Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

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

Comparison of Balancing Strategies for Classifying Guava Fruit Diseases

Putri Nabilla[1]*, Nana Suarna[2], Agus Bahtiar[3], Nining Rahaningsih[4], Willy Prihartono[5]

[1,2]Informatics Engineering, STMIK IKMI CIREBON
[3]Information System, STMIK IKMI CIREBON
[4,5]Computerized Accounting, STMIK IKMI CIREBON
putrinabilaaa7@gmail.com[1*], st_nana@yahoo.com[2], agusbahtiar038@gmail.com[3], niningr157@yahoo.co.id[4], willyprihartono@gmail.com[5]

Abstract

The problem of class imbalance often poses an obstacle in deep learning-based image classification, especially in the domain of digital agriculture. The imbalance in data distribution makes it easier for models to recognize the majority class, while performance for the minority class declines. This study aims to analyze the effectiveness of three strategies for handling class imbalance: Weighted Loss Function, Oversampling, and a combination of Weighted Loss and Oversampling, in improving the performance of image classification of guava fruit diseases using a transfer learning-based MobileNetV2 architecture. The dataset consists of 3,784 images of three disease classes, namely Anthracnose, Fruit_Fly, and Healthy_guava, which show an imbalanced distribution. The research was conducted through the stages of Exploratory Data Analysis (EDA), pre-processing, augmentation, model training with four scenarios, and evaluation using Accuracy, Precision, Recall, F1-Score, and Macro Average F1-Score. The results showed that the Combination model (Oversampling and Weighted Loss) performed best on the minority class with an F1-score of 0.9630, the highest among all models. The Oversampling strategy produced the highest Macro F1-score of 0.9617, while Weighted Loss provided a significant improvement in classification sensitivity but was still below the combination model. Thus, it can be concluded that the combination strategy is the most effective approach in improving the sensitivity of the model to minority classes, while Oversampling excels in the overall performance stability of the model.

***Keywords**: Class Imbalance, Weighted Loss Function, Oversampling, MobileNetV2, Guava Disease Classification*

1. Introduction

Advances in computer vision and deep learning technology have encouraged the use of intelligent systems to support automation in various sectors, including agriculture. In this field, Convolutional Neural Networks (CNN) are widely used to detect plant diseases based on digital images, due to their ability to recognise visual patterns accurately and efficiently [1]. This approach is in line with the concept of precision agriculture, which aims to improve the productivity and sustainability of modern agricultural systems [2].

In guava (Psidium guajava L.) commodities, early disease detection is important given their high economic value and susceptibility to diseases such as Anthracnose and Fruit Fly. However, the development of image classification models for plant disease identification often faces a fundamental problem of class imbalance in the dataset. This imbalance causes the model to focus more on learning patterns from the majority class, thereby reducing performance for the minority class even though the overall accuracy appears high [3], [4]. The results of Exploratory Data Analysis (EDA) on the guava dataset used in this study also show an uneven distribution of samples, especially in the healthy guava class, which is far fewer in number than the other two disease classes.

Various studies have proposed approaches to address this imbalance. At the algorithmic level, weighted loss functions have been shown to improve model sensitivity to minority classes [5]. At the data level, minority oversampling methods, either through conventional augmentation or generative approaches such as GAN, can enrich the variation of minority classes and improve model performance [6], [7]. Other studies show that the combination of weighted loss and oversampling results in more stable performance improvements, especially in cases of long-tail distributions and plant disease datasets [8], [9], [10].

These findings confirm that the selection of class balancing strategies is an important component in the development of image classification models, particularly in the field of agricultural AI. Based on these issues, this study aims to conduct a comparative analysis of three class imbalance handling strategies, namely Weighted Loss Function, Minority Oversampling, and a combination of both, in the task of classifying guava disease images using a transfer learning-based MobileNetV2 architecture. The evaluation was conducted using the Accuracy, Precision, Recall, F1-Score, and Macro Average F1-Score metrics to assess the effectiveness of each strategy in improving model performance for minority classes.

This research is expected to provide theoretical contributions in the form of empirical understanding of the impact of data imbalance on CNN performance, as well as practical contributions in the form of recommendations for effective data balancing strategies for the development of more accurate and reliable digital image-based plant disease detection systems.

