Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Mitigating Imbalanced Citrus Disease Image Datasets with Oversampling

Arya Gunawan*[1] , Nana Suarna[2] , Agus Bahtiar[3] , Indra Wiguna Marthanu[4] , Kaslani[5]

[1,2,3] *Department of Informatics Engineering, STMIK IKMI Cirebon*
[4,5] *Accounting Computerization Program, STMIK IKMI Cirebon*
Jl. Perjuangan No. 10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat, Indonesia
aryagunawan605@gmail.com[*1], st_anan@yahoo.com[2], agusbahtiar035@gmail.com[3], i.wiguna322@gmail.com[4], kaskalani@gmail.com[5]*

Abstract

Dataset imbalance is a critical challenge in plant disease image classification because it causes bias towards the majority class. This study evaluates the effectiveness of augmentation-based oversampling techniques on the classification performance of citrus leaf images using the MobileNetV2 architecture. The four leaf disease classes classified include Greening, Fresh, Canker, and Blackspot. The dataset was obtained from a public repository and processed through preprocessing (resize, normalization) and augmentation (rotation, flipping, zoom) stages. The model was trained and tested in two scenarios: baseline (unbalanced data) and mitigation (data balanced through augmentation). The experimental results show that the mitigation approach was able to increase accuracy from 91.92% to 93.94%. The F1-score, precision, and recall values also increased significantly, especially in the minority class. Evaluation using a confusion matrix reinforced the finding that augmentation-based oversampling is effective in reducing classification errors. This study shows that the integration of augmentation techniques and MobileNetV2-based transfer learning can significantly improve classification performance and contribute to the development of early detection systems for plant diseases in precision agriculture.

***Keywords**: image classification; data augmentation; MobileNetV2; citrus leaf disease; oversampling*

1. Introduction

Diseases in horticultural crops, especially citrus, such as Greening, Blackspot, and Canker, are serious threats that impact the quality and quantity of crop yields. These diseases are difficult to detect at an early stage by farmers, often resulting in delayed treatment. Therefore, early detection based on digital images is a potential solution in modern precision agriculture systems[1]. With advances in computer vision and artificial intelligence technology, automatic detection based on leaf images has become a rapidly developing field of research[2] .
One of the main approaches used is image classification based on Convolutional Neural Networks (CNN), which has shown outstanding performance in various domains including medicine, agriculture, and industry[3] . However, the application of CNN in plant disease image classification is not without challenges, one of which is data imbalance between classes. In many cases, data on certain diseases is available in large quantities, while rare diseases have only a few representations, causing the model to be biased towards the majority class[4].
To address this imbalance, various methods have been developed, ranging from statistical data balancing to image augmentation techniques. Augmentation techniques such as rotation, flipping, zooming, brightness shifting, and color transformation have been proven to produce realistic data variations without changing the semantic meaning, as well as helping to improve model generalization[5][6] . In addition to being easier to implement, image transformation-based augmentation is also more suitable than methods such as SMOTE, which works on numerical features rather than directly on the image domain.

On the other hand, developing efficient deep learning models for application in resource-limited environments is also an important focus. In this context, MobileNetV2 is the right choice because its architecture is designed with high computational efficiency without sacrificing classification accuracy[7]. This model is widely adopted in mobile and edge computing applications, including real-time plant disease detection in the field[8].
This study aims to apply an augmentation-based oversampling strategy to address class imbalance in citrus leaf disease image datasets and evaluate its impact on classification performance using MobileNetV2. Two testing scenarios were applied: without augmentation (baseline) and with augmentation in the minority class. The evaluation was conducted using accuracy, precision, recall, and F1-score metrics. With this approach, this study is expected to make a significant contribution to the development of an adaptive, lightweight, and accurate automatic plant disease detection system to support artificial intelligence-based precision agriculture.

