq H, Hadjiiski L, Malyarenko D, Talpaz M, Pettit K, Luker GD, Ross BD and Chenevert TL (2025) A deep learning aided bone marrow segmentation of quantitative fat MRI for myelofibrosis patients.
Front. Oncol. 15:1498832.
doi: 10.3389/fonc.2025.1498832

COPYRIGHT
© 2025 Tariq, Hadjiiski, Malyarenko, Talpaz, Pettit, Luker, Ross and Chenevert. 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.

A deep learning aided bone marrow segmentation of quantitative fat MRI for myelofibrosis patients

Humera Tariq[1*], Lubomir Hadjiiski[1], Dariya Malyarenko[1], Moshe Talpaz[2], Kristen Pettit[2], Gary D. Luker[1], Brian D. Ross[1] and Thomas L. Chenevert[1*]

[1]Department of Radiology, University of Michigan, Ann Arbor, MI, United States, [2]Department of Internal Medicine-Hematology/Oncology, University of Michigan, Ann Arbor, MI, United States

Purpose: To automate bone marrow segmentation within pelvic bones in quantitative fat MRI of myelofibrosis (MF) patients using deep-learning (DL) U-Net models.

Methods: Automated segmentation of bone marrow (BM) was evaluated for four U-Net models: 2D U-Net, 2D attention U-Net (2D A-U-Net), 3D U-Net and 3D attention U-Net (3D A-U-Net). An experienced annotator performed the delineation on in-phase (IP) pelvic MRI slices to mark the boundaries of BM regions within two pelvic bones: proximal femur and posterior ilium. The dataset comprising volumetric images of 58 MF patients was split into 32 training, 6 validation and 20 test sub-sets. Model performance was assessed using conventional metrics: average Jaccard Index (AJI), average Volume Error (AVE), average Hausdorff Distance (AHD), and average Volume Intersection Ratio (VIR). Iterative model optimization was performed based on maximizing validation sub-set AJI. Wilcoxon's rank sum test with Bonferroni corrected significance threshold of $p<0.003$ was used to compare DL segmentation models for test sub-set.

Results: 2D segmentation models performed best for iliac BM with achieved scores of 95-96% for the VIR and 87-88% for AJI agreement with expert annotations on the test set. Similar performance was observed for femoral BM segmentation with slightly better VIR but worse AJI agreement for U-Net (94% and 86%) versus A-U-Net (92% and 87%). 2D models also exhibited lower AVE variability (8-9%) and ilium AHD (16 mm). The 3D segmentation models have shown marginally higher errors (AHD of 19-20 mm for ilium and 10-12% AVE-SD for both bones) and generally lower agreement scores (VIR of 91-93% for ilium and 89-91% for femur with 85-86% AJI). Pairwise comparison across four U-Nets for three metrics (AHD, AJI, AVE) showed that AJI and AHD performance was not significantly different for 3D U-Net versus 3D A-U-Net and for 2D U-Net versus 2D A-U-Net. Except for AVE, for majority of performance metric comparisons 2D versus 3D model differences were significant in both bones ($p<0.001$).

Conclusion: All four tested U-Net models effectively automated BM segmentation in pelvic MRI of MF patients. The 2D A-U-Net was found best overall for BM segmentation in both femur and ilium.

KEYWORDS

myelofibrosis, proton density fat fraction (PDFF), pelvic MRI, in-phase (IP), proximal femur, posterior ilium, segmentation, U-Net

1 Introduction

Myelofibrosis (MF) is a chronic malignancy characterized by clonal proliferation of hematopoietic stem cells. Hallmark features of MF include bone marrow (BM) fibrosis with reticulin or collagen deposits, heterogeneous expression of inflammatory cytokines, anemia, and enlargement of the liver and spleen (hepatosplenomegaly) (1, 2). BM aspirates from ilium with needle biopsies are currently used for clinical MF patient management to assess systemic inflammation, progressive fibrosis and changes in composition. However, these procedures are painful, select limited tissue volume, subject to histopathological sampling errors, and provide inadequate assessment of BM heterogeneity (3, 4). BM heterogeneity measurement in MF patients constantly pose a challenge in hematopathology (3). A typical MF needle biopsy samples small fraction of the area visualized by a pelvic MRI examination which covers >200 cm^2 of bones encompassing iliac wings, femoral heads, trochanter, and proximal femoral shafts. The non-invasive MRI holds potential for comprehensive MF disease monitoring (3, 5, 6) at multiple anatomic bone marrow sites increasing patient comfort for repeated examinations, anatomic coverage and diagnostic accuracy.

Unlike limited biopsy access, MRI affords large volume coverage for more extensive BM survey that is critical for accurate evaluation of disease heterogeneity at MF primary site (3) and corresponding treatment efficacy during longitudinal monitoring (4–6). Non-invasive assessments of spatially heterogeneous BM pathologies would benefit from the development of robust BM MRI protocols and quantitative imaging biomarkers (QIBs) (7). Quantitative fat MRI is being investigated to assess progressive fibrosis, or reduction in blood cell production due to replacement of normal fat content by cancer cells in MF patients (5). Pre-clinical MF studies (8, 9) also investigate imaging biomarkers to establish quantitative thresholds for comprehensive monitoring of BM disease progression, heterogeneity and therapy response (6, 10). Several promising QIBs are currently investigated for MF bone marrow assessment (5, 8, 9). Proton density fat fraction (PDFF) measures fat content, T2* assesses fibrosis and iron content, apparent diffusion coefficient (ADC) reflects cellular density and microstructural changes, magnetization transfer ratio (MTR) measures the amount of free water within tissue, providing information about tissue composition and pathology, and magnetic resonance elastography (MRE) evaluates soft tissue stiffness, indicating the degree of fibrosis.

Manual delineations of BM within MF patients' pelvic bones is a major MRI QIB analysis bottleneck both for clinical and pre-clinical studies as they are time-consuming and irreproducible due to varying acquisition parameters, contrasts and intra- and inter-observer biases (11–13). Manual bone annotations are pivotal for detailed regional assessment of QIB histograms, e.g., in the posterior ilium versus proximal femur versus lumber vertebrae. However, they substantially slow down the analysis and are known to introduce inaccuracies due to operator training/bias (11, 12) which consequently have limited clinical acceptance of MF QIBs to date. The annotated data-sets available for training automated BM segmentation models are usually small and protocol-specific (11, 13) presenting a challenge for generalization (14).

The Standard U-Net deep learning (DL) models use coarse-to-fine training strategy, facilitated by skip connections that enable accurate segmentation and reproducible boundary delineation (15, 16) and therefore have shown promise for automating segmentation of quantitative imaging applications with small annotated training sets (14). The latest studies for assessment of spatial distribution of PDFF in vertebral BM also adopted U-Net models (11–13). An Attention U-Net (A-U-Net) enhances the standard U-Net by integrating attention gate mechanisms (17), which focus on relevant regions improving segmentation accuracy, especially for small or complex structures. Our group has recently successfully applied 2D Attention-U-Net (A-U-Net) for tibia segmentation of a preclinical model of myelofibrosis (12). The present study sought to evaluate the viability of 2D and 3D U-Net segmentation models to accurately localize femoral and ilium BM regions on in-phase (IP) image volumes of MF patients.

2 Materials and methods

2.1 MRI data acquisition and preparation

A single center MRI study for IRB-consented MF subjects was performed using standard imaging protocol. All MF patient MR images were acquired using clinical mDIXON-QUANT protocol

with images reconstructed on a 3T scanner including quantitative PDFF maps (Figure 1). Patient demographics and image counts for each bone site are available in Supplementary Table S1. Typical pelvic MRI scanned volume comprised 57 axial slices (3 mm thick) within 400x400 mm^2 FOV with image size of 288x288 voxels. An expert annotator performed manual segmentations of proximal femoral and posterior iliac bones using in-phase (IP) Dixon MR images. Amongst four Dixon contrasts (IP, OP, PDFF, T2*), IP was deemed the most appropriate for manual annotation as it provided good soft tissue versus cortical bone contrast largely independent of marrow fat content and thus was most consistent among subjects (Figure 1). Expert annotations were only performed for the inferior (proximal femur) and superior (posterior iliac) sub-volumes. These annotations served as a reference standard in DL aided segmentation.

The original MF patient's pelvic MRI volume did not provide a fixed number of images in each bone due to inconsistent patient positioning. The selection of sub-volumes for the pelvic bone sites therefore requires expert knowledge to determine the first and last image for each bone site. A single expert performed the systematic selection of the first and last slice of sub-volume to reduce the processing time. The expert used the most posterior point of the ilium as a reference, and the most superior slice and the most inferior slice of the segmentation were 1 cm and 3 cm from the reference point. For the femur, the top of the femur head was selected as the 1st slice, and the last inferior slice was the last slice of the femur sub-volume. The MF patient's proximal femur region extended across 14–32 images while posterior ilium spanned 11–14 images.

2.2 Segmentation workflow

The BM segmentation model development and evaluation workflow is illustrated in Figure 2. The 58 MF patient MRI dataset was split into 55% (32/58) training, 10% (6/58) validation and 34% (20/58) test sub-sets respectively. Independent reviewer subjectively scored the fat content of the femoral and iliac bone marrow (BM) by inspecting PDFF maps. The training and validation subsets were selected by independent reviewer to include balanced samples of low to high BM fat content. The pelvic bone IP images and corresponding manual annotations were used iteratively to build separate U-Net models for segmentation of proximal femoral BM and posterior iliac BM. All models were developed, evaluated and tested on NVIDIA RTX A6000 GPU with 48 GB of memory and Pytorch library (version 2.3.0+cu118).

We compared segmentation performance of four model architectures (2D and 3D U-Net and attention (A)-U-Net) for each of the bone sites (Figure 2). The models were first trained by minimizing the average Jaccard index (AJI) loss to obtain all candidate DL models. The model hyperparameters were fine-tuned and the four best models were selected for each bone site using the validation dataset and maximizing AJI for optimization (Supplementary Figures S1, S2). The optimized models were then deployed to the held-out test set and the automated segmentations were compared to the reference annotations. Results of best model deployment on the test dataset (i.e. assessment metrics for agreement with manual annotations) were recorded and compared.

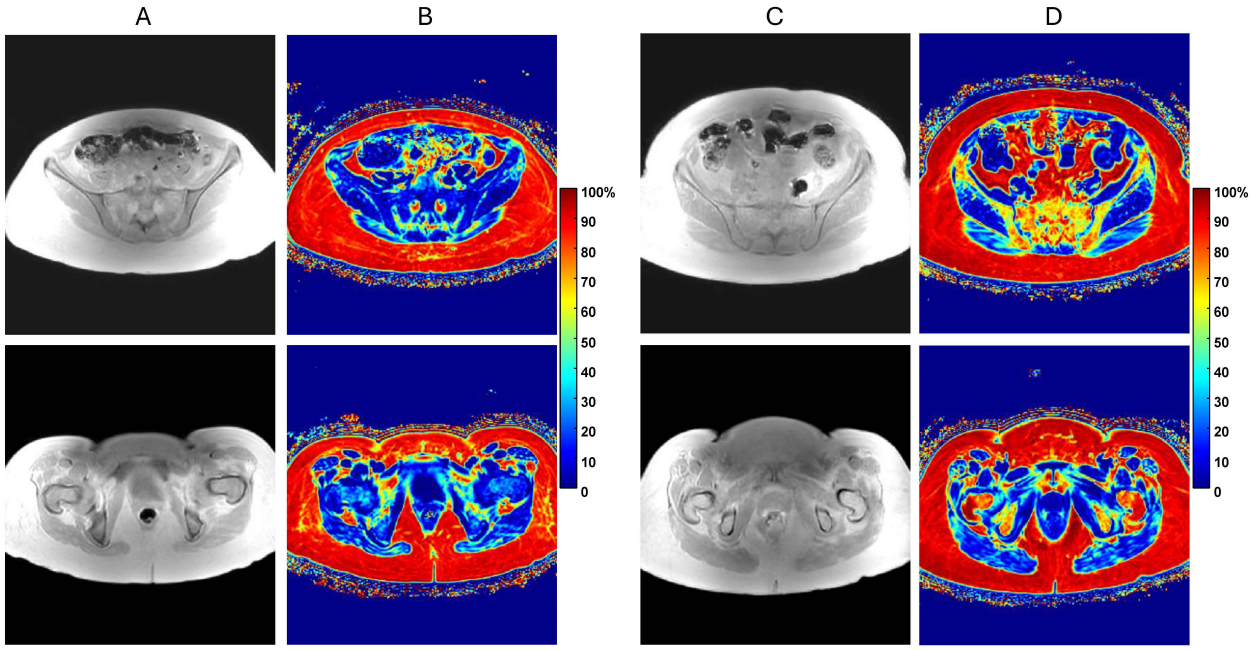


FIGURE 1
(A) In-phase (IP) pelvic MRI of myelofibrosis (MF) patient with low bone marrow fat content for axial slice through ilium (top) and femur (bottom). **(B)** Corresponding quantitative proton-density fat-fraction (PDFF) maps shown on 0 to 100% color scale. **(C)** Similar IP MRI of MF patient with relatively high fat content with **(D)** associated PDFF maps. Note, only IP images were used for segmentation by expert annotator and DL models.

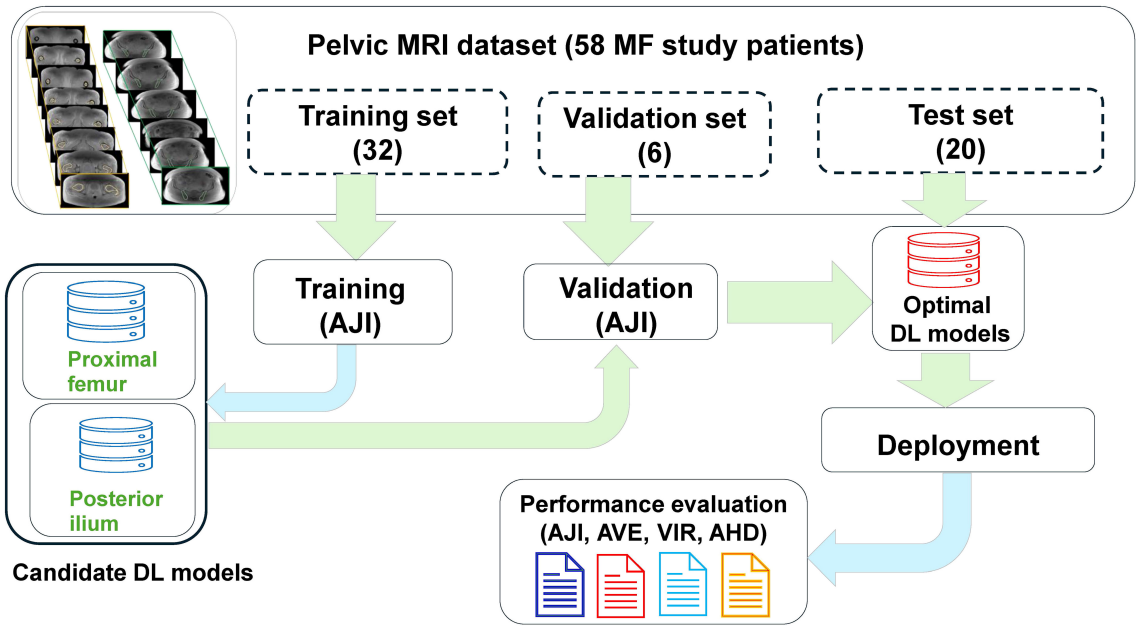


FIGURE 2
Deep learning (DL) segmentation optimization workflow. The training and validation sub-sets (with sizes are listed in parenthesis) were used for iterative model optimization at two bone sites (left workflow arm) based on average Jaccard Index (AJI) metrics. The optimal DL models were deployed for all sub-sets, including test. The segmentation performance was evaluated using volume difference and contour agreement metrics, including AJI, average Volume Error (AVE), Volume Intersection Ratio (VIR) and average Hausdorff distance (AHD), The top left inset shows example of expert annotations for femoral (left) and iliac (right) BM sub-volumes.

AJI metrics smoothness and stability were used for optimized model selection in DL-aided segmentation to aid generalizability and reliability (18, 19). Training and validation plots for four U-Net models in each bone (femur and ilium) along with brief description are provided in Supplementary Figure S1 (2D U-Net) and Supplementary Figure S2 (3D U-Net). 2D and 3D U-Net models differ from each other in terms of training time and hyperparameters, as summarized in Supplementary Table S2. Due to the difference between data premise of (3D) volume versus (2D) image input, U-Net models utilized distinct sample size, hyperparameters, weight updates (iterations) and training times. For 2D segmentation models, varying number of slices was allowed both for training and testing. To enforce consistent femoral sub-volume for 3D segmentation models, replication of the most inferior image with its associated mask was used as needed resulting in consistent femoral sub-volume of size 288x288x24 voxels. Superior iliac sub-volume needed inclusion of 2 to 3 existing iliac images resulted in consistent sub-volume size of 288x288x16 voxels for 3D models. The added replicate images were excluded from the downstream 3D model performance evaluation.

2.3 Post-processing

To improve segmentation and eliminate contouring errors for performance evaluation, post-processing was performed for hole filling and spurious noise removal after U-Net segmentation (e.g., Supplementary Figure S3). First, voxel discontinuities have been detected using nested contouring approach and seeded region growing was used to fill in the small holes. Selection of left and right bone regions have been made through connected component analysis and noise removal performed separately for left and right regions. The segmentations were visualized in 3D Slicer V4.11 and saved as meta-image header (MHD) format.

2.4 Performance evaluation metrics

Model performance was assessed on region-based and boundary-based overlap metrics. Region based overlap included average (over subjects) Jaccard Index (AJI), average Volume Intersection Ratio (VIR) and average Volume Error (AVE), while the boundary-based overlap was represented by average Hausdorff Distance (AHD). These metrics were chosen to ensure comprehensive performance analysis based on complementary measures of agreement (overlap and intersect) and bias (14, 18, 19). All metrics were calculated for individual slices of each patient MRI, and then averaged over the patients The inter-model variability for the performance metrics were assessed using standard deviation (SD). The F1-score was used to measure test's accuracy (20, 21), similar to DICE coefficient (Supplementary Materials). The relevant performance metrics calculations and definitions are summarized below. Jaccard Index: The Jaccard Index was defined as the ratio of the intersection of the predicted segmentation mask and the ground truth mask to their union. The Jaccard Index for a single slice or volume was calculated as

$$Jaccard\ \ Index = \frac{v_r \cap v_p}{v_r \cup v_p}$$

where v_r and v_p were the pixels/voxels of the segmentation mask drawn by expert as reference standard and predicted by U-Net, respectively. Additionally, the Jaccard index could be described using the concepts of true positives (TP), false positives (FP), and false negatives (FN):

True Positives (TP): The number of correctly predicted pixels/voxels of the target class.

False Positives (FP): The number of pixels/voxels incorrectly predicted as the target class.

False Negatives (FN): The number of pixels/voxels that were the target class but were not predicted as such.

$$Jaccard\ \ Index\ \ = \frac{TP}{FP + TP + FN}$$

Volume Intersection Ratio: The volume intersection ratio for the segmentation of a single scan was calculated as

$$Volume\ \ Intersection\ \ Ratio = \frac{v_r \cap v_p}{v_r}$$

where v_r and v_p were defined as above.

Volume Error: The VE quantifies the difference between reference and predicted mask volumes relative to the reference and is calculated as

$$\text{Volume Error (VE)} = 100 * (\frac{v_r - v_p}{v_r}\ )$$

where v_r and v_p are defined above. A positive value of volume error indicates under-segmentation (FN), and a negative value indicates over-segmentation (FP) by U-Net models.

Hausdorff Distance:

In our MRI analysis, each slice contains a left-right pair of segmentation (predicted, P, and reference, R) contours corresponding to the left and right bones on slice i:

$$L^P(i), L^R(i) : the\ pair\ of\ set\ of\ contour\ points\ for\ left\ bone\ on\ slice\ i$$

$$R^P(i), R^R(i) : the\ pair\ of\ set\ of\ contour\ points\ for\ right\ bone\ on\ slice\ i$$

Using our notation, the Hausdorff Distance is computed separately for each bone on every slice i. For left bone it is expressed as follows:

$$d_H^L(i) = \left\{ \max_{p \in L^P(i)} \min_{r \in L^R(i)} d(p,r),\ \ \max_{r \in L^R(i)} \min_{p \in L^P(i)} d(p,r) \right\}$$

where $d(p,r)$ is the Euclidean distance between contour points p and r.

Similarly, for right bone Hausdorff distance becomes:

$$d_H^R(i) = \left\{ \max_{p \in R^P(i)} \min_{r \in R^R(i)} d(p,r),\ \ \max_{r \in R^R(i)} \min_{p \in R^P(i)} d(p,r) \right\}$$

Since each slice provides both left and right contours, we define the representative Hausdorff distance for a slice as the maximum (worst-case) value between the two:

$$d_H^{max}(i) = \max\left\{ d_H^L(i),\ \ d_H^R(i) \right\}$$

For a patient with N slices, the overall performance is summarized by two metrics:

$$Average\ \ Hausdorff\ \ Distance\ \ (AHD) = \frac{1}{N}\ \ \sum_{i=1}^{N} d_H^{max}(i)$$

$$\text{Median Hausdorff Distance (MEDIAN _ AHD)}$$

$$= \text{median}\ \{ d_H^{max}(i)\ \mid\ i = 1, 2, \ldots . N \}\ .$$

2.5 Metric distribution analysis for model comparison

To assess the performance difference between 2D and 3D models, the pair-wise Wilcoxon rank sum test for metric comparison among AHD, AJI and AVE was performed. The performance metrics distributions were visualized using violin plots. All violin plots used kernel density estimation (KDE) for smoothing with lower bandwidth (0.3) to capture finer details (peaks and valleys) in test data results. The violin plots were generated in Python. Bonferroni correction was applied to adjust significant p-value thresholds for multiple-comparisons from initial significance level (α) 0.05 to the significant threshold $p<0.003$ for each U-Net model performance metric comparison test (22).

3 Results

The eight selected U-Net models reasonably mimicked the expert annotations from training data and generalized well on test data in both 2D and 3D for two bone-sites with good agreement VIR (90-96%) and AJI (85-88%) scores. Multi-facet evaluation of model performance on the validation sub-sets, from visual assessments to quantitative performance metrics and statistical details helped choose a preferred U-Net model for BM segmentation of the studied pelvic bone sites (proximal femur and posterior ilium). Additional details about models' selection are included in the Supplementary Materials.

3.1 Qualitative segmentation evaluation

An example of best selected models that adequately followed the reference contours and successfully localized the femoral and iliac BM is illustrated in Figure 3. The bone contours learned by U-Net models were largely consistent with reference outlines with minor variation around the boundaries. The model contours were slightly

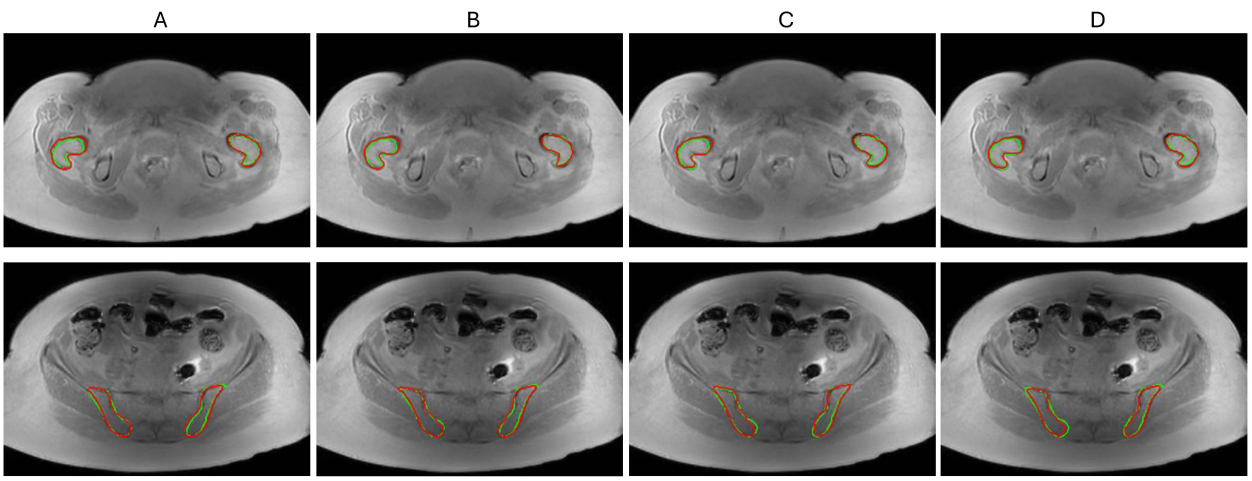


FIGURE 3
Segmentation model comparison for femoral BM (top) and iliac BM (bottom) on a test set **(A)** 2D U-Net **(B)** 2D A-U-Net **(C)** 3D U-Net **(D)** 3D A-U-Net. The reference contours are green and model segmentation contours are red.

more deviant in femoral BM as compared to iliac BM where regional overlay fully covered the reference iliac contours except in 3D A-U-Net (Figure 3D, bottom).