2. Research Metodology

2.1. Research Phases

The research stages were arranged systematically, starting from data collection, exploratory data analysis, image pre-processing and augmentation, application of MobileNetV2 architecture based on transfer learning, to class imbalance handling strategies tested through four different approaches. All of these stages are designed to ensure that the model training, validation, and evaluation processes are carried out in a structured manner so as to produce an objective comparison of the performance of the balancing methods in classifying guava diseases. The research stages carried out are shown in Figure 1.

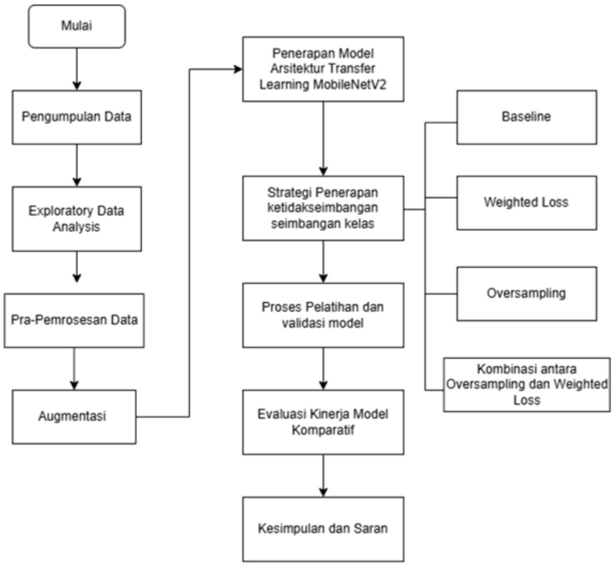

Fig. 1: Research Phase

2.1.1. Dataset Collection and Data Sources

In this study, the author used public secondary data (supervised learning). The dataset was obtained from the Kaggle website (open data). The guava fruit data consisted of three main classes, namely Anthracnose (fungal spot disease), Fruit Fly (damage caused by fruit flies) and Healthy Guava (healthy fruit). The dataset used includes 3,784 image data divided into 70% training data, 10% testing data, and 20% validation data, as shown in Table 3.1. The images have undergone pre-processing steps such as unsharp masking and CLAHE. All image data are in RGB format with uniform resolution, with a data dimension of 512 x 512 pixels. The distribution of data numbers is shown in Table 1.

Table 1: Overall Data Distribution

Set	Test	Train	val	Total
Label				
Anthracnose	156	1.080	308	1.544
Fruit_Fly	132	918	262	1.312
Healthy_guava	94	649	185	928
Total Set	382	2.647	755	3.784

2.1.2. Exploratory Data Analysis (EDA)

This stage is carried out to understand the dataset structure, visual image variations, and class distribution. The analysis covers the number of images per class, training data ratio, testing, validation, and basic statistics such as colour intensity and brightness. EDA also helps detect potential outliers or labelling errors in the dataset, measure class frequency and imbalance [11].

2.1.3. Pre-processing of Data

The pre-processing stage is carried out to ensure that the images meet the requirements of the MobileNetV2 architecture and are ready for use in the training process. All images are normalised to the range [0-1] by dividing the pixel values by 255 and resized to 224×224 pixels to match the MobileNetV2 standard input. Class labels (Anthracnose, Fruit_Fly, Healthy_guava) are converted to numerical format using one-hot encoding to facilitate processing by the softmax layer. In addition, directory structure checks and data integrity validation were performed to avoid duplication and labelling errors. This entire process was run through ImageDataGenerator, which automatically loaded, scaled, and arranged the images into batches before feeding them to the model.

2.1.4. Augmentation

The data augmentation stage is used to enrich image variation and reduce the risk of overfitting by creating artificial variations from the training data. At this stage, augmentation will be carried out by applying image rotation, horizontal and vertical shifts, shear, zoom, and horizontal flipping.

2.1.5. Implementation of the MobileNetV2 Model Architecture

The MobileNetV2 architecture is used as a transfer learning model due to its lightweight yet accurate nature, with 224×224×3 input, the use of depthwise separable convolution, inverted residual blocks, and linear bottlenecks. This model utilises pre-trained ImageNet weights as a feature extractor with frozen convolutional layers, while new classification layers GlobalAveragePooling2D, Dense with ReLU activation, and softmax are trained to adapt to the guava disease classification task. The computational efficiency of MobileNetV2, which has only about 2.4 million parameters and supports small batching and float32 data types, makes the training process stable, fast, and minimises the risk of overfitting, making it relevant for use on unbalanced datasets in this study [12], [13].

