MJI
Media Jurnal Informatika
Vol. 17, No. 2, December 2025
p-issn : 2088-2114, e-issn : 2477-2542

Classification of Banana Ripeness Using a VGG16-Based Convolutional Neural Network (CNN)

Fikri Maulana [a,1]

[a] Universitas Nusa Mandiri, East Jakarta, DKI Jakarta, 13620, Indonesia
[1] fikrimaulana85@gmail.com*

ARTICLE INFO

Article history
Received 2025/11/26
Revised 2025/12/16
Accepted 2025/12/20

Keywords
Computer Vision
VGG16 Architecture
Convolutional Neural Network
Banana Ripeness Classification
Deep Learning

ABSTRACT

The ripeness level of bananas is a crucial factor that affects the quality, taste, and selling value of the commodity, but the manual sorting process that is commonly carried out is still subjective, inconsistent, and time-consuming. This study aims to implement and evaluate the performance of a VGG16-based Convolutional Neural Network (CNN) architecture in automatically classifying the ripeness level of bananas. The research dataset consists of 5,616 digital images obtained from the Roboflow Universe platform and grouped into six specific classes: freshripe, freshunripe, overripe, ripe, rotten, and unripe. The system development methodology includes data division using stratified splitting techniques, image pre-processing with data augmentation strategies to prevent overfitting, and the application of transfer learning. The model was trained using the Stochastic Gradient Descent (SGD) optimization algorithm with a learning rate of 0.001 for 25 epochs on GPU-based hardware. Performance evaluation was conducted in depth using a confusion matrix, F1-Score metrics, and Precision-Recall curve analysis. The experimental results showed that the VGG16 model achieved an overall accuracy of 97.13%. Class-by-class analysis shows perfect performance in the freshunripe category, although there is a slight decrease in precision in the ripe class due to the similarity of visual characteristics with the overripe class. The stability of the training and validation accuracy curves also indicates that the model has good generalization capabilities. This study concludes that the VGG16 architecture is a reliable and accurate solution to support the efficiency of smart farming systems.

1. Introduction

Bananas (Musa paradisiaca L.) are one of Indonesia's main horticultural commodities, with high production levels and widespread consumption. The ripeness level of bananas is an important factor that affects the quality, taste, shelf life, and selling value of the fruit. In practice, determining the ripeness of bananas is still mostly done manually based on visual observation, which is subjective, inconsistent, and requires a lot of time and energy, especially in the context of distribution and modern agriculture.

With the development of technology, digital image processing has begun to be used as a solution to overcome the limitations of conventional methods. Initial approaches mostly utilized color-based methods, such as YCbCr color space transformation and thresholding techniques. Although these methods are computationally light, their performance is highly dependent on lighting conditions and background, making them less robust for use in real-world environments [1].

Existing research applies Convolutional Neural Networks (CNN), which have been proven capable of automatically extracting visual features and are more adaptive to image data variations. Several studies show that CNNs are effective for classifying fruit objects, including bananas. In one study that applied CNN with the VGG16 architecture for banana classification, an accuracy of 78% was achieved, proving the potential of VGG16 in fruit image processing [2]. Another study used the VGG19 architecture for banana ripeness classification and achieved a validation accuracy of up to 98%, but its implementation is still limited to static (offline) image processing [3].

On the other hand, recent research has begun to focus on real-time banana ripeness detection and classification systems using object detection models such as YOLOv8 and the MobileNetV2–KNN hybrid approach. These models are capable of achieving high accuracy and running in real time, but they have complex architectures and do not specifically examine the performance of CNN VGG16 as an end-to-end model for real-time systems [4].

Comparative research on CNN architecture in other fruits, such as apples, shows that VGG16 has higher accuracy performance than MobileNetV2, even though it requires greater computing resources [5]. This shows that VGG16 is still relevant for further discussion, especially in real-time optimization and application.

