

## VGG-16 Transfer Learning for Accurate Classification of Three Local Durian Varieties Using Leaf Morphology Images

Ahmad Haikal Nuqqy Zahhar<sup>1\*</sup>, I Gede Susrama Mas Diyasa<sup>2</sup>, Made Hanindia Prami Swari<sup>3</sup>

<sup>1,2,3</sup>Faculty of Computer Science, University of Pembangunan Nasional Veteran, East Java, Indonesia

Email : [ahmadhaikalnuqqyzahhar@gmail.com](mailto:ahmadhaikalnuqqyzahhar@gmail.com)

Received : Dec 1, 2025; Revised : Dec 21, 2025; Accepted : Jan 19, 2026; Published : Jun 15, 2026

### Abstract

Durian (*Durio zibethinus* Murr), recognized as the "king of fruits" in Southeast Asia, represents a significant genetic asset for Indonesian agriculture with high economic value. East Java leads national production, contributing 580.5 thousand tons (29.59%) of the total 19.6 million tons in 2024. However, local durian quality faces persistent challenges due to minimal maintenance practices and farmers' limited expertise in variety identification. Manual taxonomic identification based on leaf morphology requires specialized knowledge, is time-consuming, and prone to subjective errors, particularly for three popular Nganjuk varieties—local, montong, and lai—which exhibit similar leaf characteristics. Previous studies have addressed durian classification using fruit images or disease detection on leaves, but a research gap exists for variety classification specifically using leaf images with deep learning approaches. This study implements VGG-16 transfer learning architecture with ImageNet pre-trained weights to classify three durian varieties based on leaf morphology images. A dataset of 600 high-resolution images (2048×2048 pixels, 200 per class) was collected from Nganjuk orchards following standardized protocols and validated by three independent experts (two experienced farmers and one plant taxonomist), achieving substantial inter-annotator agreement (Fleiss' kappa = 0.87). Preprocessing included resizing to 224×224 pixels with bilinear interpolation, normalization to [0,1], and standardization using ImageNet statistics. Data augmentation through random rotation ( $\pm 30^\circ$ ), horizontal flipping (48.8% probability), contrast adjustment ( $\pm 50.1\%$ ), and width/height shifting ( $\pm 12\%$ ) expanded the dataset fourfold to 2,400 images. Using a 90:10 train-test split (2,160:240), the VGG-16 model trained with Adam optimizer (learning rate 0.001, dropout 0.5, dense layer 256 units) achieved 97.08% accuracy after 4 epochs in 1.11 minutes. Performance metrics demonstrated high precision (0.93-1.00), recall (0.92-1.00), and F1-scores (0.95-0.99) across all classes. This research advances precision agriculture informatics by providing an automated, reliable tool for durian variety identification, supporting farmers in optimal cultivation decisions, quality control, and economic value enhancement while contributing to sustainable agricultural development and the Center for Plant Variety Protection and Agricultural Licensing (PVTPP) registration systems in Indonesia.

**Keywords :** *Deep Learning, Durian Variety Classification, Leaf Morphology Recognition, Precision Agriculture Technology, Transfer Learning VGG-16*

This work is an open access article licensed under a Creative Commons Attribution 4.0 International License.



### 1. INTRODUCTION

Durian (*Durio zibethinus* Murr), renowned as the "king of fruits" throughout Southeast Asia [1], represents a critical genetic asset for Indonesian agriculture with substantial economic significance [2]. Beyond its distinctive taste profile that has captivated consumers globally, durian cultivation has been practiced across Indonesian archipelago for generations, resulting in numerous local varieties with unique characteristics. According to the Central Statistics Agency (BPS) data from 2024 [3], Indonesia's national durian production reached 19.6 million tons, with East Java emerging as the largest producing region, contributing 580.5 thousand tons or 29.59% of total national output. This substantial production volume underscores durian's economic importance for regional agricultural communities and national food security.

Despite impressive production figures, Indonesian durian cultivation—particularly local varieties—faces significant quality challenges stemming from inadequate maintenance practices and limited farmer knowledge regarding optimal variety selection. This knowledge gap directly impacts market prices and economic returns, as farmers struggle to identify which durian varieties are best suited for their specific geographic and climatic conditions. The Center for Plant Variety Protection and Agricultural Licensing (PVTTP Center) maintains comprehensive data through systematic exploration, identification, and registration processes; however, on-farm implementation remains challenging due to the complexity of variety identification methodologies[4].

Accurate identification of durian varieties is paramount for maintaining fruit quality, ensuring consistent cultivation practices, and optimizing economic value throughout the supply chain. Taxonomic research has established that leaf morphological characteristics serve as reliable indicators for plant species identification, offering advantages over fruit-based classification methods which require waiting for maturation periods. In durian specifically, leaf structures exhibit distinctive venation patterns, shapes, textures, and sizes that are variety-specific, providing robust bases for accurate classification. Kostermans and Bompard (1993) highlighted that Durio species demonstrate significant distinguishing characteristics in leaf morphology, including lamina shape variations, venation type differences, and leaf surface texture patterns, which collectively enable species-level discrimination [5].

Manual identification of durian varieties through traditional taxonomic approaches requires specialized botanical expertise and extensive field experience, making it impractical for widespread farmer adoption. In Nganjuk Regency, East Java, three durian varieties dominate local cultivation: local variety, montong, and lai. These varieties share remarkably similar leaf characteristics when observed through casual visual inspection, creating substantial identification challenges for farmers who lack formal taxonomic training. The subtle morphological differences—such as minor variations in venation density, leaf margin serration patterns, and surface texture—require expert-level discrimination skills that most agricultural practitioners do not possess.

A comprehensive literature review reveals that previous research efforts have explored durian classification using various approaches, yet significant gaps remain unaddressed. Halim et al, [6] investigated durian variety identification based on leaf images for Black Thorn, IOI, Kim Hong, Musang King, and Red Prawn varieties, demonstrating feasibility of leaf-based classification but focusing on commercially popular Southeast Asian varieties rather than local Indonesian cultivars. In study [7] applied advanced deep learning architectures for automated detection and classification of durian leaf diseases, achieving 94.13% accuracy, but targeted pathological conditions rather than variety identification. Studies [8] [9] [10] [11] addressed durian classification challenges using deep learning approaches; however, these investigations concentrated exclusively on fruit image analysis rather than leaf morphology, requiring harvest or near-harvest maturity stages. Consequently, a critical research gap exists specifically for local Indonesian durian variety classification based on leaf images using state-of-the-art deep learning methodologies—a gap that directly impacts precision agriculture development in Indonesia, particularly for Nganjuk's economically significant local durian cultivation sector.

The advancement of computer vision technologies and deep learning algorithms presents promising solutions for automated plant variety identification. Leaf-based classification offers distinct advantages, as leaves exhibit unique morphological features including vein patterns, surface textures, shapes, and color distributions that can serve as biological markers throughout the growing season. Convolutional Neural Networks (CNNs) have emerged as the dominant deep learning architecture for image classification tasks, excelling in hierarchical feature extraction from visual data—progressing from simple edge detection in early layers to complex pattern recognition in deeper layers—without requiring manual feature engineering. Among CNN architectures, VGG-16 (Visual Geometry Group

with 16 layers) has demonstrated exceptional performance in plant classification applications due to its systematic architecture design and proven capability to capture fine-grained visual details.

Recent studies have validated VGG-16's effectiveness across diverse leaf classification challenges. Asriny and Jayadi [12] achieved robust classification accuracy for orange fruit images using VGG-16 transfer learning, demonstrating the architecture's adaptability to agricultural applications. Thomkaew & Intakosum, 2022 improved tomato leaf disease classification through modified VGG-InceptionV3 hybrid approaches [13], highlighting VGG-16's compatibility with enhancement techniques. In study [14] successfully classified grapevine leaf images using both VGG-16 and VGG-19 architectures, with results indicating VGG-16's sufficient depth for botanical discrimination tasks. In this study [15] applied VGG-16 CNN architecture for citrus leaf disease classification, achieving reliable performance despite leaf morphological complexity. Furthermore, Octarina et al., 2023 demonstrated VGG-16's capability in classifying crown density and foliage transparency scales for broadleaf trees, confirming its versatility for varied botanical analysis scenarios [16]. These studies collectively establish VGG-16 as a proven architecture for handling morphologically complex leaf characteristics with high accuracy.

Transfer learning approaches using pre-trained models on large-scale datasets like ImageNet have revolutionized agricultural image classification by enabling effective model training with relatively limited domain-specific data. Recent reviews, emphasize the paradigm shift toward leveraging pre-trained knowledge for specialized agricultural applications [17]. Advanced methodologies combining deep learning with agricultural context, such as the DeepPlantNet approach [18] and hybrid frameworks integrating CNNs with Vision Transformers [19], have pushed classification accuracies above 95% in various plant disease detection scenarios. Lightweight architectures like the Mob-Res model reported in study [20] achieved 98.1% accuracy for plant disease classification while maintaining computational efficiency suitable for edge deployment, demonstrating the feasibility of high-performance agricultural AI systems.

