



Enhanced InceptionV3 Transfer Learning with Augmentation Strategy for Multi-Class Fruit Classification

Abdul Karim¹, Fakhri Lambardo², Rizqi Elmuna Hidayah³

¹Faculty of Science and Technology, Department of Information Technology, Universitas Labuhanbatu, Rantauprapat, Indonesia

²Faculty of Computer Science, Informatics Study Program, Universitas Sjakhyakirti, Palembang, Indonesia

³Faculty of Information Technology, Information Systems Program, Universitas Islam Kalimantan Muhammad Arsyad Al-Banjari Banjarmasin, Banjarmasin, Indonesia

*abdulkarim@ulb.ac.id

Abstract

Fruit variety recognition using digital images is an important component in developing automated systems for agricultural handling and food-industry processing. The task is difficult because different fruit types may present nearly identical visual patterns, while image acquisition factors such as illumination, viewing position, and image quality can reduce classification reliability. To address this issue, this study designed a deep learning model for 131 fruit classes by adapting InceptionV3 through a transfer learning scheme. The pre-trained feature extraction layers were retained without retraining, while the original output structure was replaced with task-specific layers consisting of global average pooling, a 1024-unit dense layer with ReLU activation, and a softmax classifier. The image data were standardized to 224×224 pixels, augmented to increase visual variation, and divided into training, validation, and testing subsets using an 80:10:10 ratio. The proposed model produced an accuracy of 99.80%, with precision, recall, and F1-score values of 0.9900. These results exceeded the performance of GoogLeNet, ResNet, and VGGNet, showing that the use of pre-trained InceptionV3 features, customized classification layers, and augmentation can improve prediction consistency and reduce classification errors. Further evaluation on unconstrained real-world images and optimization for real-time use are recommended for future development.

Keywords: CNN; fruit classification; image classification; InceptionV3; transfer learning

How to Cite: A. Karim, F. Lambardo, and R. E. Hidayah, "Enhanced InceptionV3 Transfer Learning with Augmentation Strategy for Multi-Class Fruit Classification", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 3, pp. 756 - 766, Jun. 2026.

Permalink/DOI: <https://doi.org/10.29207/resti.v10i3.7531>

Received: January 27, 2026

Accepted: May 23, 2026

Available Online: June 21, 2026

*This is an open-access article under the [CC BY 4.0 License](#)
Published by [Ikatan Ahli Informatika Indonesia](#)*

1. Introduction

The increasing adoption of information technology and artificial intelligence (AI) has accelerated the development of automated solutions in many sectors, including agriculture and the food industry [1], [2]. In these sectors, the ability to identify and classify fruit varieties accurately has become an important requirement for supporting supply chain management, product quality control, and operational efficiency in agricultural distribution and trade [3-5]. As digital agriculture continues to generate larger and more complex visual data, automated fruit recognition systems are increasingly needed to provide fast, consistent, and accurate classification results [6], [7]. Compared with manual inspection, image-based classification offers a more scalable alternative because

it reduces human dependency, processing time, and subjective errors [8], [9]. Nevertheless, fruit classification remains challenging when different varieties share similar visual properties, particularly in terms of shape, color, and surface texture. This problem becomes more difficult when images are affected by variations in illumination, viewpoint, resolution, and overall image quality [10], [11].

Visual similarity among fruit varieties is one of the main obstacles in image-based classification tasks [12]. Subtle differences between fruit classes may be difficult to capture, especially when the dataset contains heterogeneous images collected under different conditions [13-15]. Changes in lighting, camera position, image resolution, and object quality may alter the visual cues used by the model during the learning

process. In addition, environmental factors may cause deformation, color variation, or texture changes in fruit samples, further increasing the complexity of the recognition task [16], [17]. Therefore, a more adaptive deep learning approach is required to extract discriminative features from visually similar fruit varieties while maintaining robustness against image variability.

Convolutional Neural Networks (CNNs) have been widely used for image classification because of their ability to learn hierarchical visual representations from image data. Several CNN architectures, including VGGNet [18-21], ResNet [22-24], and GoogLeNet [25-28], have shown strong capability in handling complex visual recognition problems. ResNet introduces residual connections that help overcome the vanishing gradient problem, allowing the network to learn from deeper and more complex image representations [27], [29], [30]. GoogLeNet applies the Inception mechanism to capture multi-scale features, enabling the model to recognize both local details and broader contextual information within an image. Meanwhile, VGGNet uses a simple and structured convolutional design with small filters to produce rich feature representations. The performance of these architectures can be further strengthened through transfer learning and data augmentation, which help improve model adaptability under variations in lighting, viewpoint, and image quality [31-33].

Several previous studies have demonstrated the effectiveness of deep learning architectures in different image classification domains. Chang (2024) [34] applied MobileNet to identify implant-ridge relationships and reported an accuracy of 95.28% on the complete dataset and 99.21% under specific conditions. The study showed that MobileNet can provide efficient classification performance and is suitable for implementation on devices with limited computational resources, which is useful in clinical dentistry. However, the study also had limitations because it used only a single cross-sectional CBCT slice, which may not fully represent the three-dimensional structure of ridge defects. Furthermore, the analysis was limited to fully healed posterior mandibular ridges, indicating the need for further evaluation across different anatomical regions and unhealed ridge conditions.

Al-Khater (2024) [18] also investigated deep learning-based image classification by applying 3D VGG-16 and 3D ResNet-18 with the FABEMD technique for malware detection. The method achieved strong classification results, with VGG-16 obtaining 96.14% and ResNet-18 achieving 99.64% on the Malimg dataset, while the corresponding results on the MaleVis dataset reached 98.60% and 99.46%. These findings indicate that deep learning models can effectively handle complex classification problems, including previously undetected malware. However, the study also emphasized that model performance depends heavily on the availability of balanced and sufficiently large datasets. In addition, although FABEMD

increased the amount of training data, image modification may change the original characteristics of malware samples and increase the possibility of misclassification.