From previous research, a banana ripeness classification system can be developed using a VGG16-based Convolutional Neural Network (CNN). Most studies utilizing the VGG16 architecture still focus on offline image classification, while object detection-based approaches generally use other models such as YOLO or hybrid methods. Therefore, this study aims to implement and evaluate the performance of CNN VGG16 in classifying the ripeness level of bananas based on digital images. The proposed system was developed through methodological stages that included dataset collection, standardized image pre-processing, model training using transfer learning techniques, and performance evaluation based on classification metrics. To support the efficiency of the model training and testing process, this study utilizes GPU-based computing acceleration with a CUDA-enabled device hardware configuration: cuda:0 (NVIDIA GTX 1060 6GB), so that the system is expected to produce accurate, stable, and reliable classifications in supporting the application of smart farming technology.

2. Method

This research on deep learning-based system development, the data preparation stage plays an important role in determining image classification accuracy, where feature engineering is needed to recognize complex data patterns [6], [7]. In this study, the dataset used was obtained through the Roboflow Universe Platform. The Roboflow platform was chosen because the dataset was already structured, from the labeling process to the provision of essential preprocessing features before the data entered the model training stage [8], [6].

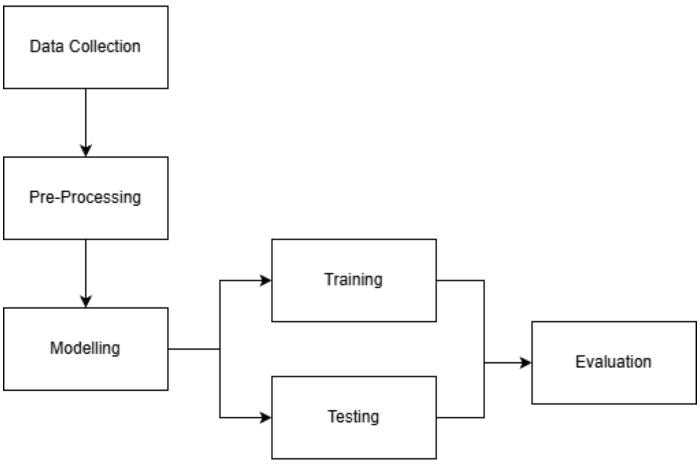

Figure 1. Method Design

2.1 Data Collection

The dataset used in this study consisted of 5,616 digital images obtained through the Roboflow Universe Platform. To enable the system to recognize specific visual characteristics of banana ripeness, the dataset was classified into six classes, namely fresh ripe, fresh unripe, overripe, ripe, rotten, and unripe. This detailed class grouping plays an important role in overcoming the challenge of visual similarities between maturity levels, which often complicates the classification process, while minimizing the inconsistency of manual assessment due to human subjectivity. Furthermore, the dataset was divided into three subsets, namely a training set of 70% (3,931 images), a validation set of 20% (1,123 images), and a test set of 10% (562 images). The data division process was carried out using a stratified splitting approach to maintain a consistent class distribution proportion in each subset, so that each maturity level remained evenly represented in the training, validation, and test data. The stratified split implementation was carried out by combining all the initial data into one set before redistributing it using the train_test_split function with the stratify parameter and random_state = 42 to ensure experiment reproducibility. This approach ensures that the class distribution in each subset is statistically identical, allowing for a more objective and unbiased evaluation of model performance [8], [6].

2.2 Pre-Processing

During pre-processing stage, this study applied a data augmentation strategy to increase feature representation variability during the training process in order to mitigate the risk of overfitting and improve the generalization ability of the VGG16 CNN model. Data augmentation on training images was performed stochastically through a combination of geometric and photometric transformations, including RandomResizedCrop with a scale of 0.8–1.0 to simulate the effects of zoom and image capture distance variations, RandomRotation up to ±15 degrees to improve the model's resilience to object orientation differences, RandomHorizontalFlip to train invariance to lateral spatial position changes, and ColorJitter which modifies brightness and contrast levels to accommodate variations in lighting conditions. All augmented images were then converted to tensor format and normalized using mean values [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225] to align with ImageNet dataset statistics, which is an important prerequisite in the transfer learning scheme on the VGG16 architecture. Meanwhile, the validation data only goes through basic pre-processing stages, which include standardizing the image size to 224×224 pixels, converting it to a tensor, and normalizing it without additional augmentation, so that model performance evaluation can be carried out objectively on data that represents the original conditions. This pre-processing approach is designed to ensure stable training convergence, gradient optimization efficiency, and improved model generalization capabilities in the task of classifying the ripeness of bananas [9].

