Bulletin of Earth Sciences of Thailand

CropNet: Leveraging SegFormer for Efficient and Scalable Crop Mapping with Sentinel-2 Data

Sathirada Phahurat , Pongthep Thongsang, Srilert Chotpantarat*

Department of Geology, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand

* Corresponding author: Srilert.c@chula.ac.th

Received: 06 Jun 2025
Revised: 15 Jul 2025
Accepted: 17 Jul 2025

Abstract

This research investigates a deep learning-based methodology for crop classification by integrating Sentinel-2 satellite imagery with SegFormer, a state-of-the-art transformer-based semantic segmentation model. The study focuses on five dominant land cover types: rice fields, sugarcane, cassava, para rubber, and pond areas within a part of Khu Mueang District, Buriram Province, Thailand. The main objectives are to develop an efficient classification method using Sentinel-2 satellite data and to evaluate the predictive performance of SegFormer in the agricultural field. Satellite images were acquired via Google Earth Engine (GEE) during the harvest season (Nov 2023–Jan 2024), complemented by ground truth data collected from field surveys and high-resolution drone imagery. Preprocessing steps included cloud filtering, image normalization, and manual pixel-level labeling in QGIS software. The dataset was divided into 512×512 pixel patches, resulting in 780 image–mask pairs allocated for training (480), validation (120), and testing (180). The SegFormer model was trained using Optuna to find the best hyperparameter settings. The model achieved 0.967 pixel-wise accuracy with a validation loss of 0.075 (cross-entropy) on the training and validation datasets, demonstrating strong learning performance during model development. It showed strong classification performance for para rubber and sugarcane. However, it faced challenges in distinguishing cassava, ponds, and bare soil due to class imbalance and spectral similarity.

Keywords: Deep learning; SegFormer; Sentinel-2 satellite image; Crop classification; UAV

1. Introduction

Agriculture continues to be a major contributor to ensuring global food security, environmental stability, and national economic development. According to the United Nations (2017) the global population is projected to increase from 7.6 billion in 2017 to 9.8 billion by 2050 which means the demand for agricultural products will increase significantly. However, the agricultural sector is still facing increasing problems such as climate change resulting in unpredictable weather conditions, limited agricultural land, and water scarcity (Food and Agriculture Organization of the United Nation, 2015). These challenges adversely affect crop productivity and compel agricultural systems to adopt more efficient and sustainable. These concerns are especially relevant in Thailand, where agriculture remains a key economic activity as well as a source of

income for people residing in the countryside. Thailand had approximately 8.7 million agricultural landholders (National Statistical Office Thailand, 2021). However, this sector has shown slower growth than other industries, with a decline in both the quantity and quality of agricultural production. Key contributing factors include the effects of climate change, limited technological adoption, and inefficient management of water and land resources. These constraints highlight the need for modern, data-driven solutions that can enhance monitoring, productivity, and long-term sustainability in Thai agriculture.

Among the key data-driven solutions, crop classification also plays a vital role in agricultural resource management. Accurate crop maps contribute to improved water management by enabling precise estimation of crop water requirements, planning of irrigation schedules, and assessment of water use efficiency. Furthermore, timely and accurate crop classification supports crop monitoring and early detection of stress conditions, such as droughts or pest infestations, which is crucial for mitigating crop losses. These capabilities are especially important for enhancing agricultural productivity and ensuring food security under climate change conditions. Therefore, a robust crop classification system has practical significance not only in terms of technological innovation but also in supporting data-driven decision-making for sustainable agriculture in Thailand.

To address the challenges in agriculture, the application of modern technology has become more important. Remote sensing technologies, particularly satellite imagery, have emerged as valuable tools for large-scale, real-time agricultural monitoring. When combined with artificial intelligence (AI) and deep learning (DL) models, these technologies can automate tasks such as crop classification, crop health monitoring, and yield estimation (Ma et al., 2019). Among various deep learning techniques, Convolutional Neural Networks (CNNs) are widely applied due to their superior performance in image-based classification tasks (Abdi, 2019). However, CNNs typically require large training datasets and might be computationally expensive.