2. Research Method

This study used a quantitative experimental approach with a pre-post test design to evaluate the effect of augmentation and oversampling techniques on the performance of citrus leaf disease image classification models. This design allows for a comparison between two scenarios: baseline (using unbalanced data) and mitigation (using augmented data for class balance). By adopting a computational experimental approach, this study aims to evaluate the effectiveness of augmentation in overcoming class imbalance in citrus leaf image datasets with a transfer learning-based MobileNetV2 model.

The research design is exploratory and structured, starting from dataset acquisition, initial data exploration, preprocessing, augmentation for minority class oversampling, MobileNetV2 model training, and model performance evaluation with classification metrics. This process is consistent with the approach in the deep learning literature for image classification, which emphasizes systematic validation of the entire pipeline[1][2].

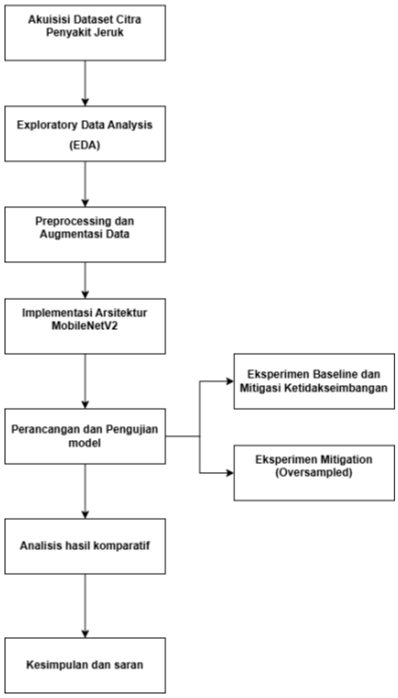

Fig.1. Research Producer

MobileNetV2 was chosen as the architecture due to its efficiency in terms of parameters and performance on devices with limited resources, as well as its adaptability through transfer learning[3] . To address data imbalance, image augmentation such as rotation, flipping, zooming, and brightness changes were applied. Evaluation was performed using accuracy, precision, recall, and F1-score metrics to obtain a comprehensive overview of the model's performance.

1. Research Object

The research object is orange leaf images from four classes: Greening, Fresh, Canker, and Blackspot. The images represent specific visual conditions of orange leaves and are used as classification model inputs. The research focuses on the imbalance in the training data distribution and its impact on classification performance. Augmentation is performed to overcome the dominance of the majority class and enrich the minority class data[3][4][5] .

This dataset reflects real-world conditions with natural backgrounds and visual noise, unlike benchmark datasets such as PlantVillage. This study aims to develop a classification model that is robust and adaptive to realistic data in order to improve the accuracy of citrus leaf disease diagnosis in the real world.

2. Data Collection Methods and Techniques

Data was obtained from the open dataset "Orange diseases dataset"

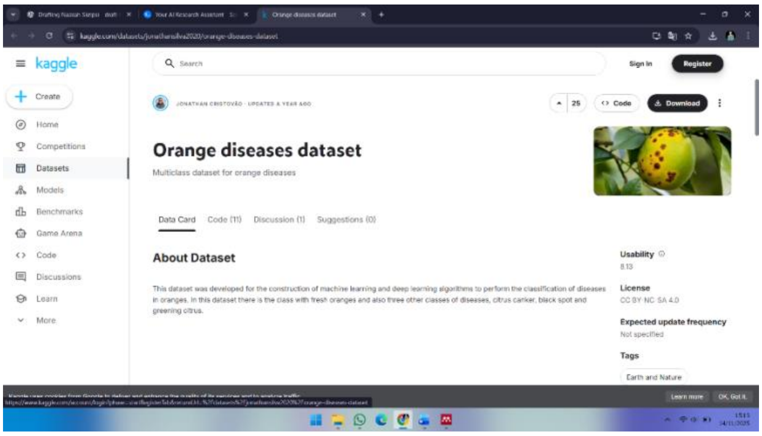

Fig. 1. Orange Diseases Dataset

(https://www.kaggle.com/datasets/saikatdatta1994/tea-leaf-disease), which contains images of citrus leaves with four disease classes. Data was collected through image quality selection that included resolution, feature clarity, and distribution between classes.