An example of large segmentation errors observed for the studied data set is shown in Figure 4 for a single test MF patient with challenging metal artifact due to a hip implant. Femoral bone (top) BM segmentations in Figure 4C follow reference contours relatively well but also miss some true BM pixels causing false negatives (FNs). The iliac bone (bottom) segmentations by all U-Net models also largely overlap with expert annotations but incur few false positives (FPs) outside the reference boundaries. This example showed that U-Net models largely trained on symmetric bones exhibited the ability to generalize well for a single-side bone segmentation. Additional examples of segmentation errors observed primarily for the marginal slices of imaged bone sub-volume are illustrated in Supplementary Figure S3. We also observed that denoising post-processing affected less than 4% of total segmentation volumes and was apparently required more often for 3D than for 2D models and for femur versus ilium, likely reflecting limited training set size for the bone site anatomy.

3.2 Quantitative performance evaluation

Tables 1 and 2 systematically compare region and contour agreement and error scores (AJI, AVE, VIR, AHD and

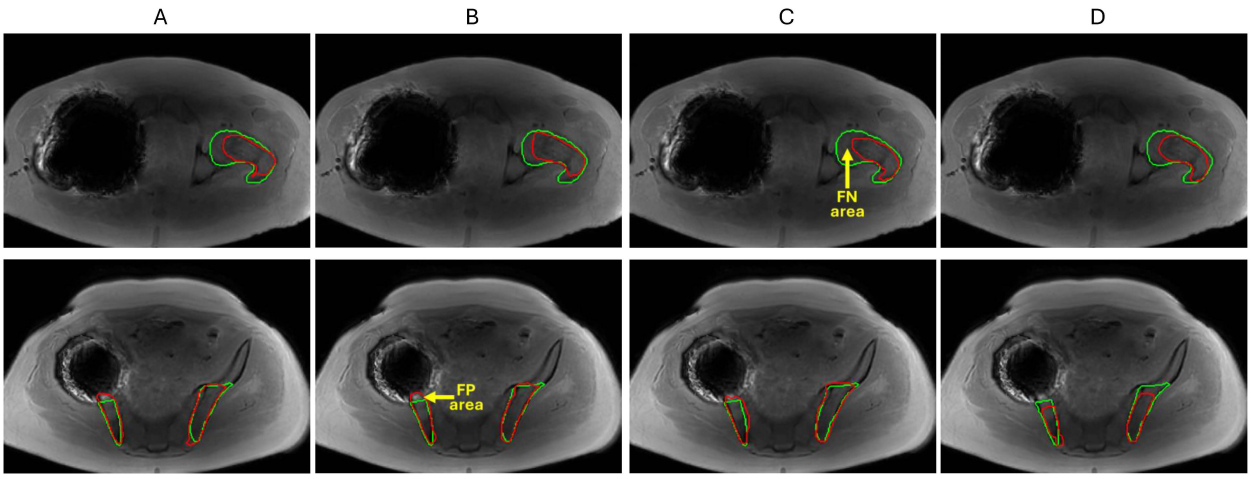


FIGURE 4
Segmentation model comparison for femoral BM (top) and iliac BM (bottom) for a challenging test patient (with hip implant artifact) **(A)** 2D U-Net **(B)** 2D A-U-Net **(C)** 3D U-Net **(D)** 3D A-U-Net. The reference contours are green and model segmentation contours are red. Yellow arrows point at examples of false negative (FN) and false positive (FP) areas.

TABLE 1 Performance metrics [mean ± standard deviation (SD)] for proximal femur segmentation with different tested U-Net models.

Bone-Site	Proximal Femur					
DL Models	Dataset	AJI%	VIR %	AVE%	AHD (mm)	MEDIAN_AHD (mm)
2D U-Net	Training	89.1 ± 2.0	94.3 ± 1.5	0.7 ± 5.5	29.7 ± 7.1	5.1
	Validation	88.5 ± 1.8	93.6 ± 2.3	0.6 ± 8.0	31.4 ± 5.3	5.6
	Test	**85.5 ± 6.7**	**93.3 ± 2.2**	**0.2 ± 8.0**	**28.4 ± 8.9**	**5.7**
2D A-U-Net	Training	89.0 ± 1.4	92.5 ± 1.6	6.0 ± 5.0	28.5 ± 6.5	5.4
	Validation	88.3 ± 2.2	92.2 ± 3.0	5.0 ± 9.0	32.0 ± 3.7	5.1
	Test	**86.9 ± 2.5**	**91.5 ± 3.5**	**5.0 ± 9.0**	**26.8 ± 9.3**	**6.2**
3D U-Net	Training	87.8 ± 2.0	93.0 ± 2.0	2.4 ± 6.4	26.7 ± 6.3	5.8
	Validation	86.0 ± 3.0	91.3 ± 3.0	2.9 ± 10.0	28.3 ± 6.9	6.8
	Test	**85.7 ± 2.3**	**91.0 ± 3.1**	**5.2 ± 10.2**	**25.5 ± 9.0**	**6.2**
3D A-U-Net	Training	87.8 ± 2.1	91.6 ± 2.1	7.3 ± 6.4	26.9 ± 7.0	5.6
	Validation	86.1 ± 3.0	89.6 ± 3.4	10.3 ± 7.7	29.7 ± 5.7	6.7
	Test	**84.9 ± 4.0**	**89.0 ± 5.0**	**11.6 ± 10.8**	**25.8 ± 9.5**	**5.8**

The values in bold indicate performance metrics for the independent test set.

MEDIAN_AHD) for proximal femur and posterior iliac BM segmentation across training, validation and test sets after denoising post-processing. This analysis confirms the moderately better performance of 2D U-Net models over 3D U-Net models which otherwise were not obvious from visual inspection of segmentations. As expected, validation and testing performances were always lower than the corresponding training performances in all metrics, except femur AVE and AHD.

Table 1 illustrates that the slice-wise 2D AHD for all four U-Net models was relatively high in femoral BM (26–32 mm) with wide range of standard deviations (SD=4-9 mm). Median AHD for test set varied less between 2D and 3D U-Net models (5.7-6.2 mm) indicating comparable performance for directed boundary distance in femur. This confirmed the likely source of higher 2D AHD values from marginal slices in the femur volume. In femoral BM, both AVE and SD(AVE) were lower for 2D versus 3D model

TABLE 2 Performance metrics [mean ± standard deviation (SD)] for posterior ilium segmentation with different tested U-Net models.

Bone-Site	Posterior Ilium					
DL Models	Dataset	AJI%	VIR %	AVE%	AHD (mm)	MEDIAN_AHD (mm)
2D U-Net	Training	90.3 ± 1.5	96.4 ± 1.1	-4.5 ± 5.2	9.6 ± 4.0	5.1
	Validation	89.4 ± 1.1	96.5 ± 1.7	-7.3 ± 7.4	10.1 ± 3.0	5.5
	Test	**87.0 ± 3.6**	**95.7 ± 2.6**	**-10.4 ± 11.5**	**18.1 ± 10.7**	**6.8**
2D A-U-Net	Training	91.0 ± 1.5	96.0 ± 1.3	-1.5 ± 5.0	8.9 ± 2.6	4.4
	Validation	89.4 ± 1.2	95.6 ± 1.9	-4.6 ± 5.6	15.0 ± 6.8	5.6
	Test	**88.0 ± 2.4**	**95.5 ± 1.8**	**-7.5 ± 9.0**	**16.1 ± 8.2**	**5.7**
3D U-Net	Training	88.0 ± 1.8	93.4 ± 2.4	-0.2 ± 6.6	12.7 ± 4.0	6.3
	Validation	86.8 ± 2.0	94.0 ± 2.8	-3.8 ± 7.4	18.9 ± 11.1	8.9
	Test	**85.0 ± 5.6**	**93.0 ± 6.0**	**-4.3 ± 12.0**	**20 ± 13.0**	**8.4**
3D A-U-Net	Training	87.8 ± 2.1	91.5 ± 3.2	6.0 ± 6.8	14.3 ± 6.3	6.9
	Validation	87.6 ± 1.9	92.3 ± 2.7	2.7 ± 6.5	16.1 ± 7.5	6.1
	Test	**85.1 ± 4.7**	**90.6 ± 5.3**	**3.5 ± 10.7**	**19.1 ± 10.3**	**6.8**

The values in bold indicate performance metrics for the independent test set.

segmentations. e.g. (AVE: 0.2-6% versus 2-12% with SD 5 – 9% versus 6-11%). Interestingly, the mean values of contour-distance (AHD) decreased for the test set as compared to training set for 2D models (e.g. from 27-28 mm versus 29–30 mm). The AHD for 3D U-Net models (26 mm) were slightly lower than for 2D models and similarly lower than training AHD (27 mm) and similar SD (9-10 mm).

Table 1 further shows that 2D U-Net achieved marginally higher VIR on the test set compared to the 2D A-U-Net in proximal femur (93% versus 92%) while the trend was reverse for AJI (86% versus 87%). The 2D models agreement metrics (AJI, VIR) reflected similar performance across training, validation and test sets (AJI: 86-89%, VIR: 92-94%), consistently better than for 3D models (AJI: 85-88%, VIR: 89-93%). The relatively small standard deviations (SDs) of 2-3% were observed for 2D A-U-Net and 3D U-Net but roughly doubled from training to test set for agreement measures of 2D U-Net and 3D A-U-Net. Improvement in agreement and lower variability between 2D and 3D models for training, test and validation set demonstrated that 2D models marginally outperformed the 3D models for femur segmentation.

Table 2 summarizes the performance of four iliac U-Net models using overlap and boundary metrics. The ilium contour errors are surprisingly lower in comparison to femur (2D Ilium AHD: 9-18 mm, 2D femur AHD: 26-32 mm). Contrary to femur, no anomalous trends on test versus training set were observed in AHD of iliac U-Net models with both mean (16–20 mm) and SDs (8–11 mm) increasing for the test set. Table 2 shows that 2D U-Net and 2D A-U-Net training, validation and testing for iliac segmentation models were consistent in both AJI and VIR and small performance decline (3-5%) was observed from training to validation and test (test AJI: 87–88%, test VIR: 95-96%). Test 2D segmentation errors for U-Net and A-U-Net in iliac bone site ranged (AVE: 7–10%, AHD: 16–18 mm). The 3D iliac models also attained higher AJI (85– 88%) and differed from each other only in VIR by 3% (e.g. test VIR: 90% versus 93%). Comparison of 2D versus 3D posterior ilium segmentations from Table 2, shows that the 3D training, validation and test accuracy were also higher for 2D iliac models (2D AJI: 87–91% versus 3D AJI: 85–88%, 2D VIR: 95–97%, versus 3D VIR: 91–94%).

Tables 1 and 2 allow comparison between the two BM bone-sites. In both bone-sites, the performance of 2D U-Net and 2D A-U-Net were satisfactory and close (2D AJI: 87-91%, 2D VIR: 91-96%). Differences in agreement accuracy between bones indicated more consistent performance in ilium (e.g. 2D iliac VIR: 95–96% versus 2D femur VIR 91-94%). Finally, comparing the 3D model segmentation performance between Tables 1 and 2 to analyze bone-site impact on 3D training, validation and testing data revealed that the AJI was similar for 3D U-Net and 3D A-U-Net in both femur and ilium (3D AJI: 85-88%). 3D models VIR performance declined more in femur as compared to ilium (4% versus 2% decline).

Except for 3D A-U-Net, all ilium segmentation models had a negative AVE bias indicating tendency to over-estimate volume compared to expert segmentations. In contrast, femur model segmentation AVE bias was largely positive consistent with the tendency to underestimate BM volume. The volume contour error (MEDIAN_AHD) for all four U-Net models attained consistent range in both pelvic bones (femur: 5–7 mm, ilium: 5-9 mm). Overall higher AVE-SD and AHD variability in 3D models further confirmed lower performance of 3D models compared to 2D U-Net and 2D A-U-Net (2D AVE SD: 2.5% versus 3D AVE SD: 3.5%, 2D AHD SD: 3–11 mm versus 3D AHD SD: 5-13 mm).

3.3 Metric distribution analysis

Figures 5 and 6 for femur and ilium provided further insight into the distribution shape, mean, and range of three most error-sensitive performance metrics (AJI, AVE and AHD) for the test data with all four selected U-Net models in each bone site.

The distribution's shape and width of violin plots revealed subtle differences in variability and performance consistency, particularly between 2D and 3D models and between U-Net and A-U-Net configurations. Multiple distribution modes were resolved for 3D and 2D model performance metrics for femur, likely reflecting the higher susceptibility for error at the bone margins due to inconsistencies of image volumes.

The comparison of AJI violin plots of femoral bones (Figure 5) and iliac bones (Figure 6) indicated uniformly high performance with maximum density around 90% for all four U-Net models in both bone-sites. Femoral bone metric distributions (Figure 5) showed comparable performance for 2D versus 3D models with marginally narrower distributions (less variability for 2D A-U-Net AJI (range: [85%,95%]). Most of 2D U-Net and A-U-Net AJI had clustered around 86% with just 2-5% narrow error bars. The femur metric distributions are complex and multi-modal likely indicating limitations of small test data set.

The low mean errors for both 2D and 3D U-Net models for AVE demonstrated generally strong agreement with reference volume. Both 2D models performed better than 3D with respect to AVE (2D femoral AVE range: [-20, 20], 2D ilium AVE SD: [-30, 10] versus 3D femoral AVE range: [-20, 35], 3D ilium range: [-25, 30]). Iliac bone metric distributions (Figure 6) followed the similar trends of higher AJI and lower AVE SD with notably narrower AHD, particularly for 2D A-U-Net model.

All four U-Net models AHD distribution in for femur versus ilium (Figures 5, 6) exhibited consistent trend of high data density at higher AHD values (> 20 mm) for femur and at lower AHD (< 18 mm) for ilium, confirming better performance for ilium than femur. The mean values being higher than the maximum width pointed out the presence of outliers, likely for the bone margins.

Figure 7 depicts the results of Wilcoxon rank sum test in the form of grouped bar plot for p-values in femoral bone and iliac BM respectively. Each metric group thus consisted of six bars representing the six unique comparisons, with a total of 18 comparisons for the three metric groups. Notably, non-significant p-values above the threshold were observed for 6 comparisons in femur and 7 in iliac segmentation models. The AJI and AVE metrics were most sensitive to the differences between the models There was no significant difference in 3D A-U-Net versus U-Net performance for iliac and femoral BM segmentation for AJI and AHD. Ilium models also showed that 3D U-Net and 3D A-U-Net are not

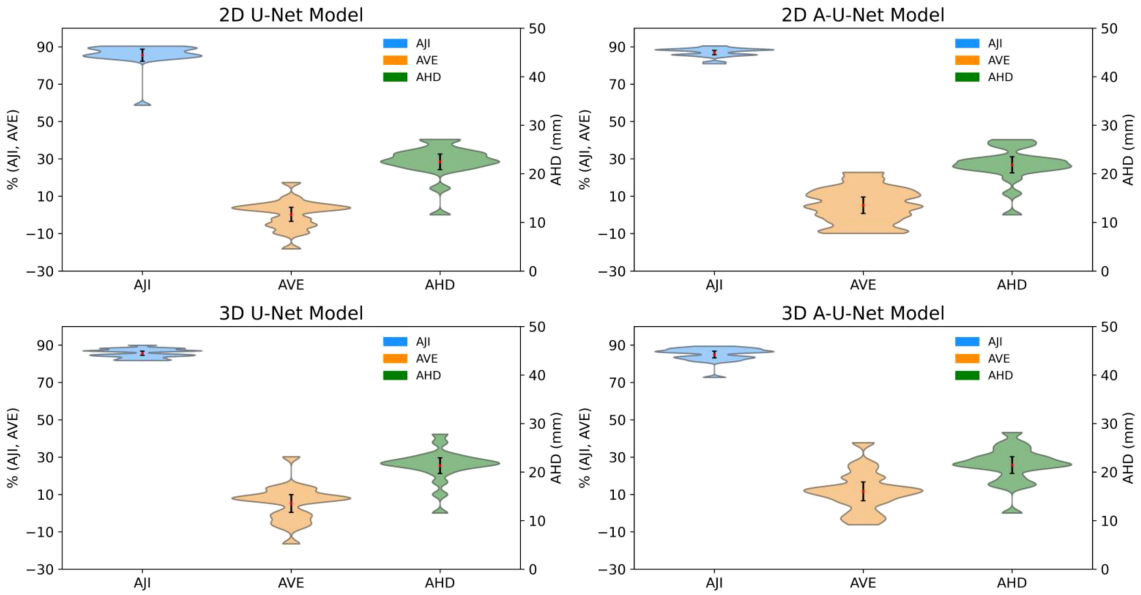


FIGURE 5
Distribution of AJI, AVE and AHD for femoral segmentation test data. Black error bars represent one standard deviation (SD) from the mean while colored triangles correspond to data ranges. The metric distribution is color-coded in the legend. The left axis is shared for AJI and AVE (%), while right axis is for AHD (mm).

significantly different. Except for AHD, the differences between 2D and 3D models were significant. Combined with absolute performance metric values, the significant differences between other models helped identify the overall most effective model for MF bone marrow segmentation as 2D A-U-Net.

Supplementary Table S3 indicated that the proximal femur and posterior ilium (U-Net and A-U-Net) 2D models exhibited stronger performance with F1 scores ranging between 89% and 90%. Close F1-scores for 2D models (~2% difference), indicated that these models achieved a similar balance between precision and recall in each bone-site. This suggested consistent accuracy and reliability in BM segmentation. Iliac 2D U-Nets had lower SD (0.02-0.03) suggesting more consistent performance across different test runs than for femoral bone. In comparison to 2D, the 3D models showed

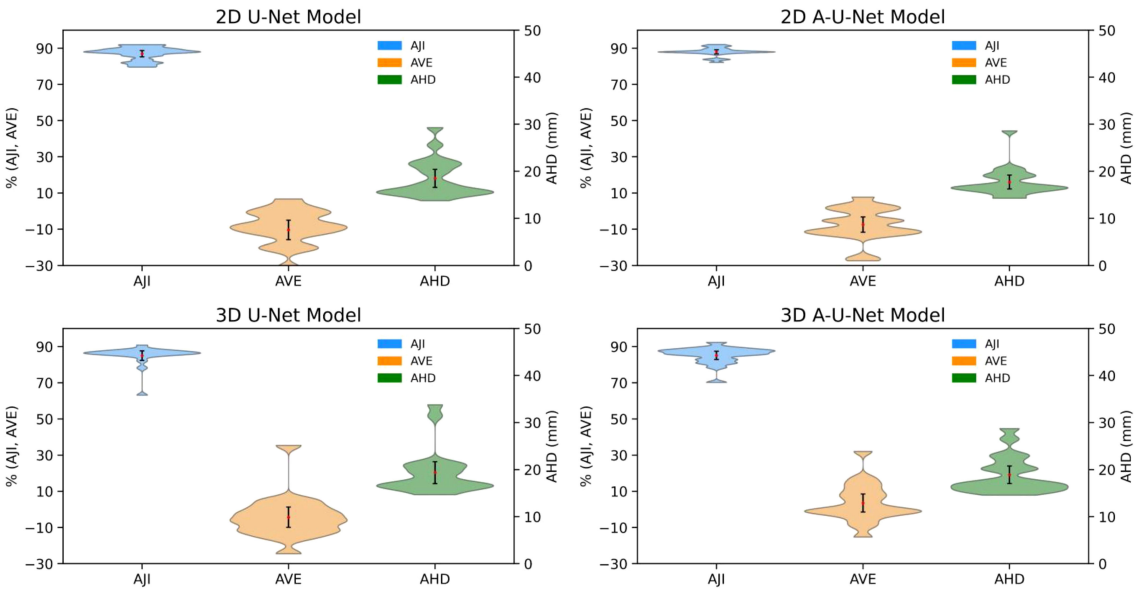


FIGURE 6
Distribution of AJI, AVE and AHD for iliac segmentation test data. Black error bars represent one standard deviation (SD) from the mean while colored triangles correspond to data ranges. The metric distribution is color-coded in the legend. The left axis is shared for AJI and AVE (%), while right axis is for AHD (mm).

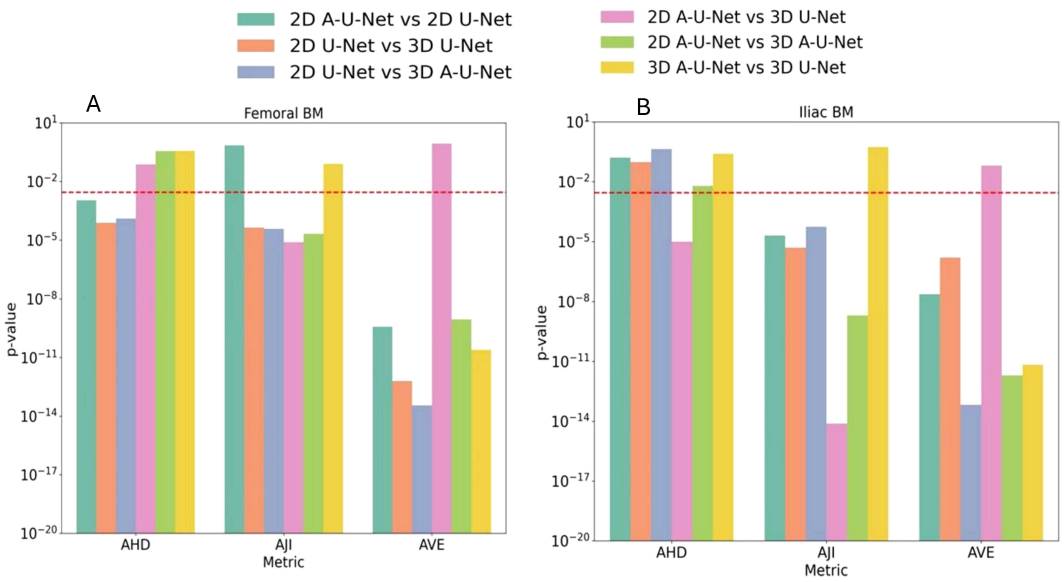


FIGURE 7
(A, B) p-values of pairwise comparison of AHD, AJI and AVE for UNET models (color-coded in the legend) for segmentation performance on a test dataset. A red dashed line in p-value plots represented the Bonferroni-corrected significance threshold p<0.003.

lower F1-score in both bones (Femur: 87% versus 89%, Ilium 87% versus 90%).

Overall, compared to 3D segmentation models, clear performance enhancement was observed for 2D A-U-Net in ilium, while 2D models in femur showed better agreement (AJI, VIR, F1, AVE-SD) but slightly higher contouring errors (AHD) and lower precision (Supplementary Table S3).

Supplementary Figure S4 illustrates Bland-Altman agreement analysis for mean fat fraction (FF) for expert versus preferred DL segmentation model (2D A-U-NET) for the test set. This analysis confirms good agreement (LOA<3%) between FF quantifications for expert versus model segmentations.

4 Discussion

The focus of our study was to find the best U-Net models to automate the BM segmentation in two pelvic bones (proximal femur and posterior ilium) for MF patient Dixon MRI. We developed, validated and compared four distinct U-Net models based on 2D and 3D U-Net and A-U-Net to identify superior models for each pelvic bone site. All eight independently validated U-Net segmentation models in our study showed strong agreement with reference BM segmentations. Overall, the optimized segmentation models in each bone achieved adequate performance and overlap accuracy of boundary delineation for complex and granular segmentation tasks on training, validation and test sets. Our 2D models had higher F1-scores compared to the 3D models (2D: 89-90% versus 3D: 86-88%), good AJI (2D: 87-88% versus 3D: 85%-86%) and excellent VIR (2D: 96% versus 3D: 89-93%). All models demonstrated low segmentation errors and consistently reproduced the reference BM volumes (AVE bias between -11% and 12%). Ilium models were predominantly under-segmenting, while femur models were prone to over-segmentation. Higher variability was observed for 3D AVE versus 2D AVE at both bone sites. Our BM segmentation results suggested that the 2D A-U-Net performed better on average and provided more reliable predictions for both pelvic bone segmentation where the training set contained variable coverage of the femoral shaft for MF patients.

Interestingly, 3D segmentations exhibited comparable performance versus 2D in boundary match (AHD) between models and reference for femur but markedly higher deviations and variability for ilium. The statistical evaluation of femoral and pelvic models reflected that 2D models were more robust than their peer 3D models. Our study further found that AJI was the most stable metric to compare alternative 2D and 3D U-Net BM segmentation models for both bone-sites. Moreover, femoral segmentation showed greater absolute contour mismatches (AHD=26-28 mm) than iliac model segmentations ((AHD=16-20 mm)) reflecting that different bones require training distinct segmentation models. In other words, a DL model trained on a single bone-site would not optimally segment every bone. In our case, the iliac BM voxels occupied larger area with more diverse orientations as compared to femoral BM voxels on individual images but had smaller scanned sub volume (400 versus 735 training images) which is a likely reason for different performance of U-Net models on different bones.