Recent advancements in transformer-based models, like SegFormer (Xie et al., 2021) have provided new direction for semantic segmentation models. SegFormer provides high accuracy and efficiency with a lightweight architecture, making it appropriate option for remote sensing applications. A Despite its advantages, the application of SegFormer in agricultural crop classification remains limited, especially in Thai farm systems, and further evaluation under local conditions is required. The study aims to address this gap by developing an efficient methodology for crop classification using Sentinel-2 Multi-Spectral Instrument, Level-2A satellite imagery. Specifically, it focuses on evaluating the performance of the SegFormer model in classifying major crop types within a selected area covering a part of Khu Mueang district, Buriram Province, Thailand.

2. Material and Method

2.1 Study area

This research was conducted in a selected agricultural area located in Khu Mueang District, Buriram Province, situated in the lower northeastern part of Thailand. The study area includes parts of three subdistricts of Khu Mueang, Hin Lek Fai, and Phon Samran,

covering approximately 4.7 square kilometers (Figure 1).

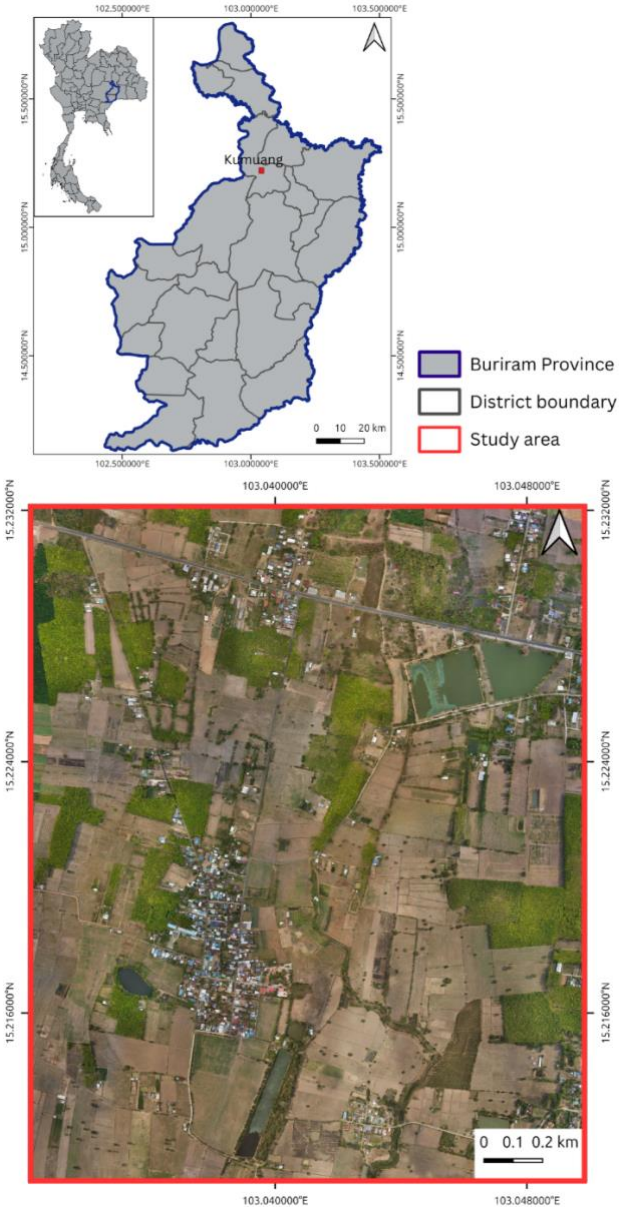


Figure 1 The location of the study area within Kumuang District, Buriram Province, Thailand. The high-resolution drone, with a 10.6 cm resolution, shows the variety of crop types, captured in 6-10 May 2024.

This region is characterized by a tropical savanna climate zone (classified as "Aw" in the Köppen–Geiger system) with distinct wet and dry seasons (Phumkokrux, 2021). The climate alternates between hot and humid conditions during the rainy season and a prolonged dry period with limited precipitation (Meteorological Development Department, 2023). The study area is predominantly composed of sandy loam soils, which are low in organic matter content and have limited nutrients. These environmental conditions strongly influence local agricultural practices, supporting the cultivation of diverse crop types including rice, sugarcane, cassava, and para rubber. These crops represent the dominant cultivated land use in Buriram Province and are considered as Thailand's major economic crops (Land Development Department, 2021). Therefore, this study area was selected based on its agricultural diversity, the availability of ground truth data, and regional land and water resource management concern.

