Pertanika J. Sci. & Technol. 34 (3): 1525 - 1549 (2026)

SCIENCE & TECHNOLOGY

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

D-Xotic: A Deep Learning Improvement for Recognising Different Durian Cultivars Based on Leaves Signature

Aulia Brilliantina[1,2], Tri Agus Siswoyo[1], Yuli Witono[3], and Bayu Taruna Widjaja Putra[3,4]*

[1]Faculty of Agriculture, University of Jember, Jl. Kalimantan No. 37, Tegalboto Campus, Jember, 68121 East Java, Indonesia
[2]Food Industry Technology Study Program, Department of Agricultural Technology, Jember State Polytechnic, Jl. Mastrip, PO Box 164, Jember, 68121 East Java, Indonesia
[3]Faculty of Agricultural Technology, University of Jember, Jl. Kalimantan No. 37, Tegalboto Campus, Jember, 68121 East Java, Indonesia
[4]Centre of Excellence in Artificial Intelligence for Industrial Agriculture, University of Jember, Jl. Kalimantan No. 37, Tegalboto Campus, Jember, 68121 East Java, Indonesia

ABSTRACT

Early identification of durian cultivars at the seedling stage remains challenging because conventional methods rely on fruit characteristics that appear only after several years, often leading to misdistribution and inefficiencies in certification. This study proposes a real-time, non-destructive deep learning framework for intra-species durian cultivar identification using leaf images. A dataset of 1,788 leaf images from five cultivars (Bawor, Kani, Monthong, Musang King, and Petruk) was collected under real-world conditions. Unlike existing studies that formulate plant varietal identification solely as an image classification task, this work introduces a detection-based formulation that explicitly models cultivar-specific leaf features at the object level. This formulation enables the model to localise and learn fine-grained morphological differences that are often overlooked in global classification approaches. To validate this approach, MobileNetV2, Xception, and YOLOv11 were systematically evaluated using K-Fold Cross Validation under the same experimental setting. The key contributions of this study are: (1) the formulation of intra-species plant cultivar identification as an object detection problem rather than a pure classification task, (2) the development of a practical, real-time framework for early-stage durian cultivar identification using leaf images in unconstrained environments, and (3) a comparative analysis demonstrating the effectiveness of detection-

ARTICLE INFO

Article history:
Received: 20 August 2025
Accepted: 03 April 2026
Published: 19 June 2026

DOI: https://doi.org/10.47836/pjst.34.3.11

E-mail addresses:
aulia_b@polije.ac.id (Aulia Brilliantina)
triagus.faperta@unej.ac.id (Tri Agus Siswoyo)
yuliwitono.ftp@unej.ac.id (Yuli Witono)
bayu@unej.ac.id (Bayu Taruna Widjaja Putra)
* Corresponding author

e-ISSN: 2231-8526

based models in capturing subtle inter-varietal differences. Experimental results show that YOLOv11 achieves over 99% validation accuracy, with precision and recall above 0.95, while maintaining an inference time of 0.2 ms per image. These findings highlight the potential of detection-based approaches for AI-assisted seedling selection in horticulture.

Keywords: Cultivar identification, deep learning, durian, leaf image, YOLO

INTRODUCTION

Durian (*Durio zibethinus*) is one of the most economically important fruit crops in Southeast Asia, serving not only as a cultural and culinary icon but also as a high-value commodity in both domestic and international markets. The rapid growth of global demand, particularly from China, has significantly increased durian production in major producing countries such as Thailand, Malaysia, and Indonesia (May et al., 2023). The commercial value of durian is strongly determined by varietal characteristics, including flavour, aroma, texture, and postharvest quality. Therefore, accurate identification of durian cultivars is essential for farmers, nurseries, certification bodies, and other stakeholders to ensure product authenticity and supply chain traceability.

In agricultural systems, varietal identification plays a critical role in farm management, quality assurance, and plant breeders' rights. Misidentification can result in inappropriate agronomic practices, economic losses, and market disputes. Moreover, cultivar-specific traits influence nutrient management, disease resistance, and harvesting time, making accurate identification essential for long-term planning in sustainable durian production (Magandhi et al., 2024).

Conventional and molecular approaches for durian varietal identification have been widely applied. Traditional methods rely primarily on morphological traits, particularly fruit characteristics, which only become observable after several years. Although leaf-based identification during the vegetative stage is possible, it is highly subjective and sensitive to environmental variations. Molecular techniques such as DNA analysis offer high accuracy but are costly, time-consuming, and require specialised laboratory infrastructure, limiting their applicability for routine field use. These limitations highlight a fundamental challenge in early-stage durian varietal identification. As a result, there is currently no rapid, non-destructive, and field-deployable method capable of accurately identifying durian cultivars at the seedling stage under real-world conditions.

Recent advances in Artificial Intelligence (AI), particularly deep learning and computer vision, have significantly improved plant identification using image-based approaches. Convolutional Neural Networks (CNNs) have demonstrated superior performance compared to traditional machine learning methods such as Support Vector Machines (SVMs), Random Forests (RFs), and K-Nearest Neighbours (KNN), which rely on manually engineered features sensitive to variations in lighting, background, and perspective (Kamilaris & Prenafeta-Boldu, 2018; Wäldchen et al., 2018). CNN-based models can automatically learn hierarchical visual features from raw images, enabling more robust recognition of morphological differences such as leaf shape and venation patterns (LeCun et al., 2015).

Advanced CNN architectures and real-time object detection frameworks have further expanded the applicability of image-based plant analysis. Architectures such as Xception and MobileNetV2 improve classification accuracy while maintaining computational efficiency, making them suitable for real-world applications (Chollet, 2017; Sandler et al., 2018). These models have been successfully applied in crop identification tasks under field conditions (Too et al., 2019; Ullah et al., 2023). In parallel, object detection frameworks such as YOLO (You Only Look Once) enable simultaneous localisation and classification of multiple objects within a single image, which is particularly advantageous in complex agricultural environments. Recent developments in YOLO architectures, including YOLOv11, have improved detection accuracy and efficiency through optimised backbone structures and loss functions (Ahmad et al., 2023; Bochkovskiy et al., 2020). However, most existing studies predominantly formulate plant identification as a single-image classification problem under controlled conditions, where each image contains a single object with minimal background variation. This limitation reduces their applicability in real-world agricultural environments, where multiple leaves may appear simultaneously with occlusion, varying illumination, and complex backgrounds. Recent studies on fruit and leaf-based deep learning, including papaya leaf disease classification, fruit bruise detection, pear leaf disease detection, and MobileNetV2-based fruit classification, further demonstrate the effectiveness of deep learning models in agricultural applications.

Despite these advancements, most existing studies focus on inter-species classification or disease detection, with limited attention to fine-grained intra-species varietal identification under real-world conditions. Several works have applied deep learning to leaf disease detection in fruit crops or fruit-level quality assessment, often using controlled datasets and experimental settings. While these studies demonstrate the effectiveness of deep learning, they provide limited insights into fine-grained intra-species classification, particularly for perennial fruit crops such as durian during the vegetative stage. Furthermore, many prior approaches formulate plant identification solely as an image classification problem, without exploring detection-based strategies that may better capture subtle visual differences under real-world conditions (Mohanty et al., 2016; Singh et al., 2016).

In practice, varietal management in Indonesia and other durian-producing countries remains largely dependent on manual and subjective identification, increasing the risk of mislabelling during seedling distribution and certification processes (Horiuchi et al., 2023). Existing image-based approaches are often developed under controlled conditions, limiting their robustness in real-world environments characterised by variable lighting, complex backgrounds, and occlusions. Consequently, their applicability in practical durian cultivation systems remains constrained.

