



Journal of Information System Exploration and Research

<https://joiser.shmpublisher.com>

p-ISSN 2964-1160 | e-ISSN 2963-6361



Application of Kratom Leaf Image Dataset for Leaf Health Classification Using Machine Learning-Based Image Augmentation Techniques

Jesy Kurniati¹, Muhammad Faqih Dzulqarnain², Hermanto³

^{1,2,3}D-III Study Program in Information Technology, Politeknik Aisyiyah Pontianak, Pontianak, Indonesia

*Corresponding Author: jesykurniati@students.politapontianak.ac.id

NPP. 6171052A2000001

Article Info

Keywords:

Kratom Leaf,
Image
Augmentation,
ImageDataGenerator,
MobileNetV2,
Machine Learning

Abstract

Identification of kratom (*Mitragyna speciosa*) leaf health in West Kalimantan is still performed visually and subjectively, risking inconsistent assessments. This study applies image augmentation to increase dataset variation without altering the leaves' main characteristics, and measures classification accuracy using a Convolutional Neural Network (CNN) across two classes (healthy and unhealthy). Augmentation was applied online (on-the-fly) using ImageDataGenerator (TensorFlow/Keras), covering random rotation, shifting, shear, zoom, and flipping. Classification used MobileNetV2 transfer learning and was evaluated with Stratified 5-Fold Cross Validation on 100 original images (50 healthy, 50 unhealthy). The combined out-of-fold predictions from the five folds achieved an accuracy of 80.00%, an AUC of 0.852, and macro-average precision, recall, and F1-score of 0.8005, 0.8000, and 0.7999, respectively. The confusion matrix showed that 41 of 50 healthy images and 39 of 50 unhealthy images were correctly classified. The combination of online augmentation and MobileNetV2 transfer learning delivered a fairly good and balanced classification performance, indicating potential to support automatic, consistent, and efficient identification of kratom leaf health.



This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

1. Introduction

Kratom (*Mitragyna speciosa*) is a tropical plant commonly found across Southeast Asia, including Indonesia, and has long been used in traditional medicine [1]. Ethnobotanical studies confirm that communities in West Kalimantan continue to rely on kratom for traditional remedies and possess detailed local knowledge of its uses, reflecting the plant's cultural and medicinal significance [2], [3]. Kratom leaves are known to contain various bioactive compounds that provide benefits such as pain relief and stimulant effects [4], and laboratory studies have further reported antifungal activity and sun-protection potential in kratom leaf extracts [5], [6]. Kratom also holds considerable economic value, particularly in West Kalimantan, one of its main producing regions [7], even as its regulatory status remains a subject of ongoing policy discussion [8].

Rising market demand has made leaf quality an important factor affecting selling price and product competitiveness [9]. Identification of kratom leaf health is currently still carried out mainly through direct visual observation of leaf color, texture, and shape [10]. This method is subjective and depends on the observer's experience, so it risks producing inconsistent assessments. Morphological similarity between leaf classes further complicates the identification process [11]. Identification errors can reduce product quality, which in turn lowers the selling value and export competitiveness of West Kalimantan kratom.

Addressing this identification problem requires a technology-based approach. Digital image processing combined with machine learning offers a way to automatically analyze leaf images and recognize the patterns and characteristics that distinguish healthy from unhealthy leaf conditions.

Convolutional Neural Network (CNN) is a deep learning architecture capable of automatically extracting image features and has been proven effective at recognizing visual characteristics such as texture, color, and pattern in plant leaf images, thereby improving classification accuracy [12].

Previous research has shown that CNN can be applied to classify healthy and diseased plant leaves based on image data [13]; That study applied data augmentation to increase image variation through transformations such as rotation, flipping, and cropping without altering the leaves' main characteristics. Leaf image classification research using machine learning has been widely conducted on various plant types, but its application to kratom remains relatively limited, leaving an opportunity to develop a technology-based classification system that supports more consistent leaf quality assessment. Classification in this study uses transfer learning with the MobileNetV2 architecture, implemented with Python and TensorFlow/Keras, chosen for its lightweight design while still delivering good classification performance without requiring heavy computational resources [11].

Given the high economic value of kratom in West Kalimantan and the still-limited automatic classification systems developed specifically for this plant, a technology-based approach capable of accurately and consistently classifying kratom leaf health conditions is needed. Applying image augmentation within a machine learning framework offers a relevant solution to the problem of limited data while also improving the generalization ability of the classification model. This study aims to (i) apply image augmentation techniques to increase the number and variation of the kratom leaf image dataset without altering the leaves' main characteristics, and (ii) measure the classification accuracy of kratom leaf images using a CNN model across two leaf-health classes, namely healthy and unhealthy.

2. Literature Review

Plant leaf health classification has been widely studied using Convolutional Neural Networks combined with data augmentation to improve model performance, including for rice leaf disease detection [14] and herbal medicinal leaf identification [15]. Ayuni et al. [16] applied the EfficientNet-B3 architecture together with data augmentation to classify rice leaf diseases; their results showed that augmentation enriched the image dataset, allowing the CNN model to learn image characteristics more effectively and achieve more optimal classification performance.

Roslan et al. [17] applied a CNN to classify Malaysian medicinal herb plants using rotation and flip augmentation, and directly compared model performance before and after augmentation. The accuracy they reported increased from 75% to 88%, a rise of 13 percentage points, with precision, recall, and F1-score also improving, showing that augmentation can strengthen a model's ability to recognize plant images.

Agustina et al. [18] used the DenseNet-121 architecture combined with cropping-based augmentation to classify three rice leaf diseases - brown spot, blast, and blight - and also compared performance with and without augmentation, concluding that augmentation effectively improved the model's ability to recognize several types of rice leaf disease simultaneously.

Another study by Fadlullah & Natalia [11] applied the MobileNetV2 architecture to classify images of traditional Dayak Kenyah medicinal leaves. Their results showed that MobileNetV2 was effectively used for leaf image classification with limited computational resources, making this architecture relevant for application in kratom leaf classification research.