Our best 2D A-U-Net (femur and iliac) segmentation model performed comparable to 3D U-Net lumbar vertebra segmentation model described in the recent quantitative PDFF MRI study (11). The vertebral segmentation study involved 30 healthy training and 12 testing subjects, which were comparable to our training and testing sets (training:32, testing:20). Our validation set AJI achieved substantial gain in both pelvic bones (femur and ilium) with an approximate increase of 8% to 10% compared to reported lumbar

vertebral AJI. Our mean F1-score on the test set were remarkable close to the cited vertebral F1-score (dice similarity) of 86-90%. The precision-recall trend in pelvic bones and vertebral bones also looked similar. However, the vertebral U-Net model training configuration (hyperparameters) and validation methodology did not include tuning and optimization. Our optimized 2D U-Net models also outperformed another Dixon lumbar vertebra segmentation study on two test sets (annotated by different observers) in both AJI and dice similarity with substantial 10% advancement in AJI and 5% increase in F1-score (13). In addition to different U-Net training and optimization approaches, a probable reason for difference in observed segmentation performance could be that vertebral bodies inter slice spatial locations on sagittal images were subject to change while both femur and iliac spatial coordinates remained consistent in all images. Importantly, the achieved agreement with the manually annotated reference for the test set in our study exceeds the reported inter-observer agreement by 15% (13).

Our present study confirmed U-Net models' potential for accurate BM segmentation as found in previous studies. Furthermore, it revealed the utility of IP image training for better model generalization for BM diseases with variable fat content for patient population like MF. Practical considerations about DL aided segmentation workflow optimization for both 2D and 3D U-Net models along with comprehensive segmentation performance evaluation and statistical analysis will also benefit other quantitative MRI studies of bone marrow disease. In addition to automated segmentation, the insights gained by our study include separate U-Net model utility for individual bone sites, empirical evidence of 2D U-Net preference over 3D U-Net for varying patient positioning with small, annotated training set, substantial difference in overlap and boundary performance metrics and practical workflow to find single best DL model for each pelvic bone-site. This segmentation automation will save about 20 minutes of expert time per subject/time point and will likely improve repeatability and remove expert-dependent bias (11, 12). Our previous study in murine model of MF using similar DL segmentation models indicated significantly higher inter-observer reproducibility and test-retest repeatability by automated segmentation (12).

Our study had several limitations. First, this was a single center rare disease (MF patient) study in specific bone-sites (proximal femur and posterior ilium) that may limit the segmentation model's applicability (22, 23). However, this study has the largest patient cohort reported to date for MF rare disease and proposed use of IP images for segmentation, makes this approach less dependent on relative BM fat content. Second, all training, validation and test subsets were acquired on a single scanner with uniform scan protocol which further limits the direct generalization of the models for different acquisition protocols. Small datasets with limited and biased annotations are prevalent in imaging research (24). Our study also utilized a small MF data set with a single expert annotation and did not assess inter-observer reproducibility. Therefore, manual quality check for independent test sets, multi-reader studies and correction of new cases would be required for comprehensive evaluation. Alternatively, model retraining and advancement in segmentation workflow automation would be an option to further reduce the manual annotation efforts. Another limitation was the absence of segmentation ground truth which was partially resolved by training U-Net models (25). Single expert delineations were considered reliable as a reference for our application, although they might not capture all relevant details due to human perception limitations. Even repeated annotations from the same expert may be inconsistent and might limit DL model performance. Furthermore, expert intervention would be required to manually define sub-volumes for the model application to new data at femur and ilium sites, which limits full workflow automation. This step can be streamlined in future applications. Another important technical issue identified during this study was inconsistency of patients positioning during MRI acquisition, resulting in image volume variability affecting the 3D U-Net model performance (26). By their nature, 2D U-Net segmentation models are less sensitive to patient positioning and thus would be preferred for pelvic BM MRI studies.

Future work will apply the best developed models to advance segmentation automation in the ongoing MF studies by leveraging limiting annotated data and extensive un-annotated data. The best DL models will be used to generate pseudo labels for expert review and adjustment. The combined annotated and pseudo-labeled data can be used iteratively to retrain the model, continuously enhancing its accuracy and reliability. This semi-supervised setting may outperform traditional transfer learning and, in some cases, self-supervision. It will also reduce dependency on extensive manual labeling to accelerate the segmentation efficiency and throughput (27, 28). The derived segmentations will be applied to quantitative PDFF maps to measure longitudinal changes in MF bone marrow. We also plan to use these segmentations for transfer to other quantitative MRI contrasts (e.g., ADC) at multiple imaging time points (currently 5–7 time points per patient). The improved automation in segmentation workflow is vital for timely development, validation and implementation of quantitative biomarkers and design therapeutics strategies in MF patients' treatment response.

5 Conclusion

This study developed, validated and tested independent U-Net models to effectively segment BM in quantitative PDFF MRI (IP volumes) of MF patients. The best selected models automatically detected, and segmented BM in pelvic bones (proximal femur and posterior ilium) and showed strong agreement with reference BM segmentations. The comparative analysis revealed that 2D U-Net and 2D A-U-Net out-performed their peer 3D models for BM segmentation in patients' target pelvic bones. 2D A-U-Net model performance was more robust in comparison to other DL models in both femur and ilium. Overall, selected U-Net models demonstrated promising performance to accelerate segmentation and accurately localize BM regions for future patient studies. The

developed BM segmentation models lay the foundation to assess heterogeneity and spatial distribution of the PDFF in pelvic bones which enhances QIB precision. The U-Net based segmentation automates the bone marrow delineation with improved accuracy to facilitate future clinical adoption of MRI for MF patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by University of Michigan institution review board. The scans were performed for consented patients older than 27 years. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

HT: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. LH: Conceptualization, Data curation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Supervision. DM: Conceptualization, Data curation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. MT: Writing – review & editing. KP: Writing – review & editing. GL: Project administration, Resources, Writing – review & editing. BR: Funding acquisition, Project administration, Resources, Writing – review & editing. TC: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the National Institute of Health grant numbers NIH/NCI R01CA190299, NIH/NCI U24CA237683, NIH/NCI R01CA238023, NIH/NCI U01CA166104.

Acknowledgments

Imaging data collection was supported by a clinical trial NCT01973881.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1498832/full#supplementary-material

References

1. Salama ME. Important pathologic considerations for establishing the diagnosis of myelofibrosis. *Hematol Oncol Clin North Am*. (2021) 35:267–78. doi: 10.1016/j.hoc.2020.11.002

2. Tefferi A. Primary myelofibrosis: 2023 update on diagnosis, risk-stratification, and management. *Am J Hematol*. (2023) 98:801–21. doi: 10.1002/ajh.26857

3. Leimkuhler NB, Gleitz HFE, Ronghui L, Snoeren IAM, Fuchs SNR, Nagai JS, et al. Heterogeneous bone-marrow stromal progenitors drive myelofibrosis via a druggable alarmin axis. *Cell Stem Cell*. (2021) 28:637–52 e8. doi: 10.1016/j.stem.2020.11.004

4. Slot S, Lavini C, Zwezerijnen GJC, Boden BJH, Marcus JT, Huisman MC, et al. Characterizing the bone marrow environment in advanced-stage myelofibrosis during ruxolitinib treatment using PET/CT and MRI: A pilot study. *Tomography*. (2023) 9:459–74. doi: 10.3390/tomography9020038

5. Luker GD, Nguyen HM, Hoff BA, Galban CJ, Hernando D, Chenevert TL, et al. A pilot study of quantitative MRI parametric response mapping of bone marrow fat for treatment assessment in myelofibrosis. *Tomography*. (2016) 2:67–78. doi: 10.18383/j.tom.2016.00115

6. Wang F, Qiu T, Wang H, Yang Q. State-of-the-art review on myelofibrosis therapies. *Clin Lymphoma Myeloma Leuk*. (2022) 22:e350–e62. doi: 10.1016/j.clml.2021.11.007

7. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging*. (2019) 49:e101–e21. doi: 10.1002/jmri.26518

8. Robison TH, Lee W, Luker KE, Pettit K, Talpaz M, Chenevert TL, et al. Quantitative MRI reveals heterogeneous impacts of treatment on diseased bone marrow in a mouse model of myelofibrosis. *Magn Reson Med*. (2024) 91:2568–78. doi: 10.1002/mrm.30016

9. Ross BD, Jang Y, Welton A, Bonham CA, Palagama DSW, Heist K, et al. A lymphatic-absorbed multi-targeted kinase inhibitor for myelofibrosis therapy. *Nat Commun*. (2022) 13:4730. doi: 10.1038/s41467-022-32486-8

10. Gianelli U, Vener C, Bossi A, Cortinovis I, Iurlo A, Fracchiolla NS, et al. The European Consensus on grading of bone marrow fibrosis allows a better prognostication of patients with primary myelofibrosis. *Mod Pathol*. (2012) 25:1193–202. doi: 10.1038/modpathol.2012.87

11. Haueise T, Stefan N, Schulz TJ, Schick F, Birkenfeld AL, Machann J. Automated shape-independent assessment of the spatial distribution of proton density fat fraction in vertebral bone marrow. *Z Med Phys*. (2024) 34:436–45. doi: 10.1016/j.zemedi.2022.12.004

12. Kushwaha A, Mourad RF, Heist K, Tariq H, Chan HP, Ross BD, et al. Improved repeatability of mouse tibia volume segmentation in murine myelofibrosis model using deep learning. *Tomography*. (2023) 9:589–602. doi: 10.3390/tomography9020048

13. Zhou J, Damasceno PF, ChaChad R, Cheung JR, Ballatori A, Lotz JC, et al. Automatic vertebral body segmentation based on deep learning of dixon images for bone marrow fat fraction quantification. *Front Endocrinol (Lausanne)*. (2020) 11:612. doi: 10.3389/fendo.2020.00612

14. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. *Adv Exp Med Biol*. (2020) 1213:3–21. doi: 10.1007/978-3-030-33128-3_1

15. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O eds. *3D U-net: Learning Dense Volumetric Segmentation from Sparse Annotation*. Cham: Springer International Publishing (2016).

16. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sc*. (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28

17. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. *Med Image Anal*. (2019) 53:197–207. doi: 10.1016/j.media.2019.01.012

18. Muller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*. (2022) 15:210. doi: 10.1186/s13104-022-06096-y

19. Way TW, Hadjiiski LM, Sahiner B, Chan HP, Cascade PN, Kazerooni EA, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. *Med Phys*. (2006) 33:2323–37. doi: 10.1118/1.2207129

20. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep-Uk*. (2022) 12. doi: 10.1038/s41598-022-09954-8

21. Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J Med Imaging (Bellingham)*. (2018) 5:015006. doi: 10.1117/1.JMI.5.1.015006

22. Hadjiiski L, Cha K, Chan HP, Drukker K, Morra L, Nappi JJ, et al. AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging. *Med Phys*. (2023) 50:e1–e24. doi: 10.1002/mp.16188

23. Mali SA, Ibrahim A, Woodruff HC, Andrearczyk V, Muller H, Primakov S, et al. Making radiomics more reproducible across scanner and imaging protocol variations: A review of harmonization methods. *J Pers Med*. (2021) 11:842. doi: 10.3390/jpm11090842

24. Candemir S, Nguyen XV, Folio LR, Prevedello LM. Training strategies for radiology deep learning models in data-limited scenarios. *Radiol Artif Intell*. (2021) 3:e210014. doi: 10.1148/ryai.2021210014

25. Azad R, Aghdam EK, Rauland A, Jia Y, Avval AH, Bozorgpour A, et al. Medical image segmentation review: the success of U-net. *IEEE Trans Pattern Anal Mach Intell*. (2024) 46:10076–95. doi: 10.1109/TPAMI.2024.3435571

26. Morris DM, Wang C, Papanastasiou G, Gray CD, Xu W, Sjostrom S, et al. A novel deep learning method for large-scale analysis of bone marrow adiposity using UK Biobank Dixon MRI data. *Comput Struct Biotechnol J*. (2024) 24:89–104. doi: 10.1016/j.csbj.2023.12.029

27. Kenney HM, Peng Y, Chen KL, Ajalik R, Schnur L, Wood RW, et al. A high-throughput semi-automated bone segmentation workflow for murine hindpaw micro-CT datasets. *Bone Rep*. (2022) 16. doi: 10.1016/j.bonr.2022.101167

28. Xu M, Zhou Y, Jin C, de Groot M, Alexander DC, Oxtoby NP, et al. Expectation maximisation pseudo labels. *Med Image Anal*. (2024) 94:103125. doi: 10.1016/j.media.2024.103125

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 18 June 2025
DOI 10.3389/fonc.2025.1480384

Check for updates

OPEN ACCESS

EDITED BY
Eric Edward Sigmund,
New York University, United States

REVIEWED BY
Radu Hristu,
National University of Science and Technology POLITEHNICA Bucharest, Romania
Muge Karaman,
University of Illinois Chicago, United States

*CORRESPONDENCE
Gabriel Kalweit
gabriel.kalweit@criion.org

RECEIVED 13 August 2024
ACCEPTED 21 May 2025
PUBLISHED 18 June 2025

CITATION
Kalweit G, Klett A, Silvestrini P, Rahnfeld J, Naouar M, Vogt Y, Infante D, Berger R, Duque-Afonso J, Hartmann TN, Follo M, Bodurova-Spassova E, Lübbert M, Mertelsmann R, Boedecker J, Ullrich E and Kalweit M (2025) Leveraging a foundation model zoo for cell similarity search in oncological microscopy across devices. *Front. Oncol.* 15:1480384. doi: 10.3389/fonc.2025.1480384

COPYRIGHT
© 2025 Kalweit, Klett, Silvestrini, Rahnfeld, Naouar, Vogt, Infante, Berger, Duque-Afonso, Hartmann, Follo, Bodurova-Spassova, Lübbert, Mertelsmann, Boedecker, Ullrich and Kalweit. 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.

Leveraging a foundation model zoo for cell similarity search in oncological microscopy across devices

Gabriel Kalweit[1,2]*, Anusha Klett[1], Paula Silvestrini[1,3], Jens Rahnfeld[1,2], Mehdi Naouar[1,2], Yannick Vogt[1,2], Diana Infante[1,4], Rebecca Berger[4], Jesús Duque-Afonso[4], Tanja Nicole Hartmann[4], Marie Follo[4,5], Elitsa Bodurova-Spassova[4,5], Michael Lübbert[4,6], Roland Mertelsmann[1,4,7], Joschka Boedecker[1,2,8], Evelyn Ullrich[1,9,10] and Maria Kalweit[1,2]

[1]Collaborative Research Institute Intelligent Oncology (CRIION), Freiburg, Germany, [2]Neurorobotics Lab, Department of Computer Science, University of Freiburg, Freiburg, Germany, [3]Laboratory of Applied Cellular and Molecular Biology, Institute of Veterinary Sciences of the Litoral, Universidad Nacional del Litoral (UNL) - National Scientific and Technical Research Council (CONICET), Esperanza, Argentina, [4]Department of Medicine I, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany, [5]Lighthouse Core Facility, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany, [6]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Freiburg, Freiburg, Germany, [7]Mertelsmann Foundation, Freiburg, Germany, [8]Intelligent Machine-Brain Interfacing Technology (IMBIT), BrainLinks-BrainTools, University of Freiburg, Freiburg, Germany, [9]Department of Pediatrics, Experimental Immunology and Cell Therapy, Goethe University Frankfurt, Frankfurt am Main, Germany, [10]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Frankfurt, Frankfurt am Main, Germany

Background: Cellular imaging analysis using the traditional retrospective approach is extremely time-consuming and labor-intensive. Although AI-based solutions are available, these approaches rely heavily on supervised learning techniques that require high quality, large labeled datasets from the same microscope to be reliable. In addition, primary patient samples are often heterogeneous cell populations and need to be stained to distinguish the cellular subsets. The resulting imaging data is analyzed and labeled manually by experts. Therefore, a method to distinguish cell populations across imaging devices without the need for staining and extensive manual labeling would help immensely to gain real-time insights into cell population dynamics. This especially holds true for recognizing specific cell types and states in response to treatments.

Objective: We aim to develop an unsupervised approach using general vision foundation models trained on diverse and extensive imaging datasets to extract rich visual features for cell-analysis across devices, including both stained and unstained live cells. Our method, Entropy-guided Weighted Combinational FAISS (EWC-FAISS), uses these models purely in an inference-only mode without task-specific retraining on the cellular data. Combining the generated embeddings in an efficient and adaptive k-nearest neighbor search allows for automated, cross

device identification of cell types and states, providing a strong basis for AI-assisted cancer therapy.

Methods: We utilized two publicly available datasets. The WBC dataset includes 14,424 images of stained white blood cell samples from patients with acute myeloid and lymphoid leukemia, as well as those without leukemic pathology. The LISC dataset comprises 257 images of white blood cell samples from healthy individuals. We generated four in-house datasets utilizing the JIMT-1 breast cancer cell line, as well as Jurkat and K562 (leukemic cell lines). These datasets were acquired using the Nanolive 3D Cell Explorer-fluo (CX-A) holotomographic microscope and the BioTek Lionheart FX automated brightfield microscope. The images from the in-house datasets were manually annotated using Roboflow software. To generate the embeddings, we used and optimized a concatenated combination of SAM, DINO, ConvNeXT, SWIN, CLIP and ViTMAE. The combined embeddings were used as input for the adaptive k-nearest neighbor search, building an approximate Hierarchical Navigable Small World FAISS index. We compared EWC-FAISS to fully fined-tuned ViT-Classifiers with DINO-, and SWIN-backbones, a ConvNeXT architecture, as well as to NMTune as a lightweight domain-adaptation method with frozen backbone.

Results: EWC-FAISS performed competitively with the baselines on the original datasets in terms of macro accuracy. Macro accuracy is the average of class-specific accuracies, treating all classes equally by averaging their individual accuracies. EWC-FAISS ranked second for the WBC dataset (macro accuracy: 97.6 $\pm$ 0.2), first for cell state classification from Nanolive (macro accuracy: 90 $\pm$ 0), and performed comparably for cell type classification from Lionheart (macro accuracy: 87 $\pm$ 0). For the transfer to out-of-distribution (OOD) datasets, which the model had not seen during training, EWC-FAISS consistently outperformed the other baselines. For the LISC dataset, EWC-FAISS achieved a macro accuracy of 78.5 $\pm$ 0.3, compared to DINO FT's 17 $\pm$ 1, SWIN FT's 44 $\pm$ 14, ConvNeXT FT's 45 $\pm$ 9, and NMTune's 52 $\pm$ 10. For the cell state classification from Lionheart, EWC-FAISS had a macro accuracy of 86 $\pm$ 1, while DINO FT, SWIN FT, and ConvNeXT FT achieved 65 $\pm$ 11, 68 $\pm$ 16, and 81 $\pm$ 1, respectively, and NMTune 81 $\pm$ 7. For the transfer of cell type classification from Nanolive, EWC-FAISS attained a macro accuracy of 85 $\pm$ 0, compared to DINO FT's 24.5 $\pm$ 0.9, SWIN FT's 57 $\pm$ 6, ConvNeXT FT's 54 $\pm$ 4, and NMTune's 63 $\pm$ 4. Additionally, building EWC-FAISS after embedding generation was significantly faster than training DINO FT ($\sim$ 6 minutes compared to > 10 hours). Lastly, EWC-FAISS performed comparably in distinguishing cancerous cell lines from Peripheral Blood Mononuclear Cells with a mean accuracy of 80 $\pm$ 5, compared to CellMixer with a mean accuracy of 79.7.

Conclusion: We present a novel approach to identify various cell lines and primary cells based on their identity and state using images acquired across various imaging platforms which vary in resolution, magnification and image quality. Despite these differences, we could show that our efficient, adaptive k-nearest neighbor search pipeline can be applied on a large image dataset containing different cell types and effectively differentiate between the cells and their states such as live, apoptotic or necrotic. There are several applications, particularly in distinguishing various cell populations in patient samples or monitoring therapy.

KEYWORDS

artificial intelligence, deep learning, foundation models, nearest neighbor search, cell imaging

Highlights

- Foundation models, even without domain-specific training, provide effective discriminative features for both stained and unstained cellular imaging. EWC-FAISS as an adaptive k-nearest neighbor search on embeddings generated from a combination of foundation models achieved high accuracy across various datasets.
- General features from a combination of foundation models demonstrated superior transferability to new experimental settings, including both stained and unstained cells across different recording devices.
- Fast inference with EWC-FAISS allows for quick development cycles, supporting efficient differentiation of cell types and states, crucial for AI-assisted cancer therapy.
- A combination of DINO, ConvNeXT and SWIN proved to be a general, robust and versatile combination leading to good transfer performances in cell state and type classification.

1 Introduction

In medical and biological research, data acquisition is often challenging and involves high costs and labor-intensive processes (1). This is especially true in cellular imaging analysis, where traditional approaches rely heavily on supervised learning techniques that require high-quality, large-scale labeled datasets, which are expensive and time-consuming to produce. Additionally, the heterogeneity of imaging equipment (e.g., different microscopes) and protocols (e.g., varying media and lighting conditions) introduces variability, complicating the task and degrading the performance of narrowly trained models. Training these models demands costly GPUs, extensive training time, and frequent retraining for new tasks. Addressing these challenges necessitates methodologies that leverage existing data more efficiently and generalize across diverse imaging conditions without extensive retraining or fine-tuning.

Recent advancements in machine learning, particularly in the development of *general* foundation models (2), present a promising solution. Particularly models like DINO (3, 4) and Segment Anything (SAM) (5), have significantly influenced medical and cellular image processing domains. MedSAM (6) has extended the utility of SAM to general medical imaging tasks, while models like UNI (7), WTC-11 DINO (8), DINOBloom (9) and scDINO (10) have adapted DINO-style approaches to histopathology and (multi-channel) cellular image analysis. The scDINO (10) model demonstrated that a k-nearest neighbor (k-NN) search using DINO features, fine-tuned and adapted to multi-channel cellular imaging, can be competitive with other methods for cell classification tasks. Israel et al. (11) introduced with CellSAM an adaptation of SAM specifically designed for cell segmentation. Also, self-supervised masked autoencoders have been shown to be capable of capturing cellular biology when trained on massive datasets (12). Despite these advancements, training foundation models specifically for medical applications often requires substantial computational resources (6, 7, 12), limiting accessibility for multiple iterations during model development. In their work, Doron et al. (8) showed that DINO features could predict expert-defined cellular phenotypes, enhance the prediction of compound bioactivity, and facilitate unbiased profiling of cellular morphology. However, this study also revealed that ImageNet features *can* generalize in some settings more effectively than fine-tuned models in the cellular domain, especially in (rather) low-data regimes. Generally, foundation models, characterized by their vast scale and versatility, are pre-trained on a variety of abstract objectives, enabling them to capture a wide array of features applicable across domains.

DINO leverages self-distillation, allowing the model to *teach* itself by comparing different versions of the same image. SAM focuses on segmentation, learning to identify specific objects within an image based on prompts such as points and bounding boxes. SWIN (13, 14) builds a layered understanding of the image through hierarchical feature maps and directs its attention to specific regions using a *shifted window* approach. ConvNeXT (15, 16) rethinks the traditional convolutional neural network architecture. CLIP (17) learns to associate image content with natural language descriptions. Finally, ViTMAE (18) employs a masked autoencoder technique, hiding parts of the image and tasking the model to reconstruct them. These diverse objectives and architectures enable these models to extract complementary and orthogonal information from the data, potentially leading to better generalization on unseen data outside the training distribution. In line with the *Platonic Representation Hypothesis* (19), we believe this makes them particularly suitable for tasks like cellular imaging analysis, where acquiring large amounts of labeled data can be challenging.