2.1.6. Class Imbalance Implementation Strategy

After implementing the MobileNetV2 architecture, four strategies were applied to address class imbalance and evaluate their effectiveness, namely the baseline model without balancing techniques, the Weighted Loss Function strategy that gives greater weight to minority classes, the Oversampling method, which increases the number of minority class samples through additional augmentation, and the Combination approach, which integrates weighted loss and oversampling to assess the potential for simultaneous performance improvement.

2.1.7. Model Performance Evaluation

The performance evaluation of the model in this study was conducted using several key metrics, namely accuracy to assess overall predictions, precision to measure the accuracy of positive class predictions, recall to see the model's ability to detect minority classes, and F1-score as a metric that balances precision and recall. In addition, a confusion matrix was used to visualise the relationship between predicted labels and actual labels, allowing for a clearer analysis of classification error patterns between classes. The entire training process was conducted using a Google Colab CPU/GPU environment so that the results of each epoch were recorded consistently and could be compared between models.

3. Results and Discussion

3.1. Results of Exploratory Data Analysis (EDA)

The dataset distribution in Figure 2 shows that the number of images in the three classes, Anthracnose, Fruit_Fly, and Healthy_guava, is unbalanced, with Anthracnose having the most samples at 1,544, followed by Fruit_Fly with 1,312 data, and Healthy_guava as the minority class with around 928. This imbalance has the potential to cause model bias because the majority class is more dominant in the learning process, so that the recall of the Healthy_guava class tends to be low. Figure 2 also shows that the imbalance remains consistent in the train, test, and val subsets even though the data division is done proportionally, so that a balancing strategy is still needed to ensure fair model performance across all classes.

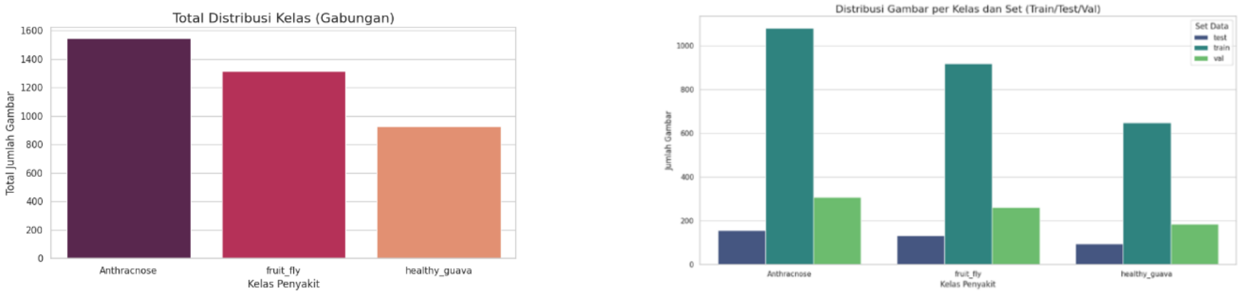

Fig. 2 : Class distribution

In the image property analysis, the statistical results of all 3,784 images have a uniform size of 512×512 pixels with an aspect ratio of 1, so that all images are square and standardised. The standard deviation value of zero confirms that there is no variation in size, while the consistent number of RGB channels and the absence of corrupted files indicate that the dataset is ready for use without additional dimensional adjustments. The histogram and scatter plot visualisations in Figure 3 also show a single distribution at 512 pixels, confirming the uniformity of image resolution. This consistency is beneficial in the training process as it does not require padding or additional size normalisation.