3. Research Instruments

The main instrument is Python 3 software, with the Keras and TensorFlow libraries for model training, ImageDataGenerator for augmentation, and Scikit-learn for evaluation. Visualization is performed with Matplotlib and Seaborn. OpenCV is used for image preprocessing (resize, filter, color conversion). Model evaluation is performed with classification metrics and confusion matrix visualization[6][7][8] .

4. Research Procedure

This research procedure began with exploration and analysis of the initial distribution of the orange leaf image dataset. This analysis included an evaluation of the number of images per class, resolution size variation, and overall visual quality. The results of the initial data exploration showed a significant imbalance in the distribution of data between classes. This condition indicated the need to apply an augmentation strategy as a measure to overcome this imbalance. The exploration was carried out using an Exploratory Data Analysis (EDA) approach using the Pandas and Matplotlib libraries in Python, which helped identify class distribution characteristics and reveal the long-tail distribution phenomenon[4] .

The next step is image data preprocessing and augmentation. The pre-processing process includes converting the image size to 224×224 pixels in accordance with the MobileNetV2 architecture input standard, normalizing pixel values to a range of 0–1, and reducing noise by applying Gaussian and median filters. In addition, visual quality is enhanced through adaptive contrast and histogram equalization techniques to improve the readability of important visual features. All of these steps aim to ensure that the data used in training is consistent and of high quality[9][10][11] . Image augmentation is then performed in real-time using the ImageDataGenerator library, with transformation techniques including random rotation, horizontal and vertical flipping, zoom, shear, and brightness adjustment to increase the diversity of visual data[6] .

The next important step is the implementation of physical oversampling techniques on minority classes. Augmentation is selectively applied to the canker and blackspot classes, which previously had less data than other classes. The purpose of this oversampling is to equalize the number of images between classes, while enriching the visual representation of minority classes without causing excessive data duplication. Augmentation is performed dynamically during the model training process, so that the resulting image variants are more diverse. To maintain class distribution balance in the training and testing processes, data partitioning is performed using the Stratified K-Fold Cross Validation method with five folds[12] .

In the final stage, classification model training and evaluation were performed using the MobileNetV2 architecture. This model was equipped with initial weights from ImageNet and modified to complete the task of classifying four classes of citrus leaf diseases. The training process was carried out with a batch size parameter of 32, Adam optimizer, categorical crossentropy loss function, and trained for 15 epochs. Model performance evaluation involved the use of classification metrics such as accuracy, precision, recall, and F1-score, as well as training curve analysis and confusion matrix. The evaluation results showed that the model achieved high accuracy, especially in the fresh and greening classes, while the classification error rate tended to be higher in the canker and blackspot classes, which have more similar visual characteristics. [13]–[15]

5. Data Analysis Techniques

The analysis was conducted by comparing performance metrics between two scenarios (baseline and oversampling). The metrics included precision, recall, and F1-score. Confusion matrix analysis and training curves were used to evaluate the stability and generalization of the model. The results showed higher performance than previous approaches in the literature, reinforcing the effectiveness of the MobileNetV2 approach with augmentation in the classification of citrus leaf disease[4], [13], [16], [17] .

The experimental results provide a strong basis for the implementation of an automatic classification system in the image-based digital agriculture sector, as well as emphasizing the importance of handling imbalanced data distribution in the development of deep learning-based image classification models.

3. Results And Discussion

1. Exploratory Data Analysis (EDA)

Exploratory analysis of the dataset was conducted to identify the initial characteristics of the data, particularly the distribution of the number of images per class. The main focus of EDA was to detect class imbalance, as this is a crucial factor that directly affects model performance. The analysis results show that the distribution is uneven, with the *fresh* and *greening* classes as the majority classes, while *blackspot* and *canker* are minority classes. This imbalance can cause the model to be biased towards the majority class and reduce the model's sensitivity to other classes.