2.3 Modelling

VGG16 (Visual Geometry Group) is a deep Convolutional Neural Network (CNN) architecture developed by K. Simonyan and A. Zisserman, consisting of a total of 16 weighted layers, including 13 convolutional layers and 3 Fully Connected layers. This architecture is designed to process image inputs that are generally 224×224 pixels in size using small 3×3 filter kernels, enabling it to produce high-quality image feature representations and achieve a top-5 test accuracy of up to 92.7% on the ImageNet dataset. In modern research implementations, VGG16 is often utilized through transfer learning techniques as a pre-trained model, where network parameters that have been trained on large-scale datasets are reused to solve classification problems on different datasets to improve computational efficiency and mitigate the risk of overfitting. The main advantage of this model lies in the simplicity of its architectural structure combined with its high generalization ability in recognizing complex visual patterns, making it the standard in object classification, detection, and localization tasks [10].

$$F_l^{(k)} = \sigma\left(\sum_{c=1}^{C_{l-1}} W_l^{(k,c)} * F_{l-1}^{(c)} + b_l^{(k)}\right)$$

Where in the VGG16 architecture, where $F_l^{(k)}$ is a is the k -th feature map feature map generated from the convolution between the input feature map $F_{l-1}^{(c)}$ and the convolution kernel $W_l^{(k,c)}$, which is summed across all input channels C_{l-1} and added bias $b_l^{(k)}$. The * operator denotes a two-dimensional convolution process, while the non-linear activation function σ(·) which in VGG16 uses Rectified Linear Unit (ReLU), serves to improve the model's representational ability in extracting complex visual patterns. With the use of small kernels 3×3 in a multi-level manner, VGG16 is capable of constructing hierarchical feature representations ranging from simple edges and textures to high-level visual characteristics, making it effective for use in image classification tasks, including fruit ripeness detection [11].

2.4 Training & Testing

Training data serves as a reference for learning variable patterns, while test data is used to validate prediction accuracy based on the proximity of feature characteristics to the training data [12], The training focused on optimizing the weight of the adaptive classifier layer using the Stochastic Gradient Descent (SGD) algorithm with momentum [13], which aims to minimize the Cross-Entropy Loss objective function over 25 iterations (epochs). Sequentially, the testing phase is performed at each epoch using a separate validation dataset to monitor the convergence of the accuracy metric and detect potential overfitting, which is ultimately verified through confusion matrix analysis to evaluate the model's classification precision on unseen data [14].

Table 1. Hyperparameter

Hyperparameter	Value
Learning Rate (LR)	0.001
Momentum	0.9
Batch Size	32
Epochs	25

Table 1 shows the training process of the VGG16-based CNN model using hyperparameter configurations designed to achieve a balance between convergence stability and optimal classification performance. The learning rate is set to 0.001 to control the speed of weight updates so that the optimization process proceeds gradually and stably without causing excessive gradient oscillations. A momentum parameter of 0.9 is applied to accelerate convergence while helping the model escape local minima during the training process. A batch size of 32 is selected to maintain a balance between computational efficiency and gradient estimation accuracy at each training iteration. Furthermore, the training process was carried out for 25 epochs to give the model sufficient opportunity to learn the visual patterns of banana ripeness optimally, while minimizing the risk of underfitting and overfitting. This combination of hyperparameter settings contributed to the achievement of stable and accurate classification performance at the evaluation stage.