Therefore, the core problem addressed in this study is the absence of a robust, accurate, and field-applicable method for early-stage durian cultivar identification under real-world conditions. To address this challenge, this study proposes a deep learning-based framework that integrates lightweight CNN classification models (MobileNetV2 and Xception) with a real-time object detection model (YOLOv11) for intra-species durian cultivar identification using leaf images. The main contributions of this study are threefold: (1) the development of a real-world durian leaf image dataset collected under field conditions; (2) the introduction of a detection-based formulation for fine-grained intra-species cultivar identification, moving beyond conventional classification-based approaches; and (3) a systematic comparison between classification-based and detection-based deep learning methods, demonstrating that detection models can more effectively capture subtle inter-varietal differences for early-stage identification.

MATERIALS AND METHODS

Study Location

Jember Regency, East Java, Indonesia, is a major centre for durian cultivation and development. The region has favourable agroclimatic conditions, including fertile soil, annual rainfall ranging from 1,800-3,000 mm, and temperatures between 24-32 °C, which support the development of durian cultivation and enable the growth of both local cultivars (Bawor and Petruk) and commercial cultivars (Musang King, Monthong, and Kani). This study selected Jember Regency as the study area because it represents an active durian seed production and certification region, providing a realistic context for evaluating a leaf image–based cultivar identification methodology.

Materials and Workflow Overview

This section presents the methodological workflow adopted in this study, encompassing image acquisition, expert-assisted labelling, data preprocessing, model training, and evaluation, as shown in Figure 1. Leaf images were collected under natural lighting conditions, while manual labelling was verified by agricultural experts to ensure annotation reliability. The dataset consisted of leaf images from five durian cultivars (Monthong, Kani, Petruk, Bawor, and Musang King) collected in Jember Regency.

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), an Intel-based CPU (10 cores, 3.7 GHz), 64 GB RAM, and running Ubuntu 20.04 LTS. Model training was performed using Python 3.9 with TensorFlow/Keras and PyTorch frameworks. For inference and deployment-level

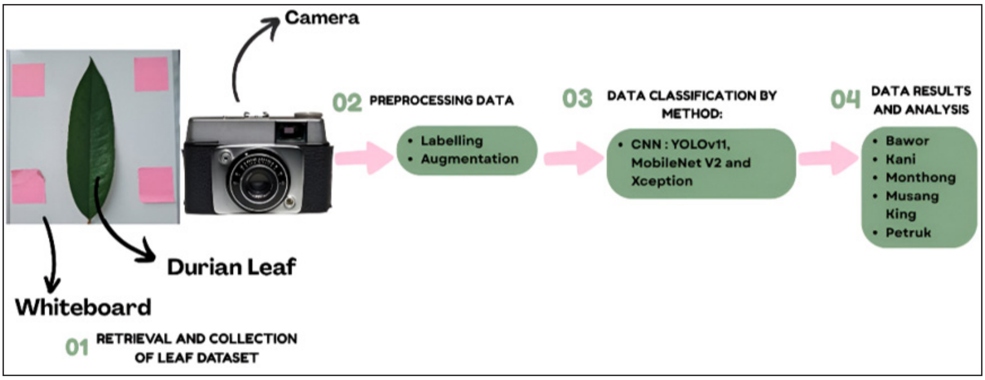

Figure 1. Research stages

evaluation, models were tested under real-time conditions using the Ultralytics YOLOv11 inference pipeline, reflecting practical field deployment scenarios.

Data Collection Procedures

The data collection process began with the identification and validation of five durian cultivars (Bawor, Kani, Monthong, Musang King, and Petruk) conducted in collaboration with local farmers and agricultural experts, as shown in Figure 2. This validation step was included as a methodological control to reduce labelling uncertainty in intra-species classification tasks.

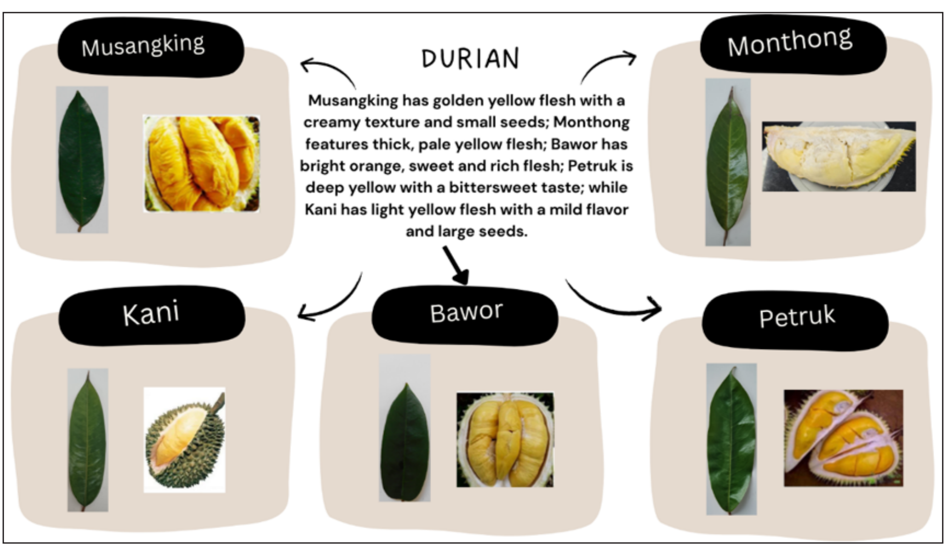

Figure 2. Durian leaf cultivars in Jember Regency (Musang King, Monthong, Kani, Bawor, and Petruk)

A total of 1,788 leaf images were captured using a Nikon Coolpix B500 digital camera (16 MP) at distances ranging from 15 to 50 cm. To minimise background interference and simplify preprocessing, leaves were positioned against a white background with pink markers. All images were collected under natural lighting conditions to reflect practical field acquisition scenarios, rather than laboratory-controlled environments. The dataset distribution across cultivars consisted of Bawor (293), Kani (600), Monthong (443), Musang King (163), and Petruk (289) images.

Data Augmentation Techniques

Data augmentation was employed to enhance model generalisation and simulate real-world variability, as commonly recommended in image-based deep learning studies (Meethongjan et al., 2022; Sengupta et al., 2023). Before augmentation, all images were standardised in terms of resolution and colour space, followed by contrast normalisation to mitigate acquisition-related bias.

Augmentation operations, including rotation, flipping, brightness modification, scaling, cropping, translation, and Gaussian blur, were applied exclusively to the training dataset, with transformation probabilities ranging from 0.3 to 0.5 (Zhang et al., 2024). This strategy was adopted to expose the model to diverse visual conditions rather than to artificially inflate the dataset size.

The dataset was divided into training (80%), validation (10%), and testing (10%) subsets (Table 1). To address class imbalance, the number of augmented samples per class was adjusted to obtain a balanced training set of approximately 600 images per durian cultivar. The validation and testing subsets consisted solely of original (non-augmented) images to prevent data leakage and ensure an unbiased performance evaluation.

Justification for Object Detection-based Approach

While durian varietal identification can be formulated as an image classification problem, practical field scenarios often involve multiple leaves appearing simultaneously within a single frame, partial occlusion, and complex backgrounds.