A more recent study by Karim et al. [19] applied a lightweight CNN architecture combined with Grad-CAM visualization to classify grape leaf disease on a real-time edge device, demonstrating that compact CNN models can achieve accurate and interpretable leaf-disease classification under constrained computational resources, an approach broadly consistent with the lightweight MobileNetV2 strategy adopted in this study. In a related direction, Hendry et al. [20] combined a hybrid Vision Transformer-ConvNeXt architecture with medical test-time augmentation for skin lesion classification, indicating that pairing augmentation strategies with modern deep learning architectures continues to improve classification performance beyond the plant-leaf domain.

Unlike previous studies that focused on rice and other medicinal plants, this study specifically focuses on classifying the health of kratom (*Mitragyna speciosa*) leaves. The dataset used was built from two sources - the Kaggle platform and direct field image acquisition in Nanga Mahap Subdistrict, Sekadau Regency, West Kalimantan. The training data in each fold was enriched using rotation and flipping augmentation to increase training data variation without altering the leaves' main characteristics. Classification was performed using MobileNetV2 transfer learning on TensorFlow/Keras with Stratified 5-Fold Cross Validation, grouping leaf conditions into two classes: healthy and unhealthy. As summarized in Table 1, previous studies are still dominated by rice and general medicinal plant objects; to date, few studies have specifically built a kratom leaf image dataset while also applying image augmentation for leaf health classification. This study seeks to fill that gap.

NPP. 6171052A2000001

Table 1. Comparison with related research

No.	Author (Year)	Method	Description
1	Ayuni et al. (2023) [16]	CNN (EfficientNet-B3) + data augmentation	Applied data augmentation to enlarge the rice leaf disease image dataset, improving CNN classification performance.
2	Roslan et al. (2023) [17]	CNN + data augmentation (rotation, flip)	Compared CNN performance before/after augmentation on Malaysian medicinal herb plants; accuracy rose from 75% to 88%.
3	Agustina et al. (2024) [18]	CNN (DenseNet-121) + data augmentation (cropping)	Classified three rice leaf diseases and compared performance with and without augmentation.
4	Fadlullah & Natalia (2026) [11]	CNN MobileNetV2	Classified images of traditional Dayak Kenyah medicinal leaves using the MobileNetV2 architecture
5	Karim et al. (2024) [19]	Lightweight CNN + Grad-CAM	Classified grape leaf disease on a real-time edge device using a lightweight CNN with Grad-CAM interpretability
6	This study (2026)	MobileNetV2 transfer learning + online image augmentation	Focuses specifically on kratom leaf health (healthy/unhealthy), built from Kaggle and field data, evaluated with Stratified 5-Fold Cross Validation.

3. Method

This research was carried out through several systematic stages: (1) dataset collection, (2) image preprocessing, (3) image augmentation, (4) image classification, and (5) model evaluation, as illustrated in Figure 1.

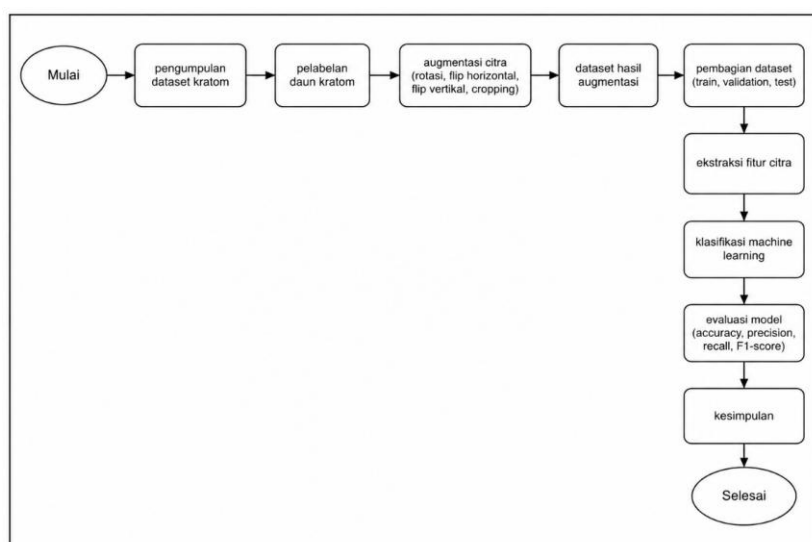


Figure 1. Research flow diagram

Figure 1 illustrates the data processing flow for kratom leaf image classification. The process begins with dataset collection, followed by labeling, after which the dataset is split into training and validation subsets before augmentation is performed, so that augmentation is applied only to the training data. The augmented training data is used to train the CNN MobileNetV2 model through transfer learning, while the validation data (original images) is used to evaluate model performance. Feature extraction is performed automatically by the MobileNetV2 convolutional layers, followed by classification through a fine-tuned dense layer. The resulting model is then evaluated using accuracy, AUC, and confusion matrix metrics before conclusions are drawn.

3.1 Image Dataset

The dataset consists of kratom (*Mitragyna speciosa*) leaf images collected from two sources: the Kaggle platform and direct field acquisition. After a selection process to retain only truly unique and representative images (removing duplicates and near-duplicates), the final dataset used in this study consisted of 100 images, evenly split between 50 healthy leaves and 50 unhealthy leaves, as shown in Table 2. Healthy leaves are characterized by an even, fresh green color, a smooth surface texture without spots or lesions, an intact shape with even edges, and clearly visible light-green to yellowish leaf veins, reflecting optimal growth with sufficient nutrient and water intake. Unhealthy leaves show recognizable visual changes such as yellowing, browning, or paling, black or brown spots caused by fungal or bacterial infection, rough, wrinkled, or perforated surfaces, and physical damage to the edges or middle of the leaf. These visual differences form the basis of the class labels used in this study.