2.5 Evaluation

The evaluation stage is a fundamental step in validating the reliability of the post-training model to ensure the stability and accuracy of the resulting classification performance. Quantitative analysis is performed using a Confusion Matrix that maps the components of True Positive, True Negative, False Positive, and False Negative as the basis for calculating the correct prediction ratio [15]. Next, Precision and Recall metrics are used to measure the accuracy and sensitivity of classification in each class, while F1-Score serves as a harmonic mean that comprehensively represents the balance of model performance. To strengthen performance validation and evaluate the model's discrimination capabilities in greater depth, the analysis also includes a Precision–Recall (PR) curve. The PR curve highlights the trade-off between Precision and Recall, particularly in conditions of unbalanced class distribution. The combination of evaluation metrics and curve analysis enables an objective and comprehensive assessment of the model's generalization quality before the system is further implemented. [16].

3. Results and Discussion

This study presents the results of experiments and discussions on the performance of the VGG16 CNN model in classifying the ripeness of bananas. Evaluations were conducted to assess the accuracy, stability, and generalization capabilities of the model based on test data, while also examining its relationship with the pre-processing and modeling methods used [17].

3.1 Data Collection

The dataset collected for this study consisted of 5,616 images obtained through the Roboflow Universe Platform. In order for the system to recognize specific visual characteristics of the fruit, the dataset was classified into six maturity classes, namely: ripe, unripe, overripe, rotten, and raw.

3.2 Pre-Processing

The pre-processing stage plays an important role in determining the quality of classification results in a CNN VGG16-based banana ripeness detection system[18]. In this phase, images are prepared through a series of standardized transformations to ensure compatibility with model architecture requirements and improve generalization capabilities during the inference process. Image dimension normalization is performed to maintain spatial representation consistency, while data augmentation is applied to enrich visual variation and reduce the risk of overfitting due to data distribution limitations. In addition, ImageNet-based statistical pixel intensity normalization serves to stabilize the gradient optimization process. With this combination of strategies, the pre-processing stage becomes an important foundation for producing accurate, stable, and reliable model performance in the evaluation stage [19].

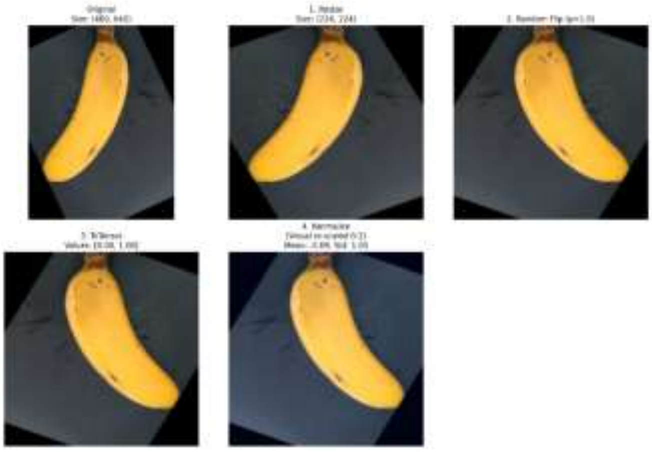

Figure 2. Augmentation Results

Figure 2 shows the image pre-processing steps applied before model training, including adjusting the image size to 224×224 pixels, augmentation through horizontal flipping, conversion to tensor format with pixel value normalization, and statistical intensity standardization to support the stability and consistency of CNN model learning.

3.1 VGG16 Architecture

VGG16 architecture was employed in this study as the primary model for analyzing the performance of banana ripeness classification in the results and discussion stages. VGG16 is a deep Convolutional Neural Network (CNN) composed of 13 convolutional layers and 3 fully connected layers, utilizing small 3×3 kernels in a sequential manner to hierarchically extract visual features. The use of an ImageNet-based pre-trained model enables a more efficient transfer learning process and enhances the model's ability to generalize to unseen data. In this section, the performance of VGG16 is analyzed based on training outcomes, evaluation of classification metrics, and visualization of model outputs to assess the effectiveness of the architecture in supporting the banana ripeness classification system. [20].