Previous studies confirm that deep learning, particularly CNN-based models, can achieve strong results in image classification and has been increasingly applied in agricultural image analysis [35-37]. Architectures such as VGGNet, ResNet, and GoogLeNet have proven effective in learning complex visual patterns. However, maintaining high classification accuracy remains difficult when the data contain visually similar objects and high image variability [38-41]. In the context of fruit variety classification, many existing studies have not specifically focused on designing a model that can distinguish classes with highly similar shapes, colors, and textures. This gap indicates the need for a classification model that combines high accuracy, robustness to image heterogeneity, and practical relevance for agricultural image processing and food industry applications [42], [43].

Although CNNs have shown strong potential in image-based pattern recognition, standard architectures may still be insufficient when applied to complex and heterogeneous fruit image datasets [4], [44]. For this reason, combining CNN architectures with transfer learning, data augmentation, and architectural optimization is considered a relevant strategy for improving classification performance [11], [45], [46]. Transfer learning allows the model to reuse pre-trained visual representations, reducing the need for extensive training while improving feature extraction capability. Therefore, this study develops a deep learning model based on InceptionV3, transfer learning, and data augmentation to improve the accuracy, robustness, and efficiency of fruit classification across 131 classes. The proposed model is expected to address the challenge of distinguishing visually similar fruit varieties and to support the development of automated classification systems for agriculture and the food industry.

2. Methods

This study adopts a quantitative experimental design to develop and assess a deep learning-based model for image classification. The main objective is to evaluate the model's capability in recognizing and classifying different fruit varieties based on visual image features. To measure its performance objectively, the proposed model is benchmarked against several widely used CNN architectures, namely VGGNet, ResNet, and GoogLeNet. The evaluation is conducted using standard classification metrics, including accuracy, precision, recall, and F1-score. The experimental process uses an image dataset containing multiple fruit varieties, which is first processed through a preprocessing stage to produce consistent and suitable input data. This stage plays an important role in improving data quality, supporting the learning

process, and enhancing the overall classification performance of the model.

2.1 Dataset

The dataset in this study comprises fruit images with diverse visual characteristics, such as variations in shape, color, surface texture, and overall appearance. The images were collected from a publicly accessible Kaggle repository and represent fruit samples captured under different lighting conditions, viewpoints, and image resolutions. This variability is important because it reflects practical classification conditions, where differences between fruit varieties are often subtle and visually difficult to separate. The dataset includes several fruit categories, ranging from common fruits such as apples, bananas, and oranges to more specific varieties, including green grapes, kiwi, and red mangoes. The image samples were divided proportionally into training, validation, and testing subsets to support model development and enable objective performance evaluation. Representative examples of the fruit images used in this study are shown in Figure 1.

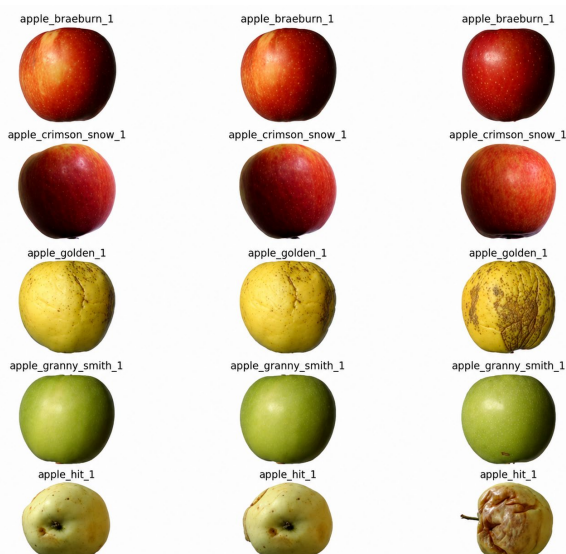


Figure 1. Images of Various Fruit Varieties in the Dataset

This research utilized an image dataset of fruits and vegetables sourced from an open-access Kaggle repository. The dataset comprises 90,483 images grouped into 131 classes, representing diverse visual properties such as shape, color, texture, and general appearance, as illustrated in Figure 1. From this dataset, 67,692 images were used for training and 22,688 images for testing. The classes include widely known fruits, such as apples, bananas, and oranges, as well as more specific varieties, including green grapes, red mangoes, kumquat, and persimmon. A structured naming system is used in the dataset, allowing different varieties of the same fruit type to be treated as independent class labels.

The source images were originally provided at 100×100 pixels and were collected under controlled

acquisition conditions. During image capture, the fruit samples were placed on a motorized rotating platform and recorded using a Logitech C920 webcam against a white background to facilitate object separation. To improve the clarity of the samples, a flood-fill-based segmentation process was applied. This process identifies background regions from the image boundaries by measuring color similarity and then converts those regions into a uniform white background. As a result, the fruit objects are more clearly separated from the background and become more suitable for model training.

Before being processed by the classification models, all images were resized to 224×224 pixels to satisfy the input format required by the CNN architectures. Data augmentation was then applied through several transformations, including rotation, horizontal flipping, zooming, and shifting. These transformations were used to increase image variability and strengthen the model's robustness to changes in lighting, object orientation, and camera perspective. Furthermore, images from the "test-multiple_fruits" folder, captured with a Nexus 5X smartphone, were included to evaluate how well the model performs under less controlled and more realistic imaging conditions.

2.2 Proposed Model

The proposed model is designed by adapting and improving the GoogLeNet-based InceptionV3 architecture to handle the classification of fruit varieties that share similar visual patterns. Several architectural components are refined to strengthen the model's ability to learn detailed and discriminative image features. In particular, the convolutional and pooling mechanisms are utilized to capture complex visual representations that are important for distinguishing fruit classes with subtle differences in shape, color, and texture.