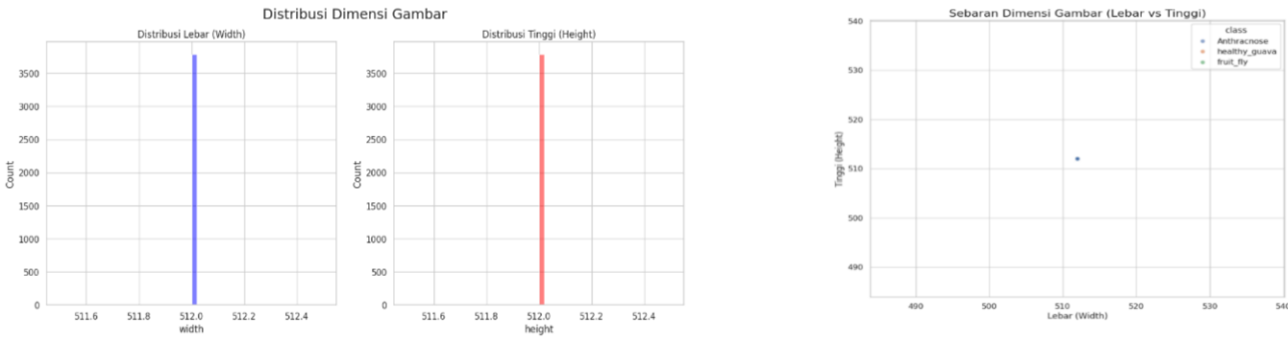

Fig. 3 : Image Dimensions

Analysis of the colour histogram and brightness distribution shows variations in lighting between classes, where Fruit_Fly has higher brightness values, Anthracnose is at a medium lighting level, and Healthy_guava has a dominance of the green (G) channel. The three brightness curves overlap, indicating differences in lighting between classes but still within reasonable limits. The colour histogram also shows the visual characteristics of each class, such as an increase in the red channel (R) in Anthracnose due to dark-coloured fungal spots, and brighter colour variations in Fruit_Fly, as shown in Figures 4 and 5.

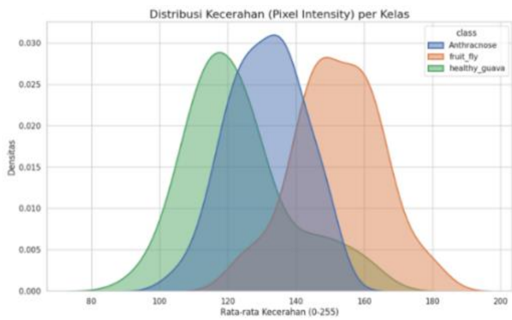

Fig. 4 : Average brightness distribution

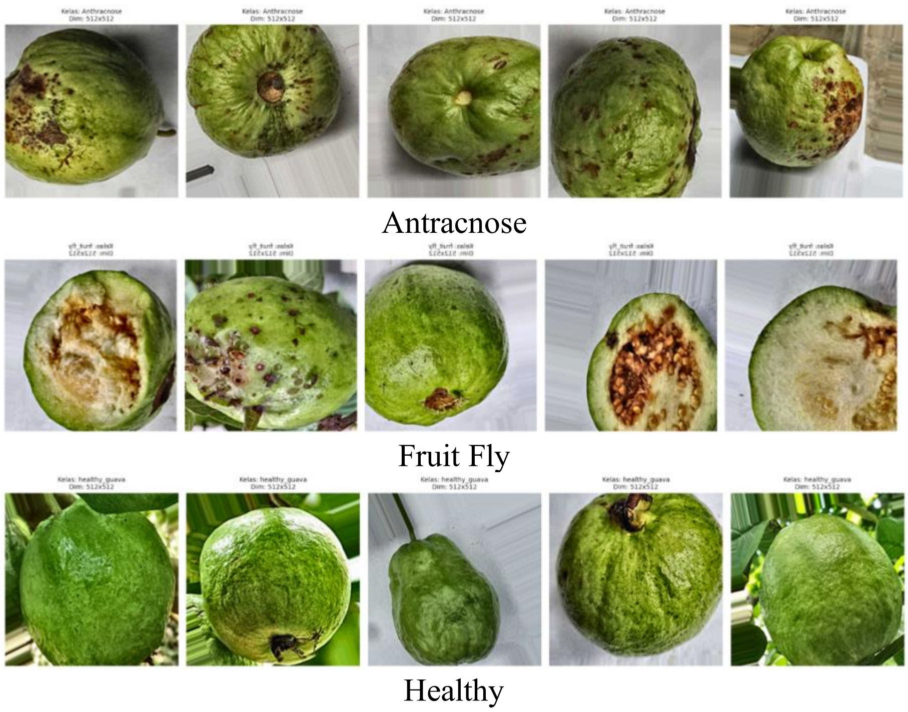

Fig. 5 : Sample data for Anthracnose, Fruit fly and Healthy classes

Overall, the dataset has good visual quality with a consistent RGB format, but still shows an imbalance in the number of samples between classes. These EDA findings form an important basis for determining the balancing strategy in the next stage, such as Weighted Loss Function, Oversampling, or a combination of both, in order to reduce model bias and improve the ability to recognise patterns from classes with limited representation.

