A privacy-preserving, on-board satellite image classification technique incorporating homomorphic encryption and transfer learning

Abhijit Roy[1], Mahendra Kumar Gourisaria[2], Rajdeep Chatterjee[2], Amitkumar V. Jha[3]*, Bhargav Appasani[3], Nicu Bizon[4]* and Alin Gheorghita Mazare[4]

[1]IIIT, Guwahati, Assam, India, [2]School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India, [3]School of Electronics Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India, [4]The National University of Science and Technology POLITEHNICA Bucharest, Piteşti University Centre, Pitesti, Romania

Satellite image classification is an important and challenging task in the modern technological age. Satellites can capture images of danger-prone areas with very little effort. However, the size and number of satellite images are very high when they are rapidly captured from space, and they require a huge amount of memory to store the data. In addition, keeping the satellite images private is another important task for security purposes. On-board, instant, accurate classification of a smaller number of satellite images is a challenging task, which is important to determine the specific condition of an area for instant monitoring. In the proposed hybrid approach, the captured images are kept secure, while the required training of the classification is done separately. Finally, the trained module is encrypted for use by the satellite to perform the on-board classification task. The Brakerski–Fan–Vercauteren (BFV)-based homomorphic encryption of EuroSAT satellite images is applied to store images in a cloud storage, where the privacy of the images can be maintained. Then, the decrypted images are used for training four transfer learning models (YOLOv8, YOLOv12, ResNet34, ResNet101, and a vision transformer classification. The best-trained module is encoded and encrypted again by using homomorphic encryption to limit the module to authorized devices. The encrypted module is decrypted and decoded to recover the trained module, which is used for instant classification of test images. Finally, the performance of the transfer learning models is evaluated from the test results. The vision transformer classifier achieved the highest accuracy of 99.65%.

KEYWORDS

Brakerski–Fan–Vercauteren scheme, transfer learning, land use land cover classification, homomorphic encryption, privacy preservation

1 Introduction

The majority of the land on Earth is occupied by humans for different purposes. Unoccupied land is a natural resource that could be used for the development of society (i.e., for industry, infrastructure development, and so on) (Zhao et al., 2020). In the modern technological age, satellite images taken from the orbit of the Earth can be used to detect land occupation by classifying land types (Sicre et al., 2020). A satellite image dataset is very important for satellite image classification, and the proper classification model also plays a very important role. As satellite images are taken from great distances, image resolution and clarity are dependent on the available light spectrum, the camera used on the satellite, and other parameters (Zhao et al., 2020). Due to the limitations of the features in satellite images, a large number of images are required for proper classification accuracy. It is a challenging task to train the classification network with a large set of images with limited features.

Several satellite image datasets are available in the literature, including Formosat-2 (Sicre et al., 2020), WorldView-2 (Pan et al., 2020), WorldView-3 (Deur et al., 2020), ImageNet (Giorgiani do Nasc et al., 2020), and more. Each dataset contains specific images. These datasets are created in the visible light spectrum, the IR spectrum, and so on. In addition, different datasets contain different types of images, such as roads, rivers, and urban scenes. Satellite images with specific features can be used to detect specific events. Vegetation classification (Zhao et al., 2020), crop identification (Sicre et al., 2020), unplanned city settlements (Pan et al., 2020), tree species classification, forest management (Deur et al., 2020), airbus ship detection (Giorgiani do Nasc et al., 2020), and post-hurricane building damage identification (Cao and Choe, 2020) are some of the situations that have been monitored based on specific satellite images.

As each dataset has a limited number of image classes, it is necessary to use different datasets for image classification. However, this increases the number of images needed to train the classification network. The large number of images requires huge amounts of memory. To counteract such a challenge, the number of images in each class should be reduced for the multiclass classification. As each dataset has different features, a good preprocessing technique is also required to standardize the datasets. In addition, it is very challenging to classify all the classes at the same time. So, a deep learning classification network can be used to overcome these challenges.

In this work, we reduced the EuroSAT database and created a customized dataset. We encrypted those images and stored them in a cloud storage for memory management. Then, the stored dataset was loaded into the proposed system, and the decrypted images were used for transfer learning-based classification. The classification-trained module of the transfer learning models was then once again encoded, encrypted, and stored in the cloud storage. Subsequently, the stored module was loaded, decrypted, and decoded, and then used for testing image classification. Finally, the test results were used for the performance evaluation of the proposed system. By doing this, we save a significant amount of satellite memory, and at the same time, we implement greater security. In short, we developed an on-board instant classification module for satellites.

The rest of the article is organized as follows:

After presenting an introduction in Section 1, related work is included in Section 2, where research gaps and novel contributions are also presented. Section 3 elaborates on the materials and methods for the proposed study. The proposed system is presented in Section 4. The results with a detailed discussion, including comparative analysis, are presented in Section 5. Finally, the conclusion and future directions of the present work are included in the last section.

2 Related works and contributions

Some of the related works in the context of the proposed study are reviewed in this section. The literature review covers the objective, methodology, results, and the research gaps on which the contributions of the present article are outlined.

2.1 Related works

In 2020, Saralioglu and Gungor (2022) used WorldView-2 multispectral satellite images for the classification of land use and land cover (LULC). Forests, shadows, hazelnuts, tea, soils, roads, and buildings were classified in the study. Images were segmented based on this classification. In addition, the performance of the studied SS-CNN model, IKONOS, Deimos-2, and Pleiades image datasets was compared. In this study, the CNN model was compared with random forest (RF) and support vector machine (SVM) architectures. The authors achieved an accuracy of 95.6% for the 3D-2D CNN model, of 89.2% for the RF model, and of 86.4% for the SVM model.

Rahman et al. (2020) used Sentinel-2, Landsat 8, and Planet image datasets for LULC classification. The RF, SVM, and stacked algorithms were used for the LULC classification. In this study, the authors achieved an overall accuracy of 0.969 and 0.983 along with overall kappa (κ) values of 0.948 and 0.968, respectively.

In 2021, Akshay et al. (2020) studied a CNN-based algorithm for the detection of unused land in images. In this work, image segmentation and classification of 20 images from six datasets with different classes were studied. The datasets used were QuickBird, IKONOS, SkySat-2, Smeggie, ISRO, and satellite images (from a search engine). Here, thresholding segmentation was used to segment the images. In addition, different transfer learning methods, such as AlexNet, LeNet-5, ResNet, and multi-SVM, were used to classify satellite images. The highest F1-score achieved was 0.9037.

