remote sensing

MDPI

Article

First Agriculture Land Use Map in Vietnam Using an Adaptive Weighted Combined Loss Function for UNET++

Ta Hoang Trung [1,2], Nguyen Vu Ky [1], Duong Cao Phan [3,4], Duong Binh Minh [1], Ho Nguyen [5,6] and Kenlo Nishida Nasahara [7,*]

[1] Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8572, Ibaraki, Japan; s2436063@u.tsukuba.ac.jp (T.H.T.); s2326060@u.tsukuba.ac.jp (N.V.K.); s2426036@u.tsukuba.ac.jp (D.B.M.)
[2] Department of Survey, Mapping and Geoinfo Viet Nam, Ministry of Agriculture and Environment, 2 Dang Thuy Tram, Hanoi 10000, Vietnam
[3] Ireland's Centre for AI, School of Computer Science, University of Dublin, Belfield Office Park, Clonskeagh, Dublin 4, D04 V2N9 Dublin, Ireland
[4] Hydraulic Construction Institute, Viet Nam Academy for Water Resources, No. 3, Alley 95, Chua Boc Street, Dong Da District, Hanoi 10000, Vietnam
[5] Department of Land Management, Dong Thap University, Cao Lanh 87000, Vietnam; nguyenho@dthu.edu.vn
[6] Institute of Landscape Ecology, University of Münster, 48149 Münster, Germany
[7] Institute of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8572, Ibaraki, Japan
* Correspondence: nasahara.kenlo.gw@u.tsukuba.ac.jp

Highlights

What are the main findings?

- The first national-scale agricultural land use maps of Vietnam for 2020 and 2024 were produced, including 15 land-cover categories, of which eight represent agricultural types, with particular emphasis on high-value plantations such as coffee and rubber.
- The proposed framework achieved overall accuracies of 83.01% ± 1.37% for 2020 and 80.09% ± 0.76% for 2024 by introducing an adaptive weighted combined loss function to mitigate class imbalance in the training dataset.

What are the implications of the main findings?

- An end-to-end and transferable deep learning framework is presented for large-scale agricultural mapping over heterogeneous landscapes, effectively addressing class imbalance.
- The resulting agricultural maps provide reliable spatial information on agricultural land, supporting data-driven decision-making and policy formulation for sustainable agricultural development in Vietnam.

Abstract

Accurate and timely agricultural mapping is essential for supporting sustainable agricultural development, resource management, and food security. Despite its importance, Vietnam lacks detailed and consistent large-scale agricultural maps. In this study, we produced the first national-scale agricultural map of Vietnam for 2024 using a UNet++ deep learning architecture that integrates multi-temporal Sentinel-1 and Sentinel-2 imagery with Global-30 DEM data. The resulting product includes 15 land-cover categories, eight of which represent the most popular agricultural types in Vietnam. We further evaluate the model's transferability by applying the 2024 trained model to generate a corresponding map for 2020. The approach achieves overall classification accuracies of 83.01 ± 1.37% (2020) and 80.09 ± 0.76% (2024). To address class imbalance within the training dataset, we introduced an adaptive weight combined loss function that automatically adjusts the

Check for updates

Academic Editor: Christoph Jörges

Received: 25 December 2025
Revised: 22 January 2026
Accepted: 24 January 2026
Published: 29 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

weight of dice loss and cross-entropy loss within a combined loss function during the model training process.

Keywords: agriculture; land use/land cover; remote sensing; semantic segmentation; UNet++; adaptive weighted combined loss function

1. Introduction

Agriculture plays a fundamental role in global development by driving economic growth, supporting poverty reduction, ensuring food security, and contributing to environmental sustainability [1,2]. In addition to food provision, it supplies essential raw materials to diverse industrial sectors [3]. Agriculture's socio-economic importance is further underscored by the fact that it supported approximately 892 million people in 2023, accounting for about 26% of the global workforce [4].

In Vietnam, agriculture remains a cornerstone of the national economy. In 2023, the sector contributed 8.84% of the country's gross domestic product (GDP) [5] and positioned Vietnam among the world's leading exporters of key agricultural commodities, including rice, coffee, tea, pepper, cashew, cassava, rubber, and aquatic products [6]. This strong agricultural performance plays a critical role in rural livelihoods and national food security.

Accurate agricultural land mapping is therefore essential for sustainable agricultural development [7], crop yield estimation [8], evidence-based policy formulation [9], and long-term food security. Despite its importance, national-scale agricultural mapping in Vietnam remains limited. Existing products predominantly emphasize broad land use/land cover (LULC) classes (e.g., urban, forest, water), while diverse orchard and plantation systems such as coffee and tea are frequently aggregated into a single cropland category [10,11]. Moreover, most studies focus on regional scales [12,13] or individual crops, such as coffee [14] or rice [15], resulting in the absence of a unified, multi-crop national assessment.

Satellite remote sensing provides an effective data source for large-scale agricultural mapping by enabling timely, consistent, and spatially explicit observations over extensive areas [16,17]. Among currently available satellite missions, Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 optical imagery, operated by the European Space Agency (ESA), are particularly suitable for national-scale agricultural applications. Both sensors offer a spatial resolution of 10 m and frequent revisit intervals of approximately five days, supporting detailed and timely monitoring of agricultural landscapes. Sentinel-2 supplies multispectral information through 13 spectral bands for characterizing vegetation conditions, while Sentinel-1 provides C-band SAR observations that are insensitive to cloud cover, enabling continuous monitoring under all weather conditions [18]. Together, these complementary characteristics make Sentinel-1 and Sentinel-2 well-suited for national-scale agricultural mapping that requires adequate spatial resolution, high temporal frequency, wide geographic coverage, and cost-effective data availability [16].

Despite the advantages of Sentinel-1 and Sentinel-2 for large-scale agricultural monitoring, reliance on a single type of satellite imagery remains insufficient to capture the full complexity of agricultural landscapes at national scales. Many existing studies have focused on agricultural land mapping using single-date optical imagery, typically selecting cloud-free scenes to minimize cloud contamination, and have therefore been applied primarily to relatively small study areas [19–23]. However, a single optical acquisition cannot adequately represent crop phenological dynamics, which limits the separability of different crop types and constrains mapping accuracy at broader spatial scales.

Integrating Sentinel-1 SAR and Sentinel-2 optical data provides an effective solution to these limitations by exploiting the complementary strengths of radar and optical observations. Specifically, Sentinel-1 SAR offers cloud-independent structural information on crops, while Sentinel-2 optical imagery captures spectral characteristics of vegetation. Numerous studies have demonstrated that this synergistic integration substantially improves classification accuracy compared with using either data source alone [24–26]. Nevertheless, most existing implementations remain confined to limited spatial extents or single-crop systems, highlighting the need to extend this integrated approach to national-scale, multi-crop agricultural mapping.

A wide range of methods has been developed for agricultural mapping using remote sensing imagery. Traditional pixel-based classifiers, such as Random Forest, Support Vector Machines, and Multilayer Perceptrons, are widely applied but may face challenges in heterogeneous agricultural landscapes due to high intra-class variability [27]. Object-based image analysis (OBIA) can partially mitigate this issue by incorporating spatial context; however, its performance is sensitive to segmentation parameters and may lack robustness across regions.

In contrast, deep learning approaches, particularly convolutional neural networks (CNNs), have demonstrated superior performance in agricultural mapping. Encoder–decoder architectures such as UNet++ effectively capture both fine-scale spatial details and high-level contextual information through multi-scale feature learning. The nested dense skip connections in UNet++ reduce the semantic gap between encoder and decoder representations, enabling more accurate boundary delineation and improved discrimination of complex agricultural patterns. Consequently, UNet++ has shown strong and consistent performance in remote sensing semantic segmentation, even under limited training data or highly heterogeneous landscape conditions [20,28,29].

Attention mechanisms have proven effective in enhancing the accuracy of remote sensing image classification by enabling deep learning models to emphasize informative features while suppressing irrelevant or noisy signals selectively. In complex landscapes, where spectrally similar land-cover types frequently coexist, treating all spatial locations and feature channels equally can limit model discrimination. Attention modules address this limitation by adaptively assigning higher weights to important features while down-weighting less informative ones, thereby enabling the model to focus more effectively on key spatial and spectral cues [30]. Recent studies have demonstrated that incorporating attention mechanisms into CNN-based architectures significantly enhances classification and segmentation performance in remote sensing applications, including LULC mapping and crop classification [9,31]. Consequently, integrating an attention mechanism represents an effective strategy for improving the accuracy and reliability of large-scale agricultural mapping products.

Within deep learning frameworks, loss function design plays a critical role in segmentation performance, particularly when training data exhibit strong class imbalance [32]. This issue is especially pronounced in agricultural mapping, where dominant land-cover classes can overshadow minority but agronomically important categories [33]. Among loss functions, the cross-entropy loss function is widely used due to its stable optimization behavior, but it tends to underrepresent minority classes. In contrast, Dice loss function emphasizes spatial overlap and is more effective at addressing class imbalance, although it may introduce training instability [21,34]. To leverage the complementary strengths of these loss functions, combined cross-entropy and Dice losses with fixed weighting coefficients have been widely adopted in remote sensing segmentation studies, demonstrating improved performance across imbalanced land-cover classification tasks [33,35–37].

Although combined cross-entropy and Dice loss functions have demonstrated strong performance in handling class imbalance, their weighting coefficients are typically predefined and remain fixed throughout training [34,38–40]. The extent to which dynamically adjusting these weights in response to model performance can further improve training stability and enhance segmentation accuracy, particularly for minority classes has not yet been systematically investigated in large-scale agricultural mapping. This observation motivates the present study to explore adaptive loss weighting strategies and evaluate their potential benefits.

Building on these considerations, this study proposes an end-to-end framework for national-scale agricultural mapping in Vietnam using multi-temporal Sentinel-1 and Sentinel-2 data and a deep learning segmentation model trained with an adaptive weighted combination of cross-entropy and Dice losses. The main contributions of this work are:

- We produced the first agricultural map that covers the entire mainland of Vietnam, with a specific focus on the topmost valuable agricultural plantation of the country (i.e., coffee, rubber tree) in 2020 and 2024. To the best of our knowledge, our product is the first and most updated agricultural map in Vietnam.
- We developed an end-to-end framework integrating data processing and a transferable deep learning model to generate agricultural maps of Vietnam.
- We designed a novel loss function that automatically adjusts the relative contributions of the cross-entropy and Dice loss functions. The corresponding weights, denoted as α and β, are dynamically updated during training based on the rate of change of the validation mean Intersection over Union (IoU), to improve both the validation mean IoU and the overall validation loss.

2. Materials and Methods

2.1. Study Area

The study area encompasses the entire mainland of Vietnam (Figure 1), a country characterized by pronounced geographical, climatic, and socio-economic diversity. Vietnam is commonly divided into three major regions: the North, Central, and South. The northern region comprises extensive mountainous areas in the north and northwest, alongside large alluvial deltas in the eastern lowlands. The Central region is defined by high mountain ranges in the west and a narrow coastal plain to the east, while the southern region is dominated by the Mekong Delta, which serves as the country's primary rice-producing area and plays a critical role in national food security.

Climatically, Vietnam can be broadly divided into two sub-regions. The North experiences four distinct seasons, whereas the South is characterized by alternating rainy and dry seasons. Combined with complex topography, this strong environmental heterogeneity has resulted in highly diverse agricultural systems across the country. While such diversity poses substantial challenges for agricultural land-cover classification, it also makes Vietnam an ideal test site for evaluating the robustness, spatial generalization, and temporal transferability of national-scale agricultural mapping approaches.