2.2 Data collection and preprocessing

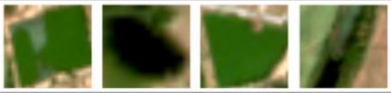
2.2.1 Sentinel-2 satellite images

A total of 26 Sentinel-2 MSI Level-2A images were collected via the Copernicus Open Access Hub using Google Earth Engine (GEE) during the harvesting period (November 2023 and January 2024). RGB images with less than 10% of cloud cover were chosen to ensure quality. The selected scenes were exported in TIFF format to preserve high-resolution and lossless geospatial quality, then converted to PNG format using Python scripts. Each image had a resolution of 10 meters and a size of 2,088×2,600 pixels. Of the total images, 20 images from November–December 2023 were used for model training and validation, while six images from January 2024 were used for the testing set. The GEE script used for image collection and export is available at:

https://code.earthengine.google.com/?scriptPath=users%2Fsathiradap1999%2FDL%3ABRR.

Table 1 presents Sentinel-2 image samples of different crop types. In this research, Sentinel-2 images were collected during the growing seasons of sugarcane and para rubber. Cassava fields exhibited mixed growth stages, while pond areas remained visually stable. Due to spectral variability across rice growth stages, only harvested rice fields were used for classification and this can be confirmed through field surveys and is assumed to remain rice-dedicated within one cropping cycle. As a result, crop growth stages were not differentiated.

Table 1 Visual characteristics of each crop type as observed from Sentinel-2 satellite imagery.

Crop type class	Sentinel-2 satellite image
Rice field (Harvested season)	
Sugarcane (Growing season)	
Para rubber (Growing season)	
Cassava (Mixed season)	
Pond	

(Note: All image samples were displayed using the same spatial scale. The reference scale bar shown at the bottom right applies to all image samples.)

As shown in Table 1, harvested rice fields typically appear as continuous pale brownish patches with rectangular shapes and medium to large field sizes, often located close together. Sugarcane fields are identifiable by their light green tones and elongated rectangular shapes, with medium to large field sizes and relatively uniform vegetation coverage. Para rubber exhibits dense, dark green coverage with blocky plot shapes and consistently large field sizes, reflecting its perennial and structured plantation characteristics. In contrast, cassava plots appear darker brownish, less structured, and more scattered, with typically small field sizes that reflect the variability in planting practices. Pond areas are marked by dark to olive green tones, irregular shapes, and varying sizes, and are often partially obscured by aquatic vegetation. These spectral and spatial differences are critical for distinguishing crop types in remote sensing–based classification.

2.2.2 Drone photography

To ensure spatial accuracy in ground truth labeling, high-resolution drone imagery was captured over the study area on 6–10 May 2024 using a flight altitude of 200 meters. This produced imagery at approximately 10.6 cm/pixel, which was used as visual reference for manual labeling in QGIS.

2.2.3 Field observations

Field surveys were conducted the same period (6–10 May 2024) to verify crop types and collect in situ reference data, which were cross-validated with both satellite and drone imagery.

2.3 Data preparation

The collected data was used to label images into five crop classes—rice, sugarcane, cassava, para rubber, and ponds—using high-resolution imagery in QGIS. These labeled images were then converted into grayscale segmentation masks suitable for deep learning input, using class-specific RGB-to-index mapping. To

address input limitations and computational efficiency, all images were split into 512×512 pixel patches. This size offers a balance between spatial context and memory constraints, making it suitable for deep learning architectures such as U-Net (Ronneberger et al., 2015) and SegFormer. The final dataset comprised 780 image-mask pairs, divided into 480 for training, 120 for validation, and 180 for testing.

2.4 Model configuration and training

The crop classification model in this study was developed using the SegFormer architecture (Figure 2) with the mit_b0 backbone. Hyperparameter tuning was conducted using Optuna (Akiba et al., 2019) to identify optimal training settings. Model training was conducted using the PyTorch framework with the AdamW optimizer with cross-entropy loss. A ReduceLROnPlateau scheduler was applied to automatically adjust the learning rate. Training configurations and hyperparameters are summarized in Table 2.

Table 2 Model training configuration and hyperparameters used in this research.

Parameter	Configuration
Number of epochs	700
Learning rate	0.0001
Weight decay	0.0001
Gamma	0.90
Step size	5
Batch size	6
Patience	6
Hardware	GPU (CUDA-enabled)
GPU used	NVIDIA RTX 3050

Model performance was evaluated on the validation set using pixel-wise accuracy and validation loss, which are commonly used metrics in semantic segmentation tasks (Singh et al., 2022).