In 2021, Zhang T. et al. (2021) used Sentinel-2 satellite image data for urban land cover classification, with L2 band images used for this purpose. In this study, crop fields, trees, roads, buildings, and water were classified using SVM and RF (Bayesian network) algorithms. The authors achieved an overall accuracy of 0.4605 for SVM, 0.8744 for random forest, and 0.8788 for optimized Random Forest. Alkhelaiwi et al. (2021) studied a privacy-preserving deep learning algorithm in which satellite image data were encrypted before training. Then, the encrypted data were used to train the classification network. Here, Paillier-based partial homomorphic encryption was studied. For image classification, a CNN-based custom network, Vgg16, Xception,

ResNet50, and DenseNet21 were used to classify the classes of vegetation, bare soil, road, and urban images. The authors achieved 0.9384 accuracy for the plain dataset and 0.9092 accuracy for the encrypted dataset. Yassine et al. (2021) studied a satellite dataset for LULC observation. The EuroSAT and Sentinel-2 datasets were used for classification purposes. In this study, a CNN-based classification model was used, and an overall accuracy of 0.9958 was reported. Thiagarajan et al. (2021) used the SAT4, SAT6, and EuroSAT datasets. A hierarchical framework and ensemble learning (HFEL), along with optimal feature selection, were used for image identification. For HFEL, CNN-based custom models such as AlexNet, LeNet-5, and ResNet were used for image classification. Correlation coefficient-based gravitational search algorithm-based feature extraction was also performed. The extracted features were used with the SVM algorithm for image classification. The authors achieved 0.9999 overall accuracy with this method. Zhang C. J. et al. (2021) used GMS1–5, GEO 9, MTSAT-1R, MTSAT-2, and Himawari 8 satellite images for tropical cyclone intensity classification. A tropical cyclone intensity classification module followed by a tropical cyclone intensity estimation module was studied. A CNN-based regression model was used for the classification module, and four TC intensity estimation models were evaluated, with the highest RMSE of 8.60.

In 2022, Basheer et al. (2022) used satellite images for LULC classification. Landsat 8, Planet, and Sentinel-2 datasets were used. SVM, RF, maximum likelihood, minimum distance, classification, and regression tree models were used for classification. The performance of the models was evaluated based on overall accuracy. The authors achieved an overall accuracy of 0.89 with the Landsat dataset, 0.91 with the Sentinel dataset, and 0.94 with the Planet dataset. Plakman et al. (2021) used Sentinel-1 and Sentinel-2 image datasets for solar park detection to monitor PV installations. The images were segmented using non-iterative clustering-based segmentation, and subsequently, RF was used to classify the images. For the performance measurement, overall accuracy, F1-score, and IoU were measured, with an 0.9997 overall accuracy. Tanim et al. (2022) used Sentinel-1 satellite images for flood detection to ensure the safety of pedestrians, damage control, and lifelines. In the study, supervised learning (i.e., RF, SVM, and maximum likelihood) and unsupervised learning (i.e., change detection that contains the Otsu algorithm, iso-clustering, and fuzzy rules methods) were used for classification. The authors achieved 0.69, 0.87, 0.83, and 0.87 accuracy for the RF, SVM, MLC, and CD models, respectively. In 2023, Kaselimi et al. (2022) used the Planet dataset of satellite images for deforestation monitoring. A vision transformer-based classification ForestViT model was studied. The performance of the model was compared with VGG, DenseNet, ResNet, and MobileNet transfer learning models. Precision and recall were evaluated for performance measurement. The research team achieved 0.80, 0.77, 0.78, 0.75, and 0.74 precision scores for ForestViT, ResNet50, VGG16, DenseNet121, and MobileNet, respectively.

Shang et al. (2023) used the LSCIDMR dataset for meteorological image classification. A channel–dilation–concatenation network was studied, and the network was compared with AlexNet, SqueezeNet, VGG16, ResNet18, ShuffleNetv2, and MobileNetv3. Accuracy and F1-score were used to evaluate performance measurement. The authors achieved 0.9356 accuracy. Ouchra et al. (2023) used the Landsat 8 dataset for a land usage-related case study in Morocco. The study analyzed SVM, classification and regression trees, RF, minimum distance, gradient tree, and decision tree for the dataset. The overall accuracy of the models was evaluated. The researchers achieved an improved overall accuracy of 0.93. Tarasiou et al. (2023) studied a temporo-spatial vision transformer for time series processing of satellite images. The study was compared with the literature by using overall accuracy. The authors achieved an overall accuracy of 0.95 for segmentation and 0.947 for object classification.

In 2024, Le (2024) studied satellite images for Earth observation. The classification process was performed on board by using MobileViTv2 and EfficientViT-M2 models. The study was compared with CNN- and ResNet-based models by evaluating precision, recall, accuracy, and specificity. The author achieved 0.9876 accuracy for the improved model. PushpaRani et al. (2024) used the MBSRC satellite-based dataset to extract geological information using a U-Net architecture. Precision and F1-score were evaluated. They achieved 0.95 precision.

Dhande et al. (2024) used real-time data from Google Earth Engine for crop analysis. The TRSAITL model was studied for real-time images. The model was compared with HCNN, MSRPS, and CNN TSS from the literature. Precision and recall were plotted in the study, which achieved 0.978 accuracy. Shendy and Nalepa (2024) used on-board satellite image classification. Real-world satellite image classification was performed based on OPS-SAT data, which was operated by the European Space Agency. An ensemble learning-based, data-centric, and model-centric technology was studied. The authors advanced the field of satellite image categorization despite the limitations of nanosatellite operations by illuminating efficient model training techniques and highlighting the complex issues present in deep learning for real-world Earth observation.

Zhang et al. (2025) improved tree species classification based on satellite images. The Spatiotemporal Entropy-based Change Resistance Filter (STECR-F) algorithm was used as the central algorithm of this study's lightweight spatiotemporal categorization system. The STECR-F algorithm incorporates the idea of spatiotemporal entropy (STE) and reduces classification uncertainty using weighted spatiotemporal neighborhood information. The performance of STECR-F was thoroughly assessed in this work from three perspectives: STE, transfer change, and classification accuracy, and it was contrasted with alternative approaches. The total accuracy of STECR-F was improved to 0.9135. All things considered, the STECR-F algorithm handles the uncertainty and interannual dynamics in tree species categorization findings well. Chaturvedi et al. (2025) studied a fire-smoke detection model by using the UTSC SmokeRS and IIITDMJ Smoke datasets. A classification was conducted to detect fire smoke for normal weather conditions, and under conditions such as fog, storms, clouds, hurricanes, and snow. A multi-attention network was studied, which is made up of a CNN and a vision transformer network. For performance evaluation, accuracy, precision, recall, F1-score, and FAR were evaluated. They achieved a 0.9022 best accuracy. A comprehensive summary of the related work is presented in Table 1, focusing on different key aspects.

TABLE 1 Summary of literature review.