Table 1
Dataset statistics and class distribution before and after augmentation

Durian Cultivar	Original Images	Augmented Images (Training Only)	Total Training	Validation	Testing
Bawor	293	366	600	29	30
Kani	600	120	600	60	60
Monthong	443	246	600	44	45
Musang King	163	470	600	16	17
Petruk	289	369	600	29	29
Total	1,788	1,571	3,000	178	181

Object detection models, such as YOLOv11, offer the ability to localise and classify multiple leaf instances in real time, which is not achievable using pure image-level classification approaches. This capability is particularly relevant for deployment in plantation-scale monitoring, mobile-based scouting, and automated nursery inspection systems. Therefore, object detection was investigated not as a replacement for classification, but as a scalable and deployment-oriented alternative suitable for real-world agricultural environments.

Deep Learning Architectures

YOLOv11: Object Detection Framework

YOLOv11 is a single-stage object detection framework that formulates object detection as a unified regression problem, simultaneously predicting bounding box coordinates and class probabilities (Redmon et al., 2016). In this study, YOLOv11 was employed primarily to assess its suitability as an object detection module for identifying durian leaves within complex image scenes, rather than to emphasise detection performance outcomes shown in Figure 3.

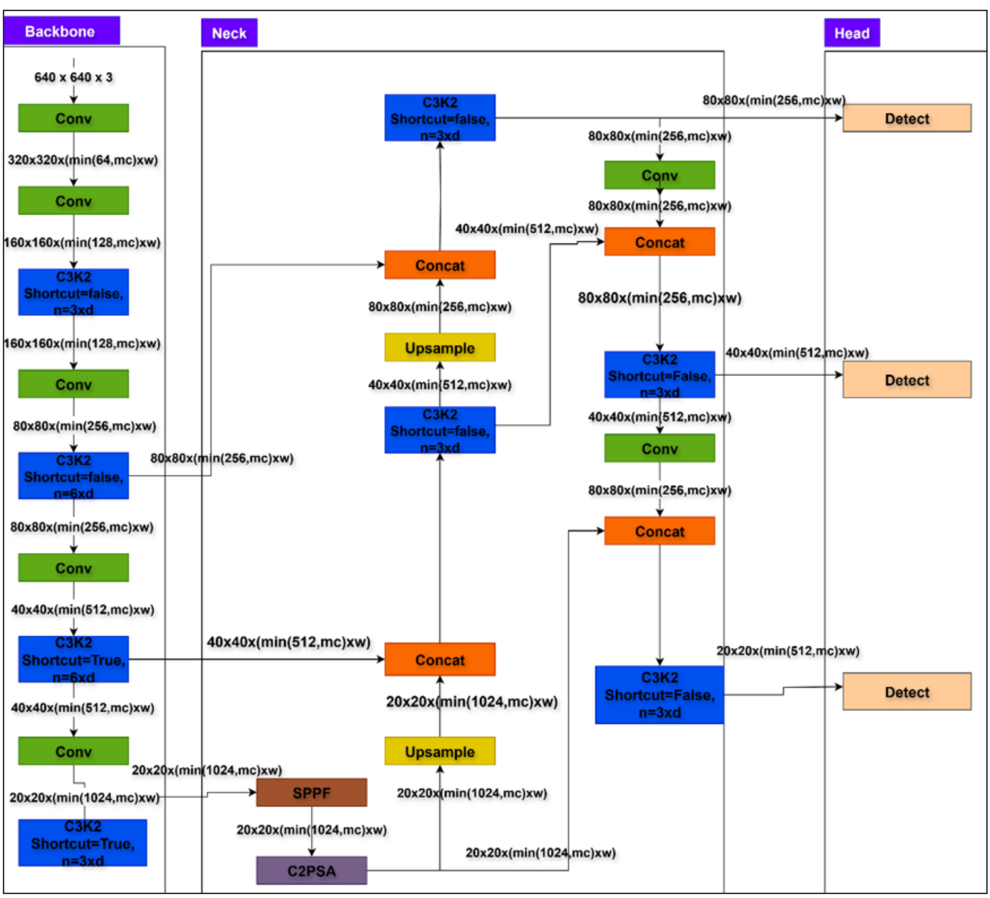

Figure 3. YOLOv11 architecture

This study does not propose architectural modifications but instead focuses on evaluating the applicability of an existing detection framework in real-world agricultural scenarios.

YOLOv11 incorporates a convolutional neural network (CNN) backbone composed of Conv–BatchNorm–SiLU blocks, as defined in the original YOLOv11 architecture, which have been reported to improve training stability and convergence behaviour (Maaroof & Bouhlel, 2024). Spatial pyramid pooling is included as part of the standard YOLOv11 design to preserve spatial information in lower-resolution feature maps, as documented in prior object detection studies. No architectural modifications were introduced to the YOLOv11 backbone in this study.

The model was implemented using the Ultralytics YOLOv11 framework. Training parameters were set to an initial learning rate of 0.001, a batch size of 32, and the Stochastic Gradient Descent (SGD) optimiser (Bottou, 2012). These parameters were kept constant across all experiments to ensure methodological consistency rather than performance optimisation.

MobileNetV2: Lightweight CNN for Resource-constrained Applications

MobileNetV2 is a lightweight convolutional neural network architecture that employs inverted residual blocks and linear bottlenecks to reduce computational complexity while preserving representational capacity (Sandler et al., 2018). In this study, MobileNetV2 was selected to evaluate the feasibility of deploying durian varietal classification models on resource-constrained platforms, such as mobile or edge devices, shown in Figure 4. This architecture has been widely adopted in agricultural image classification tasks, including plant species recognition and fruit quality assessment, due to its favourable trade-off between classification accuracy and computational efficiency (Kazi, 2024; Kusumaningtyas et al., 2024).

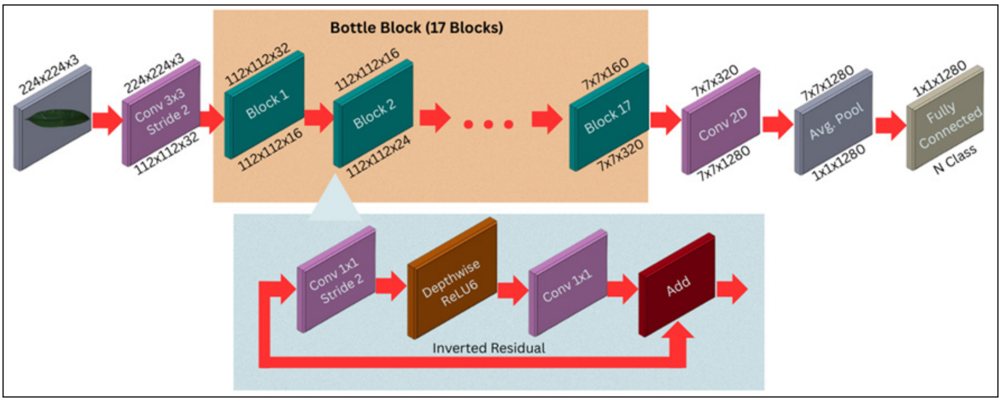

Figure 4. MobileNet V2 architecture

Xception: Deep Feature Extraction Model

Xception is a deep convolutional neural network architecture that replaces standard convolution operations with depthwise separable convolutions, enabling independent learning of spatial and channel-wise features and thereby improving computational efficiency and representational power (Chollet, 2017). In this study, Xception was included as a high-capacity feature extraction model to serve as a comparative benchmark for capturing fine-grained morphological differences among durian leaf cultivars. This architecture has been extensively applied across diverse image analysis domains, including medical imaging, demonstrating its effectiveness in learning complex visual patterns shown in Figure 5 (Maesaroh et al., 2024; Tejada et al., 2024; Zeynali et al., 2024).

