Pertanika J. Sci. & Technol. 34 (3): 1933 - 1951 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Multi-task Deep Learning Pipeline for Rice Field Classification and Growth Monitoring Using Drone Imagery

Youssef Yasser Salaheldin Elhammamy[1], Chung Gwo Chin[2]*, Gan Ming Tao[3], Chan Kah Yoong[4], and Pang Wai Leong[5]

[1]Department of Information Technology, Nityo Infotech Services Sdn. Bhd., Q Sentral, No. 2A, Jalan Stesen Sentral 2, Kuala Lumpur Sentral, 50470 Kuala Lumpur, Malaysia
[2]Department of Electrical and Electronic Engineering, Centre of Excellence for Intelligent Networks, Faculty of Artificial Intelligence and Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia
[3]Department of Artificial Intelligence, Centre of Excellence for Artificial Intelligence, Faculty of Artificial Intelligence and Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia
[4]Department of Electrical and Electronic Engineering, Centre of Excellence for Robotics and Sensing Technologies, Faculty of Artificial Intelligence and Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia
[5]Department of Electrical and Electronic Engineering, School of Engineering, Faculty of Innovation and Technology, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia

ABSTRACT

A booming population around the world raises the concern of shortages of food resources in this new era. Thus, monitoring and managing crop production is extremely essential, especially rice crops, as they are the fundamental food source for most countries. Several challenges need to be addressed in this case, such as the classification of farmland from various land usages, precise monitoring of rice seedlings, and segmentation of rice growth. By leveraging advanced technologies such as drone imagery and machine learning, this paper proposed a new integrated pipeline for rice field classification and growth monitoring: a combination of convolutional neural networks (CNNs), You Only Look Once (YOLO), and modified U-Net models. These models were used in stages, specifically for paddy field classification, rice seedling detection, and rice growth segmentation. Substantial measurements and analysis have been carried out to verify the performance of the proposed system, including an accuracy of at least 85%, low classification/segmentation loss below 0.35, and high detection recall above 0.9. Thus, the findings highlight how combining different machine learning models with aerial photography can revolutionise conventional farming methods for better efficacy.

Keywords: Convolutional neural network, deep learning, drone, modified U-Net, rice field classification, YOLOv8

ARTICLE INFO

Article history:
Received: 08 June 2025
Accepted: 20 May 2026
Published: 25 June 2026

DOI: https://doi.org/10.47836/pjst.34.3.27

E-mail addresses:
yousefyassersalah@gmail.com (Youssef Yasser Salaheldin Elhammamy)
gcchung@mmu.edu.my (Chung Gwo Chin)
mtgan@mmu.edu.my (Gan Ming Tao)
kychan@mmu.edu.my (Chan Kah Yoong)
waileong.pang@taylors.edu.my (Pang Wai Leong)
* Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

With a rapidly changing climate and a fast-growing global population, the presence of food security and agricultural sustainability is highly demanded in this new era. Among other crops, rice stands out as a staple food, especially in developing nations like Malaysia. According to the statistics provided by the Food and Agriculture Organisation of the United Nations (FAOSTAT), rice is one of the main crops grown in the world (Food and Agriculture Organisation of the United Nations, n.d.) and is gradually increasing yearly, underscoring its importance and the need for better management practices. However, the task of monitoring and managing the rice crop presents major obstacles, especially considering the vast expanses of land dedicated to its cultivation, in addition to the reliance on huge manpower and time usage (Sinha & Dhanalakshmi, 2022). The first challenge is the categorisation of land use, particularly differentiating between regions with rice plantations and other uses. The complexity of this undertaking stems from the variations in spectral fingerprints resulting from various farming techniques and geographical features. Local knowledge, like crop management and phonological details, greatly influences the subtleties of these spectral signatures. The second challenge is the precise monitoring of rice seedlings. Although comprehensive imagery can be obtained from satellite and high-resolution optical data, its implementation is hindered by inaccuracies in the early phases of crop growth and the challenge of mapping crops across wide areas. The third challenge is segmenting rice crops' growth stages. Effective segmentation aids in tracking crop development and optimising agricultural interventions. However, limited motivations have been found in precise farming for this purpose.

To overcome these challenges, a modern automated system that does not require more manpower and setup costs is a better solution. Hence, the development of drone technology has opened new avenues for agricultural study and surveillance (Velusamy et al., 2021). The cost of installing drone equipment is much lower than that of satellite and surveillance cameras on the farmland, especially in rural areas. Crop pictures are also more adaptable and easier to capture than optical data. Drone imagery offers an airborne perspective on crop fields that are difficult to observe from the ground. This form of application is known as photogrammetry (Jiménez-Jiménez et al., 2021), and it appears to be a better way to obtain image data for precision agriculture. Photogrammetry has been demonstrated in research to be effective in monitoring agricultural fields over time with high spatial and temporal resolution. For instance, researchers in Costa Rica employed it to observe bean crop advancement and degradation with accuracies in the cm range (Arriola-Valverde et al., 2020). Another paper has used this technology too to create a 3D map of fruit crop canopies for precision input control (Sinha et al., 2022).

Despite the capability of drone imagery to deliver high-resolution crop images from multiple angles, it requires sophisticated image processing, including computer vision,

to accurately extract significant features from the crops due to the intricate structures of the geographical terrain and vegetation (Velusamy et al., 2021). Hence, machine learning has been proposed to mitigate this problem. A comprehensive review was presented in (Parashar et al., 2024) to analyse the effectiveness of various input parameters considered in crop yield prediction models. Convolutional neural networks (CNNs) are one of the simplest and most efficient machine learning models for performing landscape classification. For example, Pan et al. (2020) suggested a CNN-based multi-spectral Light Detection and Ranging (LiDAR) land-cover classification framework. This system incorporated pre-processing of multi-spectral 3D LiDAR data into 2D pictures, reducing the number of layers utilised in the CNN model while maintaining good performance. Another approach using CNN, as reported by (Kussul et al., 2017), achieved an accuracy of more than 85% for agricultural land-use and crop classification. In addition, Hsieh and Kiang (2020) developed different versions of CNNs to categorise hyperspectral images (HSIs) of crop and mixed vegetation agricultural fields. The CNN model achieved the maximum accuracy with only a one-dimensional (1D) CNN, demonstrating the efficacy of CNN in picture categorisation. Meanwhile, a CNN variant (AlexNet) was also presented in (Ajayi et al., 2024), and its performance outperformed conventional CNN for mixed-crop classification.