3.2. Pre-processing Data Results

The pre-processing stage is applied to ensure that all images have a consistent format, size, and quality before being used in model training. All images are resized to 224 × 224 pixels to match the MobileNetV2 input while maintaining a balance between computational efficiency and visual feature clarity. This process allows the model to learn image patterns stably without losing important details on disease classes. Next, all images are normalised to a pixel range of [0, 1] by dividing the RGB values by 255. This normalisation equalises the data scale, accelerates model convergence, and improves gradient calculation stability during training. With standardised data, the training process becomes more efficient and the model results more consistent. The preprocessing results are shown in Figure 6.

Fig. 6 : Preprocessing results

3.3. Augmentation Results

Data augmentation was applied to enrich image variation, particularly in minority classes such as Healthy_guava, using a combination of geometric and photometric transformations to enable the model to better adapt to changes in object orientation, position, lighting, and distance. All baseline models, weighted loss, oversampling, and combinations use the same augmentation parameters, including pixel normalisation (rescale 1/255), rotation up to 30°, horizontal and vertical shifts of 20%, shear of 20°, zoom of 20%, horizontal flipping, and fill_mode="nearest" to keep the images natural. With this configuration, the resulting image variations become more diverse, making MobileNetV2 more robust to real-world conditions, helping to reduce overfitting, and improving generalisation capabilities, especially for classes with low sample counts. Figure 7 shows the results after the augmentation process.

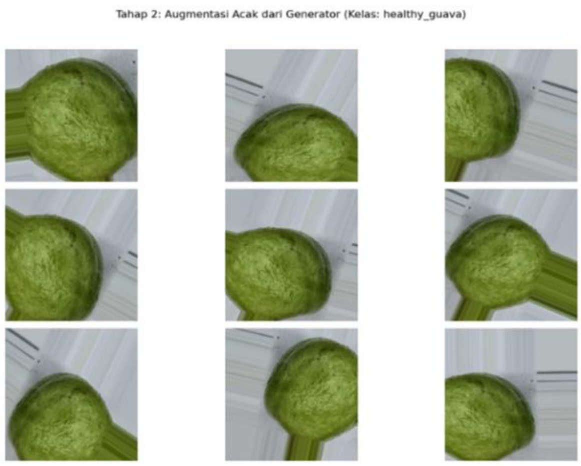

Fig. 7 Example of results after augmentation

3.4. Baseline Model Results

The baseline model shows high accuracy (0.9398), but its performance on minority classes is lower. The healthy_guava class obtained an F1-score of 0.94, still below the majority class Anthracnose. This indicates model bias towards an imbalanced data distribution. The confusion matrix shows that the baseline model trained without class balancing strategies performs well on the majority class Anthracnose, with 155 out of 156 images classified correctly. However, its accuracy decreases in the Fruit_Fly class, where only 117 out of 132 images are correctly recognised, and the rest are misclassified as Anthracnose or Healthy_guava.

Meanwhile, in the Healthy_guava class, the model only classified 87 of the 94 images correctly, with several incorrect predictions to other classes. This pattern confirms that the baseline model is biased towards the majority class and less sensitive to the minority class, requiring a balancing strategy to make the classification performance more balanced. Figure 8 shows the confusion matrix results of the baseline model.

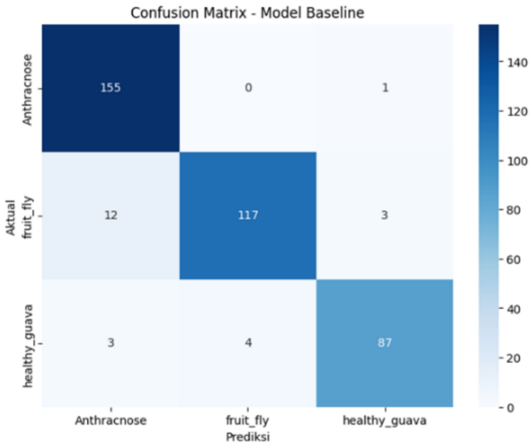

Fig. 8: Confusion matrix model baseline

3.5 Weighted Loss Function Model Results

The Weighted Loss Function was applied in this study by adding class weights to the loss function to balance the contribution of each class during training. Class weights were determined based on the inverse class frequency, so that classes with less data received a greater penalty. Mathematically, the weights were calculated using the following formula:

$$\omega_i = \frac{N}{n_i} \tag{1}$$

ω_i = weight for class i
N = total number of samples in the dataset
n_i = number of samples in class i