In addition to architectural refinement, the model applies transfer learning, extensive data augmentation, and hyperparameter optimization to improve classification performance. Transfer learning enables the model to reuse feature representations learned by InceptionV3 from large-scale image data, allowing more efficient adaptation to the fruit dataset while reducing training time. The complete structure of the proposed model is illustrated in Figure 2.

Figure 2 presents the visual architecture of the modified InceptionV3 model. To improve dataset diversity, an extensive data augmentation strategy is applied, including rotation, flipping, zooming, shifting, and contrast adjustment. These techniques enhance the model's robustness against variations in lighting conditions, camera angles, and image quality. Hyperparameter tuning is also performed to optimize the learning process and improve model accuracy and stability.

A key modification in the proposed model is the freezing of early layers in the original GoogLeNet-

based InceptionV3 architecture, followed by the addition of custom layers tailored to the fruit classification task. These custom layers enable the

model to focus on domain-specific features while preserving the general feature extraction capability of InceptionV3.

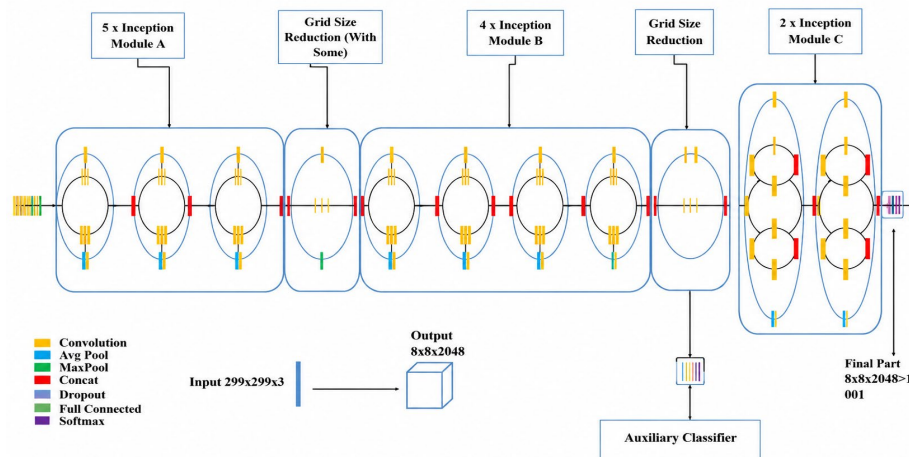


Figure 2. Proposed Method Architecture

The detailed architectural modifications, including frozen layers and additional customized layers, are summarized in Table 1. This structured approach helps

the model maintain high classification accuracy while improving its adaptability and robustness in real-world fruit classification scenarios.

Table 1. Proposed Model Overview

Aspect	Description
Base Architecture	The proposed architecture is built on InceptionV3, which belongs to the GoogLeNet family. This backbone is selected because its Inception blocks can process image information at several feature scales within the same network structure.
Transfer Learning Strategy	The feature extractor is initialized using weights obtained from large-scale image training, such as ImageNet. By using these learned representations, the model can begin training with existing visual knowledge instead of learning all features from the beginning.
Frozen Early Layers	The lower-level layers of the InceptionV3 backbone are kept fixed during training. These layers are excluded from weight updates so that their previously learned visual representations remain unchanged.
Rationale for Freezing	The early convolutional blocks usually detect basic image cues, including edges, textures, contours, and simple geometric patterns. Because these features are generally useful across different image datasets, they can be reused without additional retraining.
Added Custom Layers	The default classification part of InceptionV3 is replaced with a new head designed for fruit classification. Global Average Pooling is used to summarize feature maps and reduce model complexity. This is followed by a 1024-neuron dense layer with ReLU activation to learn task-related patterns, and a softmax layer to produce class probabilities.
Custom Layer Configuration	The output features from InceptionV3 are passed into a Global Average Pooling operation, followed by a 1024-unit dense layer using ReLU activation. The final layer applies softmax activation, with the number of output neurons adjusted to the total number of fruit classes.
Optimizer and Loss Function	The training process uses Adam as the optimizer because it can adjust parameter updates adaptively during learning. For the objective function, categorical cross-entropy is applied since the task involves classification into multiple fruit categories.
Advantages of the Proposed Model	This design improves efficiency by reducing parameter complexity through Global Average Pooling. It also benefits from reusable visual features learned by the pre-trained backbone, while the added classification layers allow the model to focus on visual characteristics that are specific to fruit varieties.

Table 1 presents the technical configuration used in the proposed architecture. The model employs InceptionV3 from the GoogLeNet family as the main feature-learning backbone because its Inception blocks can process image information at different spatial scales. To shorten the training process and retain transferable visual knowledge, the backbone is initialized with pre-trained weights, while the early feature-extraction layers are kept fixed. These lower layers are preserved because they commonly encode general image cues, including edges, contours,

textures, and simple shapes, which remain useful for fruit image recognition.

The default classification component of InceptionV3 is replaced with a task-oriented head for fruit variety classification. This head applies Global Average Pooling to summarize feature maps and reduce model complexity, followed by a 1024-neuron dense layer with ReLU activation to learn more specific fruit-related patterns. The final softmax layer generates class probabilities for multi-class prediction. The training process uses the Adam optimizer together with categorical cross - entropy loss to support stable

learning. Overall, this configuration is designed to reduce unnecessary trainable parameters, improve generalization, and strengthen the model's ability to distinguish visually similar fruit varieties.

2.3. Method Framework

This study is arranged as an experimental workflow for developing and validating an image-based fruit classification model. The proposed system is designed to learn subtle visual cues, including color distribution, shape structure, and surface texture, while remaining robust to image variability caused by illumination changes, camera viewpoint, and image quality. Based on these conditions, the research process is divided into several sequential stages.