Both MobileNetV2 and Xception models were initialised with ImageNet pre-trained weights. Transfer learning was applied by freezing the early feature extraction layers and fine-tuning the task-specific classification layers, following established best practices for small to medium-sized datasets (Yosinski et al., 2014).

Experimental Setup

All leaf images were resized to 640 × 640 pixels to ensure consistent input dimensions across both classification and detection models. Model training was conducted on a workstation equipped with an NVIDIA RTX 3090 GPU, using TensorFlow/Keras and PyTorch frameworks, with training progress monitored via TensorBoard.

Training parameters included an initial learning rate of 0.001, a batch size of 32, and a maximum of 100 epochs. Early stopping was employed as an experimental control strategy, whereby training was terminated if the validation loss failed to improve for 10 consecutive epochs, to mitigate overfitting and enhance reproducibility. Under selected experimental conditions, leaf segmentation was applied as a preprocessing step to reduce background interference, thereby enabling the models to focus on salient leaf morphological features.

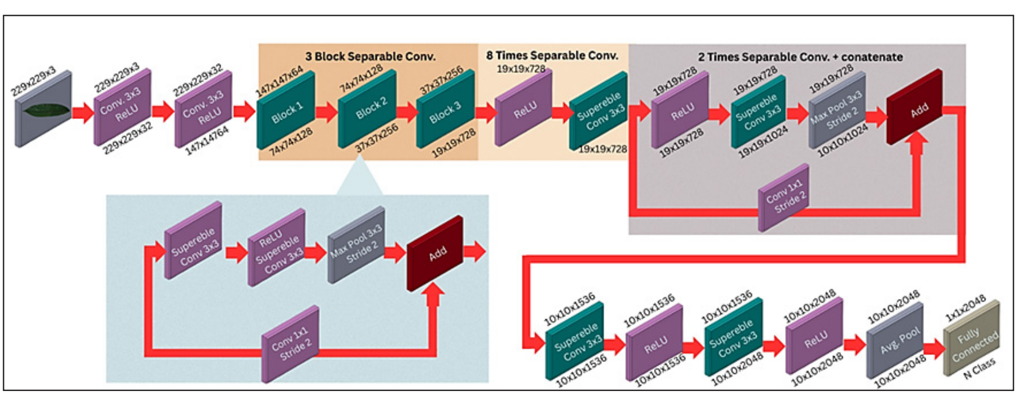

Figure 5. Xception architecture

Evaluation Metrics and Validation Strategy

Model performance was evaluated using a confusion matrix-based approach, from which accuracy, precision, recall, and F1-score were computed (Equations 1 to 4). These metrics were selected to provide a balanced assessment of classification behaviour rather than to emphasise peak performance values.

$$\text{Accuracy} = \frac{TP + TN}{TP+TN+FP+FN} \quad [1]$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad [2]$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad [3]$$

$$\text{F1 – Score} = 2 \times \frac{Precision\ \ x\ Recall}{Precision\ \ +Recall} \quad [4]$$

Description:

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

To reduce partitioning bias and improve generalisation reliability, K-Fold Cross Validation (K-Fold CV) was employed, which has been widely adopted in CNN-based classification studies with limited datasets (Gorriz et al., 2024; Hasan et al., 2025; Uzer, 2025). The average accuracy formulation of K-Fold Cross Validation was used to compute the final classification accuracy, Equation 5. For object detection tasks, mean Average Precision (mAP) was used to summarise the precision-recall relationship across all durian cultivars, following standard object detection evaluation practices. The average accuracy formulation of K-Fold Cross Validation:

$$\text{Accuracy\_Kfold} = 1/\text{K} + \sum_{n=1}^{K} Accuracy \quad [5]$$

Performance of YOLO-based detectors is typically reported using mean Average Precision (mAP), which summarises the precision-recall trade-off across all classes. Let C be the number of classes. mAP is the mean of the per-class Average Precision (AP), as expressed in Equation 6:

$$\text{mAP} = 1/\text{N} \sum_{i=1}^{N} AccuracyAPi \quad [6]$$

where N is the number of object classes, and APi is the area under the precision-recall curve for class i.

RESULTS AND DISCUSSION

Model Performance Comparison

MobileNetV2 Performance: Accuracy and Training Characteristics

MobileNetV2 is designed to achieve high classification efficiency while maintaining low computational complexity. In this study, MobileNetV2 was employed to classify durian leaf cultivars, achieving classification accuracies ranging from 90% to 95%. The use of depthwise separable convolutions enables a substantial reduction in model parameters and computational cost without compromising classification performance.

Owing to its lightweight architecture, MobileNetV2 is particularly suitable for deployment on resource-constrained devices, such as smartphones and embedded systems. During training, the model effectively learned discriminative patterns from the durian leaf dataset; however, multiple training iterations were required to achieve stable accuracy, consistent with observations reported in previous studies (Huang, Sarabi, & Ragab, 2024; Rahman et al., 2023). Using an input image resolution of 224 × 224 pixels, MobileNetV2 successfully extracted salient leaf features, allowing reliable differentiation among durian cultivars. The achieved loss value of 0.00024 indicates a very low classification error, reflecting strong convergence during training. In addition to its classification performance, the relatively small model size of MobileNetV2 offers advantages in storage efficiency and reduced energy consumption compared to conventional convolutional neural networks.

Nevertheless, optimal performance with MobileNetV2 is highly dependent on the availability of high-quality training data and effective data augmentation strategies. This dependency is consistent with prior work on plant disease detection using MobileNetV2, where enhanced feature extraction through appropriate preprocessing and augmentation yielded classification accuracies of up to 99.71% (Chen et al., 2021). Overall, the results demonstrate that MobileNetV2 provides a reliable and computationally efficient solution for durian leaf classification tasks.

Xception Performance: Accuracy and Training Characteristics

The Xception model achieved high classification performance in durian leaf identification, with accuracy values ranging from 92% to 97%. Its deep network architecture enables the extraction of fine-grained and discriminative visual features, thereby contributing to improved classification accuracy. Similar performance gains have been reported in other domains; for example, Yoon (2025) employed the Xception architecture for brain tumour classification and achieved an accuracy of 99.09%, demonstrating its capability to model complex image patterns.

During training, the classification accuracy of Xception increased steadily with the number of epochs, indicating stable convergence and effective feature learning. The model exhibited strong capability in capturing morphological characteristics of durian leaves, making it well-suited for image classification tasks that require detailed shape and texture recognition.

However, due to its deeper and more complex architecture, Xception requires greater computational resources during both training and inference compared to lightweight models. This trade-off between classification accuracy and computational cost highlights the importance of model selection based on deployment constraints and application requirements.

YOLOv11 Performance: mAP50, mAP50-95, and Inference Speed

YOLOv11 is a single-stage object detection architecture specifically designed for real-time object localisation and classification in complex image scenes. It enables simultaneous object identification and bounding box regression within a unified framework, resulting in high detection accuracy and fast inference performance (Nimma et al., 2025; Sulzbach et al., 2025).

In this study, YOLOv11 was employed to detect durian leaves in natural image environments. The model achieved a mAP50 value approaching 100%, while mAP50–95 values ranged from 97.6% to 99.9%, indicating highly consistent detection performance across varying intersection-over-union thresholds. To the best of our knowledge, the application of the YOLOv11 architecture for durian leaf detection has not been extensively explored in prior studies.

Previous research has demonstrated the effectiveness of earlier YOLO variants in plant-related object detection tasks. For instance, Abid et al. (2024) reported a mean Average Precision (mAP) of 98% using the YOLOv8 architecture for plant leaf disease detection in Bangladesh, highlighting the suitability of YOLO-based frameworks for high-precision agricultural image analysis.