The application of class-weighted loss improves the model's sensitivity to minority classes. The F1-score value for the healthy_guava class increased to 0.9570, while Macro F1 also increased from 0.9388 to 0.9549. The model was able to balance the contributions between classes even without changes in data distribution. The confusion matrix of the model with the Weighted Loss Function strategy shows more balanced classification performance across the three guava disease classes. The model successfully recognised 154 out of 156 Anthracnose images, 122 out of 132 Fruit_Fly images, and 89 out of 94 Healthy_guava images. The dominance of values on the main diagonal indicates that Weighted Loss effectively increases sensitivity to minority classes without reducing the accuracy of majority classes, resulting in more stable performance across all classes. Figure 9 shows the results of the weighted loss model confusion matrix.

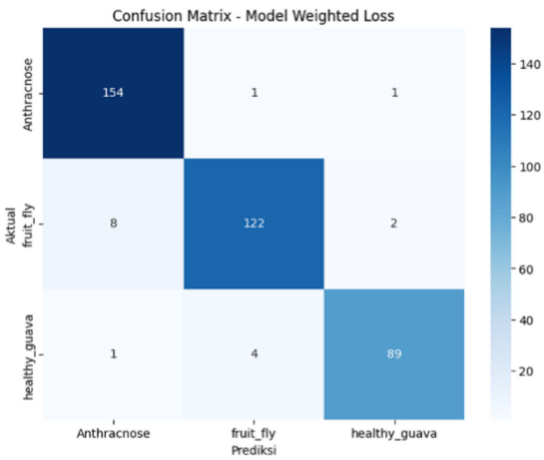

Fig. 9 : Confusion matrix model weighted loss

3.6. Oversampling Model Results

Oversampling improves the distribution of the dataset so that the model can learn minority class patterns better. The F1-score for the healthy_guava class increased to 0.9565, and the Macro F1 increased to 0.9617, the highest value among all models. This shows that adding data variation through augmentation contributes positively to model performance. The confusion matrix of the model with the oversampling strategy shows high and more balanced performance across the three guava disease classes. The model successfully classified 155 out of

156 Anthracnose images, 125 out of 132 Fruit_Fly images, and 88 out of 94 Healthy_guava images correctly. The dominance of diagonal values indicates that Oversampling effectively strengthens the model's ability to recognise minority classes without reducing accuracy in majority classes, resulting in a more stable performance distribution across classes. Figure 10 shows the results of the Oversampling model confusion matrix.

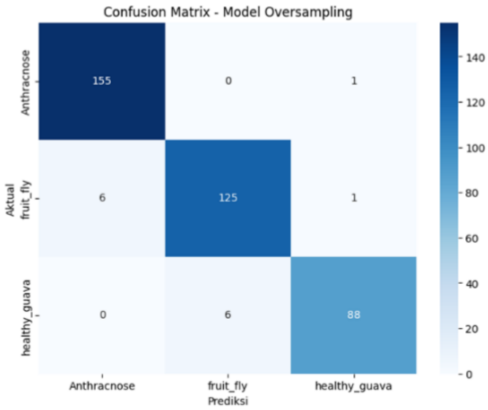

Fig. 10: Confusion matrix model oversampling

3.7. Results of Weighted Loss and Oversampling Combination

This hybrid model combines a data-level and algorithm-level approach. Although Macro F1 (0.9600) is slightly lower than the oversampling method alone, this model still shows stable performance across all classes and the highest accuracy (0.9607). This combined strategy has proven effective in reducing bias and improving model generalisation. The confusion matrix in the combination model (Weighted Loss and Oversampling) shows excellent and more balanced performance across all guava disease classes. The model successfully classified 155 out of 156 Anthracnose images correctly, 121 out of 132 Fruit_Fly images accurately, and 91 out of 94 Healthy_guava images, which is a minority class. The dominance of predictions on the main diagonal confirms that the combination strategy is able to reduce bias towards the majority class while increasing detection accuracy in the minority class, resulting in stable and more optimal classification performance compared to the single approach. Figure 11 shows the results of the Combination model confusion matrix.