The workflow shown in Figure 3 begins with dataset preparation. In this stage, fruit images are collected

from public sources and processed to obtain standardized input data. The images are resized to 224×224 pixels, cleaned through background removal, and enriched using augmentation techniques such as rotation, flipping, zooming, and shifting. The next stage focuses on constructing the proposed CNN architecture by combining transfer learning, customized classification layers, and hyperparameter adjustment. After that, the model is trained and evaluated using standard classification indicators, namely accuracy, precision, recall, and F1-score. In the final stage, the proposed model is compared with VGGNet, ResNet, and GoogLeNet to determine its effectiveness in classifying fruit varieties with challenging visual similarities.

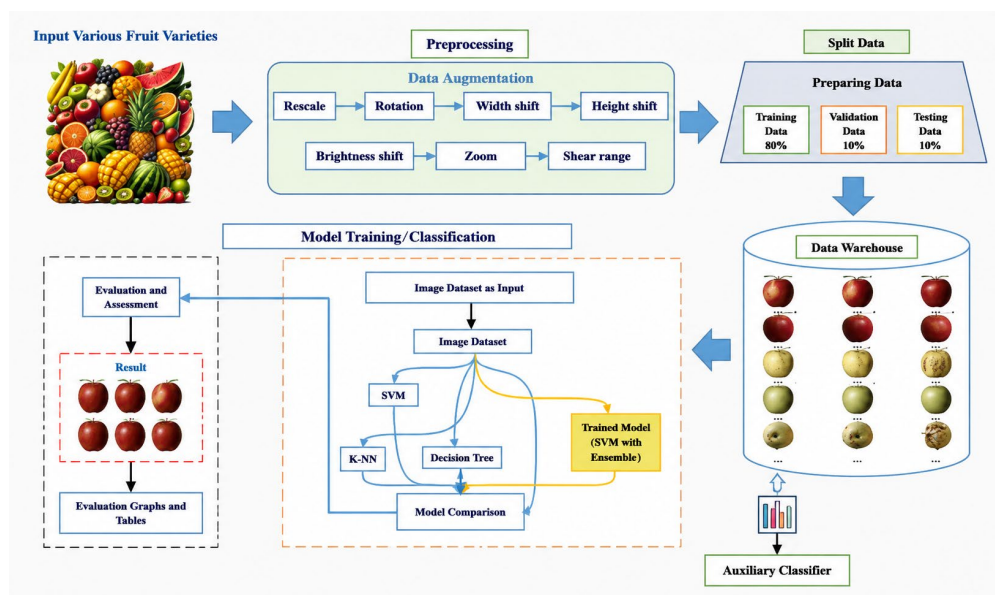


Figure 3. The Method Framework

3. Results and Discussions

3.1 Data Preprocessing and Augmentation

The image dataset used in this experiment contains fruit samples with diverse visual properties, including differences in color, surface pattern, shape, and overall appearance. Before the training stage, preprocessing was carried out to make the input data more consistent, while augmentation was applied to expand the visual variation of the training images. Pixel normalization was performed through rescaling so that the model could process the image values more stably. Rotation and zoom transformations were introduced to simulate changes in viewing distance and angle, whereas width and height shifting allowed the model to learn fruit objects appearing in different image positions. Horizontal flipping was also applied to increase orientation diversity.

Through these transformations, the training data became more varied and better suited for improving model robustness. Figure 4 shows examples of

augmented images generated from the original fruit samples. The augmented variations help the model avoid depending on fixed object positions or specific image orientations. As a result, the model can learn more general fruit patterns and become more adaptable when handling images captured under different visual conditions.

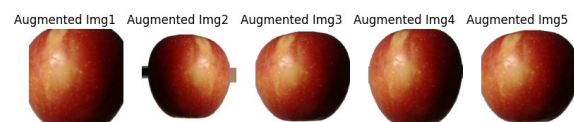


Figure 4. Sample Images of Augmented Data

After the augmentation process, the dataset was separated into training, validation, and testing subsets. The training subset, representing 80% of the data, was used to build the model's ability to recognize visual patterns from fruit images. The validation subset, representing 10%, was used during training to observe model behavior and control overfitting. The remaining 10% was assigned as the testing subset to measure the

final performance of the trained model on unseen samples. Stratified sampling was applied so that each subset maintained proportional class representation, allowing the evaluation process to remain fair and consistent across all fruit varieties.

3.2 Model Training Classification

Four CNN architectures were evaluated in this study: ResNet, GoogLeNet, VGGNet, and the Proposed Model. The purpose of this comparison was to observe how each architecture learned visual patterns from the

fruit dataset and how effectively it performed the classification task. Figure 5 presents the accuracy curves during training, which show the development of model performance across epochs.

Figure 5 presents the accuracy development of ResNet, GoogLeNet, VGGNet, and the Proposed Model throughout the training process. The curves show that each architecture learns at a different rate because of differences in network structure and optimization behavior.