Table1 Overall Data Distribution

class	*test*	*train*	Total
blackspot	22	184	206
canker	22	179	201
fresh	33	295	328
grenning	22	353	375
Total sets	99	1,011	1,110

Based on the table above, the *greening* class has the most images (375 images), followed by *fresh* (328 images). Meanwhile, *blackspot* and *canker* have far fewer images. This imbalance is the main reason for the need for augmentation and oversampling techniques to reduce bias during training.

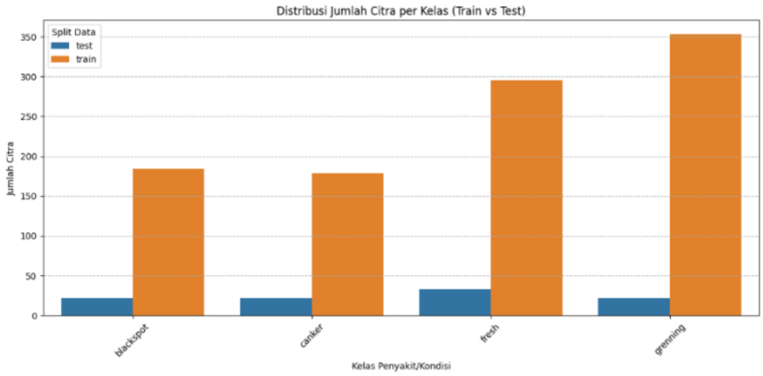

Fig.2 Distribution of Image Counts Per Class (train vs test)

The dataset splitting ratio is:

a) 91.1% training data,

b) 8.9% test data.

This proportion is in line with deep learning best practices for relatively small datasets, so that the model obtains sufficient visual representation to learn before being tested. However, even with an optimal ratio, class imbalance must still be addressed to prevent bias in the model.

2. Research Results

This section presents the results of the entire experimental process, from preprocessing, augmentation, oversampling, and MobileNetV2 model training to the performance evaluation of two scenarios: baseline and mitigation.

a. Acquisition and Distribution of Initial Dataset

The dataset consists of four main classes: *blackspot, canker, fresh,* and *greening*. Before augmentation and oversampling, the dataset showed a clear class imbalance as presented in Table 2

Table3. Class Distribution in the ORIGINAL Training Data (EDA Results)

Class	Count	Percentage
grenning	353	34.92
fresh	295	29.18
blackspot	184	18.2
canker	179	17.71

This imbalance confirms the need for augmentation techniques to increase the number of minority class samples.

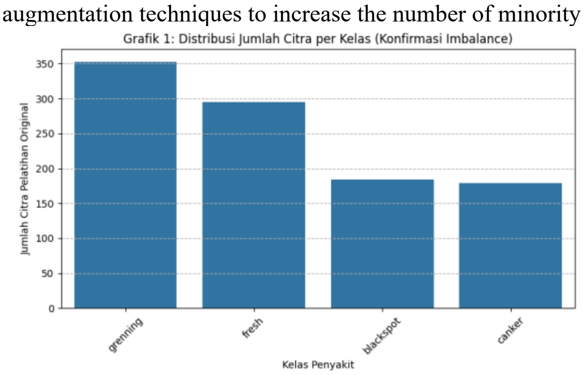

Fig.3 Distribution of Image Counts per Class (Confirmation of Imbalance)

Table 3. Class Distribution after Oversampling (Balanced)

Class	Count
Blackspot	35
Canker	353
Fresh	353
Grenning	353
Total	**1,422**

The augmentation-based oversampling process has successfully equalized the number of images in each class.

b. Data Preprocessing and Augmentation

The preprocessing stage includes:

1. Resizing all images to 224×224 pixels.
2. Pixel normalization (0–1).
3. Minority class augmentation: rotation, shear, zoom, horizontal flip.
4. One-hot encoding for labels.
5. Data splitting into train (91.1%) and test (8.9%) sets.