Aside from that, machine learning has demonstrated amazing success in detecting or identifying objects in pictures for precision agriculture. CNN and the You Only Look Once (YOLO) algorithm are the most widely used machine learning models for object recognition. In (Vishnoi et al., 2022), a CNN model was claimed to be well-suited to identify apple leaf diseases and achieved 98% classification accuracy. On the other hand, a survey was conducted in (Badgujar et al., 2024), demonstrating how YOLO provides real-time detection performance with high accuracy in a variety of agricultural jobs such as monitoring, surveillance, sensing, automation, and robotics operations. Meanwhile, a unique weed detection method using YOLOv5 was described in (Alrowais et al., 2022) by allowing Internet of Things (IoT) devices to record farm photographs and transfer them to a cloud server for model training. Furthermore, YOLOv4 with Mish activation was implemented for rice seedling detection, counting, and location labelling, as documented in (Yeh et al., 2024), which was effective for small-object detection in UAV imagery. More variants of YOLOs (YOLOv8n, YOLOv9t, and YOLOv10n) were presented in (Chen et al., 2025) to analyse the performance of robust seedlings' detection in paddy fields.

In addition, more advanced machine learning models, such as the U-Net architecture, can address the difficulty of crop segmentation using pictures. U-Net is a deep learning model that is widely used in medical image segmentation because it can extract very exact characteristics from an image using several convolutional layers (Siddique et al., 2021).

Hence, it has also been used in agriculture for various purposes, such as plant disease detection and growth stage identification. An improved U-Net (MU-Net) was constructed for plant diseased leaf image segmentation (Zhang & Zhang, 2023). Another paper has also applied the U-Net for the segmentation of wheat farmland using drone images (Liu et al., 2022). The study showed that the modified U-Net achieved significantly better segmentation results compared to classical models.

Finally, the Internet of Things (IoT) is another latest technology integrated with machine learning for sustainable precision agriculture. As reported by Ramu et al. (2026), a lightweight CNN, long short-term memory (LSTM) model, along with whale optimisation algorithm (WOA), was designed to categorise the health status of crops, optimise irrigation and spreading of fertilisers, and enhance sustainability performance based on spatiotemporal field information. Another approach proposed a machine learning-based fertiliser recommendation methodology according to the real-time soil fertility context captured through IoT-assisted soil fertility mapping to improve the accuracy of the fertiliser recommendation system (Khan et al., 2022a). This study has been extended to include context-aware monthly evapotranspiration (ET) estimations for saline soil reclamation with the effective leaching process using LSTM models (Khan et al., 2022b) and to determine the raw reference evapotranspiration (Eto) with only temperature and optional optimisation of the raw ETo according to the availability using an artificial neural network (ANN) model (Bashir et al., 2023).

Therefore, machine learning has become a key technique in the fast-developing field of precision farming, especially when integrated with drone imagery. Table 1 summarises the benefits of machine learning above conventional approaches (without machine learning) from earlier research. This paper aims to analyse the performance of machine learning-based rice field classification and growth monitoring using drone-captured imagery. The proposed system integrated a new deep learning-based pipeline into three stages of rice field analysis: paddy field classification, rice seedling detection, and rice growth segmentation. In the first stage, a CNN model is applied to distinguish between rice plantation land and other usages of land. In the second stage, a YOLOv8 model is used to detect and identify individual rice seedlings after the paddy field has been recognised. In the last stage, a modified U-Net model is utilised to determine various growth stages of rice crops. Finally, the performance of these models is then evaluated extensively to measure the accuracy, consistency, and reliability of the system. With the contribution of this research, rice crop monitoring can be gradually transformed into automated management, resulting in better decision-making and sustainable farming methods.

Table 1
Comparison of machine learning and traditional methods

Aspects	Machine Learning Techniques	Traditional Methods
Accuracy and Precision	Higher accuracy and precision in classifying elements, ability to learn complex data patterns.	Lower accuracy and precision may miss complex patterns.
Complex Visual Data	Excellent at processing complex visual data, including variability in lighting and crop density.	Struggle with complex visual data variability.
Automation and Scalability	Can automate tasks and process large datasets quickly once trained, scalable for extensive agricultural areas.	Manual or less efficient automated processes are not as scalable.
Adaptability	Models can be updated with new data to refine predictions and adapt to changes, enhancing accuracy and relevance.	May require significant adjustments to handle new data types or improve accuracy.
Multi-task Learning	Simultaneous multi-task analysis (classification, segmentation, detection), providing comprehensive agricultural insights.	Often single-task focused, requiring separate analyses for different tasks.
Reduction in Manual Labour and Subjectivity	Reduces reliance on manual labour and minimises subjectivity, leading to more objective outcomes.	Rely more on manual surveying and interpretation, subject to human error and bias.

METHODOLOGY

Figure 1 describes the system flow of the proposed methodology. It began with the collection of drone images of rice crops. The selected paddy field was at Sekinchan, Malaysia, due to its well-known and widespread rice plantation. There were nearly 1000 images collected online freely (Roboflow) or captured using the DJI Mavic Air 2 drone during the seedling period in February 2024 and the transplanting period in March 2024. These images have a maximum size of 1280 × 720 pixels. The altitude of the drone was maintained at around 10m above the ground to obtain roughly a ground solution distance (GSD) of 5 cm/pixel. The images were captured at different spots on different days for 2 months (the actual date varied due to the need for clear weather). However, only 250 images (around 125 from online and 125 from the drone) were selected at several view angles for the rice crops' detection after pre-processing of unwanted and low-quality images. The distribution of images obtained from drones and online sources was comparable under imaging conditions, such as resolution, lighting, and type of paddies. Although the online images did not have the altitude specification, image filtering was eventually done to match the size of the paddy with the drone imagery. The selected images were then resized to a resolution of 416 × 416 pixels, which is important for the machine learning models to train the dataset efficiently and quickly for all stages. Figure 2 shows some samples of the image dataset. To further reduce noise, open-source software named LabelMe was used to perform image filtering and processing, such as image masking and augmentation. By using augmentation approaches, the machine learning models are more resistant to overfitting and are trained on a variety of data circumstances. In this case, alterations such as rotation, flipping, scaling, brightness

adjustment, noise injection, etc., generated modified versions of the original images while preserving their class label.