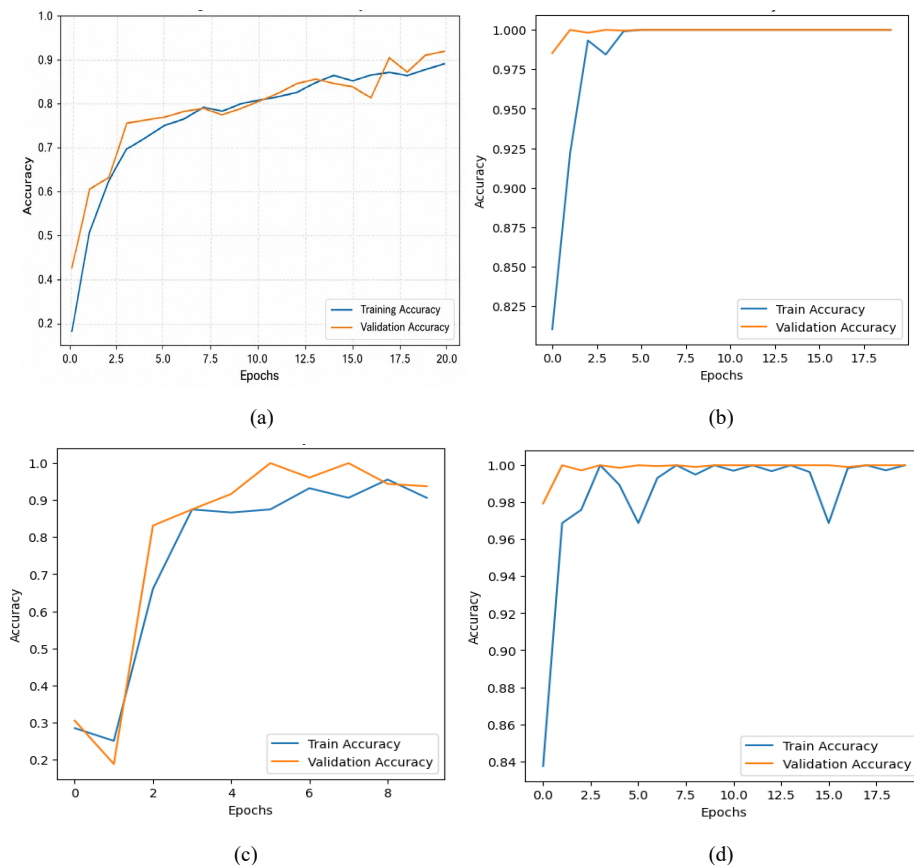


Figure 5. Training Accuracy Curve: (a) ResNet, (b) GoogleNet, (c) VGGNet, (d) Proposed Model

ResNet (Figure 5.a) improves gradually in both training and validation accuracy, indicating a stable learning trend and acceptable generalization. However, its slower increase suggests that additional epochs are needed before the model reaches its best performance. GoogLeNet (Figure 5.b) obtains high validation accuracy during the early training stage, showing rapid convergence. Even so, the early flattening of its validation curve may indicate that the model reaches its learning capacity quickly and has limited improvement in later epochs. VGGNet (Figure 5.c) shows unstable accuracy at the beginning of training before becoming more consistent. This early variation suggests that regularization techniques, such as dropout or weight decay, may be useful for improving its training stability.

The Proposed Model (Figure 5.d) shows the most stable and effective accuracy pattern among the evaluated models. It reaches high accuracy within

fewer epochs and maintains a close distance between the training and validation curves. This indicates that the model learns efficiently while reducing the possibility of overfitting. The use of transfer learning, customized classification layers, and augmentation strategies helps the model capture more relevant fruit features and improve classification performance. Compared with VGGNet, the Proposed Model produces a smoother accuracy curve, and compared with ResNet, it reaches strong performance more quickly.

Figure 6 shows the loss curves of the training and validation processes for each model. These curves describe how classification errors decrease as training progresses. ResNet (Figure 6.a) produces a steady reduction in loss, which indicates that the model gradually learns useful image patterns. However, the slight increase in validation loss during the later epochs suggests a possible early sign of overfitting.

GoogLeNet (Figure 6.b) reduces loss quickly at the beginning of training, with both training and validation loss approaching very low values. This result confirms its fast learning ability, although the very low validation loss also suggests early convergence. VGGNet (Figure 6.c) decreases loss more slowly and

shows fluctuations, especially in the initial epochs. Although the loss continues to decline, the early difference between training and validation loss indicates that VGGNet requires better training stability to achieve stronger generalization.

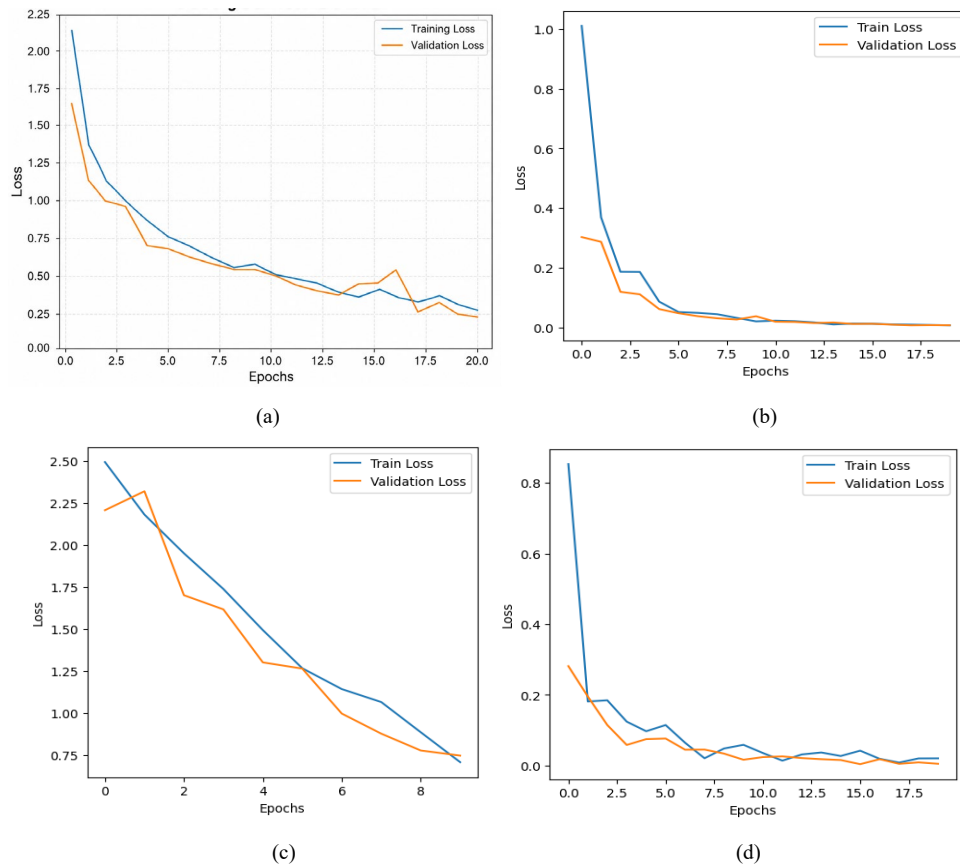


Figure 6. Training Loss Curve: (a) ResNet, (b) GoogleNet, (c) VGGNet, (d) Proposed Model

The Proposed Model (Figure 6.d) demonstrates the most consistent loss reduction among the evaluated architectures. Its training and validation loss curves remain close throughout the training process, indicating that the model can reduce classification errors without losing its ability to perform well on unseen images. Compared with the other models, the validation loss of the Proposed Model is lower and more stable, showing that the applied transfer learning scheme, refined layer configuration, and augmentation strategy contribute to a more controlled and effective learning process.