In addition to its detection accuracy, YOLOv11 exhibited a fast inference speed of approximately 0.2 ms per image, underscoring its capability for real-time deployment. The one-stage detection paradigm allows YOLOv11 to identify multiple durian leaf cultivars within a single image without requiring separate processing steps. These characteristics make YOLOv11 particularly well-suited for real-time applications, such as automated durian orchard monitoring and camera-based inspection systems.

Effect of Data Augmentation on Model Performance

Data augmentation, including rotation, horizontal flip and brightness, was used in training the model to increase robustness. Integration of these augmented techniques achieved high validation accuracy: 99.02% and low validation loss: 0.0341, demonstrating excellent generalisation performance on the unseen dataset. Previous works also find that models trained only with data augmentation are not as generalisable since the diversity of input images is small. The high accuracy of validation obtained in this study indicates that data augmentation is an essential measure to overcome overfitting and be more robust to leaves with variation in orientation and illumination conditions.

Training and Validation Curve Analysis

The accuracy and loss curves of training and validation are shown in Figure 6. The curves show an orderly convergence behaviour in that both training and validation accuracy follow each other closely, with a steady decline in both training and validation loss without clear signs of overfitting.

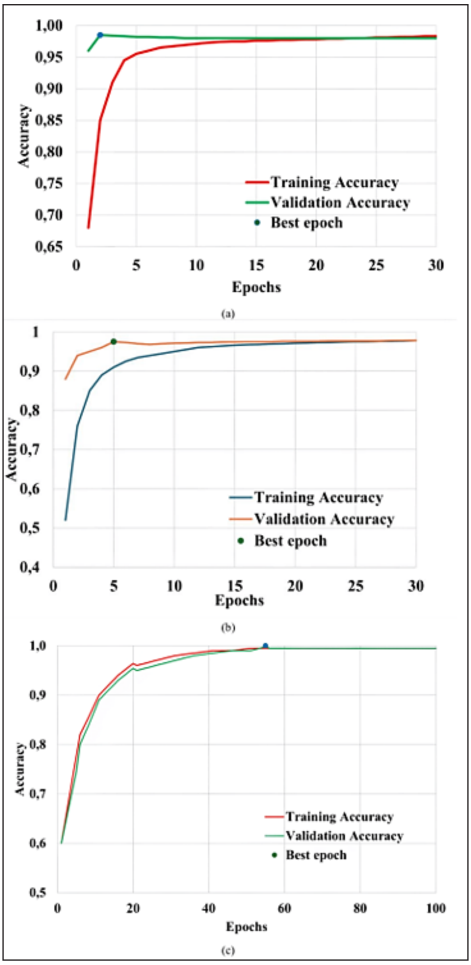

Figure 6. Training and validation accuracy results using the architectures (a) MobileNetV2; (b) Xception; and (c) YOLOv11

Such behaviour is indicative of the model not being overfit. The lack of a wide margin between the training and validation curves suggests that model generalisation is strong, a conclusion supported by the results of K-fold cross-validation analysis.

Comparative Analysis of the Three Architectures in Terms of Accuracy and Speed

The comparative analysis reveals clear performance differences among YOLOv11, MobileNetV2, and Xception for durian leaf detection and classification. Among the evaluated models, YOLOv11 demonstrated the strongest overall performance, achieving a mAP50 value close to 100% across all durian cultivars (Bawor, Kani, Monthong, Musang King, and Petruk), along with consistently high mAP50-95 values ranging from 97.6% to 99.9%. In addition, YOLOv11 exhibited a rapid inference speed of approximately 0.2 ms per image, highlighting its suitability for real-time object detection applications.

In comparison, MobileNetV2 achieved a classification accuracy of approximately 90–95%, which, while competitive, remained lower than that of YOLOv11. The MobileNetV2 architecture is optimised for computational efficiency using depthwise separable convolutions, making it lightweight and well-suited for deployment on resource-constrained platforms such as smartphones and edge devices. However, a key limitation of MobileNetV2 is its relatively slower inference speed compared to YOLOv11, as the model is designed primarily for image classification rather than real-time object detection.

Xception demonstrated higher classification accuracy than MobileNetV2, with performance ranging from 92% to 97%, but at the expense of increased computational complexity. Its deeper architecture and more intricate depthwise separable convolutional operations enhance feature representation, particularly for fine-grained morphological patterns. Nevertheless, this increased complexity results in slower inference compared to both MobileNetV2 and YOLOv11. Consequently, although Xception outperforms MobileNetV2 in classification accuracy, it remains less efficient than YOLOv11 in terms of speed and overall deployment feasibility.

These performance disparities can be primarily attributed to architectural design differences. YOLOv11 is explicitly engineered for object detection, employing a one-stage detection paradigm that performs object classification and bounding box regression simultaneously, thereby enabling high detection accuracy and fast inference. In contrast, MobileNetV2 and Xception are designed for image classification tasks, focusing on categorising a single dominant object per image rather than detecting multiple objects within complex scenes.

The YOLOv11 model used in this study comprises approximately 2.59 million parameters with a computational cost of 6.4 GFLOPs. This relatively low model complexity, compared to deeper convolutional architectures, contributes to its fast inference speed while maintaining high detection accuracy. These characteristics further underscore the suitability of YOLOv11 for real-time deployment on resource-constrained platforms.

Overall, YOLOv11 emerges as the most effective solution for durian leaf detection due to its superior accuracy and significantly faster inference speed compared to MobileNetV2 and Xception. MobileNetV2 remains a viable alternative when strict hardware constraints are present, while Xception is preferable in scenarios where classification accuracy is prioritised over inference speed. A quantitative summary of the comparative performance of MobileNetV2, Xception, and YOLOv11 is presented in Table 2.

Table 2
Performance comparison of MobileNetV2, Xception, and YOLOv11

Model	Accuracy	Speed	Excellence
MobileNet V2	90-95%	Slower than YOLOv11	Lightweight, suitable for mobile devices
Xception	92-97%	Slower than MobileNet V2 & YOLOv11	High accuracy, but computationally heavier
YOLOv11	97-100% (mAP50-95)	0.2ms per image	Real-time detection, extremely fast and accurate

K-Fold Cross Validation Results

Average Validation Accuracy and Loss Achieved Through K-Fold Cross Validation

The model was assessed through the K-Fold Cross-Validation (CV) method, which can provide more robust estimates of accuracy and mitigate arbitrariness from data partitioning (Refaeilzadeh et al., 2009). The accuracy for the validation results had an average performance of 99.02% with a loss value of 0.0341. This indicates that the model achieves high precision in distinguishing data into five classes, including bawor, kani, monthong, musang king and petruk. Table 3 shows the detailed classification report with precision, recall, and F1-score of each class. The F1-score values for all classes are in the range of 0.95~1.00, suggesting a strong balance between precision and recall (Sokolova & Lapalme, 2009).

Table 3
Precision, recall, and f1-score for each durian cultivar across folds

	Precision	Recall	F1-Score	Support
Bawor	1,00	1,00	1,00	234
Kani	1,00	1,00	1,00	480
Monthong	1,00	0,99	1,00	354
Musang King	0,99	0,92	0,95	131
Petruk	0,95	1,00	0,98	231

The bawor, kani and petruk classes achieved 100% of precision and recall, which means that the classifier identified all samples from these categories properly. On the other hand, there was a significant decrease in recall of the monthong and musang king classes, with both values equal to 0.99 and 0.95, respectively. It is important to mention that for musang king, 11 of the 131 samples were misclassified as petruk by the model. This misclassification indicates that musang king and petruk present the same morphology that could lead to confusion in sample interpretation.