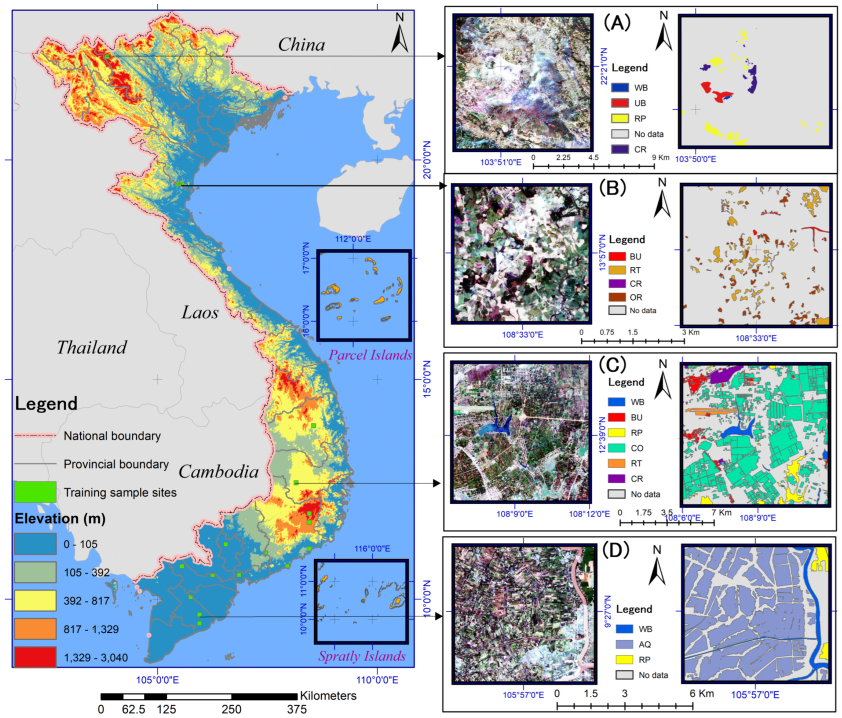


Figure 1. The study area encompasses 17 training sample sites, indicated by green boxes. Panels (**A**–**D**) present the true color Sentinel-2 image composite and training data spatial distribution in 4 different sites ((**A**) the Northern Mountainous region; (**B**) the Central Coastal region; (**C**) the Central Highlands, and (**D**) the Mekong Delta). Elevation information was derived from the Copernicus Global-30 DEM.

2.2. Classification Scheme

A classification scheme is chosen to highlight the top-most valuable agricultural products of Vietnam, including coffee, rubber trees, rice, aquaculture, etc. [5]. The name and description of categories are shown in Table 1.

Table 1. Classification scheme and definition.

No	Code	Name of Category	Definition
1	RP	Rice paddy field	Agricultural land intentionally flooded or irrigated to cultivate rice.
2	RT	Rubber tree	Land dominated by plantation perennial rubber tree *(Hevea brasiliensis)*
3	CO	Coffee	Land dominated by plantation perennial coffee tree *(Coffea)*
4	OR	Orchard	Land dominated by plantation perennial fruit trees and not removed after each harvest, excluding coffee, rubber tree, cashew and coconut
5	CR	Crop	Cultivated agricultural land used for growing seasonal plants, which are harvested within a single growing cycle
6	CC	Coconut	Land dominated by plantation perennial coconut palm trees *(Cocos nucifera)*
7	CA	Cashew	Land dominated by plantation perennial tropical evergreen tree cashew *(Anacardium occidentale)*

Table 1. *Cont.*

No	Code	Name of Category	Definition
8	AQ	Aquaculture	Artificial surface water for cultivating aquatic organisms such as fish, shrimp, shellfish, and aquatic plants
9	GR	Grassland	Land dominated by natural or managed grasses and herbaceous vegetation
10	ME	Melaleuca	Perennial woody tree species occurring primarily in freshwater wetland environments
11	MA	Mangrove	Perennial woody tree species occurring primarily in saline wetland environments
12	EB	Evergreen broad-leaved tree	Land dominated by natural or plantation broad-leaved trees that retain their foliage throughout the year.
13	WB	Water body	Natural surface water features such as rivers, streams, and natural ponds that retain water seasonally or permanently, excluding artificial ponds used for aquaculture.
14	BU	Built-up area	Regions characterized by human settlement and infrastructure, including buildings, roads, and other constructed features.
15	BR	Barren	Abandoned land, including lands that are temporarily abandoned

2.3. Satellite Acquisition and Processing

The region of interest was divided into 171 (0.5° × 0.5°) tiles to facilitate the downloading and processing of satellite imagery. Satellite images were obtained using Geemap (version 0.35.3) which is a Python (version 3.12) package for interacting with Google Earth Engine [41]. After downloading, satellite images were processed by Python packages such as numpy, geopandas, and Geospatial Data Abstraction Library (GDAL).

2.3.1. Sentinel-1 Data

Sentinel-1 is a C-band Synthetic Aperture Radar (SAR) mission that operates independently of atmospheric conditions, making it particularly suitable for monitoring phenological dynamics in cloud-prone regions. In this study, Ground Range Detected (GRD) Sentinel-1 data with dual polarizations (VV and VH) were used. These products are provided with standard preprocessing, including radiometric calibration to backscatter coefficients and ortho-rectification, and therefore require no additional preprocessing. Data were acquired at two-month intervals from January to December 2024, resulting in six temporal composites. For each interval, the mean of all available scenes was calculated, producing one representative single-band image per polarization and a total of 12 single-band images per tile over the study period.

2.3.2. Sentinel-2 Data

Harmonized Sentinel-2 MultiSpectral Instrument (MSI) surface reflectance data were acquired for this study. Five spectral bands, including Blue (B2), Green (B3), Red (B4), Near-Infrared (B8), and Shortwave Infrared 1 (B11) were selected from images acquired during the dry season (January to March) with cloud cover below 20%. All bands were processed at a final spatial resolution of 10 m; therefore, the native 20 m B11 band was automatically resampled to the target resolution within the Google Earth Engine environment using the default nearest-neighbor method to ensure spatial alignment across all bands. Cloud masking was applied using the cloud probability layer provided with the product. Pixels

were removed by cloud masking were gap-filled using observations from the same spatial location and seasonal period in alternate years. Finally, a median composite was generated for each band over the selected time period to further reduce residual cloud and shadow effects. Details of the Sentinel-2 data used in this study are summarized in Table 2.

Table 2. Sentinel-2 data used in this study.

Channel	Name of Bands	Central Wavelength (nm)	Spatial Resolution (m)	Times of Acquisition
B2	Blue	490	10	January–March 2024
B3	Green	560	10	
B4	Red	665	10	
B8	Near Infrared	842	10	
B11	Shortwave Infrared	1610	20	

Based on the selected Sentinel-2 bands, four spectral indices were derived to enhance the discrimination of spectrally similar land-cover classes. The Normalized Difference Vegetation Index (NDVI) was applied to enhance the classification of vegetation. The calculation of NDVI is shown in Equation (1). In addition, the Normalized Difference Water Index (NDWI) [42], and Normalized Difference Pond Index (NDPI) [43] were used to distinguish water bodies and aquaculture farms. NDWI and NDPI were calculated by Equations (2) and (3), respectively. Finally, the Normalized Difference Built-up Index (NDBI) [44] was utilized to improve the detection of built-up areas (Equation (4)).

$$NDVI = \frac{B8 - B4}{B8 + B4} \tag{1}$$

$$NDWI = \frac{B3 - B8}{B3 + B8} \tag{2}$$

$$NDPI = \frac{B11 - B3}{B11 + B3} \tag{3}$$

$$NDBI = \frac{B11 - B8}{B11 + B8} \tag{4}$$

In addition to spectral bands and indices, texture features were incorporated to further improve the discrimination of land-cover types with similar spectral signatures but distinct spatial patterns [45–47]. In this study, entropy and correlation were calculated from the gray-level co-occurrence matrix of Band 8, providing additional structural information to support the classification model. Correlation measures the linear dependency of gray levels in the image, while entropy provides information about the average level of uncertainty of pixels [48]. In this study, a 3×3 pixel sliding window was formed to calculate correlation and entropy based on Equations (5) and (10), respectively.

$$\text{Correlation} = \frac{\sum_i \sum_j (a_i - \mu_r)(a_j - \mu_c) P(i,j)}{\sigma_r \sigma_c} \tag{5}$$

$$\mu_r = \sum_i a_i \left(\sum_j P(i,j) \right) \tag{6}$$

$$\mu_c = \sum_j a_j \left(\sum_i P(i,j) \right) \tag{7}$$

$$\sigma_r{}^2 = \sum_i \left((a_i - \mu_r)^2 \sum_j P(i,j) \right) \tag{8}$$

$$\sigma_c{}^2 = \sum_j ((a_j - \mu_c)^2 \sum_i P(i,j)) \tag{9}$$

$$\text{Entropy} = -\sum_i \sum_j P(i,j) \log_2 P(i,j) \tag{10}$$

where:
a_i, a_j grey level in a window;
$P(i,j)$ co-occurrence probability of combination a_i and a_j;
μ_r, μ_c means of row and column grey levels in a window;
σ_r, σ_c standard deviations of row and column grey levels.

2.3.3. DEM Data

To enhance vegetation type differentiation, the Copernicus DEM GLO-30 digital surface model was integrated into the model training process. In addition to elevation, slope was extracted using the terrain function in Google Earth Engine to provide topographic information to the model. Elevation and slope were then resampled to a 10 m spatial resolution within Google Earth Engine using the default nearest-neighbor method to ensure spatial alignment with the other input bands. This topographic context helps distinguish vegetation types with similar spectral signatures but different terrain preferences, such as rubber plantations and plantation forests, where rubber is typically established in flatter areas, while plantation forests are more commonly distributed on steeper slopes.

After downloading, all images are normalized into $[0, 1]$. Sentinel-2 spectral bands were normalized using Equation (12). The original Sentinel-2 reflectance values are provided as scaled integers ranging from 0 to 10,000, and linear normalization was applied to map these values into a common range suitable for model input. Spectral indices derived from Sentinel-2 bands were used in their original form without additional normalization.

Sentinel-1 HH and HV bands were linearly normalized based on an empirical inspection of sampled backscatter distributions across representative regions and seasons in Vietnam. Most valid Sentinel-1 backscatter values were observed to fall approximately within the range of -60 dB to $+10$ dB, thus Sentinel-1 bands were normalized by Equation (11).

Texture information correlation normalization was calculated by Equation (13), whereas the entropy value remains un-normalized. Elevation and slope were normalized by Equations (14) and (15), respectively.

$$\text{Sentinel-1}_{\text{norm}} = \frac{DN + 60}{70} \tag{11}$$

$$\text{Sentinel-2}_{\text{norm}} = \frac{DN}{10,000} \tag{12}$$

$$\text{Correlation}_{\text{norm}} = \frac{Correlation + 1}{2} \tag{13}$$

$$\text{DEM}_{\text{norm}} = \frac{elevation}{3143\text{m}} \tag{14}$$

$$\text{Slope}_{\text{norm}} = \frac{img}{90^\circ} \tag{15}$$

Finally, all normalized images were stacked into a 25-band image composite. The feature space is shown in Table 3.

Table 3. Feature space before normalization.

Sensor/Index	Name of Band	Total Bands in Feature Space
Sentinel-1	VH (Jan, Mar, May, Jul, Sep, Nov) VV (Jan, Mar, May, Jul, Sep, Nov)	6 6
Sentinel-2	B2, B3, B4, B8, B11	5
Spectral index	NDVI, NDWI, NDPI, NDBI	4
Texture	Correlation, Entropy	2
Copernicus Global DEM-30	Elevation, Slope	2

2.4. Training Data

We selected 17 representative and feature-rich sites distributed across the major geographic regions of Vietnam (North, Central, and South) to construct the training dataset. Each site covers an area of $0.1^\circ \times 0.1^\circ$ (approximately 121 km^2) and was deliberately chosen to capture dominant agricultural systems while maximizing land cover diversity within each tile. This selection strategy emphasizes spatial continuity in the training data by preserving both inter-class boundaries between different land-cover types and intra-class parcel boundaries (e.g., between adjacent rice paddy fields). Such spatial context is particularly important for training semantic segmentation models such as UNet++, which rely on learning object-level geometry and boundary delineation rather than area-proportional pixel frequencies. The spatial distribution of the selected training sites is shown in Figure 1. In total, the training sites cover 2057 km^2, corresponding to approximately 0.62% of the total study area.

The training dataset was generated through manual vectorization of Sentinel-2 imagery based on visual interpretation. When the visual interpretation of Sentinel-2 imagery was inconclusive, ambiguous regions were verified using the Planet global quarterly base map of the same year and field photographs. The final output consisted of a polygon vector file delineating distinct land-cover classes, with each class assigned a unique integer label. This vector dataset was subsequently converted to raster format (TIFF), with a no-data value of 255 assigned to pixels corresponding to areas excluded from the training process.

The training images were subsequently divided into patches of 128×128 pixels using a sliding window approach with a stride of 25 pixels. The 128×128 patch size was selected to provide sufficient spatial context for representing agricultural field patterns while maintaining manageable computational cost during training. Larger patch sizes increase memory usage and training time and, in the presence of no-data regions, lead to a higher proportion of discarded samples when enforcing data-quality thresholds. Conversely, smaller patches provide more limited contextual information. To reduce noise and avoid training on non-informative samples, patches containing fewer than 10% valid pixels were excluded from the dataset. From the remaining patches, 90% were used for model training and 10% were randomly selected for validation. In total, 25,963 image patches were used for training and 3814 patches were reserved for validation.