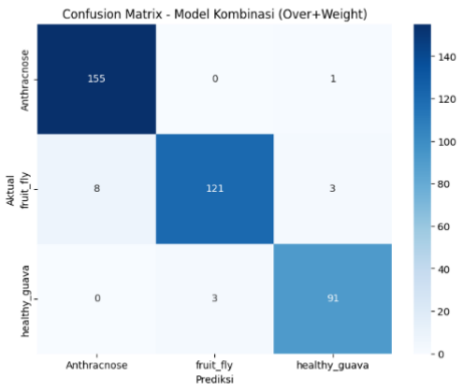

Fig. 11: Confusion matrix model combination

3.8. Comparative Results Evaluation

A comparative evaluation of four models Baseline, Weighted Loss, Oversampling, and Combination shows significant differences in performance in handling class imbalance in the guava fruit disease dataset, as shown in Table 2. The Baseline model obtained a macro average F1-score of 0.9388 and an accuracy of 0.9398, indicating a bias towards the majority class. The application of Weighted Loss improved performance to a macro F1-score of 0.9549 and an accuracy of 0.9555, indicating that class weighting was able to improve sensitivity to the minority class.

Oversampling provided the best results for overall performance with a macro F1-score of 0.9617 and an accuracy of 0.9634, indicating that balancing data between classes through augmentation effectively improves prediction stability. However, the combination of oversampling and weighted loss produced a macro F1-score of 0.9600 and an accuracy of 0.9607, slightly lower than oversampling alone, although still in the high performance category.

Table 2 : Test results for all models

Model	Anthracnose	fruit_fly	healthy_guava	Macro Avg F1	Accuracy
Baseline	0,9509	0,9249	0,9405	0,9388	0,9398
Weighted Loss	0,9655	0,9421	0,9570	0,9549	0,9555
Oversampling	0,9779	0,9506	0,9565	0,9617	0,9634
Kombinasi (Over+Weight)	0,9718	0,9453	0,9630	0,9600	0,9607

When looking specifically at the Healthy_guava minority class, the Combination strategy provides the highest F1-score of 0.9630, surpassing Weighted Loss (0.9570) and Oversampling (0.9565). This shows that combining these two techniques has a synergistic effect in improving the model's sensitivity to minority classes. Overall, Oversampling is the most effective strategy for improving the overall balance of model performance, while Combination is most optimal for improving performance specifically on minority classes. The comparison curve of validation accuracy and loss for all models is shown in Figure 12.

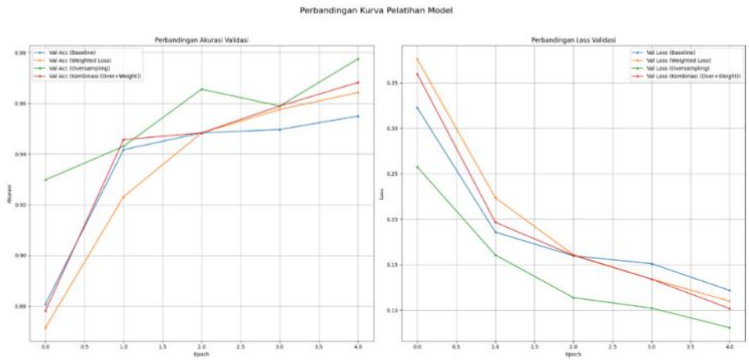

Fig. 12: Accuracy val and loss val comparison curve for all models

3.9. Discussion

Class imbalance in the dataset causes the baseline model to be biased towards the majority class, resulting in poorer performance for the minority classes, especially Healthy_guava. This can be seen from the lower baseline F1-score in the Fruit_Fly and Healthy_guava classes and the number of prediction errors in the confusion matrix. The application of the Weighted Loss Function is able to reduce this bias by increasing the penalty for minority class errors, thereby increasing the Healthy_guava F1-score from 0.9405 to 0.9570. However, this improvement is still limited because the data distribution has not changed.

On the other hand, the Oversampling strategy provided the best overall performance improvement with a Macro F1-score of 0.9617 thanks to the addition of visual variation in the minority class. Meanwhile, the Combination strategy (Oversampling and Weighted Loss) provided the most optimal results for the Healthy_guava class, with the highest F1-score of 0.9630. Although its Macro F1-score was slightly below Oversampling, the Combination approach remained the most effective method in achieving the main objective of the study, which was to increase the model's sensitivity to minority classes in conditions of data imbalance.