Reference	Objective	Database used	ML methods	Performance parameters
Saralioglu and Gungor (2022)	LULC classification	WorldView-2	CNN, SVM, and RF	Overall and average accuracy
Rahman et al. (2020)	LULC classification	Sentinel-2, Landsat 8, and Planet	RF, SVM, and stacked algorithm	OA
Akshay et al. (2020)	Unused land detection	Landsat, EuroSAT, QuickBird, IKONOS, SkySat-2, Smeggie, and the ISRO dataset	CNN, AlexNet, LeNet-5, ResNet, and multi SVM	Precision, recall, F2-score, and accuracy
Zhang et al. (2021a)	Urban land cover classification	Sentinel-2	SVM and RF	OA
Alkhelaiwi et al. (2021)	Privacy-preserving deep learning	SPOT satellite dataset	CNN, VGG16, Xception, ResNet50, and DenseNet21	Accuracy
Yassine et al. (2021)	LULC observation	EuroSAT and Sentinel-2	CNN	Accuracy
Thiagarajan et al. (2021)	Satellite image classification	SAT4, SAT6, and EuroSAT	CNN, AlexNet, LeNet-5, ResNet, and SVM	Accuracy, precision, and recall
Zhang et al. (2021b)	Tropical cyclone intensity classification	Himawari 8 dataset	CNN	MAE and RMSE
Basheer et al. (2022)	LULC classification	Landsat 8, Sentinel-2, and Planet	SVM and RF	OA
Plakman et al. (2022)	Solar park detection	Sentinel-1 and Sentinel-2	RF	OA
Tanim et al. (2022)	Flood detection	Sentinel-1	RF, SVM, MLE, the Otsu algorithm, iso-clustering, and fuzzy rules	Accuracy, recall, precision, and F1-score
Kaselimi et al. (2022)	Deforestation monitoring	Planet dataset	VGG, DenseNet, ResNet, and MobileNet	Precision and recall
Shang et al. (2023)	Meteorological image classification	LSCIDMR dataset	Channel–dilation–concatenation network, AlexNet, SqueezeNet, VGG16, ResNet18, ShuffleNetv2, and MobileNetv3	Accuracy and F1-score
Ouchra et al. (2023)	Land usage detection	Landsat 8	SVM and regression trees, RF, minimum distance, gradient tree, and decision tree	Accuracy
Tarasiou et al. (2023)	Time series satellite image processing	Satellite images	Temporo-spatial Vision Transformer	OA
Le (2024)	On-board Earth observation	Satellite	MobileViTv2 and EfficientViT-M2	Precision, recall, accuracy, and specificity
PushpaRani et al. (2024)	Extraction of geological information	MBSRC dataset	U-net	Model precision
Dhande et al. (2024)	Crop analysis	Google Earth Engine dataset	TRSAITL, HCNN, MSRPS, and CNN TSS	Precision and recall
Shendy and Nalepa (2024)	On-board classification	OPS-SAT images	Ensemble learning	Accuracy
Zhang et al. (2025)	Tree species classification	Sentinel-1 and Sentinel-2	STECR-F algorithm	OA and F1-score
Chaturvedi et al. (2025)	Fire-smoke detection	UTSC SmokeRS and IIITDMJ Smoke dataset	CNN and ViT	Accuracy, precision, recall, F1-score, and FAR
Proposed work	HE (BFV scheme), TL for the hybrid approach, and cloud-based on-board instant monitoring	EuroSAT dataset	YOLOv8, YOLOv12, ResNet34, ResNet101, and ViT	Accuracy, precision, recall, specificity, and F1-score

2.2 Research gaps

Satellite images have been used for many different purposes. Various satellite images contain different information regarding trees, forests, industry, urban areas, rural areas, natural resources, and so on. Based on image features, different event monitoring tasks were performed in the literature. Research has been conducted on satellite images for monitoring and detection of challenging scenarios, where the images were taken from a satellite to understand the situation. However, none of the available research focuses on security aspects. To the best of the authors' knowledge, we have not only covered the security aspects but also improved significantly the memory requirements. Furthermore, the proposed system is time-efficient compared to existing research, which is a very important aspect for mission-critical real-time applications requiring minimal delay.

FIGURE 1
Dataset image sample from the 10 classes.
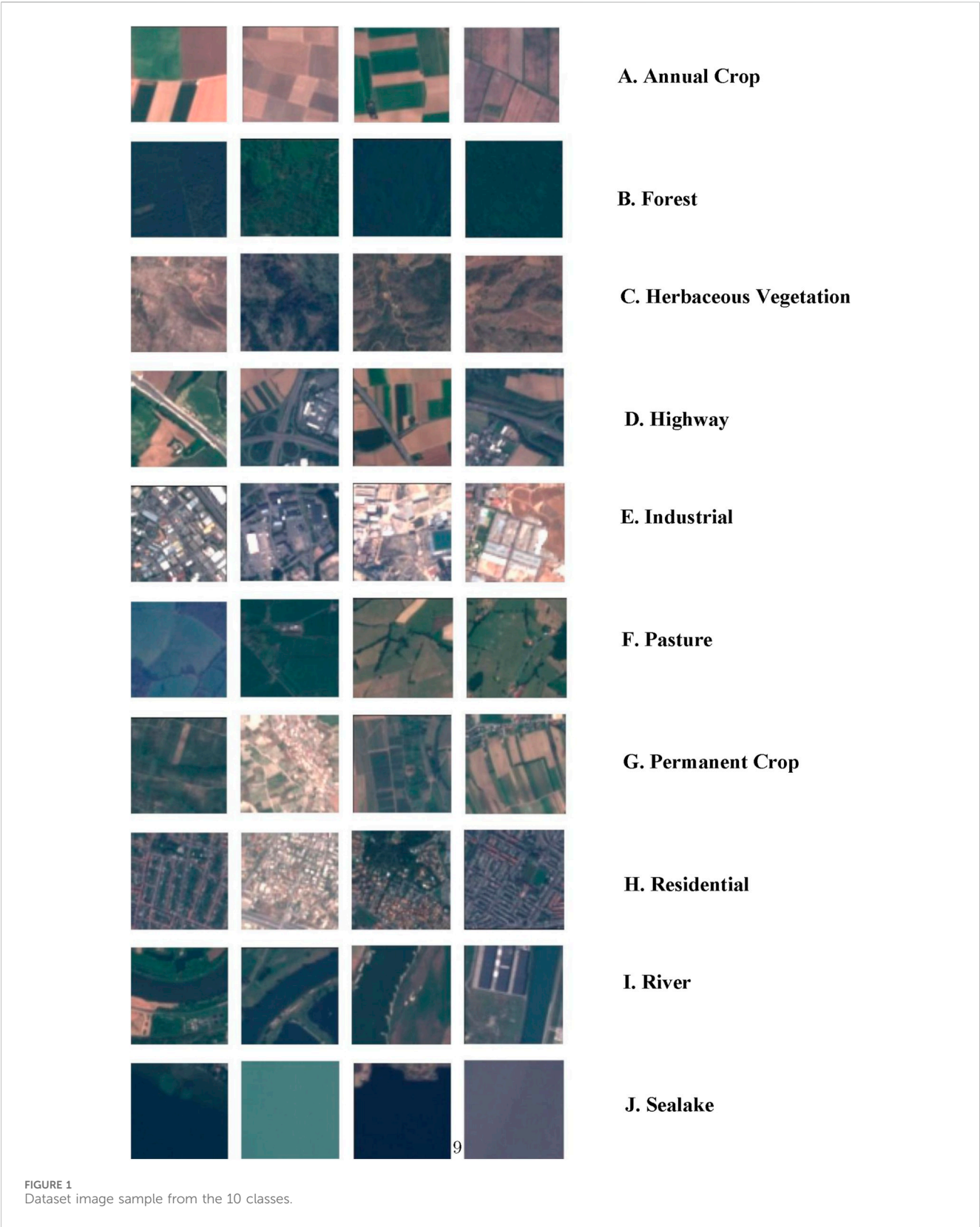
2.3 Contribution

The present work aims to fill the research gap in the existing literature through the proposal of an efficient and secure image classification module for satellite images, which combines image encryption, transfer learning-based LULC classification, and the cloud-based, instant application of the encrypted trained module. The key contributions are enumerated below:

- The proposed module uses the cloud to store satellite images in a dynamically varying environment.
- The proposed module applies homomorphic encryption (HE) (Brakerski–Fan–Vercauteren (BFV) scheme) to satellite images to secure the privacy of a nation while monitoring various situations.
- The proposed module employs trained module creation for instant classification of unknown situations.
- We propose an encoded HE (BFV scheme) method to keep trained classification modules secure from unauthorized personnel.
- The proposed module is well-suited for use as an on-board instant classification module.