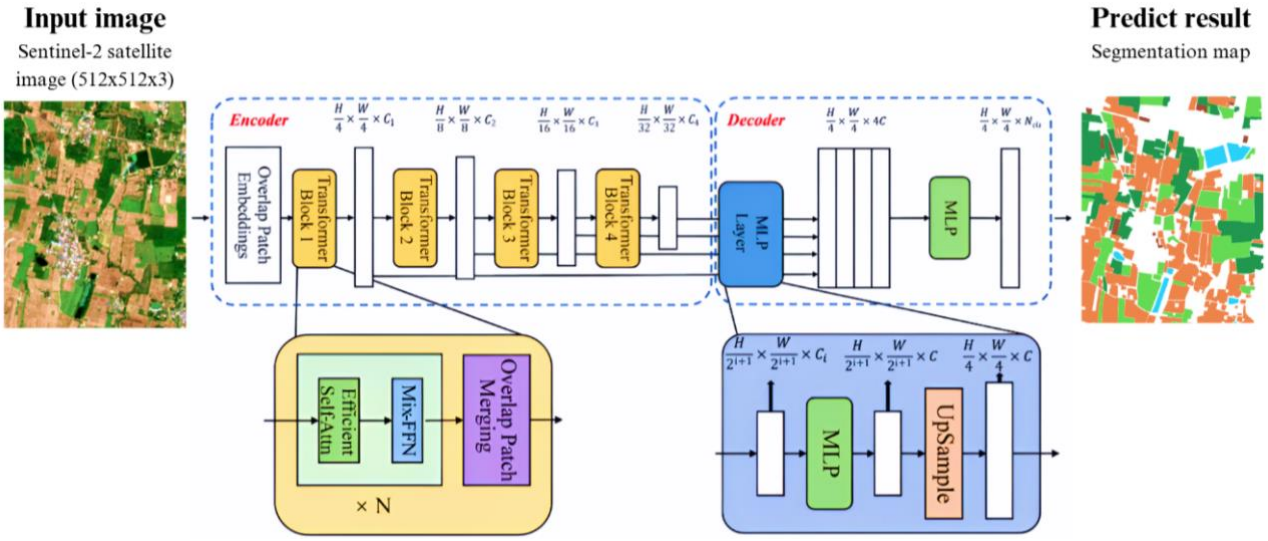


Figure 2 Overview of the SegFormer semantic segmentation model architecture, consisting of a transformer-based encoder and an All MLP-based decoder (adapted from Song et al. (2023)).

2.5 Inference and post-training process

After training, the SegFormer model was loaded from a saved checkpoint to retain its learned weights and biases. Test images 512×512 pixels were passed through the model to make a classification prediction.

To improve spatial consistency, predicted patches were reconstructed into full-sized images and refined using a majority color process based on the argmax function. This process reduces noise and ensures that a single dominant class coherently represents each polygon. The dominant class C^* for polygon P is defined by:

$$C^* = argmax \sum_{i \in P}^{c} 1(y_i = c)$$

where:

C^* is the class with the highest frequency in P, ensuring that the entire polygon is filled with this dominant class.

P represents the set of pixels within a polygon,

y_i is the predicted class for pixel i,

$1(y_i = c)$ is an indicator function that counts occurrences of class c, defined as:

$$1(\mathrm{y_i} = \mathrm{c}) = \begin{cases} 1, & \text{if } \mathrm{y_i} = \mathrm{c} \\ 0, & \text{otherwise} \end{cases}$$

3. Results

3.1 Performance of the SegFormer model

The training process was monitored over 700 epochs, and model checkpoints were saved based on validation loss improvement. As shown in Figure 3, the model exhibited a consistent decline in both training and validation loss, indicating successful convergence without signs of overfitting.

The best-performing model checkpoint achieved a pixel-wise accuracy of 0.967 on the validation set, demonstrating its ability to learn spatial–spectral patterns of various crop types in the region. This model was subsequently used for inference on a separate test set, which was used solely for qualitative evaluation. No quantitative metrics were reported, as the test set was used only for visual inference.