Following this rationale, this study explores the utility of various foundation models without fine-tuning for the task of cellular imaging analysis. In this context, we developed an automated pipeline, *Entropy-guided Weighted Combinational FAISS* (EWC-FAISS), combining different foundation models as pre-trained feature-extractors to generate concatenated embeddings, which are then used to build an approximate Hierarchical Navigable Small World (HNSW) (20) FAISS index (21) for an efficient k-NN search (see Figure 1). To enhance robustness, we propose an entropy-based search for the optimal number of neighbors at runtime, and to alleviate class imbalance through distribution reweighting. Recent research has investigated how to best select foundation models and hyperparameters for cost-efficient fine-tuning for the task at hand (22) and how to make the general features learned from foundation models more robust for downstream tasks via covariance and dominant singular value regularization (23). Our proposed approach stands orthogonal to this line of research by leveraging a combination of features from various foundation models as feature extractors, even when trained on non-domain specific data. This methodology aims to achieve better generalization and adaptability in cellular imaging analysis without any fine-tuning typically required. By building a FAISS index, model iteration can be executed much faster compared to

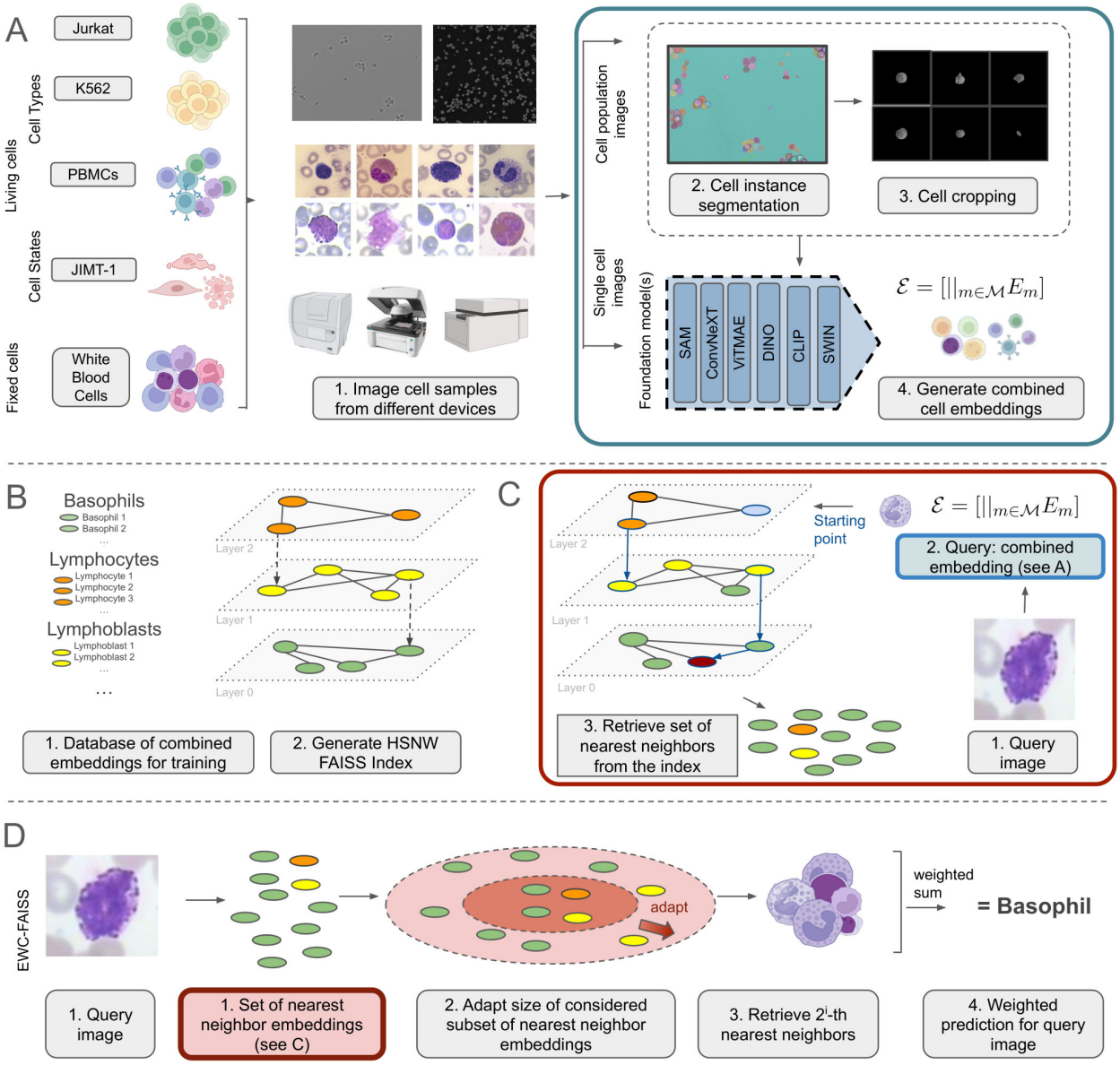


FIGURE 1
(A) Unsupervised cell embedding generation from cell images using a combination of foundation models, with embeddings generated in pure inference-only mode without task-specific fine-tuning. **(B)** Generation of training database and HNSW FAISS index from cell embeddings. **(C)** Retrieving the set of nearest neighbors from the index for a new query image. **(D)** Inference of EWC-FAISS for the prediction of a new image.

training a full parameterized classifier while still being able to benefit from the generalization capabilities of sophisticated feature extractors (cf. Figure 2). We were able to demonstrate, that combining features from multiple foundation models, trained on natural images, can outperform single-model approaches, including DINO, in terms of performance and transferability.

Finally, in this work we successfully validated the effectiveness of our approach in multiple scenarios. First, we performed an evaluation on the WBC dataset (24) containing stained blood cell smears, as well as a transfer to the LISC dataset (25). Second, we conducted an analysis of EWC-FAISS on live cell state classification, with a transfer from the Nanolive CX-A to the BioTek Lionheart FX microscopes. Third, we challenged our approach on live cell type classification, with a transfer from the BioTek Lionheart FX to the Nanolive CX-A. Lastly, we evaluated NMTune (23) in these domains.

2 Materials and methods

2.1 Data

We utilized two publicly available datasets, i.e. WBC (24) and LISC (25), and created four new datasets (CELL DEATH NANOLIVE, CELL DEATH LIONHEART, CELL TYPE NANOLIVE and CELL TYPE LIONHEART, cf. Figure 3). For all datasets, external and in-house, the labeling was performed by a

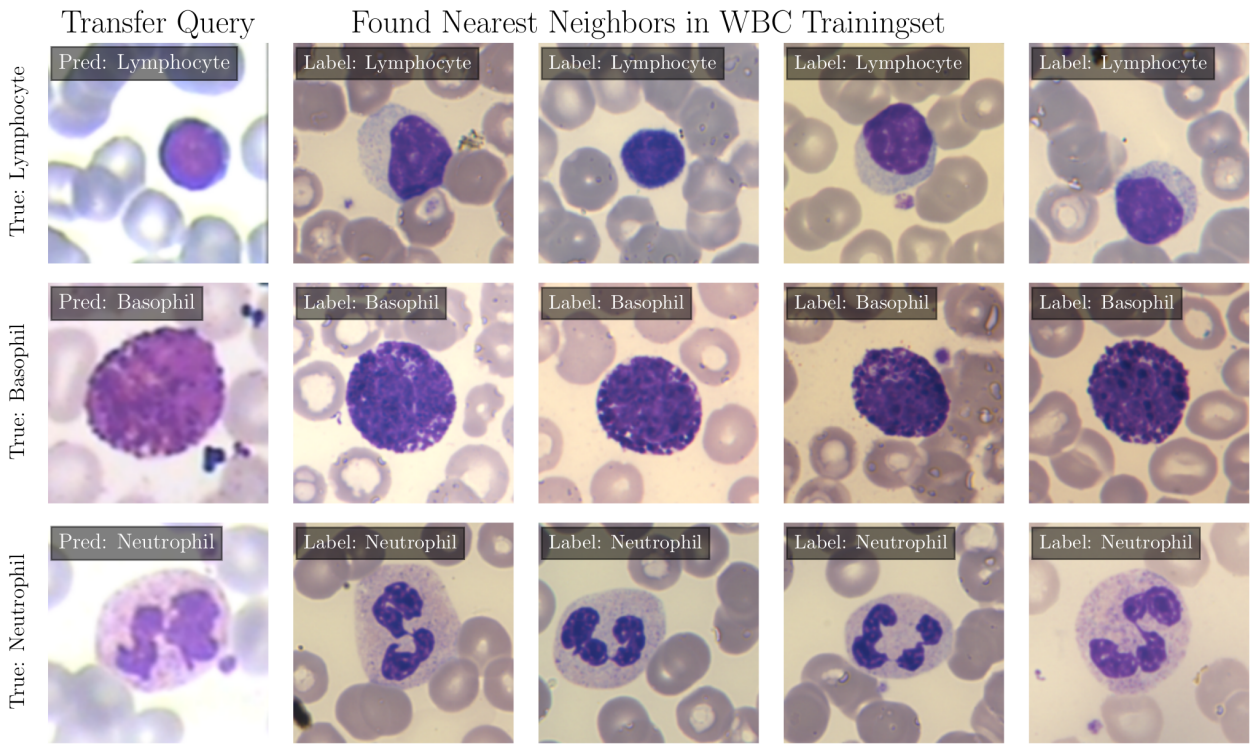


FIGURE 2
Prediction of EWC-FAISS on query images from the LISC dataset, the four nearest neighbors in the WBC dataset and the ground truth labels.

human domain expert (24, 25). The distribution can be found in Figure 3.

2.1.1 Stained white blood cells

The WBC (24) dataset includes 14,424 cell images from microscopic blood smear images from 36 leukemic and 45 non-leukemic peripheral blood smears, collected from 78 anonymized patients. This cohort includes 18 patients with acute myeloid leukemia, 15 with acute lymphoid leukemia, and 45 with no leukemic pathology. Blood smears were stained using May-Grünwald and Giemsa-Romanowski solutions, and blast cell lineage was determined by flow cytometry. Images were captured using an Olympus BX51 brightfield microscope with a Basler acA5472-17uc camera, achieving a resolution of approximately 42 pixels per 1μm under a magnification of 100×. The dataset contains nine different annotated blood cell types: neutrophil segments and neutrophil bands (3300), eosinophils (1017), basophils (1023), lymphocytes (3046), monocytes (2040), normoblasts (510), myeloblasts (2534), and lymphoblasts (2557). Due to the low number of neutrophil bands, we have merged them with the neutrophil segments. The LISC (25) dataset contains hematological images from peripheral blood of 8 healthy individuals, resulting in 257 white blood cell images from 100 microscope slides. These slides were stained using the GismoRight technique, imaged with a Zeiss Axioskop 40 brightfield microscope at 100× magnification, and recorded by a Sony SSCDC50AP digital camera in BMP format. We converted the images to grayscale. Each image was collected from the Hematology-Oncology and BMT Research Center at Imam Khomeini Hospital, Tehran. A hematologist classified the images into five normal leukocyte categories: basophil (55), eosinophil (39), lymphocyte (59), monocyte (48), and neutrophil (56).

2.1.2 Live cell state imaging

The CELL DEATH NANOLIVE dataset comprises 7,420 images of JIMT-1 cells, captured at 60× magnification using a Nanolive CX-A microscope. The microscope generates several cross sections which are combined to generate a high-resolution 3D holotomographic projection and can be represented as a 2D maximum intensity projection. The dataset is categorized into Dead (728), Living (4,613), Apoptotic (707), and Necrotic (1,372) cells. An additional test set includes 1,122 images, with Dead (255), Living (500), Apoptotic (99), and Necrotic (268) cells. The images are 2D projections of 3D volumes. JIMT-1 cells were cultivated using Dulbecco's modified eagle medium (DMEM) FluoroBrite (Gibco), supplemented with 10% fetal bovine serum (FBS; Gibco), 1x L-glutamine (Gibco), and 1% Pen Strep (10,000 Units/ml penicillin, 10,000 µg/ml streptomycin; Gibco). JIMT-1 cells were either non treated or treated with 2.5 µM, 5 µM and 10 µM of Raptinal (Sigma Aldrich) for 24h to induce cell death. 300 nM of Biotracker Apo15 (Sigma) was used as a fluorescence marker to detect early stage apoptosis. The cells were seeded in a µ dish 35 mm, low glass bottom (Ibidi). Brightfield Images were acquired every 15 min and fluorescence every 3h. For each image, we used

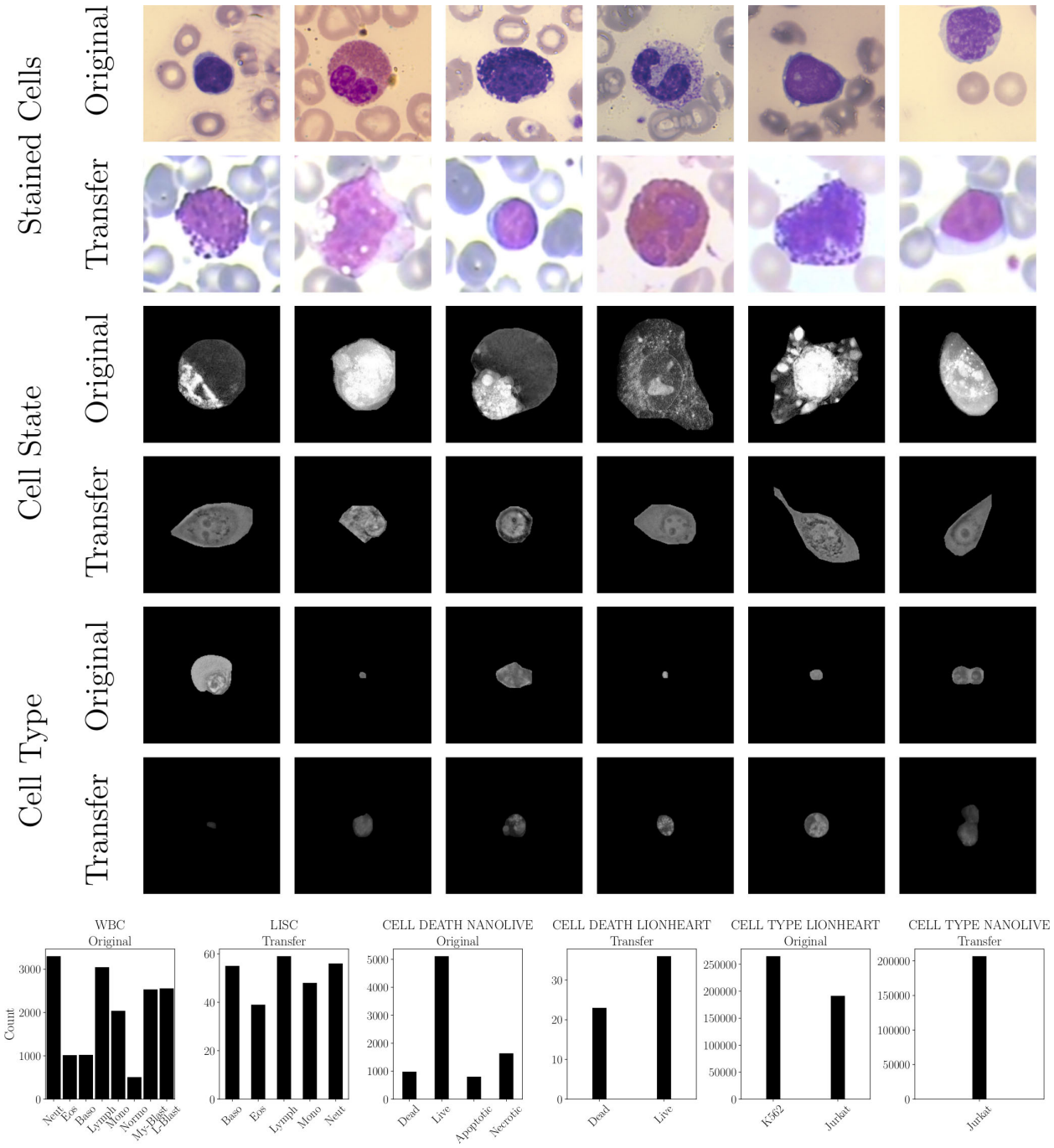


FIGURE 3
Used datasets: stained white blood cells of the WBC (Original) and LISC dataset (Transfer), live cell state recorded by the Nanolive (Original, CELL DEATH NANOLIVE) and BioTek Lionheart FX microscopes (Transfer, CELL DEATH LIONHEART), and two cell lines recorded by BioTek Lionheart FX (Original, CELL TYPE LIONHEART) and Nanolive (Transfer, CELL TYPE NANOLIVE). (top) Exemplary images. (bottom) Class distributions of all datasets.

contrast limited adaptive histogram equalization (CLAHE) to normalize its contrast. We used Roboflow for annotations. The CELL DEATH LIONHEART dataset contains 59 annotated test images at 20×, categorized into dead (23) and living cells (36). The breast carcinoma cell line JIMT-1 (ACC 589, DSMZ) was used as adherent cells. JIMT-1 cells were cultivated using Dulbecco's modified eagle medium (DMEM) (Gibco), supplemented with 10% fetal bovine serum (FBS; Gibco), 1x L-glutamine (Gibco), and 1% Pen Strep (10,000 Units/ml penicillin, 10,000 µg/ml streptomycin; Gibco). Cells were incubated at 37°C in a humidified atmosphere containing 5% CO2 and passaged twice a week. For treatment, JIMT-1 cells were seeded in an 8-well chip (Ibidi) and either left untreated or treated with 25 µM of Etoposide (Sigma Aldrich) for 72 hours. Propidium Iodide (0.25 µg/ml, Sigma Aldrich) was used as a fluorescent marker to stain dead cells. Brightfield and fluorescence images were acquired every 2 hours

using a BioTek Lionheart FX automated brightfield microscope. Examples can be found in Figure 4.

2.1.3 Live cell type imaging

The CELL TYPE LIONHEART dataset includes 456,366 images of homogeneous K562 (264,904) and Jurkat cell images (191,462), captured using a BioTek Lionheart FX automated brightfield microscope at 20× magnification. The CELL TYPE NANOLIVE dataset consists of 206,742 images of Jurkat cells captured with the Nanolive CX-A holotomographic microscope at 60×. Each image was segmented using SAM, and the type of each crop was assigned accordingly. Contrast normalization was applied to each image using CLAHE. Jurkat and K562 cells were cultivated in RPMI 1640 medium (Gibco), supplemented with Advanced RPMI, 4% FBS, 1x L-glutamine, and 1% Pen Strep. All cells were incubated at 37°C in a humidified atmosphere containing 5% CO2 and passaged twice a week. Peripheral Blood Mononuclear Cells or PBMCs were isolated from LRS chambers obtained from healthy donors from the Institut fur¨ Transfusionsmedizin und Gentherapie, Medical Center - University of Freiburg. Contents of the LRS chambers were diluted in Histopaque®-1077 (Sigma Aldrich) and centrifuged according to the manufacturer's protocol. The isolated PBMCs were counted and viability was estimated using Trypan blue exclusion dye using a Neubauer chamber. PBMCs were maintained in RPMI 1640 medium (Gibco) supplemented with 10% FBS, 1x L-glutamine and 1% Pen Strep. PBMCs were seeded in an 18 well chip (Ibidi) either separately or mixed with Jurkat and K562 cell lines at various ratios (25:75, 75:25 and 50:50) and brightfield images were acquired using Biotek Lionheart automated microscope.

Unless single-cell images were already available, cell images were segmented and masked using SAM, where we cropped 224 × 224 pixel crops around the centers of the found cell masks. Examples can be found in Figure 4. We split the original datasets to 90% train, 9% validation, and 1% test sets and used the transfer datasets only for evaluation.

2.1.4 Magnification adaption

We adapted the resolution between Lionheart and Nanolive microscopes. The Lionheart microscope, operating at 20× with a field of view of 291x394µm^2, generates images of 904x1224 pixels, corresponding to approximately 0.322µm/pixel. In contrast, the Nanolive microscope, set at 60× with a field of view of 85x85µm^2, produces images of 448x448 pixels, resulting in a resolution of about 0.19µm/pixel. To match the resolution of the images from the Nanolive microscope to those from the Lionheart microscope, a scaling factor of 0.59 was applied, calculated by dividing the Nanolive pixel size by the Lionheart pixel size. Hence, the Nanolive images were downscaled to match the pixel size of the Lionheart.

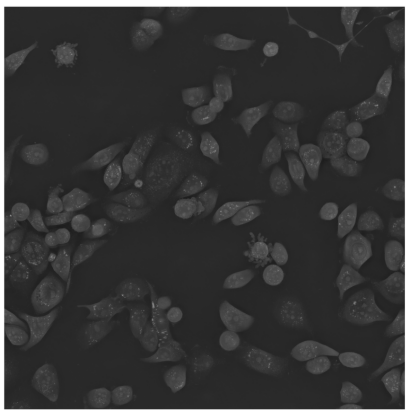


Figure 4a.

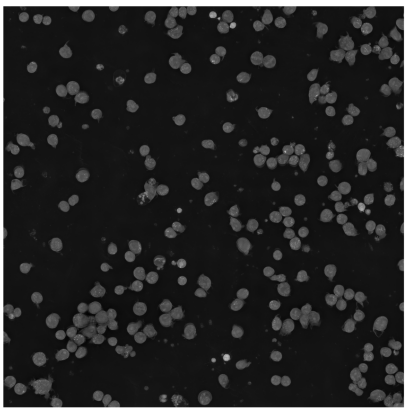


Figure 4b.

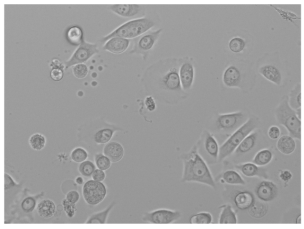


Figure 4c.

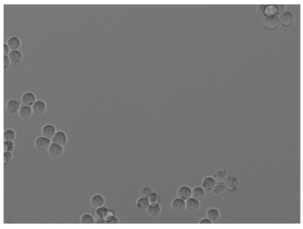


Figure 4d.

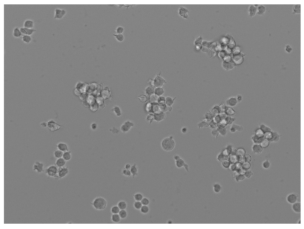


Figure 4e.

FIGURE 4
2D maximum intensity projection of the 3D holotomographic Nanolive microscope and images of the brightfield Lionheart microscope showing various cell types. **(a)** CELL DEATH NANOLIVE image of JIMT-1 breast cancer cells. **(b)** CELL TYPE NANOLIVE image of Jurkat cells. **(c)** CELL DEATH LIONHEART image of JIMT-1 cells. **(d)** CELL TYPE LIONHEART image of K562 cells. **(e)** CELL TYPE LIONHEART image of Jurkat cells.

2.2 Models

For DINO, we used facebook/dinov2-giant, for ConvNeXT we used facebook/convnextv2-large-22k-224, for SWIN we used microsoft/swinv2-large-patch4window12-192-22k, for CLIP we used openai/clip-vit-large-patch14 and for ViTMAE we used facebook/vit-mae-huge, with their respective AutoImageProcessor from the HuggingFace Transformers Library (26). For SAM, we used samvith4b8939 from the official implementation at https://github.com/facebookresearch/segment-anything.

2.3 EWC-FAISS

Next, we outline how to build, train, and query a latent embeddings database using EWC-FAISS.

2.3.1 Foundation model embedding generation

We utilize a set of foundation models as encoders to generate embeddings for our data. Each foundation model M_i in the set $\{M_1, M_2, \ldots, M_n\}$ processes the input data X to produce a corresponding embedding $E_i = M_i(X)$. Fora given encoder subset $\mathcal{M} \subseteq \{M_1, M_2, \ldots, M_n\}$, we concatenate the embeddings E_i from each encoder $M_i \in \mathcal{M}$ to form a full feature representation $\mathcal{E} = [||_{m \in \mathcal{M}} E_m]$. This concatenated embedding $\mathcal{E}$ serves as the input for subsequent tasks.

2.3.2 Database and FAISS index construction

We construct a database $\mathcal{D}$ consisting of embeddings $\{E_i|_{i=1}^{n}\}_c$ and labels for each cell c (cf. Figures 5, 6 (top)). This results in $\mathcal{D} = \left\{E_i|_{i=1}^{n}\right\}_{c=1}^{N}$, where N is the total number of cells. $\mathcal{D}$ is then used to train a HNSW FAISS index (20) on concatenated embeddings $\mathcal{E}_c$.

2.3.3 Class weight and entropy calculation

To address the issue of class imbalance in our training data, class weights were calculated based on the frequency of each class once after adding the embeddings to the index. The total number of samples was divided by the product of the number of classes and the count of each class. The fixed class weight for class i is given by

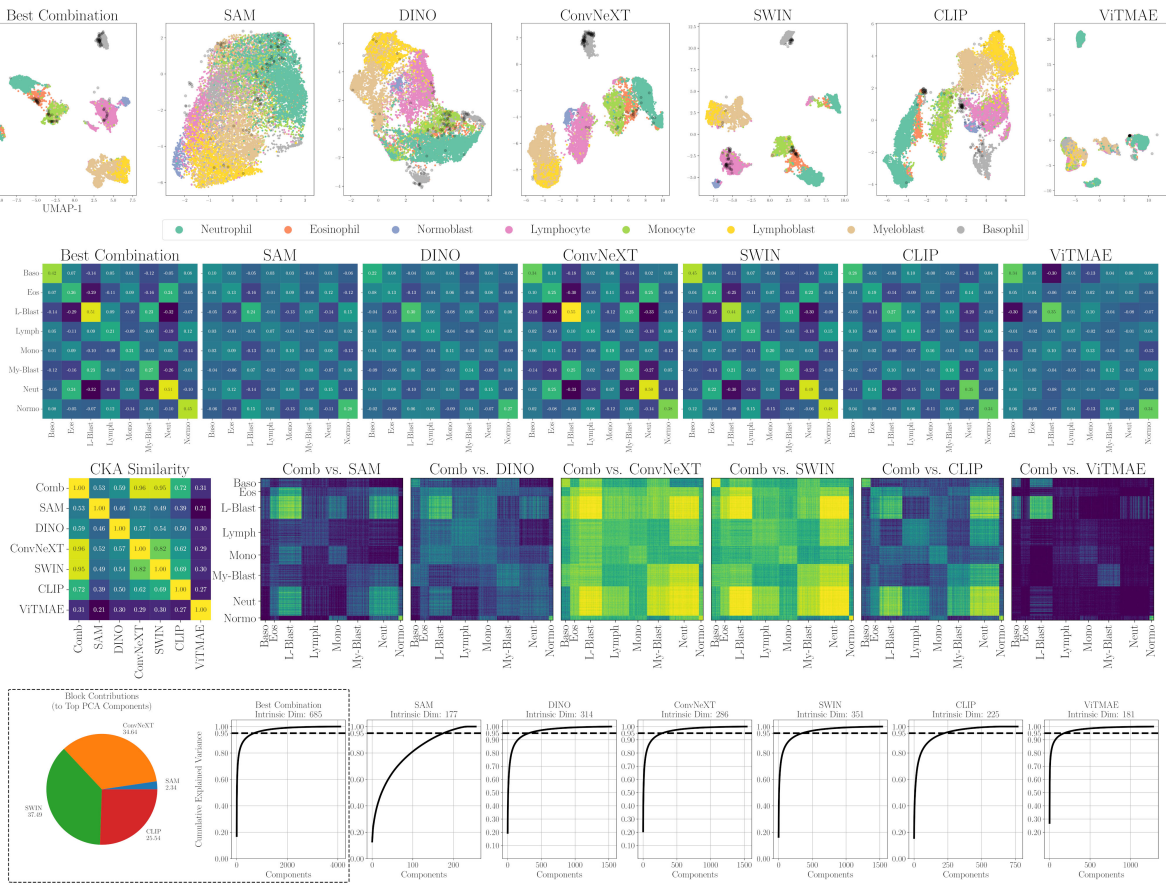


FIGURE 5
Alignment, similarity and dimensionality analysis of foundation model embeddings for WBC. (first row) 2D projections of individual encoders and SAM +ConvNeXT+SWIN+CLIP (Best Combination). Samples from LISC in black. (second row) Mean cosine similarity scores per cell type or state. (third row) RBF CKA similarity matrix and pairwise comparisons of the combined embedding (Comb) to individual models. Blue indicates low similarity, yellow high similarity. (fourth row) PCA-based intrinsic dimensionality plots, showing the number of components required to explain 95% of the variance, and intra-dataset block contributions of individual encoders to the top PCA components.