3 Materials and methods

3.1 Dataset description

The EuroSAT dataset was created using the Sentinel-2 satellite imagery database, which is taken from Kaggle and has a ground sampling of 10 m. The dataset contains images from 13 bands, including the RGB band, four red edge bands, aerosol, two shortwave infrared bands, water vapor, near-infrared (NIR), and cirrus. The dataset contains a total of 27,000 images divided into 10 classes. These classes include sea, river, forest, pasture, herbaceous vegetation, annual crop, permanent crop, residential areas, highway, and industrial areas. Consequently, the dataset can be used for multiclass classification. For our study, we created a subset of the RGB images from the EuroSAT database. Of the 27,000 images, our customized dataset contains only 2,000 images with 10 classes, with each class containing 200 images. For training and validation purposes, we selected a total of 1,800 images (90%), and for testing purposes, 200 (10%) images were selected. For the classification model, the image size was kept at 200 × 200. The sample images from 10 classes are shown in Figure 1. The link to the actual dataset (referred to in Kaggle) and our own customized dataset are provided in the Data Availability section.

3.2 Methods

In this section, the different methods used in the study are described.

3.2.1 Encryption and decryption
- BFV scheme: Homomorphic encryption (HE) is an outstanding technology in the field of encryption for cloud-based data preservation. The HE is divided into two subbranches, labeled HE (LHE) and fully HE (FHE). The BFV encryption is an LHE encryption scheme. The scheme consists of a plaintext space, which includes plaintext messages. The ciphertext space consists of encrypted messages. These depend on the polynomial degree parameters. The secret and public keys are generated by key-generating functions. An encrypted function is used to encrypt, and the decrypted function is used to decrypt the

encrypted messages by using the secret key (Al Badawi et al., 2019; Clet et al., 2021). In this study, BFV-based HE was applied to encrypt and decrypt an RGB image and a trained module.

For the plaintext space, $\mathbb{Z}[X]/(X^N + 1)$, where p is an integer that represents a plaintext modulus and N represents the polynomial degree, which is a power of 2. Here, q is the ciphertext modulus. The cryptosystem is represented by

- Secret Key generation: From a random distribution of the subspace $\mathbb{Z}[X]/(X^N + 1)$, the secret key s is generated.
- Public Key generation: From a uniform distribution and error distribution over $\mathbb{Z}[X]/(X^N + 1)$, the elements a and e are generated, respectively. Here, the public key is represented by $pk = ([-a \times s + e]_q), a = (p_0, p_1)$.
- Encryption: Given a message $\mu \in \mathbb{Z}[X]/(X^N + 1)$, let u, e_1, e_2 be small errors. The encrypted message $c = ([p_0 \times u + e_1 + \Delta\mu]_q, [p_1 \times u + e_2]_q) = (c_0, c_1)$.
- Decryption: The message can be retrieved using $[[\lfloor \frac{p \times [c_0 + c_1 \times s]_q}{q} \rceil]]_p$. Here, for proper operation, the noise must be lower than $\Delta/2$ (Clet et al., 2021).
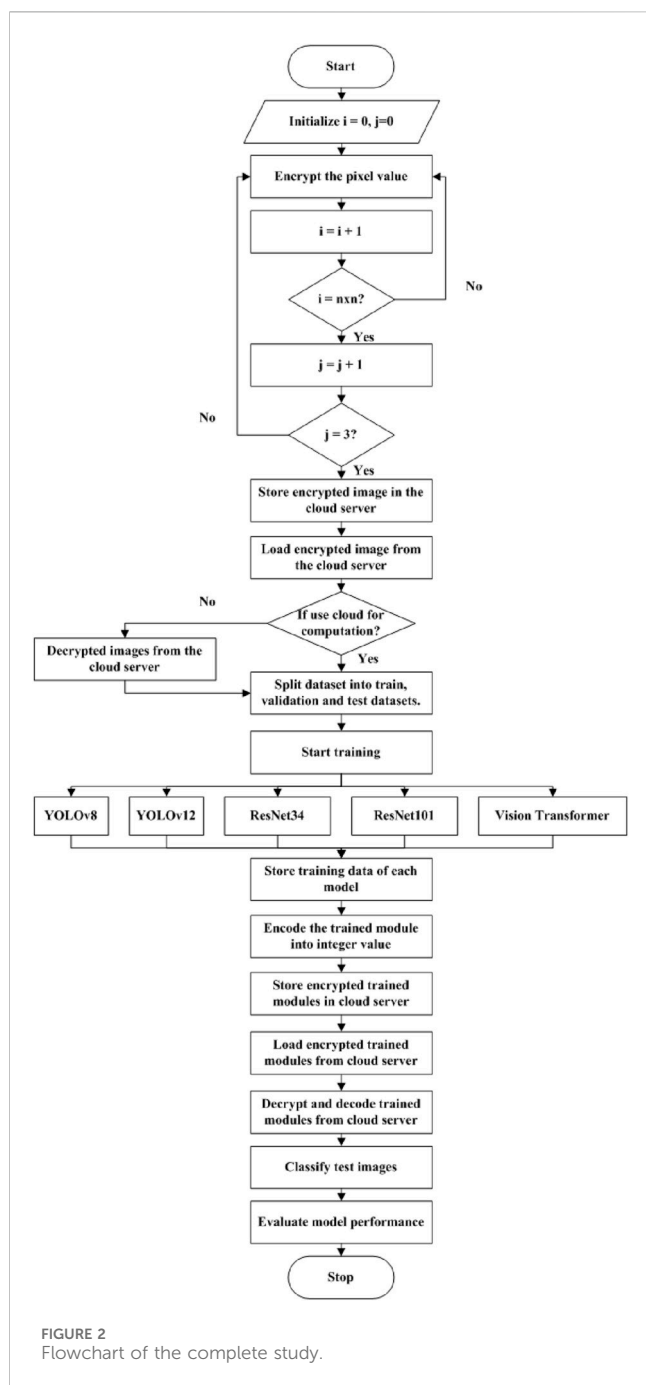
3.2.2 Classification models
LULC classification requires a large number of images for each class, which increases the memory requirement for the appropriate classification. Classifying images with less training data, a complex classification module with several layers is required. However, the complexity of the classification module will also increase the computational memory requirement. Therefore, it is challenging to train a classification model with fewer images in each class and with fewer layers to classify the images accurately. Some of the transfer learning models, which are capable of classifying images with less training data, are presented below (Li et al., 2022; Barbelian et al., 2021).