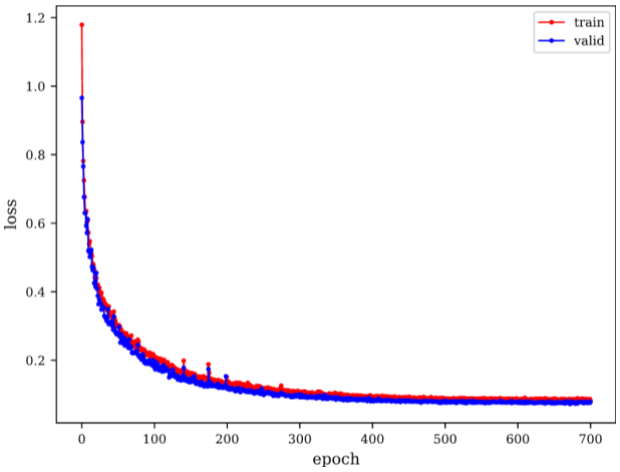


Figure 3 Training (red line) and validation (blue line) loss during 700 epochs of model training process.

3.2 Segmentation and model inference result

Figures 4 and 5 illustrate representative prediction masks and segmentation results from the training and validation sets. The four panels (a–d) presented in Figure 4 highlight representative patches selected for quality control (QC) and visualization of the model's behavior during the training process. These examples demonstrate how the model effectively distinguishes between crop types under learned conditions.

Figure 4, the predicted masks from the training and validation sets are closely aligned

with the ground truth images with the following highlights:

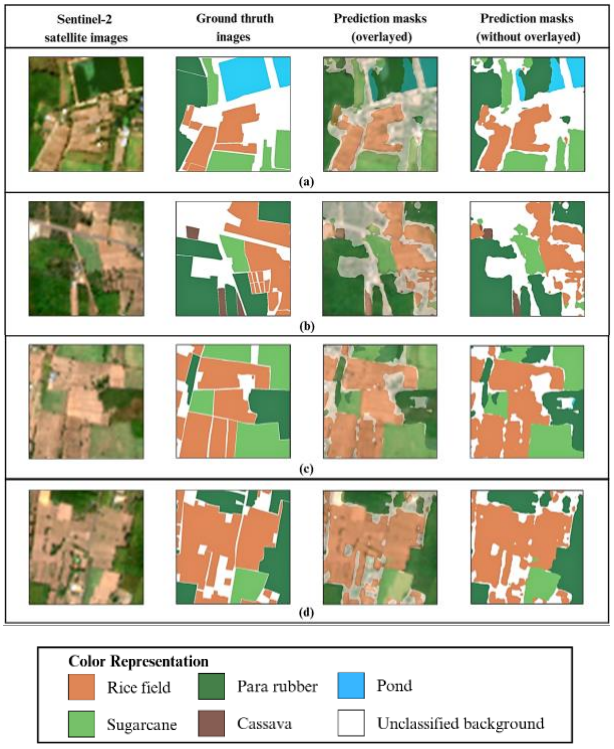


Figure 4 Inference comparison using the training and validation datasets. Panels a–d represent different patch locations. The colors highlight different classes: rice (orange), sugarcane (light green), para rubber trees (dark green), cassava (brown), and pond (blue).

- The pond class (panel a), especially when covered by aquatic vegetation, is frequently misclassified as para rubber due to the similarity in spectral reflectance between the two classes.
- The cassava class (panel b) is difficult to detect due to its occurrence in small, scattered plots and limited representation in the training data.
- The model generally distinguishes rice fields, sugarcane, and para rubber with high accuracy, with class boundaries closely aligning with the ground truth. However, confusion still occurs between sugarcane and para rubber (panel c) due to similar spectral features, suggesting potential areas for model refinement.

Figure 5 presents a full-scene segmentation map reconstructed from patch-level predictions and refined using the majority-vote process.

As shown in Figure 5, the model performs well in distinguishing crop types with stable spectral and structural characteristics, such as para rubber and sugarcane. Para rubber, with its dense canopy and consistent year-round reflectance, is the most accurately classified class. Sugarcane is also effectively segmented, though some misclassifications with para rubber occur in areas with similar leaf coloration or mixed vegetation. Harvested rice fields are reasonably well identified due to their distinct brown tones, which contrast with other land covers. Conversely, cassava remains one of the most challenging classes due to its scattered distribution and limited training data. Pond areas are also difficult to classify, particularly when small water bodies are covered by vegetation or have similar spectral features to tree canopies. These results show that while the model performs well on dominant crop types, it still struggles with classes that are either less represented in the training set or have spectral characteristics that closely resemble other land covers.