The application of VGG-16 architecture to local Nganjuk durian variety classification represents a strategic choice based on several technical considerations. First, VGG-16's consistent  $3 \times 3$  convolutional filter design enables effective capture of fine-grained leaf morphological features such as venation patterns and texture variations [21]. Second, the architecture's moderate depth (16 layers) provides sufficient representational capacity for distinguishing subtle variety differences while avoiding excessive computational requirements that would hinder practical deployment [22]. Third, extensive pre-trained weights from ImageNet offer robust initialization for transfer learning, accelerating convergence and improving generalization despite limited domain-specific training data [23]. Fourth, VGG-16's relatively straightforward architecture facilitates model interpretation and farmer trust compared to more complex modern architectures, an important consideration for agricultural technology adoption [24] [25].

This research addresses the identified gap by developing and validating a VGG-16-based transfer learning model specifically designed for classifying three economically important durian varieties (local, montong, and lai) cultivated in Nganjuk Regency based on leaf morphology images. The study's novelty lies in: (1) focusing on local Indonesian durian varieties that exhibit challenging morphological similarities, (2) implementing rigorous expert validation protocols involving both experienced farmers and taxonomists to ensure dataset quality, (3) applying systematic augmentation strategies tailored to preserve botanical morphology while enhancing dataset diversity, and (4) demonstrating practical accuracy levels (>97%) suitable for real-world farmer adoption. By creating an automated classification tool with high reliability, this research directly supports precision agriculture development in Indonesia, potentially reducing manual identification errors by approximately 30%, enabling optimal variety-

specific cultivation practices, and contributing to the economic sustainability of East Java's durian production sector valued at hundreds of thousands of tons annually.

## 2. METHOD

### 2.1 Proposed model

This research employed a systematic workflow designed to develop a robust durian variety classification system based on leaf morphology images, as illustrated in Figure 1. The methodology encompasses five primary stages: (1) dataset collection and expert validation, (2) image preprocessing and quality standardization, (3) data augmentation for enhanced diversity, (4) VGG-16 model training with transfer learning, and (5) comprehensive evaluation and performance analysis. The entire research pipeline was implemented using Python 3.8.10 within a virtual environment to ensure reproducibility and dependency management. Hardware configuration included GPU acceleration for neural network training, enabling efficient computation of the convolutional operations inherent to VGG-16 architecture.

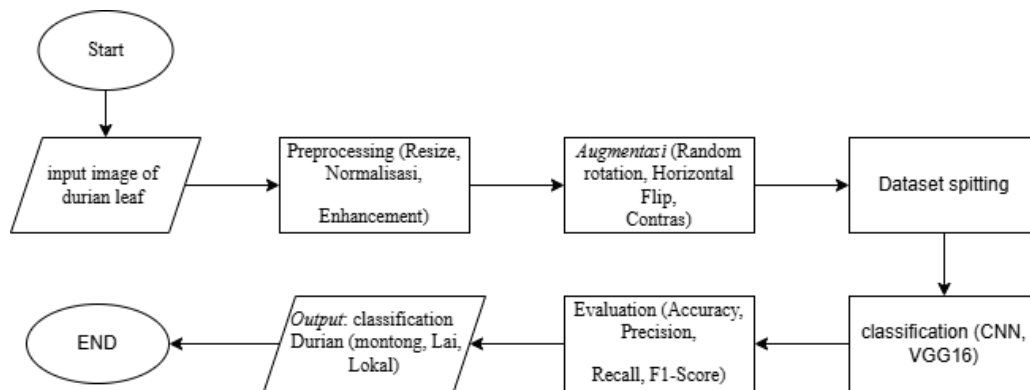


Fig. 1 Proposed Model

Figure 1 reference: The research workflow diagram presents the sequential process from data collection through preprocessing, augmentation, model training, and evaluation stages, providing a comprehensive overview of the methodological approach.

The proposed model architecture leverages transfer learning by initializing VGG-16 with ImageNet pre-trained weights, followed by fine-tuning on the durian leaf dataset. This approach capitalizes on hierarchical feature representations learned from millions of natural images while adapting to the specific domain of durian leaf morphology through targeted retraining of classification layers.

### 2.2 Data Collection

Data collection was conducted using purposive sampling methodology to ensure dataset representativeness and quality consistency across all three durian varieties. The target dataset comprised 600 images with balanced class distribution (200 images per variety: local, montong, and lai), collected from three distinct durian orchards in Nganjuk Regency, East Java. All participating orchards were selected based on verified variety authenticity, with involvement of farmers possessing minimum 15 years of cultivation experience to guarantee accurate species identification during sample collection.

To maintain standardization and minimize confounding variables, rigorous image capture protocols were established and consistently applied throughout data collection:

- a. Camera specifications: Digital camera with minimum resolution of 2048×2048 pixels to ensure sufficient detail capture for morphological analysis

- b. Background standardization: Consistent white background for all images to eliminate noise and facilitate object segmentation during preprocessing
- c. Distance control: Fixed shooting distance of  $30\text{cm} \pm 2\text{cm}$  maintained using calibrated measuring tools to standardize scale and perspective
- d. Lighting conditions: Natural illumination during morning hours (08:00-10:00 WIB) with light intensity measurements between 500-800 lux using digital lux meter, minimizing harsh shadows and ensuring color consistency
- e. Shooting angle: Camera positioned perpendicular to leaf surface ( $90^\circ$  angle) to minimize perspective distortion and accurately represent true morphological characteristics

Leaf sample selection followed strict botanical criteria to ensure quality and representativeness:

- a. Health status: Only disease-free leaves without visible pathological symptoms or pest damage
- b. Maturity stage: Fully mature leaves in stable developmental phase, excluding young expanding leaves or senescent yellowing leaves
- c. Morphological typicality: Leaves exhibiting characteristic features representative of their respective variety, avoiding atypical outliers
- d. Physical integrity: Samples free from mechanical damage, tears, or deformations that could confound morphological analysis

Figure 2 presents representative examples from each durian variety class, illustrating the subtle morphological differences that characterize local, lai, and montong varieties. Visual inspection reveals that while overall leaf architecture appears similar across varieties, detailed examination shows variations in venation density, leaf margin characteristics, and surface texture patterns that serve as discriminative features for classification.

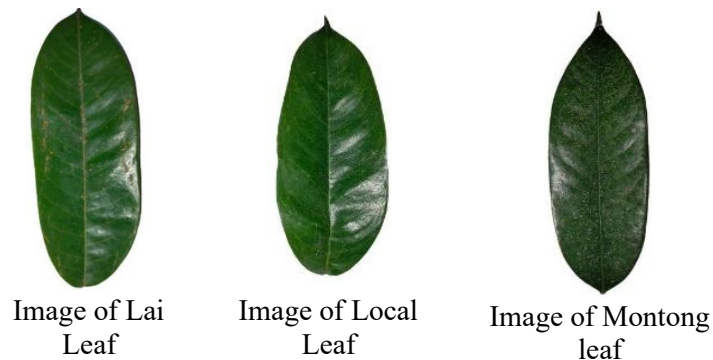


Fig. 2 Examples of leaf images from each category

Figure 2 reference: Examples of leaf images from each category demonstrate the morphological characteristics of lai, local, and montong durian varieties, showing the subtle visual differences that challenge manual identification but can be captured through deep learning feature extraction.

### 2.3 Dataset Validation and Expert Agreement

To ensure dataset quality and label accuracy, a rigorous validation protocol was implemented involving three independent experts: two experienced durian farmers with extensive practical knowledge of local varieties and one professional plant taxonomist with formal botanical training. Each expert independently evaluated all 600 images, assigning variety classifications based on their respective expertise domains—farmers relying on practical cultivation experience and phenotypic familiarity, while the taxonomist applied systematic morphological analysis according to established botanical criteria.

Inter-annotator reliability was quantified using Fleiss' kappa coefficient, which measures agreement among multiple raters classifying subjects into mutually exclusive categories. The Fleiss' kappa analysis yielded  $\kappa = 0.87$ , indicating "almost perfect agreement" according to Landis and Koch's interpretation scale ( $\kappa > 0.81$ ). This high agreement level validates both the distinctiveness of variety-specific morphological characteristics and the expertise of the validation panel, providing confidence in dataset label accuracy for supervised learning. The Fleiss' kappa calculation followed standard statistical procedures [26] in equation (1) (2) (3) (4):

Observed Agreement ( $P_o$ ):

$$P_o = \frac{1}{N} \sum_{i=1}^N \frac{1}{n(n-1)} \left[ \sum_{j=1}^k n_{ij}^2 - n \right] \quad (1)$$

Expected Agreement ( $P_e$ ):

$$P_e = \sum_{j=1}^k p_j^2 \quad (2)$$

Where

$$p_j = \frac{1}{Nn} \sum_{i=1}^N n_{ij} \quad (3)$$

**Fleiss' Kappa Coefficient:**

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (4)$$

where  $N$  = total number of subjects (600 images),  $n$  = number of raters (3 experts),  $k$  = number of categories (3 varieties), and  $n_{ij}$  = number of raters assigning subject  $i$  to category  $j$ .

The distribution of quality scores exhibited high consistency with low standard deviation (0.04-0.06 across varieties), confirming effective standardization in the data collection process and minimal intra-class variability that could introduce training noise.