Table 2. Number of kratom leaf images used

Class	Number of Images
Healthy	50
Unhealthy	50
Total	100

This limited dataset size reflects the continuing difficulty of obtaining genuinely valid (non-duplicate) kratom leaf images, which is the main reason image augmentation techniques are needed to enrich the variation of the training data.

3.2 Image Preprocessing

All collected images were preprocessed using Python, including resizing to 224x224 pixels and pixel value normalization using the built-in `preprocess_input` function of the MobileNetV2 architecture, applied when data was loaded from the training and validation folders. This stage ensures the dataset has a uniform size and format and is ready for feature extraction and classification by the CNN.

3.3 Image Augmentation

Image augmentation is an important step for increasing the amount and diversity of training data without needing to collect new field data. Augmentation is applied only after the original dataset is split into training and validation subsets, and only to the training data, so that the validation data still consists of original images the model has never seen, ensuring the evaluation reflects genuine generalization ability.

Augmentation is implemented online (on-the-fly) using the ImageDataGenerator class from TensorFlow/Keras, which generates random transformations directly from the original images each time a training batch is formed, so no augmented image files are written to disk. The techniques applied include random rotation of up to 90 degrees, horizontal/vertical width/height shifts of 10%, shear of 10%, zoom of 15%, and horizontal and vertical flipping. Because each transformation combination is chosen randomly for every training batch, the resulting image variation is dynamic (different at each epoch), rather than a fixed set of variations generated once at the start. This combination enriches variation in angle, position, scale, and orientation while still preserving the leaves' main characteristics and, therefore, their class identity. This choice is consistent with prior findings that carefully tuned augmentation parameters can substantially improve model performance on small datasets [21], that systematic evaluation of augmentation techniques improves CNN-based classification under limited-data conditions [22], and that simple geometric transformations remain effective augmentation strategies even outside the image domain [23].

3.4 Classification Model

Classification uses a transfer learning approach with the MobileNetV2 architecture, pre-trained on ImageNet, as the base model. The final layer of the base model is replaced with a Global Average Pooling layer, dropout, a dense layer (128 neurons, ReLU activation), dropout, and a final dense layer (1 neuron, sigmoid activation) to produce a binary classification (healthy or unhealthy).

Training was carried out in two stages. The first stage (feature extraction) trains the new classification layers with the entire MobileNetV2 base model frozen, using the Adam optimizer with a learning rate of $1e-4$. The second stage (fine-tuning) unfreezes some of the upper layers of the base model and retrains them with a smaller learning rate ($1e-5$) so the model can adapt more specific features to kratom leaf characteristics. Training was repeated five times following the 5-Fold Cross Validation scheme, with the model trained from scratch in each fold. EarlyStopping and ReduceLROnPlateau callbacks were applied to stop training when validation performance stagnated and to automatically reduce the learning rate. After all five folds were completed, the out-of-fold predictions for all 100 images were combined and the optimal classification threshold was determined from the ROC curve. Finally, the model was retrained once more using all 100 original images (100% of the data) to produce the final model used for prediction/deployment.

3.5 Evaluation Method

Given the limited size of the original dataset (100 images), model evaluation uses a Stratified 5-Fold Cross Validation approach, in which each original image serves as validation data exactly once. Model performance is evaluated using accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), calculated on the combined out-of-fold predictions across all original images. The optimal classification threshold is determined from the ROC curve using the Youden J statistic, so that the model's classification decisions align with the discriminative ability shown by the AUC.

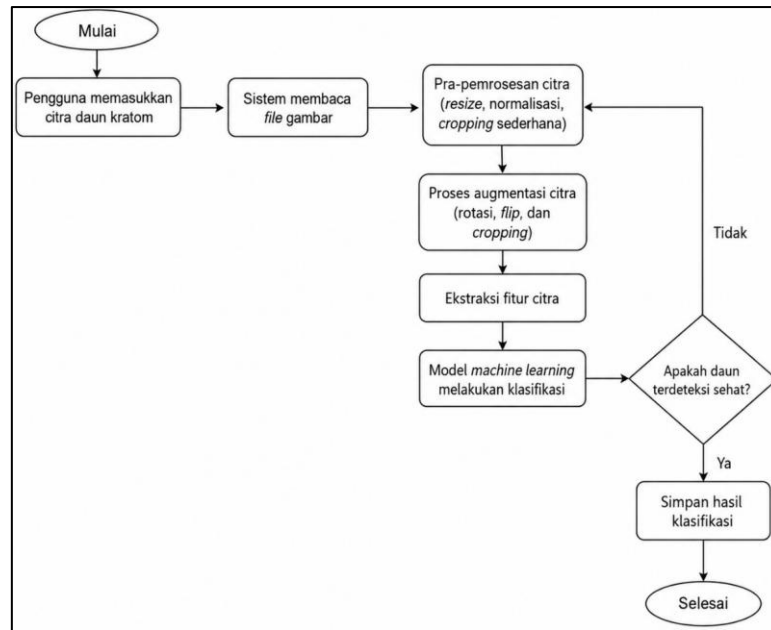


Figure 2. System flow diagram

Figure 2 depicts the system flow diagram centered on a Python/TensorFlow-Keras script as the main engine for processing and analyzing kratom leaf images. Users input leaf images through a command-line interface, after which the images are preprocessed (resized and normalized). The system applies ImageDataGenerator-based augmentation (rotation, width/height shift, shear, zoom, flip) to the training data, then extracts features through the MobileNetV2 convolutional layers before classifying the leaf through the classification dense layer. The classification results, including the model's confidence level, a Grad-CAM visualization highlighting focus areas, detected spot areas, and automatic captions, are displayed to the user through a Matplotlib-based dashboard and can be saved as an image for further analysis.