Despite the overall accuracy, some segmentation errors remain, particularly in areas with complex boundaries or spectrally similar classes like cassava and ponds. The black outlines in panel (c) highlight regions where predictions differ from the ground truth image, reflecting potential misclassifications. These

discrepancies may arise from class imbalance in the training data or subtle spectral similarities between certain crop types. These issues suggest directions for further model refinement.

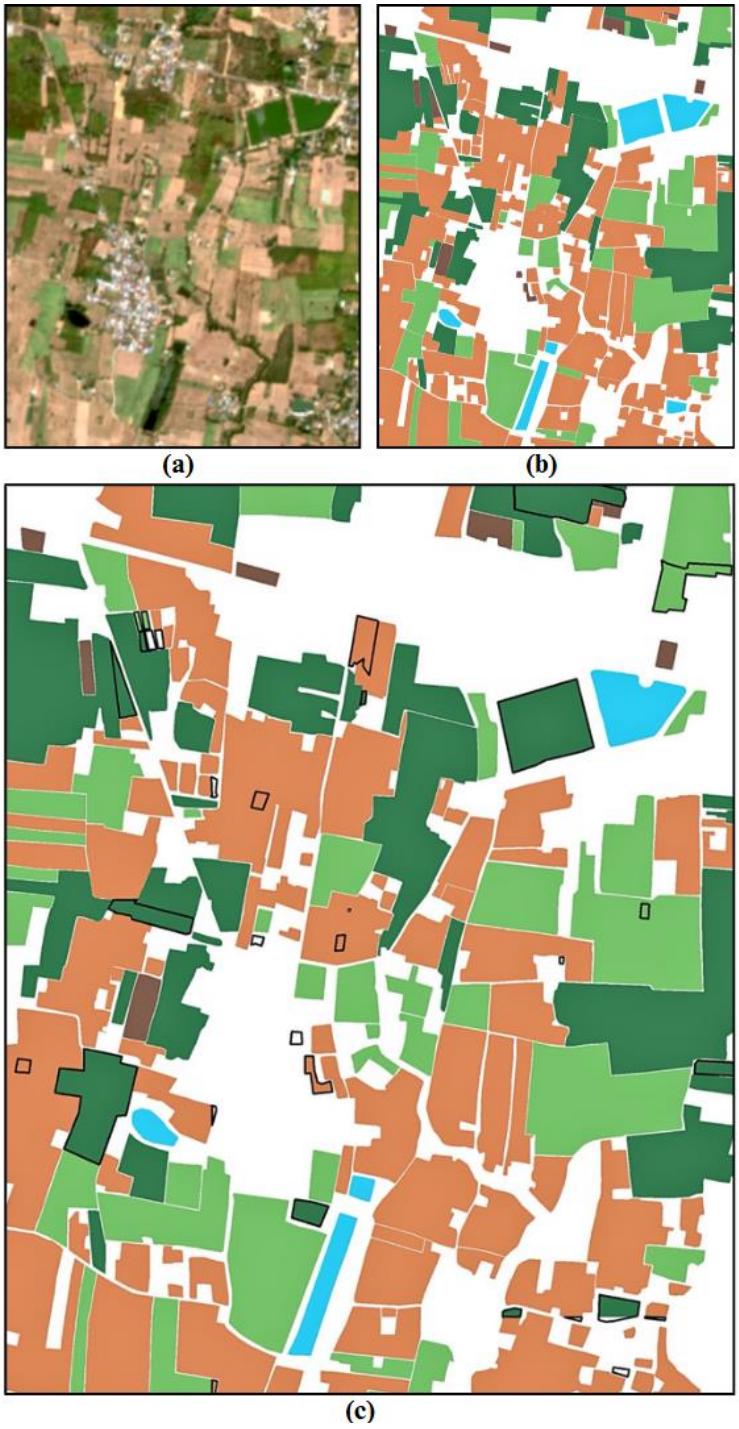


Figure 5 Comparison of the input Sentinel-2 satellite image (a), annotated segmentation (ground truth labeled) image (b), and final segmentation map (c) at a resolution of 2,088 x 2,600 pixels, derived from the training and validation datasets after reconstruction and majority color processing

Figure 6 shows the final segmentation output generated from the test dataset, which was not used during training or validation process.