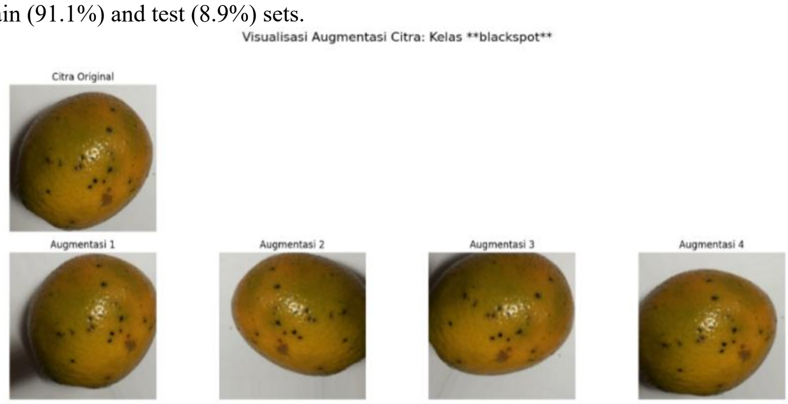

Fig.4 Visualization of Image Augmentation Results – Blackspot Class

c. Implementation of MobileNetV2 Architecture

Hyperparameter configuration is shown in Table 4.

Table4 Testing Parameters

Hyperparameters	Value/Configuration
Model Architecture	MobileNetV2
Input Shape	(224,224,3)
Optimizer	Adam
Loss Function	Categorical cross-entropy
Batch Size	32
Epoch	15
Evaluation *Matrix*	Accuracy, Recall, *Macro* avg F1-score, *Precision*

The baseline model (imbalanced) and mitigation model (balanced) were trained and tested using the same configuration to ensure comparative validity.

d. Baseline vs Oversampled Experiment

Table5 Model Performance Comparison (Baseline vs Oversampled)

Model	Accuracy	Precision	Recall	F1-Score
Baseline	87.31	87.11	87.24%	87.14%
Oversampled	95.79	95.78	95.79	95.78

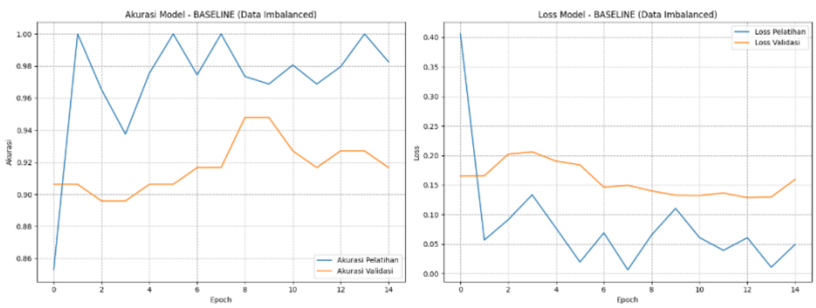

Fig.5 Baseline Model Training Graph (imbalanced)

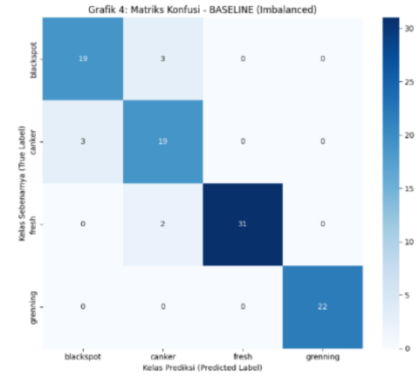

Fig.6 Confusion Matrix - BASELINE (imbalanced)

The baseline model experiences overfitting and bias toward the majority class.