After that, the drone images were split into a 60:20:20 ratio (training data, validation data, and test data). The data size was limited due to time limitations and environmental factors, such as cultivation events that have already begun in April 2024 in the study field. However, data augmentation was applied to enrich the dataset and generate slightly altered versions of images for each epoch during the training process. The proposed method divided the rice field analysis into three continuous stages: paddy field classification, rice seedling detection, and rice growth segmentation. Different machine learning models were proposed at different stages based on their capability and suitability. At the first stage, a CNN model was used. All 250 images were used in this case (200 paddy fields and 50 non-paddy fields). The classified paddy land was then trained with the YOLOv8 model for rice seedling detection. The last stage was to determine the rice growth using a modified U-Net model. For the last 2 stages, only 200 images with paddy fields were used. All models utilised the same dataset for the purpose of fair performance comparison. They were trained using a computer with an Intel Core i7 processor and 16 GB RAM, and Python 3.10. A performance analysis was then conducted to determine the effectiveness of these models on rice crop detection, which includes training and validation loss, distributional focal loss, mean average precision (mAP), and recall. The next part goes into detail about the proposed steps, which are based on three different machine learning models in this study.

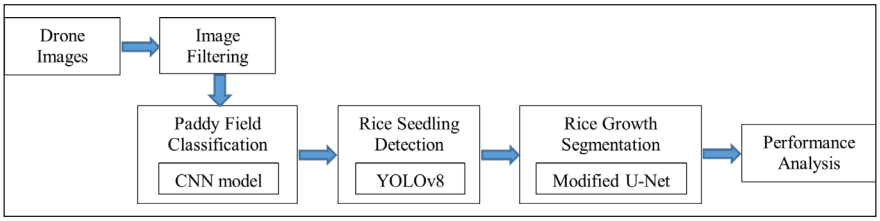

Figure 1. Block diagram

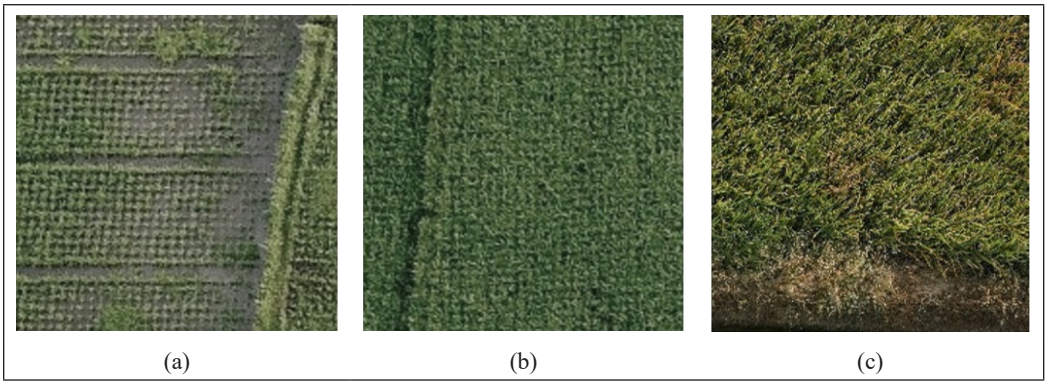

Figure 2. Samples of the image dataset obtained from (a) a drone; (b) online; and (c) online

Rice Field Classification

The first stage is to distinguish rice crops from other uses of land. The CNN model was used as the training model for this classification because of its accuracy and simplicity (Milosevic, 2020). Binary classification was implemented, which consisted of rice field and non-rice field land. Figure 3 shows the architecture of the CNN used to create the classification model in this paper.

The architecture started with an input layer, which supplied raw image data (dimensions of 416 × 416 pixels) into the network. It comprised five convolutional layers with 64, 128, 256, 512, and 512 filters, respectively, each followed by ReLU activation and max-pooling operations. These layers extracted numerous features from pictures. The feature extraction was then followed by two fully connected layers containing 4096 neurons each and a final output layer with two neurons for binary classification. It was expressly decided to employ a limited number of epochs of less than 10 and a dropout rate of 0.5 after each fully connected layer during the training of the model with the intention of avoiding overfitting, a common problem where a model learns too much from the training set, including noise and outliers. Adam was chosen as the optimiser due to its adaptive learning rate, which adjusted automatically during training to achieve faster convergence. To balance training stability and convergence speed, the learning rate was set to 0.001, the batch size was selected as 8, and the early stopping criterion was set to 5 epochs, allowing the model's accuracy to progressively rise over time.