To increase the robustness and generalization capability of the model, data augmentation was applied to the training images. This process generates a more diverse set of synthetic samples while preserving the key characteristics of the original data [49]. Specifically, 50% of the image patches were randomly flipped in either the horizontal or vertical direction, and 50% were randomly scaled within a range of $[0.9, 1.1]$. All augmentation operations were implemented using the Albumentations Python library, which is widely used for image augmentation in deep learning applications.

The test dataset was constructed independently from the training and validation datasets to ensure an unbiased performance assessment (Figure 2). Test points were randomly generated across the study area while explicitly excluding all locations used to construct the training dataset (the predefined $0.1^\circ \times 0.1^\circ$ sample plots), thereby avoiding spatial dependence between training and testing data. We used as many suitable geotagged field photographs as possible to construct the test dataset. For locations where field photographs were unavailable, test points were labeled through expert visual interpretation using high-resolution Planet global quarterly basemap imagery, supported by local knowledge. Due to limitations in available ground-truth photographs for earlier years, the 2020 evaluation relied on 2838 test points, whereas a larger and more comprehensive reference dataset comprising 10,388 test points was available for 2024.

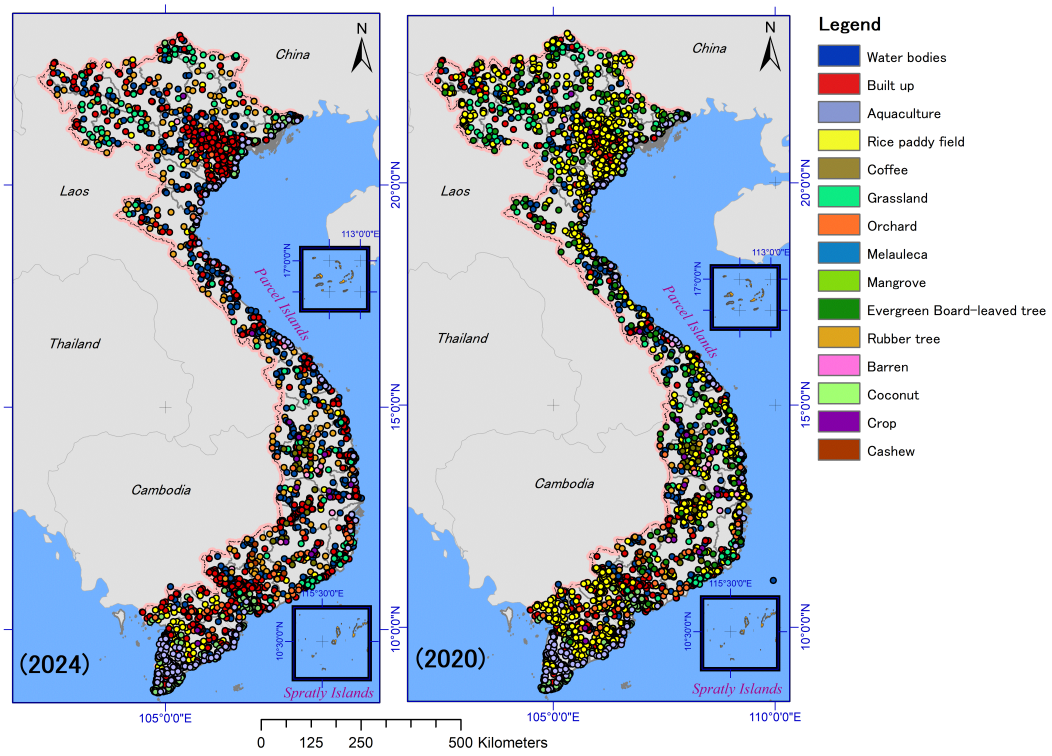


Figure 2. Spatial distribution of the test dataset for years 2024 and 2020 across mainland Vietnam.

2.5. Field Survey Data

Field surveys were conducted in February 2020 and February 2025 across the Mekong Delta and the Central Highlands of Vietnam to collect ground-truth data for model validation. The surveys focused on acquiring reference information for the target land-cover and crop categories defined in the classification scheme. Ground truth photographs were collected along predefined survey routes using multiple platforms, including a Casio GPS-mounted camera (Casio Computer Co., Ltd., Tokyo, Japan), GoPro Max (GoPro, Inc., San Mateo, CA, USA), Theta X camera (RICOH Co., Ltd., Tokyo, Japan), and a drone DJI Air 3S (SZ DJI Technology Co., Ltd., Shenzhen, China). All images were geotagged with geographic coordinates (latitude and longitude) to ensure accurate spatial referencing.

In 2020, approximately 1200 geotagged photographs were collected using a GPS-mounted camera. In 2025, a substantially larger dataset of 12,931 geotagged photographs was acquired, consisting of 10,526 images from the GoPro Max, 1444 from the Theta camera, and 961 from drone surveys. These ground-truth photographs were used as reference data for constructing both the training and independent test datasets. The survey routes and representative GPS-tagged photographs are shown in Figure 3.

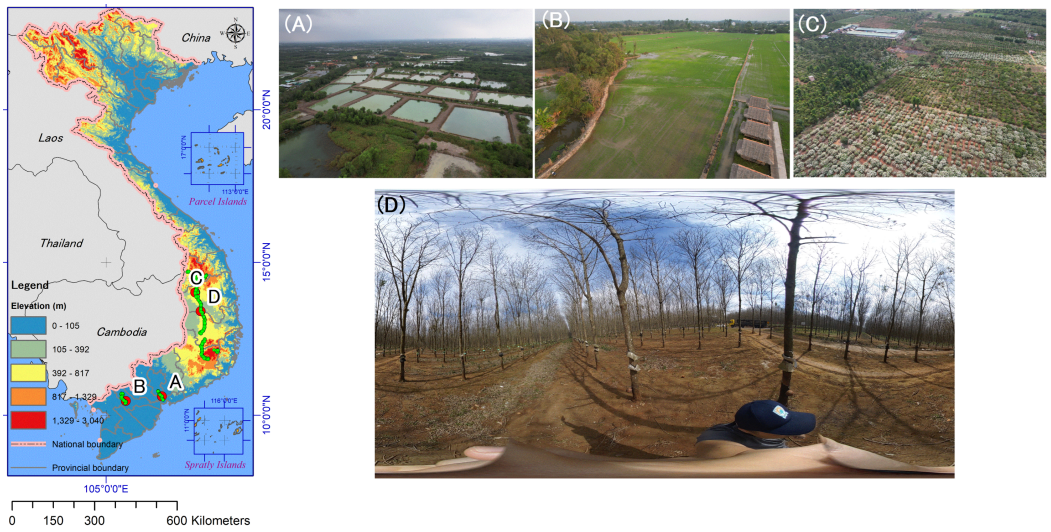


Figure 3. Field survey route in 2025. The dots (green and red) indicate the location of GPS photos were taken during the field trip in the Mekong Delta and Central Highland in 2025. Totally, 12,931 geotagged photos had been captured. The red dots denote the location of pictures of four representative plantations. (**A**) (106°49′38″, 10°37′12″)—Aquaculture; (**B**) (105°38′59″, 10°28′17″)—Rice paddy field; (**C**) (107°55′08″, 14°01′01″)—Coffee plantation; (**D**) (108°09′08″, 11°54′35″)—Rubber tree.

2.6. Training Model

Our overall methodology consists of three main stages: data download and processing, label data, and Training model. An overview of the proposed workflow is shown in Figure 4.

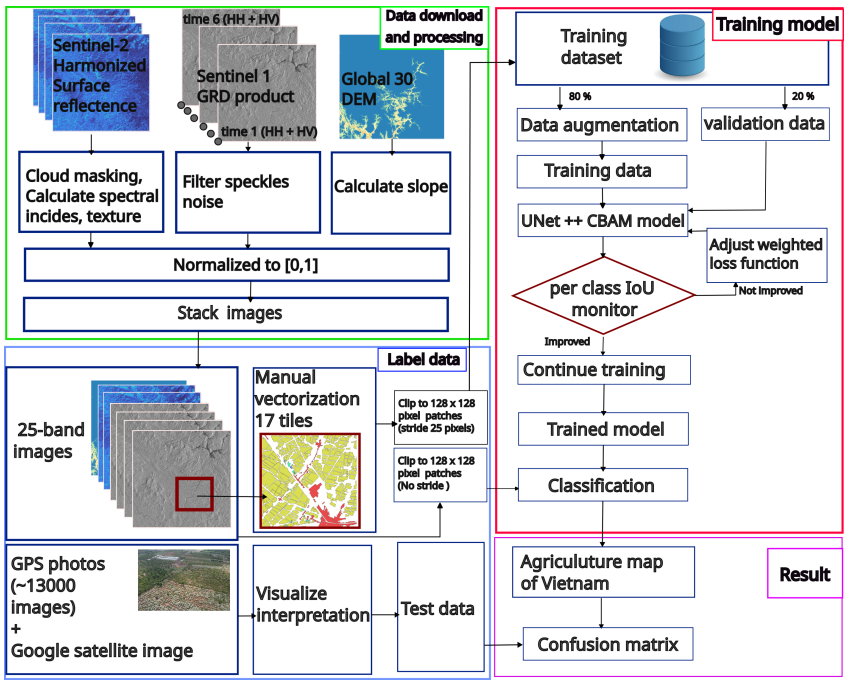


Figure 4. Workflow for generating the agricultural land-use map in this study.

To perform semantic segmentation for agricultural land classification, the UNet++ architecture was employed to classify multi-source remote sensing data. UNet++ is an advanced extension of the original UNet architecture and has been widely adopted for semantic segmentation tasks. It follows an encoder-decoder structure designed to capture

both fine-scale spatial details and high-level semantic information. In the encoder, successive 3 × 3 convolutional layers with a stride of 1 are applied to extract hierarchical feature representations, followed by max-pooling operations that progressively reduce spatial resolution while increasing the number of feature channels. At the center of the network, a bottleneck layer captures high-level abstract features with an expanded receptive field. The decoder then restores spatial resolution through upsampling operations, enabling pixel-level classification.

Compared with the original UNet, UNet++ introduces nested dense skip connections that link encoder and decoder feature maps at multiple semantic levels. These connections reduce the semantic gap between low-level spatial features and high-level contextual representations, enhancing multi-scale feature fusion and improving boundary delineation. This design is particularly effective for preserving object shapes and detecting irregular agricultural field boundaries. In this study, the encoder channel configuration was set to 16, 32, 64, 128, 256, 512, 1024, and 2048. To further enhance feature representation, a Convolutional Block Attention Module (CBAM) was embedded within the U-Net++ architecture. Specifically, CBAM was integrated at the end of each convolutional block in both the encoder and decoder pathways. After the standard sequence of convolution, batch normalization, activation, and dropout operations, the feature maps were sequentially refined using channel and spatial attention mechanisms. The overall network architecture is illustrated in Figure 5.

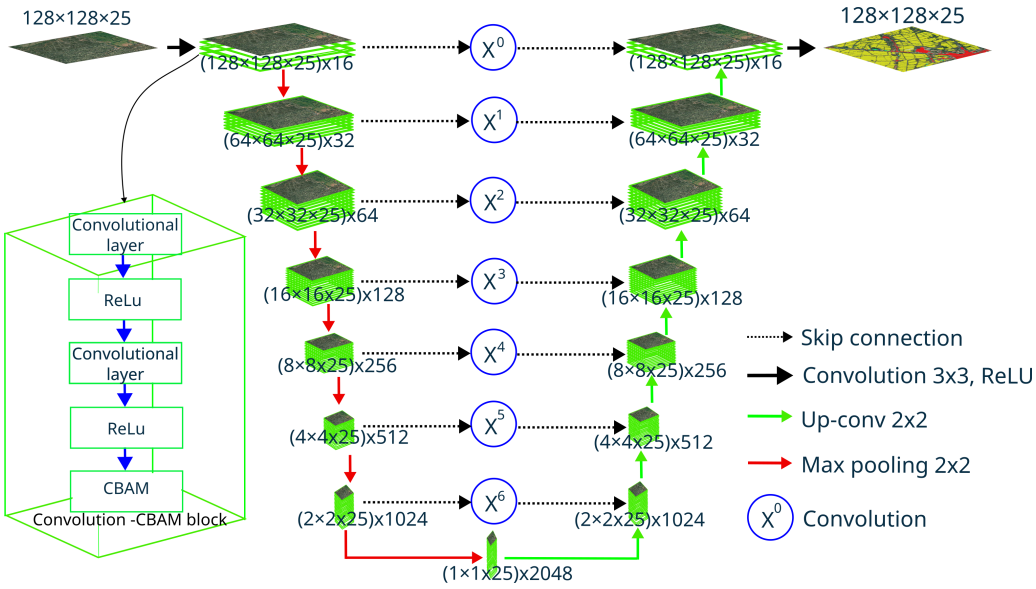


Figure 5. UNET ++ CBAM model used in this study. The color-coded arrows denote the main operations: black arrows represent 3 × 3 convolution followed by ReLU activation, red arrows represent 2 × 2 max pooling, green arrows represent 2 × 2 up-convolution for upsampling, and dashed black lines represent skip connections. The curved callout indicates a zoomed-in view of the convolution–CBAM block used in this study.