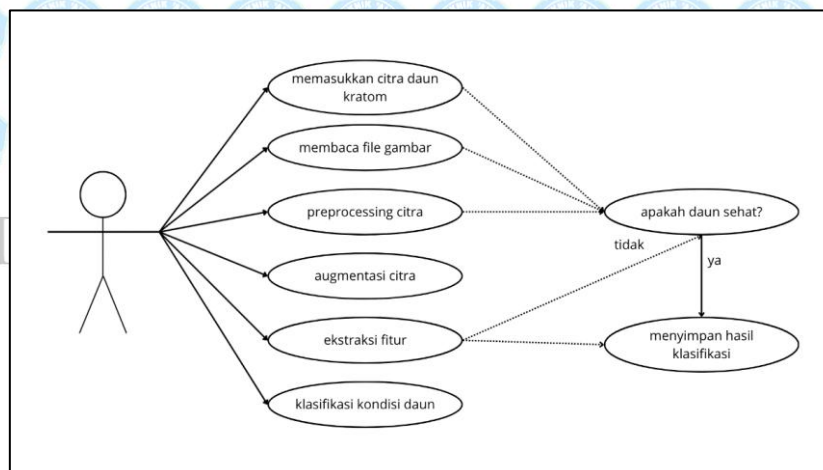


Figure 3. Use case diagram

The use case diagram in Figure 3 represents the interaction between the user (actor) and the system across each processing stage, from image input, preprocessing (resize, normalization, simple cropping), rotation/flip augmentation, and feature extraction, to leaf-condition classification. If the leaf is not detected as healthy, the process can be repeated; if detected as healthy, the classification result is saved as output.

Table 3. Test design

No.	Test Type / Scenario	Expected Result
1	Dataset splitting and augmentation: the system splits the dataset into 5 folds (Stratified 5-Fold CV), then runs	The training data variation of each fold increases without

	online augmentation (rotation, width/height shift, shear, zoom, flip) only on the training data of each fold.	altering leaf characteristics, while the validation data of each fold remains the original images.
2	Image classification: the system classifies kratom leaf images using the CNN MobileNetV2 model.	The system successfully classifies leaf condition (healthy/unhealthy) correctly and produces accuracy, precision, recall, F1-score, and AUC values.
3	System compatibility: the Python script is run on the target computing environment to ensure consistency of data format and pipeline.	The dataset and classification pipeline run without compatibility issues in the environment used.

4. Results and Discussion

4.1 Dataset Collection Results

Kratom leaf images were collected from the Kaggle platform (Figure 4) and from direct field acquisition (Figure 5). After a selection process - including removal of duplicate and near-duplicate images - and labeling, the dataset consisted of 50 healthy images and 50 unhealthy images, for a total of 100 images.

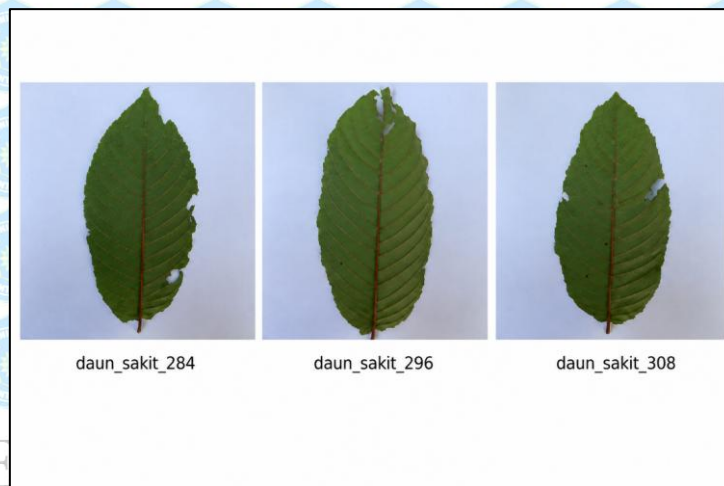


Figure 4. Kratom leaf image dataset from the Kaggle platform

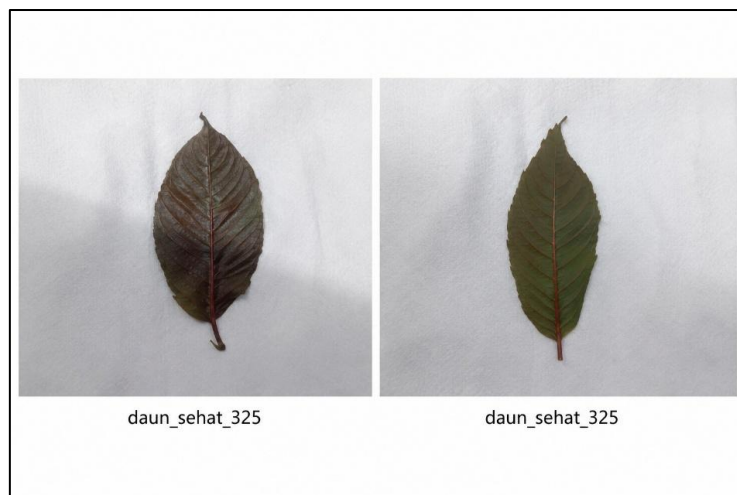


Figure 5. Kratom leaf image dataset from field observation

4.2 Dataset Splitting

Dataset splitting was performed on the original images (before augmentation) using Stratified 5-Fold Cross Validation with a fixed random seed to ensure reproducibility, so that class proportions remained balanced in each fold. Splitting was done before augmentation to prevent data leakage - that is, variations from the same source image spreading into both the training and validation subsets, which would bias the evaluation. In each fold, about 80 images (80%) were used for training and 20 images (20%) for validation, repeated five times with a different training/validation combination each time, so that each of the 100 original images served as validation data exactly once, as summarized in Table 4. This careful splitting strategy is important because prior work has shown that the composition of training/validation splits can significantly affect reported classification performance [24].