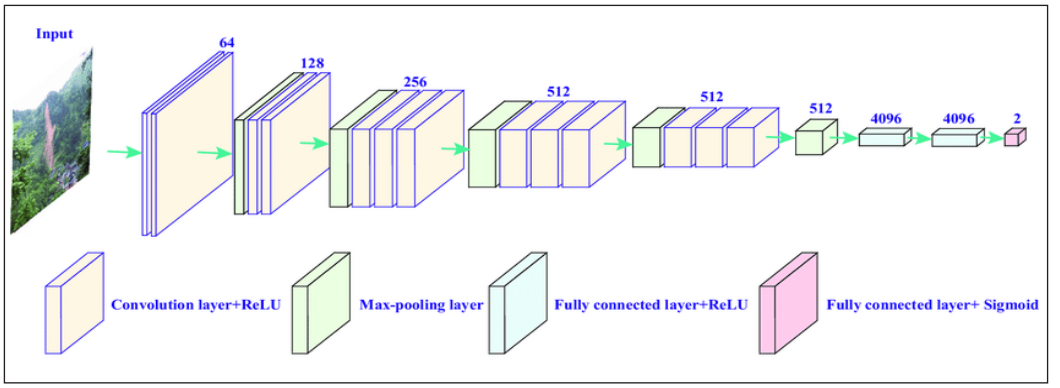

Figure 3. CNN architecture of the classification model

Rice Seedling Detection

Individual seedlings were found and identified in the paddy field during the second stage. Object detection stands as a crucial component in the realm of precision farming, enabling the automated analysis of vast crop fields with speed and accuracy.

As a result, a well-known object recognition deep learning model, YOLOv8, was chosen because it can identify objects in pictures with a variety of backdrops. The architecture of YOLOv8 is very complicated (Jiang et al., 2022) but generally can be described as a series of convolutional layers with batch normalisation and Leaky ReLU activation functions. The real object detection happened in the detection layers at the end of the architecture, in which the bounding boxes, object classes, and confidence ratings for each detected object in the image were predicted using the processed feature maps. The proposed model employs a sophisticated ensemble of layers and approaches designed expressly to tackle the hard problem of detecting objects inside a context, balancing computing efficiency and accuracy in detection. Finally, the dataset for this stage contained 4,500 annotated rice seedlings distributed across 250 drone images, with an average of 21 seedlings per image. Seedling sizes ranged from approximately 15 × 15 to 65 × 65 pixels. Most seedlings were clearly visible, although moderate crowding and partial occlusion were observed in densely planted regions.

Rice Growth Segmentation

A more complex machine learning model is required to identify each developing stage of the rice plant, from pre-booting to maturity in the third stage. As a result, the deep learning model modified U-Net was selected for this task. The U-Net model's architecture is also a convolutional network, with a contracting path to record context and a symmetric expanding path to allow for precise picture localisation. The modified U-Net used to create the segmentation model in this paper was acquired from Siddique et al. (2021). The architecture started with an input layer that received drone photos, which were typically in RGB format. It then went through a few convolutional layers to extract features at different resolutions. These convolutional layers used filters of various sizes to capture a variety of features, with the number of filters growing as the network progressed deeper, as shown by the expanding breadth of the layers Siddique et al. (2021). Activation functions such as ReLU were used after each convolutional process to add non-linearity to the model. To better meet the requirements of the agricultural imagery being analysed, parameter adjustment was made, as reported later in the result section, to the modified U-Net architecture, guaranteeing that the model can segment the images with high precision and effectiveness.

RESULTS AND DISCUSSION

Paddy Field Classification

Figure 4 presents the trajectory of loss values for the training and validation sets across the epochs during the model's learning process. It is a metric used to measure how well the model fits the training data. The training loss is traced by the blue line,

commencing at a higher value and indicating the model's initial inaccuracy in predicting the training data. As the epochs progress, this line shows a steep descent, reflecting a rapid improvement in the model's performance on the training data. Simultaneously, the validation loss, represented by the orange line, begins its course slightly above the training loss. The proximity of the two lines suggests that the model is achieving a similar level of performance on both the training and unseen validation data, which is a desirable outcome indicating good generalisation. As the epochs advance, both lines gradually level off and approach a state of minimal change, indicating that the model has reached a stable state. This plateau signals that the model has learned the dominant patterns in the data to a satisfactory degree and that additional training is unlikely to result in meaningful improvements.

Besides that, Figure 5 shows the binary precision-recall curve of the image classification model for the training set. The graph is designed to evaluate the trade-off between precision, the accuracy of positive predictions, and recall, the ability to identify all the positive instances. The area under the curve (AUC) is annotated as being 1.00, which is the maximum possible value, indicating a perfect precision-recall balance. The curve itself hugs the top right corner of the plot, reflecting an exceptional scenario where both precision and recall metrics are at their optimal level of 1.0 across all thresholds. This means that the model succeeds in identifying all positive instances (recall of 1.0) without any false positives (precision of 1.0), a rare and ideal outcome in practice. To avoid chances of overfitting, data augmentation has been applied during the training process. This graph verifies that the model is exceptionally well-tuned for the binary classification task at hand.

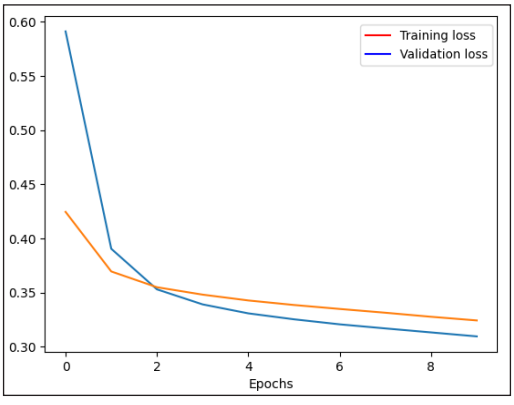

Figure 4. Training and validation loss for the classification model

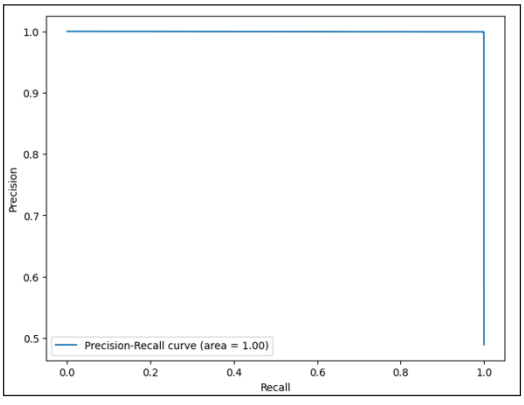

Figure 5. Binary precision-recall curve for classification model

Rice Seedling Detection

Figure 6 depicts the training and validation box loss for the YOLOv8 model used in individual rice seedling detection in this paper. The training box loss, or the difference between the predicted bounding box locations and the ground truth during the training phase, is displayed on the graph on the left. There is a noticeable decreasing trend, which suggests that as training went on, the model's accuracy in predicting the location of rice seedlings improved dramatically. On the other hand, the validation box loss, which shows how well the model's predictions on unseen data match the actual locations, is shown on the graph on the right. The overall trend is downward, despite some noticeable swings, which could be attributable to the inherent instability in the validation data. The decrease in loss for both training and validation highlights the model's learning efficacy, while the ongoing low loss on validation data indicates the model's robustness and potential for real-world applications.