Figure 7 presents the confusion matrix results for ResNet, GoogLeNet, VGGNet, and the Proposed Model. These matrices provide a detailed view of how each model assigns image samples to the correct or incorrect fruit classes. Correct predictions are concentrated along the main diagonal, while values outside the diagonal represent classification mistakes. Therefore, the confusion matrix helps reveal how well each architecture separates fruit categories with similar visual appearances.

ResNet (Figure 7.a) produces moderate classification results, as several samples are still assigned to incorrect classes outside the main diagonal. This indicates that ResNet has difficulty separating fruit varieties with close visual resemblance. GoogLeNet (Figure 7.b) performs better, with most predictions appearing along the diagonal, suggesting stronger feature learning and better generalization. VGGNet (Figure 7.c) shows less consistent results, with more errors than GoogLeNet, indicating that it may be less effective in learning complex fruit-related visual patterns.

The Proposed Model (Figure 7.d) achieves the clearest classification pattern among all models. Most predictions are concentrated on the main diagonal, and only a small number of samples are misclassified. This result shows that the model can distinguish fruit varieties more accurately. The improvement is supported by the combination of transfer learning, fine-tuned classification layers, and optimized augmentation. Overall, the confusion matrix confirms that the Proposed Model provides the most reliable classification performance, while ResNet produces

moderate results and VGGNet shows higher instability.

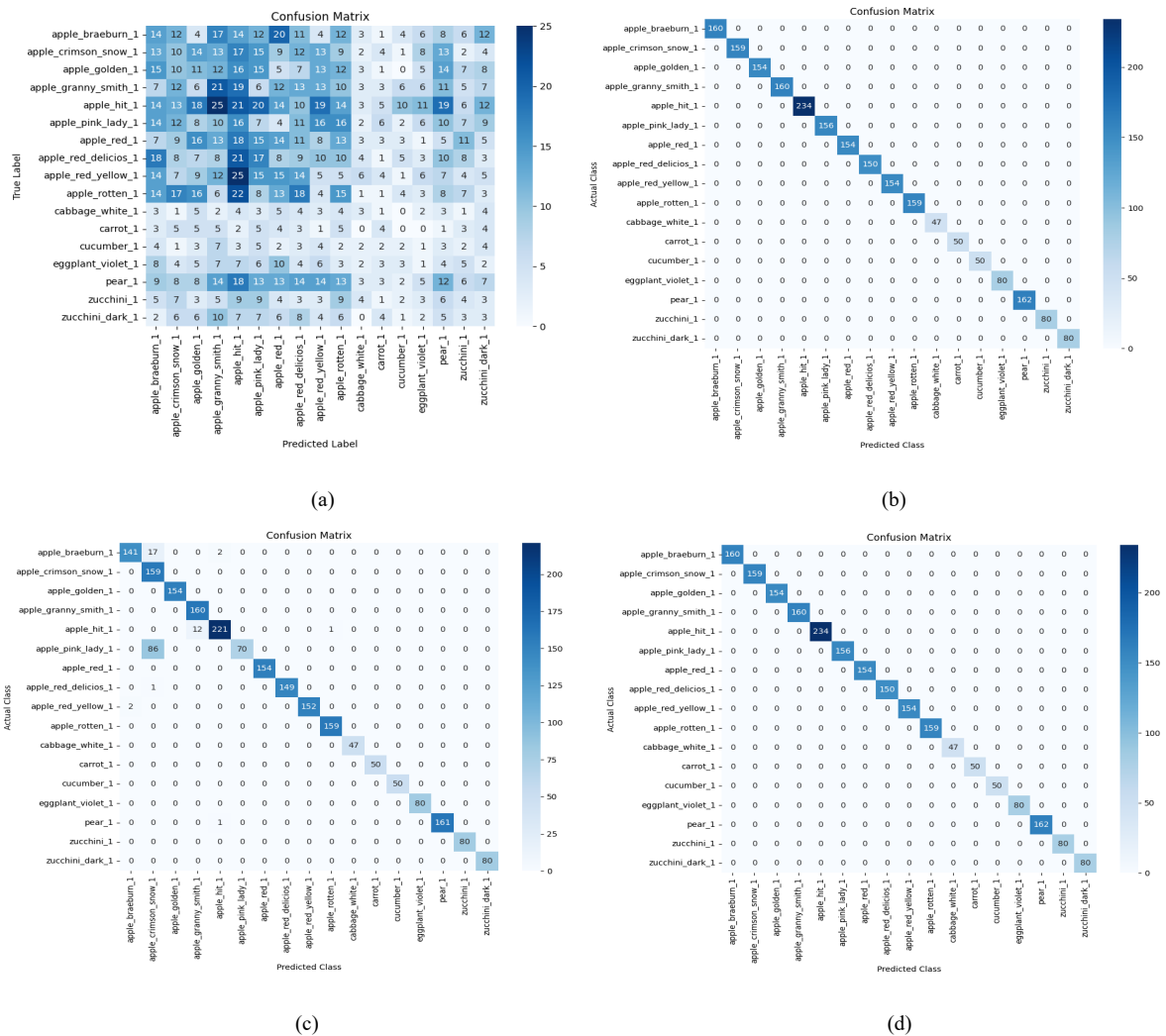


Figure 7. Confusion Matrix (a) ResNet, (b) GoogleNet, (c) VGGNet, (d) Proposed Model

3.3 Discussion

This section discusses the comparative performance of the four evaluated architectures, namely ResNet, GoogLeNet, VGGNet, and the Proposed Model, in

classifying different fruit varieties. The results indicate that each model produces a distinct learning pattern, which influences its classification accuracy, convergence behavior, and ability to generalize to unseen fruit images.