Table 4. Dataset splitting

Condition	Total Dataset	Training Data (80%)	Validation Data (20%)
Average per fold (5-Fold CV)	100	80 (online augmentation)	20
Final model (100% data)	100	-	-

4.3 Model Evaluation Results

The MobileNetV2 model, evaluated on the combined out-of-fold predictions from all 100 original images, achieved an accuracy of 80.0% and an AUC of 0.852, with an optimal classification threshold of 0.4571 obtained through ROC curve tuning, as shown in Table 5.

Table 5. MobileNetV2 evaluation results (5-Fold Cross Validation)

Model	AUC	Accuracy	F1-score	Precision / Recall
MobileNetV2 (5-Fold CV)	0.852	0.800	0.800	0.800 / 0.800

The macro-average precision, recall, and F1-score were 0.8005, 0.8000, and 0.7999, respectively, indicating relatively balanced classification performance between the Healthy and Unhealthy classes. Because this result combines the evaluation across the entire available dataset rather than a single training/validation split, it is more representative than an evaluation based on only one data split.

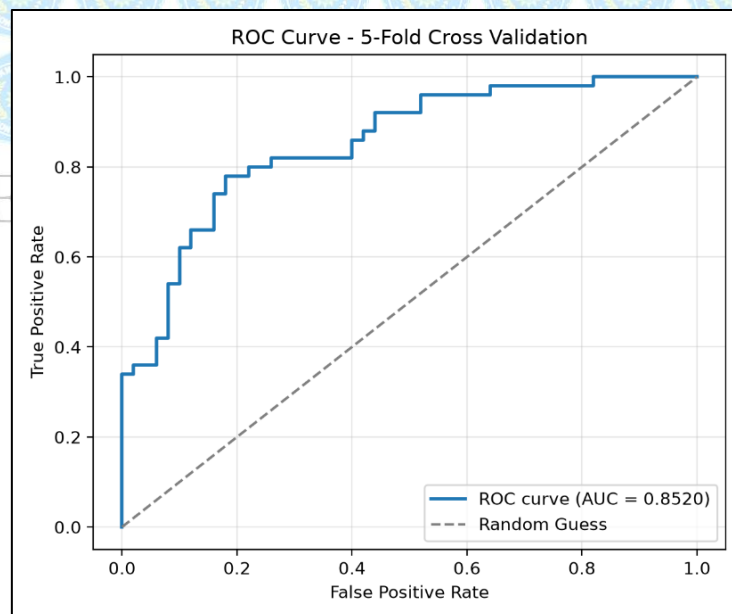


Figure 6. ROC curve of the MobileNetV2 model for classifying healthy and unhealthy kratom leaves

The ROC curve in Figure 6 shows the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) of the MobileNetV2 model in distinguishing healthy from unhealthy kratom leaves. The resulting AUC of 0.852 falls within the 0.7-0.9 range generally associated with good classification ability. The curve rises sharply toward the upper-left corner early in the plot, indicating that the model

achieves a high TPR at a low FPR - good sensitivity in detecting unhealthy leaves without a high false-positive rate on healthy leaves. The diagonal dashed line represents the random-guess baseline (AUC = 0.5); the model curve lying well above it confirms that classification performance is significantly better than random guessing.

Table 6. Classification report

Class	Precision	Recall	F1-score	Support
Healthy	0.7885	0.8200	0.8039	50
Unhealthy	0.8125	0.7800	0.7959	50
Accuracy			0.8000	100
Macro average	0.8005	0.8000	0.7999	100
Weighted average	0.8005	0.8000	0.7999	100

Table 6 shows that precision for the Unhealthy class (0.8125) is higher than for the Healthy class (0.7885), indicating that the model less often mislabels unhealthy leaves as healthy. Recall for the Healthy class (0.8200) is higher than for the Unhealthy class (0.7800), showing that the model is better at finding all genuinely healthy leaves. The F1-scores are fairly close between classes (0.8039 for Healthy and 0.7959 for Unhealthy), and the macro-average metrics confirm this balance, indicating no strong bias toward either class.

4.4 Confusion Matrix

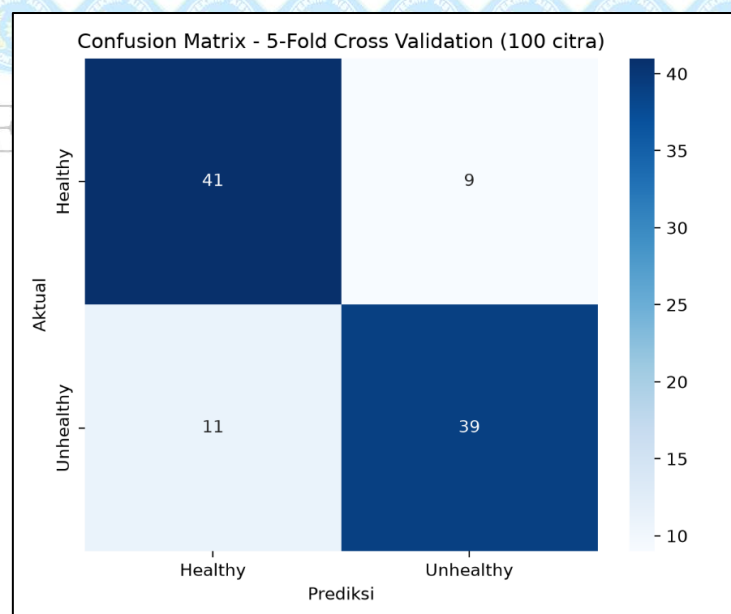


Figure 7. Confusion matrix of the MobileNetV2 model based on the combined 5-Fold Cross Validation predictions

Figure 7 shows the confusion matrix based on the combined out-of-fold predictions across all 100 original images. Of the 50 healthy images, 41 were correctly classified and 9 were misclassified as unhealthy. Of the 50 unhealthy images, 39 were correctly classified and 11 were misclassified as healthy. Classification errors occurred slightly more often in the Unhealthy class (11 of 50), indicating that some unhealthy leaves - particularly those with mild symptoms or in early stages - share visual characteristics with healthy leaves, making them harder for the model to distinguish. The error distribution between the two classes is not strongly biased toward either class: 20 of 100 images (20%) were misclassified overall, with 9 errors in the Healthy class and 11 in the Unhealthy class.