- ResNet: The deep residual network is an ensemble of several shallow networks, instead of a single deep learning network. For a residual ensemble unit j, if y_{j-1} is the input and $f(j)$ is trainable d dimensional convolution stages, w_j is the trainable weight parameter, then the output of the ensemble unit j is defined as Equation 1.

$$y_j \equiv f_j(y_{j-1}, w_j) + y_{j-1}. \tag{1}$$

The pre-activation section contains the batch normalization and the rectified linear units within the convolution network. Different versions of ResNet contain different numbers of sub-network ensembles. These shallow networks add effective depth for feature extraction from images during training. Thus, an increased number of shallow networks increases the training performance (Wu et al., 2019). Here, the ResNet34 and ResNet101 versions of the ResNet model were used.

- Vision transformer: The vision transformer model, which is used for language processing, also works for image classification. The classification model breaks an image into several tokens of the same sequence with a fixed length.

FIGURE 2
Flowchart of the complete study.

Multiple transformer layers are used on the tokens to train the model to recognize global relationships in image classification. The classification model generates the structure for an image by recursively generating an aggregate of each neighboring token to a particular token, which models the surrounding tokens by modifying the local structure. The token length of the model is reduced by the process, which in turn reduces the parameter count of the classification model (Yuan et al., 2021).

- YOLO: The YOLO model is a modified architecture of the CSPDarknet53 network model. This CNN-based model is built based on cross-stage partial connections, which perform the transmission enhancement of information

to each layer inside the network. A spatial pyramid pooling faster module is used here with multiple layers of CNN. The YOLO consists of unsampled layers that increase resolution for feature extraction. Bounding box and class probability predictions are made by analyzing feature maps with CNN layers, followed by a linear layer. Finally, the high-dimensional features are used to detect bounding boxes and predict object classes (Sohan et al., 2024). Here, YOLOv8 and YOLOv12 are used for the YOLO model.

4 Proposed methods

The satellite image dataset contains sensitive information that must be kept secure from unauthorized access. The EuroSAT dataset is public, but for the proof of concept of security-based event monitoring, we have customized this dataset by applying encryption. The customized dataset images were encrypted first. Then, the dataset was classified by different transfer learning techniques after decrypting the images. Then the trained module was encrypted again for use by authorized officials and to prevent illegal access. Finally, the decrypted trained module was used to test the test images.

4.1 Work flow diagram

Figure 2 shows the flowchart of the study, which consists of four sections. First, the images were encrypted from the created dataset. Next, the encrypted images were stored in a cloud storage. Then, the stored data were retrieved, decrypted, and used for the LULC classification by four transfer learning models. Next, the trained modules were encoded and encrypted. Subsequently, the trained module was stored in a cloud server, which is decrypted and decoded for the application. Finally, the trained modules were used to classify the test images. In this study, the retriever module consists of a decrypted algorithm that decrypts the images and a trained module for further processing. Finally, the performance of each module was evaluated. The details of the different technique modules are as follows.

4.2 Image encryption and decryption

An image is a set of three matrices, which are the R, G, and B matrices that form the image. Each integer in each matrix is a part of a pixel. Pixel intensity depends on three values. The encryption of each value can ensure the security of the pixel data. An image can be represented by $n \times n$, for a square image. Each color image contains $3n^2$ integer values that can encrypt the image. The methods of image encryption and decryption followed are presented in Algorithm 1. This algorithm represents the encryption of a square RGB image, where the image is represented by three matrices. Each matrix contains an integer value for each pixel, which is encrypted by the BFV scheme. The encryption time cost is 0.27 s per image.

```
Require: An image with n × n dimensions.
Ensure: The image is an RGB color image.
    1. Generate the ternary polynomial secret key
       Sk ∈-1,0,1 and a pair of polynomial public
       keys (Pk,Pk1) = (-[a,Sk + e]q,a) where a,Sk ←— Re
       and e ←— χerror.
    2. Generate (l + 1) pairs of evaluation keys
       Ek = (wi S2 - [ai,Sk + ei]q,ai) for o ≤ i ≤ l
    3. Generate a tuple by Sk,Pk and Ek
    4. For each n × n matrix, the plaintext message
       mj ∈ R for 1 ≤ j ≤ n2 generate ciphertext
       ctj = ([Δmj + Pk[0]u + e1]q,Pk[1]u + e2]q), where
       Δ = q/t. Do the same for all three matrices.
    5. Generate mj = [t/q[ctj[0] + ctj[1]Sk]q]t
```

Algorithm 1. Encryption and decryption of the RGB color image.

4.3 LULC classification

The encrypted images for each class were loaded and decrypted for classification using transfer learning methods. Here, no preprocessing method was used as the data are clean. However, it is recommended to have suitable preprocessing techniques based on the data for achieving better results. The classification training methods used in this study are YOLOv8, YOLOv12, ResNet34, ResNet101, and the vision transfer classification model. After training, a trained module was generated for each classification model for further processing.

4.4 Trained module encryption and decryption

A trained module of a classification training model is a file that is used for the classification of a certain type of training. Large amounts of memory and time are required for training. After the trained module is generated, it is used without the requirement of a large amount of memory. As the module contains very important data for certain types of classification, it is important to keep it private from unauthorized personnel. For this purpose, the Encoded HE (BFV) scheme was used. Algorithm 2 illustrates the process of encrypting the trained module. The algorithm encrypts the matrix generated by the trained module. The matrix is encrypted by using the BFV scheme.

```
Require: A matrix.
Ensure:   The  values  of  the  matrix  are
alphanumeric values.
    1. Encode the alphanumeric values as
       numeric values.
    2. Generate the ternary polynomial secret key
       Ske ∈-1,0,1 and a pair of polynomials public
       key     (Pke0,Pke1) = (-[a,Ske + e]q,a)     where
       a,Ske ←— Re and e ←— χerror.
    3. Generate (l + 1) pairs of evaluation keys
       Eke = (wi S2 - [ai,Ske + ei]q,ai) for o ≤ i ≤ l
    4. Generate a tuple by SkePke and Eke
```

```
    5. For each n × 1 matrix, the plaintext message
       mj ∈ R for 1 ≤ j ≤ n generate ciphertext
       ctj = ([Δmj + Pke[0]u + e1]q,Pke[1]u + e2]q) where
       Δ = q/t. Do the same for all three matrices.
    6. Generate mj = [t/q[ctj[0] + ctj[1]Ske]q]t
    7. Decode  the  mj  and  generate  the
       trained module.
```

Algorithm 2. Encoded Encryption and Decryption of the trained module.

4.5 Classification model testing

A total of 200 images different from the training and validation images were taken from all 10 classes. The regenerated trained modules were used to classify the test images from all 10 classes.