Discussion of the Performance Consistency Across Different Cultivars

The evaluation results indicate that the proposed model can predict most cultivars with a relatively small margin of error. The Bawor, Kani, and Monthong cultivars exhibit near-optimal performance across all evaluation metrics. In contrast, the Musang king cultivar shows a lower recall value compared to the other cultivars, indicating a higher likelihood of misclassification, where Musang king leaves are occasionally predicted as other types.

The Petruk cultivar achieved a recall value of 1.00, demonstrating that the model successfully identified all Petruk samples. However, its precision value was slightly lower, at 0.95, suggesting that a small number of samples from other cultivars were incorrectly classified as Petruk.

Overall, the model demonstrates consistent and robust performance across all folds and cultivars, reflecting a well-balanced dataset and a stable classification model. Musang king was the only cultivar that exhibited a higher tendency for misclassification. This behaviour is likely attributable to similarities in morphological characteristics between Musang king and the Monthong or Petruk cultivars, particularly under certain lighting conditions and image orientations.

Confusion Matrix Analysis

Visual Representation of Prediction Accuracy and Errors

Figure 7 presents the confusion matrix, which clearly illustrates that most of the model's predictions lie along the main diagonal, indicating a high level of classification accuracy. Only a limited number of misclassifications are observed, with most errors occurring in the Musang king class. These off-diagonal entries indicate instances where the predicted labels do not correspond to the ground truth. Such fine-grained classification errors are common in image-based recognition tasks when morphological cues are subtle or visually overlapping (Krause et al., 2015; Wang et al., 2020).

Identification of Specific Misclassification Patterns, Particularly for the Musang King Cultivar

Further analysis of the confusion matrix, as shown in Figure 7, confirms that misclassification errors are predominantly associated with the Musang king class.

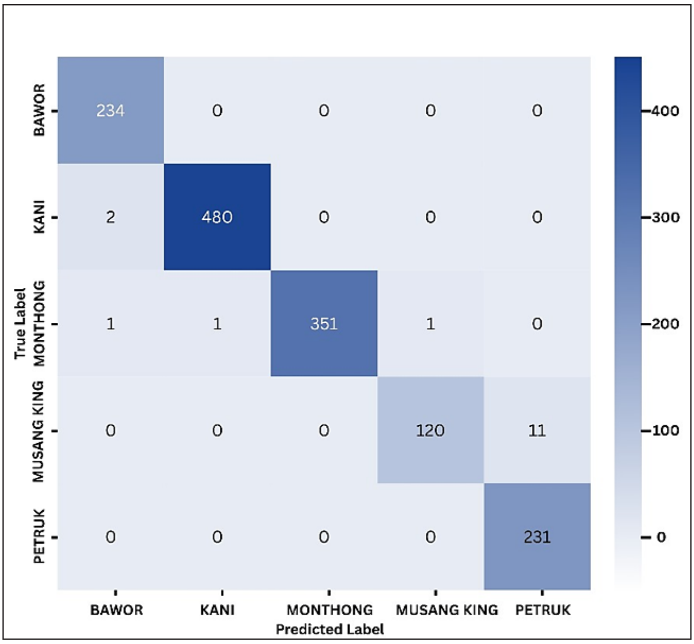

Figure 7. Confusion matrix

As noted by Chicco and Jurman (2020), confusion matrix analysis is essential for identifying class-specific error patterns. The model demonstrates stable and consistent performance for the remaining classes, achieving prediction accuracies exceeding 99%, with no notable cross-class prediction errors. The Musang king cultivar exhibits the highest frequency of misclassification, with some samples being incorrectly classified as Petruk. This confusion is likely attributable to similarities in leaf texture and edge morphology between these two cultivars, particularly when images are captured under comparable growth conditions, lighting environments, and viewing angles. Similar observations have been reported by Zhang and Wu (2012), who noted that cultivars within the same species often share highly similar visual characteristics under analogous imaging conditions.

Interpretation of High Prediction Accuracy for Other Cultivars

The consistently high classification performance observed for the Bawor, Kani, and Petruk cultivars highlights the model's strong capability to distinguish leaf images with clearly differentiated morphological traits. The pronounced intra-class consistency and inter-class separability of these cultivars provide sufficient visual cues, such as vein structures, leaf margins, and blade proportions, enabling the model to extract discriminative features effectively. This reliable differentiation underscores the robustness of the proposed model's feature-learning mechanism, particularly under controlled image acquisition conditions. These findings are consistent with Barré et al. (2017), who demonstrated that high classification accuracy in plant identification tasks can be achieved when morphological traits are both consistent and visually distinct. Similarly, Too et al. (2019) reported that convolutional neural networks perform exceptionally well in fine-grained classification tasks when trained on datasets with clear inter-class boundaries.

Insights into Visual Similarities Contributing to Misclassifications

The misclassification patterns observed in the Musang king class further emphasise the presence of visual morphological similarities among certain cultivars, a well-documented challenge in plant image classification (Zhang & Wu, 2012). These limitations may be addressed through the incorporation of additional training samples, enhanced data augmentation strategies, or the exploration of more discriminative feature extraction techniques. Despite these challenges, the results indicate that the proposed model architecture exhibits strong generalisation capability and performs effectively across all five durian cultivars. Nevertheless, improving classification performance for the Musang king class remains an important direction for future work.

ROC and AUC Analysis

To further assess the robustness of the classification model, Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) analyses were conducted using a one-vs-rest strategy for each durian cultivar. The model achieved AUC values exceeding 0.98 across all classes, indicating excellent separability among the cultivars. These high AUC scores confirm the model's strong discriminative ability, even in scenarios involving visually similar leaf structures.

Visual Explainability Using Grad-CAM

Grad-CAM visualisation was employed to examine the regions contributing most significantly to the model's classification decisions. The resulting heatmaps indicate that the model primarily focuses on discriminative leaf regions, including vein patterns, margins, and texture characteristics. This behaviour suggests that the model learns meaningful botanical features rather than relying on background artefacts, thereby enhancing both interpretability and reliability of the classification outcomes.

Discussion of Findings and Comparison with Existing Literature

Architectural Factors Behind the Superior Performance of YOLOv11

The superior real-time detection and classification performance of YOLOv11 observed in this study can be primarily attributed to its one-stage detection architecture, which enables simultaneous feature extraction, localisation, and classification within a single forward pass.

Unlike classification-based models that operate on pre-cropped or single-object images, YOLOv11 is inherently designed to handle spatial variability and background noise, conditions that commonly occur in field-acquired agricultural images.

Although MobileNetV2, Xception, and YOLOv11 all achieved high classification accuracy exceeding 99% shown in Figure 6, their training behaviour and deployment characteristics differed substantially. MobileNetV2 converged rapidly, reaching its peak accuracy within only two epochs, which reflects its lightweight design and reduced parameter space optimised for computational efficiency (Sandler et al., 2018). However, this efficiency comes at the cost of limited representational capacity, restricting its ability to capture highly nuanced intra-species variations.

Xception, by contrast, required more training epochs to reach optimal performance, owing to its deeper architecture and extensive use of depthwise separable convolutions, which enhance its ability to model subtle spatial and textural differences among visually similar leaf cultivars (Chollet, 2017). This architectural depth, referring to the greater number of convolutional layers, explains its improved accuracy relative to MobileNetV2, but also accounts for its increased computational burden and reduced inference speed.