Figure 6 Segmentation map generated from the test dataset after combining patch-level predictions and applying a majority color processing.

Compared to the results from the training and validation datasets shown in Figure 5, the final segmentation output from the test dataset (Figure 6) shows a noticeable decline in classification quality.

While the model still correctly identifies major crop types such as para rubber and sugarcane in several regions, misclassifications are more frequent in the test dataset, particularly for cassava, harvested rice fields, and pond areas. These classes tend to exhibit spectral features that overlap with other land cover types. In addition, the model showed limited ability to detect small agricultural plots. Even when such plots were detected, predictions were often inaccurate. Notably, fields smaller than approximately 5,200 square meters were more likely to be misclassified or missed, possibly because the 10-meter resolution of Sentinel-2 imagery was too coarse to capture their shapes clearly.

This result highlights the model's limited generalization capacity when applied to unseen areas, emphasizing the importance of dataset diversity and class balance in future model development.

4. Discussion

Although SegFormer was originally developed for general-purpose semantic segmentation in urban and structured scenes, its application in agricultural settings has begun to emerge, albeit with limited. This study demonstrates that, with appropriate adaptation, SegFormer can be effectively applied to crop classification in real-world agricultural settings.

Transformer-based deep learning models have recently gained momentum in remote sensing and semantic segmentation, with architectures such as SegFormer offering notable advantages in capturing global spatial features, improved boundary segmentation, and efficient processing with fewer parameters compared to traditional models (Xie et al. (2021); Li et al. (2023).

Nonetheless, recent studies suggest that Transformer-based models can outperform CNNs in agricultural applications when appropriately adapted. For instance, Gallo et al. (2024) reported that a Swin UNETR model achieved higher accuracy and faster training than traditional CNNs when applied to time-series data. Similarly, Zhao et al. (2018) showed that transformer-based models effectively captured both spatial and temporal complexities in crop classification tasks. In the application of SegFormer, Song et al. (2023) applied the model for crop classification in Bengbu, China—an area characterized by large-scale, irrigated farmlands with relatively continuous field structures. Their study relied on curated ground truth data derived from official land-use maps, field surveys, and NDVI-based phenological masks, which ensured high annotation quality. SegFormer outperformed CNN-based models and RF, achieving the highest overall accuracy and segmentation consistency especially when applied to well-structured agricultural landscapes with clearly defined cropping patterns and balanced training data. While Li et al. (2023) founds that U-Net achieved the highest overall accuracy for crop classification in their study over structured farmlands, they also acknowledged the strong performance of SegFormer. While its accuracy was slightly lower than U-Net, SegFormer demonstrated notable strengths in spatial clarity and boundary delineation, particularly when well-prepared ground truth data were used.

In contrast to previous studies that applied SegFormer model to structured farmlands using curated datasets, this research investigates more complex field conditions. These conditions are characterized by fragmented and heterogeneous plots with varying shapes and sizes—typical of many agricultural landscapes in Thailand—which increase the complexity of crop classification from satellite imagery.

The model's performance in this study, supported by visual assessments, high-quality training data, and spatial refinement techniques suggests that the SegFormer can be adapted to meet the challenges of remote sensing applications in real-world agricultural. To highlight the unique contributions of this research, the key findings are summarized below:

- Real-world applicability in unstructured farming landscapes: the model was tested under field conditions featuring fragmented plot patterns and heterogeneous cropping practices, providing insights into its robustness beyond idealized or curated datasets.
- Post-inference refinement using majority-vote process: improved spatial consistency in the final segmentation map and reduced misclassification noise at field boundaries.
- Integration of UAV-labeled training data: high-resolution drone imagery was used to generate precise polygon labels, improving the quality of training data and enhancing the model's ability to detect small, irregular field boundaries, particularly in plots larger than approximately 5,200 square meters.
- Qualitative insights into class-specific learning: the visual progression from training and validation to test datasets were analyzed to assess the model's learning behavior. These observations offer valuable insights for refining training data and enhancing model development in future research.

Collectively, these insights underscore the potential of SegFormer when applied with appropriate adaptation and highlight practical considerations for future development.