4. Conclusion

Based on the results of the study, it can be concluded that class imbalance in the dataset causes the baseline model to be biased towards the majority class, resulting in high accuracy but a low F1-score in the Healthy_guava class. This condition indicates that the model is unable to optimally recognise visual patterns of the minority class. The application of the Weighted Loss function successfully increased the model's sensitivity to the Healthy_guava class by improving the F1-score from 0.9405 to 0.9570, demonstrating a significant improvement compared to the baseline model, although it has not yet proven to be the most effective strategy overall. Meanwhile, the combination strategy between Oversampling and Weighted Loss achieved the best performance with the highest F1-score of 0.9630 for the Healthy_guava class, as it was able to balance the data distribution while simultaneously strengthening the error penalty on the minority class, making it the most effective approach in achieving the research objectives. Based on these findings, further research is recommended to explore more advanced balancing methods such as Focal Loss, Class-Balanced Loss, or LDAM Loss, as well as generative techniques including GANs, Diffusion Models, or CycleGAN to increase the diversity of minority-class data. Future studies may also evaluate more modern architectures such as EfficientNet, ResNet, or Vision Transformer (ViT) and utilise larger and more varied datasets to improve the model's generalisation capability. These recommendations are expected to support the development of a more accurate deep learning-based plant disease classification system that is ready to be implemented in the context of precision agriculture in Indonesia.

References

[1] H. N. Ngugi, A. A. Akinyelu, and A. E. Ezugwu, "Machine learning and deep learning for crop disease diagnosis: performance analysis and review," *Agronomy*, vol. 14, no. 12, p. 3001, 2024, doi: 10.3390/agronomy14123001.

[2] A. O. Adegbenjo and M. O. Ngadi, “Handling the imbalanced problem in agri-food data analysis,” *Foods*, vol. 13, no. 20, p. 3300, 2024, doi: 10.3390/foods13203300.

[3] M. Xu *et al.*, “Embracing limited and imperfect training datasets: opportunities and challenges in plant disease recognition using deep learning,” *Front Plant Sci*, vol. 14, p. 1225409, 2023, doi: 10.3389/fpls.2023.1225409.

[4] D. Dablain, B. Krawczyk, and N. Chawla, “Towards a holistic view of bias in machine learning: bridging algorithmic fairness and imbalanced learning,” *Discover Data*, vol. 2, p. 4, 2024, doi: 10.1007/s44248-024-00007-1.

[5] S. Albahli and M. Masood, “Efficient attention-based CNN network (EANet) for multi-class maize crop disease classification,” *Front Plant Sci*, vol. 13, p. 1003152, 2022, doi: 10.3389/fpls.2022.1003152.

[6] H. Alghamdi and T. Turki, “PDD-Net: Plant disease diagnoses using multilevel and multiscale convolutional neural network features,” *Agriculture*, vol. 13, no. 5, p. 1072, 2023, doi: 10.3390/agriculture13051072.

[7] J. Mu, Q. Feng, J. Yang, J. Zhang, and S. Yang, “Few-shot disease recognition algorithm based on supervised contrastive learning,” *Front Plant Sci*, vol. 15, p. 1341831, 2024, doi: 10.3389/fpls.2024.1341831.

[8] X. Wang and W. Cao, “GACN: Generative Adversarial Classified Network for Balancing Plant Disease Dataset and Plant Disease Recognition,” *Sensors*, vol. 23, no. 15, Aug. 2023, doi: 10.3390/s23156844.

[9] B. Min, T. Kim, D. Shin, and D. Shin, “Data Augmentation Method for Plant Leaf Disease Recognition,” *Applied Sciences (Switzerland)*, vol. 13, no. 3, Feb. 2023, doi: 10.3390/app13031465.

[10] O. Khare, S. Mane, H. Kulkarni, and N. Barve, “LeafNST: an improved data augmentation method for classification of plant disease using object-based neural style transfer,” *Discover Artificial Intelligence*, vol. 4, no. 1, Dec. 2024, doi: 10.1007/s44163-024-00150-3.

[11] V. Da Poian *et al.*, “Exploratory data analysis (EDA) machine learning approaches for ocean world analog mass spectrometry,” *Frontiers in Astronomy and Space Sciences*, vol. 10, p. 1134141, 2023, doi: 10.3389/fspas.2023.1134141.

[12] M. Pradnyatama, “MobileNetV2 and Xception for butterfly species classification,” *Jurnal Masyarakat Informatika*, vol. 16, no. 1, 2025, doi: 10.xxx/xxxxx.

[13] L. Yong, L. Ma, D. Sun, and L. Du, “Application of MobileNetV2 to waste classification,” *PLoS One*, vol. 18, no. 3, p. e0282336, 2023, doi: 10.1371/journal.pone.0282336.