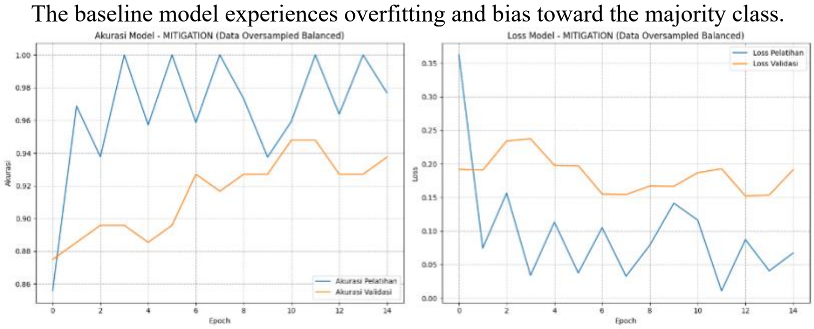

Fig.7*MITIGATION Model Training Results (Oversampled)*

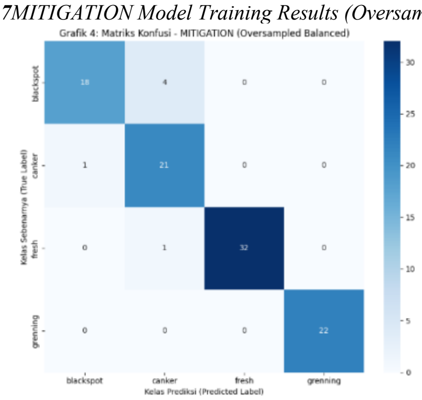

Fig.8 Confusion Matrix - MITIGATION (Oversampled Balanced)

The oversampled results show:

a) Better training stability
b) More stable validation loss
c) Reduced classification error between classes

Table6 Classification Report

Class	***precision***	**recall**	**F1-score**	**support**
blackspot	0.8636	0.8636	0.8636	22
canker	0.7917	0.8636	0.8261	22
fresh	1	0.9394	0.9688	33
grenning	1	1	1	22
accuracy			0.9192	99
macro average	0.9138	0.9167	0.9146	99
weighted average	0.9234	0.9192	0.9206	99

3. Discussion

This section discusses in depth the research findings based on the results of EDA, preprocessing, oversampling, model training, and performance evaluation. The discussion is organized according to the methodology flow and is directly related to the baseline and mitigation (oversampling) experiment results.

a. The Effect of Data Augmentation on Performance Improvement

Augmentation proved effective in increasing accuracy from 91.92% to 93.94% and the macro F1-score from 0.9146 to 0.9344. Rotation, zoom, shear, and flipping transformations enriched image variation, making the model more robust.

b. Impact of Data Imbalance on the Model

The baseline model is biased towards the majority class. Oversampling improves the recall of the minority class and reduces cross-classification errors.

c. Effectiveness of MobileNetV2 in Limited Data Conditions

MobileNetV2 proved to be stable and efficient. This model successfully learned disease features despite the limited initial data.

d. Comparative Analysis of Baseline vs. Oversampled Models
The comparison is shown in Table 7.

Table7 Model Performance Comparison (Baseline vs Oversampled)

Metric	*Baseline* (Imbalanced)	*MITIGATION* (*Oversampled*)
Accuracy	0.9192	0.9394
F1-Score (*Macro* Avg)	0.9146	0.9344
Precision (*Macro* Avg)	0.9138	0.9388
Recall (*Macro* Avg)	0.9167	0.9356

All metrics improved after oversampling → more balanced learning.

e. Implications for Precision Agriculture Systems
The model can be integrated into mobile and IoT applications for rapid detection of citrus leaf disease.

f. Comparison with Previous Research
Table 9 shows that your research:
a) Is more focused on *addressing class imbalance*
b) Uses a lightweight model (MobileNetV2)
c) Produces competitive performance (95.79%)

Table8 Comparison with Previous Research

Researcher	Method/Model	Dataset	Augmentation Technique	Accuracy	Comparison of Findings	Research Gap
(Ali et al., 2024)	CNN Ensemble (EfficientNet, DenseNet, ResNet, Inception)	PlantVillage (87,000 images, 38 classes)	Not used (class-weighted)	99.89	Ensemble models excel in multi-class classification using large datasets.	The effects of physical augmentation or oversampling on small minority classes have not been tested.
(G. Han et al., 2025)	CNN + GAN	Field images and synthetic images from GAN	Yes, via GAN	Not explicitly mentioned	Focus on visual enhancement and generalization for minority classes.	No direct evaluation of a single CNN model without GAN.
(Agustina et al., 2025)	DenseNet-121	Citrus fruit (greening, canker, blackspot, healthy)	Not mentioned	99.31	The model successfully classified with limited data and a robust architecture.	Does not discuss class distribution or the effect of augmentation on minorities.
(Firdaus et al., 2023)	DenseNet-169 + feature combination	LDI (3,000 images)	Yes	96.67	Feature combination improves accuracy by 5.33%.	Augmentation was performed, but the effects of balanced distribution were not discussed.
Penelitian saat ini (2025)	MobileNetV2 + augmentation and oversampling	1,011 images (4 orange leaf classes)	Yes (rotation, zoom, shear, flip)	95.79	Oversampling and augmentation significantly improve performance.	Provides empirical contributions in addressing class imbalance through structured and systematic physical augmentation.