Model training was performed for a maximum of 150 epochs. Early stopping with a patience of 10 epochs was applied to prevent overfitting. The Adam optimizer was used, with an adaptive learning rate adjusted based on the training loss at each epoch. Model training was performed using standard Python deep learning utilities provided by the TensorFlow/Keras framework. A maximum of 150 training epochs was specified; however, the effective number of epochs was determined automatically through an early-stopping mechanism based on validation loss, with a patience of 10 epochs. In addition, a built-in learning-rate scheduling callback was used to reduce the learning rate when validation performance plateaued, and the best-performing model was automatically

saved using a validation–loss–based checkpoint strategy. While a comprehensive ablation study or exhaustive hyperparameter optimization was not performed, key training parameters were adjusted during model development based on observed trends in training and validation loss, validation mIoU, and convergence stability to obtain a stable and well-performing configuration.

To improve training stability, we used an adaptive weighted combined loss function (AWCLF) that integrates cross-entropy and Dice loss. The relative weights of the two loss terms are adjusted during training based on validation performance, rather than being fixed in advance. Together with early stopping and learning-rate scheduling, AWCLF helps ensure a stable and reproducible training process. The detailed formulation and implementation of AWCLF are presented in the following section.

All experiments were conducted on an Ubuntu Linux 22.04.5 LTS operating system. Model training and inference were implemented in Python 3.12 using the TensorFlow 2.19 framework. To accelerate deep learning training and large-scale image classification, an NVIDIA RTX A6000 GPU with 49 GB of memory was utilized, together with CUDA version 12.2.

2.7. Adaptive Weighted Combined Loss Function

In this study, we propose AWCLF that integrates cross-entropy loss and Dice loss through dynamically adjusted weighting coefficients. This strategy aims to improve classification stability and robustness to class imbalance training dataset by tuning the contribution of each loss function during training. The AWCLF is defined as follows:

$$\mathcal{L} = \alpha_{\text{adaptive}}\,\mathcal{L}_{\text{CE}} + \beta_{\text{adaptive}}\,\mathcal{L}_{\text{Dice}}, \tag{16}$$

where $\mathcal{L}$ denotes the total loss value at each training epoch, $\mathcal{L}_{\text{CE}}$ represents the cross-entropy loss, and $\mathcal{L}_{\text{Dice}}$ denotes the Dice loss.

In Equation (16), the weighting coefficients α_{adaptive} and β_{adaptive} are dynamically updated during training based on the changes of the validation mean Intersection over Union (mIoU). The mIoU metric, which quantifies the spatial overlap between predicted and validated segmentation masks, is first computed for each class and then averaged across all classes. When the validation mIoU decreases for four consecutive epochs, indicating a potential degradation in segmentation performance, the rate of change of mIoU is estimated by fitting a first-order linear regression to the most recent four mIoU values. The resulting slope, denoted as δ, is used to characterize the performance trend and is computed using Equations (17)–(19) with the NumPy regression utilities [50].

$$\delta = \frac{\sum_{i=1}^{4}(i-\bar{i})(mIoU_i - \overline{\text{mIoU}})}{\sum_{i=1}^{4}(i-\bar{i})^2} \tag{17}$$

$$\bar{i} = \frac{1}{4}\sum_{i=1}^{4} i = 2.5 \tag{18}$$

$$\overline{\text{mIoU}} = \frac{1}{4}\sum_{i=1}^{4} mIoU_i \tag{19}$$

where

$i \in \{1,2,3,4\}$ denotes the epoch index, corresponding to the first, second, third, and fourth epochs, respectively.

$mIoU_i$ is the mean IoU of ith epoch.

After calculating the rate of change, it is applied to adjust α and β in 2 directions, increasing and decreasing, by Equations (20) and (21), respectively.

$$case1 = \begin{cases} \alpha_{new} = \alpha_{\text{adaptive}} + \delta \\ \beta_{new} = \beta_{\text{adaptive}} - \delta \end{cases} \tag{20}$$

$$case2 = \begin{cases} \alpha_{new} = \alpha_{\text{adaptive}} - \delta \\ \beta_{new} = \beta_{\text{adaptive}} + \delta \end{cases} \tag{21}$$

Each case is applied temporarily, and the corresponding validation mIoU is recalculated to evaluate the impact of changing to the mIoU. The adjustment yielding the higher validation performance is retained for subsequent training epochs. Finally, the new α and β were updated to the loss function by Equation (22). This bidirectional update strategy ensures flexibility in the optimization process, allowing the model to adaptively emphasize the more informative loss component based on validation feedback.

$$\mathcal{L} = \alpha_{\text{new}} \mathcal{L}_{\text{CE}} + \beta_{\text{new}} \mathcal{L}_{\text{Dice}} \tag{22}$$

The α_{new} and β_{new} are explicitly constrained to predefined bounds and clipped to the range [0.1, 0.9] to prevent extreme dominance of either loss component, ensuring stability during training.

The AWCLF is integrated into the standard training workflow and does not require additional backpropagation steps. Because the adaptive update involves only simple numerical operations on scalar validation metrics, the additional computational overhead is negligible. The adaptive mechanism is triggered only when a decline in validation mIoU is observed over four consecutive epochs. This window length was selected to balance sensitivity and stability. Shorter windows tend to be sensitive to short-term fluctuations in validation mIoU and may lead to frequent, unstable interventions, whereas longer windows delay corrective action and reduce the practical effectiveness of the adaptive strategy.

The adaptive mechanism is activated only when a decline in validation mIoU is observed over four consecutive epochs. In the initial stage of training in this study, the validation mIoU consistently increased as the model learned relevant feature representations, and the activation condition was therefore not satisfied. As a result, the AWCLF adjustment was not applied during the early training phase and was invoked only when a sustained degradation in validation performance was detected at later stages of training.

2.8. Accuracy Assessment

In this study, we assessed the accuracy during the training process using the rate of change of mIoU to adjust the weights of the cross-entropy and dice loss functions, thereby impacting the final combined loss function. IoU and mIOU are calculated based on Equations (23) and (24).

$$\text{IoU}_i = \frac{TP_i}{TP_i + FP_i + FN_i} \tag{23}$$

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \text{IoU}_i = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{24}$$

After finishing classification, we built the confusion matrix using test data, then measured the accuracy of the products based on the matrix. Overall accuracy (OA) was calculated by Equation (25). Uncertainty associated with the accuracy metrics was derived

directly from the confusion matrix and the size of the independent validation dataset. The standard error was computed by Equation (26):

$$OA = \frac{\sum_{i=1}^{K} n_{ii}}{N} \tag{25}$$

where:
K is the total number of class;
n_{ii} denotes the correct classification for class i;
N is the total samples in the test dataset.

$$SE = \sqrt{\frac{OA\,(1 - OA)}{N}} \tag{26}$$

and the reported uncertainty values (±) correspond to the 95% confidence interval, calculated as $\pm 1.96 \times SE$. This confidence interval reflects sampling uncertainty in the independent validation data.

In addition, the Kappa coefficient is employed to measure the agreement between testing and classification data. Kappa coefficient is calculated by Equations (27)–(29):

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{27}$$

$$p_0 = \frac{\sum_i nii}{N} \tag{28}$$

$$p_e = \frac{1}{N^2} \sum_i R_i \times C_i \tag{29}$$

where:
p_0 observed agreement;
p_e expected agreement by chance;
R_i total samples in row i (ground truth class i);
C_i total samples in column i (predicted class i);
n_{ii} element at row i and column i on the confusion matrix;
N total samples in the test dataset.

3. Results

3.1. Agricultural Land-Use Mapping Across Mainland Vietnam in 2020 and 2024

We produced a national-scale agricultural land-use map of mainland Vietnam by integrating multi-temporal, multi-source remote sensing data with a UNet++ semantic segmentation model. To further assess model robustness and temporal generalization, we evaluated model transferability by applying a model trained using 2024 data to classify remotely sensed imagery from 2020. The spatial distribution of agricultural land use for both years is shown in Figure 6. The resulting maps have a spatial resolution of 10 m, with overall accuracies of 83.01% ± 1.37% for 2020 and 80.09% ± 0.76% for 2024. The corresponding kappa coefficients are 0.77 and 0.80, respectively. These results demonstrate the capability of the proposed framework to produce consistent and reliable agricultural maps at the national scale, providing a valuable basis for large-scale agricultural monitoring and decision-making.

The generated maps clearly show the main agricultural regions of Vietnam, including the Red River Delta in the north, the narrow coastal plains along the eastern margin, the Central Highlands, and the Mekong Delta in the south. In the Red River Delta, rice cultivation dominates the landscape and is closely mixed with residential areas, orchards,

and aquaculture. In contrast, agricultural land use in the Mekong Delta shows clearer spatial separation. Rice paddies are mainly found in the upper delta, while orchards are concentrated along downstream river branches, where tropical fruits such as durian, dragon fruit, and pomelo are grown. Aquaculture areas are primarily located along southern coastal zones and river networks, reflecting intensive shrimp and fish farming activities.

Rubber plantations are predominantly located in transitional zones between the southeastern lowlands (e.g., Binh Phuoc, Tay Ninh, Dong Nai, and Binh Duong) and the Central Highlands (e.g., Gia Lai, Dak Lak, Dak Nong, and Kon Tum), where ecological conditions are particularly favorable. The Central Highlands and southeastern regions also represent Vietnam's primary coffee-growing areas, supported by fertile basalt soils, suitable elevations ranging from 500 to 1200 m, and a climate characterized by distinct wet and dry seasons. These spatial patterns are well captured by the proposed mapping approach and are consistent with known agro-ecological conditions across the country.

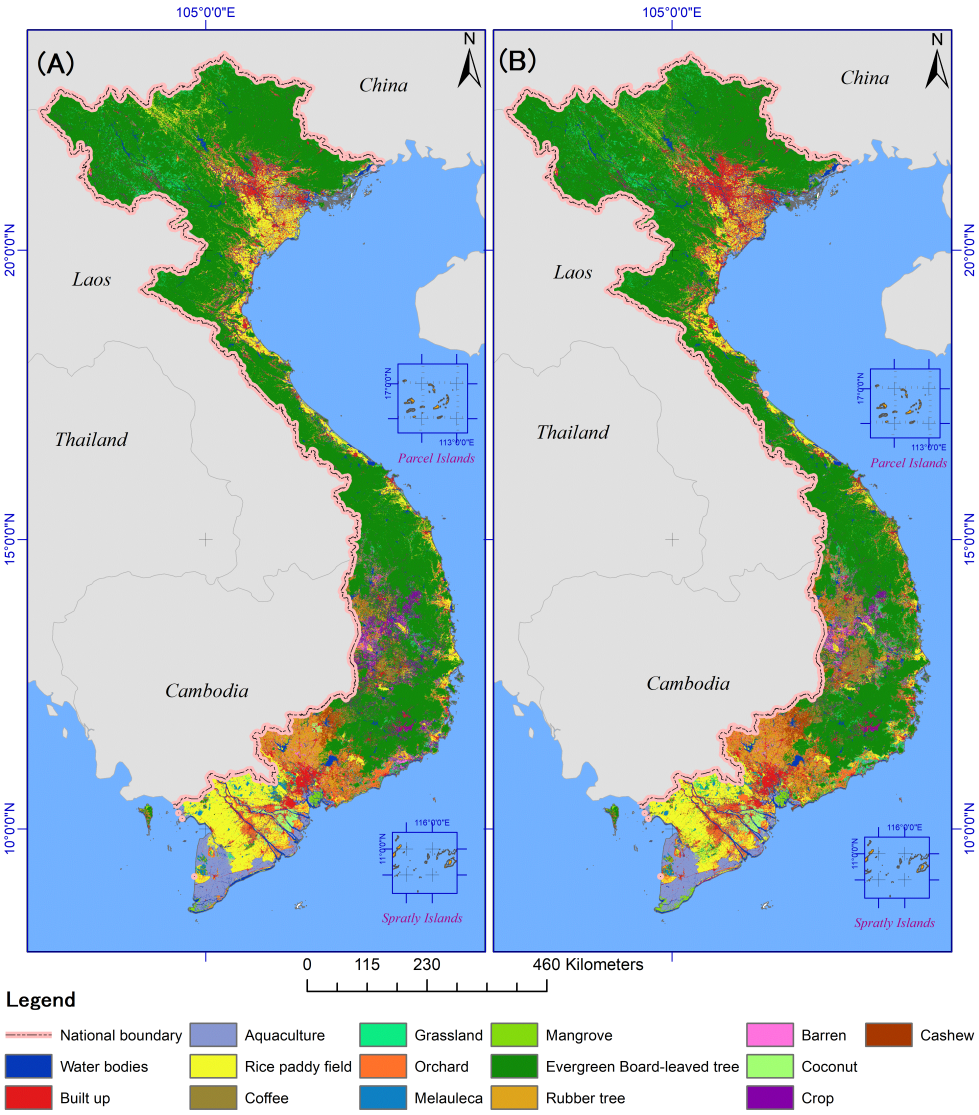


Figure 6. Agricultural land use map across the mainland Vietnam in 2020 (**A**) and 2024 (**B**).