Table 2. Classification Report

Kelas	Precision	Recall	F1-Score	Support
Freshripe	0.9574	0.9890	0.9730	182
Freshunripe	1.0000	1.0000	1.0000	127
Overripe	0.9752	0.9704	0.9728	203
Ripe	0.9160	0.9083	0.9121	120
Rotten	0.9854	0.9740	0.9797	346
Unripe	0.9697	0.9697	0.9697	33
Accuracy	-	-	0.9713	1011
Macro Avg	0.9673	0.9686	0.9679	1011
Weighted Avg	0.9714	0.9713	0.9713	1011

Table 2 shows that the VGG16-based Convolutional Neural Network (CNN) model has excellent classification performance with an overall accuracy of 97.13%. High precision, recall, and F1-score values in almost all classes indicate that the model is capable of effectively extracting and representing the visual characteristics of banana ripeness. The fresh unripe class achieved perfect performance with precision, recall, and F1-score values of 1.0000, indicating that the visual characteristics of fresh unripe bananas can be clearly distinguished by the model. In addition, the fresh ripe, overripe, and rotten classes also showed very stable performance with F1-scores above 0.97, indicating the model's ability to recognize consistent visual patterns at these levels of ripeness. The macro average F1-score of 0.9679 and weighted average F1-score of 0.9713 further confirm that the model's performance remains optimal in both relatively balanced and unbalanced data distribution conditions.

However, relatively lower performance was observed in the ripe class, which had a precision value of 0.9160 and an F1-score of 0.9121. This condition indicates classification errors that are likely influenced by the similarity of visual characteristics between physiologically adjacent classes, particularly between the ripe and overripe classes. In addition, the unripe class has a very limited amount of data (support = 33), which can affect the stability of the evaluation even though the precision, recall, and F1-score values in this class remain high. Overall, these evaluation results indicate that the VGG16 CNN model has strong and stable generalization capabilities in classifying the ripeness of bananas, and has the potential to be further improved through the addition of data and augmentation strategies that are more focused on classes with high visual similarity.

Figure 3. Model Performance Metrics by Class

Figure 3 shows the performance of the VGG16-based CNN model for each banana ripeness class based on precision, recall, and F1-score metrics. The visualization results show that the model achieves very high and consistent performance across almost all classes, with metric values generally above 0.95. The freshunripe class shows the best performance with precision, recall, and F1-score values of 1.00. Meanwhile, the ripe class has relatively lower performance compared to other classes, indicating similarities in visual characteristics with adjacent classes. Overall, the VGG16 model shows excellent and stable classification capabilities across various levels of banana ripeness.

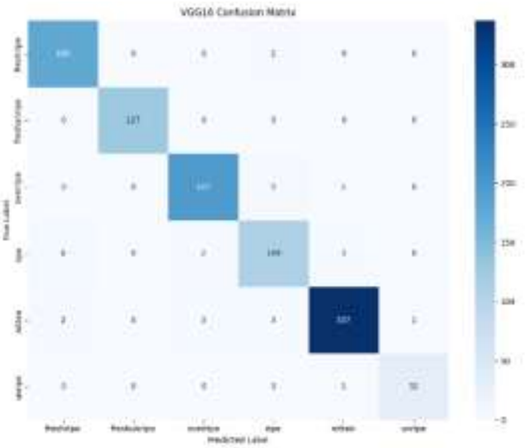

Figure 4. Confusion Matrix

Figure 4 shows the confusion matrix of the VGG16-based CNN model, indicating that most samples were classified correctly, as shown by the dominance of values on the main diagonal. The freshunripe class was classified perfectly without any prediction errors, while the rotten class also showed a very high rate of correct predictions with a minimal number of errors. Classification errors occur relatively more frequently in classes with similar visual characteristics, particularly between the overripe and ripe classes, and a small number of errors between the ripe and freshripe classes. These findings indicate that the VGG16 CNN model has good feature discrimination capabilities, with a low error rate that is still concentrated in classes with visual similarities.