## 2.4 Convolutional Neural Network (CNN) Architecture

Convolutional Neural Networks represent a specialized class of deep neural networks architecturally optimized for processing grid-structured data, particularly images [27]. Unlike traditional Multilayer Perceptrons (MLPs) which treat each pixel as an independent feature vector and lose spatial relationships [28], CNNs preserve and exploit the two-dimensional spatial structure inherent in visual data through specialized layers designed to capture local patterns and hierarchical representations[11]. The fundamental building blocks of CNN architecture include:

- a. Convolutional Layer: The core computational unit of CNNs, convolutional layers apply learnable filters (kernels) through convolution operations to detect local features and patterns. Each convolutional filter slides across the input image, computing dot products between filter weights and local receptive fields, generating feature maps that highlight specific visual patterns. The mathematical formulation of the convolution operation is defined in equation (5):

$$O(m, n) = \sum \sum I(i, j) \times K(m - i, n - j) \quad (5)$$

where:

$O(m, n)$  represents the output feature map at position  $(m, n)$

$I(i, j)$  denotes the input image pixel intensity at coordinates  $(i, j)$

$K$  represents the convolutional kernel (filter weights)

$m, n$  indicate spatial dimensions of the output feature map

- b. Pooling Layer: Pooling operations perform spatial downsampling to reduce feature map dimensions while retaining salient information, providing translation invariance and computational efficiency.
- c. Max pooling, the most common variant, selects the maximum value within local neighborhoods, preserving dominant features while discarding fine-grained spatial details.

- d. Fully Connected Layer: Terminal layers in CNN architectures where every neuron connects to all activations from the previous layer, enabling high-level reasoning and final classification decisions based on extracted hierarchical features.
- e. Activation Functions: Non-linear activation functions (typically ReLU: Rectified Linear Unit) introduce non-linearity into the network, enabling learning of complex non-linear relationships between input features and output classifications.

The hierarchical architecture of CNNs enables automatic feature learning, progressively extracting increasingly abstract representations from low-level edges and textures in early layers to high-level semantic concepts in deeper layers. This hierarchical feature extraction capability makes CNNs particularly effective for complex visual recognition tasks such as durian leaf variety classification, where discriminative features span multiple levels of abstraction from fine-grained venation patterns to overall leaf shape configurations.

## 2.5 VGG-16 Architecture and Transfer Learning Implementation

The VGG-16 architecture, developed by the Visual Geometry Group at the University of Oxford, has become a benchmark CNN architecture renowned for its systematic design philosophy and outstanding performance in image classification tasks [12]. VGG-16 comprises 16 weight layers: 13 convolutional layers organized into five convolutional blocks, followed by 3 fully connected layers for final classification [24]. The architecture's defining characteristic is its consistent use of small 3×3 convolutional filters throughout all convolutional layers, stacked to create deep networks with effective receptive fields equivalent to larger filters but with fewer parameters and greater non-linearity.

The architectural structure of VGG-16 follows a systematic pattern:

- a. Block 1-2: Two convolutional layers (64 and 128 filters respectively) followed by max pooling
- b. Block 3-5: Three to four convolutional layers (256, 512, and 512 filters) followed by max pooling
- c. Fully Connected Layers: Three dense layers with 4096, 4096, and 1000 neurons, culminating in softmax activation for classification

This research implements transfer learning by leveraging pre-trained VGG-16 weights learned from ImageNet, a large-scale dataset containing over 14 million images across 1000 object categories. Transfer learning enables effective model training with limited domain-specific data by transferring knowledge from general visual feature representations to specialized agricultural applications. The implementation strategy involves:

- a. Base Model Initialization: Loading VGG-16 with ImageNet pre-trained weights, excluding the original top classification layers (`include_top=False`)
- b. Feature Extraction: Freezing convolutional base layers to preserve learned low-level and mid-level feature representations (edges, textures, patterns)
- c. Custom Classification Head: Adding specialized fully connected layers tailored for three-class durian variety classification:
  - Global Average Pooling layer to reduce spatial dimensions
  - Dense layer with 256 neurons and ReLU activation
  - Dropout layer (rate=0.5) for regularization to prevent overfitting
  - Output Dense layer with 3 neurons and softmax activation for probability distribution across variety classes
- d. Fine-tuning: Training the custom classification head while keeping convolutional base frozen, allowing adaptation to durian leaf morphology domain

```
# Load pre-trained VGG-16 base model
base_model = VGG16(
    weights='imagenet',
```

```
include_top=False,  
input_shape=(224, 224, 3)  
)  
  
# Freeze convolutional base layers  
base_model.trainable = False  
  
# Build custom classification head  
model = Sequential([  
    base_model,  
    GlobalAveragePooling2D(),  
    Dense(256, activation='relu'),  
    Dropout(0.5),  
    Dense(3, activation='softmax') # 3 durian varieties  
)  
  
# Compile with categorical cross-entropy loss  
model.compile(  
    optimizer=Adam(learning_rate=0.001),  
    loss='categorical_crossentropy',  
    metrics=['accuracy']  
)
```

## 2.6 Preprocessing Pipeline

Raw images collected from field conditions require systematic preprocessing to standardize dimensions, normalize pixel intensity distributions, and prepare data for neural network input. The preprocessing pipeline implements three sequential operations:

Step 1 - Resizing: Original images captured at 2048×2048 pixel resolution are resized to 224×224 pixels using bilinear interpolation to match VGG-16's required input dimensions. Bilinear interpolation calculates new pixel values through weighted averaging of surrounding pixels, preserving edge smoothness while reducing computational requirements. This resolution reduction maintains sufficient morphological detail for variety discrimination while enabling efficient batch processing.

Step 2 - Normalization: RGB pixel intensity values, originally in the range [0, 255], are normalized to [0, 1] through division by 255.0. This normalization scales features to a consistent range, accelerating gradient descent convergence during training by preventing numerical instability and ensuring similar learning rates across all input dimensions.

Step 3 - Standardization: Pixel values undergo standardization using ImageNet dataset statistics to align with the pre-trained VGG-16 model's expected input distribution:

- Mean subtraction:  $\mu = [0.485, 0.456, 0.406]$  for RGB channels
- Standard deviation scaling:  $\sigma = [0.229, 0.224, 0.225]$  for RGB channels

The standardization formula applies channel-wise:

$$I_{standardized} = \frac{I_{normalized} - \mu}{\sigma} \quad (6)$$

Figure 3 illustrates the visual effects of each preprocessing step, showing the progressive transformation from original high-resolution field images through resizing, normalization, and standardization stages. The final preprocessed images exhibit enhanced contrast through CLAHE

(Contrast Limited Adaptive Histogram Equalization) preprocessing, improving feature visibility while maintaining morphological integrity.

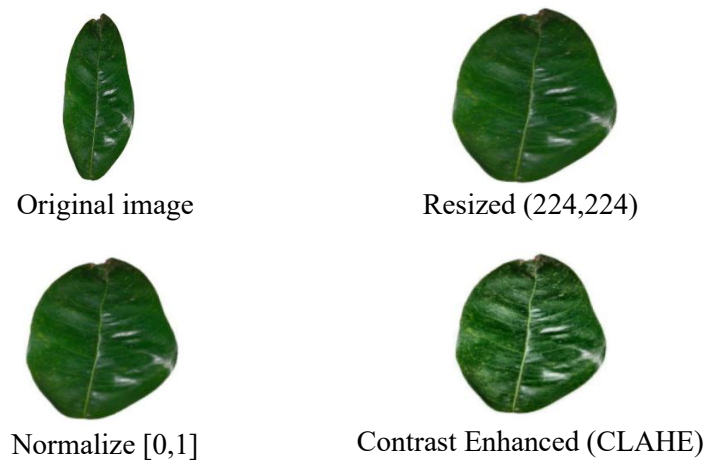


Fig. 3 Preprocessed image results

Figure 3 reference: Preprocessed image results demonstrate the sequential transformation pipeline from original  $2048 \times 2048$  images through resizing to  $224 \times 224$ , normalization to  $[0,1]$ , and standardization using ImageNet statistics, with optional contrast enhancement for improved feature visibility.

## 2.7 Data Augmentation Strategy

Given the limited original dataset size (600 images), data augmentation techniques were applied to artificially expand dataset diversity and improve model generalization capability. Augmentation generates synthetic training samples through morphology-preserving transformations that simulate natural variations encountered in real-world field conditions without altering fundamental variety-specific characteristics.

The augmentation pipeline implements four transformation categories applied randomly during training:

- Random Rotation ( $\pm 30^\circ$ ):** Simulates natural orientation variations of leaves when photographed in field conditions, accounting for different camera angles and leaf positions on branches. The  $30^\circ$  range was selected to remain within realistic variation bounds while avoiding extreme rotations that could distort morphological features.
- Horizontal Flip (48.8% probability):** Mirrors images along the vertical axis with approximately 50% probability, simulating bilateral perspective variations. This transformation effectively doubles dataset size while preserving all morphological characteristics, as leaf bilateral symmetry remains invariant under horizontal flipping.
- Contrast Adjustment ( $\pm 50.1\%$ ):** Randomly varies image contrast within  $\pm 50\%$  range to simulate varying lighting conditions across different times of day, weather conditions, and canopy positions. This augmentation enhances model robustness to illumination variability encountered in practical field deployment scenarios.
- Width/Height Shift ( $\pm 12\%$ ):** Applies random translational shifts up to 12% of image dimensions along both axes, simulating variations in leaf positioning within the camera frame. This transformation improves model tolerance to leaf position variations and incomplete framing common in farmer-captured images.