YOLOv11 exhibited a longer convergence process compared to MobileNetV2, which is consistent with observations reported in prior object detection research (Xie et al., 2017). Nevertheless, its ability to jointly learn localisation and classification features provides a decisive advantage in real-world scenarios where object scale, orientation, and illumination vary significantly (Bochkovskiy et al., 2020; Redmon et al., 2016). These findings align with previous studies emphasising the inherent trade-offs between model depth, inference speed, and feature discrimination in agricultural image analysis (Kamilaris & Prenafeta-Boldu, 2018; Saleem et al., 2019). Consequently, MobileNetV2 is most suitable for lightweight edge deployments, Xception for applications prioritising feature richness, and YOLOv11 for robust real-time detection and classification in dynamic agricultural environments.

Technical Rationale for the Effectiveness of the Web-Based YOLOv11 System

The successful implementation of a web-based durian leaf classification system using YOLOv11 further demonstrates the practical significance of the proposed approach, as shown in Figure 8. The system's ability to process images uploaded through a browser interface or captured via a live camera is enabled by YOLOv11's low-latency inference pipeline.

The high confidence scores achieved for cultivars such as Monthong (up to 99.9%) are not solely a function of model accuracy but also reflect the consistency between learned feature representations and the morphological characteristics of the input images. The integration of detection and classification within a single model minimises intermediate processing steps, thereby reducing error propagation and improving prediction stability.

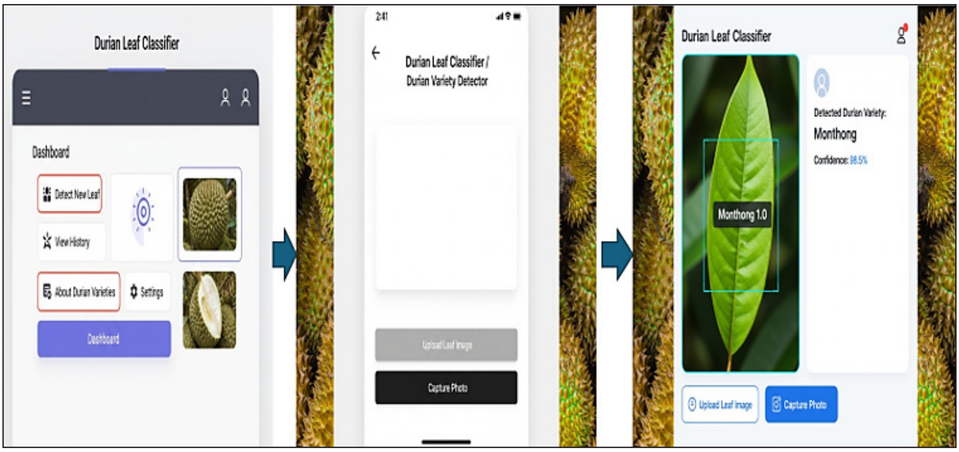

Figure 8. Web-based durian leaf classification

This architectural efficiency explains why YOLOv11 is particularly well-suited for browser-based and real-time agricultural applications, where responsiveness and reliability are critical.

Importance of Real-time Inference Speed in Agricultural Field Applications

Real-time inference speed is a decisive factor in the deployment of computer vision systems in agricultural contexts. The high processing speed demonstrated by YOLOv11 enables immediate feedback during field operations, such as seedling verification, orchard monitoring, and visual quality inspection.

From a practical standpoint, rapid inference reduces dependency on offline laboratory analysis and expert visual assessment, both of which are time-consuming and prone to subjectivity. By enabling near-instantaneous cultivar identification, YOLOv11 facilitates faster decision-making and enhances operational efficiency across multiple stages of the agricultural supply chain. This advantage is particularly relevant in large-scale durian cultivation systems, where timely interventions directly impact productivity and quality control.

Addressing the Challenge of Intra-species Leaf-based Classification

Intra-species classification presents a considerably greater challenge than inter-species classification due to high morphological similarity among cultivars. Durian leaf cultivars often share overlapping features in shape, texture, and venation patterns, which complicates discriminative feature learning.

The high classification performance achieved in this study indicates that deep learning architectures, when combined with effective feature extraction strategies, can capture subtle intra-class distinctions. This finding extends existing literature that has predominantly focused on interspecific classification and demonstrates the feasibility of applying advanced deep learning models to fine-grained, within-species agricultural identification tasks.

Implications for Durian Seed Certification and Agribusiness Management

The results of this study have important implications for durian seed certification, orchard management, and cultivar control within the broader agribusiness ecosystem. By reducing reliance on manual observation and expert judgment, the proposed approach enhances objectivity and repeatability in cultivar identification.

The integration of real-time detection capabilities with high classification accuracy enables faster verification processes, minimises misidentification risks, and supports scalable deployment across production and distribution stages. As a result, this system has the potential to improve traceability, standardisation, and quality assurance throughout the durian agribusiness value chain.

CONCLUSION

This study addresses the challenge of durian cultivar identification based on leaf morphology, which is inherently difficult due to high intra-species similarity and dependence on expert visual assessment. The results demonstrate that deep learning, particularly the YOLOv11 architecture, effectively overcomes this limitation by learning discriminative spatial features, enabling reliable and non-destructive identification under real-world conditions.

The proposed approach directly addresses the research problem by combining high classification accuracy with real-time inference capability, making it suitable for practical agricultural deployment. To support real-world implementation, a web-based system was developed that allows stakeholders, including nursery operators, orchard managers, and certification authorities, to capture or upload leaf images using standard mobile devices for immediate verification.

Within agricultural workflows, the system can be applied to seedling certification, orchard establishment, and varietal quality control across the durian supply chain. Despite practical constraints such as variable lighting and leaf orientation, the model demonstrates strong robustness and fast inference, enabling efficient field-level decision-making. Overall, this study provides a scalable and application-ready solution that bridges the gap between research and operational needs in durian agribusiness.

ACKNOWLEDGEMENT

We would like to extend our utmost gratitude to the IT Centre of The University of Jember for providing the Supercomputer NVIDIA Station DGX A 100 for this project.

REFERENCES

Abid, S., Jahan, B., Al Mamun, A., Jakir, H., & Hossain Mazumder, S. (2024). Bangladeshi crop leaf disease detection using YOLOv8. *Heliyon, 10*(18), Article e36694. https://doi.org/10.1016/j.heliyon.2024.e36694

Ahmad, A., Saraswat, D., & El Gamal, A. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for the development of appropriate tools. *Smart Agricultural Technology, 3*, Article 100083. https://doi.org/10.1016/j.atech.2022.100083

Barré, P., Stöver, B. C., Müller, K. F., & Steinhage, V. (2017). LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics, 40*, 50-56. https://doi.org/10.1016/j.ecoinf.2017.05.005

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv*. https://doi.org/10.48550/arXiv.2004.10934

Bottou, L. (2012). Stochastic gradient descent tricks. In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (2nd ed., pp. 421-436). Springer. https://doi.org/10.1007/978-3-642-35289-8_25

Chen, J., Zhang, D., Suzauddola, M., & Zeb, A. (2021). Identifying crop diseases using attention embedded MobileNet-V2 model. *Applied Soft Computing, 113*, Article 107901. https://doi.org/10.1016/j.asoc.2021.107901

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*(1), Article 6. https://doi.org/10.1186/s12864-019-6413-7

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251-1258). IEEE. https://doi.org/10.1109/CVPR.2017.195

Gorriz, J. M., Clemente, R. M., Segovia, F., Ramirez, J., Ortiz, A., & Suckling, J. (2024). Is k-fold cross-validation the best model selection method for machine learning? *arXiv*. https://doi.org/10.48550/arXiv.2401.16407