4. Conclusion

Based on the entire series of research covering Exploratory Data Analysis (EDA), preprocessing, augmentation and oversampling, MobileNetV2 model training, and performance evaluation through two scenarios (baseline and mitigation), the following conclusions were obtained:
Augmentation and oversampling techniques have been proven effective in improving the classification performance of citrus leaf images. Data imbalance has a significant impact on the generalization ability of the model. MobileNetV2 has proven effective for classifying citrus leaf disease images under limited data conditions. The combination of preprocessing, augmentation, and efficient CNN architecture can produce models that are ready for field application. Based on the research results and limitations encountered, several recommendations for further research are as follows:
1. There is a need for more diverse field data variations.
2. Model testing should be expanded using other CNN architectures for comparison.
3. Advanced augmentation techniques and synthetic data generation should be explored.
4. Real-time system implementation on mobile devices or IoT.

References

[1] J. Han *et al.*, "Classifying Neovascular Age-Related Macular Degeneration With a Deep Convolutional Neural Network Based on Optical Coherence Tomography Images," *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-05903-7.
[2] H. Zhang, H. Sun, and T. Yuan, "Network Art Image Classification and Print Propagation Extraction Based on Depth Algorithm," *Wirel. Commun. Mob. Comput.*, vol. 2022, no. 1, 2022, doi: 10.1155/2022/2546015.
[3] Q. Lv, W. Feng, Y. Quan, G. Dauphin, L. Gao, and M. Xing, "Enhanced-Random-Feature-Subspace-Based Ensemble CNN for the Imbalanced Hyperspectral Image Classification," *Ieee J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 3988–3999, 2021, doi: 10.1109/jstars.2021.3069013.