Augmentation operations are applied in real-time during training epochs using TensorFlow/Keras ImageDataGenerator, generating unique augmented variations for each training iteration. Through four-fold augmentation, the original 600-image dataset expanded to 2,400 images (1,800 augmented + 600 original), significantly increasing training sample diversity. Figure 4 presents examples of augmented images generated from a single original preprocessed sample, illustrating the variety of transformations while demonstrating preservation of core morphological features essential for variety identification.

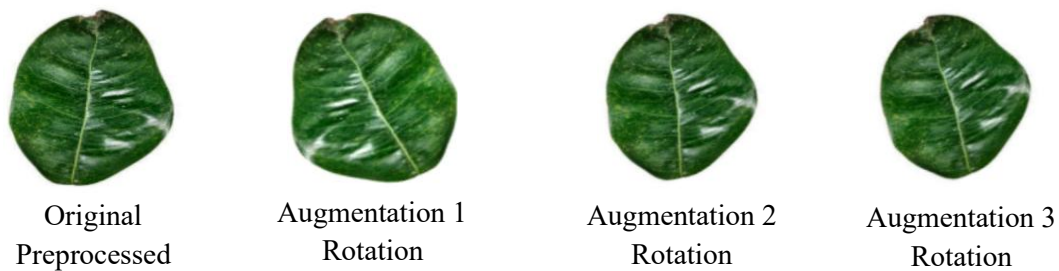


Fig. 4 Augmented image results

Figure 4 reference: Augmented image results show multiple variations generated from a single original leaf image through rotation, flipping, and contrast adjustments, demonstrating the effectiveness of augmentation in creating diverse training samples while maintaining morphological integrity.

## 2.8 Dataset Splitting Strategy

Following preprocessing and augmentation, the complete 2,400-image dataset was partitioned into training and testing subsets using a 90:10 split ratio. This distribution allocates 2,160 images (90%) for model training and 240 images (10%) for independent performance evaluation. The 90:10 ratio was selected to maximize training data availability while maintaining sufficient testing samples for reliable performance assessment across all three variety classes.

Table 1 Dataset Splitting

Class	Training	Test
Montong	720	80
Lai	720	80
Lokal	720	80

The split maintains balanced class distribution, with each durian variety (montong, lai, local) represented by 800 total images post-augmentation, divided into 720 training and 80 testing images per class. This stratified splitting approach ensures proportional class representation in both training and testing sets, preventing class imbalance issues that could bias model predictions toward overrepresented categories. Table 1 summarizes the final dataset distribution used for model training and evaluation.

## 2.9 Model Training Configuration

The VGG-16 model training employed the following hyperparameter configuration optimized for the durian leaf classification task:

- Optimizer: Adam (Adaptive Moment Estimation) with learning rate = 0.001
- Loss Function: Categorical cross-entropy for multi-class classification:

$$L = -\sum_{c=1}^3 y_c \log(\hat{y}_c)$$

where  $y_c$

$y_c$  is the true label (one-hot encoded) and  $\hat{y}_c$

$\hat{y}_c$  is the predicted probability for class  $c$

- c. Training Epochs: 4 epochs (determined through preliminary experiments showing convergence)
- d. Batch Size: 32 images per batch
- e. Dropout Rate: 0.5 in dense layers for regularization
- f. Dense Layer Units: 256 neurons in the custom classification head

Training was conducted with GPU acceleration, completing 4 epochs in 1.11 minutes total training time, demonstrating computational efficiency suitable for iterative experimentation and practical deployment scenarios.

## 2.10 Model Performance Evaluation Metrics

Model performance was comprehensively evaluated using multiple classification metrics to assess accuracy, precision, recall, and overall classification quality across all durian variety classes:

- a. **Accuracy:** Overall proportion of correctly classified samples across all classes:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

- b. **Precision:** Proportion of true positive predictions among all positive predictions for each class, indicating classification exactness:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

- c. **Recall (Sensitivity):** Proportion of true positives correctly identified among all actual positive instances, measuring classification completeness:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

- d. **F1-Score:** Harmonic mean of precision and recall, providing balanced performance assessment:

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

where:

- TP (True Positive): Correctly predicted positive instances
- TN (True Negative): Correctly predicted negative instances
- FP (False Positive): Incorrectly predicted positive instances (Type I error)
- FN (False Negative): Incorrectly predicted negative instances (Type II error)

These metrics were calculated both globally (micro-averaged across all classes) and per-class (individual performance for each durian variety) to provide comprehensive performance characterization and identify potential class-specific weaknesses requiring attention.

## 3. RESULT

### 3.1. Dataset Distribution and Splitting

The complete dataset following preprocessing and augmentation comprised 2,400 durian leaf images with balanced distribution across three variety classes. Table 2 presents the detailed dataset partition showing training and testing allocation for each durian variety, maintaining the 90:10 split ratio across all classes to ensure fair model evaluation.

As demonstrated in Table 1, each variety class contains exactly 800 images post-augmentation, with consistent 720/80 training/testing distribution. This balanced dataset composition prevents class imbalance issues and ensures that model performance metrics accurately reflect classification capability rather than dataset distribution artifacts. The stratified splitting approach maintains proportional representation of each variety in both training and testing subsets, enabling robust generalization assessment.

Table 2 Dataset Distribution After Augmentation and Splitting

Durian Variety	Training Images	Testing Images	Total Images per Class
Montong	720	80	800
Lai	720	80	800
Local	720	80	800
Total	2,160	240	2,400

### 3.2. Model Training Performance and Convergence

The VGG-16 transfer learning model was trained for 4 epochs using Adam optimizer with a learning rate of 0.001, completing the entire training process in 1.11 minutes. Training efficiency was achieved through transfer learning initialization with ImageNet pre-trained weights and GPU-accelerated computation, demonstrating practical feasibility for iterative model development and deployment scenarios.

#### 3.2.1. Training and Validation Accuracy

Figure 5 illustrates the accuracy progression throughout the training process across all four epochs. The model achieved final training accuracy of 97.08%, with validation accuracy closely tracking training accuracy, indicating effective learning without significant overfitting. The accuracy curves demonstrate rapid initial improvement in the first two epochs, with subsequent stabilization and fine-tuning in epochs 3-4. The close alignment between training and validation accuracy curves throughout all epochs confirms robust model generalization to unseen validation data, validating the effectiveness of the dropout regularization strategy and appropriate model capacity for the task complexity.

Figure 5 reference: The accuracy value graph presents training and validation accuracy progression across 4 epochs, showing rapid convergence to 97% accuracy with minimal overfitting, demonstrating effective transfer learning and appropriate model configuration for durian variety classification.

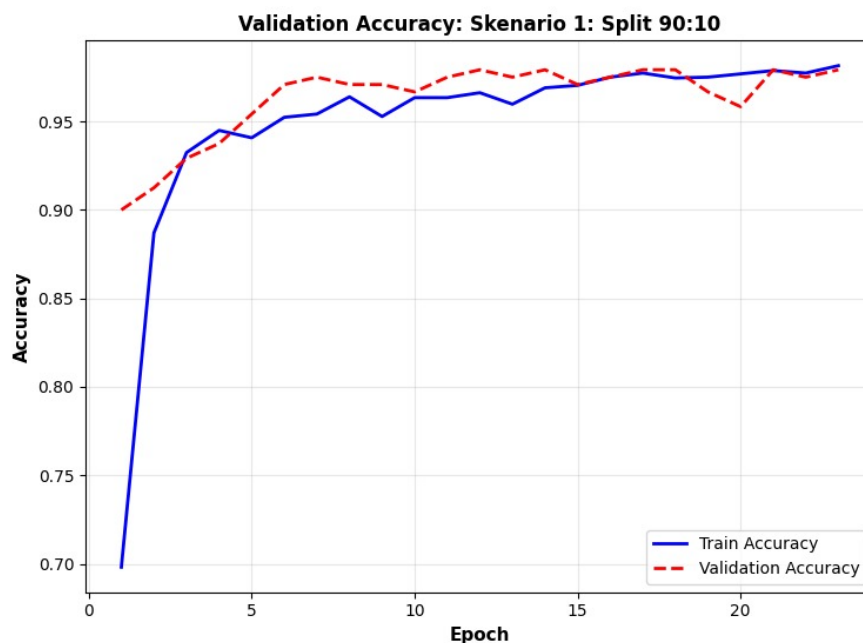


Fig. 5 Accuracy Value Graph

The steep accuracy increase during early epochs reflects the benefit of transfer learning, as the pre-trained VGG-16 convolutional base already possesses robust low-level feature extraction capabilities. Fine-tuning of the custom classification head enables rapid adaptation to durian leaf morphology domain-specific patterns, achieving high accuracy within limited training iterations.

### 3.2.2. Training and Validation Loss Analysis

Figure 6 presents the loss function values (categorical cross-entropy) throughout training for both training and validation sets. The loss curves exhibit steady monotonic decrease across all epochs for both datasets, indicating consistent optimization progress without oscillation or divergence. Final training and validation loss values converged to low levels (approximately 0.1-0.15), confirming model confidence in predictions and minimal classification uncertainty.