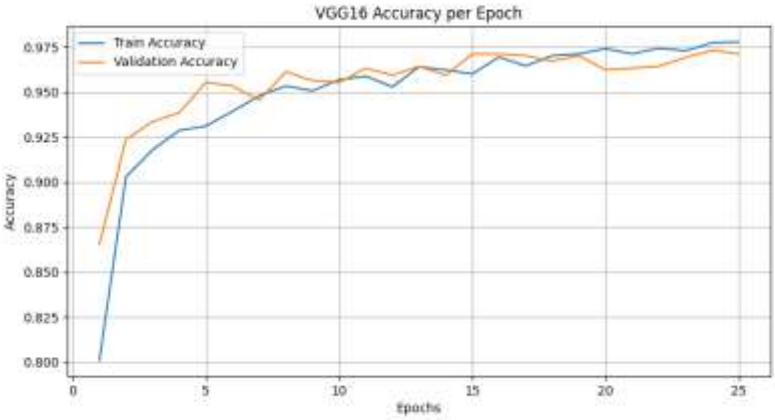

Figure 5. Accuracy per Epoch

Figure 5 shows the training and validation accuracy graphs of the VGG16-based CNN model at each epoch. It can be seen that the training data accuracy increases sharply in the early epochs, then gradually increases until it reaches a stable condition close to the maximum value in the next epoch, indicating that the model convergence process is going well. The validation accuracy also shows a consistent upward trend with relatively small fluctuations throughout the training process, without any significant decline. The difference in accuracy between the training data and the validation data appears to be relatively small and stable, indicating that the model does not experience significant overfitting. Overall, this curve pattern reflects the stability of the learning process and the good generalization ability of the VGG16 CNN model to data that has not been trained.

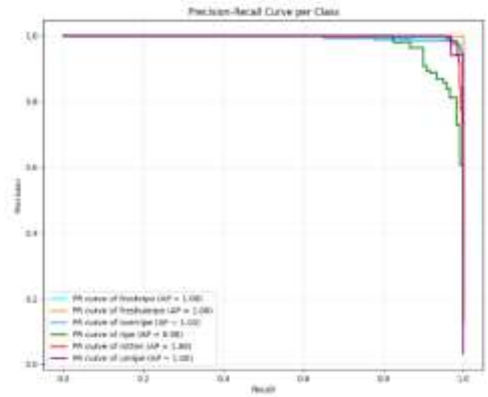

Figure 6. Precision-Recall Curve per Class

Figure 6 shows the Precision–Recall Curve for each banana ripeness class in the VGG16-based CNN model. The curve shows that almost all classes have very high performance, as indicated by the curve being close to the upper right corner of the graph. The Average Precision (AP) value reaches 1.00 in the freshripe, freshunripe, overripe, rotten, and unripe classes, indicating the model's ability to maintain high precision at various recall levels. The ripe class shows a slight decline in performance with an AP value of 0.98, which is reflected in a decrease in precision at very high recall values. This pattern indicates classification challenges due to visual similarities with other nearby classes. Overall, this curve confirms that the VGG16 model has excellent discrimination capabilities and a low error rate in classifying banana ripeness levels.

4. Conclusion

This study concludes that the implementation of Convolutional Neural Network (CNN) using the VGG16 architecture proved to be very effective in classifying six levels of banana ripeness, with an overall accuracy of 97.13% supported by a data augmentation strategy to prevent overfitting. An in-depth analysis through the confusion matrix and Precision-Recall curve shows that the model has strong and stable generalization capabilities, where the freshunripe class was detected perfectly, although there were still slight challenges in classifying the ripe class due to its high visual similarity to the overripe class. The model's success in maintaining performance consistency between training and validation data confirms that the transfer learning approach with SGD optimization applied is capable of producing a reliable detection system. Thus, this proposed model is recommended as a computer vision-based solution to support the efficiency and objectivity of post-harvest sorting processes in smart farming ecosystems.

References

[1] A. Hermawan, T. R. Effendi, and N. Fadillah, "Deteksi kematangan buah pisang berdasarakan kulit menggunakan metode multi-level thresholding dan ycbcr," vol. 02, no. 02, pp. 105–108, 2021.

[2] F. Huda and M. P. K. Putra, "Klasifikasi jenis buah pisang menggunakan algoritma convolutional neural network," vol. 1, no. 3, pp. 100–105, 2023.

[3] G. D. Cahyo, R. Helilintar, and I. N. Farida, "Pengolahan citra kematangan pisang menggunakan convolutional neural network vgg19," *INOTEK*, vol. 9, pp. 1507–1515, 2025.