In addition, Figure 7 shows the training and validation classification loss for the YOLOv8 used to detect individual rice seedlings. The classification loss is the model's error in correctly classifying the objects within the bounding boxes during training. The training classification loss (on the left) decreases significantly, as the model rapidly reduces its error rate in the initial epoch. This suggests that the model's predictions on the training data are becoming more accurate with time. The validation classification loss (on the right) has a general decrease in loss but with more pronounced spikes. However, the downward trends in both training and validation classification losses show that the model is successfully learning to classify the rice seedlings within the drone images. The sporadic spikes in the validation loss highlight how crucial it is to use a representative and diverse training dataset to guarantee the model's resilience in an actual agricultural environment.

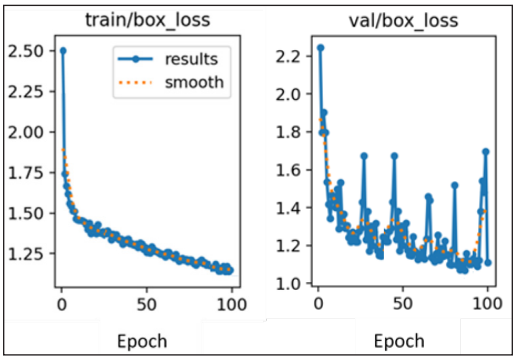

Figure 6. Training and validation box loss

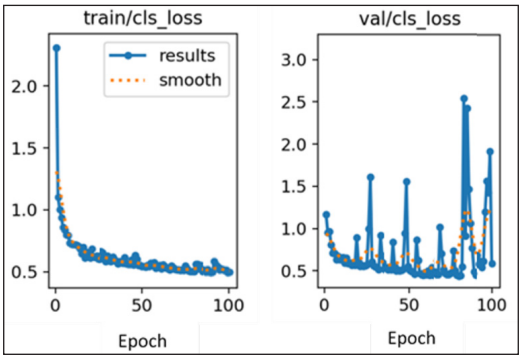

Figure 7. Training and validation classification loss

Besides that, Figure 8 represents the training and validation losses using the distributional focal loss (dfl_loss) function in a machine learning model. The dfl_loss is a part of the standard training framework in YOLOv8. It is particularly relevant when dealing with imbalanced datasets, as it helps to focus the model's learning on harder-to-classify examples, aiming to reduce the imbalance by adjusting the focus towards difficult or misclassified cases within the training process. In the training loss graph, we observe a sharp decline in loss initially, with the model quickly learning and improving its predictions on the training data. The loss values stabilise close to 1.0 as the epochs increase, indicating that the model has effectively learned the primary features and patterns necessary for the task at hand. The validation loss graph shows more variability, which is typical when the model is exposed to previously unseen data. Despite the fluctuations, there is a general downward trend in loss values, suggesting that the model can generalise the patterns it has learned to new data. However, the spikes in validation loss also highlight instances where the model struggles with certain validation examples, which could be areas where the model's predictions are less certain.

In addition, Figure 9 presents two key performance metrics for the YOLOv8 model: mAP at 50% IoU threshold (mAP50) on the left and mAP across IoU thresholds from 50% to 95% (mAP50-95) on the right. The mAP50 metric shown on the left indicates the model's precision at a specific intersection over union (IoU) threshold, a common benchmark for object detection models, where a 50% overlap with the ground truth is considered a correct detection. The model's mAP50 is predominantly high, staying close to 1.0 through most of the training epochs. This tells us that the model is highly accurate when a moderate overlap with the true object is considered sufficient. The mAP50-95 on the right averages the precision across a range of stricter IoU thresholds, from 50% up to 95%. This is a more rigorous evaluation of model performance, as it requires the model's predicted bounding boxes to closely match the ground truth. While the values here are naturally lower due to the increased difficulty, they still show the model performing well, with most values above 0.5. In both metrics, we observe occasional dips, which suggest moments where the model's performance on the validation set fluctuated, possibly due to more challenging data points or anomalies within the dataset.

Other than that, Figure 10 captures the recall metric for the YOLOv8 model across training epochs. Recall is an important measure in object detection scenarios, reflecting the model's ability to identify all relevant objects within the image. The graph indicates that the model's recall remains consistently high, predominantly above 0.9, which means the model is correctly identifying over 90% of the actual rice seedlings present in the images. This level of recall is impressive and suggests that the model is highly sensitive to the presence of seedlings, with few being missed.

However, rare severe drops in recall imply that the model failed to recognise some seedlings in specific epochs. Such changes can be attributable to variations in the complexity of the images in the validation set or slight quirks in the learning process at a few locations. The model's ability to sustain high memory is critical for practical applications, ensuring that fewer seedlings are overlooked, which could be especially useful for complete field study and yield estimation.

Finally, Figure 11 shows an example of detecting individual rice seedlings in a paddy area. Every seedling detected by the model is encompassed in a red bounding box, labelled "RiceSeedling" and accompanied by a confidence score indicating the model's level of conviction. The confidence scores fluctuate, showing that the model is more confident in certain seedlings than others. These scores are most likely dependent on the features that the model learns during training, such as shape, colour, and texture. The model's ability to detect and categorise items in drone images as rice seedlings is demonstrated graphically, and this ability may be useful for agricultural activities such as counting rice seedlings to assess planting success or production estimates.