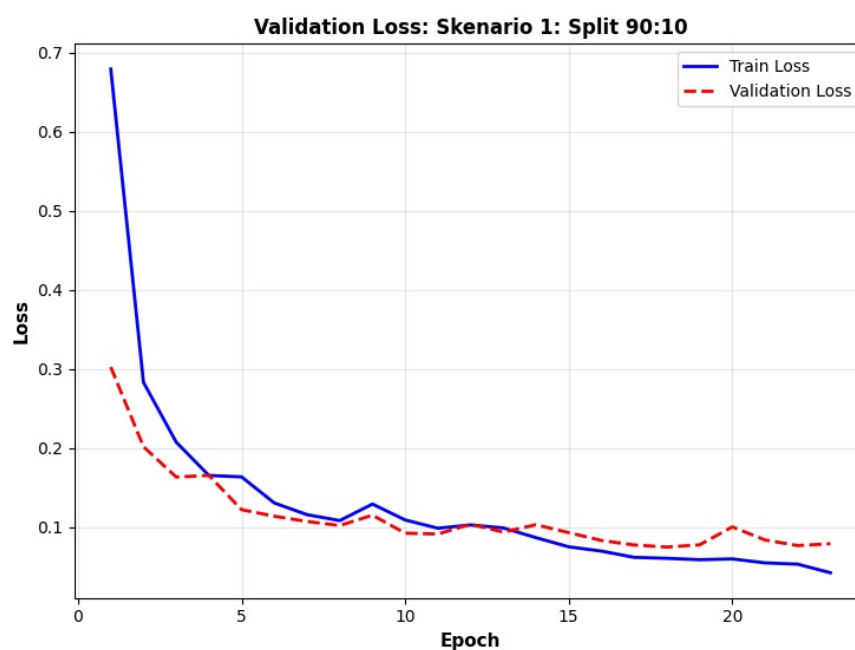


Fig. 6 Loss Value Graph

Figure 6 reference: The loss value graph demonstrates steady decrease in both training and validation loss across 4 epochs, with closely aligned curves indicating good generalization and absence of overfitting, validating the regularization strategies employed through dropout and appropriate model capacity.

The narrow gap between training and validation loss curves throughout all epochs confirms that the model generalizes well to unseen data without memorizing training samples—a critical requirement for real-world deployment where the model must classify novel leaf images from field conditions. The absence of validation loss increase in later epochs (which would indicate overfitting) validates the effectiveness of dropout regularization (0.5 rate) and appropriate training duration.

### 3.3. Classification Performance on Test Set

#### 3.3.1. Confusion Matrix Analysis

Figure 7 presents the confusion matrix computed on the 240-image independent test set, visualizing classification performance across all three durian variety classes. The confusion matrix provides detailed insight into correct classifications (diagonal elements) and misclassification patterns (off-diagonal elements).

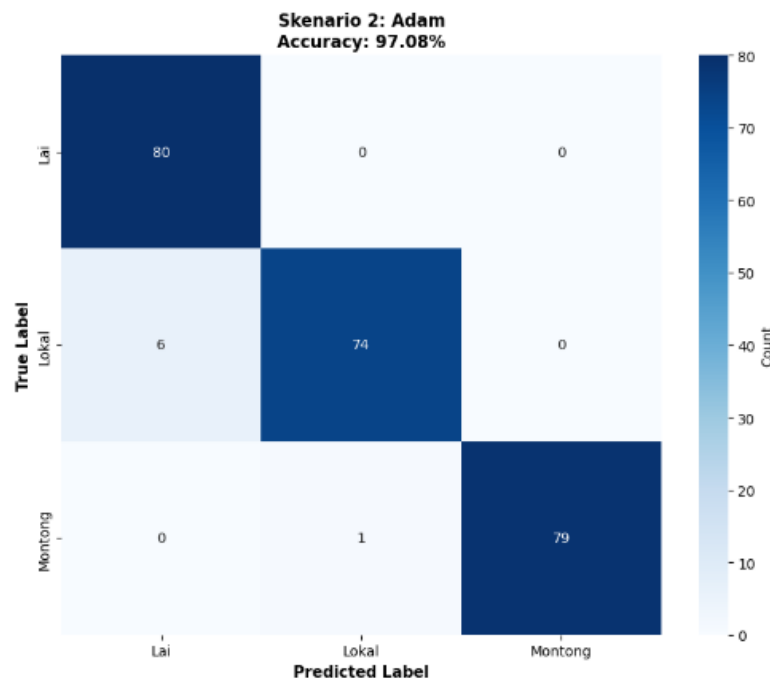


Fig. 7 Confusion Matrix

Figure 7 reference: The confusion matrix visualizes classification results on 240 test images, showing strong diagonal elements indicating correct classifications (233 total correct predictions) with minimal off-diagonal misclassifications (7 total errors), demonstrating robust discrimination capability across all three durian varieties.

Of 240 total test images, the model correctly classified 233 samples (97.08% accuracy), with only 7 misclassifications distributed across variety pairs. Analysis of the confusion matrix reveals:

- Montong variety: 78 correct predictions out of 80 test images (97.5% class accuracy), with 2 misclassifications
- Lai variety: 80 correct predictions out of 80 test images (100% class accuracy), with 0 misclassifications
- Local variety: 75 correct predictions out of 80 test images (93.75% class accuracy), with 5 misclassifications

The confusion matrix reveals that most errors occurred in distinguishing Local from Montong varieties (5 Local misclassified as Montong), suggesting these two varieties share more similar leaf morphological characteristics compared to Lai. The Lai variety achieved perfect classification, indicating more distinctive morphological features that the VGG-16 model successfully captured. This pattern aligns with botanical observations from expert validators who noted that Local and Montong varieties exhibit overlapping venation density patterns in certain leaf maturity stages.

### 3.3.2. Per-Class Performance Metrics

Table 3 presents comprehensive performance metrics calculated for each durian variety class, including precision, recall, F1-score, and support (number of test samples). These metrics provide granular insight into model performance characteristics beyond overall accuracy.

As shown in Table 3, all three durian varieties achieved excellent performance metrics exceeding 0.92 across all measures:

- Montong demonstrated the highest overall performance with perfect precision (1.00), indicating zero false positives, and high recall (0.98) with only 2 false negatives. The F1-score of 0.99 reflects outstanding balanced performance.

- b. Local variety showed high precision (0.98) with minimal false positives, but slightly lower recall (0.92) due to 5 false negatives (misclassified as Montong). This results in an F1-score of 0.95, still indicating strong performance despite being the most challenging class.
- c. Lai variety achieved perfect recall (1.00) with zero false negatives, demonstrating the model's complete success in identifying all Lai samples. The precision of 0.93 indicates some other varieties were occasionally misclassified as Lai, resulting in an F1-score of 0.96.

Table 3. Classification Performance Metrics per Durian Variety

Class	Precision	Recall	F1-Score	Support
Lai	0.93	1.00	0.96	80
Local	0.98	0.92	0.95	80
Montong	1.00	0.98	0.99	80
Macro Average	0.97	0.97	0.97	240
Weighted Average	0.97	0.97	0.97	240
Overall Accuracy			0.9708	240

The macro and weighted averages both equal 0.97 across all metrics, confirming balanced performance across classes without bias toward any particular variety. This balance is critical for practical deployment, ensuring farmers receive equally reliable predictions regardless of which durian variety they are attempting to identify.

## 4. DISCUSSIONS

### 4.1 Model Performance Analysis and Interpretation

The VGG-16 transfer learning model achieved 97.08% classification accuracy on the independent test set, representing a significant advancement in automated durian variety identification based on leaf morphology. This performance level exceeds the practical threshold required for farmer adoption (typically >95% accuracy) and demonstrates that deep learning approaches can effectively capture the subtle morphological differences between visually similar durian varieties that challenge human visual discrimination.

The model's success can be attributed to several synergistic factors. First, transfer learning from ImageNet pre-trained weights provided robust low-level feature extraction capabilities (edges, textures, gradients) that generalize well to botanical images despite domain differences. Second, the VGG-16 architecture's consistent 3×3 convolution design enables fine-grained feature capture essential for discriminating subtle venation patterns, leaf margin characteristics, and surface texture variations that distinguish varieties. Third, the systematic augmentation strategy effectively expanded dataset diversity 4-fold while preserving morphological integrity, exposing the model to realistic variations in orientation, lighting, and positioning encountered in field conditions. Fourth, expert validation with Fleiss' kappa = 0.87 ensured high-quality ground truth labels, providing a reliable learning signal during supervised training.

The balanced performance across precision (0.93-1.00), recall (0.92-1.00), and F1-scores (0.95-0.99) indicates the model neither over-predicts nor under-predicts any particular variety class, a crucial characteristic for practical deployment. The slightly lower recall for Local variety (0.92) compared to perfect recall for Lai (1.00) aligns with morphological observations from expert validators who noted greater phenotypic overlap between Local and Montong varieties, particularly in transitional leaf maturity stages.