10.3389/fonc.2025.1480384

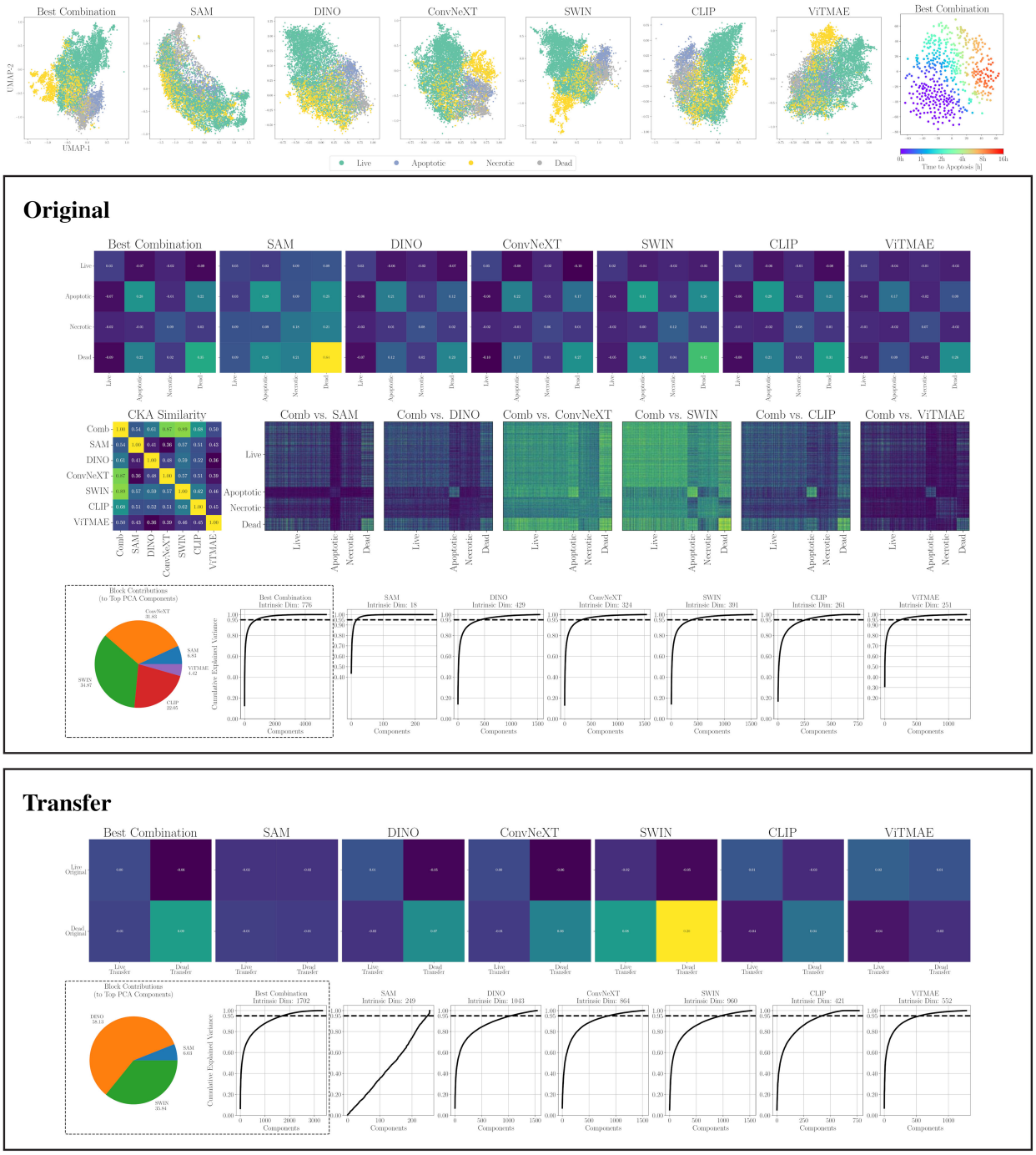


FIGURE 6
(top) 2D projections of individual encoders and the best combination for CELL DEATH NANOLIVE (left). Embeddings from 15 cells tracked until apoptosis. Color indicates time to apoptosis smoothed by nearest neighbors (right). (middle) Similarity and dimensionality analysis for CELL DEATH NANOLIVE: mean cosine similarity scores (first row); RBF CKA matrix and pairwise comparisons of the combined embedding (Comb) to individual models (second row; blue = low, yellow = high); PCA plots showing components explaining 95% variance and intra-dataset encoder contributions to top components (fourth row). (bottom) Similarity and dimensionality analysis for transfer to CELL DEATH LIONHEART: mean cosine similarity scores (first row); PCA plots and inter-dataset contributions (second row).

$w_i = N/(C \cdot N_i)$, where N is the total number of samples, C is the number of classes, and N_i is the number of samples in class i in $\mathcal{D}$. This approach ensures that less frequent classes receive higher weights, thereby reducing the impact of imbalance. We then quantify the normalized entropy of labels from the nearest neighbors per sample to estimate the uncertainty in label distribution independent from the re-weighting by Equation 1:

$$H_{\text{norm}} = \frac{-\sum_{i=1}^{m} p_i \log(p_i)}{\log(k)}, \tag{1}$$

where p_i is the probability of the i-th class and k is the number of nearest neighbors. By normalizing the entropy, it is scaled between 0 and 1 regardless of the number of drawn neighbors. A lower entropy indicates a higher purity of the neighborhood in terms of class labels, which is desirable for accurate classification.

2.3.4 Nearest neighbor search and prediction

The core of our method involves an adaptive nearest neighbor search to determine the optimal number of neighbors (k) for classification. Starting from a minimum value $k_{\min}$, k is increased exponentially until the entropy of the nearest neighbors falls below a pre-defined threshold or $k_{\max}$ is reached. This adaptive approach balances the need for accuracy and computational efficiency by dynamically adjusting k based on the neighborhood's label distribution. During the search process, which approximately doubles the runtime per test sample, the class weights are used to perform a weighted vote among the nearest neighbors to account for minority classes. The predicted class is then determined by the class with the highest weighted vote as indicated in Equation 2:

$$y_{\text{pred}} = \arg\max_{l} \sum_{j=1}^{k} w_j \cdot \mathbb{1}\{y_j = l\}, \tag{2}$$

where $\mathbb{1}\{y_j = l\}$ is an indicator function that is 1 if the label of the j-th nearest neighbor y_j is l and w_j is the weight of the j-th nearest neighbor's class label.

2.4 Experimental setup and hyperparameters

We optimized EWC-FAISS to use the best combination of foundation models (on a validation set) and compared to multiple fine-tuned classification models. Since most related work uses DINO to represent cellular morphology (8, 10), we compared to a (fully-)fine-tuned DINOv2-based vision transformer model (DINO FT), as well as fully-fine-tuned variants of the encoders SWIN and ConvNeXT (SWIN FT and ConvNeXT FT). Furthermore, we compared to a domain-adaption method, NMTune, optimizing a hidden layer following the frozen encoders, as a lightweight addendum to foundation models aiming at making performances more robust on (unseen) downstream tasks.

2.4.1 EWC-FAISS

To create an adaptive and high-performing configuration for EWC-FAISS, we optimized several parameters, including the subset of foundation models used. Given our set of up to six foundation models, we evaluated every possible model combination (all subsets) to find the one that maximized performance. For each subset, we measured macro accuracy on the validation set, selecting the combination with the highest accuracy as the final or *incumbent* configuration. The parameters for $k_{\min}$, $k_{\max}$ and the entropy threshold were selected empirically based on validation results in preliminary experiments. We did not apply (L2-)normalization or dimensionality reduction, as preliminary validation results also indicated better performance without these steps.

The best-performing index for the WBC dataset used a combination of SAM, ConvNeXT, SWIN, and CLIP with k between $k_{\min} = 3$ and $k_{\max} = 1000$, an entropy threshold of 0.3, and L2-distance. For the transfer to LISC, we used a combination of DINO, ConvNeXT, SWIN, and ViTMAE. For the cell state classification from the Nanolive microscope, we used a combination of SAM, ConvNeXT, SWIN, CLIP and ViTMAE with k between $k_{\min} = 3$ and $k_{\max} = 100$, an entropy threshold of 0.6 and Canberra distance. For the transfer to the Lionheart microscope, we set k between 10 and 1000 and an entropy threshold of 0.1 with a combination for SAM, DINO and SWIN. For the cell type classification from the Lionheart microscope, as well as for the transfer to the Nanolive microscope, we used a combination of SAM, DINO, ConvNeXT, and CLIP, with k between 20 and 1000, an entropy threshold of 0.2 and L2-distance. For the segmentation experiment for real Jurkat and K562 mixtures with PBMCs, we used a combination of SAM, DINO, ConvNeXT, SWIN and CLIP with k between 33 and 1000 and an entropy threshold of 0.2 without reweighting.

2.4.2 Fine-tuning baselines

To contextualize our results, we compare against standard fine-tuning approaches using state-of-the-art vision transformers. Specifically, we fine-tune a pre-trained DINOv2 model for 50 epochs using the AdamW optimizer with cosine learning rate decay and warm-up (from an initial learning rate of 10^{-5} to 0). Additionally, we apply the same training setup to SWIN and ConvNeXT backbones. During training, we use horizontal flipping, normalization, and color jitter as data augmentations.

2.4.3 Domain adaption baseline

As a baseline for domain adaptation, we followed a strategy in which the encoder parameters were kept fixed and only an intermediate hidden layer was fine-tuned. Specifically, we employed NMTune, which has been shown to be more robust than conventional layer-wise fine-tuning (23). Following the recommended setup in (23), we set $\lambda = 0.01$ to balance the regularization terms for feature consistency, covariance alignment, and dominant singular value preservation. We applied NMTune to the optimized embeddings described in Section 2.4.1, as these outperformed the individual encoder embeddings. To reduce computational costs, embeddings were first reduced to 200 principal components using PCA. A two-layer MLP with 800 hidden units was trained for 10 epochs using the Adam optimizer with a learning rate of 10^{-3}.

3 Results

First, we utilized a set of foundation models as encoders to generate cell sample embeddings as described in Section 2.3.1. Then, we used these embeddings to construct a database and resulting FAISS index as described in Section 2.3.2 to use this as a k-nearest neighbor classifier.

3.1 Separation of cell sample embeddings generated from a combination of foundation models

A first intuition into the discriminative capabilities of the general features from foundation models can be seen at the top of Figure 5, which depicts a UMAP 2D projection of the latent representations for the different foundation models and the best combination for the WBC dataset. Despite being untrained on domain-specific data, all foundation models except for ViTMAE preserve the local neighborhoods of the different classes, as indicated by the class-specific color coding. When projecting samples from the LISC dataset (i.e., for transfer), which contains only samples from healthy donors, it is evident that ConvNeXT, among all single foundation models, distinguishes best between blasts and non-blasts, with only two miscategorized examples. The best-found combination then has only a single misclassified example according to the 2D visualization. We compared the similarities among cell embeddings within one foundation model exemplary for the WBC dataset in Figure 5 (second row) by the mean of the cosine similarity scores of all embeddings per cell type projected via Principal Component Analysis (PCA) to 100 dimensions. The best combination of foundation models achieved the highest intra-class similarity with a mean diagonal similarity of 0.355 and the lowest inter-class similarity with a mean off- diagonal similarity of −0.043. In contrast, the ViTMAE model showed the noisiest results, indicating less distinct feature separation. Additionally, we studied the similarity across foundation model representations, using Radial Basis Function (RBF) Centered Kernel Alignment (27) in Figure 5 (third row). ConvNeXT and SWIN achieved the highest similarity compared to the best combination of models (being part of the best combination SAM+ConvNeXT +SWIN+CLIP). The fourth row of Figure 5 shows PCA-based intrinsic dimensionality estimates, revealing that the best combination has substantially higher intrinsic dimensionality than any individual model, indicating that it aggregates complementary, non-redundant features across models. To assess how each encoder contributes to the combined embedding, we projected the concatenated representation onto its principal components and quantified the average loading of each encoder's feature block across the top components, revealing their relative influence on the informative variance. Similarly, a projection of embeddings for cell states from CELL DEATH NANOLIVE can be found at the top left of Figure 6. The latent representation from the best combination of foundation models shows the best overall separability between classes, as also depicted in Figure 6 (middle, first and second row). The third row of Figure 6 (middle) further shows that this combined embedding retains a higher intrinsic dimensionality, indicating complementary information across models. Lastly, the first row of Figure 6 (bottom) shows the cosine similarities between states from CELL DEATH NANOLIVE and CELL DEATH LIONHEART. It can be seen, that the different encoders have complementary strengths and weaknesses for separating the single cell states. The found best combination offers a sweet-spot with best overall separation. The bottom row of Figure 6 (bottom) shows PCA-based estimates of intrinsic dimensionality for embeddings transferred from CELL DEATH NANOLIVE to CELL DEATH LIONHEART. The best combination retains a substantially higher number of informative components, coming from multiple different encoders, than any individual model, suggesting robust generalization.

Tables 1, 2 summarize the mean intra-class similarity (average of the diagonal), mean inter-class similarity (average of the off-diagonal), and class separability, defined as the average per class of the difference between the diagonal entry and the sum of off-diagonal entries. These metrics are based on cosine similarity and are reported for all foundation models and their best combination, using embeddings from the WBC and CELL DEATH NANOLIVE datasets.

Finally, we present a 2D projection of latent representations from the trajectories of 15 cells taken from the CELL DEATH NANOLIVE dataset and tracked over time by a human domain expert until apoptosis in Figure 6 (top right). The color indicates the time to apoptosis, smoothed across the respective nearest neighbors. While this does not constitute definitive proof of correct alignment, it illustrates the general tendency of foundation models to capture even subtle morphological differences related to cell state. We further investigate this hypothesis by evaluating classification performance on cell states in Section 3.2.

3.2 EWC-FAISS classification performance compared to fine-tuned methods on the original dataset and on transfer datasets

An overview of the classification results over five runs is given in Table 3. In terms of within-dataset performance (i.e., trained and tested on the same dataset), EWC-FAISS performs on par or better than the fine-tuned baselines: it achieves the highest accuracy on WBC (97.6%), ties for best on CELL DEATH Nanolive (90%), and performs competitively on CELL TYPE Lionheart (87%). In the transfer setting, EWC-FAISS demonstrates clearly superior generalization. It outperforms all fine-tuned and domain-adapted models by large margins on LISC (78.5% vs. 52% for the next best), CELL DEATH Lionheart (86% vs. 81%), and CELL TYPE Nanolive (85% vs. 63%). Notably, models like DINO FT and SWIN FT degrade significantly in these transfer scenarios, especially on LISC, where DINO FT drops to 17% and SWIN FT to 44%. NMTune, while more than the fine-tuned models, still trails behind EWC-FAISS across all transfer tasks.

In addition to performance, EWC-FAISS offers substantial computational advantages. Once the embeddings are computed, building EWC-FAISS is multiple orders of magnitudes faster than training DINO FT (∼ 6 minutes compared to > 10 hours, cf. Figures 7, 8) . Lastly, we created real mixtures of Jurkat and K562 cell lines with PBMCs from healthy donors. Our results demonstrate that EWC-FAISS performs comparably to (semi-) supervised segmentation methods (28) in identifying cancerous

TABLE 1 Mean intra-class and inter-class similarities, as well as class separability for the WBC dataset.

Model	Intra-class similarity	Inter-class similarity	Class separability
Best Combination	0.355	−0.043	**0.654**
SWIN	0.349	−0.041	0.637
ConvNeXT	0.33	−0.04	0.612
CLIP	0.244	−0.027	0.431
ViTMAE	0.179	−0.021	0.329
DINO	0.183	−0.014	0.283
SAM	0.144	−0.019	0.274

All metrics are based on cosine similarity. *Class Separability* is defined as the average per class of the difference between the diagonal entry and the sum of off-diagonal entries. Values in bold indicate best separability.

TABLE 2 Mean intra-class and inter-class similarities, as well as class separability for different cell states in the CELL DEATH NANOLIVE dataset.

Model	Intra-class similarity	Inter-class similarity	Class separability	Class separability (Transfer)
Best Combination	0.188	0.009	**0.162**	**0.0772**
ConvNeXT	0.147	-0.004	0.159	0.0626
CLIP	0.176	0.007	0.154	0.0596
ViTMAE	0.130	-0.007	0.152	0.0168
DINO	0.140	-0.001	0.143	0.0769
SWIN	0.215	0.032	0.120	0.0736
SAM	0.290	0.128	-0.093	-0.0003

All metrics are based on cosine similarity. *Class Separability* is defined as the average per class of the difference between the diagonal entry and the sum of off-diagonal entries. Values in bold indicate best separability.

cell lines within the collection of healthy PBMCs (cf. Table 4; Figure 9).

3.3 Application demonstration: classification of fixed cells

We implemented EWC-FAISS as a web application to provide a subtype detection for white blood cells using our model trained on WBC. The streamlit[1] application, shown in Figure 10 allows to upload an image and shows the prediction together with a certainty measure and the $m = 5$ nearest neighbors of the training set. As measure of certainty, the normalized entropy H_{norm} was calculated from the label counts of the k nearest neighbors. This entropy was normalized relative to the number of nearest neighbors to produce a certainty percentage as an intuitive measure of classification confidence:

$$C = (1 - H_{\text{norm}}) \times 100\,\%$$

This application serves as an initial step toward practical clinical use, with the potential to be integrated with a microscope for real-time WBC subtype classification in clinical diagnostics.

1 https://streamlit.io/

3.4 Application demonstration: longitudinal classification of cell type and state during live cell analysis

The subtype detection classifier trained on the CELL DEATH NANOLIVE dataset was used to classify the cell state for all images at each time point $t = 1, \ldots, t_e$, where t_e is the final time frame of the experiment. Additionally, as a control signal, we calculate the fluorescence intensity by thresholding each Apo-15 fluorescence image using the Otsu method, which determines the optimal threshold by maximizing inter-class variance. A binary mask was created from the thresholded image to highlight regions of interest. These regions were then labeled, and the total fluorescence intensity was quantified by summing the area of the labeled regions. This analysis provides a quantitative measure of the cellular response over the course of the experiment. The analysis of a live cell experiment for JIMT-1 cells treated with 10 µM Raptinal is shown in Figure 11.

4 Discussion

Using AI-based approaches to optimize cancer therapy, in general, is gaining momentum (29–31). However, distinguishing between different cell types or states from microscopic images,

10.3389/fonc.2025.1480384

TABLE 3 Classification results for models with fully fine-tuned feature extraction (FE), domain adaptation using a frozen backbone (NMTune), and EWC-FAISS with unsupervised FE.

Feature extraction	Method	Macro accuracy	Macro precision
Original: WBC (8 classes)			
	DINO FT	94.0 ± 0.9	98 ± 2
FINE-TUNED	SWIN FT	92.9 ± 0.8	99.4 ± 0.1
	ConvNeXT FT	92.7 ± 0.5	98 ± 1
DOMAIN-ADAPTED	NMTune	98.4 ± 0.5	98.1 ± 0.9
UNSUPERVISED	EWC-FAISS	97.6 ± 0.2	98 ± 0
Transfer: LISC (5 classes)			
	DINO FT	17 ± 1	33 ± 2
FINE-TUNED	SWIN FT	44 ± 14	35 ± 3
	ConvNeXT FT	45 ± 9	49 ± 8
DOMAIN-ADAPTED	NMTune	52 ± 10	59 ± 6
UNSUPERVISED	EWC-FAISS	78.5 ± 0.3	81.9 ± 0.5
Original: Cell Death Nanolive (4 classes)			
	DINO FT	89.5 ± 0.9	89 ± 1
FINE-TUNED	SWIN FT	90 ± 4	91 ± 2
	ConvNeXT FT	89 ± 2	91 ± 2
DOMAIN-ADAPTED	NMTune	88.4 ± 0.6	89 ± 1
UNSUPERVISED	EWC-FAISS	90 ± 0	92 ± 0
Transfer: Cell Death Lionheart (2 classes)			
	DINO FT	65 ± 11	80 ± 21
FINE-TUNED	SWIN FT	68 ± 16	75 ± 14
	ConvNeXT FT	81 ± 1	85 ± 2
DOMAIN-ADAPTED	NMTune	81 ± 7	80 ± 6
UNSUPERVISED	EWC-FAISS	86 ± 1	84.9 ± 0.8
Original: Cell Type Lionheart (2 classes)			
	DINO FT	91.9 ± 0.1	92.0 ± 0.1
FINE-TUNED	SWIN FT	90.6 ± 0.2	90.7 ± 0.2
	ConvNeXT FT	92.32 ± 0.08	92.33 ± 0.05
DOMAIN-ADAPTED	NMTune	92.1 ± 0.2	91.9 ± 0.1
UNSUPERVISED	EWC-FAISS	87 ± 0	87 ± 0
Transfer: Cell Type Nanolive (1 class)			
	DINO FT	24.5 ± 0.9	–
FINE-TUNED	SWIN FT	57 ± 6	–
	ConvNeXT FT	54 ± 4	–
DOMAIN-ADAPTED	NMTune	63 ± 4	–
UNSUPERVISED	EWC-FAISS	85 ± 0	–

Models were trained on the ORIGINAL set and evaluated on both the test and TRANSFER sets. Results are reported over five runs as mean ± standard deviation, rounded to one significant digit.

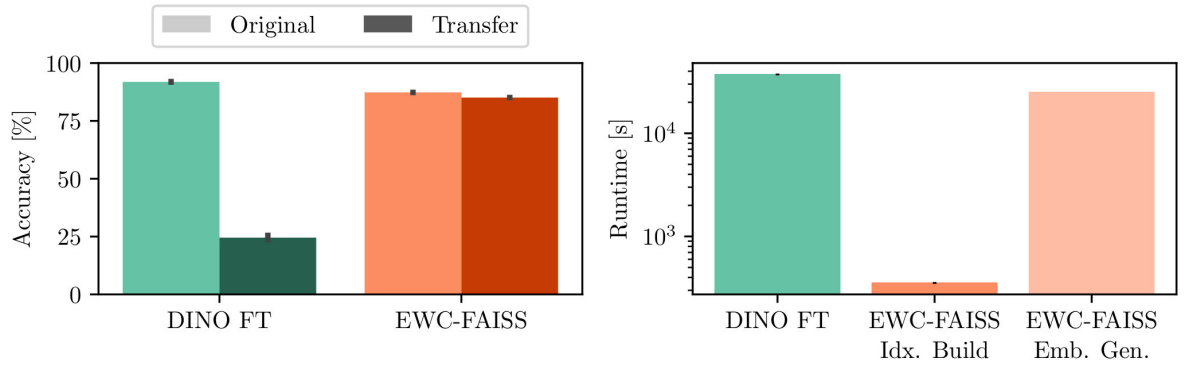


FIGURE 7
Results on the transfer from a BioTek Lionheart FX to the Nanolive 3D Cell Explorer. (left) EWC-FAISS is robust to distribution shift induced by the second device compared to DINO FT. (right) Once the embeddings are generated (Emb. Gen.), EWC-FAISS is multiple orders of magnitude faster (Idx. Build).

TABLE 4 Results of EWC-FAISS on mixtures of PBMCs and Jurkats and PBMCs and K562s.

Method	PBMC +Jurkats	PBMC +K562s	∅
	mAcc	mAcc	mAcc
CellMixer (28)	95.8	63.7	79.7
EWC-FAISS	85.3 ± 0.7	75.2 ± 0.5	80 ± 5

Results are reported as mean ± SD (denoted with one significant digit).
CellMixer results as reported in (28).

especially in a heterogeneous sample population, still pose significant challenges. Importantly, our results provide evidence that adaptive k-NN search on fixed features from combinations of foundation models (cf. Figure 8) can yield matching or better performances in the domain of cellular imaging to distinguish between different cell types or cell states. However, combining features from foundation models also is no panacea and except for the WBC dataset, we had to adapt the combinations and hyperparameters when transferring to a new device. Nonetheless, we were always able to find a good working combination in the realm of minutes instead of hours of training fully fine-tuned models, even for full iterations over the whole power set. This highlights the versatility and accessibility of the proposed framework, as performing several development cycles is very cost-effective and fast; all experiments have been executed on consumer hardware, specifically an AMD Ryzen 9 7950X3 CPU and an NVIDIA GeForce RTX 4090. NMTune on the best found set of foundation models also is a well-performing alternative to a fully-trained classification model, although the generalization capabilities toward out-of-distribution samples appear to be inferior compared to approximate k-NN search. While an exhaustive search over the power set is reasonably cost-effective for smaller sets of foundation models, optimizing the best combination may be challenging in resource-constrained environments or with larger candidate sets. In such cases, the combination of DINO, ConvNeXT, and SWIN has demonstrated consistent performance, particularly for transfer tasks across all evaluated datasets.