To measure the accuracy of the result, we employed a confusion matrix and calculated producer accuracy and user accuracy of each class. In addition, the overall accuracy and the kappa coefficient were also calculated to assess the accuracy of the clas-

sification results. The confusion matrices for the 2020 and 2024 results are shown in Table 4 and Table 5, respectively.

Table 4. Confusion matrix of the agriculture map land use of Vietnam in 2020.

		Predict															
	Class	**WB**	**BU**	**AQ**	**RP**	**CO**	**GR**	**OR**	**ME**	**MA**	**EB**	**RT**	**BR**	**CC**	**CR**	**CA**	**PA (%)**
Actual	**WB**	288	9	7	18	0	1	0	0	0	2	1	3	1	2	0	86.75
	BU	8	586	9	6	0	0	3	0	0	7	2	8	0	3	1	92.58
	AQ	14	0	235	5	0	0	0	0	0	0	0	0	1	0	0	92.16
	RP	10	33	28	577	10	19	17	0	1	10	0	3	1	2	1	81.04
	CO	0	0	0	0	50	0	1	0	0	11	0	0	0	4	0	75.76
	GR	1	2	0	1	0	53	6	5	0	8	4	1	0	10	6	54.64
	OR	2	3	0	3	1	3	80	0	1	10	4	1	0	4	2	70.18
	ME	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0	100.00
	MA	23	3	15	29	0	0	14	1	72	3	0	0	8	0	0	42.86
	EB	1	0	0	5	1	5	2	0	0	137	2	0	0	0	1	88.96
	RT	0	0	0	0	0	0	4	0	0	1	40	0	0	2	0	85.11
	BR	0	1	0	2	0	2	0	0	0	3	0	37	0	1	0	80.43
	CC	0	0	0	0	0	0	0	0	0	0	0	0	40	0	0	100.00
	CR	0	1	0	3	0	1	1	0	0	8	0	0	0	99	0	87.61
	CA	0	0	0	0	0	0	0	0	0	0	0	0	1	0	30	96.77
	UA (%)	83.00	91.85	79.93	88.91	80.65	63.10	62.50	86.96	97.30	68.50	75.47	69.81	76.92	77.95	73.17	83.01

WB: Water body; BU: Built-up; AQ: Aquaculture; RP: Rice paddy; CO: Coffee; GR: Grassland; OR: Orchard; ME: Melaleuca; MA: Mangrove; EB: Evergreen broad-leaved; RT: Rubber tree; BR: Barren; CC: Coconut; CR: Cropland; CA: Cashew; PA: Producer accuracy; UA: User accuracy.

In 2020, the classification achieved an overall accuracy of 83.01%. Examination of the user's accuracy indicated substantial variation among land-cover classes. Mangrove attained the highest user's accuracy at approximately 97%, reflecting a low level of commission error. In contrast, grassland and orchard exhibited the lowest user's accuracies, at 63.10% and 62.50%, respectively, indicating notable over-classification. Orchard pixels were frequently confused with rice paddy and mangrove, whereas grassland was predominantly misclassified as rice paddy. Regarding the producer's accuracy, Melaleuca achieved the highest value at 100%, indicating that all Melaleuca pixels present in the testing data were correctly identified by the classifier. In contrast, mangrove recorded the lowest producer's accuracy, 42.86%, meaning that a substantial portion of mangrove pixels were omitted and incorrectly assigned to other categories. Interestingly, mangrove also reached the highest user's accuracies, suggesting that although the class was rarely over-classified, the model tended to underestimate the true extent of mangrove.

Table 5. Confusion matrix of the agriculture map land use of Vietnam in 2024.

		Predict															
	Class	**WB**	**BU**	**AQ**	**RP**	**CO**	**GR**	**OR**	**ME**	**MA**	**EB**	**RT**	**BR**	**CC**	**CR**	**CA**	**PA (%)**
Actual	**WB**	278	45	7	26	18	6	21	0	3	11	0	0	3	3	0	66.03
	BU	16	1324	1	46	39	5	37	2	1	4	3	7	5	17	4	87.62
	AQ	16	7	392	24	0	0	2	0	1	0	0	0	0	0	0	88.69
	RP	4	293	9	1501	10	14	29	1	0	5	4	3	4	36	0	78.46
	CO	0	8	0	0	832	1	0	0	0	20	0	1	0	23	1	93.91
	GR	2	8	1	20	1	154	32	7	2	49	2	1	0	2	4	54.04
	OR	6	52	5	63	8	3	592	13	9	61	30	3	29	6	7	66.74
	ME	1	0	2	2	0	0	3	41	3	4	0	0	1	0	1	70.69
	MA	6	5	23	12	0	0	67	0	1555	11	0	0	4	0	0	92.39
	EB	2	53	0	19	4	8	66	3	2	348	1	5	4	39	4	62.37
	RT	0	2	0	1	3	0	2	0	0	3	188	1	0	0	8	90.38
	BR	4	46	1	43	5	5	35	0	0	39	33	152	2	76	7	33.93
	CC	3	7	0	0	0	0	0	0	1	0	2	0	329	0	0	96.20
	CR	5	15	0	24	15	2	4	0	0	20	1	8	1	505	0	84.17
	CA	0	0	0	0	0	0	0	0	0	0	7	0	0	0	89	92.71
	UA (%)	81.05	70.99	88.89	84.28	88.98	77.78	66.52	61.19	98.60	60.52	69.37	83.98	86.13	71.43	71.20	80.09

WB: Water body; BU: Built-up; AQ: Aquaculture; RP: Rice paddy; CO: Coffee; GR: Grassland; OR: Orchard; ME: Melaleuca; MA: Mangrove; EB: Evergreen broad-leaved; RT: Rubber tree; BR: Barren; CC: Coconut; CR: Cropland; CA: Cashew; PA: Producer accuracy; UA: User accuracy.

In 2024, the highest classification performance was achieved for the mangrove forest class, with a user accuracy of 98.6% and a producer accuracy of 92.4%. Aquaculture and coffee followed, each attaining approximately 90% for both accuracy metrics. Rice paddy fields, barren land, and coconut plantations showed moderate performance, with accuracies around 80%. In contrast, evergreen broad-leaved forest and orchard classes exhibited the lowest overall accuracies among the evaluated land-use types.

Producer accuracy, which reflects omission errors, varied substantially across classes. The barren class showed the lowest producer accuracy at approximately 34%, indicating extensive omission errors. A considerable proportion of barren pixels were misclassified as built-up areas, forest, orchard, rubber plantations, and rice paddy fields. A similar pattern was observed for grassland, which achieved a producer accuracy of 54.04%, with many grassland pixels incorrectly classified as rice paddy fields.

User accuracy, which reflects commission errors, was lowest for orchard and evergreen broad-leaved forest classes, with values of approximately 66% and 60%, respectively. These commission errors are likely driven by the pronounced heterogeneity of orchard systems in Vietnam, where orchards vary widely in species composition, canopy structure, and management practices across regions, presenting challenges for accurate class separation by the segmentation model.

Notably, the overall accuracy of the 2020 classification is higher than that of the 2024 classification. This difference is associated with variations in the independent test datasets used for validation, which differ in size and composition due to practical constraints on ground-truth data availability. The 2020 validation relied on a smaller set of geo-referenced field photographs, whereas the 2024 assessment used a larger reference dataset. Although the same validation protocol was applied and both test sets were independent of the training data, overall accuracy is sensitive to sample size and class composition, as reflected by the wider confidence interval for 2020 ($\pm$1.37%) compared with 2024 ($\pm$0.76%). Therefore, the higher overall accuracy observed for 2020 should be interpreted in the context of validation dataset characteristics rather than as evidence of intrinsically superior model performance.

In addition to overall accuracy, the kappa coefficient (κ) is reported to account for agreement beyond chance. The higher κ value obtained for 2024 indicates a stronger chance-corrected agreement despite its slightly lower overall accuracy, providing a complementary perspective on classification performance across years when considered together with class-wise accuracy metrics.

3.2. Impact of the Adaptive Weighted Combined Loss Function on Minority Class Performance

A major difficulty in producing large-scale agricultural maps is the pronounced class imbalance in training data. In this study, the training dataset exhibits substantial imbalance among land-use classes, as summarized in Table 6. Coconut and Melaleuca are the least represented classes, accounting for only 0.51% and 0.56% of the total training pixels, respectively. In contrast, evergreen broad-leaved forest constitutes nearly one-fifth of the training dataset, making it the dominant class, followed by coffee and mangrove, each contributing approximately 12%. All remaining classes individually account for less than 5% of the total training data.

In this study, we integrated an Adaptive Weighted Combined Loss Function (AWCLF) that combines Dice loss and cross-entropy loss to improve semantic segmentation performance, particularly for minority classes. Model performance was evaluated using multiple loss functions to identify the most effective approach for handling the strongly imbalanced training dataset. Because all loss functions were evaluated using the same overlapping patch extraction and validation strategy, comparisons of overall accuracy (OA) and mean

Intersection over Union (mIoU) are interpreted in a relative sense to assess training stability and class separability, rather than as unbiased estimates of generalization accuracy.

Table 6. Pixel count and percentage distribution of land cover classes.

Category	Number of Pixels	Percentage (%)
Water bodies	14,644,362	4.77
Built-up	10,093,008	3.29
Aquaculture	12,791,499	4.17
Rice paddy	13,514,324	4.41
Coffee	36,953,476	12.05
Grassland	11,535,745	3.76
Orchard	16,587,473	7.29
Melaleuca	1,709,784	0.56
Mangrove	38,940,090	12.70
EBF	62,582,252	20.33
Rubber tree	53,082,929	17.37
Crop	17,875,902	3.30
Barren	5,143,562	1.68
Coconut	1,570,171	0.51
Cashew	11,686,829	3.81

The weighting coefficients α and β were initialized to 0.5 at the first training epoch. At epoch 64, which yielded the best validation performance, the coefficients were adaptively adjusted to $\alpha = 0.25$ and $\beta = 0.75$, resulting in a mean Intersection over Union (mIoU) of 0.7999 (Table 7), the highest value among all tested loss functions. In terms of overall accuracy, AWCLF ranked second, achieving an overall accuracy of 84.92%, which is comparable to the highest accuracy of 85.53% obtained using the Dice loss alone. These results indicate that AWCLF provides a favorable balance between overall accuracy and improved segmentation performance for minority classes.

Table 7. Comparison of loss functions using mean Intersection over Union (mIoU) and Overall Accuracy (OA).

Loss Function	mIoU	OA (%)
Cross-entropy	0.4862	71.72
Dice loss	0.7390	85.53
Cross-entropy and Dice loss	0.7758	84.01
AWCLF	0.7999	84.92

Figure 7 presents the class-wise model performance achieved using different loss functions. Among them, AWCLF yields the most pronounced performance improvements, producing clearer and more accurate predictions for eight land-cover categories while maintaining stable performance for the remaining classes. The largest gains are observed for the built-up, barren, coconut, and cashew categories, which represent only 3.29%, 1.68%, 0.51%, and 3.81% of the training dataset, respectively. These results demonstrate the effectiveness of dynamically weighted loss functions in improving segmentation performance for underrepresented classes in strongly imbalanced training datasets.

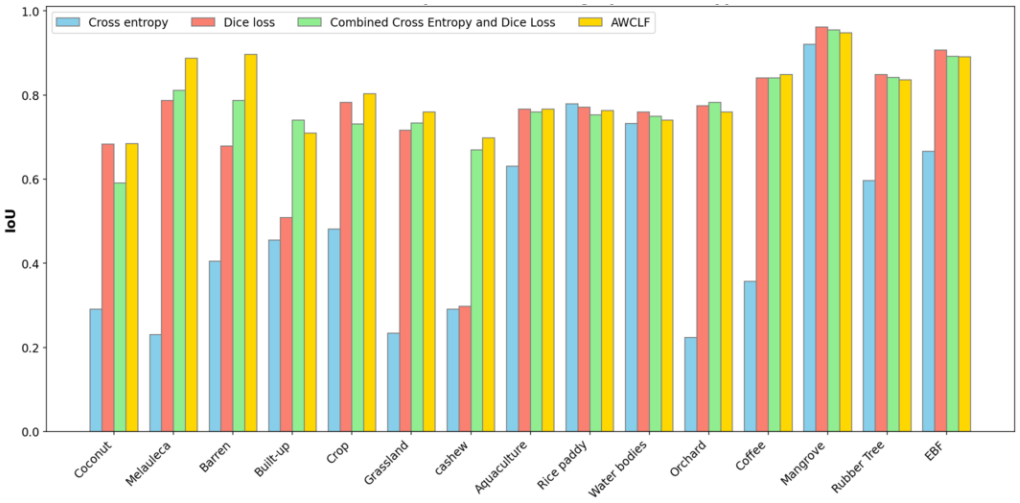


Figure 7. Class-wise Intersection over Union comparison across different loss functions, with land-cover classes ordered by ascending class frequency.

The effectiveness of AWCLF is most clearly demonstrated for the two most underrepresented classes in the dataset, namely Melaleuca and coconut, based on visual inspection of the classification maps. For Melaleuca, substantial improvements were observed in U Minh Ha National Park, a well-known Melaleuca-dominated landscape in the Mekong Delta of Vietnam. Prior to applying AWCLF, a large proportion of Melaleuca areas were incorrectly classified as rubber plantations (Figure 8(A1)). After incorporating AWCLF, this misclassification was substantially reduced, resulting in a more accurate spatial delineation of Melaleuca stands (Figure 8(A2)).

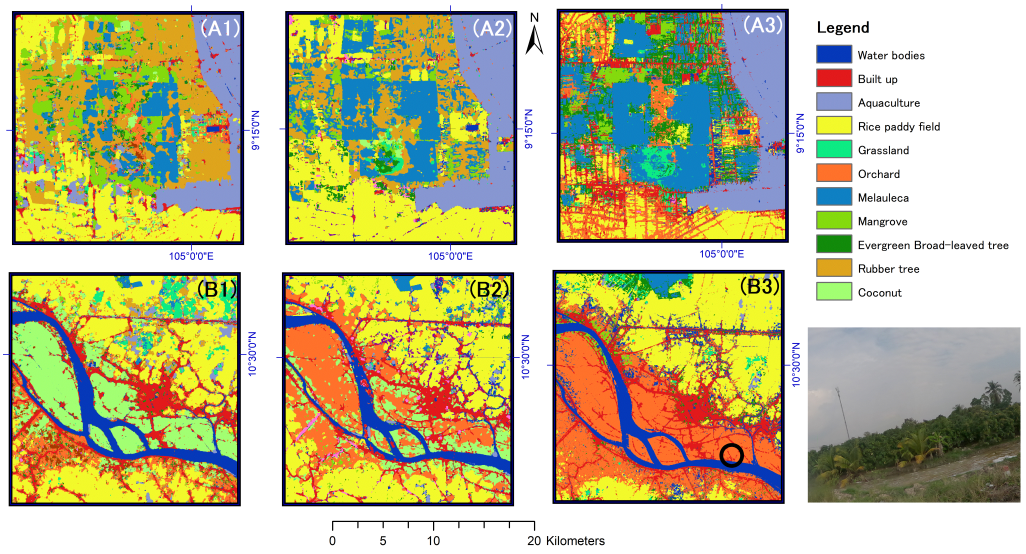


Figure 8. Improvements achieved by the Adaptive Weighted Combined Loss Function (AWCLF) for underrepresented categories. (**A1**,**B1**) show the baseline results without AWCLF; (**A2**,**B2**) present the results after applying AWCLF; and (**A3**,**B3**) illustrate the outcomes when both AWCLF and the attention module are incorporated. The geo-referenced field photograph collected during the 2025 field campaign (coordinates: 105°42′25″E, 10°24′35″E); indicated by the black circle.

A similar improvement was observed for the coconut class. In Dong Thap Province, located in the upper Mekong Delta, the baseline UNet++ model exhibited a tendency to misclassify coconut plantations as orchards (Figure 8(B1)). The application of AWCLF effectively mitigated this confusion, leading to more accurate coconut classification (Figure 8(B2)). These improvements were further supported by geo-tagged photographs

collected during field surveys (Figure 8), which confirmed that the canopy structures visible in the field images corresponded to orchard species rather than coconut, consistent with the corrected model predictions.

Overall, these qualitative results demonstrate that AWCLF substantially enhances classification performance for minority classes that are sparsely represented in the training dataset. The findings underscore the importance of adaptive, class-sensitive loss functions in improving the robustness and reliability of semantic segmentation for heterogeneous agricultural landscapes.

4. Discussion

4.1. Semantic Segmentation Application at the National Scale

Previous agricultural mapping efforts in Vietnam have largely focused on single crop types or orchard systems within relatively small and homogeneous regions, such as coffee plantations [51] or paddy rice fields [12]. In these localized settings, pixel-based classifiers often achieve high accuracy because spectral and structural variability is limited. However, when mapping is extended to broader and more heterogeneous regions, landscape complexity increases substantially. Many agricultural categories exhibit similar spectral reflectance and spatial patterns (e.g., coffee, dragon fruit, and tea [52]; rubber plantations and natural forests [53]), making reliable discrimination difficult for traditional pixel-based approaches. Moreover, pixel-based classification typically produces noisy and fragmented outputs because spatial context and object-level information are not explicitly incorporated during feature extraction [54]. As a result, reliance on spectral or textural features alone becomes insufficient for large-scale agricultural mapping.

In contrast, the proposed semantic segmentation framework demonstrates improved capability in separating spectrally and texturally similar land-cover types at the national scale. By integrating spatial context, object shape, and multi-temporal spectral information, the model reduces salt-and-pepper noise and produces more spatially coherent agricultural maps. These results highlight the advantages of semantic segmentation over traditional pixel-based classification for large-scale agricultural applications, particularly in heterogeneous landscapes where object structure and contextual relationships play a critical role.

Despite these improvements, the overall classification accuracy remains lower than that reported in some studies focusing on smaller areas or fewer crop types. Misclassifications are most pronounced among orchard, grassland, and coffee classes, which can be partly attributed to the medium spatial resolution of the input data. In northern Vietnam, for example, terraced rice fields are often narrower than the effective 10 m pixel size, resulting in frequent confusion with grassland. In addition, the max-pooling operations within the UNet++ architecture, while effective for capturing high-level features, inevitably reduce fine spatial detail, which can further affect the delineation of narrow or fragmented agricultural features.

Most previous studies applying UNet++ to land use and land cover mapping have relied on very high spatial resolution (VHR) imagery, such as 25 cm aerial photographs [55] or 1 m Earth-i imagery [56], due to their ability to clearly represent object boundaries and fine-scale structures [57]. Although VHR data enable precise boundary detection, their high acquisition costs, limited spatial coverage, and often restricted spectral information constrain their suitability for national-scale applications [56]. Furthermore, optical VHR imagery is highly susceptible to cloud contamination, making the acquisition of consistent multi-temporal datasets particularly challenging in tropical regions.

Our results indicate that integrating medium–spatial resolution Sentinel-1 and Sentinel-2 data offers a practical and scalable alternative for national-scale semantic segmentation. While these sensors provide coarser spatial detail than VHR imagery, their frequent revisit

intervals and complementary radar–optical information enable effective characterization of vegetation phenology and structural dynamics. In particular, multi-season Sentinel-1 time series capture temporal variations in vegetation structure that are often missed by single-date optical imagery. When combined with spectral, textural, and topographic features, this multi-source approach enhances class separability and mitigates the limitations associated with cloud cover and spectral confusion.

4.2. Comparison with Other Products

We compared the classification results with official land-use statistics for corresponding categories and found a generally consistent relationship between the two sources. According to the General Statistics Office of Vietnam, the national paddy field area in 2023 was approximately 3.93 million hectares [5], while our product estimated about 3.7 million hectares, corresponding to a deviation of 5.46%. A similar level of agreement was observed for rubber plantations, for which our estimate was 961,000 hectares compared to the official figure of 911,000 hectares [5], representing a difference of approximately 6%. The largest discrepancy was found for coffee plantations, where our product underestimated the area by about 144,000 hectares (approximately 20%). A comprehensive comparison between our estimates and official statistics is provided in Table 8.

The underestimation of coffee plantation is likely related to the widespread practice of intercropping coffee with large canopy, high-value shade trees, which has been actively promoted by the Vietnamese government to enhance climate resilience and sustainable production [58]. As these agroforestry systems mature, they often form dense canopy cover that increases spectral and structural similarity to evergreen broad-leaved forests in satellite observations [59], leading to potential misclassification and underestimation of coffee areas. In addition, the area estimates reported in this study are derived directly from the classified products and do not explicitly account for omission and commission errors, introducing further uncertainty in the comparison. It should also be noted that official agricultural statistics, which are commonly compiled from surveys and administrative reports, are themselves subject to reporting uncertainty. Consequently, the observed discrepancy between mapped coffee area and reported statistics reflects the combined effects of classification uncertainty, landscape complexity, and uncertainty in reference statistics, rather than a single source of model bias.

Table 8. Comparison between official statistical data and result of this study (Unit: 1000 ha).

Code	Category	Statistical Data	This Study Data	Difference	Ratio (%)
RP	Rice paddy field	3930.4	3715.6	214.8	5.46
CO	Coffee	718.6	574.4	144.2	20.0
RT	Rubber tree	911.2	961.34	−50.1	−6.0

We further compared our results with an existing national-scale LULC product for Vietnam and identified several areas where the proposed approach provides improved spatial representation. In the Mekong Delta, particularly near the mouth of the Hau River, large orchard areas were misclassified as mangrove forests in the JAXA LULC map [10]. While mangrove vegetation is indeed present in coastal zones near the river mouth, upstream areas are predominantly characterized by orchard systems rather than mangrove forests. A similar issue was observed at the mouth of the Tien River, where orchard areas were misclassified as rice paddy fields. These misclassification patterns are illustrated in Figure 9. Taken together, these comparisons suggest that the proposed

approach improves the discrimination of complex agricultural systems in coastal and riverine transition areas, while further refinement is required for specific crop types and mixed land-use systems.

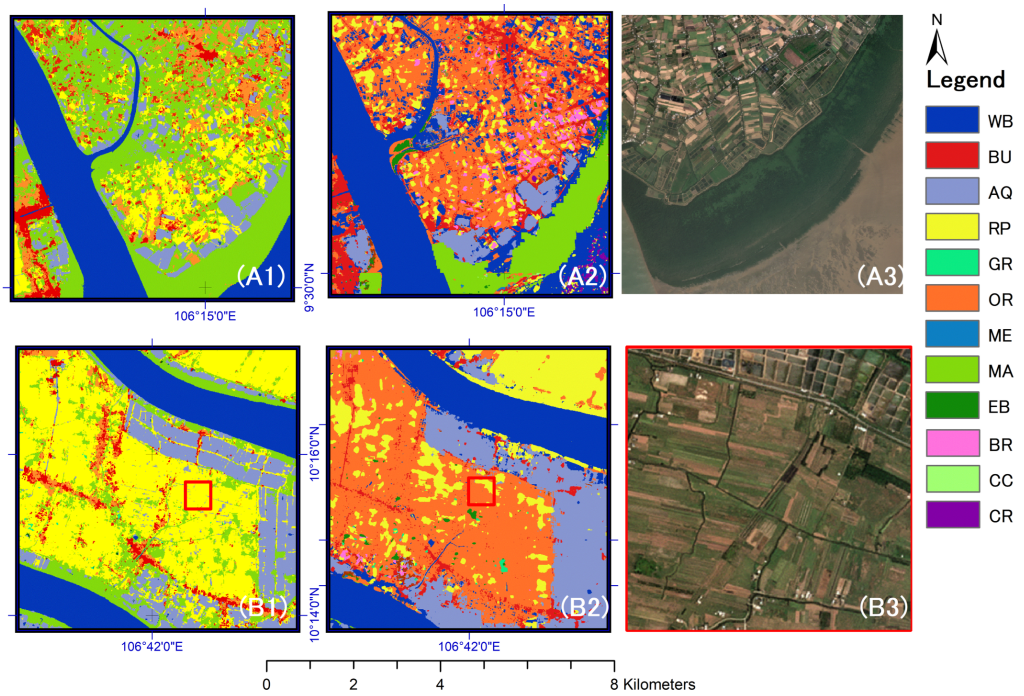


Figure 9. JAXA LULC map [10] in 2020 (**A1**,**B1**), our product in 2020 (**A2**,**B2**) and Planet global quarterly base map January to April 2020 of the same location (**A3**,**B3**).The red box in (**B1**,**B2**) delineates the area selected for detailed inspection; (**B3**) presents an enlarged view of this same area.

4.3. Effect of Attention Module on Classification Result

The integration of the Convolutional Block Attention Module (CBAM) effectively mitigated several classification challenges by enabling the network to emphasize informative features while suppressing irrelevant background signals [60]. In this study, the model incorporated a diverse set of input variables, including SAR imagery, multispectral data, and ancillary information such as spectral indices (NDVI, NDWI, NDPI, UI), texture features, and topographic attributes (elevation and slope). These variables provide important contextual cues; for example, rubber plantations are typically located in relatively flat terrain, making topographic features valuable predictors for their identification [61]. By integrating CBAM, the model was able to more effectively exploit these heterogeneous inputs, resulting in noticeable accuracy improvements for several classes, including water bodies, built-up areas, melaleuca, crops, coconut, and cashew.

The attention mechanism also enhanced the model's sensitivity to topographic and contextual information. Melaleuca plantations, which are primarily distributed in the flat lowlands of the Mekong Delta, were frequently confused with planted forests in mountainous areas or with rubber plantations due to similarities in spectral and textural characteristics. Although the inclusion of topographic variables was expected to reduce this confusion, the model without CBAM continued to misclassify a substantial proportion of Melaleuca pixels as rubber plantations, which are typically found at higher elevations (Figure 8(A2)). When CBAM was applied, these misclassifications were markedly reduced (Figure 8(A3)). This observation is consistent with previous findings showing that attention-based architectures can more effectively leverage topographic information to improve crop mapping performance [31]. Overall, these results highlight the importance of attention

mechanisms for integrating multi-source and contextual information in large-scale land-cover mapping.

4.4. Limitations

Semantic segmentation shows a stronger capability to distinguish spectrally similar land-cover classes than traditional pixel-based classification methods by incorporating spatial context and object-level information. However, the use of max-pooling operations in convolutional neural networks might lead to the loss of fine spatial details in the final predictions. Despite this limitation, the model performed notably well in classifying orchard areas, suggesting that segmentation benefits not only from pixel-level spectral information but also from spatial structure and shape patterns learned by the network.

Although the proposed model successfully differentiated several classes with similar spectral and textural characteristics, the classification accuracy for certain categories remained lower than that reported in some previous studies. This difference can be partly attributed to the moderate spatial resolution of the satellite data used in this study, as well as the high heterogeneity of the national-scale study area. In particular, orchard and cropland classes encompass diverse species compositions and management practices across regions, increasing intra-class variability and posing challenges for accurate classification.

In this study, a low producer's accuracy for the barren land class was observed in the 2024 classification (33.93%), indicating that the model tends to overestimate cropland extent, as a substantial proportion of barren reference samples were misclassified as cropland. This limitation is likely influenced by regional climatic and land-use conditions. In Vietnam's Central Highlands, the dry season lasts for approximately six months (from October to April), during which prolonged drought frequently leaves post-harvest fields temporarily exposed, particularly in cassava, sugarcane, and single-crop rice systems with limited irrigation. In addition, perennial crops such as coffee and rubber are commonly replanted after reaching the end of their economic life cycle, resulting in extensive areas of temporarily bare land following tree removal. Under these conditions, cropland surfaces often exhibit spectral characteristics similar to barren land in dry-season optical imagery, which increases confusion for the classification model.

Another contributing factor is the composition of the training dataset, in which the cropland class includes parcels that were already harvested and therefore exhibit reflectance characteristics similar to barren surfaces. More generally, this pattern suggests that the proposed approach has limited capability to explicitly capture short-term temporal dynamics associated with cropping cycles.

In this study, phenological information was partially incorporated through six bi-monthly Sentinel-1 SAR composites spanning January to December, which capture temporal variations related to canopy structure and surface moisture. The relatively high classification accuracy achieved for rubber plantations, a deciduous perennial crop with a pronounced seasonal cycle and strong agreement with statistical data indicates that Sentinel-1 data can effectively represent phenological changes for certain vegetation types. However, the lower performance observed for harvested cropland suggests that Sentinel-1 backscatter signals may be less informative for capturing short-term phenological transitions in annual crops under prolonged dry-season conditions, or that the current model configuration is not sufficiently effective at exploiting SAR-based temporal information for this purpose. This limitation highlights the need for improved temporal feature representations, more advanced time-series modeling approaches, or the integration of complementary multi-temporal optical data in future work to better characterize crop growth and harvest dynamics.

In addition, the reliance on moderate-resolution imagery inherently limits the detection of features smaller than 10 m. Consequently, narrow roads and small canals, especially those partially or fully covered by vegetation canopy, were frequently misclassified as vegetation rather than built-up areas. This issue highlights the trade-off between spatial coverage and spatial detail when using medium-resolution satellite data and underscores the challenges of detailed land-cover mapping in complex agricultural landscapes.

Similarly, certain agroforestry systems remain difficult to characterize accurately using medium-resolution imagery. In particular, shaded coffee plantations exhibited notable underestimation (approximately 20%), as coffee intercropped with large-canopy shade trees is spectrally and structurally similar to evergreen forest in both optical and SAR data. These results indicate that moderate-resolution multispectral and SAR observations may be insufficient to fully resolve complex vertical and horizontal vegetation structures. Overcoming this limitation will likely require complementary information, for example higher-resolution satellite imagery and/or explicit canopy-structure measurements (e.g., LiDAR). In practice, however, acquiring wall-to-wall LiDAR coverage over large areas is often difficult, and GEDI provides footprint-level samples rather than continuous observations. Therefore, future work may benefit from leveraging higher-resolution imagery where available and incorporating LiDAR/GEDI structural information when feasible (e.g., via upscaling with wall-to-wall predictors), together with more advanced representation-learning approaches such as attention-based networks or foundation models.

5. Conclusions

This study demonstrated an end-to-end national-scale agricultural mapping framework for mainland Vietnam and produced the first agricultural land-use map of the country, focusing on its most important agricultural systems. Using freely available multi-temporal Sentinel-1 and Sentinel-2 data together with Copernicus GLO-30 topographic variables, we implemented a UNet++ semantic segmentation model enhanced with CBAM and an adaptive weighted combined loss function (AWCLF) to address class imbalance and improve delineation of minority classes. The resulting 10 m products provide wall-to-wall agricultural land-use maps for 2024 and a temporally transferred map for 2020, enabling operational and repeatable national monitoring.

The proposed approach achieved overall accuracies of $80.09 \pm 0.76\%$ (2024) and $83.01 \pm 1.37\%$ (2020), with corresponding κ values of 0.80 and 0.77, respectively. Beyond overall accuracy, AWCLF increased the segmentation accuracy of minority classes, particularly coconut and Melaleuca, and reduced common misclassifications compared with the baseline loss setting. These results indicate that adaptive loss reweighting improves model robustness when mapping heterogeneous agricultural landscapes with strong class imbalance. The resulting maps reproduce major agricultural regions and spatial patterns, providing useful information for agricultural planning and policy support.

Despite these advantages, this study has several limitations that should be acknowledged. First, some classes remain difficult to separate, especially those with high internal variability and similar appearance (e.g., orchard and grassland). This problem is more pronounced in smallholder areas where fields are small, narrow, and fragmented, so 10 m imagery cannot always capture clear boundaries and mixed pixels occur frequently. Second, confusion between barren land and post-harvest cropland increases during prolonged dry-season periods because both surfaces can show similar reflectance and radar backscatter. Third, the coffee area was underestimated compared with official statistics, likely because shaded and intercropped coffee agroforestry systems are harder to distinguish from surrounding tree cover and other perennial crops at medium spatial resolution.

Future research should focus on the two directions. First, hybrid methods that combine pixel-based and semantic segmentation may improve the mapping of small and fragmented fields. Second, integrating geospatial foundation models and, where available, structural information (e.g., LiDAR) could better represent complex mixed cropping systems while reducing reliance on large labeled datasets.

Author Contributions: Conceptualization, T.H.T. and D.C.P.; methodology, T.H.T.; software, T.H.T. and D.B.M.; formal analysis, T.H.T.; investigation, T.H.T.; resources, T.H.T. and K.N.N.; data curation, T.H.T., D.B.M. and N.V.K.; writing—original draft preparation, T.H.T. and N.V.K.; writing—review and editing, T.H.T., D.C.P., N.V.K., D.B.M., H.N. and K.N.N.; visualization, T.H.T.; supervision, T.H.T. and K.N.N.; project administration, T.H.T.; funding acquisition, K.N.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was mainly funded by Japanese Grant Aid for Human Resource Development (JDS), number B0012024VNMD06 and Japan Aerospace Exploration Agency (JAXA) Commissioned Research "FY2024 Study on Advanced High-Resolution Land Use and Land Cover Classification for Consideration of Satellite Lidar Data Use" (JX-PSPC-573690; Principal Investigator: Kenlo Nasahara).

Data Availability Statement: Agricultural maps of Vietnam in 2020 and 2024, Trained model, source code used in this study is available at https://github.com/tatrung/AgrUNet accessed on 20 January 2026). Geotagged photos are available at https://www.mapillary.com/app/user/trung (accessed on 20 January 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Bank. Agriculture and Food. 2025. Available online: https://www.worldbank.org/en/topic/agriculture (accessed on 2 June 2025).
2. Byerlee, D.; de Janvry, A.; Sadoulet, E. Agriculture for Development: Toward a New Paradigm. *Annu. Rev. Resour. Econ.* **2009**, *1*, 15–31. [CrossRef]
3. Guillou, M.; Matheron, G. Producing Other Goods. In *The World's Challenge: Feeding 9 Billion People*; Guillou, M., Matheron, G., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 77–91. [CrossRef]
4. Food and Agriculture Organization of the United Nations. *World Food and Agriculture—Statistical Yearbook 2024*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2024. [CrossRef]
5. General Statistics Office of Viet Nam. *Statistical Yearbook of Viet Nam 2023*; General Statistics Office of Vietnam: Hanoi, Vietnam, 2023.
6. Ministry of Agriculture, Nature and Food Quality (Netherlands). Shaping Vietnam's Agricultural Future with Sustainable Growth and Organic Fertilizers. 2025. Available online: https://www.agroberichtenbuitenland.nl/actueel/nieuws/2025/04/24/as14-vietnam (accessed on 28 August 2025).
7. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
8. Reisi-Gahrouei, O.; Homayouni, S.; McNairn, H.; Hosseini, M.; Safari, A. Crop Biomass Estimation Using Multi Regression Analysis and Neural Networks from Multitemporal L-Band Polarimetric Synthetic Aperture Radar Data. *Int. J. Remote Sens.* **2019**, *40*, 6822–6840. [CrossRef]
9. Zheng, Y.; Dong, W.; Yang, Z.; Lu, Y.; Zhang, X.; Dong, Y.; Sun, F. A New Attention-Based Deep Metric Model for Crop Type Mapping in Complex Agricultural Landscapes Using Multisource Remote Sensing Data. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *134*, 104204. [CrossRef]
10. Truong, V.T.; Hirayama, S.; Phan, D.C.; Hoang, T.T.; Tadono, T.; Nasahara, K.N. JAXA's New High-Resolution Land Use Land Cover Map for Vietnam Using a Time-Feature Convolutional Neural Network. *Sci. Rep.* **2024**, *14*, 3926. [CrossRef]
11. Phan, D.C.; Trung, T.H.; Truong, V.T.; Sasagawa, T.; Vu, T.P.T.; Bui, D.T.; Hayashi, M.; Tadono, T.; Nasahara, K.N. First Comprehensive Quantification of Annual Land Use/Cover from 1990 to 2020 across Mainland Vietnam. *Sci. Rep.* **2021**, *11*, 9979. [CrossRef]
12. Ngo, T.X.; Bui, N.B.; Phan, H.D.T.; Ha, H.M.; Nguyen, T.T.N. Paddy Rice Mapping in the Red River Delta, Vietnam, Using Sentinel-1/2 Data and Machine Learning Algorithms. *J. Spat. Sci.* **2024**, *69*, 103–119. [CrossRef]