5 Results and discussion

5.1 Performance analysis

Different classification models were considered for the performance evaluation of the proposed system. To measure the efficacy of the proposed system under different models, a set of standard indicators was considered, which included confusion matrices, accuracy, precision, sensitivity, specificity, and F1-score (Bengio et al., 2017).

The results of the proposed system for different transfer learning models, that is, YOLOv8, YOLOv12, ResNet34, ResNet101, and vision transformer classifier, are reported in Table 2. Table 2 shows that the vision transformer classifier achieved the highest result in terms of sensitivity, specificity, precision, accuracy, and F1-score. Table 3 shows the classwise performance of the vision transformer classifier, where accuracy, precision, sensitivity, specificity, and F1-score are evaluated for each class.

Figure 3 shows a graphical representation of all the transfer learning models, that is, YOLOv8, YOLOv12, ResNet34, ResNet101, and the Vision Transformer classifiers. The vision transformer model achieved a sensitivity of 0.9801, a specificity of 0.9983, a precision of 0.9850, an accuracy of 0.9965, and an F1-score of 0.9825 and performed as the best model.

Confusion matrices were generated after classifying the test images for each model. This matrix compares the predicted value against the actual value. Figure 4 shows the confusion matrix for the vision transformer model-based classification, which is the best model. The testing of the vision transformer model produces a testing result divided into 10 different classes. Each class consists of 20 images, some of which were mispredicted to another class. The class-wise prediction of each class is shown in Figure 4. Figure 5 shows the receiver operating characteristic (ROC) curve of the vision transformer model.

5.2 Comparative analysis

Table 2 shows that, of the four classification networks, the Vision Transformer model performs best in terms of all

TABLE 2 Performance evaluation of the transfer learning models.

Sl. no.	Classification models	Sensitivity	Specificity	Precision	Accuracy	F1-score
1	YOLOv8	0.9500	0.9939	0.9452	0.9895	0.9476
2	YOLOv12	0.9700	0.9972	0.9749	0.9945	0.9724
3	ResNet34	0.9552	0.9956	0.9600	0.9915	0.9576
4	ResNet101	0.9602	0.9961	0.9650	0.9926	0.9626
5	Vision transformer classifier	0.9801	0.9983	0.9850	0.9965	0.9825

TABLE 3 Class-wise performance evaluation of the vision transformer classification model.

Sl. no.	Class	Accuracy	Precision	Sensitivity	Specificity	F1-score
1	Annual crop	1	1	1	1	1
2	Forest	1	1	1	1	1
3	Herbaceous vegetation	0.9950	1	0.9500	1	0.9744
4	Highway	0.9900	0.9091	1	0.9889	0.9524
5	Industrial area	1	1	1	1	1
6	Pasture	1	1	1	1	1
7	Permanent crop	0.9950	0.9523	1	0.9945	0.9756
8	Residential area	1	1	1	1	1
9	River	0.9900	1	0.9000	1	0.9474
10	Sea lake	1	1	1	1	1

FIGURE 3
Performance of different transfer learning models.
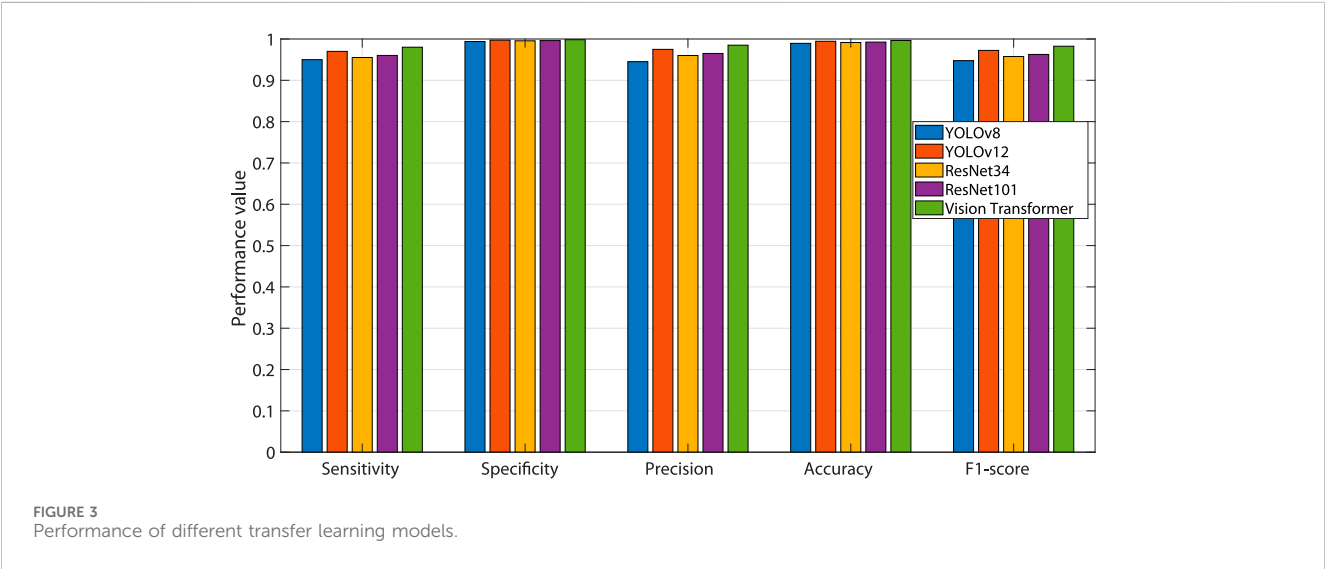
parameters, that is, sensitivity, specificity, precision, accuracy, and F1-score. Therefore, according to the structure of transfer learning models, token-based and shallow network-based models are more efficient than CNN-based models with large layers. The token and shallow network-based models work independently of other layers, whereas in the case of the CNN models, each layer depends directly on other layers. Due to the network architecture, the vision transformer classifier is efficient and effective.

We know that often, natural calamities (i.e., wildfires, earthquakes, floods, tsunamis, soil erosion, water level increment, and so on) are the main causes of national

FIGURE 4
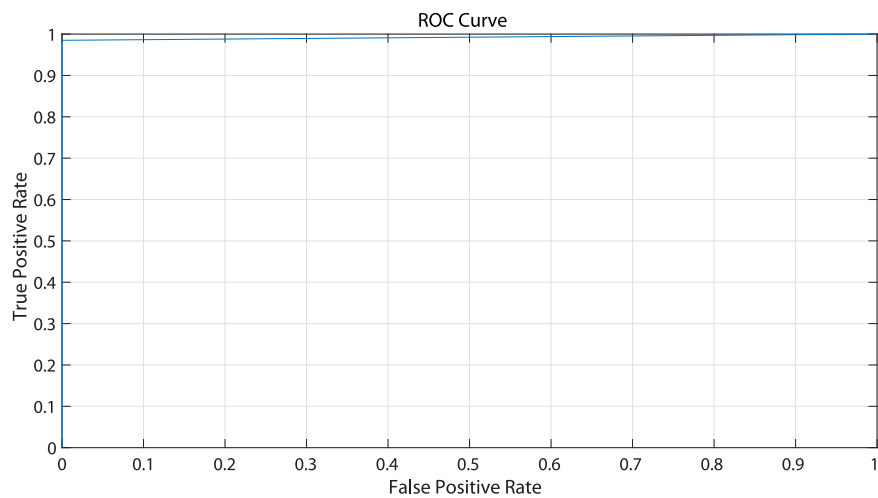Confusion matrix of the vision transformer classification.