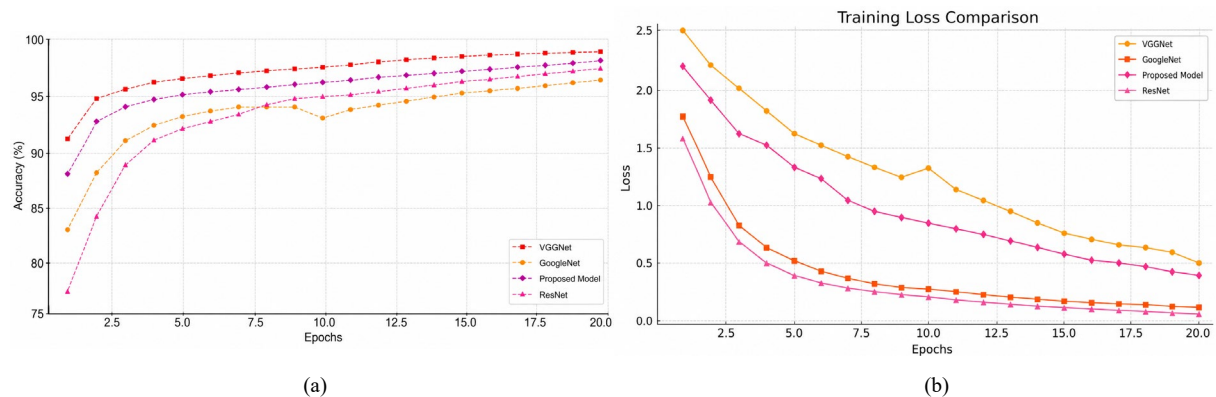


Figure 8. Comparison of Training Results: (a) Accuracy; (b) Loss

Figure 8 compares the training performance of VGGNet, GoogLeNet, ResNet, and the Proposed Model in terms of accuracy and loss. As shown in

Figure 8(a), the Proposed Model achieves the most consistent improvement in accuracy and reaches the highest final performance among all tested

architectures. GoogLeNet also shows rapid learning during the early epochs, but its accuracy curve becomes relatively stable after several iterations. ResNet demonstrates gradual improvement, indicating a slower but steady learning process. Meanwhile, VGGNet records the lowest accuracy trend, suggesting that it requires more training effort to extract discriminative features from the fruit image dataset.

Figure 8(b) illustrates how the loss values of the four evaluated architectures change during training. The proposed architecture shows the most favorable loss pattern, as its loss decreases more quickly and remains lower than the other models across most epochs. This indicates that the model can reduce prediction errors more effectively during the learning process. GoogLeNet also produces relatively low loss values, although its curve tends to become stable earlier. ResNet shows a more gradual loss reduction, suggesting that it requires longer training to approach optimal convergence. In contrast, VGGNet maintains higher and more fluctuating loss values, indicating less stable learning behavior.

The combined accuracy and loss trends in Figure 8 confirm that the proposed architecture achieves the strongest balance between rapid convergence, high classification accuracy, and stable error reduction. These results indicate that the integration of transfer learning, customized layers, and data augmentation enables the model to learn fruit-related visual patterns more effectively than the baseline CNN architectures. The testing performance of ResNet, GoogLeNet, VGGNet, and the Proposed Method is summarized in Table 2.

Table 2. Performance Evaluation of All Models

Model	Recall	Precision	F1-Score	Accuracy
Propose Method	0,9900	0,9900	0,9900	0,9980
GoogleNet	0,9800	0,9800	0,9800	0,9870
ResNet	0,9200	0,9100	0,9100	0,9120
VGGNet	0,8800	0,8700	0,8700	0,8850

Table 2 reports the testing results of all evaluated models using recall, precision, F1-score, and accuracy. The Proposed Method produces the highest scores, with recall, precision, and F1-score values of 0.9900 and an accuracy of 0.9980. These results show that the model can classify fruit varieties with very high consistency and a low error rate.

GoogLeNet ranks second, achieving 0.9800 for recall, precision, and F1-score, with an accuracy of 0.9870. Although this result indicates strong classification capability, it remains slightly below the Proposed Method. ResNet obtains moderate performance, with an accuracy of 0.9120, while VGGNet shows the weakest result, with an accuracy of 0.8850. The lower performance of ResNet and VGGNet suggests that these models are less effective in representing fine visual differences among fruit varieties. Overall, the results demonstrate that the proposed transfer learning-based architecture provides the most reliable

performance among the tested models. Figure 9 shows several prediction examples generated by the Proposed Method.

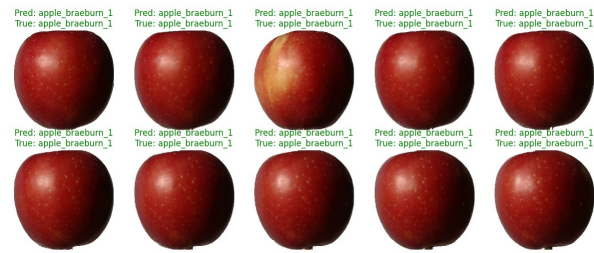


Figure 9. Classification Results of the Proposed Model

Figure 9 presents several classification examples produced by the Proposed Method. Each image is accompanied by the predicted label and the actual label, allowing direct observation of the model's prediction correctness. In the displayed samples, all fruit images are classified correctly, as indicated by the matching predicted and true labels. This result shows that the model can recognize the visual characteristics of the tested fruit samples accurately.

Although the displayed examples confirm strong prediction performance on this subset, further evaluation remains necessary using images captured under more varied real-world conditions. Since the dataset was mainly obtained in a controlled setting, additional testing with uncontrolled lighting, background variation, and camera viewpoints is important to assess the model's practical generalization ability.