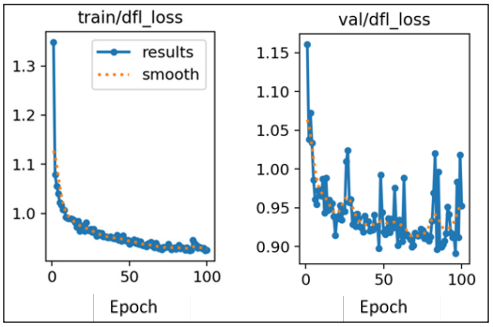

Figure 8. Training and validation of distributional focal loss

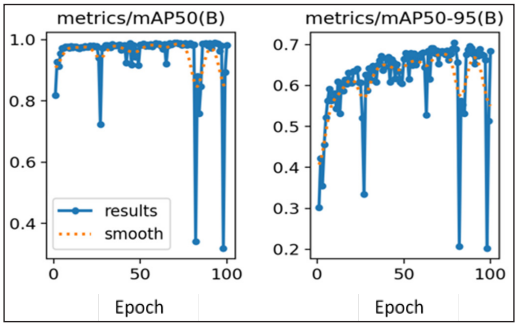

Figure 9. mAP at different IoU thresholds for seedling detection

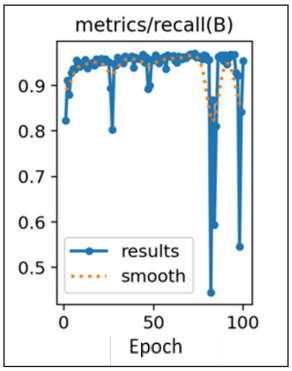

Figure 10. Recall for seedling detection

Figure 11. Example of rice seedling detection

Rice Growth Segmentation

In Figure 12, the loss values for the segmentation model during training are plotted across 10 epochs, applying an Adam optimiser with a learning rate of 0.001. It is observed that both training loss and validation loss decrease in the initial epochs, with the training loss showing a steeper decline than validation loss before gradually plateauing. This convergence of training and validation loss is indicative of a model that is learning effectively without overfitting significantly, as overfitting would typically be represented by a decrease in training loss but an increase in validation loss as epochs progress. The relatively stable and low values of validation loss in later epochs suggest that the model generalises well to new data. However, both training and validation final loss values are still over 0.25, which is considerably high and leads to a lower accuracy. To improve the loss performance, a lower learning rate of 0.0001 is employed. The training and validation losses are plotted over an extended period of 100 epochs as shown in Figure 13. From observation, the loss for both training and validation swiftly declines and levels out, with the training loss slightly lower than the validation loss. The extended training period has allowed the model to thoroughly learn from the data without fitting too closely to the noise within the training set. The reduction in the learning rate and the increased epoch count have evidently contributed to a model that performs well and is stable over time, as seen in the minimal gap between the training and validation lines and lower loss values.

Nevertheless, Figure 14 illustrates a sample output from the second segmentation model, showcasing its ability to predict the growth stages of rice crops from drone imagery. The leftmost image is the input, a high-resolution field image. The centre image is the true mask, which denotes the actual growth stages of the rice crop as identified by experts. The experts consisted of five paddy farmers in Sekinchan, Malaysia. They were assigned to select different stages of rice's growth among all the tested samples. Only those samples with 100 per cent matching would serve as true masks. The stages are colour-coded from 0 to 4, where '0' represents other non-crop elements, and stages '1' to '4' correspond to the increasing growth levels, with '4' being the harvest stage. The rightmost image is the model's predicted mask. It closely mirrors the true mask, indicating that the model can accurately identify and segment the various growth stages. The colours in the predicted mask align well with the true mask, with only minor discrepancies around the edges, reflecting the model's precision and the effectiveness of its learning from the model.

As compared to the state-of-the-art methods mentioned in the previous studies, our proposed methodology can incorporate the paddy field classification, rice seedling detection, and rice growth segmentation in a single system with considerably low prediction loss, high mAP, and high consistency in performance. Table 2 summarises the accuracy of each machine learning model used in this paper.

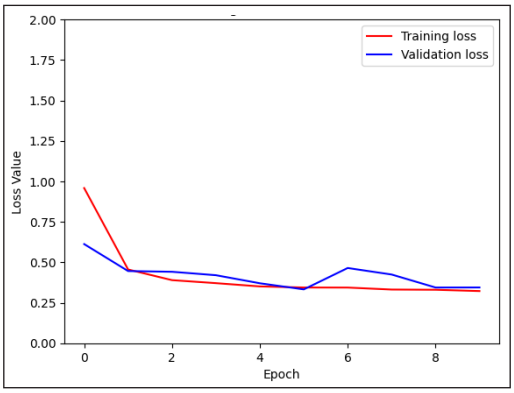

Figure 12. Training and validation loss for the first segmentation model

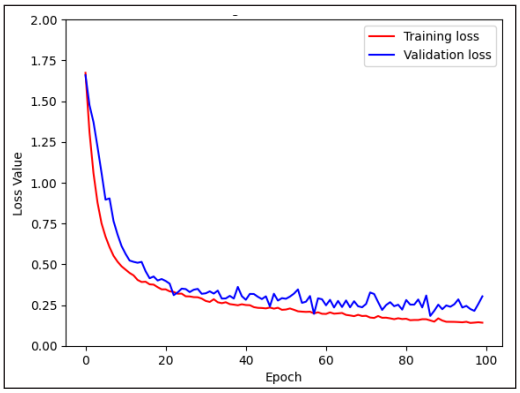

Figure 13. Training and validation loss for the second segmentation model

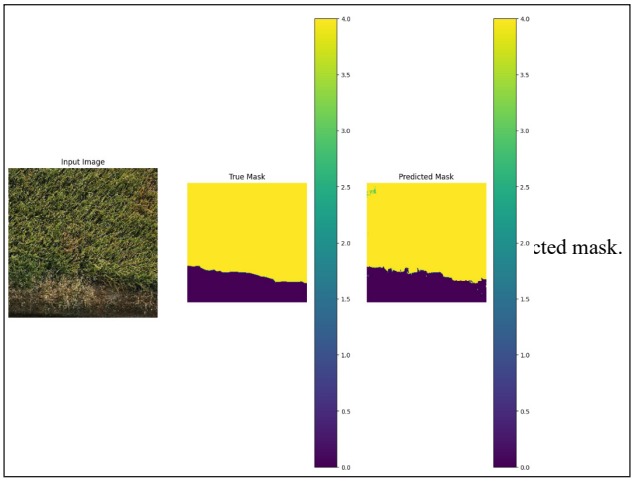

Figure 14. Sample prediction image with true mask and predicted mask

Table 2
Accuracy of the models

Stages	Accuracy
Paddy Field Classification	≈ 99%
Rice Seedling Detection	> 85%
Rice Growth Segmentation	> 90%