4.5 Grad-CAM Visualization

Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations were generated as a qualitative complement to the quantitative evaluation above, highlighting the image areas that most influenced the model's classification decision. Grad-CAM is not part of the classification process itself, but serves as an interpretation aid.

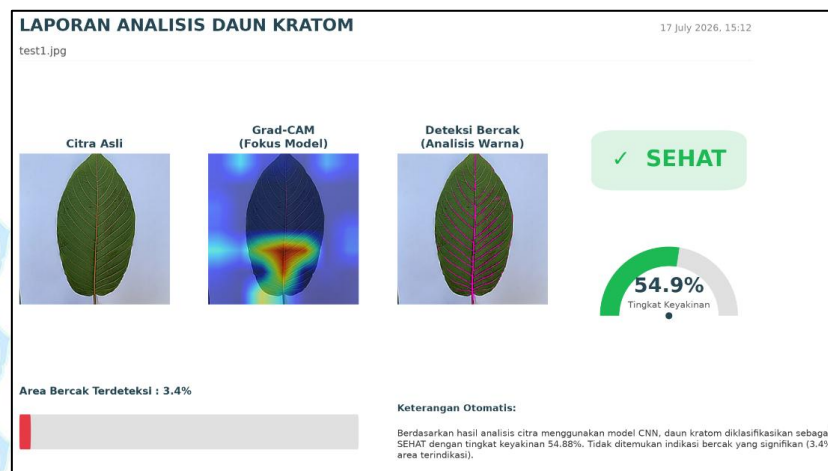


Figure 8. Grad-CAM visualization for an image predicted as healthy (confidence level 54.88%)

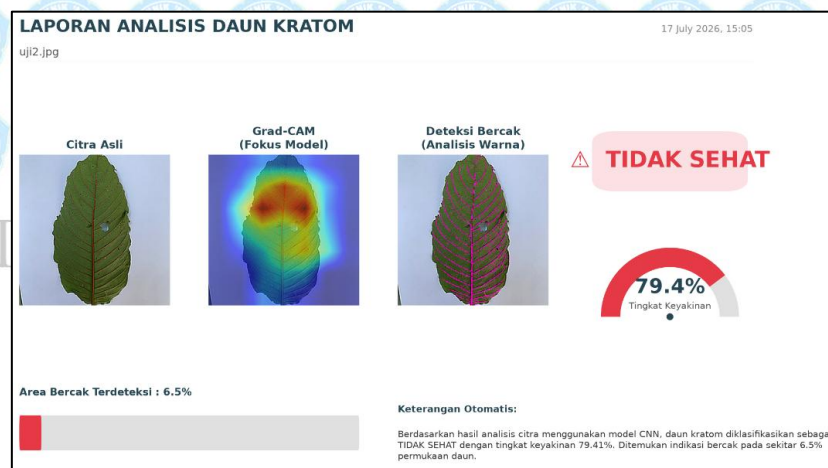


Figure 9. Grad-CAM visualization for an image predicted as unhealthy (confidence level 79.41%)

Figure 8 shows a test image classified as Healthy with a confidence level of 54.88%, where the model's focus is scattered and only slightly concentrated near the leaf base. Figure 9 shows a test image classified as Unhealthy with a confidence level of 79.41%, where the model's focus is more concentrated on a specific area of the leaf surface, indicating that the model assigns greater weight to that area in its decision-making. The difference in focus patterns between these two predictions shows that classification is not performed randomly but is based on specific image areas, complementing the quantitative results in Tables 5 and 6 rather than replacing them.

4.6 Discussion

The original dataset of 100 kratom leaf images, evenly split between the healthy and unhealthy classes, is fairly limited for training a deep-learning-based classification model, which is why online (*on-the-fly*) ImageDataGenerator augmentation was used to enrich the training data variation without needing additional field data collection. MobileNetV2 transfer learning was chosen for its efficient use of general visual features already learned from ImageNet, enabling faster training despite the limited dataset. Training was carried out through feature extraction and fine-tuning stages so the model's features could be adapted more specifically to kratom leaf characteristics.

Splitting the dataset into training and validation subsets before augmentation prevents data leakage, so the validation data still represents images the model has genuinely never seen and the evaluation remains objective. With Stratified 5-Fold Cross Validation, the model achieved an accuracy of 80.00% and an AUC of 0.852 based on the combined predictions for all 100 original images, with relatively balanced performance between the Healthy class (precision 0.7885, recall 0.8200, F1-score 0.8039) and the Unhealthy class (precision 0.8125, recall 0.7800, F1-score 0.7959). The optimal classification threshold of 0.4571, obtained through the Youden J statistic on the ROC curve in place of the default value of 0.5, helps balance the model's sensitivity and specificity, which is reflected in the balanced per-class performance in Table 6.

Using Stratified 5-Fold Cross Validation provides a more representative evaluation than a single training/validation split, because the reported results combine predictions across the entire available dataset rather than only a small held-out portion (e.g., 20%); as a result, misclassification of one or two images will not disproportionately change the reported accuracy, as can happen when the validation data is very small. Overall, the combination of online image augmentation, MobileNetV2 transfer learning, 5-Fold Cross Validation, and threshold tuning produced a kratom leaf health classification model with fairly good and class-balanced performance. The relatively higher error rate in the Unhealthy class suggests that future research should increase image variation for that class, particularly for early-stage disease symptoms.