5. Conclusion and recommendation

The research successfully developed a modern methodology for crop classification using Sentinel-2 satellite imagery in combination with the SegFormer model, achieving a high pixel-wise accuracy of 0.967 on the validation set, demonstrating its effectiveness in learning spatial and spectral patterns from the training data.

However, segmentation performance varied across crop classes. Para rubber and sugarcane were classified with high accuracy due to their distinct and relatively stable spectral features. Harvest rice fields, while showing moderate accuracy, experienced some confusion with bare soil. Cassava posed the greatest challenge due to its limited representation in the dataset and scattered planting patterns, resulting in frequent misclassifications. These outcomes underscore the importance of balanced and diverse training data for robust model generalization.

This study also demonstrated the benefits of post-processing techniques such as majority-vote filtering, which improved spatial consistency and reduced noise along field boundaries. Additionally, the integration of high-resolution UAV-labeled training data enhanced the model's ability to detect small and irregular field plots, further contributing to overall performance. Based on visual

assessment, the model was able to correctly classify small fields starting from approximately 5,200 square meters. For plots smaller than this threshold, detection was inconsistent, and misclassifications were more likely. This limitation is likely due to the coarse spatial resolution of Sentinel-2 imagery (10 meters), which reduces the model's ability to distinguish fine-scale features in very small plots.

While the application of deep learning to satellite-based crop classification holds strong potential, challenges remain. Spectral similarity among certain classes (e.g., para rubber vs. pond areas) and the impact of seasonal variation require careful data preparation and model tuning. The findings of this study confirm that SegFormer can be effectively tailored for large-scale agricultural mapping, provided that preprocessing, label quality, and class balance are appropriately addressed. Future work may incorporate multi-temporal data to enhance classification accuracy.

References

Abdi, A. M. (2019). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, *57*(1), 1-20. https://doi.org/10.1080/15481603.2019.1650447

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna* Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. https://doi.org/10.1145/3292500.333070

Food and Agriculture Organization of the United Nation. (2015). *Climate change and food security: risks and responses*. Retrieved 10 December 2024 from https://openknowledge.fao.org/server/api/core/bitstreams/a4fd8ac5-4582-4a66-91b0-55abf642a400/content

Land Development Department. (2021). *Guidelines for appropriate agricultural promotion in Buriram Province: AGRI-MAP*. https://www.ldd.go.th/agri-map/Data/NE/brm.pdf

Li, G., Han, W., Dong, Y., Zhai, X., Huang, S., Ma, W., Cui, X., & Wang, Y. (2023). Multi-Year Crop Type Mapping Using Sentinel-2 Imagery and Deep Semantic Segmentation Algorithm in the Hetao Irrigation District in China. *Remote Sensing*, *15*(4). https://doi.org/10.3390/rs15040875

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *152*, 166-177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Meteorological Development Department. (2023). *Climate of Buriram Province*. http://climate.tmd.go.th/

National Statistical Office Thailand. (2021). *Agricultural Census 2023*. Retrieved 10 November 2023 from https://www.nso.go.th/nsoweb/main/summano/P7

Phumkokrux, N. (2021). Köppen-Geiger Climate System Classification and Forecasting in Thailand. *Folia Geographica*, *63*(2), 110.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18,

Singh, G., Singh, S., Sethi, G., & Sood, V. (2022). Deep Learning in the Mapping of Agricultural Land Use Using Sentinel-2 Satellite Data. *Geographies*, *2*(4), 691-700. https://doi.org/10.3390/geographies2040042

Song, W., Feng, A., Wang, G., Zhang, Q., Dai, W., Wei, X., Hu, Y., Amankwah, S. O. Y., Zhou, F., & Liu, Y. (2023). Bi-Objective Crop Mapping from Sentinel-2 Images Based on Multiple Deep Learning Networks. *Remote Sensing*, *15*(13). https://doi.org/10.3390/rs15133417

United Nations. (2017). *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100*. Retrieved 10 Dec 2023 from https://www.un.org/en/desa/world-

population-projected-reach-98-billion-2050-and-112-billion-2100

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, *34*, 12077-12090.

Zhao, W., Zhang, H., Yan, Y., Fu, Y., & Wang, H. (2018). A Semantic Segmentation Algorithm Using FCN with Combination of BSLIC. *Applied Sciences*, *8*(4). https://doi.org/10.3390/app8040500