Discussion of Findings

Table 3 highlights that existing studies predominantly concentrate on individual tasks, including crop categorisation, seedling detection, or field segmentation. Although these methods demonstrated robust performance, their usefulness is typically confined to a particular phase of agricultural monitoring. The proposed system enhances prior research by integrating CNN-based land-use categorisation, YOLOv8-based seedling identification, and a modified U-Net for growth-stage segmentation into an interconnected UAV-based pipeline. It achieved competitive performance, with accuracies ranging from 85% to 99%, recall values above 0.90, and low classification and segmentation losses. While additional validation on bigger and more diverse datasets is necessary, the findings obtained suggest that the pr method attains comparable performance while offering a more thorough solution for rice field monitoring and precision agricultural applications.

Table 3
Benchmarking of previous and current studies

Studies	Technique Used	Application	Performance Reported	Advantages	Limitations
(Kussul et al., 2017)	CNN	Agricultural land-use and crop classification	Accuracy > 85%	Effective automatic feature learning without manual feature engineering	Require large, labelled datasets for robust performance
(Ajavi et al., 2024)	CNN (AlexNet)	Mixed-crop classification from high-resolution UAV imagery	Accuracy ≈ 72%	Outperformed conventional CNN across all tested effective automatic feature extraction from UAV images epochs;	Validation accuracy remained substantially lower than training accuracy, indicating overfitting; performance decreased after 60 epochs
(Yeh et al., 2024)	YOLOv4 with Mish activation	Rice seedling detection, counting, and location labelling from UAV imagery	F1-score = 0.91; average precision improved by 4.96% compared with the previous YOLOv4	Effective for small-object detection in UAV imagery; integrates detection, counting, and coordinate labelling in a single framework	Limited dataset (44 original UAV images); evaluation focused on a competition dataset rather than multi-location field validation

Table 3 (*continued*)

Studies	Technique Used	Application	Performance Reported	Advantages	Limitations
(Chen et al., 2025)	YOLOv8n, YOLOv9t, and YOLOv10n	Paddy seedling counting and detection from UAV imagery	mAP@50 values of 0.964, 0.936, and 0.944 for YOLOv8n, YOLOv9t, and YOLOv10n	High mAP, robust detection of seedlings; good counting reliability	Limited dataset (100-200 UAV images); performance degrades under severe occlusion and dense planting
(Liu et al., 2022)	Improved U-Net (MGA + Interaction Mechanism)	Wheat land segmentation from UAV imagery	Dice = 93.88%, better than U-Net & DeepLabV3	High segmentation accuracy; improved semantic gap	Require more computation resources; sensitive to data quality
Proposed method	CNN + YOLOv8 + Modified U-Net	Land-use classification, seedling detection, and growth-stage segmentation	Accuracy ranging from 85% to 99%, classification/ segmentation loss < 0.35, and recall > 0.9	Integrated framework addressing three rice-monitoring tasks within a single workflow	Limited dataset size and need for broader external validation

CONCLUSION

This paper successfully analyses the performance of several machine learning models used for paddy field classification, rice seedling detection, and growth stage segmentation using drone-captured images. The functionalities provided by CNN, YOLOv8, and modified U-Net models illustrate very promising performance. Due to 99% accuracy of field classification, the seedling detection was able to achieve at least 85% accuracy. Hence, the accuracy of growth segmentation has been improved to over 90%. However, there are practical challenges and limitations on real-world implementation of the proposed system, including dependency on high-quality drone imagery, environmental influences such as lighting and weather conditions, and the computational cost involved in training machine learning models. Future research might delve into fine-tuning parameters such as dfl_loss or explore its application in conjunction with other techniques to further improve model robustness and generalisation, especially in diverse agricultural settings. Nevertheless, the limited dataset size may affect the robustness of the model, and more evaluation can be done on larger and more diverse external datasets to further validate the model's generalisability. In addition, these models' flexibility would be tested, and their scalability might be revealed by including more real-time images from different regions

and expanding the dataset to cover a range of climates, crop diseases, and growth phases. By doing so, these models can be fully utilised and be instrumental in precision farming, helping farmers to make informed decisions about irrigation, fertilisation, and harvesting times.

ACKNOWLEDGEMENT

This research was not supported by any funding. We thank our colleagues from the Faculty of AI & Engineering who provided insight and expertise that greatly assisted the research, although they may not agree with all the interpretations/conclusions of this paper.