The adaptive similarity search on general-purpose features from foundation models proved robust across various microscopy scenarios and devices. This robustness is further demonstrated by achieving not only the highest transfer performance in terms of mean macro-accuracy but also the lowest associated standard deviations. In the LISC transfer task, EWC-FAISS shows exceptional stability with a standard deviation of only ±0.3, compared to NMTune's ±10, SWIN FT's ±14 or ConvNeXT FT's ±9, and less than half of Dino FT's ±1. For transfer of cell state detection, EWC-FAISS and ConvNeXT FT achieve a variance up to an order of magnitude smaller than Dino FT or SWIN FT (± 1 vs. ± 11 and ±16) and notably lower than NMTune's ±7. Finally, in the cell type detection transfer task, EWC-FAISS shows negligible variance, compared to NMTune's ±4, underscoring its consistent performance across various transfer tasks. Finding robust representations of cells is crucial for AI-assisted analysis of the effects of cancer therapies. This robustness is particularly important given the common challenges in medical research, such as small sample sizes and limited data availability. The ability to accurately classify different cell types and tracking changes in real-time using AI-based methods is essential for assessing and accelerating therapeutic impacts. In our study, we successfully distinguished multiple cell types, including leukemic cell lines and PBMCs. This capability is significant, as PBMCs from healthy donors can serve as a model for the tumor microenvironment, providing a heterogeneous mixture of cells similar to patient tumor samples. This aspect of our study underscores the potential application of our method in clinical settings, where differentiating malignant cells from normal immune cells can provide insights into immune infiltration and tumor biology. Furthermore, our methodology demonstrates the capability to identify cell states such as live, apoptotic, necrotic, or dead cells, which for example is vital for evaluating the efficacy of cancer therapies. This functionality is particularly relevant in the context of cancer therapy. Our cell similarity search approach could potentially streamline this process,

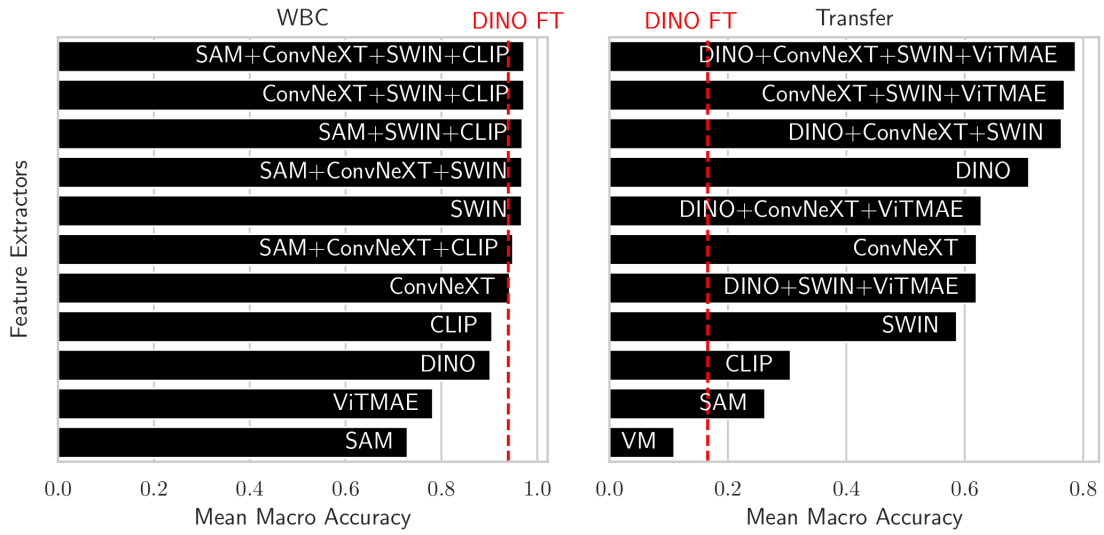


FIGURE 8
Balanced accuracies of the foundation models and the best combination using EWC-FAISS in black and DINO FT in red on WBC (left) and on the transfer to LISC (right).

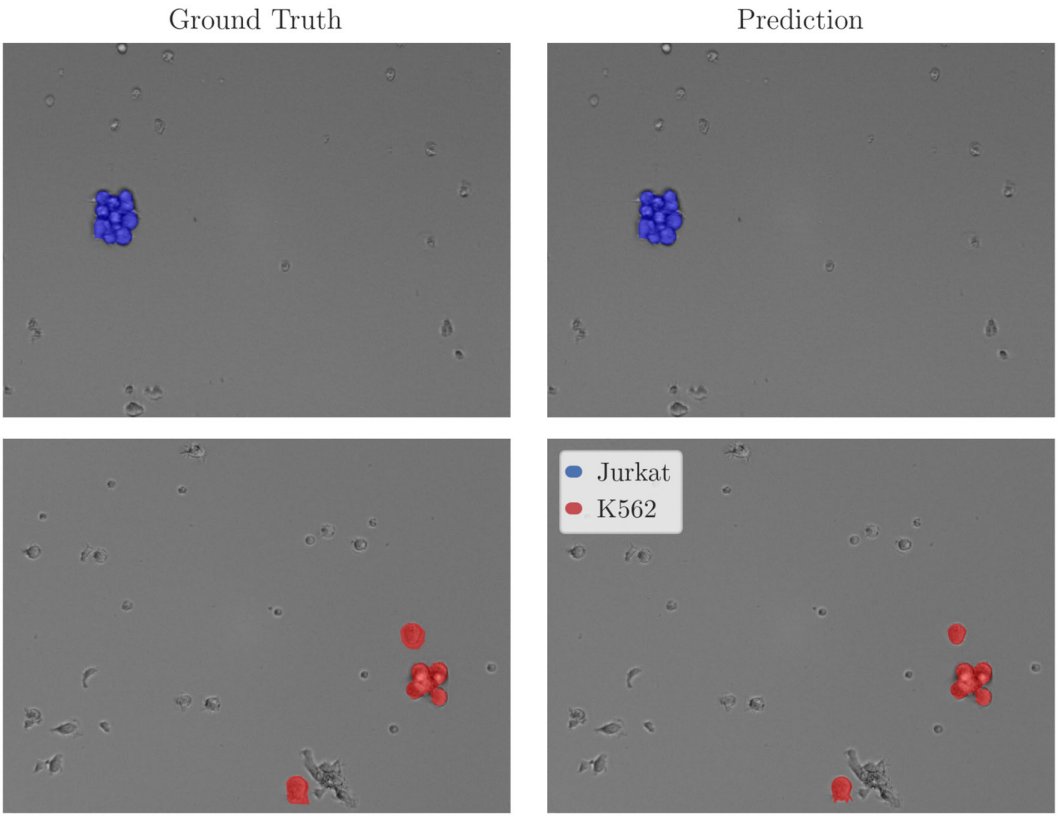


FIGURE 9
Exemplary inference results for EWC-FAISS evaluated on real mixtures of PBMCs with Jurkats and PBMCs and K562s. Expert annotations were non-exhaustively given for Jurkats and K562s, PBMCs are not annotated.

FIGURE 10
WBC EWC-FAISS classifier within a Streamlit web application predicting a monocyte correctly, showing the classification certainty and the nearest neighbors found in the database.

facilitating the identification of donor cells with strong anti-tumor activity and thus optimizing the preparation and effectiveness of immunotherapies.

In our study, we validated this approach across diverse datasets encompassing various staining techniques and imaging platforms, from conventional microscopy to advanced holotomography-based imaging. The consistent performance across these different datasets highlights our method's robustness, particularly in clinical contexts involving complex, heterogeneous cell populations. This robustness also suggests potential beyond oncology, where our method's ability to distinguish molecular characteristics in stained cells could accelerate cytological investigations, helping to identify clinically relevant features in settings such as point-of-care diagnostics. For example, a stained blood sample could be imaged with a basic microscope, uploaded via a web application, and analyzed in real time, providing rapid diagnostic insights. While our study primarily utilized optical and hologram-based microscopy, our approach could generalize effectively to other high-resolution imaging methods. Techniques, which capture high-contrast, visually distinct features, align well with our approach. The data efficiency of EWC-FAISS further supports adaptability, even in domains with limited image throughput.

Going forward, the rapid evolution of foundation models presents a valuable opportunity to enhance our framework. With

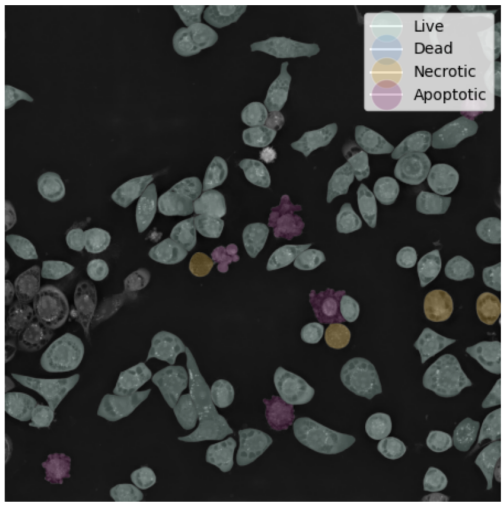


Figure 11a.

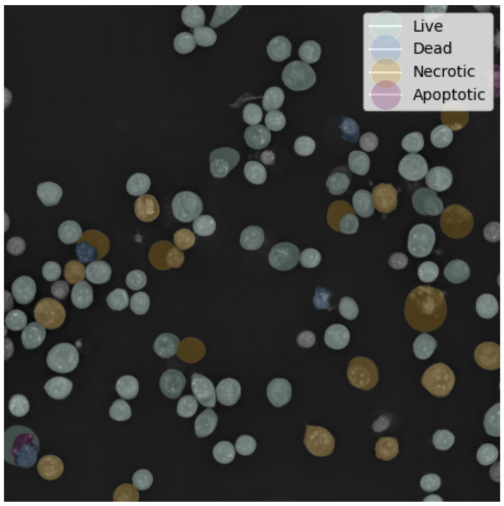


Figure 11b.

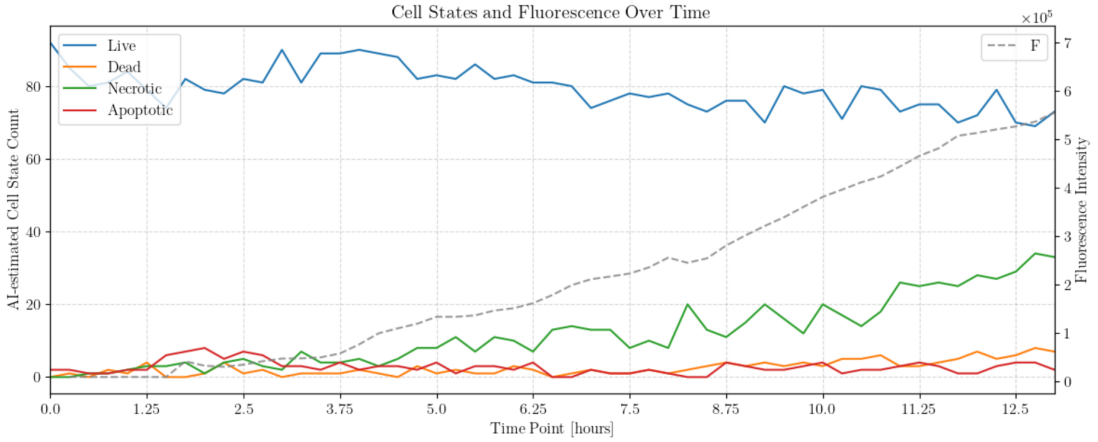


Figure 11c.

FIGURE 11
AI-Detection of JIMT-1 cell states using EWC-FAISS trained on NANOLIVE CELL DEATH. The cells were treated with Raptinal (10 μM). **(a)** Timepoint 1.75 hours. **(b)** Timepoint 9.25 hours. **(c)** AI-assisted cell state counts over 12 hours and the fluorescence intensity.

EWC-FAISS, foundation models can be seamlessly replaced as new models emerge: embeddings can be generated from updated models, and the index rebuilt without altering the underlying workflow. Models such as CellSAM, which have recently emerged, may offer richer feature representations that could further improve the accuracy and robustness of our method. Incorporating more advanced models into our pipeline will likely yield significant benefits, particularly in handling the increasingly complex data generated in modern biomedical research. However, as the diversity of visual foundation models continues to expand, there is a growing need for sophisticated hyperparameter optimization techniques to identify the most effective model combinations.

Lastly, exploring metroid-based solutions offers a promising direction for future research. These methods could provide more efficient alternatives to the full (approximate) k-NN search currently employed, particularly in scenarios involving even large datasets or limited computational resources. By optimizing the computational efficiency of our approach, we can further lower the barriers to its widespread adoption and application in diagnostics and therapy response monitoring.

5 Conclusion

In this study, we presented a robust and versatile method for distinguishing between various cell types and states using adaptive k-NN search on fixed features derived from combinations of foundation models. Our approach demonstrated high adaptability and efficiency across different microscopic imaging devices,

highlighting its potential for broad applicability in both research and clinical settings. The method's ability to rapidly adapt to different datasets and experimental conditions, while maintaining accurate classification performance, underscores its utility in dynamic and resource-constrained environments.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human samples were approved by the Ethics Commission of the Albert-Ludwigs-University of Freiburg (Approval number: 22-1442-S1). The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by-product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

GK: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AK: Conceptualization, Data curation, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. PS: Data curation, Writing – review & editing. JR: Methodology, Writing – review & editing. MN: Methodology, Validation, Writing – review & editing. YV: Methodology, Writing – review & editing. DI: Data curation, Writing – review & editing. RB: Methodology, Writing – review & editing. JD: Methodology, Writing – review & editing. TH: Methodology, Writing – review & editing. MF: Methodology, Resources, Writing – review & editing. EB: Data curation, Writing – review & editing. ML: Methodology, Writing – review & editing. RM: Conceptualization, Funding acquisition, Methodology, Resources, Writing – review & editing. JB: Funding acquisition, Resources, Writing – review & editing. EU: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. MK: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This project was funded by the Mertelsmann Foundation.

Acknowledgments

We thank Florian Wolpert for his help in implementing the web applications.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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.

References

1. Johnson TR, Bourne PE. The biological data sustainability paradox. *arXiv*. (2023).

2. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, et al. On the opportunities and risks of foundation models. *ArXiv*. (2021). doi: 10.48550/arXiv.2108.07258

3. Caron M, Touvron H, Misra I, Jegou H, Mairal J, Bojanowski P, et al. (2021). Emerging properties in self-supervised vision transformers, in: *IEEE/CVF International Conference on Computer Vision, ICCV 2023*, Paris, France, October 1-6, 2023. New York, NY, USA: IEEE.

4. Oquab M, Darcet T, Moutakanni T, Vo HV, Szafraniec M, Khalidov V, et al. DINOv2: Learning robust visual features without supervision. *Trans Mach Learn Res*. (2024).

5. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. (2023). Segment anything, *IEEE/CVF International Conference on Computer Vision, ICCV 2023*, Paris, France, October 1-6, 2023. New York, NY, USA: IEEE.

6. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. *Nat Commun*. (2024) 15:654. doi: 10.1038/s41467-024-44824-z

7. Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. *Nat Med*. (2024) 30:850–62. doi: 10.1038/s41591-024-02857-3

8. Doron M, Moutakanni T, Chen ZS, Moshkov N, Caron M, Touvron H, et al. Unbiased single-cell morphology with self-supervised vision transformers. (2023). doi: 10.1101/2023.06.16.545359

9. Koch V, Wagner SJ, Kazeminia S, Sancar E, Hehr M, Schnabel JA, et al. "Dinobloom: A foundation model for generalizable cell embeddings in hematology." In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. MICCAI 2024. Lecture Notes in Computer Science*. Springer, Cham (2024) vol. 15012. doi: 10.1007/978-3-031-72390-2_49.

10. Pfaendler R, Hanimann J, Lee S, Snijder B. Self-supervised vision transformers accurately decode cellular state heterogeneity. *bioRxiv*. (2023). doi: 10.1101/2023.01.16.524226

11. Israel U, Marks M, Dilip R, Li Q, Schwartz M, Pradhan E, et al. A foundation model for cell segmentation. *bioRxiv*. (2023). doi: 10.1101/2023.11.17.567630

12. Kraus O, Kenyon-Dean K, Saberian S, Fallah M, McLean P, Leung J, et al. "Masked autoencoders for microscopy are scalable learners of cellular biology." In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA (2024). pp. 11757–11768. doi: 10.1109/CVPR52733.2024.01117.

13. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows, in: *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, . pp. 9992–10002. IEEE.

14. Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, et al. (2022). Swin transformer v2: Scaling up capacity and resolution, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*, New Orleans, LA, USA, June 18-24, 2022. New York, NY, USA: IEEE.

15. Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S. (2022). A convNet for the 2020s, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*, New Orleans, LA, USA, June 18-24, 2022. New York, NY, USA: IEEE.

16. Woo S, Debnath S, Hu R, Chen X, Liu Z, Kweon IS, et al. (2023). ConvNeXt v2: Co-designing and scaling ConvNets with masked autoencoders, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023*. Vancouver, BC, Canada: IEEE.

17. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. (2021). Learning transferable visual models from natural language supervision, in: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*. Cambridge, MA, USA: PMLR. pp. 8748–63.

18. He K, Chen X, Xie S, Li Y, Dollar P, Girshick R. (2022). Masked autoencoders are scalable vision learners, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*, New Orleans, LA, USA, June 18-24, 2022. New York, NY, USA: IEEE.

19. Huh M, Cheung B, Wang T, Isola P. (2024). The platonic representation hypothesis, in: *Proceedings of the 41st International Conference on Machine Learning, ICML 2024*. Cambridge, MA, USA: PMLR

20. Malkov YA, Yashunin DA. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE Trans Pattern Anal Mach Intell*. (2020) 42:824–36. doi: 10.1109/TPAMI.34

21. Douze M, Guzhva A, Deng C, Johnson J, Szilvasy G, Mazaré P-E, et al. *The faiss library*. *arXiv*. (2024).

22. Arango SP, Ferreira F, Kadra A, Hutter F, Grabocka J. (2024). Quick-tune: Quickly learning which pretrained model to finetune and how, in: *The Twelfth International Conference on Learning Representations, ICLR 2024*, Vienna, Austria, May 7-11, 2024. Amherst, MA, USA: OpenReview.net

23. Chen H, Wang J, Shah A, Tao R, Wei H, Xie X, et al. (2024). Understanding and mitigating the label noise in pre-training on downstream tasks, in: *The Twelfth International Conference on Learning Representations*, .

24. Bodzas A, Kodytek P, Zidek J. A high-resolution large-scale dataset of pathological and normal white blood cells. *Sci Data*. (2023) 10:466. doi: 10.1038/s41597-023-02378-7

25. Rezatofighi H, Khaksari K, Soltanian-Zadeh H. Automatic recognition of five types of white blood cells in peripheral blood. (2010) 6112:161–72. doi: 10.1007/978-3-642-13775-417

26. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. (2020). Transformers: Stateof-the-art natural language processing, in: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Stroudsburg, PA, USA: Association for Computational Linguistics pp. 38–45.

27. Kornblith S, Norouzi M, Lee H, Hinton G. Similarity of neural network representations revisited. In: Chaudhuri K, Salakhutdinov R, editors. *Proceedings of the 36th International Conference on Machine Learning*, vol. 97 . Cambridge, MA, USA: Proceedings of Machine Learning Research (2019). p. 3519–29.

28. Naouar M, Kalweit G, Klett A, Vogt Y, Silvestrini P, Ramirez DLI, et al. *Cellmixer: Annotation-free semantic cell segmentation of heterogeneous cell populations*. *arXiv*. (2023).

29. Netherton TJ, Cardenas CE, Rhee DJ, Court LE, Beadle BM. The emergence of artificial intelligence within radiation oncology treatment planning. *Oncology*. (2020) 99:124–34. doi: 10.1159/000512172

30. Xu Z, Wang X, Zeng S, Ren X, Yan Y, Gong Z. Applying artificial intelligence for cancer immunotherapy. *Acta Pharm Sin B*. (2021) 11:3393–405. doi: 10.1016/j.apsb.2021.02.007

31. Fiste O, Gkiozos I, Charpidou A, Syrigos NK. Artificial intelligence-based treatment decisions: A new era for nsclc. *Cancers*. (2024) 16. doi: 10.3390/cancers16040831

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 24 November 2025
DOI 10.3389/fonc.2025.1464884

Check for updates

OPEN ACCESS

EDITED BY
Sharon R. Pine,
University of Colorado Anschutz Medical Campus, United States

REVIEWED BY
Chenbin Liu,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Yingjian Yang,
Shenzhen Lanmage Medical Technology Co., Ltd, China

*CORRESPONDENCE
Jing Cai
jing.cai@polyu.edu.hk

†PRESENT ADDRESS
Yuanpeng Zhang,
Department of Medical Informatics, Nantong University, Nantong, Jiangsu, China

‡These authors have contributed equally to this work

RECEIVED 15 July 2024
ACCEPTED 27 October 2025
PUBLISHED 24 November 2025

CITATION
Wang Y, Zwanenburg A, Zhang J, Teng X, Lam S-K, Cao J, Ma Z, Zhou T, Zhang Y, Ge H and Cai J (2025) Establishment of a CT-based radiomic feature robustness databank for OPC patients via image perturbation in a multi-institutional study: a practical method to safeguard model generalizability. *Front. Oncol.* 15:1464884. doi: 10.3389/fonc.2025.1464884

COPYRIGHT
© 2025 Wang, Zwanenburg, Zhang, Teng, Lam, Cao, Ma, Zhou, Zhang, Ge 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.

Establishment of a CT-based radiomic feature robustness databank for OPC patients via image perturbation in a multi-institutional study: a practical method to safeguard model generalizability

Yongqiang Wang[1,2‡], Alex Zwanenburg[3,4,5,6‡], Jiang Zhang[1], Xinzhi Teng[1], Sai-Kit Lam[7,8], Jin Cao[1], Zongrui Ma[1], Ta Zhou[1], Yuanpeng Zhang[1†], Hong Ge[2] and Jing Cai[1,8,9*]

[1]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [2]Department of Radiation Oncology, The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, China, [3]National Center for Tumor Diseases Dresden/University Cancer Center (NCT/UCC), German Cancer Research Center Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany, [4]Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden (TUD) Dresden University of Technology, Dresden, Germany, [5]Department of Radiation Physics, Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Dresden, Germany, [6]OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden (TUD) Dresden University of Technology, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany, [7]Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [8]Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [9]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China

Purpose: To guide the preselection of highly repeatable radiomic features (RFs) in downstream analysis without further analysis its repeatability, a detailed radiomic feature robustness databank (RF-RobustDB) was established via image perturbation.

Methods: Data on 1,274 oropharyngeal carcinoma (OPC) patients who had undergone pretreatment computed tomography (CT) imaging, collected from a public dataset. The original images and corresponding masks underwent systematic perturbations to simulate potential variations encountered during CT image rescanning, including translational shifts, rotational changes, random noise additions, and contour modifications. For each radiomic feature (RF), including unfiltered, wavelet-filtered, and Laplacian-of-Gaussian (LoG)-filtered features, we systematically quantified robustness against these perturbations by intraclass correlation coefficients (ICCs).

Results: Out of 1395 first- and high-order RFs, 470 demonstrated excellent repeatability, i.e., a mean ICC of greater than 0.9. The use of these preselected highly repeatable RFs in model development improved the mean concordance (C) index in two external validation cohorts and reduced the mean C index gap between the training and external validation cohorts. These results demonstrate

that the preselected high repeatable RFs from RF-RobustDB can effectively enhance radiomic model generalizability.

Conclusions: The methodology employed to establish the RF-RobustDB is highly transferable to other tumor sites and different imaging modalities, which will facilitate the creation of RF-RobustDBs to guide the development of universally applicable radiomic models.

KEYWORDS

radiomics, feature repeatability, model generalizability, oropharyngeal carcinoma, progression-free survival

1 Introduction

The role and potential of radiomics in cancer management have been constantly expanding over the past decades, such as the distant metastases prediction of advanced nasopharyngeal carcinoma (NPC) (1), performing risk stratification of oropharyngeal cancer (OPC) (2), breast cancer risk estimation (3), and prediction of treatment response in non-small-cell lung cancer (NSCLC) (4). However, model generalizability remains the prime stumbling block for bend-to-bedside translation of radiomic models. To enhance the generalizability of radiomic models, concerted efforts have been made to enhance repeatability and reproducibility of radiomic features (RFs) for primary model generation (5–7). RF extraction, implemented prior to modeling process (8–11), is crucial for ensuring model reliability and generalizability. Although the Image Biomarker Standardization Initiative (IBSI) provides standardized guidelines for RF extraction (12), RF repeatability and reproducibility remain limited across institutions and imaging protocols (13–15). Consequently, these limitations represent fundamental challenges that need to be addressed before RFs can be effectively incorporated into modeling workflows.