Hasan, M. N., Miah, M. S., Ghose, P., Jannat, T., Bhuyain, M. M. H., Chowdhury, M. S. A., ... & Harun-Ar-Rashid, M. (2025). A Deep Learning Approach for Brain Tumor Diagnosis: Combining an 8 Layer CNN with Rigorous K-Fold Validation. *Mathematical Modelling of Engineering Problems*, *12*(12). https://doi.org/10.18280/mmep.121227

Horiuchi, A., Masuda, K., Shirasawa, K., Onoue, N., Matsuzaki, R., Tao, R., Kubo, Y., Ushijima, K., & Akagi, T. (2023). Genetic basis of lineage-specific evolution of fruit traits in hexaploid persimmon. *DNA Research, 30*(5), Article dsad015. https://doi.org/10.1093/dnares/dsad015

Huang, C., Sarabi, M., & Ragab, A. E. (2024). MobileNet-V2/IFHO model for accurate detection of early-stage diabetic retinopathy. *Heliyon, 10*(17), Article e37293. https://doi.org/10.1016/j.heliyon.2024.e37293

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70-90. https://doi.org/10.1016/j.compag.2018.02.016

Kazi, A. (2024). Multiclass classification using enhanced MobileNetV2 architecture. *International Journal of Scientific Research in Engineering and Management, 8*(6), 1-5. https://doi.org/10.55041/IJSREM35845

Krause, J., Jin, H., Yang, J., & Fei-Fei, L. (2015). Fine-grained recognition without part annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5546-5555). IEEE. https://doi.org/10.1109/CVPR.2015.7299173

Kusumaningtyas, E. M., Ramadijanti, N., & Rijal, I. H. K. (2024). Convolutional neural network implementation with MobileNetV2 architecture for Indonesian herbal plants classification in mobile app. In *2024 International Electronics Symposium (IES)* (pp. 521-527). IEEE. https://doi.org/10.1109/IES63037.2024.10665862

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. https://doi.org/10.1038/nature14539

Maaroof, M. K. A., & Bouhlel, M. S. (2025). Real-time object detection using YOLO-8 model: A drone-based approach. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16*(1), 190-204. https://doi.org/10.58346/JOWUA.2025.I1.011

Maesaroh, I. S., Syaja'ah, K. N., Perkasa, Y. S., Hadianti, S., Riana, D., & Nurmalasari, R. R. (2024). Cervical cancer classification from Pap smear using Xception model. In *2024 10th International Conference on Wireless and Telematics (ICWT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICWT62080.2024.10674721

Magandhi, M., Sobir, W. Y., & Sudarmono, M. D. (2024). DNA barcoding of endemic durian kura-kura in West Kalimantan, Indonesia. *SABRAO Journal of Breeding and Genetics, 56*(5). https://doi.org/10.54910/sabrao2024.56.5.5

May, O. S., Ismail, W. N. A. B. W., & Jun, C. (2023). The king of fruits dilemma in Malaysia: Discovering durian export challenges to China. *Journal of Engineering Research and Education, 15*, 61-75. https://doi.org/10.58915/jere.v15.2023.652

Meethongjan, K., Hoang, V. T., & Surinwarangkoon, T. (2022). Data augmentation by combining feature selection and colour features for image classification. *International Journal of Electrical and Computer Engineering, 12*(6), 6172-6177. https://doi.org/10.11591/ijece.v12i6.pp6172-6177

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, Article 1419. https://doi.org/10.3389/fpls.2016.01419

Nimma, D., Al-Omari, O., Pradhan, R., Ulmas, Z., Krishna, R. V. V., El-Ebiary, T. Y. A. B., & Rao, V. S. (2025). Object detection in real-time video surveillance using attention-based transformer-YOLOv8 model. *Alexandria Engineering Journal, 118*, 482-495. https://doi.org/10.1016/j.aej.2025.01.032

Rahman, M. M., Basar, M. A., Shinti, T. S., Khan, M. S. I., Babu, H. M. H., & Uddin, K. M. M. (2023). A deep CNN approach to detect and classify local fruits through a web interface. *Smart Agricultural Technology, 5*, Article 100321. https://doi.org/10.1016/j.atech.2023.100321

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779-788). IEEE. https://doi.org/10.1109/CVPR.2016.91

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 532-538). Springer. https://doi.org/10.1007/978-0-387-39940-9_565

Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants, 8*(11), Article 468. https://doi.org/10.3390/plants8110468

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4510-4520). IEEE. https://doi.org/10.1109/CVPR.2018.00474

Sengupta, P., Mehta, A., & Rana, P. S. (2023). Enhancing performance of deep learning models with a novel data augmentation approach. In *2023, the 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICCCNT56998.2023.10308298

Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science, 21*(2), 110-124. https://doi.org/10.1016/j.tplants.2015.10.015

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

Sulzbach, E., Scheeren, I., Torres Veras, M. S., Tosin, M. C., Ellert Kroth, W. A., Merotto, A., & Markus, C. (2025). Deep learning model optimisation methods and performance evaluation of YOLOv8 for enhanced weed detection in soybeans. *Computers and Electronics in Agriculture, 232*, Article 110117. https://doi.org/10.1016/j.compag.2025.110117

Tejada, C., Espinoza, G., & Subauste, D. (2024). Use of Xception architecture for the classification of skin lesions. In *Proceedings of the International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2024)* (pp. 129-134). https://doi.org/10.54808/IMCIC2024.01.129

Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture, 161*, 272-279. https://doi.org/10.1016/j.compag.2018.03.032

Ullah, N., Khan, J. A., Almakdi, S., Alshehri, M. S., Al Qathrady, M., Aldakheel, E. A., & Khafaga, D. S. (2023). A lightweight deep learning-based model for tomato leaf disease classification. *Computers, Materials & Continua, 77*(3), 3969-3992. https://doi.org/10.32604/cmc.2023.041819

Uzer, M. S. (2025). Deep learning-based classification consisting of pre-trained models and the proposed model using k-fold cross-validation for pistachio species. *Applied Sciences, 15*(8), Article 4516. https://doi.org/10.3390/app15084516

Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated plant species identification: Trends and future directions. *PLoS Computational Biology, 14*(4), Article e1005993. https://doi.org/10.1371/journal.pcbi.1005993

Wang, N., Ma, S., Li, J., Zhang, Y., & Zhang, L. (2020). Multistage attention network for image inpainting. *Pattern Recognition, 106*, Article 107448. https://doi.org/10.1016/j.patcog.2020.107448

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1492-1500). IEEE. https://doi.org/10.1109/CVPR.2017.634

Yoon, S. (2025). Brain tumour classification using a hybrid ensemble of Xception and parallel deep CNN models. *Informatics in Medicine Unlocked, 54*, Article 101629. https://doi.org/10.1016/j.imu.2025.101629

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems* (Vol. 27, pp. 3320-3328). https://doi.org/10.48550/arXiv.1411.1792

Zeynali, A., Tinati, M. A., & Tazehkand, B. M. (2024). Hybrid CNN-transformer architecture with Xception-based feature enhancement for accurate breast cancer classification. *IEEE Access, 12*, 189477-189493. https://doi.org/10.1109/ACCESS.2024.3516535

Zhang, Y.-D., & Wu, L. (2012). An MR brain image classifier via principal component analysis and kernel support vector machine. *Progress in Electromagnetics Research, 130*, 369-388. https://doi.org/10.2528/PIER12061410

Zhang, Z., Hou, J., & Lin, Z. (2024). Exploring integration methods for image data augmentation. In *2024 6th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI)* (pp. 707-713). IEEE. https://doi.org/10.1109/IoTAAI62601.2024.10692583