REFERENCES

Ajayi, O. G., Iwendi, E., & Adetunji, O. O. (2024). Optimising crop classification in precision agriculture using AlexNet and high-resolution UAV imagery. *Technology in Agronomy, 4*, Article e011. https://doi.org/10.48130/tia-0024-0009

Alrowais, F., Asiri, M. M., Alabdan, R., Marzouk, R., Hilal, A. M., Alkhayyat, A., & Gupta, D. (2022). Hybrid leader-based optimisation with deep learning-driven weed detection on Internet of Things-enabled smart agriculture environment. *Computers and Electrical Engineering, 104*, Article 108411. https://doi.org/10.1016/j.compeleceng.2022.108411

Arriola-Valverde, S., Villagra-Mendoza, K., Méndez-Morales, M., Solórzano-Quintana, M., Gómez-Calderón, N., & Rimolo-Donadio, R. (2020). Analysis of crop dynamics through close-range UAS photogrammetry. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-5). IEEE. https://doi.org/10.1109/ISCAS45731.2020.9181285

Badgujar, C. M., Poulose, A., & Gan, H. (2024). Agricultural object detection with You Only Look Once (YOLO) algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture, 223*, Article 109090. https://doi.org/10.1016/j.compag.2024.109090

Bashir, R. N., Khan, F. A., Khan, A. A., Tausif, M., Abbas, M. Z., Shahid, M. M. A., & Khan, N. (2023). Intelligent optimisation of reference evapotranspiration (ETo) for precision irrigation. *Journal of Computational Science, 69*, Article 102025. https://doi.org/10.1016/j.jocs.2023.102025

Chen, S., Li, W., Chen, D., Xie, Z., Zhang, S., Cen, F., Tu, L., Zhao, Q., & Gao, Z. (2025). Recognition of rice seedlings count in UAV remote sensing images via the YOLO algorithm. *Smart Agricultural Technology, 12*, Article 101107. https://doi.org/10.1016/j.atech.2025.101107

Food and Agriculture Organisation of the United Nations. (n.d.). *Rice*. Retrieved July 1, 2024, from https://www.fao.org/markets-and-trade/commodities/rice/en/

Hsieh, T.-H., & Kiang, J.-F. (2020). Comparison of CNN algorithms on hyperspectral image classification in agricultural lands. *Sensors, 20*(6), Article 1734. https://doi.org/10.3390/s20061734

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A review of YOLO algorithm developments. *Procedia Computer Science, 199*, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135

Jiménez-Jiménez, S. I., Ojeda-Bustamante, W., Marcial-Pablo, M., & Enciso, J. (2021). Digital terrain models generated with low-cost UAV photogrammetry: Methodology and accuracy. *ISPRS International Journal of Geo-Information, 10*(5), Article 285. https://doi.org/10.3390/ijgi10050285

Khan, A. A., Faheem, M., Bashir, R. N., Wechtaisong, C., & Abbas, M. Z. (2022a). Internet of Things (IoT)-assisted context-aware fertiliser recommendation. *IEEE Access, 10*, 129505-129519. https://doi.org/10.1109/ACCESS.2022.3228160

Khan, A. A., Nauman, M. A., Bashir, R. N., Jahangir, R., Alroobaea, R., Binmahfoudh, A., Alsafyani, M., & Wechtaisong, C. (2022b). Context-aware evapotranspiration (ETs) for saline soils reclamation. *IEEE Access, 10*, 110050-110063. https://doi.org/10.1109/ACCESS.2022.3206009

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters, 14*(5), 778-782. https://doi.org/10.1109/LGRS.2017.2681128

Liu, G., Bai, L., Zhao, M., Zang, H., & Zheng, G. (2022). Segmentation of wheat farmland with an improved U-Net on drone images. *Journal of Applied Remote Sensing, 16*(3), Article 034511. https://doi.org/10.1117/1.JRS.16.034511

Milosevic, N. (2020). *Introduction to convolutional neural networks: With image classification using PyTorch* [Video course]. Apress. https://doi.org/10.1007/978-1-4842-5648-0

Pan, S., Guan, H., Chen, Y., Yu, Y., Nunes Gonçalves, W., Marcato Junior, J., & Li, J. (2020). Land-cover classification of multispectral LiDAR data using CNN with optimised hyperparameters. *ISPRS Journal of Photogrammetry and Remote Sensing, 166*, 241-254. https://doi.org/10.1016/j.isprsjprs.2020.05.022

Parashar, N., Johri, P., Khan, A., Gaur, N., & Kadry, S. (2024). An integrated analysis of yield prediction models: A comprehensive review of advancements and challenges. *Computers, Materials & Continua, 80*(1), 389-431. https://doi.org/10.32604/cmc.2024.050240

Ramu, S. C., Raman, D., Ramana, K., Krishna, A. V., Khan, A. A., Khan, S., & Abdu, S. M. (2026). A lightweight deep learning and whale optimisation framework for sustainable precision agriculture. *Discover Computing, 29*, Article 73. https://doi.org/10.1007/s10791-026-09952-8

Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access, 9*, 82031-82057. https://doi.org/10.1109/ACCESS.2021.3086020

Sinha, B. B., & Dhanalakshmi, R. (2022). Recent advancements and challenges of Internet of Things in smart agriculture: A survey. *Future Generation Computer Systems, 126*, 169-184. https://doi.org/10.1016/j.future.2021.08.006

Sinha, R., Quirós, J. J., Sankaran, S., & Khot, L. R. (2022). High-resolution aerial photogrammetry-based 3D mapping of fruit crop canopies for precision inputs management. *Information Processing in Agriculture, 9*(1), 11-23. https://doi.org/10.1016/j.inpa.2021.01.006

Velusamy, P., Rajendran, S., Mahendran, R. K., Naseer, S., Shafiq, M., & Choi, J.-G. (2022). Unmanned aerial vehicles (UAVs) in precision agriculture: Applications and challenges. *Energies, 15*(1), Article 217. https://doi.org/10.3390/en15010217

Vishnoi, V. K., Kumar, K., Kumar, B., Mohan, S., & Khan, A. A. (2023). Detection of apple plant diseases using leaf images through a convolutional neural network. *IEEE Access, 11*, 6594-6609. https://doi.org/10.1109/ACCESS.2022.3232917

Yeh, J. F., Lin, K. M., Yuan, L. C., & Hsu, J. M. (2024). Automatic counting and location labelling of rice seedlings from unmanned aerial vehicle images. *Electronics, 13*(2), Article 273. https://doi.org/10.3390/electronics13020273

Zhang, S., & Zhang, C. (2023). Modified U-Net for plant diseased leaf image segmentation. *Computers and Electronics in Agriculture, 204*, Article 107511. https://doi.org/10.1016/j.compag.2022.107511