The 80.00% accuracy achieved in this study is comparable to the findings of Roslan et al. [17], who reported an accuracy increase from 75% to 88% after augmentation on Malaysian medicinal plants. Differences in the research object (kratom vs. Malaysian medicinal plants) and dataset size (100 vs. 64 images) are factors influencing the difference in accuracy achieved. Agustina et al.'s [18] study on rice leaf disease with DenseNet-121 also demonstrated the effectiveness of augmentation in improving accuracy, indicating that image augmentation is a generally effective strategy for improving the performance of leaf image classification models. Similar transfer-learning approaches have also proven effective in other image classification domains beyond plant leaves, such as diabetic retinopathy detection from fundus images [25] and external wound-type classification using pre-trained CNN models [26], further supporting the suitability of transfer learning combined with augmentation for image classification under limited-data conditions.

5. Conclusion

Online image augmentation (*on-the-fly*) using ImageDataGenerator from TensorFlow/Keras - including random rotation of up to 90 degrees, horizontal/vertical shifting, shear, zoom, and horizontal/vertical flipping - was successfully applied to the training data of each fold, dynamically enriching training image variation (around 80 images per fold) at every training batch without altering the leaves' main characteristics and without writing augmented image files to disk. Augmentation was applied only after the training/validation split within the Stratified 5-Fold Cross Validation scheme, thereby preventing data leakage.

The MobileNetV2 transfer learning classification model successfully distinguished healthy from unhealthy kratom leaves, achieving, based on the combined out-of-fold 5-Fold Cross Validation predictions across all 100 original images, an accuracy of 80.00%, an AUC of 0.852, and macro-average precision, recall, and F1-score of 0.8005, 0.8000, and 0.7999, respectively. Applying the optimal classification threshold (0.4571) obtained from the ROC curve helped balance classification performance between the Healthy and Unhealthy classes.

The approach developed in this study shows potential as a basis for an automatic kratom leaf health detection tool. The use of Stratified 5-Fold Cross Validation provides a more representative evaluation than a single data split, since it covers all 100 available original images. Nevertheless, given

the still-limited overall dataset size, further research with a larger dataset is still needed before this approach can be fully relied upon to support leaf quality selection for kratom farmers and industry practitioners. Future research is recommended to (i) collect more original images, particularly unhealthy leaves with more varied symptoms - including early stages - to improve the model's real-world generalization ability; (ii) develop the two-class scheme into a more specific classification based on disease type or damage severity; and (iii) develop this classification system into a web- or mobile-based application so it can be used directly by kratom farmers in the field.

Acknowledgements

The authors would like to thank the D-III Information Technology Study Program, Politeknik Aisyiyah Pontianak, and the kratom farmers in Nanga Mahap Subdistrict, Sekadau Regency, West Kalimantan, for their support and cooperation during the field data collection process.

Credit Authorship Contribution Statement

Jesy Kurniati: Conceptualization, Methodology, Software, Investigation, Data Curation, Writing - Original Draft. Muhammad Faqih Dzulqarnain: Supervision, Validation, Writing - Review & Editing. Hermanto: Supervision, Validation, Writing - Review & Editing.

Declaration of Competing Interests

The authors declare that they have no financial conflicts of interest or personal relationships that could have influenced the research results reported in this article.

Data Availability

The kratom leaf image dataset and analysis scripts used in this study are available from the corresponding author upon reasonable request.

Use of Artificial Intelligence

In preparing this manuscript, the authors used an AI-based writing assistant (Claude, Anthropic) to help restructure and translate the content of the authors' original final-year project report into this journal article format, and to assist with reference formatting. All research design, data, experiments, results, and conclusions are the authors' original work. The authors have reviewed and edited all AI-assisted text and take full responsibility for the content of this publication.

References

- [1] W. A. Pratama, "Regulasi terhadap penggunaan kratom (*Mitragyna speciosa*) sebagai narkotika jenis baru," Undergraduate Thesis, Universitas Islam Indonesia, 2023.
- [2] R. Herawatiningsih, R. Sabtiani, and T. G. Eva, "Etnobotani tanaman kratom (*Mitragyna speciosa*) oleh masyarakat di sekitar DAS Labian di Desa Labian Ira'ang Kecamatan Batang Lupar Kabupaten Kapuas Hulu," *J. Hutan Lestari*, vol. 12, no. 1, pp. 1-12, 2024.
- [3] R. Syarma, S. M. Kartikawati, and D. Setyawati, "Karakteristik dan pengetahuan masyarakat Desa Entibab tentang pemanfaatan tumbuhan kratom (*Mitragyna speciosa*) di Kabupaten Kapuas Hulu," Undergraduate Thesis, 2023.
- [4] E. Prevete, K. P. C. Kuypers, E. L. Theunissen, G. Esposito, J. G. Ramaekers, M. Pasquini, and O. Corazza, "Clinical implications of kratom (*Mitragyna speciosa*) use: A literature review," *Curr. Addict. Rep.*, vol. 10, no. 2, pp. 317-334, 2023, doi: 10.1007/s40429-023-00478-3.
- [5] Normaidah, M. Najahidin, M. Rahmah, Fadlilaturrhamah, and H. Izma, "Uji Sun Protection Factor (SPF) Ekstrak Etanol Daun *Mitragyna speciosa* Korth," *J. Pharmascience*, vol. 10, no. 2, pp. 386-393, 2023.
- [6] D. Susanto, D. Natalia, and A. Widiyantoro, "Aktivitas antijamur ekstrak etanol daun kratom (*Mitragyna speciosa*) terhadap jamur *Microsporum canis*," Undergraduate Thesis, 2023.
- [7] F. F. Rabbani, M. D. I. D. Arta, A. Rofiq, and F. R. Mukhsin, "E-Commerce dan Legalitas Kratom: Implikasi Hukum bagi Pelaku Usaha di Pasar Domestik dan Internasional," *Indones. J. Soc. Sci. Humanit.*, vol. 4, no. 2, pp. 65-74, 2024.