Multiple variables influence the repeatability and reproducibility of RFs throughout the imaging process (13–27), such as scanner model (13, 14), scanner type (13, 15), scanning parameters (16), segmentation (17, 18), reconstruction (25), and preprocessing methods (26, 27). However, clear guidelines for selecting highly repeatable RFs in multi-institutional datasets remain unavailable. Test-retest methods pose additional challenges, as they may increase patients' radiation exposure and consume medical resources unnecessarily. Manual re-segmentation further burdens radiologists with additional workload. Although phantom-based studies offer a radiation-free alternative for evaluating RF selection (28), their clinical applicability is limited due to imperfect simulation of human tissues. Given the practical constraints of test-retest studies and manual re-segmentation across institutions, there is an urgent need for a cost-effective, efficient, easily implementable, and clinically transferable RF robustness assessment method. Fortunately, A software-based image perturbation method proposed by Zwanenburg et al. offers a promising way to simulate the test-retest and re-segmentation process (29). This method simulates patient positioning during imaging, manual segmentation randomization, and varying noise levels of the imaging device. The effectiveness of image perturbation has been demonstrated by improved performance in radiomic models. For example, Teng et al. (30, 31) and Zhang et al. (6) applied image perturbation to select highly reproducible RFs that improved the reliability and generalizability of radiomic models. Moreover, image perturbation has been shown to achieve the same optimal reliability as test-retest imaging for constructing radiomic models (32).

Since the perturbation method demonstrates encouraging/promising capabilities in assessing feature stability, in this study, we aim to establish a reliable RF robustness databank (RF-RobustDB) via perturbation method for guiding the downstream development of radiomic models. Specifically, we included a large-scale of CT images of OPC patients from a total of 7 medical institution. The OPC dataset was obtained from the Cancer Imaging Archive (TCIA) (33). RFs from CT images with and without applications of popular imaging filters were analyzed. The repeatability of the RFs in the RF-RobustDB was quantified by one-way intra-class correlation coefficients (ICCs) (34). We adopted CT dataset for this study mainly owing to its wide-spreading popularity in the cancer management for pre-treatment planning, mid-treatment monitoring, and post-treatment evaluation, as well as the availability of dataset in the community.

Through systemic analysis, the cohort size effects on feature repeatability ensured that the sample size is sufficient to maintain the reliability of RF-RobustDB. Meanwhile, the RF-RobustDB-enhanced selection of highly repeatable RFs significantly improved the generalizability of progression-free survival (PFS) model. These results support the reliability of the comprehensive CT-based RF-RobustDB for OPC, offering a valuable insight into RF repeatability. Moreover, this study provided a comprehensive and generalized methodology for establishing an extensive RF-RobustDB applicable to diverse tissue sites and imaging modalities.

2 Materials and methods

2.1 Patient cohort

This retrospective study analyzed a dataset of pretreatment CT images from 1,418 head-and-neck cancer patients obtained from TCIA (33). The dataset included patients from seven medical institutions: 137 patients from the single-institution HEAD-NECK-RADIOMICS-HN1 (HN1) study (35, 36), 524 patients from the single-institution Radiomic Biomarkers in Oropharyngeal Carcinoma (RBOPC) study (37, 38), 298 patients from four institutions in the Head-Neck-PET-CT (HNPET) study (39, 40), and 459 patients from the single-institution Head and Neck Squamous Cell Carcinoma (HNSCC) study (41–43). To maintain consistency, only OPC patients with primary gross tumor volume (GTV) data were included, resulting in a final cohort of 1,274 OPC patients for establishing the site-specific RF-RobustDB. The baseline characteristics of the selected OPC patient are presented in Table A1. Due to the retrospective nature of this study, informed consent was not required.

2.2 Image perturbation

To simulate the inevitable variabilities in patient setup during image acquisition, a validated image perturbation method was used to mimic patient setup, randomized noise, and manual segmentation diversity. Translational and rotational perturbations were applied to the original (unfiltered) images and tumor masks to mimic patient position. Randomized noise was added to the original images to simulate noise variations during image acquisition. Contour randomization was applied to the tumor mask to mimic variations in manual tumor segmentation.

The image perturbation settings were based on previous studies on repeatability evaluation via image perturbation (6, 29, 31): translation distances were set to 0, 0.4, and 0.8 pixels; rotation angles were set to -20°, 0°, and 20°; noise levels were increased to 0, 1, 2, and 5 times the original noise level; and a three-dimensional random displacement field was used to deform segmented masks, resulting in randomized contours. For each voxel point, a random field vector component in each dimension is generated from a uniform distribution between -1 and 1. All z-components of the field vectors on the same slice are set to the same value to mimic the uniform inter-slice contour variations resulting from slice-by-slice contouring. The field vectors are then normalized in each dimension by the root mean square. Finally, they are smoothed using a Gaussian filter with a defined sigma value of 10 to ensure continuous changes in the random displacement field and to avoid sharp changes in the deformed contours. Sixty different perturbations were performed to enhance the reliability of our results, as previous studies have suggested that 40 different perturbations are sufficient (6, 30). During each perturbation operation, parameters from the four perturbation modes were randomly combined to simulate the uncertainty in variables during image rescanning.

TABLE A1 Baseline patient characteristics of the dataset in different cohorts.

Data cohort	Sex	Median age	Overall stage
HNSCC	Male: 395	57(28-87)	I-IV
	Female:64		
HN1	Male: 67	60(44-80)	I-IVb
	Female:21		
RBOPC	Male: 423	60(33-89)	I-IVb
	Female:101		
HNPET	Male: 151	63(34-90)	I-IVb
	Female:52		

2.3 RF extraction

Image pre-processing and RF extraction were conducted in accordance with the IBSI guidelines (12). Before RF extraction, all images were resampled to a $1 \times 1 \times 1$ mm^3 resolution, and re-segmentation was performed to limit pixel values between -150 and 180 HU, effectively excluding non-tumor tissue (such as air and bone) within the volume of interest (31). As a previous study suggested, using the fixed bin number between 8 and 128 discretize images can reduce the infinite possible number of intensity values to a finite set and image noise (44). Hence, a fixed bin number of 30 was used for image discretization in this study. RF extraction was performed using PyRadiomics v2.2 (45) in Python v3.7. Shape-based features, first-order features, and high-order features from the gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighboring gray-tone difference matrix (NGTDM) were extracted from the GTVs in original, Laplacian-of-Gaussian (LoG)-filtered (with sigma values of 1, 2, 3, 4, 5, and 6 mm) and Coiflet-1 wavelet-filtered images.

Fourteen shape-based features were extracted from each tumor mask, and 93 first-order and high-order features were extracted from each of the unfiltered, LoG-filtered, and wavelet-filtered images. Following image perturbation, we additionally extracted corresponding feature sets from all perturbed image variants. Finally, the RF-RobustDB contained 1,316 unfiltered, LoG-filtered, and wavelet-filtered RFs, and 78,960 perturbed features were extracted for ICC analysis.

2.4 RF repeatability assessment

Since the feature is extracted from different perturbated mask region, the assignment of perturbation parameters is independent to patients. Therefore, the robustness of each RF was quantified in

terms of a one-way, random, absolute-agreement ICC, which was calculated using Equation 1, as follows (34).

$$ICC(1,1) = \frac{MS_n - MS_W}{MS_n + (k+1)MS_W} \quad (1)$$

where MS_n is the mean square for different patients, MS_W is the mean square for residual sources of variance, and k is the number of perturbation times plus one for the unperturbed image. As recommended by a previous study (34), features with an ICC< 0.5 were regarded as having poor repeatability, ICC between 0.50 and 0.75 were regarded as having moderate repeatability, ICC between 0.75 and 0.90 were regarded as having good repeatability, and ICC > 0.9 were regarded as having excellent repeatability.

2.5 Establishment of the RF-RobustDB

To determine the reliable patient sample size (n) required for constructing the RF-RobustDB, the mean ICC values as a function of patient numbers were systematically analyzed. The methodology was implemented as follows: Starting with 10 patients, the sample size was incrementally increased by 10 patients up to 100, followed by 100-patient increments thereafter. For each specific patient-number subgroup, 10 rounds of random resampling were performed from the oropharyngeal carcinoma (OPC) datasets to calculate the corresponding mean ICC values. The six radiomic features (RFs) demonstrating the greatest variations in mean ICC values across different sample sizes were selected to illustrate the sample size dependency of ICC metrics. Based on this analysis, a patient number that showed a stabilized trend in mean ICCs was identified and ultimately used for establishing the RF-RobustDB.

2.6 RF selection and PFS model development

To demonstrate the efficacy of the RF-RobustDB, six PFS models were constructed using RFs selected through different ICC thresholds: (1) ICC > 0.9, (2) ICC > 0.85, (3) ICC > 0.8, (4) ICC > 0.75, (5) ICC > 0.5, and (6) non-preselected RFs. PFS events were

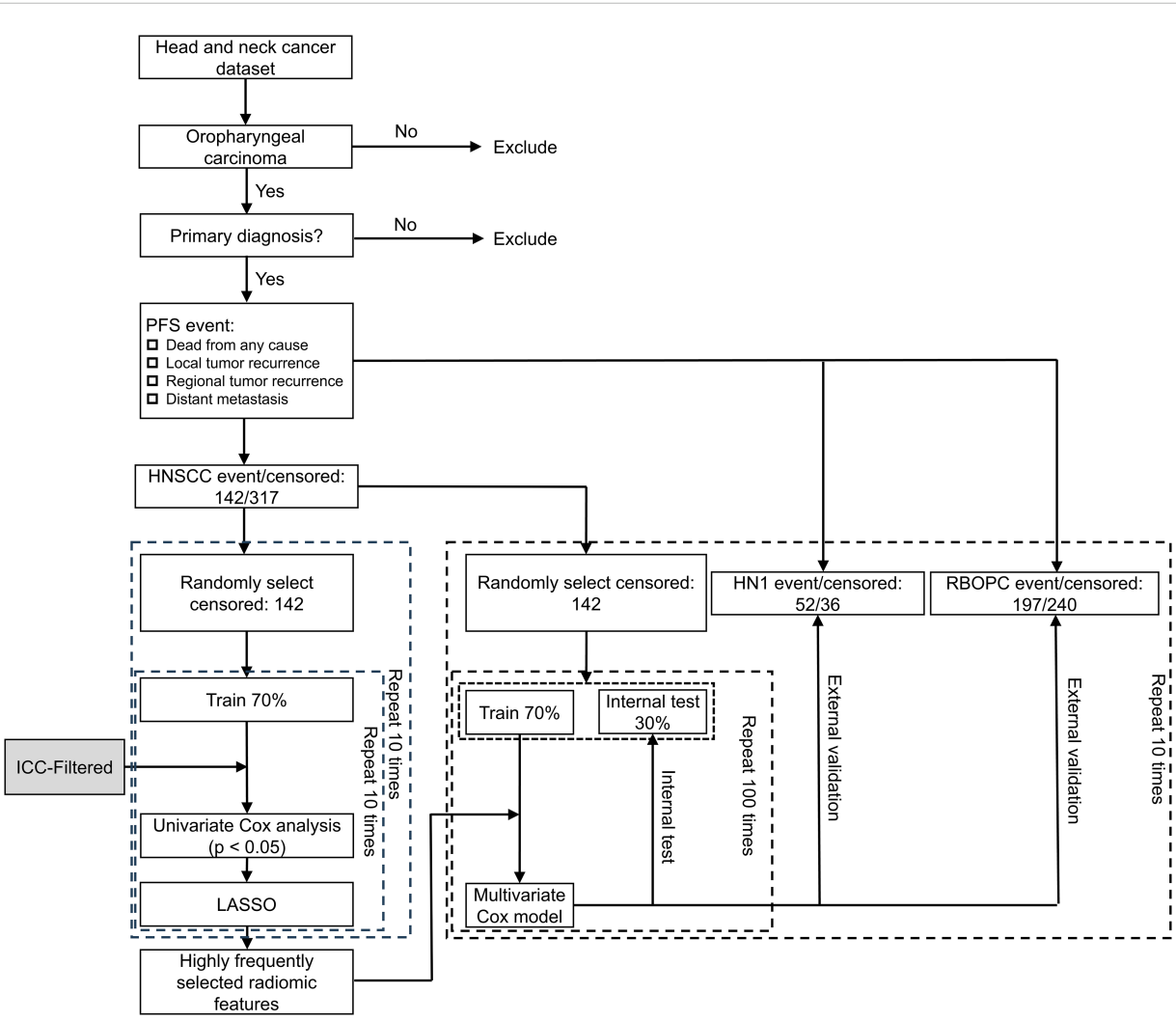


FIGURE 1
Workflow used for building progression-free survival (PFS) models based on CT images of oropharyngeal carcinoma patients. ICC = intraclass correlation coefficient; LASSO = least absolute shrinkage and selection operator; RF = radiomic feature; HNSCC = Head and Neck Squamous Cell Carcinoma; HN1 = HEAD-NECK-RADIOMICS-HN1; RBOPC = Radiomic Biomarkers in Oropharyngeal Carcinoma.

defined as local/regional recurrence, distant metastasis, or death from any cause. Figure 1 demonstrate the complete feature selection and modeling workflow. For the feature selection procedure, the highly reliable RFs derived from unperturbed images were initially chosen based on their robustness, as defined by the mean ICC. Subsequently, univariate Cox analysis was utilized to identify the RFs associated with PFS events within the pre-selected RFs in the training group. RFs with a p-value of less than 0.05 were considered significant. Finally, the least absolute shrinkage and selection operator algorithm was employed to select RFs with non-zero coefficients in the training group. The training groups were randomly bootstrapped 10 times from the 10 resampled balanced HNSCC dataset. The features that appeared frequently were selected to construct the PFS models.

To determine the optimal number of RFs for modeling, the relationship between the feature number and model performance was systematically investigated in the HNSCC dataset (Figure A1). The results revealed that the model constructed using non-preselected RFs exhibited optimal performance in the internal testing group when the feature number reached five (Figure A1 (A)). Specifically, the mean C index exhibited the highest value in the testing cohorts, indicating the superior predictive ability of the model at this feature threshold. Moreover, using a greater number of RFs resulted in larger mean C index gaps between the training and internal testing groups. Similar trends of optimal performance were observed in the other five experiments (Figure A1(B)-(F)). Therefore, the top five RFs that most frequently appeared in all experiments were ultimately selected to ensure that the feature number would not introduce conflicts into the final results. The selected RFs for each experiment are listed in Table A2.

Multivariate Cox regression was employed to model the survival risks for PFS in the HNSCC dataset. The performance of the developed PFS models was evaluated by concordance (C) index for the training, internal testing, and external validation (HN1, RBOPC) cohorts. The HNPET dataset was excluded from external validation due to insufficient follow-up data on local/regional recurrence and distant metastasis. To assess the robustness of the models, the mean C index and its 95% confidence intervals were calculated in 100 bootstrap experiments on the 10 randomly resampled balanced datasets.

2.7 Model generalizability assessment

To assess model generalizability across external validation (EV) cohorts, a generalizability index (G) that quantifies the absolute difference in C index values between training and EV groups. The G index is defined by Equation 2:

$$G = \sum_{m}^{M} \frac{\left| C_{Train} - C_{EV_m} \right|}{M} \quad (2)$$

where C_{Train} represents the C index of the trained model, C_{EV_m} is the C index for the m^{th} EV cohort, and M is the total number of EV cohorts. Lower G-index values indicate superior model generalizability, reflecting smaller performance discrepancies between training and validation datasets. The mean G index and corresponding 95% confidence intervals were calculated from 1,000 cross validation models to comprehensively evaluate the distribution of model generalizability performance.

3 Results

3.1 Patient-number dependence analysis

Figure 2 depicts the relationship between the number of patients and the mean ICCs of six RFs, which were selected as the top six RFs exhibiting the most significant variance changes with varying patient numbers. As the number of patients increased, the mean ICCs of the *Firstorder_Maximum* and *GLCM_ClusterTendency* features initially increased and then stabilized. In contrast, the mean ICCs of the remaining four selected features first fluctuated before eventually stabilizing. All six features tended to stabilize once the patient count reached 200. This stabilization trend of the mean ICCs demonstrates that the patient sample size used in this study was sufficient for constructing the RF-RobustDB. Specifically, 800 resampled patients were subjected to 100 iterations of resampling to compute the mean ICCs and their corresponding 95% confidence intervals.

3.2 Establishment and validation of the RF-RobustDB

3.2.1 The repeatability of shape-based RFs

Figure 3 presents a comprehensive evaluation of the ICCs for shape-based RFs across multiple simulated test conditions. Notably, all shape-based features demonstrated consistently high repeatability, with mean ICC values surpassing the 0.9 threshold (range: 0.955-0.999) across various perturbation scenarios. This robust performance supports the clinical applicability of shape-based RFs, as their measurements consistently reflect the tumor's shape characteristics and are less influenced by other clinical factors.

3.2.2 The repeatability of first-order and textural RFs

Figure 4 presents the ICCs of the first-order and textural RFs. The left panel of Figure 4 displays the mean ICCs and their 95% confidence intervals for each unfiltered RF. The right panel of Figure 4 gives the mean ICCs of the unfiltered, wavelet-filtered and LoG-filtered RFs. Collectively, Figures 3 and 4 establish the RF-RobustDB. Detailed information on the mean ICCs and their 95% confidence intervals can be found in the Appendix.

3.2.3 RF-RobustDB help improving the generalizability of PFS models

Figure 5A, B illustrates the selection of highly reproducible radiomic features (RFs) using the established RF-RobustDB. Figure 5A quantifies the absolute counts, while Figure 5B presents the relative proportions of these reproducible RFs across six feature

TABLE A2 Final selected radiomic features for each PFS model.

ICC threshold	Feature name
ICC>0	log-sigma-2-mm-3D_glszm_LowGrayLevelZoneEmphasis_30_binCount
	log-sigma-5-mm-3D_glszm_LowGrayLevelZoneEmphasis_30_binCount
	log-sigma-6-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis_30_binCount
	wavelet-LLH_firstorder_Skewness_30_binCount
	log-sigma-3-mm-3D_glszm_ZoneEntropy_30_binCount
ICC>05	wavelet-HHH_firstorder_Energy_30_binCount
	wavelet-LHH_firstorder_Energy_30_binCount
	log-sigma-5-mm-3D_firstorder_Energy_30_binCount
	log-sigma-4-mm-3D_firstorder_Energy_30_binCount
	wavelet-LLH_firstorder_Energy_30_binCount
ICC>075	wavelet-LHH_glrlm_GrayLevelNonUniformityNormalized_30_binCount
	wavelet-LHH_glcm_Idn_30_binCount
	log-sigma-6-mm-3D_glcm_ClusterShade_30_binCount
	original_glcm_Correlation_30_binCount
	wavelet-LLL_glcm_JointAverage_30_binCount
ICC>08	wavelet-LHH_glcm_Idn_30_binCount
	log-sigma-6-mm-3D_glcm_ClusterShade_30_binCount
	log-sigma-5-mm-3D_glszm_ZoneEntropy_30_binCount
	original_glcm_Correlation_30_binCount
	log-sigma-4-mm-3D_glszm_SizeZoneNonUniformity_30_binCount
ICC>085	wavelet-LHH_gldm_SmallDependenceEmphasis_30_binCount
	wavelet-LLL_glrlm_ShortRunHighGrayLevelEmphasis_30_binCount
	wavelet-LLL_firstorder_Range_30_binCount
	log-sigma-2-mm-3D_firstorder_90Percentile_30_binCount
	log-sigma-6-mm-3D_glrlm_RunLengthNonUniformity_30_binCount
ICC>09	log-sigma-1-mm-3D_firstorder_Maximum_30_binCount
	log-sigma-2-mm-3D_glszm_SizeZoneNonUniformity_30_binCount
	original_glcm_Correlation_30_binCount
	log-sigma-3-mm-3D_glszm_SizeZoneNonUniformity_30_binCount
	log-sigma-6-mm-3D_glcm_DifferenceVariance_30_binCount

classes: first-order, GLCM, GLRLM, GLSZM, GLDM, and NGTDM. Among 1,395 first-order and high-order RFs analyzed, 470 features (33.7%) demonstrated excellent repeatability, defined by mean ICCs > 0.9. Notably, unfiltered and LoG-filtered features exhibited significantly higher repeatability rates compared to wavelet-filtered features.

Using the established RF-RobustDB, six PFS models were systematically evaluated by recommended RFs at varying mean ICC thresholds. The C index in the training and internal testing groups first decreased and then increased as the ICC increased, as the red and black curve shown in Figure 5C. However, it is noteworthy that the EV cohorts (HN1 and RBOPC) demonstrated superior discriminative performance for models constructed using ICC preselected RFs, as shown by the pink and blue lines in Figure 5C. In addition, there was a smaller C index gap between the training and EV cohorts in the mean ICC preselected RF groups than in the non-preselected RF group, as shown by the G index trend in Figure 5D. The smallest mean G index was obtained in the RF groups preselected with a mean ICC of 0.5, and the largest mean G index was obtained in the non-preselected RFs groups. By taking into account the large C index and small G index of the RFs preselected with mean ICCs, it was found that the PFS models

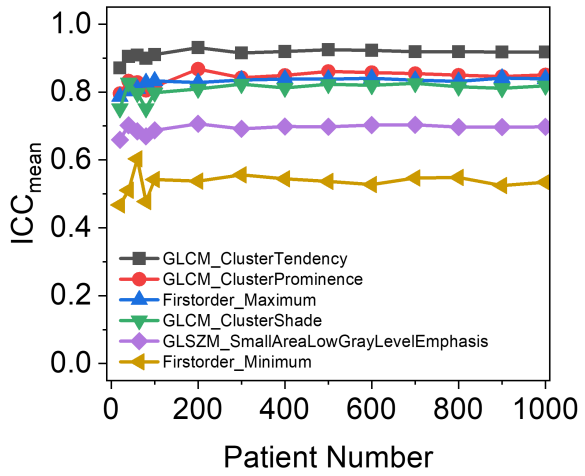


FIGURE 2
Patient-number dependence of mean intraclass correlation coefficients (ICCs) in six unfiltered radiomic features, which were selected based on the top six radiomic features with the largest variance in mean ICCs between different patient numbers.

incorporating preselected RFs demonstrated significantly higher generalizability compared to those using non-preselected RFs.

4 Discussion

4.1 Result analysis

A robust RF-RobustDB of pretreatment CT-derived RFs in OPC patients was established through image perturbation. This database effectively guides the preselection of repeatable RFs and enhances the generalizability of multi-cohort PFS studies. A new G index was introduced to quantitatively evaluate the generalizability of the constructed PFS models. The methodology developed in this study can be easily extended to other anatomical sites and imaging modalities, providing a feasible solution for establishing standardized RF-RobustDBs to comprehensively assess RF repeatability across various clinical scenarios.

Our analysis of mean ICC dependence on patient cohort size (Figure 2) revealed that the six selected unfiltered RFs showing the highest variance across different sample sizes achieved stabilization when the patient number exceeded 200. This suggests that a minimum of 200 patients provides sufficient data for reliable ICC-based assessment of RF robustness, confirming that the sample size in this study (1,274) ensured the reliability of the RF-RobustDB. The RF-RobustDB evaluation demonstrated significant differences in feature repeatability between filtering methods: only 11.7% of wavelet-filtered RFs exhibited excellent repeatability (ICC>0.9), compared to 56.3% of Laplacian-of-Gaussian (LoG)-filtered RFs. This substantial disparity (44.6%) establishes the superior robustness of LoG-filtered features. The low repeatability of wavelet-filtered RFs likely stems from the characteristics of wavelet filtering, image resampling strategies, and perturbation settings (6). Therefore, radiomic model construction requires more stringent selection criteria for wavelet-filtered RFs compared to their LoG-filtered counterparts.

4.2 RF-RobustDB reliability analysis

The clinical utility of the RF-RobustDB was evaluated through external validation using two independent cohorts derived from separate institutions. Importantly, these validation cohorts were

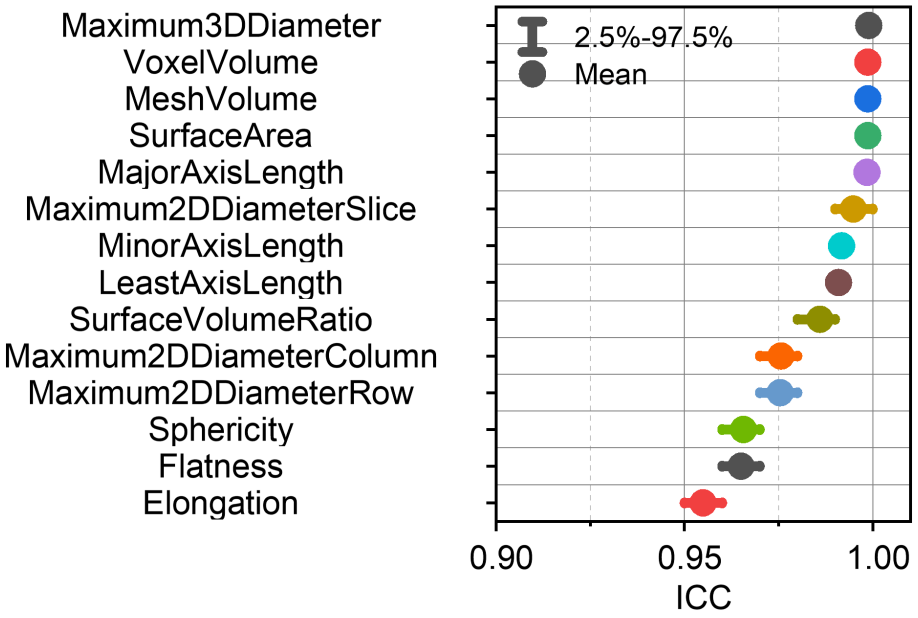


FIGURE 3
Mean intraclass correlation coefficients (ICCs) with 95% confidence intervals for each shape-based feature.