FIGURE 5
Receiver operating characteristics (ROC) of the vision transformer classifier.

disasters. LULC detection plays a very important role in detecting unauthorized activity. Consequently, it leads to the following challenges:

- The data must be secure as it contains very sensitive information, which could lead to dire consequences if accessed by unauthorized users.
- It is necessary to detect a disaster or unusual activity instantly. This requires minimal delay.
- As the number of satellites for disaster management is limited, it is important to develop a satellite capable of performing several classifications with a very small amount of memory. This requires the proposed system to be memory-efficient.

These requirements are challenging, but our proposed system overcomes all these challenges when compared with the existing literature. Moreover, the system performance is also better than that of existing work, as discussed below.

In the literature, segmentation and classification were performed for specific event monitoring in the majority of cases. Of those, LULC and unused land coverage classification were addressed by several authors (Saralioglu and Gungor, 2022; Rahman et al., 2020; Akshay et al., 2020; Zhang T. et al., 2021; Yassine et al., 2021; Basheer et al., 2022). Another author used RF, SVM, MLE, and CNN-based machine learning algorithms in addition to AlexNet, LeNet, ResNet, VGG16, Xception, and DenseNet21 transfer learning methods. Alkhelaiwi et al. (2021) introduced image encryption for privacy-preserving deep learning by using the Paillier scheme. However, the model was partially homomorphic and was insufficient for the on-board classification task. Later, Le (2024) and Shendy and Nalepa (2024) introduced an on-board image classification, but it was not found to be a secure model. For the true on-board memory-efficient model, cloud dependencies are also required. In addition, the security of the image data is also highly important for cloud-based memory usage. Numerically, the work done by Saralioglu and Gungor (2022) achieved an accuracy of 0.956, and the work by Rahman et al. (2020) achieved an overall accuracy of 0.983. Akshay et al. (2020) achieved an F1-score of 0.9037, and the work done by Zhang T. et al. (2021) achieved an accuracy of 0.8788. The work by Yassine et al. (2021) achieved an overall accuracy of 0.9958, and the work by Basheer et al. (2022) achieved an accuracy of 0.94. The accuracies achieved by Alkhelaiwi et al. (2021), Le (2024), and Shendy and Nalepa (2024) were 0.9384, 0.9876, and 0.8400, respectively.

In our study, all these drawbacks are solved by our hybrid architecture, with the captured satellite images stored in cloud memory in an encrypted form, and the classification models are also trained by authorized personnel who have the secret key to decrypt the images. After training the transfer learning modules, the best-trained module is encrypted and stored in the cloud. The satellite can access the trained module after decrypting the module, and finally, the satellite can perform on-board classification tasks with very little memory requirement in an efficient way. In this process, the task is performed by authorized personnel and the satellite itself in a very secure way, ensuring savings in memory requirements and minimal delay.

6 Conclusion and future work

In this work, transfer learning models were proposed that are memory and computationally efficient. In the proposed architecture, every data element (i.e., images and trained modules) was encrypted and used by the BFV scheme of the HE, which required less built-in memory, while privacy was also maintained as an important factor. The encrypted trained modules were then used for testing purposes, replacing the requirement of the inbuilt training module. The proposed module can be used on any satellite for the instant detection of LULC and other important classifications for monitoring the Earth from space. In our study, the vision transformer-based classification model performed the best of all transfer learning models. The simulation results show that the vision transformer model achieves classification accuracy, precision, sensitivity, specificity, and F1-score of 99.65%, 98.5%, 98.01%, 99.83%, and 98.25%, respectively. This validates the efficacy of the proposed system. The results of this study can be extended to any specific application-oriented study involving different classification tasks, such as disaster management and national security monitoring purposes.

In the future, we could implement this work in real-life situations, where cloud storage and an expert laboratory on Earth would use real-life satellite data. The proposed model can be used for the observation of flood-, earthquake-, forest fire-, and cyclone-affected areas. Border activity can also be observed for national security purposes. In the future, the model will be investigated, and different performance parameters will be evaluated, including latency and detailed error analysis, while also considering atmospheric disturbance.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Author contributions

AR: Conceptualization, Data curation, Methodology, Writing – original draft. MG: Formal Analysis, Project administration, Software, Validation, Writing – original draft. RC: Data curation, Formal Analysis, Project administration, Writing – original draft. AJ: Supervision, Visualization, Writing – review and editing. BA: Formal Analysis, Visualization, Writing – review and editing. NB: Formal Analysis, Funding acquisition, Investigation, Project administration, Writing – review and editing. AM: Funding acquisition, Investigation, Supervision, Writing – review and editing.

Funding

National University of Science and Technology POLITEHNICA, Bucharest.

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 authors 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.

References

Akshay, S., Mytravarun, T. K., Manohar, N., and Pranav, M. A. (2020). "Satellite image classification for detecting unused landscape using CNN," in *2020 international conference on electronics and sustainable communication systems* (IEEE: ICESC), 215–222.

Al Badawi, A., Polyakov, Y., Aung, K. M. M., Veeravalli, B., and Rohloff, K. (2019). Implementation and performance evaluation of RNS variants of the BFV homomorphic encryption scheme. *IEEE Trans. Emerg. Top. Comput.* 9 (2), 941–956. doi:10.1109/tetc.2019.2902799

Alkhelaiwi, M., Boulila, W., Ahmad, J., Koubaa, A., and Driss, M. (2021). An efficient approach based on privacy-preserving deep learning for satellite image classification. *Remote Sens.* 13 (11), 2221. doi:10.3390/rs13112221

Barbelian, M. A., Cornel, D. I. N. U., and Venera, C. (2021). Deep learning approach on shark attack risk assessment using real-time autonomous surveillance systems. *UPB Sci. Bull. Ser. D.* 83. Available online at: https://www.scientificbulletin.upb.ro/rev_docs_arhiva/fulldb9_373484.pdf.

Basheer, S., Wang, X., Farooque, A. A., Nawaz, R. A., Liu, K., Adekanmbi, T., et al. (2022). Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques. *Remote Sens.* 14 (19), 4978. doi:10.3390/rs14194978

Bengio, Y., Goodfellow, I., and Courville, A. (2017). *Deep learning*. Cambridge, MA: MIT press, 23–24.