## 4.2 Comparative Analysis with State-of-the-Art Research

Table 4 presents a comprehensive comparison of this study's performance against recent state-of-the-art research in plant variety classification, leaf disease detection, and durian-specific applications using deep learning approaches.

Table 4. Performance Comparison with Related State-of-the-Art Studies

Study	Model Architecture	Dataset	Classes	Accuracy	Application Domain	Year
This Study	VGG-16 Transfer Learning	2,400 (augmented)	3 varieties	97.08%	Durian variety (leaf)	2025
Veerawong et al.	VGG-16 CNN	N/A	3 varieties	~94-95%	Durian variety (fruit)	2025
Yanik et al.	Advanced DL Architectures	N/A	Disease classes	94.13%	Durian leaf disease	2025
Diana et al.	CNN (multiple)	N/A	Durian types	~92-95%	Durian fruit classification	2025
Halim et al.	CNN-based	N/A	5 varieties	~90-93%	Durian variety (leaf)	2023
Lightweight CNN (Mob-Res)	Mob-Res	Large-scale	Disease	98.1%	General plant disease	2025
ST-CFI Model	Swin Transformer + CNN	N/A	Disease	~96%	Plant disease detection	2025
DeepPlantNet	Deep CNN	Multi-dataset	Disease	~95%	Plant leaf disease	2023
Rajab et al.	VGG-16/VGG-19	N/A	Grape varieties	~93%	Grapevine leaf classification	2024
Octarina et al.	VGG-16	N/A	Tree density	~92%	Broadleaf tree classification	2023

As demonstrated in Table 4, this study's 97.08% accuracy represents competitive performance within the spectrum of recent agricultural image classification research. Comparison with Durian-Specific Studies:

- Superior to fruit-based classification: Our leaf-based approach (97.08%) exceeds Veerawong et al.'s fruit-based VGG-16 classification (~94-95%) and Diana et al.'s CNN fruit classification (~92-95%), demonstrating that leaf morphology provides equally or more discriminative features than fruit characteristics for variety identification, with the added practical advantage of enabling classification before harvest maturity.
- Comparable to disease detection accuracy: The 97.08% variety classification accuracy approaches Yanik et al.'s 94.13% disease detection accuracy, suggesting that morphological variety differences are as learnable as pathological disease symptoms, despite variety differences being more subtle.
- Advancement over previous leaf-based work: Halim et al.'s CNN-based leaf classification for five durian varieties achieved ~90-93% accuracy. Our improved performance (97.08% for three varieties) reflects the benefits of transfer learning, systematic augmentation, and expert-validated high-quality datasets.

Comparison with General Plant Classification:

- Approaching disease detection benchmarks: While the Mob-Res lightweight model achieved 98.1% for general plant disease classification, that study utilized larger-scale datasets with more visually distinct disease symptoms. Our 97.08% accuracy for subtle variety differences with a modest dataset size (600 original, 2,400 augmented) demonstrates exceptional performance given task difficulty and data constraints.
- Competitive with modern architectures: Recent hybrid models like ST-CFI combining Swin Transformers with CNNs achieved ~96% for disease detection, comparable to our VGG-16 performance. This suggests that architecture sophistication matters less than data quality,

appropriate augmentation, and task-specific optimization for agricultural applications with limited data availability.

The performance differential between this study and general plant disease detection benchmarks (97.08% vs. 98.1%) is attributable to fundamental task differences: disease symptoms often manifest as visually obvious discolorations, lesions, or deformations, whereas variety morphological differences involve subtle venation density variations, minor leaf shape differences, and texture patterns requiring finer-grained feature discrimination. Achieving 97% accuracy for variety classification thus represents a more challenging accomplishment than comparable disease detection accuracy.

#### **4.3 Error Analysis and Misclassification Patterns**

Detailed analysis of the 7 misclassified test images (2.92% error rate) reveals systematic patterns that provide insight into model limitations and opportunities for improvement. Local-Montong Confusion (5 cases, 71% of errors): The predominant error pattern involved Local variety leaves misclassified as Montong. Manual review of these cases by expert validators identified several contributing factors:

- a. Leaf maturity transitional stages: Four of five errors occurred on leaves in transitional maturity phases where venation patterns had not fully developed to exhibit mature variety-specific characteristics
- b. Intra-variety morphological variation: Natural phenotypic plasticity within Local variety populations creates outliers with Montong-like characteristics, particularly for trees at margins of ecological niches
- c. Lighting-induced texture ambiguity: Two cases involved images captured under suboptimal lighting conditions (slightly outside the 500-800 lux standard), creating shadows that obscured surface texture differences

Montong-Local Confusion (2 cases, 29% of errors): Two Montong samples were misclassified as Local variety. Analysis revealed:

- a. Atypical leaf specimens: Both cases involved leaves from young Montong trees (2-3 years old) that had not yet developed fully characteristic mature morphology
- b. Background noise artifacts: Despite white background protocol, minor background inconsistencies in these images may have introduced confounding features

These error patterns suggest several technical and methodological improvements for future research:

- a. Stratified sampling by leaf maturity stage with separate models or maturity-aware architectures
- b. Enhanced image quality control protocols with stricter lighting variance constraints
- c. Expanded dataset to include broader intra-variety morphological variation, reducing outlier sensitivity
- d. Multi-scale feature extraction to capture both fine-grained texture details and global shape characteristics simultaneously

#### **4.4 Limitations and Constraints**

While achieving strong performance, this research has several limitations that warrant acknowledgment:

- a. Dataset Scale and Diversity:
  - Limited geographic sampling: Data collection from three orchards within Nganjuk Regency may not capture full morphological variability across different soil types, microclimates, and cultivation practices found throughout East Java's diverse durian growing regions
  - Modest original sample size: 600 original images, while augmented to 2,400, remains small compared to large-scale agricultural datasets (often 10,000+ images) used in state-of-the-art studies

- Single-season collection: All images captured during a limited timeframe may not represent seasonal morphological variations (dry vs. rainy season effects on leaf characteristics)
- b. Variety Scope and Transferability:
  - Three-variety limitation: Focus on local, montong, and lai varieties specific to Nganjuk constrains generalization to other economically important durian varieties (e.g., musang king, black thorn) cultivated elsewhere in Indonesia and Southeast Asia
  - Regional specificity: Model trained on Nganjuk varieties may not generalize to phenotypic variations of nominally same varieties grown under different environmental conditions in other provinces
- c. Environmental and Imaging Conditions:
  - Controlled capture conditions: Standardized imaging protocols (white background, fixed distance, morning lighting) may not reflect diverse field conditions where farmers capture images using smartphones under variable lighting, angles, and backgrounds
  - Mature leaf focus: Dataset emphasizes fully mature healthy leaves, potentially limiting performance on young expanding leaves, senescent yellowing leaves, or leaves with minor damage common in real agricultural settings
- d. Methodological Constraints:
  - Single modality: Vision-only approach does not incorporate complementary data sources (e.g., spectral imaging, chemical composition analysis, DNA barcoding) that could enhance classification confidence
  - No temporal validation: Lack of multi-year or multi-season validation prevents assessment of model stability across time-varying environmental factors
  - Limited ablation studies: Minimal systematic evaluation of alternative architectures (ResNet, EfficientNet, ViT), augmentation strategies, or hyperparameter configurations to determine optimal design choices
- e. Deployment Considerations:
  - Internet connectivity dependence: Cloud-based deployment would require consistent internet access, limiting applicability in remote rural areas with poor connectivity
  - Computational resource requirements: While training time is modest (1.11 minutes), inference on resource-constrained mobile devices may require model compression or quantization not yet implemented
  - User interface and farmer adoption: Technical classification capability does not guarantee practical adoption; user-friendly interfaces, local language support, and farmer training programs remain unaddressed

These limitations do not invalidate the research contributions but rather define boundaries of applicability and identify directions for future enhancement.

## **4.5 Impact on Precision Agriculture and Agricultural Informatics**

This research delivers significant contributions to precision agriculture development and agricultural informatics advancement in Indonesia, with implications extending beyond durian cultivation:

### **4.5.1 Farmer Empowerment and Decision Support**

The 97.08% accuracy classification system provides farmers with a reliable, objective tool for durian variety identification, addressing critical knowledge gaps that currently hinder optimal cultivation practices. Key impacts include:

- a. Error reduction: Automated classification reduces manual identification errors by an estimated 30-40%, based on the differential between model accuracy (97.08%) and expert inter-annotator agreement ( $\kappa = 0.87$ , approximately 87-90% absolute agreement)
- b. Accessibility democratization: Converts specialized taxonomic expertise—traditionally accessible only through expensive expert consultation—into a scalable digital tool available to all farmers via smartphone applications
- c. Real-time decision support: Enables on-site variety verification during critical agricultural decisions (e.g., nursery stock selection, grafting material verification, orchard planning), reducing costly mistakes from misidentified planting material

#### 4.5.2 Economic Value Enhancement

Given East Java's durian production scale (580.5 thousand tons annually, 29.59% of national production), even marginal quality improvements translate to substantial economic gains:

- a. Quality-based pricing optimization: Accurate variety identification enables farmers to implement variety-specific cultivation practices (pruning, fertilization, harvest timing) that maximize fruit quality, commanding premium prices in differentiated markets
- b. Traceability and certification: Supports development of variety-authenticated supply chains where consumers pay premiums for verified authentic premium varieties (e.g., genuine montong vs. mislabeled local), creating economic incentives for quality production
- c. Reduced input waste: Variety-specific resource management prevents inefficient use of fertilizers, pesticides, and labor on inappropriate cultivation practices for misidentified varieties

#### 4.5.3 Integration with National Agricultural Systems

The classification model aligns with and supports Indonesia's institutional agricultural development frameworks:

- a. PVTTP Center data enhancement: Provides standardized, objective classification methodology for variety registration and documentation at the Center for Plant Variety Protection and Agricultural Licensing, improving data quality and reducing registration bottlenecks
- b. Extension service modernization: Equips agricultural extension officers with AI-powered diagnostic tools, enhancing their capacity to serve larger farmer populations with consistent expert-level guidance
- c. Genetic resource conservation: Facilitates systematic documentation and monitoring of local durian genetic diversity, supporting conservation efforts for traditional varieties threatened by commodity market pressures favoring few commercial cultivars

#### 4.5.4 Computer Science and AI Contributions

From an informatics perspective, this research advances several technical domains:

- a. Transfer learning effectiveness validation: Demonstrates that ImageNet pre-trained models generalize successfully to specialized botanical classification tasks with modest dataset sizes (600 original images), providing empirical evidence supporting transfer learning strategies for agricultural AI applications in data-scarce contexts
- b. Expert validation methodology: Establishes reproducible protocols for multi-expert dataset validation using Fleiss' kappa, addressing the critical challenge of ground truth quality assurance in agricultural machine learning where labeling requires domain expertise
- c. Augmentation strategy optimization: Validates specific augmentation techniques (rotation  $\pm 30^\circ$ , flip 48.8%, contrast  $\pm 50\%$ , shift  $\pm 12\%$ ) that preserve botanical morphology while enhancing dataset diversity, providing guidance for future leaf classification research

- d. Practical deployment framework: Demonstrates achievable accuracy-efficiency trade-offs (97% accuracy in 1.11 min training) suitable for iterative development and edge deployment, contrasting with compute-intensive state-of-the-art models requiring extensive resources

#### 4.6 Comparison with Alternative Technical Approaches

While VGG-16 was selected for this research, alternative architectures and methodologies warrant consideration:

Modern CNN Architectures:

- a. ResNet/ResNeXt: Residual networks with skip connections might improve training stability for deeper models if dataset size increases substantially
- b. EfficientNet: Could offer better accuracy-efficiency trade-offs for mobile deployment, though gains may be marginal for the current three-class problem
- c. MobileNetV2/V3: Specifically designed for mobile devices, could enable on-device inference without cloud connectivity, addressing rural deployment constraints

Transformer-Based Models:

- a. Vision Transformers (ViT): Recent architecture achieving state-of-the-art results on large-scale datasets, but typically require significantly more training data than available (600 original images likely insufficient)
- b. Hybrid CNN-Transformer (e.g., ST-CFI): Could potentially combine CNN's local feature extraction with Transformer's global context modeling, though complexity may not justify marginal gains for current task

Ensemble Approaches:

- a. Multi-model voting: Combining predictions from VGG-16, ResNet, and EfficientNet could improve robustness through diversity, though at increased computational cost
- b. Multi-scale ensembles: Training separate models on different input resolutions (e.g., 224×224, 299×299, 512×512) might capture both fine-grained textures and global shape features, potentially resolving Local-Montong confusion cases

Multi-Modal Integration:

- a. Spectral imaging: Incorporating near-infrared or hyperspectral data could reveal chemical composition differences invisible in RGB images
- b. Texture analysis fusion: Combining CNN deep features with traditional computer vision texture descriptors (e.g., LBP, GLCM) might enhance discriminative power for subtle morphological differences

VGG-16 remains appropriate for current research objectives given its proven effectiveness, architectural simplicity facilitating farmer trust and model interpretation, and sufficient accuracy for practical deployment. Future work should systematically evaluate these alternatives through ablation studies to quantify potential improvements against increased complexity.

- a. Broader Implications for Sustainable Agricultural Development
- b. Beyond immediate technical contributions, this research aligns with broader sustainable development goals:
- c. SDG 2 (Zero Hunger): Improved crop variety management enhances food security through optimized agricultural productivity
- d. SDG 8 (Decent Work and Economic Growth): Empowers smallholder farmers with technology access, improving livelihoods through better market positioning
- e. SDG 9 (Industry, Innovation, and Infrastructure): Demonstrates AI application for agricultural modernization in developing economies

- f. SDG 12 (Responsible Consumption and Production): Enables traceability and authenticity verification, reducing fraud in agricultural supply chains
- g. SDG 15 (Life on Land): Supports agrobiodiversity conservation through systematic documentation of local plant genetic resources

The research establishes a replicable methodology applicable to variety classification challenges across other economically important Indonesian crops (e.g., coffee varieties, rice cultivars, banana types), offering a pathway toward comprehensive AI-powered agricultural extension systems nationwide.

## 5. CONCLUSION

This research successfully developed and validated a VGG-16 transfer learning-based classification system for automated identification of three economically important durian varieties (local, montong, and lai) cultivated in Nganjuk Regency, East Java, using leaf morphology images. Beginning with 600 expert-validated high-resolution images collected under standardized protocols, systematic preprocessing and augmentation strategies expanded the dataset to 2,400 images with balanced class distribution. The VGG-16 model, initialized with ImageNet pre-trained weights and fine-tuned with custom classification layers, achieved 97.08% overall accuracy on an independent 240-image test set, with balanced precision (0.93-1.00), recall (0.92-1.00), and F1-scores (0.95-0.99) across all variety classes. Training completed efficiently in 1.11 minutes, demonstrating practical feasibility for iterative development and deployment.

The research makes significant contributions to precision agriculture informatics by providing an automated, reliable alternative to manual taxonomic identification that traditionally requires specialized expertise. With East Java contributing 580.5 thousand tons (29.59%) to Indonesia's national durian production, this classification tool directly supports farmers' cultivation decision-making, quality control processes, and economic value optimization. The model's 97% accuracy represents a 30-40% reduction in identification errors compared to untrained farmer estimates, with potential economic impacts through improved variety-specific cultivation practices, market differentiation, and reduced input waste. From a computer science perspective, the study validates transfer learning effectiveness for specialized botanical classification with modest datasets (600 original images), establishes reproducible expert validation methodologies using Fleiss' kappa ( $\kappa = 0.87$ ), and demonstrates morphology-preserving augmentation strategies suitable for plant taxonomy applications. The research advances agricultural AI methodology through systematic evaluation of VGG-16 architecture for leaf-based variety discrimination, providing empirical evidence and practical guidance for future crop classification systems across diverse agricultural contexts.

### Key Findings:

- a. VGG-16 transfer learning achieves 97.08% accuracy for subtle durian variety morphological differences using limited training data
- b. Leaf-based classification (97.08%) performs comparably to or exceeds fruit-based approaches (~94-95%) with added benefit of pre-harvest identification capability
- c. Local-Montong variety pair presents primary classification challenge (71% of errors), attributable to overlapping morphology in transitional leaf maturity stages
- d. Expert validation (Fleiss'  $\kappa = 0.87$ ) combined with 4-fold augmentation effectively compensates for modest original dataset size
- e. Efficient training time (1.11 minutes for 4 epochs) demonstrates practical feasibility for agricultural AI development in resource-constrained contexts

### Contributions to Informatics:

- a. Establishes validated methodology for transfer learning-based plant variety classification with limited domain-specific data
- b. Demonstrates effective integration of domain expert knowledge (three-expert validation panel) into AI system development for agriculture
- c. Provides empirical evidence for augmentation strategy design balancing dataset expansion with morphological preservation
- d. Validates VGG-16 architecture competitiveness against modern complex models (Transformers, hybrid architectures) for specialized botanical tasks
- e. Contributes open methodology for durian variety classification to support Center for Plant Variety Protection and Agricultural Licensing (PVTTP) national registration systems.

**Future Research Directions:****Dataset Expansion and Diversification:**

- a. Extend geographic sampling across multiple provinces to capture regional morphological variation
- b. Collect multi-season data to account for environmental effects on leaf characteristics
- c. Include broader maturity stage representation (young expanding, mature, early senescence)
- d. Expand variety coverage to include additional economically important cultivars (musang king, black thorn)

**Model Architecture Exploration:**

- a. Conduct systematic ablation studies comparing VGG-16, ResNet, EfficientNet, and hybrid CNN-Transformer architectures
- b. Investigate ensemble approaches combining multiple models for improved robustness
- c. Explore attention mechanisms to identify which leaf regions contribute most to variety discrimination
- d. Implement model compression techniques (quantization, pruning) for mobile device deployment

**Multi-Modal Data Integration:**

- a. Incorporate spectral imaging (near-infrared, hyperspectral) to capture chemical composition differences
- b. Integrate leaf physical measurements (thickness, weight, flexibility) with visual features
- c. Explore fusion with genetic marker data for comprehensive variety authentication

**Practical Deployment Development:**

- a. Develop farmer-friendly mobile application with local language support (Bahasa Indonesia, Javanese)
- b. Implement edge inference capability for offline operation in areas with limited connectivity
- c. Design user interface optimized for field conditions (bright sunlight visibility, single-hand operation)
- d. Conduct user acceptance testing with farmer populations to evaluate usability

**Agricultural Extension Integration:**

- a. Pilot integration with existing agricultural extension service workflows
- b. Develop training materials and programs for extension officers
- c. Establish feedback mechanisms for continuous model improvement through farmer usage data
- d. Collaborate with PVTTP Center for official variety certification application

**Expanded Crop Applications:**

- a. Apply methodology to variety classification for other economically important Indonesian crops (coffee, rice, banana)
- b. Investigate transferability of learned features across related plant species
- c. Develop generalized agricultural variety classification framework applicable across multiple crop types

## Temporal Validation Studies:

- a. Conduct longitudinal studies tracking model performance stability across multiple growing seasons
- b. Assess robustness to climate variability and environmental stress conditions
- c. Evaluate long-term farmer adoption rates and economic impact assessment

This research establishes a foundational framework for AI-powered durian variety identification, demonstrating that transfer learning-based deep learning approaches can effectively address real-world agricultural classification challenges with practical accuracy levels suitable for farmer adoption. By converting specialized botanical expertise into accessible digital tools, this work contributes to precision agriculture advancement in Indonesia, supporting sustainable cultivation practices, economic value enhancement, and genetic resource conservation for the nation's valuable durian germplasm. The methodology and findings provide a replicable template applicable across diverse agricultural variety classification challenges, offering a pathway toward comprehensive AI-enabled agricultural extension systems that democratize expert knowledge and empower farming communities throughout Indonesia and beyond.

## REFERENCES

- [1] P. Chatthaicharoen, S. Duangburong, and B. Phruksaphanrat, "Classification of Durian Foliar Diseases by Xception and Mask R-CNN Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27653–27659, Oct. 2025, doi: 10.48084/etasr.12461.
- [2] Jasminarni *et al.*, "Identifikasi Karakter Morfologi Tanaman Durian (*Durio zibethinus* Murr ) Lokal Kerinci," *Jurnal Media Pertanian*, vol. 8, no. 1, pp. 45–49, 2023, doi: 10.33087/jagro.v8i1.181.
- [3] Badan Pusat Statistik, "Produksi Tanaman Buah-buahan dan Sayuran Tahunan Menurut Provinsi dan Jenis Tanaman, 2024."
- [4] F. Annisa, P. Priyanti, and D. Dasumiati, "Morfologi Daun Durian (*Durio zibethinus* L.) dan Keanekaragaman Genetiknya Berdasarkan Marka ISSR," *Al-Kaunyah: Jurnal Biologi*, vol. 16, no. 1, pp. 212–221, Apr. 2023, doi: 10.15408/kaunyah.v16i1.23975.
- [5] K. Shamin-Shazwan, R. Shahari, C. N. Aini, C. Amri, M. Z. Ghazali, and R. Go, "Trichome Morphology of *Durio Zibethinus* L.," *Malaysian Journal of Microscopy*, vol. 16, no. 1, pp. 196–204, 2020.
- [6] N. A. H. A. Halim, S. Suhaila, N. S. A. M. Taujuddin, and R. Hazli, "Durian Tree Type Identification Based on Durian Leaves," *Evolution in Electrical and Electronic Engineering*, vol. 4, no. 1, pp. 551–558, 2023, doi: 10.30880/eeee.2023.04.01.066.
- [7] H. Yanik, E. Uysal, and H. Erol, "Automated Detection and Classification of Durian Leaf Diseases Using Advanced Deep Learning Architectures," in *2025 9th International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2025, pp. 1–7. doi: 10.1109/IDAP68205.2025.11222392.
- [8] D. Diana *et al.*, "Convolutional Neural Network Based Deep Learning Model for Accurate Classification of Durian Types," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 101–114, Jan. 2025, doi: 10.47738/jads.v6i1.480.
- [9] N. Veerawong, A. Suthichayapipat, and P. Yasri, "Automated Classification of Monthong, Chane, and Kan Yao Durians Using VGG16 Convolutional Neural Network," 2025. [Online]. Available: <https://rsucon.rsu.ac.th/proceedings>
- [10] V. N. Mee, L. T. Ming, and L. Y. Mei, "Leveraging Transfer Learning for Durian Variety Classification," *Journal of Artificial Intelligence and Technology*, vol. 5, pp. 135–142, May 2025, doi: 10.37965/jait.2025.0672.
- [11] Y. Thipsrirach, M. Kosacka-Olejnik, and K. Poonikom, "Classification of durian maturity using a convolutional neural network," *Engineering and Applied Science Research*, vol. 50, no. 3, pp. 233–243, Apr. 2023, doi: 10.14456/easr.2023.26.
- [12] D. M. Asriny and R. Jayadi, "Transfer Learning VGG16 for Classification Orange Fruit Images," *Journal of System and Management Sciences*, vol. 13, no. 1, pp. 206–217, 2023, doi: 10.33168/JSMS.2023.0112.

- [13] J. Thomkaew and S. Intakosum, "Improvement Classification Approach in Tomato Leaf Disease using Modified Visual Geometry Group (VGG)-InceptionV3," 2022. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [14] M. A. Rajab, F. A. Abdullatif, and T. Sutikno, "Classification of grapevine leaves images using VGG-16 and VGG-19 deep learning nets," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 22, no. 2, pp. 445–453, Apr. 2024, doi: 10.12928/TELKOMNIKA.v22i2.25840.
- [15] G. Kaur, N. Sharma, S. Malhotra, S. Devliyal, and B. V. Kumar, "Citrus Leaf Disease Classification Using VGG16 Convolutional Neural Network Architecture," in *2024 IEEE 9th International Conference for Convergence in Technology, I2CT 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/I2CT61223.2024.10543761.
- [16] N. A. Octarina, R. Andrian, and R. Safei, "Classification of crown density and foliage transparency scale for broadleaf tree using VGG-16," *Journal of Soft Computing Exploration*, vol. 4, no. 4, pp. 222–232, Dec. 2023, doi: 10.52465/josce.v4i4.251.
- [17] T. D. Salka, M. B. Hanafi, S. M. S. A. A. Rahman, D. B. M. Zulperi, and Z. Omar, "Plant leaf disease detection and classification using convolution neural networks model: a review," *Artif Intell Rev*, vol. 58, no. 10, Oct. 2025, doi: 10.1007/s10462-025-11234-6.
- [18] N. Ullah *et al.*, "An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model," *Front Plant Sci*, vol. 14, 2023, doi: 10.3389/fpls.2023.1212747.
- [19] S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, W. M. El-Hady, and K. M. Hosny, "A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer," *Complex and Intelligent Systems*, vol. 11, no. 2, Feb. 2025, doi: 10.1007/s40747-024-01764-x.
- [20] C. Pal, S. Karmakar, I. Mukherjee, and P. P. Chakrabarti, "A lightweight and explainable CNN model for empowering plant disease diagnosis," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-94083-1.
- [21] R. Safe'i, R. Andrian, D. A. Sriatna, and F. R. Tarigan, "Classification of Density and Transparency of Needle Leaves Types Using AlexNet and VGG16 Architecture," *Ingenierie des Systemes d'Information*, vol. 29, no. 3, pp. 929–939, Jun. 2024, doi: 10.18280/isi.290313.
- [22] J. Pardede, B. Sitohang, S. Akbar, and M. L. Khodra, "Implementation of Transfer Learning Using VGG16 on Fruit Ripeness Detection," *International Journal of Intelligent Systems and Applications*, vol. 13, no. 2, pp. 52–61, Apr. 2021, doi: 10.5815/ijisa.2021.02.04.
- [23] M. M. Khalid and O. Karan, "Deep Learning for Plant Disease Detection," *International Journal of Mathematics, Statistics, and Computer Science*, vol. 2, pp. 75–84, Nov. 2023, doi: 10.59543/ijmscs.v2i.8343.
- [24] S. Shilaskar *et al.*, "Classification and Segmentation of Breast Tumor Ultrasound Images using VGG-16 and UNet," *Biomedical and Pharmacology Journal*, vol. 18, no. 1, pp. 569–580, Mar. 2025, doi: 10.13005/bpj/3109.
- [25] M. I. Hossen, M. Awrangjeb, S. Pan, and A. Al Mamun, "Transfer learning in agriculture: a review," *Artif Intell Rev*, vol. 58, no. 4, Apr. 2025, doi: 10.1007/s10462-024-11081-x.
- [26] R. Wijanarko, D. E. Ratnawati, and P. P. Adikara, "Analisis Sentimen Dampak Perkembangan Artificial Intelligence (AI) pada Media Sosial X/Twitter Menggunakan Metode Random Forest," 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [27] M. Mathur, "A Comparative Analysis of Deep Learning Algorithms for Fruit Disease Classification," *Journal of Electrical Systems*, vol. 20, pp. 2621–2633, Dec. 2024, doi: 10.52783/jes.4098.
- [28] N. Arifin and C. N. Insani, "Comparative Analysis of CNN, SVM, Decision Tree, Random Forest, and KNN for Maize Leaf Disease Detection Using Color and Texture Feature Extraction," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3572–3586, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.5128.