- [8] M. Z. Muttaqien, "Pandangan BNN terhadap penggunaan tanaman kratom pada komunitas Dayak di Kalimantan Barat," Undergraduate Thesis, 2024.
- [9] I. Supriadi, M. Mashudi, and J. H. Matsum, "Analisis Ekonomi Keluarga Petani Kratom (*Mitragyna speciosa*) Desa Nanga Embaloh Kecamatan Embaloh Hilir Kabupaten Kapuas Hulu," *J. Pendidik. Pembelajaran Khatulistiwa*, vol. 12, no. 1, pp. 180-192, 2023, doi: 10.26418/jppk.v12i1.61721.
- [10] F. Kurniawati, D. Fitria, A. R. U. Nainggolan, and P. A. Wardani, "Identifikasi dan karakterisasi tanaman kratom melalui pendekatan profil kandungan senyawa penanda secara LC-HRMS QToF dan penetapan nilai retention index secara GCMS," *Eruditio: Indones. J. Food Drug Saf.*, vol. 3, no. 1, pp. 79-90, 2023, doi: 10.54384/eruditio.v3i1.141.
- [11] A. Fadlullah and T. Natalia, "Sistem klasifikasi citra daun obat tradisional Dayak Kenyah menggunakan metode convolutional neural network dengan arsitektur MobileNetV2," *Infomatek*, vol. 28, no. 1, pp. 121-128, 2026, doi: 10.23969/infomatek.v28i1.36796.
- [12] M. A. Syaharani, T. A. C. Budianto, and R. I. Adam, "Klasifikasi buah segar dan busuk menggunakan algoritma convolutional neural network (CNN)," *JATI (J. Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 10823-10827, 2024, doi: 10.36040/jati.v8i5.11132.
- [13] S. N. Nugraha, R. Pebrianto, and E. Fitri, "Penerapan Deep Learning Pada Klasifikasi Tanaman Paprika Berdasarkan Citra Daun Menggunakan Metode CNN," *Inf. Syst. Educ. Prof. J. Inf. Syst.*, vol. 8, no. 2, pp. 133-142, 2023, doi: 10.51211/isbi.v8i2.2671.
- [14] S. Sheila, I. Permata Sari, A. Bagas Saputra, M. Kharil Anwar, and F. Restu Pujianto, "Deteksi penyakit pada daun padi berbasis pengolahan citra menggunakan metode convolutional neural network (CNN)," *Multinetics*, vol. 9, no. 1, pp. 27-34, 2023, doi: 10.32722/multinetics.v9i1.5255.
- [15] H. K. Zulfiani, Sucipto, and A. Abdullah, "Identifikasi daun herbal untuk tanaman obat menggunakan metode convolutional neural network (CNN)," *JATI (J. Mhs. Tek. Inform.)*, vol. 9, no. 6, pp. 10429-10435, 2025, doi: 10.36040/jati.v9i6.16481.
- [16] D. Putri Ayuni, Jasril, M. Irsyad, F. Yanto, and S. Sanjaya, "Augmentasi data pada implementasi convolutional neural network arsitektur EfficientNet-B3 untuk klasifikasi penyakit daun padi," *ZONasi: J. Sist. Inf.*, vol. 5, no. 2, pp. 239-249, 2023, doi: 10.31849/zn.v5i2.13874.
- [17] N. A. M. Roslan, N. M. Diah, Z. Ibrahim, Y. Munarko, and A. E. Minarno, "Automatic plant recognition using convolutional neural network on Malaysian medicinal herbs: The value of data augmentation," *Int. J. Adv. Intell. Inform.*, vol. 9, no. 1, pp. 136-147, 2023, doi: 10.26555/ijain.v9i1.1076.
- [18] A. Agustina, F. Yanto, E. Budianita, I. Iskandar, and F. Syafria, "Klasifikasi penyakit tanaman padi menggunakan metode CNN arsitektur DenseNet-121 dan augmentasi data," *JOISIE (J. Inf. Syst. Informatics Eng.)*, vol. 8, no. 1, pp. 124-134, 2024, doi: 10.35145/joisie.v8i1.4256.
- [19] M. J. Karim, M. O. F. Goni, M. Nahiduzzaman, M. Ahsan, J. Haider, and M. Kowalski, "Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-66989-9.
- [20] Hendry, F. Govert Anwar, D. Chow, A. Saputra, and M. Khaerul Naim Mursalin, "Arsitektur Hybrid Vision Transformer-ConvNeXt dengan Multi-Task Focal Loss dan Medical Test-Time Augmentation untuk Klasifikasi Lesi Kulit Berbasis Citra," *J. Digit. Ecosyst. Nat. Sustain.*, vol. 5, no. 2, pp. 55-66, 2025, doi: 10.63643/jodens.v5i2.325.
- [21] M. H. Putri, U. Zaky, and B. A. Prabawa, "Optimizing Data Augmentation Parameters in YOLOv11 for Enhanced Rip Current Detection on Small Datasets from Depok-Parangtritis Coastline," *J. Tek. Inform. (JUTIF)*, vol. 6, no. 5, 2025, doi: 10.52436/1.jutif.6.5.5352.
- [22] D. Husen, "Evaluasi teknik augmentasi data untuk klasifikasi tumor otak menggunakan CNN pada citra MRI," Undergraduate Thesis, 2024.
- [23] Q. M. M. Zarandah, S. Mohd Daud, and S. S. Abu-Naser, "Spectrogram flipping: A new technique for audio augmentation," *J. Theor. Appl. Inf. Technol.*, vol. 15, no. 11, 2023.
- [24] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *J. Sains dan Inform.*, pp. 19-28, 2023, doi: 10.34128/jsi.v9i1.622.
- [25] R. A. Pratiwi, "Pendeteksian penyakit diabetic retinopathy melalui citra fundus menggunakan model CNN InceptionResNet-V2," Undergraduate Thesis, 2025.
- [26] Z. Ramadhanti, "Deteksi dan identifikasi jenis luka luar berdasarkan image feature menggunakan CNN dengan variasi pre-trained model," Undergraduate Thesis, 2024.