[4] M. R. Pratama and J. Aryanto, "Deteksi dan klasifikasi kematangan pisang menggunakan yolov8 dan model hibrida mobilenetv2 – knn," vol. 9, no. 3, pp. 787–796, 2025, doi: 10.29408/edumatic.v9i3.32597.

[5] A. Rizky *et al.*, "Perbandiangan vgg16 dan mobilenetv2 untuk klasifikasi tingkat kematangan buah apel," pp. 30–39.

[6] S. Real-time and R. Artikel, "Implementasi algoritma yolov8 (you only look once) untuk deteksi jenis buah pisang," vol. 15, pp. 9–17, 2026.

[7] I. Najiah, I. Hariyanti, A. R. Sanjaya, A. R. Sanjaya, K. P. Celak, and D. K. Pisang, "Deteksi jenis dan kematangan pisang menggunakan metode extreme learning," vol. 2, no. 2, pp. 232–242, 2020.

[8] N. Sary, H. Handoko, S. Pasaribu, R. J. Situmeang, and R. Darmawan, "Implementasi algoritma yolov11 dan roboflow untuk deteksi tingkat kematangan anggur berbasis web," pp. 264–274, 2025, doi: 10.47002/metik.v9i2.1121.

[9] D. Liandaputra and A. Zahra, "Klasifikasi tingkat kematangan buah nanas menggunakan metode deep learning," vol. 8, no. 2, pp. 1091–1103, 2024.

[10] U. N. Putra *et al.*, "Klasifikasi tingkat kematangan buah jeruk menggunakan arsitektur vgg16," 2025.

[11] R. Suhendra, R. S. Ayu, R. S. Qaisa, I. Juliwardi, and N. Astrianda, "Penerapan cnn arsitektur vgg16 untuk deteksi kesegaran ikan berdasarkan citra digital," pp. 19–25, 2025.

[12] A. I. Syirojuddin, A. A. Riadi, A. Susanto, U. M. Kudus, K. Kudus, and J. Tengah, "Klasifikasi tingkat kematangan parijoto berdasarkan karakteristik warna menggunakan algoritma k-nearest neighbor," pp. 30–40, 2025.

[13] F. Tangguh and Y. Islami, "Analisis performa algoritma stochastic gradient descent (sgd) dalam mengklasifikasi tahu berformalin," vol. 3, no. 1, pp. 1–8, 2022.

[14] M. Rijal, A. Muhammad Yani, and A. Rahman, "Deteksi citra daun untuk klasifikasi penyakit padi menggunakan pendekatan deep learningdengan model cnn," vol. 10, no. 1, pp. 56–62, 2024.

[15] M. Kamil, T. Endra, and E. Tju, "Naïve bayes dan confusion matrix untuk efisiensi analisa intrusion detection system alert," vol. 02, pp. 81–88, 2022.

[16] L. Lusiana, A. Wibowo, and K. Dewi, "Implementasi algoritma deep learning you only look once (yolov5) untuk deteksi buah segar dan busuk," vol. 11, no. 1, pp. 123–130, 2023.

[17] H. Jayadi and A. Nurul Fajar, "Klasifikasi pisang berbasis algoritma vgg16melalui metode cnn deep

learning," vol. 15, no. 1, pp. 1–17, 2023.

[18] A. I. Hanifah and A. Hermawan, "Klasifikasi kematangan pisang menggunakan metode convolutional neural network classification of banana fruit maturity levels using the convolutional neural network method," vol. 12, no. 148, pp. 169–176, 2023, doi: 10.34010/komputika.v12i2.9999.

[19] D. Husen, "Evaluasi teknik augmentasi data untuk klasifikasi tumor otak menggunakan cnn pada citra mri," pp. 219–228, 2024.

[20] R. Dzikri Fauzi and M. Fachrie, "Sistem prediksi tingkat kematangan buah tomat mengguanakan arsitektur vgg16," vol. 5, no. 3, pp. 401–414, 2024, doi: 10.46576/djtechno.