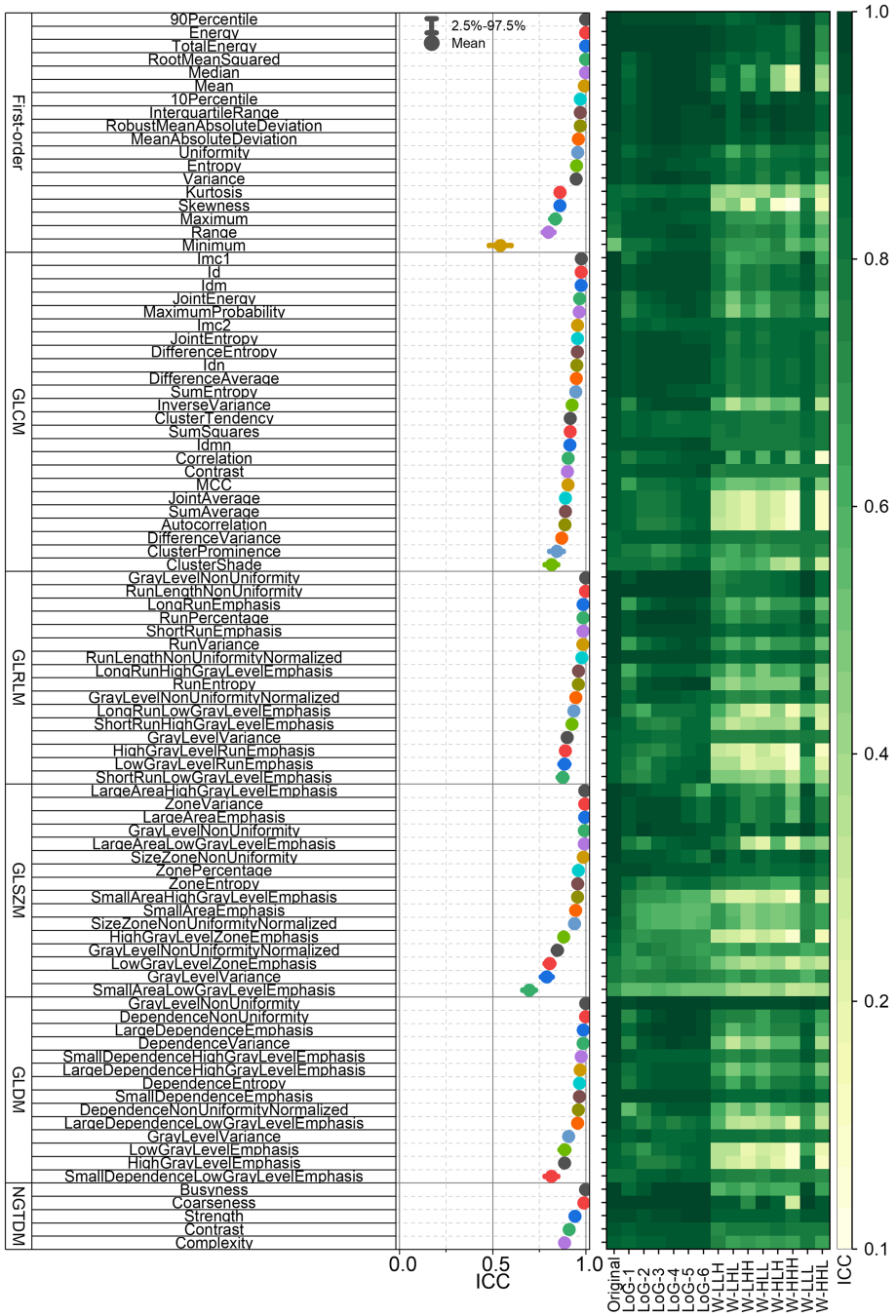


FIGURE 4
Intraclass correlation coefficients (ICCs) of first-order and textural radiomic features. GLCM = gray-level co-occurrence matrix; GLRLM = gray-level run-length matrix; GLSZM = gray-level size-zone matrix; GLDM = gray-level dependence matrix; NGTDM = neighboring gray-tone difference matrix.

exclusively utilized for feature selection and model training phases, thereby maintaining the integrity of the validation process. Comparative analysis revealed that models incorporating RF-RobustDB-preselected features demonstrated superior performance in external validation, as evidenced by the higher mean concordance indices compared to models using non-preselected features and a reduction in the performance gap between training and validation cohorts. Zhang and colleagues improved the generalizability of a disease-free survival model for head and neck cancer by pre-selecting highly reproducible RFs using the perturbation method (6). The study by Gong et al. also provides compelling evidence supporting the critical importance of

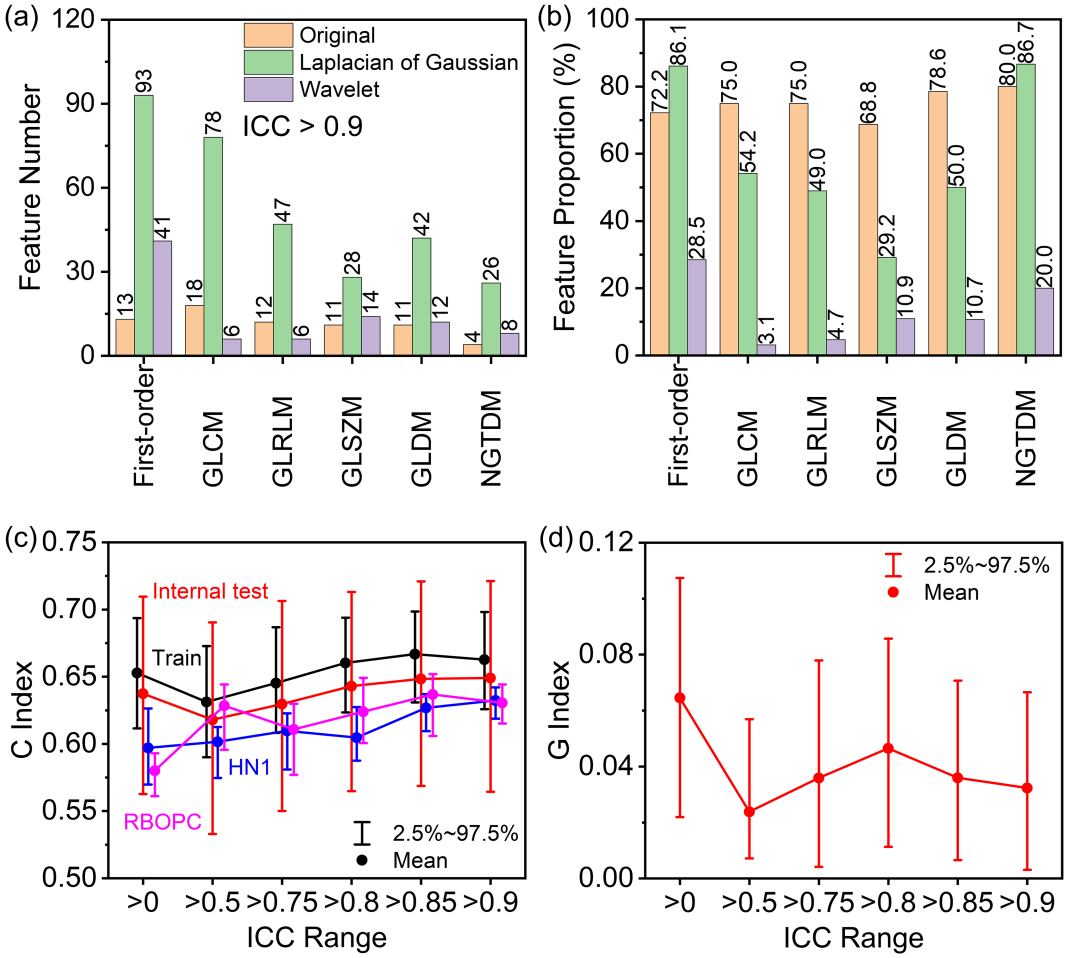


FIGURE 5
(A) Numbers and **(B)** proportions of excellent repeatable radiomic features in first-order, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighboring gray-tone difference matrix (NGTDM) features selected based on mean ICCs. **(C)** Performance of progression-free survival (PFS) models built using preselected highly repeatable radiomic features with different mean intraclass correlation coefficient (ICC) thresholds in the training (Train), internal testing (Test), and two external validation (HN1 and RBOPC) cohorts. **(D)** Generalizability index **(G)** of the PFS models across the training and EV groups based on the concordance **(C)** index gap.

feature stability in radiomic analyses (46). Through systematic perturbation analysis of CT-derived imaging features in esophageal squamous cell carcinoma, their findings substantiate that incorporating high-stability features can enhance model performance in the external validation set. Thomas Louis et al. reported that robust features demonstrated superior predictive potential compared to non-robust features in predicting the outcomes of an external validation dataset (47). These studies collectively highlight the critical importance of feature stability for model generalizability, demonstrating findings consistent with our own results. This convergence of evidence further substantiates the fundamental value of establishing comprehensive RF-RobustDB to support subsequent radiomics research.

The reproducibility of RFs in clinical practice is subject to multiple influencing factors, such as patient positioning variability, segmentation quality, the noise level of medical imaging devices, and the variations between the performance of devices depending on their model and vendor, leading to differences in CT number values (13). Additionally, the reproducibility of RFs can be affected by scanning parameters and reconstruction algorithms (16, 25). This multifactorial variability explains the observed reduction reliability and generalizability when applying RFs in multi-cohort studies versus single-cohort studies. However, unlike variations in CT scanners or scanning modalities, variations in patient positioning, segmentation, and random noises are similar in various clinical circumstances, which allows perturbation methods to be a universal tool for assessing RF repeatability. Therefore, employing image perturbation across multiple institutions is a feasible approach for evaluating the robustness of RFs, as demonstrated by the enhanced generalizability of our PFS models. Considering these factors, the methodology employed in this study is both feasible and applicable for establishing RF-RobustDBs for other tumor sites and imaging

modalities. The methodology is a promising approach for assessing the repeatability and enhancing the generalizability of radiomic models, thereby facilitating the development of more reliable and robust radiomic models with enhanced clinical translatability.

4.3 Existing limitations analysis

This study has several limitations that should be acknowledged. First, although our perturbation method simulated key variability sources including translations, rotations, random noise, and contour variations, they could not fully replicate all potential sources of variability encountered in clinical practice. For example, transient signal fluctuations that may occur during repeated scans under identical acquisition parameters cannot be effectively modeled. This inherent limitation underscores that image perturbation methods cannot entirely replace traditional test-retest validation approaches. Second, the optimal ICC thresholds for establishing reliable radiomic models remain controversial in the field. Additional investigations are required to establish evidence-based cutoff values for robust feature selection in clinical applications. Third, while our multi-institutional study design strengthened the generalizability of findings, the persistent effects of inter-scanner variability and acquisition parameter differences on feature reproducibility warrant further investigation. Furthermore, the dual use of our dataset for both RF-RobustDB construction and PFS model development may introduce circularity. Future validation should incorporate independent multi-institutional datasets to more rigorously assess the RF-RobustDB's clinical utility. Addressing these limitations through continued research will be essential for optimizing the RF-RobustDB's performance and expanding its applicability across diverse clinical implementations.

5 Conclusion

We have established a RF-RobustDB using an image perturbation approach for CT-derived RFs in OPC patients. The ICCs were calculated to quantify the reliability and repeatability of RFs. Through multi-cohort PFS experiments, we demonstrated the reference value of the RF-RobustDB, demonstrating that preselected highly repeatable RFs improved PFS model generalizability. To quantitatively assess model performance, we introduced a generalizability metric (G-index). The methodology we employed is cost-effective and easily applicable across different institutions, suggesting its potential extension to other lesion areas and imaging modalities. The comprehensive RF-RobustDB can facilitate robust RF selection when only small training datasets or single-institutional data are available, thereby enhancing the reliability, reproducibility, and generalizability of radiomic predictive models.

Data availability statement

The raw image date is available in the cancer image archive (https://www.cancerimagingarchive.net/collection/head-neck-radiomics-hn1/, https://www.cancerimagingarchive.net/collection/head-neck-pet-ct/, https://www.cancerimagingarchive.net/collection/hnscc/, https://www.cancerimagingarchive.net/collection/opc-radiomics/).

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

YW: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. AZ: Formal Analysis, Writing – review & editing. JZ: Data curation, Methodology, Software, Writing – review & editing. XT: Data curation, Methodology, Writing – review & editing. S-KL: Formal Analysis, Writing – review & editing. JiC: Software, Visualization, Writing – review & editing. ZM: Methodology, Visualization, Writing – review & editing. TZ: Visualization, Writing – review & editing. YZ: Visualization, Writing – review & editing. HG: Supervision, Writing – review & editing. JCa: Writing – review & editing, Conceptualization, Formal Analysis, Funding acquisition, Project administration, Resources, Supervision.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. The Mainland-Hong Kong Joint Funding Scheme (MHKJFS) (MHP/005/20) support evaluating the robustness of radiomic features, Shenzhen Basic Research Program (JCYJ20210324130209023) of Shenzhen Science and Technology Innovation Committee support development of the feature perturbation method, Project of Strategic Importance Fund (P0035421) and Projects of RISA (P0043001) of The Hong Kong Polytechnic University, and Health and Medical Research Fund (HMRF 09200576), the Health Bureau, the Government of the Hong Kong Special Administrative Region supports the study design, data collection, data analysis, and interpretation of the results.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1464884/full#supplementary-material

References

1. Zhang Y, Lam S, Yu T, Teng X, Zhang J, Lee FK, et al. Integration of an imbalance framework with novel high-generalizable classifiers for radiomics-based distant metastases prediction of advanced nasopharyngeal carcinoma. *Knowl Based Syst*. (2022) 235:107649. doi: 10.1016/j.knosys.2021.107649

2. Wang P, Wang X, Zhang M, Li G, Zhao N, Qiao Q. Combining the radiomics signature and HPV status for the risk stratification of patients with OPC. *Oral Dis*. (2024) 30:272–80. doi: 10.1111/odi.14386

3. Acciavatti RJ, Cohen EA, Maghsoudi OH, Gastounioti A, Pantalone L, Hsieh M-K, et al. Incorporating robustness to imaging physics into radiomic feature selection for breast cancer risk estimation. *Cancers*. (2021) 13:5497. doi: 10.3390/cancers13215497

4. Chetan MR, Gleeson FV. Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur Radiol*. (2021) 31:1049–58. doi: 10.1007/s00330-020-07141-9

5. Cheong E-N, Park JE, Park SY, Jung SC, Kim HS. Achieving imaging and computational reproducibility on mul-tiparametric MRI radiomics features in brain tumor diagnosis: phantom and clinical validation. *Eur Radiol*. (2023) 34:2008–23. doi: 10.1007/s00330-023-10164-7

6. Zhang J, Lam S-K, Teng X, Ma Z, Han X, Zhang Y, et al. Radiomic feature repeatability and its impact on prognostic model generalizability: A multi-institutional study on naso-pharyngeal carcinoma patients. *Radiol Oncol*. (2023) 183:109578. doi: 10.1016/j.radonc.2023.109578

7. Park JiE, Park SY, Kim HJ, Kim HoS. Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives. *Korean J Radiol*. (2019) 20:1124–37. doi: 10.3348/kjr.2018.0070

8. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: A systematic review on artificial intelligence and radiomics. *Eur J Nucl Med Mol Imaging*. (2019) 46:2656–72. doi: 10.1007/s00259-019-04372-x

9. Koçak B, Durmaz E.Ş, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: A practical guide for beginners. *Diagn Interv Radiol*. (2019) 25:485–95. doi: 10.5152/dir.2019.19321

10. Zhang Y-P, Zhang X-Y, Cheng Y-T, Li B, Teng X-Z, Zhang J, et al. Arti-ficial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. *Military Med Res*. (2023) 10:22. doi: 10.1186/s40779-023-00458-8

11. Chaddad A, Liang X. Stability of radiomic models and strategies to enhance reproducibility. *IEEE Trans Radiat Plasma Med Sci*. (2024) 8:540–55. doi: 10.1109/TRPMS.2024.3365778

12. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295:328–38. doi: 10.1148/radiol.2020191145

13. Chen Y, Zhong J, Wang L, Shi X, Lu W, Li J, et al. Robustness of CT radiomics features: consistency within and between single-energy CT and dual-energy CT. *Eur Radiol*. (2022) 32:5480–90. doi: 10.1007/s00330-022-08628-3

14. Rinaldi L, De Angelis SP, Raimondi S, Rizzo S, Fanciullo C, Rampinelli C, et al. Reproducibility of radiomic features in CT images of NSCLC patients: an integrative analysis on the impact of acquisition and reconstruction parameters. *Eur Radiol Exp*. (2022) 6:2. doi: 10.1186/s41747-021-00258-6

15. Pandey U, Saini J, Kumar M, Gupta R, Ingalhalikar M. Normative baseline for radiomics in brain MRI: evaluating the robustness, regional variations, and reproducibility on FLAIR images. *J Magn Reson Imaging*. (2021) 53:394–407. doi: 10.1002/jmri.27349

16. Muenzfeld H, Nowak C, Riedlberger S, Hartenstein A, Hamm B, Jahnke P, et al. Intra-scanner repeatability of quantitative imaging features in a 3D printed semi-anthropomorphic CT phantom. *Eur J Radiol*. (2021) 141:109818. doi: 10.1016/j.ejrad.2021.109818

17. Haniff NSM, Abdul Karim MK, Osman NH, Saripan MI, Che Isa IN, Ibahim MJ. Stability and reproducibility of radiomic features based various segmentation technique on MR images of hepatocellular carcinoma (HCC). *Diagnostics*. (2021) 11:1573. doi: 10.3390/diagnostics11091573

18. Tunali I, Hall LO, Napel S, Cherezov D, Guvenis A, Gillies RJ, et al. Stability and reproducibility of computed tomography radiomic features extracted from peritumoral regions of lung cancer lesions. *Med Phys*. (2019) 46:5075–85. doi: 10.1002/mp.13808

19. Wang H, Zhou Y, Wang X, Zhang Y, Ma C, Liu Bo, et al. Reproducibility and repeatability of CBCT-derived radiomics features. *Front Oncol*. (2021) 11:773512. doi: 10.3389/fonc.2021.773512

20. Jha AK, Mithun S, Jaiswar V, Sherkhane UB, Purandare NC, Prabhash K, et al. and reproducibility study of radiomic features on a phantom and human cohort. *Sci Rep*. (2021) 11:2055. doi: 10.1038/s41598-021-81526-8

21. Bologna .M, Tenconi C, Corino VDA, Annunziata G, Orlandi E, Calareso G, et al. Repeatability and reproducibility of MRI-radiomic features: A phantom experiment on a 1.5 T scanner. *Med Phys*. (2023) 50:750–62. doi: 10.1002/mp.16054

22. Fiset S, Welch ML, Weiss J, Pintilie M, Conway JL, Milosevic M, et al. Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiotherapy Oncol*. (2019) 135:107–14. doi: 10.1016/j.radonc.2019.03.001

23. Kocak B, Durmaz ES, Kaya OK, Ates E, Kilickesmez O. Reliability of single-slice–based 2D CT texture analysis of renal masses: influence of intra- and interobserver manual segmentation variability on radiomic feature reproducibility. *Am J Roentgenol*. (2019) 213:377–83. doi: 10.2214/AJR.19.21212

24. Traverso A, Wee L, Dekker A, Robert G. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys*. (2018) 102:1143–58. doi: 10.1016/j.ijrobp.2018.05.053

25. Escudero Sanchez L, Rundo L, Gill AB, Hoare M, Mendes Serrao E, Sala E. Robustness of radiomic features in CT images with different slice thickness, comparing liver tumour and muscle. *Sci Rep*. (2021) 11:8262. doi: 10.1038/s41598-021-87598-w

26. Li Y, Tan G, Vangel M, Hall J, Cai W. Influence of feature calculating parameters on the reproducibility of CT ra-diomic features: A thoracic phantom study. *Quant Imaging Med Surg*. (2020) 10:1775–85. doi: 10.21037/qims-19-921

27. Scalco E, Belfatto A, Mastropietro A, Rancati T, Avuzzi B, Messina A, et al. T2w-MRI signal nor-malization affects radiomics features reproducibility. *Med Phys*. (2020) 47:1680–91. doi: 10.1002/mp.14038

28. Euler A, Laqua FC, Cester D, Lohaus N, Sartoretti T, Pinto dos Santos D, et al. Virtual monoen-ergetic images of dual-energy CT—Impact on repeatability, reproducibility, and classification in radiomics. *Cancers*. (2021) 13:4710. doi: 10.3390/cancers13184710

29. Zwanenburg A, Leger S, Agolli L, Pilz K, Troost EGC, Richter C, et al. Assessing robustness of radiomic features by image perturbation. *Sci Rep*. (2019) 9:614. doi: 10.1038/s41598-018-36938-4

30. Teng X, Zhang J, Zwanenburg A, Sun J, Huang Y, Lam S, et al. Building reliable radiomic models using image perturbation. *Sci Rep*. (2022) 12:10035. doi: 10.1038/s41598-022-14178-x

31. Teng X, Zhang J, Ma Z, Zhang Y, Lam S, Li W, et al. Improving radiomic model reliability using robust features from perturbations for head-and-neck carcinoma. *Front Oncol*. (2022) 12:974467. doi: 10.3389/fonc.2022.974467

32. Zhang J, Teng X, Zhang X, Lam S-K, Lin Z, Liang Y, et al. Comparing effectiveness of image perturbation and test retest imaging in improving radiomic model reliability. *Sci Rep*. (2023) 13:18263. doi: 10.1038/s41598-023-45477-6

33. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. (2013) 26:1045–57. doi: 10.1007/s10278-013-9622-7

34. Koo TK, Li MYA. Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012

35. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. (2014) 5:4006. doi: 10.1038/ncomms5006

36. Wee L, Dekker A. Data from HEAD-NECK-RADIOMICS-HN1 [Data set. *Cancer Imaging Arch*. (2019). doi: 10.7937/tcia.2019.8kap372n

37. Kwan JYY, Su J, Huang SH, Ghoraie LS, Xu W, Chan B, et al. Radiomic biomarkers to refine risk models for distant metastasis in HPV-related oropharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. (2018) 102:1107–16. doi: 10.1016/j.ijrobp.2018.01.057

38. Kwan JYY, Su J, Huang SH, Ghoraie LS, Xu W, Chan B, et al. Data from radiomic biomarkers to refine risk models for distant metastasis in oropharyngeal carcinoma. *Cancer Imaging Arch*. (2019). doi: 10.7937/tcia.2019.8dho2gls

39. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep*. (2017) 7:10117. doi: 10.1038/s41598-017-10371-5

40. Vallières M, Kay-Rivest E, Perrin LéoJ, Liem X, Furstoss C, Khaouam N, et al. Data from head-neck-PET-CT. *Cancer Imaging Arch*. (2017). doi: 10.7937/K9/TCIA.2017.8oje5q00

41. Grossberg AJ, Mohamed ASR, Elhalawani H, Bennett WC, Smith KE, Nolan TS, et al. Imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. *Sci Data*. (2018) 5:180173. doi: 10.1038/sdata.2018.173

42. Elhalawani H, Mohamed ASR, White AL, Zafereo J, Wong AJ, Berends JE, et al. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. *Sci Data*. (2017) 4:170077. doi: 10.1038/sdata.2017.77

43. Grossberg A, Elhalawani H, Mohamed A, Mulder S, Williams B, White AL, et al. HNSCC [Dataset. *Cancer Imaging Arch*. (2020). doi: 10.7937/k9/tcia.2020.a8sh-7363

44. Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep*. (2015) 5:11075. doi: 10.1038/srep11075

45. van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339

46. Gong J, Wang Q, Li J, Yang Z, Zhang J, Teng X, et al. Using high-repeatable radiomic features improves the cross-institutional generalization of prognostic model in esophageal squamous cell cancer receiving definitive chemoradiotherapy. *Insights into Imaging*. (2024) 15:239. doi: 10.1186/s13244-024-01816-3

47. Louis T, Lucia F, Cousin F, Mievis C, Jansen N, Duysinx B, et al. Identification of CT radiomic features robust to acquisition and segmentation variations for improved prediction of radiotherapy-treated lung cancer patient recurrence. *Sci Rep*. (2024) 14:9028. doi: 10.1038/s41598-024-58551-4

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact