

Multi-Sensor Deep Learning for Smallholder Paddy Rice Mapping with Sentinel-1 SAR and Sentinel-2 Optical Imagery

Alex Okiemute Onojeghuo

alex.onojeghuo@jolexyenviro.com

Jolexy Environmental Services Limited

Ajoke Ruth Onojeghuo

Jolexy Environmental Services Limited

Research Article

Keywords: deep learning, Sentine-1, Sentinel-2, U-Net, DeepLabV3, DeepLabV3+, PSPNet, Google Earth Engine, Ensemble

Posted Date: September 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10944003/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Multi-Sensor Deep Learning for Smallholder Paddy Rice Mapping with Sentinel-1 SAR and Sentinel-2 Optical Imagery

Alex Okiemute Onojeghuo^{1,*} and Ajoke Ruth Onojeghuo¹

¹ Research and Development Unit, Jolexy Environmental Services Limited, Edmonton, AB, Canada

*Corresponding author: alex.onojeghuo@jolexyenviro.com

Abstract

Smallholder paddy rice systems are central to food security across tropical and subtropical regions, yet their fragmented field patterns, diverse management practices, and rapid phenological transitions make accurate mapping challenging. This study evaluates deep learning approaches for delineating smallholder paddy rice fields using multi-temporal Sentinel-1 SAR and Sentinel-2 optical-derived phenological metrics spanning vegetative, reproductive, and ripening growth stages. Four semantic segmentation architectures—U-Net, DeepLabV3, DeepLabV3+, and PSPNet—were trained using fused SAR–optical inputs and validated with independent field data. Results indicated that the deep learning models integrating multi-scale contextual reasoning and strong spatial-detail recovery (namely, U-Net and PSPNet) outperformed the other evaluated architectures, achieving mean F1 scores of 0.80 and intersection-over-union (IoU) values of 0.87. To further improve the paddy rice map products, a majority-vote ensemble fusion was performed using U-Net and PSPNet predictions to enhance classification robustness, yielding the highest overall accuracy (0.94), rice F1-score (0.888), and Kappa coefficient (0.847). These findings demonstrate the value of adopting multi-sensor, multi-temporal phenological fusion and ensemble deep learning for operational smallholder paddy rice mapping and highlight pathways for potential upscaling across heterogeneous agroecological landscapes.

Keywords: deep learning, Sentine-1, Sentinel-2, U-Net, DeepLabV3, DeepLabV3+, PSPNet, Google Earth Engine, Ensemble

Key findings

- The combination of multi-temporal SAR and optical fusion, which captures stage-specific rice phenology, improves delineation of paddy rice fields.
- Combining complementary model architectures (U-Net and PSPNet) yields a more stable thematic product than relying on a single model.
- Multi-scale contextual information helps separate rice from spectrally or hydrologically similar land-cover classes.

1. Introduction

Smallholder paddy rice systems are crucial for food security and sustaining rural livelihoods across tropical and subtropical regions. Timely, accurate local crop yield estimates are crucial for programs that focus on allocating agricultural subsidies and implementing food security policies at localized scales [35]. The fragmented nature of field patterns, diverse management practices, and rapid changes in rice phenology present unique challenges to the timely and accurate delineation of paddy rice fields. Remote sensing offers a scalable and efficient means of monitoring the spatial extent and dynamics of rice fields. However, for smallholder-scale mapping, the challenges of frequent cloud cover, mixed pixels, and high degrees of spatial heterogeneity make adopting a multi-sensor and multi-date approach viable and sustainable [3].

The integration of Sentinel-1 (S1) synthetic aperture radar (SAR) and Sentinel-2 (S2) multispectral imagery has proven effective for rice crop mapping, as both sensors provide complementary information on crop growth dynamics. S1 SAR data offer all-weather, day-and-night observations and are sensitive to crop structure, surface roughness, flooding conditions, and canopy moisture during key stages of rice development, including transplanting and vegetative growth [31, 38]. In contrast, S2 multispectral imagery provides a rich repository of spectral information capable of capturing time-series optical measures of crop phenology, vegetation health, and seasonal growth patterns through vegetation and water-related indices [22, 38].

With the fusion of SAR and optical-derived series phenological measures, it's now possible to obtain an improved information source that reduces the degree of uncertainty associated with cloud contamination and consequently enhances classification accuracy compared with a single-sensor approach [31, 38]. The combination of a multi-sensor approach with reliable, field-based data enables a robust, scalable method for accurate paddy rice mapping across diverse agroecological settings, supporting smallholders' crop monitoring and sustainable agricultural management. The phenological pattern extracted from the S1-SAR backscatter coefficient values is an invaluable information source for understanding varying changes in paddy rice crop over its growing stages (particularly, vegetative, reproductive, and ripening), making it useful in paddy rice field delineation [16, 20]. Similarly, studies have demonstrated the value of combining optical sensor-derived indices (like Normalized Difference Vegetation Index (NDVI) [27], Land Surface Water Index (LSWI) [2], and Enhanced Vegetation Index (EVI) [12, 13]) as proxies for discriminating paddy rice over the different growth stages [1, 20, 21, 26, 32].

Previous studies have employed shallow learning (such as Random Forest and Support Vector Machine) and deep convolutional neural network (CNN) semantic image segmentation techniques for rice crop mapping. The literature cited in Table 1 demonstrates a clear progression from traditional shallow learning approaches (such as RF) towards more advanced deep learning semantic segmentation methods (such as U-Net, DeepLabV3, DeepLabV3+, and PSPNet) for rice crop mapping. While shallow learning classifiers remain effective and computationally efficient tools for large-scale rice mapping, recent advances in deep semantic segmentation networks (such as U-Net, Pyramid Scene Parsing Network (PSPNet), DeepLabV3, and DeepLabV3+) have demonstrated superior capabilities in extracting spatially coherent rice field boundaries across complex landscape patterns. This is further substantiated by the increasing availability of high-temporal and high-spatial open-access Sentinel imagery and commercial satellite imagery, which is vital in building an operational rice crop monitoring system.

The approach and results of this study are built on outcomes of AO Onojeghuo et al. [21], who demonstrated the utility of combining shallow and deep learning approaches with multi-temporal Sentinel-1 SAR and Sentinel-2 optical imagery for rice mapping; this study advances the methodology adopted by focusing exclusively on deep learning-based semantic segmentation.

Specifically, this study evaluates the capacity of four deep learning architectures —U-Net, DeepLabV3, DeepLabV3+, and PSPNet—for delineating smallholder paddy rice fields using multi-temporal S1 SAR and S2 optical-derived phenological metrics acquired across the vegetative, reproductive, and ripening stages of rice crops. The research objectives focused on the following components:

1. Evaluate the performance of four deep learning semantic segmentation models, U-Net, DeepLabV3, DeepLabV3+, and PSPNet, for paddy rice crop field delineation with multi-date and multi-source remotely sensed imagery.
2. Based on the outcomes of Objective 1, investigate the value of adopting an ensemble fusion of the best-performing deep learning architectures as part of the proposed workflow.
3. Discuss the implications of the results obtained for potential upscaling.

2. Materials and Methods

2.1. Study Area

The study area, Jiansangjiang, is situated in the Sanjiang Plain, Heilongjiang Province, Northeast China, and is an alluvial plain fed by three main rivers (namely, Heilong, Songhua, and Wusuli) and shares borders with Siberia in the north and east (Figure 1). The study area has a sub-humid continental monsoon climate with marked seasonality, a mean annual temperature of approximately 2.5 °C (ranging from −20 °C in January to 22 °C in July), and an annual precipitation averaging approximately 550 mm, falling predominantly during the summer monsoon period [20, 28]. AO Onojeghuo et al. [20] provides a comprehensive description of the study area's topography, land-use/land-cover characteristics, and paddy rice crop-growing pattern. The rice crop calendar was based on field investigations conducted across the Jiansangjiang Science and Technology Backyard (STB) site. Existing literature describes the rice growing season as three stages: the vegetative stage (approximately 65 days, consisting of transplantation, tillering and stem elongation phases), the reproductive stage (approximately 30-35 days, consisting of panicle initiation, booting/heading, and flowering phases), and ripening stage (approximately 30 days, consisting of milk, dough and mature phases), and the harvesting stage (an additional 18 days) [14, 30].

Table 1. Research on paddy rice mapping using shallow and deep learning with remote sensing data

Study	Data Source	Methodology	Project area	Key contribution
Rice Mapping in Training Sample Shortage Regions Using a Deep Semantic Segmentation Model Trained on Pseudo-Labels [29]	Sentinel-1 SAR	U-Net K-Means and random forest	China and USA	This study developed a workflow for extracting rice distribution information using a deep semantic segmentation U-Net model trained on pseudo-labels. The combination of K-Means and random forest was used to generate pseudo-labels.
Evaluation and Comparison of Semantic Segmentation Networks for Rice Identification Based on Sentinel-2 Imagery [33]	Sentinel-2	U-Net, DeepLabV3, Swin Transformer	China	Performed a systematic comparison of semantic segmentation architectures for rice identification and documented the advantages of deep learning over traditional classifiers.
Semantic Segmentation of Sentinel-2 Satellite Imagery for Rice Growth Phase Classification Using Deep Learning [19]	Sentinel-2	U-Net + CNN	Indonesia	For this study, the U-Net segmentation model was used for rice field segmentation while CNNs were used to classify rice growth phases from S2 multispectral imagery.
Semantic Segmentation of Rice Fields in Sub-Meter Satellite Imagery Using an HRNet-CA-Enhanced DeepLabV3+ Framework [25]	Sub-meter optical satellite imagery	DeepLabV3+	China	Improved rice field boundary delineation and achieved high segmentation accuracy through an enhanced DeepLabV3+ architecture.
Improved Phenology-Based Rice Mapping Algorithm by Integrating Optical and Radar Data [38]	Sentinel-1, Sentinel-2	Hybrid machine learning (ML) framework	Global-scale evaluation	Developed the rice mapping framework (named Rice-Sentinel) that integrates Sentinel-1 SAR and Sentinel-2 optical-derived phenology for robust rice mapping across diverse environments. The developed algorithm combines the strengths of both optical and SAR time series data and leverages biophysical principles to generate robust rice maps without relying on any prior ground truth samples.

Study	Data Source	Methodology	Project area	Key contribution
Deep ResU-Net Convolutional Neural Networks Segmentation for Smallholder Paddy Rice Mapping Using Sentinel 1 SAR and Sentinel 2 Optical Imagery [21]	Sentinel-1, Sentinel-2	ResU-Net CNN Random forest	China	Developed an operational workflow that employs similar phenological data to identify paddy rice distribution with deep learning.
Pyramid Scene Parsing Net Model for Automated Paddy Field Map using SPOT 6 Satellite Images [11]	SPOT 6	Pyramid Scene Parsing Net model	Indonesia	This paper presents a method for generating paddy field maps using a semantic image segmentation approach, in which the Pyramid Scene Parsing Net model is used for segmenting satellite imagery. The generated paddy map can be used as a basis for decision-making, especially in the agricultural sector

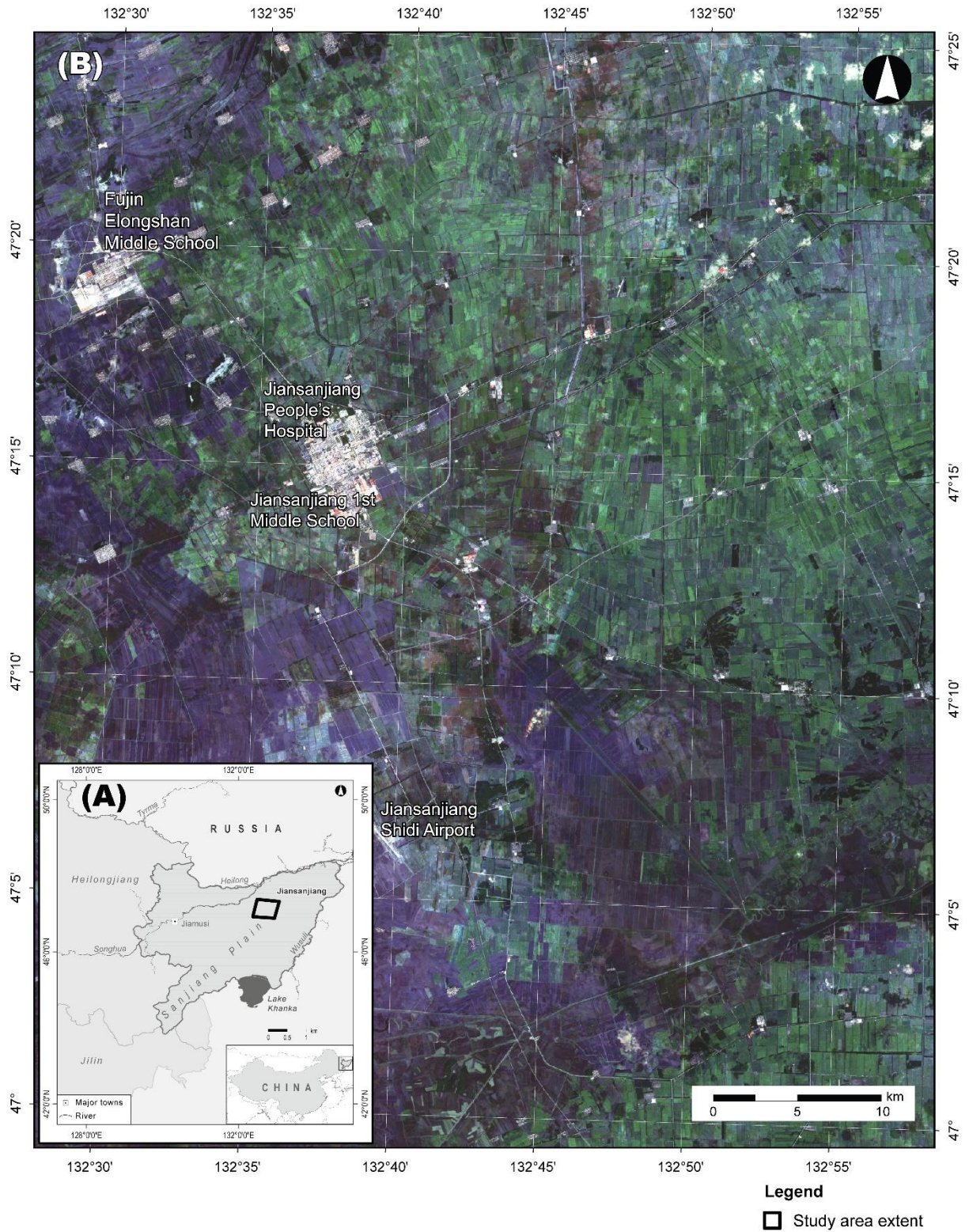


Figure 1. Study area map illustrating (A) the location of the Jiansanjiang region within Heilongjiang Province, China, and (B) a true-colour composite derived from multi-temporal Sentinel-2 optical imagery acquired in 2018.

2.2. Proposed Methodology

The proposed workflow for paddy rice mapping is presented in Figure 2 and comprises four major components: (1) preparation of remote sensing data, (2) model training, (3) model prediction and accuracy assessment, and (4) ensemble fusion using a majority voting approach. The data preparation stage provides a complete inventory of the satellite imagery and reference data used in the study, including all preprocessing steps required to generate analysis-ready inputs. All the satellite imagery used in the study was sourced from the GEE cloud computing platform. The model development stage involves training and evaluating four deep learning segmentation architectures—U-Net, PSPNet, DeepLabV3, and DeepLabV3+—to identify the most effective models for delineating paddy rice fields. Based on their comparative performance, the two best-performing architectures were selected for ensemble fusion. In the prediction and assessment stage, each paddy rice map product was evaluated using standard accuracy metrics. Finally, the ensemble majority vote approach was applied to the outputs of the two best-performing models to produce a more stable and reliable classification. These workflow components are described in detail in the following sections.

2.2.1. Preparation of Remote Sensing Data

For this study, a fusion of high-temporal S1 SAR and S2 optical data, sourced from the cloud computing platform Google Earth Engine (GEE) [10], was explored in the proposed workflow. The pre-processed GEE satellite imagery was acquired during the rice crop growing season (i.e., the vegetative, reproductive, and ripening stages). The multi-temporal optical and SAR imagery was used to calculate a series of SAR- and optical-derived indices valuable for extracting phenological information defining critical cyclic and seasonal growing patterns associated with rice crops. With this approach, the proposed workflow can capture unique phenological patterns vital in accurately mapping paddy rice crop fields.

The S1 SAR imagery collection, sourced from GEE, consists of backscatter coefficient values processed from Level 1 Ground Range Detected scenes, which have undergone pre-processing steps (such as radiometric calibration, terrain correction, image pre-projection, and speckle removal) [9, 10]. The mean monthly composites of the S1-SAR image collection (acquired over the paddy rice growing stages from April to October in 2018) were calculated using the GEE platform. Table 2 presents details of the Sentinel satellite data explored in this study.

The S2 optical imagery used in the study was also sourced from GEE and was acquired during the same time frame as the S1 imagery. For April to October of 2018, a total of 511 scenes were used to calculate optical-derived vegetation indices (VIs) to capture phenological information over the rice crop growing season. The S2-derived optical indices explored in the study were the NDVI [27], LSWI [2], and EVI [12, 13]. Table 2 outlines the S2 bands and optical indices used in the study. The optical bands used to calculate the optical indices were pre-processed in GEE, including cloud masking and image compositing. Using the median reducer, the median pixel values of the S2 image collection over the rice growth stages were calculated and used to generate the optical-derived VIs.

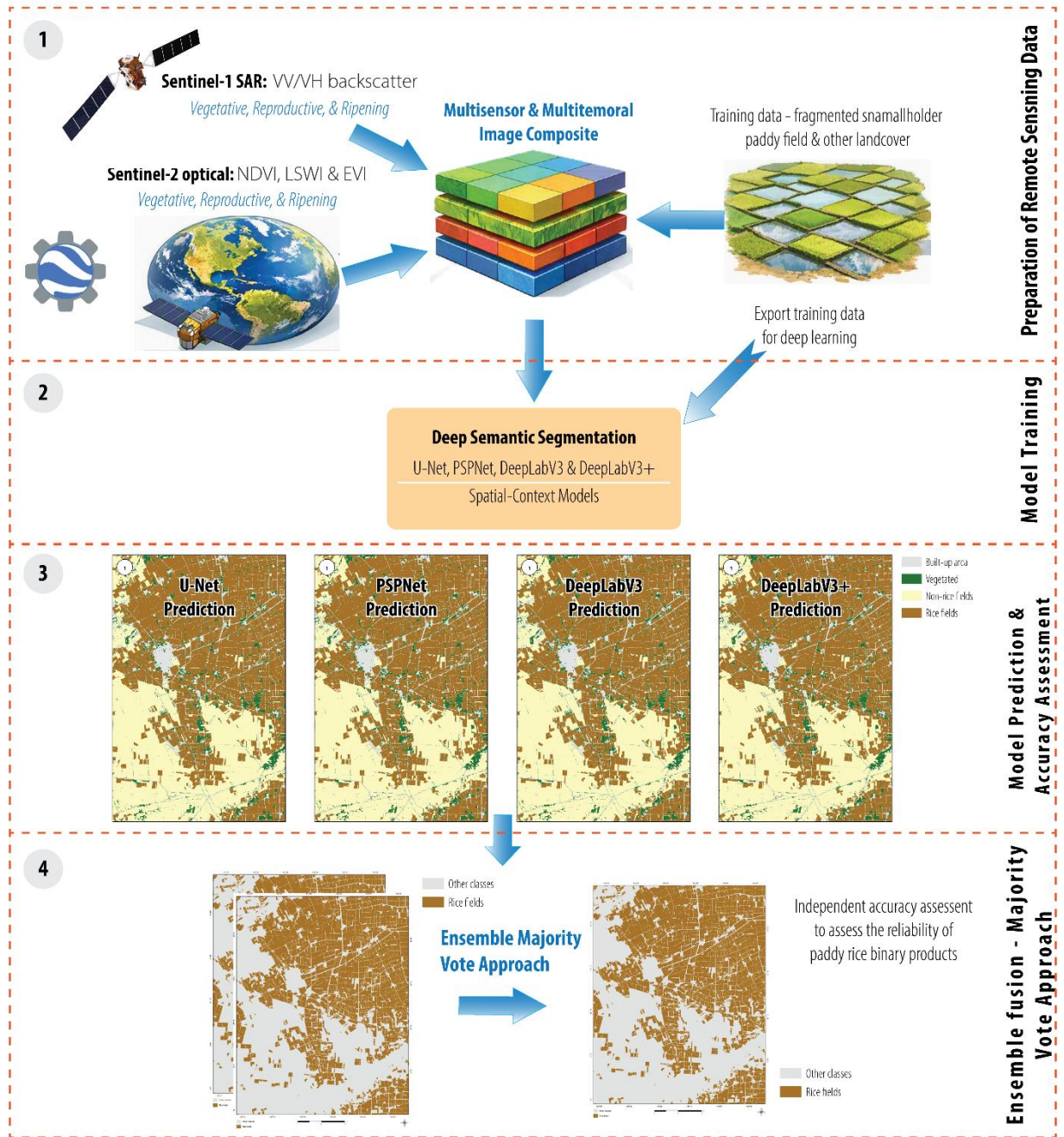


Figure 2. Proposed workflow for paddy rice mapping using deep learning-based semantic segmentation, illustrating preparation of remote sensing data, model training, model prediction and accuracy assessment, and ensemble fusion leading to the final binary rice map.

Table 2. Description of Sentinel-1 SAR and Sentinel-2 optical data and derived variables explored in the study. Where B=band, NDVI=Normalized Difference Vegetation Index, LSWI=Land Surface Water Index, and EVI=Enhanced Vegetation Index.

Sentinel-1 SAR		
Sensor	Sentinel 1A	
Spatial resolution	10 metres	
Wavelength	C-band	
Polarization	Dual-polarization bands—vertical transmission/horizontal receiver—VH/VV	
Orbital properties	Descending	
Instrument mode	Interferometric Wide (IW) swath mode	
# of VH/VV scenes	2219/2233	
SAR variables	VH _{VEG+REP+RIP} + VV _{VEG+REP+RIP}	
Sentinel-2 optical		
Sensor	Sentinel 2A MSI (Multispectral Instrument)	
Spatial resolution	10 / 20 metres	
Band and wavelength (μm)	B2 (Blue)—490 μm	B5 (Red edge 1)—704 μm
	B3 (Green)—560 μm	B6 (Red edge 2)—740 μm
	B4 (Red)—665 μm	B7 (Red edge 3)—780 μm
	B8 (Near-infrared)—842 μm	B8A (Red edge 4)—865 μm
		B11 (SWIR-1)—1614 μm
		B12 (SWIR-2)—2202 μm
# of image scenes	511	
Optical variables	$NDVI = (B8 - B4)/(B8 + B4)$	
	$LSWI = (B8 - B11)/(B8 + B11)$	
	$EVI = 2.5 * ((B8 - B4)/(B8 + 6 * B4 - 7.5 * B2 + 1))$	
Date Range	Vegetative (VEG): 10 April–25 June (2018)	
	Reproductive (REP): 28 June–10 August (2018)	
	Ripening (RIP): 10 August–10 October (2018)	

2.2.2. Deep Learning Models for Rice Crop and Landcover Mapping: Overview

The adoption of deep learning-based semantic segmentation models has become increasingly popular in modern remote sensing workflows, with applications in agricultural cropland delineation and detailed land-cover classification in heterogeneous landscapes. Examples of the most common and influential architectures are U-Net, DeepLabV3, DeepLabV3+, and PSPNet [19, 21, 25, 29, 31, 33, 35, 38], all of which have been shown to perform strongly for both crop-specific applications—such as paddy rice mapping—and broader land-use/land-cover (LULC) classification.

The U-Net classifier architecture consists of an encoder-decoder structure with skip connections, enabling it to retain fine spatial detail while learning hierarchical features [23]. This architecture can effectively delineate irregular and small agricultural parcels, such as rice paddy fields [19, 21]. The encoder component of the U-Net model consists of several convolutional layers followed by rectified linear unit (ReLU) activations and max-pooling layers. The encoder's segmented images are reconstructed into segmented images in the decoder component of the U-Net. Figure 3 shows the structure of the U-Net architecture. The U-Net has been applied to high-

resolution imagery for land-cover classification [35], rice field extraction using multi-temporal S2 imagery [21, 36], and crop classification across agricultural landscapes [17].

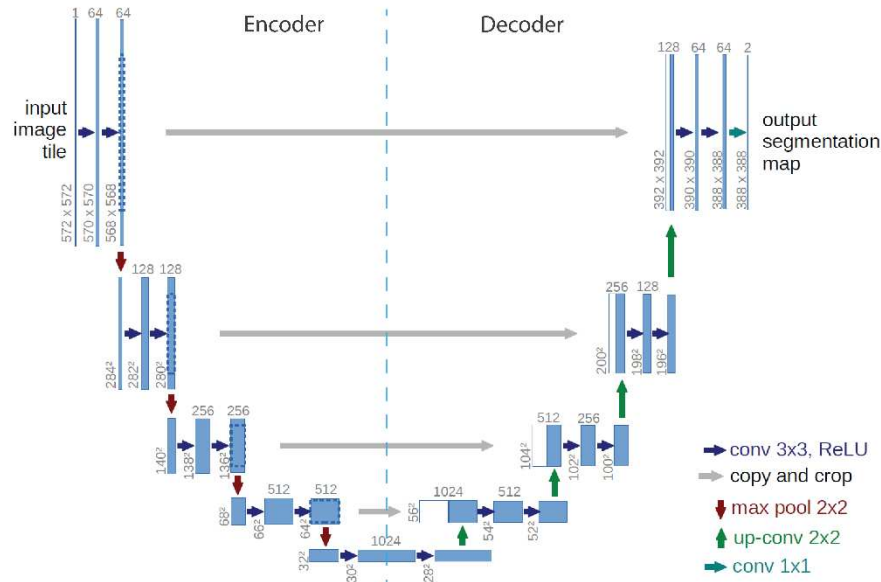


Figure 3. U-Net architecture for image segmentation used in the study, adapted from O Ronneberger et al. [23]

The DeepLabV3 architecture (Figure 4) employs atrous (dilated) convolutions (AC) and Atrous spatial pyramid pooling (ASPP). AC increases the receptive field of convolutional layers without any loss of spatial resolution, allowing the network to capture more contextual information while downsampling the feature maps. The ASPP is designed to gather multi-scale contextual information by simultaneously applying Atrous convolutions with different dilation rates. This process enables the model to recognize objects across multiple scales, improving the accuracy of its image segmentation outputs. For this study, the ResNet34 backbone architecture was used to build the DeepLabV3 semantic model.

DeepLabV3+, an advanced version of the DeepLab series, uses atrous convolution to capture multi-scale context, improving segmentation accuracy. By introducing a decoder module to refine segmentation results, particularly along object boundaries, the capabilities of typical DeepLabV3 are extended. By further exploring the Xception model and applying depth-wise separable convolution to the ASPP and decoder modules, a faster and stronger encoder-decoder network is realized [5].

The PSPNet architecture is centred on a pyramid pooling module (PPM) which aggregates global context information at multiple scales of the input dataset [18, 37]. As with a typical segmentation model, the PSPNet architecture includes an encoder and a decoder. The PSPNet encoder consists of a CNN backbone with dilated convolutions and the PPM. In this architecture, the traditional convolutional layers are replaced with dilated convolutions [34], which help increase the receptive field. As shown in Figure 5, the dilated convolution layers are placed in the last two blocks of the backbone, resulting in richer features at the end of the backbone. After the

encoder has completed feature extraction from the input image dataset, the decoder converts these features into predictions by passing them through its layers. For the PSPNet, an 8x bilinear-upsampling decoder is employed.

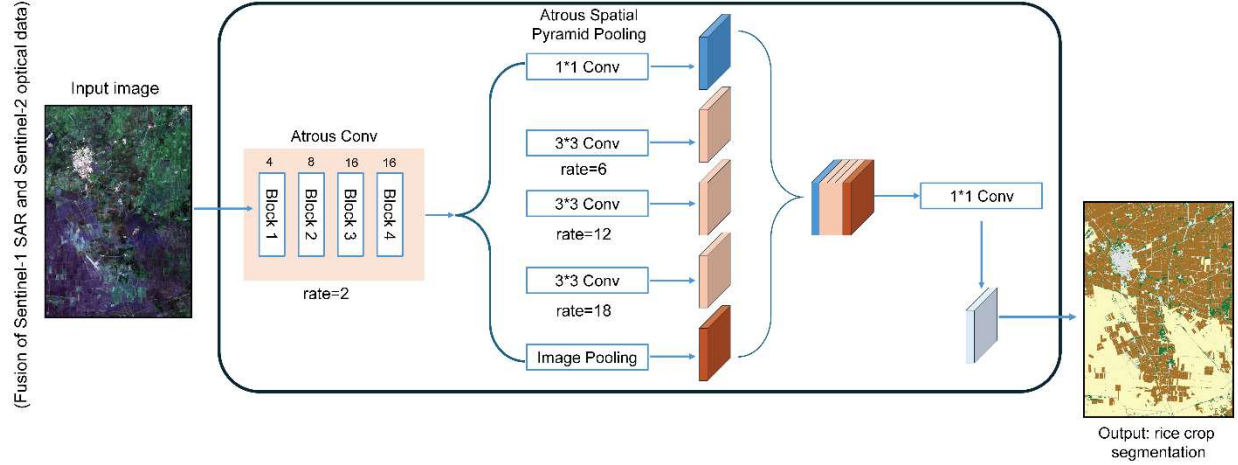


Figure 4. A schematic showing the structure of the DeepLabV3 deep learning model, adapted from Esri Developers [7]

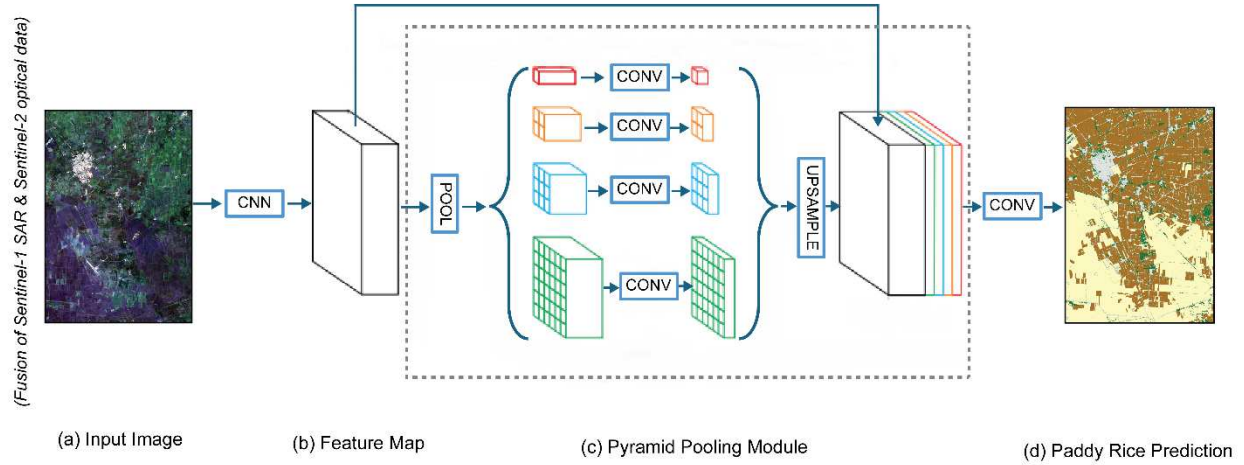


Figure 5. A schematic showing the structure of the PSPNet deep learning model, adapted from Esri Developers [8]

2.2.3. Training data and deep learning parameters

For this study, multi-sensor imagery (consisting of S1-SAR and S2-optical-derived variables) and training and testing data sourced from an existing paddy rice inventory were used to generate image chips and corresponding labels for the deep learning models evaluated. The image chips and labels had dimensions of 256-pixel rows and 256-pixel columns. The metadata format

employed was classified tiles, which allow exporting one classified chip per input image and are suitable for pixel-based classifications. Figure A.1 in the Supplementary Information shows the spatial extent of training/testing data used to build deep learning models assessed in the study.

2.2.4. Accuracy Assessment of Models

The study's accuracy assessment component was conducted using a combination of error metrics. The accuracy of the three DL classifiers (U-Net, DeepLab, and PSPNet) and the RF model was assessed on independent validation data. For this study, the error matrix tables were used to compute paddy rice accuracy metrics, namely precision, recall, F1 score, and overall accuracy (OA) (see equations 1-4).

$$\text{User's Accuracy (Precision)} = \frac{\text{True Positives (TP)}}{\text{True Positives} + \text{False Positives (FP)}} \quad (1)$$

$$\begin{aligned} \text{Producer's Accuracy (Recall)} \\ = \frac{\text{True Positives (TP)}}{\text{True Positives} + \text{False Negatives (FN)}} \end{aligned} \quad (2)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{OA} = \frac{S_d}{n} * 100 \quad (4)$$

Where S_d = total correctly classified pixels and n = the total number of validation pixels.

2.2.5. Ensemble Fusion

For this study, an ensemble classification approach based on majority voting was used to combine the predictions of the two most accurate independent DL models. This method exploits the strengths of the most accurate DL predictions. In this study, the final class assigned to each pixel was determined from the degrees of agreement between two classification outputs. By adopting this approach, the model-specific errors are reduced, and the robustness of the final classification is increased. The majority-voting ensembles have been shown to relatively improve the classification performance of individual classifier models by reducing variance and mitigating misclassification associated with spectral confusion and heterogeneous land-cover conditions [6, 15]. The last step of this proposed workflow was to perform an ensemble classification using the two most effective DL models, as indicated by the accuracy assessment results. To assess the performance of this approach, the ensemble output and the two optimal DL model predictions were reclassified into a binary product (non-rice and rice fields), and their performance was evaluated on an independent set of ground validation data points.

3. Results

3.1. Comparison of deep learning models based on training and validation loss dynamics

The three evaluated semantic segmentation architectures—U-Net, DeepLabV3, and PSPNet—showed distinct learning patterns during model optimization, as reflected in their respective training and validation loss trajectories (see Figure 6). For all three DL models, the ResNet34 backbone architecture was adopted to allow for cross-comparison. The differences in the train and validation trajectories were attributed to the architectural design choices, which are known to drive feature reuse, multi-scale context extraction, and the understanding of global scenes, factors that are crucial for accurately delineating smallholder paddy rice fields from surrounding heterogeneous landscapes.

Overall, the U-Net model demonstrated the fastest reduction in training loss, consistent with its encoder-decoder structure and dense skip connections, which facilitated efficient gradient flow and rapid learning of fine-scale spatial detail. For the validation loss, the U-Net model plateaued earlier and maintained a larger gap compared to the training loss. The observed pattern suggests a high susceptibility to overfitting, particularly in landscapes where rice fields exhibit spectral similarities to surrounding land-cover types. In this study, this scenario is mitigated by the combined use of SAR and optical proxies that capture rice crop phenology across the key growing stages (vegetative, reproductive, and ripening). Though the U-Net model is known to effectively capture boundary-level detail, its limited capability for multiscale contextual modelling hampers its ability to generalize across spatially fragmented agricultural field parcels.

Compared with the other DL models, DeepLabV3 exhibited a more gradual decline in both training and validation loss (Figure 6). The similarities between the curves of the DeepLabV3 model were indicative of a stronger generalization performance. The ASPP module in the DeepLabV3 architecture enables the model to integrate multi-scale contextual information, reducing confusion due to spectral similarities between classes and improving robustness across varying rice phenological stages. The resilience of DeepLabV3 to noise and intra-class variability, common in Sentinel-1 SAR and Sentinel-2 optical data, is evident in the reduced oscillations in the validation loss chart.

The enhanced variant, DeepLabV3+, showed a more stable, tightly coupled training-validation loss pattern than the evaluated models. By including a decoder module, DeepLabV3+ showed improved boundary refinement while preserving the multi-scale contextual advantages of ASPP. The training and validation loss chart showed a consistently lower validation loss than training loss throughout optimization (indicating efficient generalization and reduced overfitting).

The PSPNet model showed steady convergence with the validation loss, which subsequently outperformed the training loss during the early epochs. The pyramid pooling module of PSPNet effectively combines global contextual information (a vital component that captures the irregularity of paddy rice fields), mixed land-use patterns, and variable hydrological conditions in the study area. The global-context feature of the PSPNet contributed to smoother validation-loss behaviour and reduced class confusion, particularly when distinguishing rice paddies from surrounding moisture-related land-cover classes.

Overall, the comparison of the three evaluated DL models' loss profiles showed that while the U-Net excels in feature learning, the DeepLabV3+ and PSPNet exhibit more stable and generalizable performance for paddy rice and LULC mapping. The strong alignment between the

training and validation losses for DeepLabV3+ and PSPNet further highlights their suitability for operational workflows, given their robustness across diverse datasets.

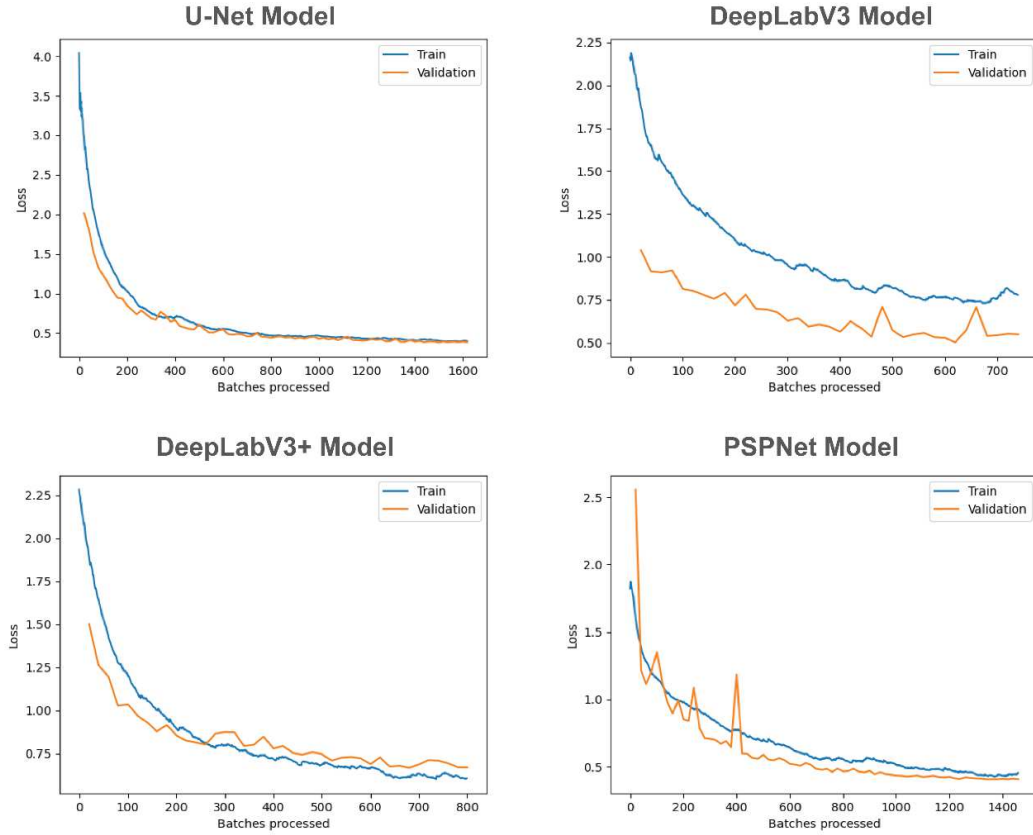


Figure 6. Training and validation loss curves for U-Net, DeepLabV3, DeepLabV3+, and PSPNet. U-Net shows rapid training-loss convergence but a larger train–validation gap. DeepLabV3 exhibits stable, aligned loss reduction, while DeepLabV3+ maintains the closest coupling between curves, indicating strong generalization. PSPNet converges smoothly, with a consistently stable validation loss, due to effective global-context aggregation.

3.2. Optimal deep learning paddy rice predictions

Figure 7 presents the accuracy metrics of the DL paddy rice predictions evaluated in the study. Based on the accuracy assessment results from this study, the four evaluated DL model architectures showed that the U-Net and PSPNet demonstrated the strongest and most balanced segmentation performance, each achieving a Mean F1 of 0.80 and an IoU of 0.87. The U-Net model achieved the highest F1 score (0.94), indicating superior per-class discrimination and boundary preservation, both of which can be attributed to its dense skip-connection design. PSPNet performed similarly to U-Net, with a high F1 score of 0.93, reflecting the effectiveness of its pyramid pooling module in capturing global contextual information. The DeepLabV3+ generated comparative results (Mean F1 = 0.79; Rice F1 = 0.93; IoU = 0.87), with slightly lower Mean F1 score. Compared with the other DL models, DeepLabV3 had the lowest performance

across all metrics (Mean F1 = 0.71; Rice F1 = 0.91; IoU = 0.83), indicating the importance of decoder refinement for accurate boundary reconstruction. Overall, the results highlight that architectures capable of combining strong multi-scale context with effective spatial detail recovery (such as U-Net and PSPNet) provide the most reliable segmentation performance for multi-date and multi-sensor satellite data as explored in this study.

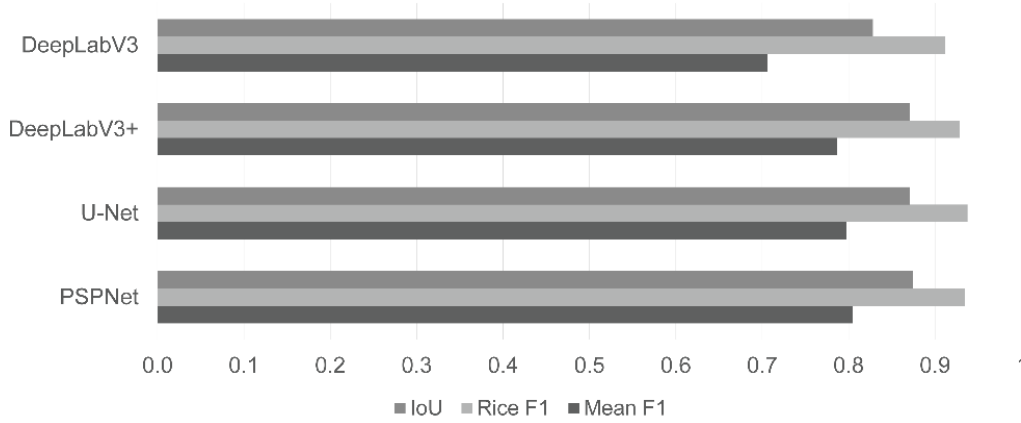


Figure 7. Classification accuracy results of evaluated deep learning models U-Net, DeepLabV3, DeepLabV3+, and PSPNet

3.3. Ensemble Analysis Results

Based on the proposed workflow, the two best-performing DL models (U-Net and PSPNet) were combined using a majority-vote ensemble approach. This approach has been shown to improve classification accuracy, spatial consistency, and the reliability of thematic products by leveraging complementary information from multiple DL approaches. Although the individual architectures (U-Net and PNPNet) demonstrated good performance, the ensemble of the two models yielded more accurate paddy rice predictions (see Figure 8). Figure 8 shows the accuracy assessment chart of binary paddy rice map products derived using the ensemble approach, U-Net, and PNPNet models. Figure A.2 shows the spatial distribution of ground validation data points used for an independent assessment of the paddy rice binary products generated in this component of the study. The accuracy assessment results indicated that the U-Net DL model had a better combination of overall accuracy, mean F1-score, and rice-specific F1-score (0.934, 0.916, and 0.878) compared to the PNPNet DL model (with overall accuracy = 0.924, mean F1-score = 0.916, and rice F1-score = 0.878). The U-Net DL model reflects its ability to preserve fine spatial detail through skip connections, while the PNPNet model benefits from its enhanced multi-scale contextual feature extraction. The ensemble of the U-Net and PNPNet performed better than both models, with an overall accuracy of 0.940, a mean F1 score of 0.923, a rice-specific F1 score of 0.888, and the highest Kappa value (0.847), indicating improved class agreement and reduced misclassification. Table A.1 in the Supplementary Information shows the error matrix of paddy rice binary products generated in the study (U-Net, PSPNet, DeepLabV3, DeepLabV3+, and an ensemble majority-vote product based on U-Net and PSPNet). The ensemble approach results highlight the value of combining architectures optimized for spatial detail (U-Net) and contextual representation (PSPNet), yielding more robust and generalizable segmentation performance than either model alone. Figure 9 shows DL segmented maps of an urban center in Jiansanjiang

surrounded by paddy rice fields, vegetated landscape, and non-rice fields. Figure 10 shows the ensemble-fused paddy rice map generated from U-Net and PSPNet, which shows improved spatial coherence and classification accuracy compared with the individual model outputs.

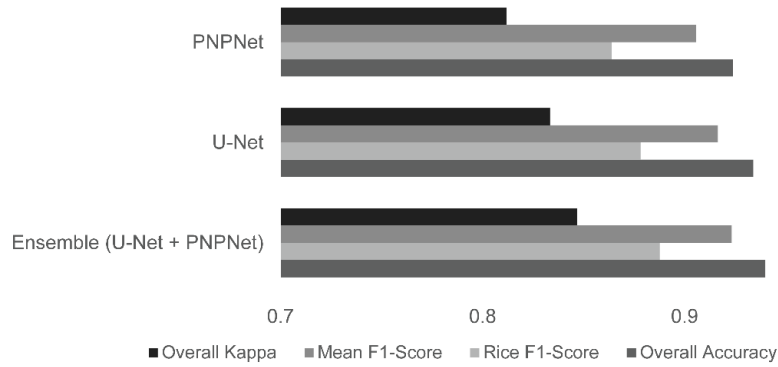


Figure 8. Accuracy assessment results for binary paddy rice maps generated using the ensemble model, U-Net, and PSPNet. The chart summarizes overall accuracy, rice-specific F1-scores, mean F1-scores, and Cohen’s Kappa values.



Figure 9. Paddy rice segmentation outputs produced by four deep learning architectures— (1) U-Net, (2) DeepLabV3, (3) DeepLabV3+, and (4) PSPNet—alongside (5) an ensemble map derived from the

best-performing models (PSPNet and U-Net). All predicted maps are compared against (6) a true-colour Sentinel-2 composite from the final evaluation dataset.



Figure 10. Ensemble binary paddy rice map generated from U-Net and PSPNet, illustrating improved spatial coherence and classification accuracy compared with the individual model outputs

4. Discussion

4.1. *Comparison of Deep Learning Model Architectures*

This study analyzed differences among deep learning models, including U-Net, PSPNet, DeepLabV3, and DeepLabV3+, primarily arising from their architectural designs. U-Net’s encoder–decoder architecture with dense skip connections effectively recovers fine spatial details, enhancing class discrimination and boundary accuracy, making it well-suited for map segmentation [23]. PSPNet integrates pyramid pooling to gather global context across multiple scales, aiding in the resolution of large, coherent patterns and maintaining stable generalization [37]. DeepLabV3 uses atrous convolutions and ASPP to enlarge the receptive field without losing resolution but lacks a decoder, limiting its ability to accurately delineate object boundaries and resulting in lower IoU and F1 scores [4], as observed here. These issues are addressed in DeepLabV3+, which features a lightweight decoder that refines segmentation edges and boundary localization, achieving more balanced performance across classes [5]. Overall, U-Net and PSPNet demonstrate superior boundary and class-specific metrics compared to DeepLabV3 and DeepLabV3+, which tend to underperform due to the absence of decoder refinement. consistently underperform due to the lack of decoder refinement functionality.

This study shows that adopting an ensemble majority-vote approach provides a statistically robust means of improving the thematic accuracy of DL-based land-cover products. Ensemble fusion has been widely recognized to mitigate model-specific limitations and stabilize predictions across diverse scenarios, harnessing the strengths of individual architectures [24]. For this study, combining U-Net and PSPNet DL models through a majority-vote ensemble approach resulted in a more coherent and accurate paddy rice map than either of the two identified models. This approach reflects the value of combining complementary fine-scale feature recovery from U-Net and PSPNet’s multi-scale contextual representation. The accuracy assessment results indicate that the ensemble fusion of the U-Net and PSPNet models reduced systematic misclassifications while enhancing generalization, supporting its use as a practical, operationally robust approach for large-scale agricultural mapping and monitoring.

4.2. *Potential upscaling of proposed workflow*

Based on the proposed operational workflow, the study highlights the potential for expansion and adoption among stakeholders and policymakers.

- **Exploiting cloud computing platforms for regional and national scales:**

The proposed workflow builds on Google Earth Engine, making it suitable for scaling across larger paddy-rice regions, as demonstrated in AO Onojeghuo et al. [21]. By extending the approach to multi-year time series, it is then possible to support nationwide crop monitoring and early warning systems. Other platforms, such as the Microsoft SpatioTemporal Asset Catalogue (STAC) API, can be used to access indexed geospatial products, such as remote sensing data, for similar applications. To demonstrate the applicability of this workflow to upscaling, the U-Net deep learning package developed for the study area was upscaled to parts of Qianjin District. To facilitate visualization of the upscaled paddy rice map product, an interactive map application was developed for

external stakeholders. This can be accessed using the link:

<https://experience.arcgis.com/experience/dc91e8b0ecea41a49b6d8ba1f91fba82>.

- **Integration of high-resolution commercial products:**

For future studies, incorporating commercial products such as PlanetScope, SPOT-6/7, or WorldView imagery has the potential to produce finer-resolution paddy rice delineated maps that better represent highly fragmented smallholder landscapes and better support precision-agriculture applications.

- **Developing region-specific training models:**

Considering the unique agroecological zones across Asia, West Africa, and Latin America, training models specific to these regions will be advantageous for supporting global rice mapping frameworks with improved spatial detail compared to commonly used phenology-only approaches.

- **Integration of paddy rice with agro-hydrological and economic models:**

The inclusion of high-resolution paddy rice map products derived from remote sensing data provides invaluable information on the spatial distribution of the crop fields, cropping intensity, and phenology dynamics, all of which are valuable inputs for agro-hydrological and economic model development. By incorporating paddy rice extent map information into watershed- or basin-scale models, one can better estimate or simulate irrigation demand, surface water runoff, groundwater recharge, and controlled irrigation. Information from rice extent maps has immense potential for integration with irrigation scheduling, water-use modelling, carbon accounting, and rural livelihood assessments. This can form part of a future study in which paddy rice proxies can be integrated in economic models.

5. Conclusion

Overall, this study demonstrates the value of integrating multi-temporal Sentinel-1 SAR and Sentinel-2 optical-derived phenological metrics with modern deep learning segmentation architectures to provide a robust and scalable framework for delineating smallholder paddy rice fields and for monitoring. The combination of a multi-sensor approach with reliable, field-based data enables a robust, scalable method for accurate paddy rice mapping across diverse agroecological settings. The results show that, of the four evaluated DL models, U-Net and PSPNet consistently exhibited the best and most balanced performance, benefiting, respectively, from architectural features, fine-detail preservation, and multi-scale contextual reasoning. Furthermore, fusing the optimal paddy rice products using the ensemble majority-vote approach improved paddy rice classification with thematic consistency, reduced misclassification, and produced the highest overall accuracy, rice-specific F1-score, and Kappa values, highlighting the value of combining complementary architectural strengths.

Furthermore, the results of this study underscore the importance of leveraging information from multi-sensor phenological fusion in heterogeneous smallholder landscapes, where cloud cover,

mixed pixels, and rapid phenological transitions pose challenges for traditional mapping approaches. The improvement in segmentation accuracy from adapting DL model architectures indicates that the resulting operational deployment system is useful for agricultural monitoring, food security, and subsidy allocation programs.

Supplementary Materials: The following supporting information can be downloaded at: xxx. The Supplementary Materials include a list of figures and tables related to the results of this study. In addition, the document includes a link to an interactive mapping application that shows examples of the study's results. Figure A.1: Map showing spatial extent of training/testing data used to build deep learning models assessed in the study (adopted from AO Onojeghuo et al. [21]). Figure A.2: Map showing spatial distribution of ground validation points used for an independent assessment of the paddy rice binary products. Table A.1: Error matrix of paddy rice binary products generated in the study. The evaluated architectures included U-Net, PSPNet, DeepLabV3, DeepLabV3+, and an ensemble majority-vote product based on U-Net and PSPNet.

Author Contributions: Conceptualization, A.O.O.; software, A.O.O.; validation, A.O.O.; formal analysis, A.O.O.; investigation, A.O.O.; resources, A.O.O.; data curation, A.O.O.; writing—original draft preparation, A.O.O.; review and editing, A.O.O. and A.R.O.; funding acquisition, A.O.O and A.R.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data generated or analyzed during this study are included in the manuscript. An inventory of the Earth observation data and results generated in this study is available as an ArcGIS Online experience builder app for dissemination. The interactive map can be accessed via this link:

<https://experience.arcgis.com/experience/dc91e8b0ecea41a49b6d8ba1f91fba82> (accessed on 28 August 2026).

Acknowledgments: The author would like to thank the anonymous reviewers and editors for their invaluable contributions to improving the quality of this manuscript.

Conflicts of Interest: The author declares that no known competing financial interests or personal relationships could have appeared to influence the work reported in this paper.

References

1. Boschetti M, Nutini F, Manfron G et al. (2014) Comparative analysis of normalised difference spectral indices derived from MODIS for detecting surface water in flooded rice cropping systems. *PloS one* 9:e88741. <https://doi.org/10.1371/journal.pone.0088741>
2. Chandrasekar K, Sesha Sai M, Roy P et al. (2010) Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product. *International Journal of Remote Sensing* 31:3987–4005. <https://doi.org/10.1080/01431160802575653>
3. Chen C, Liu J, Deng Y et al. (2026) Integrating Multi-Source and Multi-Temporal UAV Observations to Improve Wheat Yield Prediction Using Machine Learning. *Plants* 15:1345. <https://doi.org/10.3390/plants15091345>
4. Chen L-C, Papandreou G, Kokkinos I et al. (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. In *IEEE transactions on*

- pattern analysis and machine intelligence, vol. 40, no., p 834–848, <https://doi.org/10.1109/TPAMI.2017.2699184>.
5. Chen L, Zhu Y, Papandreou G et al. (2018) Deeplab: Deep labelling for semantic image segmentation. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, p 834–848, 1 April 2018, <https://doi.org/10.1109/TPAMI.2017.2699184>.
 6. Cui G, Lv Z, Li G et al. (2018) Refining Land Cover Classification Maps Based on Dual-Adaptive Majority Voting Strategy for Very High Resolution Remote Sensing Images. Remote Sensing 10:1238. <https://doi.org/10.3390/rs10081238>
 7. Esri Developers (2026) How DeepLabV3 Works? ESRI. <https://developers.arcgis.com/python/latest/guide/how-deeplabv3-works/>. Accessed 2 August 2026.
 8. Esri Developers (2026) How PSPNet works? ESRI. <https://developers.arcgis.com/python/latest/guide/how-pspnet-works/>. Accessed August 2, 2026.
 9. Fitrzyk M, Engdahl M, Fernandez D (2019) ESA Copernicus Sentinel-1 Exploitation Activities. In IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, vol. no., p 5389–5392, 28 July–2 Aug. 2019, <https://doi.org/10.1109/IGARSS.2019.8898633>.
 10. Gorelick N, Hancher M, Dixon M et al. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202:18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
 11. Heryadi Y, Irwansyah E, Miranda E et al. (2022) Pyramid Scene Parsing Net Model for Automated Paddy Field Map using SPOT 6 Satellite Images. In 2022 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS), vol. no., p 104–108, <https://doi.org/10.1109/AGERS56232.2022.10093406>.
 12. Huete A, Didan K, Miura T et al. (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83:195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
 13. Huete A, Liu H, Batchily K et al. (1997) A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote sensing of environment 59:440–451.
 14. Irri (2016) Important Management Factors by Growth Stage. International Rice Research Institute (IRRI). <http://www.knowledgebank.irri.org/decision-tools/growth-stages-and-important-management-factors>. Accessed August 1, 2026.
 15. Kulkarni K, Vijaya PA (2023) A Majority Voting Ensemble Approach for LULC Classification of Satellite Images. Journal of The Institution of Engineers (India): Series B 104:327–333. <https://doi.org/10.1007/s40031-023-00865-4>
 16. Le Toan T, Ribbes F, Wang L-F et al. (1997) Rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results. IEEE Transactions on Geoscience and Remote Sensing 35:41–56. <https://doi.org/10.1109/ACCESS.2020.3041038>
 17. Li G, Cui J, Han W et al. (2022) Crop type mapping using time-series Sentinel-2 imagery and U-Net in early growth periods in the Hetao irrigation district in China. Computers and Electronics in Agriculture 203:107478. <https://doi.org/10.1016/j.compag.2022.107478>
 18. Long X, Zhang W, Zhao B (2020) PSPNet-SLAM: A Semantic SLAM Detect Dynamic Object by Pyramid Scene Parsing Network. IEEE Access 8:214685–214695. <https://doi.org/10.1109/ACCESS.2020.3041038>
 19. Mubarak MJ, Yuniarno EM, Sihombing YO et al. (2024) Semantic Segmentation of Sentinel-2 Satellite Image for Rice Growth Phase Classification Using Deep Learning. In 2024 IEEE International Conference on Imaging Systems and Techniques (IST), vol. no., p 1–6, 14–16 Oct. 2024, <https://doi.org/10.1109/IST63414.2024.10759168>.
 20. Onojeghuo AO, Blackburn GA, Wang Q et al. (2018) Mapping paddy rice fields by applying machine learning algorithms to multi-temporal Sentinel-1A and Landsat data. International journal of remote sensing 39:1042–1067. <https://doi.org/10.1080/01431161.2017.1395969>

21. Onojeghuo AO, Miao Y, Blackburn GA (2023) Deep ResU-Net Convolutional Neural Networks Segmentation for Smallholder Paddy Rice Mapping Using Sentinel 1 SAR and Sentinel 2 Optical Imagery. *Remote Sensing* 15:1517. <https://doi.org/10.3390/rs15061517>
22. Ramadhani F, Pullanagari R, Kereszturi G et al. (2020) Automatic Mapping of Rice Growth Stages Using the Integration of Sentinel-2, MOD13Q1, and Sentinel-1. *Remote Sensing* 12:3613. <https://doi.org/10.3390/rs12213613>
23. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*. vol 9351. Springer International Publishing, Cham, p 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
24. Sarkar TK, Roy DK, Kang YS et al. (2024) Ensemble of Machine Learning Algorithms for Rice Grain Yield Prediction Using UAV-Based Remote Sensing. *Journal of Biosystems Engineering* 49:1–19. <https://doi.org/10.1007/s42853-023-00209-6>
25. Shao Y, Pan P, Zhao H et al. (2025) Semantic Segmentation of Rice Fields in Sub-Meter Satellite Imagery Using an HRNet-CA-Enhanced DeepLabV3+ Framework. *Remote Sensing* 17:2404. <https://doi.org/10.3390/rs17142404>
26. Singha M, Dong J, Zhang G et al. (2019) High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. *Scientific Data* 6:26. <https://doi.org/10.1038/s41597-019-0036-3>
27. Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8:127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
28. Wang F-M, Huang J-F, Wang X-Z (2008) Identification of Optimal Hyperspectral Bands for Estimation of Rice Biophysical Parameters. *Journal of Integrative Plant Biology* 50:291–299. <https://doi.org/10.1111/j.1744-7909.2007.00619.x>
29. Wei P, Huang R, Lin T et al. (2022) Rice Mapping in Training Sample Shortage Regions Using a Deep Semantic Segmentation Model Trained on Pseudo-Labels. *Remote Sensing* 14:328. <https://doi.org/10.3390/rs14020328>
30. Wu F, Wang C, Zhang H et al. (2011) Rice Crop Monitoring in South China With RADARSAT-2 Quad-Polarization SAR Data. *IEEE Geoscience and Remote Sensing Letters* 8:196–200. <https://doi.org/10.1109/LGRS.2010.2055830>
31. Xiao W, Xu S, He T (2021) Mapping Paddy Rice with Sentinel-1/2 and Phenology-, Object-Based Algorithm—A Implementation in Hangjiahu Plain in China Using GEE Platform. *Remote Sensing* 13:990. <https://doi.org/10.3390/rs13080990>
32. Xiao X, Boles S, Liu J et al. (2005) Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote sensing of environment* 95:480–492. <https://doi.org/10.1016/j.rse.2004.09.005>
33. Xu H, Song J, Zhu Y (2023) Evaluation and Comparison of Semantic Segmentation Networks for Rice Identification Based on Sentinel-2 Imagery. *Remote Sensing* 15:1499. <https://doi.org/10.3390/rs15061499>
34. Yu F, Koltun V, Funkhouser T (2017) Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. no., p 472–480, <https://doi.org/10.48550/arXiv.1705.09914>.
35. Zhang P, Hu S, Li W et al. (2020) Parcel-level mapping of crops in a smallholder agricultural area: A case of central China using single-temporal VHSR imagery. *Computers and Electronics in Agriculture* 175:105581. <https://doi.org/10.1016/j.compag.2020.105581>
36. Zhao E, Zhang W, Wang Y et al. (2026) A Spatial-Temporal Attention-Based U-Net for Crop Mapping from Time-Series Sentinel-2 Imagery: A Case in Sanjiang Plain. *Remote Sensing* 18:1677. <https://doi.org/10.3390/rs18111677>
37. Zhao H, Shi J, Qi X et al. (2017) Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. no., p 2881–2890, <https://doi.org/10.1109/CVPR.2017.310>

38. Zhao Z, Dong J, Zhang G et al. (2024) Improved phenology-based rice mapping algorithm by integrating optical and radar data. *Remote Sensing of Environment* 315:114460.
<https://doi.org/10.1016/j.rse.2024.114460>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)