13. Karila, K.; Nevalainen, O.; Krooks, A.; Karjalainen, M.; Kaasalainen, S. Monitoring Changes in Rice Cultivated Area from SAR and Optical Satellite Images in Ben Tre and Tra Vinh Provinces in the Mekong Delta, Vietnam. *Remote Sens.* **2014**, *6*, 4090–4108. [CrossRef]
14. Son, N.T.; Chen, C.F.; Chen, C.R.; Cheng, Y.S.; Chen, S.H. Multidecadal Evaluation of Changes in Coffee-Growing Areas Using Landsat Data in the Central Highlands, Vietnam. *Geocarto Int.* **2023**, *38*, 2204099. [CrossRef]
15. Guan, X.; Huang, C.; Liu, G.; Meng, X.; Liu, Q. Mapping Rice Cropping Systems in Vietnam Using an NDVI-Based Time-Series Similarity Measurement Based on DTW Distance. *Remote Sens.* **2016**, *8*, 19. [CrossRef]
16. Wardlow, B.D.; Egbert, S.L. Large-Area Crop Mapping Using Time-Series MODIS 250 m NDVI Data: An Assessment for the U.S. Central Great Plains. *Remote Sens. Environ.* **2008**, *112*, 1096–1116. [CrossRef]
17. Liu, J.; Liu, M.; Tian, H.; Zhuang, D.; Zhang, Z.; Zhang, W.; Tang, X.; Deng, X. Spatial and Temporal Patterns of China's Cropland during 1990–2000: An Analysis Based on Landsat TM Data. *Remote Sens. Environ.* **2005**, *98*, 442–456. [CrossRef]
18. Segarra, J.; Buchaillot, M.L.; Araus, J.L.; Kefauver, S.C. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy* **2020**, *10*, 641. [CrossRef]
19. Blaschke, T. Object Based Image Analysis for Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
20. Clabaut, É.; Foucher, S.; Bouroubi, Y.; Germain, M. Synthetic Data for Sentinel-2 Semantic Segmentation. *Remote Sens.* **2024**, *16*, 818. [CrossRef]
21. Yao, J.; Jin, S. Multi-Category Segmentation of Sentinel-2 Images Based on the Swin UNet Method. *Remote Sens.* **2022**, *14*, 3382. [CrossRef]
22. Zhou, Y.; Yang, L.; Yuan, L.; Li, X.; Mao, Y.; Dong, J.; Lin, Z.; Zhou, X. High-Precision Tea Plantation Mapping with Multi-Source Remote Sensing and Deep Learning. *Agronomy* **2024**, *14*, 2986. [CrossRef]
23. Wang, M.; Wang, J.; Cui, Y.; Liu, J.; Chen, L. Agricultural Field Boundary Delineation with Satellite Image Segmentation for High-Resolution Crop Mapping: A Case Study of Rice Paddy. *Agronomy* **2022**, *12*, 2342. [CrossRef]
24. Zhang, H.; Peng, D.; Dou, C.; Lou, Z.; Zhang, X.; Yu, L.; Song, K.; Zhang, Y.; Hu, J.; Zheng, S.; et al. Enhancing Early-Season Soybean Identification through Optical and SAR Time-Series Integration. *Front. Plant Sci.* **2025**, *16*, 1656628. [CrossRef]
25. Valero, S.; Arnaud, L.; Planells, M.; Ceschia, E. Synergy of Sentinel-1 and Sentinel-2 Imagery for Early Seasonal Agricultural Crop Mapping. *Remote Sens.* **2021**, *13*, 4891. [CrossRef]
26. Blickensdörfer, L.; Schwieder, M.; Pflugmacher, D.; Nendel, C.; Erasmi, S.; Hostert, P. Mapping of Crop Types and Crop Sequences with Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data for Germany. *Remote Sens. Environ.* **2022**, *269*, 112831. [CrossRef]
27. Lu, D.; Weng, Q. A Survey of Image Classification Methods and Techniques for Improving Classification Performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]
28. Pan, Z.; Xu, J.; Guo, Y.; Hu, Y.; Wang, G. Deep Learning Segmentation and Classification for Urban Village Using a WorldView Satellite Image Based on U-Net. *Remote Sens.* **2020**, *12*, 1574. [CrossRef]
29. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
30. Ghaffarian, S.; Valente, J.; van der Voort, M.; Tekinerdogan, B. Effect of Attention Mechanism in Deep Learning-Based Remote Sensing Image Processing: A Systematic Literature Review. *Remote Sens.* **2021**, *13*, 2965. [CrossRef]
31. Wang, Y.; Zhang, Z.; Feng, L.; Ma, Y.; Du, Q. A New Attention-Based CNN Approach for Crop Mapping Using Time Series Sentinel-2 Images. *Comput. Electron. Agric.* **2021**, *184*, 106090. [CrossRef]
32. Xu, H.; He, H.; Zhang, Y.; Ma, L.; Li, J. A Comparative Study of Loss Functions for Road Segmentation in Remotely Sensed Road Datasets. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103159. [CrossRef]
33. Chen, P.; Liu, Y.; Ren, Y.; Zhang, B.; Zhao, Y. A Deep Learning-Based Solution to the Class Imbalance Problem in High-Resolution Land Cover Classification. *Remote Sens.* **2025**, *17*, 1845. [CrossRef]
34. Yang, Z.; Wu, Q.; Zhang, F.; Zhang, X.; Chen, X.; Gao, Y. A New Semantic Segmentation Method for Remote Sensing Images Integrating Coordinate Attention and SPD-Conv. *Symmetry* **2023**, *15*, 1037. [CrossRef]
35. Li, K.; Ji, H.; Li, Z.; Cui, Z.; Liu, C. AFNE-Net: Semantic Segmentation of Remote Sensing Images via Attention-Based Feature Fusion and Neighborhood Feature Enhancement. *Remote Sens.* **2025**, *17*, 2443. [CrossRef]
36. Li, J.; Zhang, H.; Chen, L.; He, B.; Chen, H. CSNet: A Remote Sensing Image Semantic Segmentation Network Based on Coordinate Attention and Skip Connections. *Remote Sens.* **2025**, *17*, 2048. [CrossRef]
37. Zhang, Z.; Liu, Q.; Liu, X.; Zhang, Y.; Du, Z.; Cao, X. PMNet: A Multi-Branch and Multi-Scale Semantic Segmentation Approach to Water Extraction from High-Resolution Remote Sensing Images with Edge-Cloud Computing. *J. Cloud Comput.* **2024**, *13*, 76. [CrossRef]
38. Yu, S.; Tao, C.; Zhang, G.; Xuan, Y.; Wang, X. Remote Sensing Image Change Detection Based on Deep Learning: Multi-Level Feature Cross-Fusion with 3D-Convolutional Neural Networks. *Appl. Sci.* **2024**, *14*, 6269. [CrossRef]
39. Wang, Y.; Pan, Y.; Lei, H.; Jin, D.; Chen, J. SPSIS: Single-Point Supervised Instance Segmentation for Remote Sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 26267–26282. [CrossRef]

40. Yao, H.; Li, Y.; Feng, W.; Zhu, J.; Yan, H.; Zhang, S.; Zhao, H. CAGM-Seg: A Symmetry-Driven Lightweight Model for Small Object Detection in Multi-Scenario Remote Sensing. *Symmetry* **2025**, *17*, 2137. [CrossRef]
41. Wu, Q. Geemap: A Python Package for Interactive Mapping with Google Earth Engine. *J. Open Source Softw.* **2020**, *5*, 2305. [CrossRef]
42. McFeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
43. Lacaux, J.P.; Tourre, Y.M.; Vignolles, C.; Ndione, J.A.; Lafaye, M. Classification of Ponds from High-Spatial Resolution Remote Sensing: Application to Rift Valley Fever Epidemics in Senegal. *Remote Sens. Environ.* **2007**, *106*, 66–74. [CrossRef]
44. Zha, Y.; Gao, J.; Ni, S. Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]
45. Gini, R.; Sona, G.; Ronchetti, G.; Passoni, D.; Pinto, L. Improving Tree Species Classification Using UAS Multispectral Images and Texture Measures. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 315. [CrossRef]
46. Mohammadpour, P.; Viegas, D.X.; Viegas, C. Vegetation Mapping with Random Forest Using Sentinel-2 and GLCM Texture Features: A Case Study of the Lousã Region, Portugal. *Remote Sens.* **2022**, *14*, 4585. [CrossRef]
47. Mishra, V.N.; Prasad, R.; Rai, P.K.; Vishwakarma, A.K.; Arora, A. Performance Evaluation of Textural Features in Improving Land Use/Land Cover Classification Accuracy of Heterogeneous Landscapes Using Multi-Sensor Remote Sensing Data. *Earth Sci. Inform.* **2019**, *12*, 71–86. [CrossRef]
48. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [CrossRef]
49. Sierra, S.; Ramo, R.; Padilla, M.; Cobo, A. Optimizing Deep Neural Networks for High-Resolution Land Cover Classification through Data Augmentation. *Environ. Monit. Assess.* **2025**, *197*, 423. [CrossRef]
50. NumPy Developers. Numpy.Polyval. 2025. Available online: https://numpy.org/doc/stable/reference/generated/numpy.polyval.html (accessed on 21 December 2025).
51. Maskell, G.; Chemura, A.; Nguyen, H.; Gornott, C.; Mondal, P. Integration of Sentinel Optical and Radar Data for Mapping Smallholder Coffee Production Systems in Vietnam. *Remote Sens. Environ.* **2021**, *266*, 112709. [CrossRef]
52. Nagori, R. Discrimination of Mango Orchards in Malihabad, India Using Textural Features. *Geocarto Int.* **2021**, *36*, 1060–1074. [CrossRef]
53. Dong, J.; Xiao, X.; Chen, B.; Torbick, N.; Jin, C.; Zhang, G.; Biradar, C. Mapping Deciduous Rubber Plantations through Integration of PALSAR and Multi-Temporal Landsat Imagery. *Remote Sens. Environ.* **2013**, *134*, 392–402. [CrossRef]
54. Cheng, X.; Lei, H. Semantic Segmentation of Remote Sensing Imagery Based on Multiscale Deformable CNN and DenseCRF. *Remote Sens.* **2023**, *15*, 1229. [CrossRef]
55. Clark, A.; McKechnie, J. Detecting Banana Plantations in the Wet Tropics, Australia, Using Aerial Photography and U-Net. *Appl. Sci.* **2020**, *10*, 2017. [CrossRef]
56. Flood, N.; Watson, F.; Collett, L. Using a U-Net Convolutional Neural Network to Map Woody Vegetation Extent from High Resolution Satellite Imagery across Queensland, Australia. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101897. [CrossRef]
57. Boonpook, W.; Tan, Y.; Nardkulpat, A.; Torsri, K.; Torteeka, P.; Kamsing, P.; Sawangwit, U.; Pena, J.; Jainaen, M. Deep Learning Semantic Segmentation for Land Use and Land Cover Types Using Landsat 8 Imagery. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 14. [CrossRef]
58. Rigal, C.; Tuan, D.; Cuong, V.; Le Van, B.; Trung, H.Q.; Long, C.T.M. Transitioning from Monoculture to Mixed Cropping Systems: The Case of Coffee, Pepper, and Fruit Trees in Vietnam. *Ecol. Econ.* **2023**, *214*, 107980. [CrossRef]
59. Hunt, D.; Tabor, K.; Hewson, J.; Wood, M.; Reymondin, L.; Koenig, K.; Schmitt-Harsh, M.; Follett, F. Review of Remote Sensing Methods to Map Coffee Production Systems. *Remote Sens.* **2020**, *12*, 2041. [CrossRef]
60. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *arXiv* **2018**. [CrossRef]
61. Tridawati, A.; Wikantika, K.; Susantoro, T.M.; Harto, A.B.; Darmawan, S.; Yayusman, L.F.; Ghazali, M.F. Mapping the Distribution of Coffee Plantations from Multi-Resolution, Multi-Temporal, and Multi-Sensor Data Using a Random Forest Algorithm. *Remote Sens.* **2020**, *12*, 3933. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.