[4] M. H. Aabidi, A. E. Makrani, B. Jabir, and I. Zaimi, “A Model Proposal for Enhancing Leaf Disease Detection Using Convolutional Neural Networks (CNN),” *Int. J. Online Biomed. Eng.*, vol. 19, no. 12, pp. 127–143, 2023, doi: 10.3991/ijoe.v19i12.40329.
[5] O. Abayomi-Alli, R. Damaševičius, S. Misra, and R. Maskeliūnas, “Cassava Disease Recognition From <scp>low-quality</Scp> Images Using Enhanced Data Augmentation Model and Deep Learning,” *Expert Syst.*, vol. 38, no. 7, 2021, doi: 10.1111/exsy.12746.
[6] A. Ayub and H. Kim, “GAN-Based Data Augmentation With Vehicle Color Changes to Train a Vehicle Detection CNN,” *Electronics*, vol. 13, no. 7, p. 1231, 2024, doi: 10.3390/electronics13071231.
[7] B. d. O. Maciel *et al.*, “Artificial Intelligence Tools: A Practice Report From the International Center for Software Technology,” *Rev. Contemp.*, vol. 4, no. 12, p. e6973, 2024, doi: 10.56083/rcv4n12-156.
[8] A. Pirmatov, B. A. Azimov, and S. Kamalov, “Artificial Intelligence Using Python: Technologies and Applications,” *Bull. Sci. Pract.*, vol. 9, no. 11, pp. 288–295, 2023, doi: 10.33619/2414-2948/96/37.
[9] A. Li, J. P. Winebrake, and K. D. Kovacs, “Facilitating Deep Learning Through Preprocessing of Optical Coherence Tomography Images,” *BMC Ophthalmol.*, vol. 23, no. 1, 2023, doi: 10.1186/s12886-023-02916-2.
[10] M. Yousif, “Enhancing the Accuracy of Image Classification Using Deep Learning and Preprocessing Methods,” *Artif. Intell. Robot. Dev. J.*, 2024, doi: 10.52098/airdj.2023348.
[11] G. Jin, S. Oh, Y. Lee, and S. Shin, “Extracting Weld Bead Shapes From Radiographic Testing Images With U-Net,” *Appl. Sci.*, vol. 11, no. 24, p. 12051, 2021, doi: 10.3390/app112412051.
[12] S. Prusty, S. Patnaik, and S. K. Dash, “SKCV: Stratified K-Fold Cross-Validation on ML Classifiers for Predicting Cervical Cancer,” *Front. Nanotechnol.*, vol. 4, 2022, doi: 10.3389/fnano.2022.972421.
[13] A. F. Boy, A. Akhyar, T. Y. Arif, and S. Syahrial, “Artificial Intelligence for Dental Caries Detection: A Mixup, Fine-Tuning, and Quantization Approach on the MobileNetV2 Model,” *J. Conserv. Dent. Endod.*, vol. 28, no. 8, pp. 764–771, 2025, doi: 10.4103/jcde.jcde_362_25.
[14] M. Iqbal, F. Rozi, N. O. Adiwijaya, and D. I. Swasono, “Identifikasi Kinerja Arsitektur Transfer Learning VGG16 , ResNet-50 , dan Inception-V3 Dalam Pengklasifikasian Citra Penyakit Daun Tomat Identification of VGG16 , ResNet-50 , and Inception-V3 Transfer Architecture Performance in Image Classification of Tomato Leaf Diseases,” vol. 5, no. 2, pp. 145–154, 2023.
[15] H. Bechinia, D. Benmerzoug, and N. Khlifa, “Approach Based Lightweight Custom Convolutional Neural Network and Fine-Tuned MobileNet-V2 for ECG Arrhythmia Signals Classification,” *Ieee Access*, vol. 12, pp. 40827–40841, 2024, doi: 10.1109/access.2024.3378730.
[16] S. Albahli and M. Nawaz, “DCNet: DenseNet-77-based CornerNet Model for the Tomato Plant Leaf Disease Detection and Classification,” *Front. Plant Sci.*, vol. 13, 2022, doi: 10.3389/fpls.2022.957961.
[17] P. Sharma and P. Abrol, “Agricultural Advancements Through Machine Learning Technologies,” *Environ. Ecol.*, vol. 42, no. 2B, pp. 775–779, 2024, doi: 10.60151/envec/krws4781.
[18] A. H. Ali, A. Youssef, M. Abdelal, and M. A. Raja, “An ensemble of deep learning architectures for accurate plant disease classification,” *Ecol. Inform.*, vol. 81, no. September 2023, p. 102618, 2024, doi: 10.1016/j.ecoinf.2024.102618.
[19] G. Han, D. Kwaku, P. Asiedu, and K. Ebo, “Heliyon Plant disease detection with generative adversarial networks,” *Heliyon*, vol. 11, no. 7, p. e43002, 2025, doi: 10.1016/j.heliyon.2025.e43002.
[20] S. Agustina, A. Hidayat, Satriansyah, and R. Kurniawan, “Klasifikasi Penyakit pada Buah Jeruk Berdasarkan Citra dengan Pendekatan Transfer Learning Menggunakan Arsitektur Densenet-121,” *J. Ilm. Inform.*, vol. 10, pp. 42–47, Jul. 2025, doi: 10.35316/jimi.v10i1.41-47.
[21] M. H. Firdaus, E. Utami, and D. Ariatmanto, “Detection And Classification of Citrus Diseases Based on A Combination of Features Using the Densenet-169 Model,” *Sinkron*, vol. 8, no. 4, pp. 2592–2601, 2023, doi: 10.33395/sinkron.v8i4.12974.