Cao, Q. D., and Choe, Y. (2020). Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks. *Nat. Hazards* 103 (3), 3357–3376. doi:10.1007/s11069-020-04133-2

Chaturvedi, S., Thakur, P. S., Khanna, P., Ojha, A., Song, Y., and Awange, J. L. (2025). Satellite image-based surveillance and early wildfire smoke detection using a multiattention interlaced network. *IEEE Trans. Industrial Inf.* 21, 3806–3815. doi:10.1109/tii.2025.3528549

Clet, P. E., Stan, O., and Zuber, M. (2021). "BFV, CKKS, TFHE: which one is the best for a secure neural network evaluation in the cloud?," in *Applied cryptography and network security workshops: ACNS 2021 satellite workshops, AIBlock, AIHWS, AIoTS, CIMSS, Cloud S&P, SCI, SecMT, and SiMLA, Kamakura, Japan, June 21–24, 2021, proceedings* (Springer International Publishing), 279–300.

Deur, M., Gašparović, M., and Balenović, I. (2020). Tree species classification in mixed deciduous forests using very high spatial resolution satellite imagery and machine learning methods. *Remote Sens.* 12 (23), 3926. doi:10.3390/rs12233926

Dhande, A. P., Malik, R., Saini, D., Garg, R., Jha, S., Nazeer, J., et al. (2024). Design of a high-efficiency temporal engine for real-time spatial satellite image classification using augmented incremental transfer learning for crop analysis. *SN Comput. Sci.* 5 (5), 585. doi:10.1007/s42979-024-02939-6

Giorgiani do Nascimento, R., and Viana, F. (2020). *Satellite image classification and segmentation with transfer learning.* Aiaa Scitech 2020 Forum, 1864.

Kaselimi, M., Voulodimos, A., Daskalopoulos, I., Doulamis, N., and Doulamis, A. (2022). A vision transformer model for convolution-free multilabel classification of satellite imagery in deforestation monitoring. *IEEE Trans. Neural Netw. Learn. Syst.* 34 (7), 3299–3307. doi:10.1109/tnnls.2022.3144791

Le, T. D. (2024). On-board satellite image classification for earth observation: a comparative study of pre-trained vision transformer models. *arXiv Prepr. arXiv: 2409.03901.*

Li, Z., Xue, M., Sun, Q., Liu, C., Guo, Q., Wang, F., et al. (2022). Pedestrian attribute recognition based on multi-task deep learning and label correlation analysis. *UPB Sci. Bull. Ser. C* 84, 53–70.

Ouchra, H. A. F. S. A., Belangour, A., and Erraissi, A. L. L. A. E. (2023). Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: morocco case study. *IEEE Access* 11, 71127–71142. doi:10.1109/access.2023.3293828

Pan, Z., Xu, J., Guo, Y., Hu, Y., and Wang, G. (2020). Deep learning segmentation and classification for urban village using a worldview satellite image based on U-Net. *Remote Sens.* 12 (10), 1574. doi:10.3390/rs12101574

Plakman, V., Rosier, J., and Van Vliet, J. (2022). Solar park detection from publicly available satellite imagery. *GIScience Remote Sens.* 59 (1), 462–481. doi:10.1080/15481603.2022.2036056

PushpaRani, K., Roja, G., Anusha, R., Dastagiraiah, C., Srilatha, B., and Manjusha, B. (2024). "Geological information extraction from satellite imagery using deep learning," in *2024 15th international conference on computing communication and networking technologies (ICCCNT)* (IEEE), 1–7.

Rahman, A., Abdullah, H. M., Tanzir, M. T., Hossain, M. J., Khan, B. M., Miah, M. G., et al. (2020). Performance of different machine learning algorithms on satellite image classification in rural and urban setup. *Remote Sens. Appl. Soc. Environ.* 20, 100410. doi:10.1016/j.rsase.2020.100410

Saralioglu, E., and Gungor, O. (2022). Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network. *Geocarto Int.* 37 (2), 657–677. doi:10.1080/10106049.2020.1734871

Shang, S., Zhang, J., Wang, X., Wang, X., Li, Y., and Li, Y. (2023). Faster and lighter meteorological satellite image classification by a lightweight channel-dilation-concatenation net. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 16, 2301–2317. doi:10.1109/jstars.2023.3243915

Shendy, R., and Nalepa, J. (2024). Few-shot satellite image classification for bringing deep learning on board OPS-SAT. *Expert Syst. Appl.* 251, 123984. doi:10.1016/j.eswa.2024.123984

Sicre, C. M., Fieuzal, R., and Baup, F. (2020). Contribution of multispectral (optical and radar) satellite images to the classification of agricultural surfaces. *Int. J. Appl. Earth Observation Geoinformation* 84, 101972. doi:10.1016/j.jag.2019.101972

Sohan, M., Sai Ram, T., and Rami Reddy, C. V. (2024). "A review on yolov8 and its advancements," in *International conference on data intelligence and cognitive informatics* (Singapore: Springer), 529–545.

Tanim, A. H., McRae, C. B., Tavakol-Davani, H., and Goharian, E. (2022). Flood detection in urban areas using satellite imagery and machine learning. *Water* 14 (7), 1140. doi:10.3390/w14071140

Tarasiou, M., Chavez, E., and Zafeiriou, S. (2023). "Vits for sits: vision transformers for satellite image time series," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10418–10428.

Thiagarajan, K., Manapakkam Anandan, M., Stateczny, A., Bidare Divakarachari, P., and Kivudujogappa Lingappa, H. (2021). Satellite image classification using a hierarchical ensemble learning and correlation coefficient-based gravitational search algorithm. *Remote Sens.* 13 (21), 4351. doi:10.3390/rs13214351

Wu, Z., Shen, C., and Van Den Hengel, A. (2019). Wider or deeper: revisiting the resnet model for visual recognition. *Pattern Recognit.* 90, 119–133. doi:10.1016/j.patcog.2019.01.006

Yassine, H., Tout, K., and Jaber, M. (2021). Improving lulc classification from satellite imagery using deep learning–eurosat dataset. *Int. Archives Photogrammetry, Remote Sens. Spatial Inf. Sci.* 43, 369–376. doi:10.5194/isprs-archives-xliii-b3-2021-369-2021

Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z. H., et al. (2021). "Tokens-to-token vit: training vision transformers from scratch on imagenet," in *Proceedings of the IEEE/CVF international conference on computer vision*, 558–567.

Zhang, T., Su, J., Xu, Z., Luo, Y., and Li, J. (2021a). Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier. *Appl. Sci.* 11 (2), 543. doi:10.3390/app11020543

Zhang, C. J., Wang, X. J., Ma, L. M., and Lu, X. Q. (2021b). Tropical cyclone intensity classification and estimation using infrared satellite images with deep learning. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 14, 2070–2086. doi:10.1109/jstars.2021.3050767

Zhang, B., Wang, Z., Liang, B., Dong, L., Feng, Z., He, M., et al. (2025). A lightweight spatiotemporal classification framework for tree species with entropy-based change resistance filter using satellite imagery. *Int. J. Appl. Earth Observation Geoinformation* 138, 104449. doi:10.1016/j.jag.2025.104449

Zhao, F., Wu, X., and Wang, S. (2020). Object-oriented vegetation classification method based on UAV and satellite image fusion. *Procedia Comput. Sci.* 174, 609–615. doi:10.1016/j.procs.2020.06.132