4. Conclusions

This study proposed and evaluated a CNN-based deep learning model for fruit variety classification using image data. The experimental results show that the proposed architecture achieved the best overall performance compared with ResNet, GoogLeNet, and VGGNet. The model obtained an accuracy of 99.80%, with precision, recall, and F1-score values of 0.9900, indicating highly accurate and consistent classification with a very low misclassification rate.

The superior performance of the proposed model is supported by the integration of transfer learning, refined convolutional components, and extensive data augmentation. These strategies enable the model to capture detailed visual characteristics and improve its ability to distinguish fruit classes with similar appearances. The comparative results further indicate that GoogLeNet achieved competitive performance but tended to converge earlier, ResNet required a longer training process, and VGGNet produced lower results due to weaker feature representation.

The confusion matrix, accuracy curve, and loss curve confirm that the proposed architecture learns stably, converges efficiently, and generalizes well to unseen data. Therefore, the model offers a promising solution for automated fruit classification, particularly for

agricultural monitoring, post-harvest sorting, and food industry automation. Future research should evaluate the model using more diverse real-world images, optimize the architecture for real-time implementation, and incorporate explainable artificial intelligence to improve model transparency and interpretability.

References

- [1] R. A. Saputra and K. M. Lhaksana, "Prediction Retweet Using User-Based and Content-Based with Artificial Neural Network-Harmony Search," vol. 5, no. 2, pp. 407–413, 2023, doi: [10.47065/bits.v5i2.4079](https://doi.org/10.47065/bits.v5i2.4079).
- [2] Z. Khan, "Modeling and simulation of Maxwell nanofluid flows in the presence of Lorentz and Darcy–Forchheimer forces: toward a new approach on Buongiorno's model using artificial neural network (ANN)," *European Physical Journal Plus*, vol. 138, no. 1, 2023, doi: [10.1140/epjp/s13360-022-03583-w](https://doi.org/10.1140/epjp/s13360-022-03583-w).
- [3] A. P. Windarto, T. Herawan, and P. Alkhairi, "Prediction of Kidney Disease Progression Using K-Means Algorithm Approach on Histopathology Data," in *Artificial Intelligence, Data Science and Applications*, Cham: Springer Nature Switzerland, 2024, pp. 492–497, doi: https://doi.org/10.1007/978-3-031-48465-0_66.
- [4] M. V. Appalanaidu, "Rice plant nutrient deficiency classification using modified MOBILENET convolutional neural network," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 14, no. 1, 2023, doi: [10.1142/S1793962322430036](https://doi.org/10.1142/S1793962322430036).
- [5] I. Shafi, "Internal defects detection and classification in hollow cylindrical surfaces using single shot detection and MobileNet," *Measurement: Journal of the International Measurement Confederation*, vol. 202, 2022, doi: [10.1016/j.measurement.2022.111836](https://doi.org/10.1016/j.measurement.2022.111836).
- [6] S. Hossain, "Comparison of YOLO (V3,V5) and MobileNet-SSD (V1,V2) for Person Identification Using Ear-Biometrics," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1259–1271, 2024, doi: [10.12785/ijcds.150189](https://doi.org/10.12785/ijcds.150189).
- [7] K. Lavanya, "Real Time Digital Face Mask Detection using MobileNet-V2 and SSD with Apache Spark," *International Journal of Performance Engineering*, vol. 18, no. 8, pp. 598–604, 2022, doi: [10.23940/ijpe.22.08.p8.598604](https://doi.org/10.23940/ijpe.22.08.p8.598604).
- [8] S. Defit, A. P. Windarto, and P. Alkhairi, "Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization," *Telematika*, vol. 17, no. 1, pp. 52–67, 2024, doi: <https://doi.org/10.35671/telematika.v17i1.2824>.
- [9] P. Alkhairi and A. P. Windarto, "Classification Analysis of Back propagation-Optimized CNN Performance in Image Processing," *Journal of Systems Engineering and Information Technology (JOSEIT)*, vol. 2, no. 1, pp. 8–15, 2023, doi: <https://doi.org/10.29207/joseit.v2i1.5015>.
- [10] A. E. K. Gunawan, "Stock Price Movement Classification Using Ensembled Model of Long Short-Term Memory (LSTM) and Random Forest (RF)," *International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2255–2262, 2023, doi: [10.30630/joiv.7.4.1640](https://doi.org/10.30630/joiv.7.4.1640).
- [11] E. Rahimi, "The Efficiency of Long Short-Term Memory (LSTM) in Phenology-Based Crop Classification," *Korean Journal of Remote Sensing*, vol. 40, no. 1, pp. 57–69, 2024, doi: [10.7780/kjrs.2024.40.1.6](https://doi.org/10.7780/kjrs.2024.40.1.6).
- [12] I. Ranggadara, U. M. Buana, A. Ratnasari, and U. M. Buana, "Backpropagation Neural Network for Predict Sugarcane Stock Availability," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 5, pp. 8279–8284, 2020, doi: [10.30534/ijatcse/2020/197952020](https://doi.org/10.30534/ijatcse/2020/197952020).
- [13] W. Chen, "Identification Rice Varieties and Broken Rice Based on GoogLeNet," *Journal of the Chinese Cereals and Oils Association*, vol. 38, no. 2, pp. 146–152, 2023, doi: [10.1016/j.lwt.2024.117175](https://doi.org/10.1016/j.lwt.2024.117175).
- [14] B. Gunawan, M. E. Al-rivan, P. S. Informatika, U. Multi, and D. Palembang, "2ND Mdp Student Conference (MSC) 2023 E-Issn: 2985-7406 Klasifikasi Jenis Beras Putih Menggunakan CNN Residual Network Optimizer SGD," pp. 128–132, 2023, doi: <https://jurnal.mdp.ac.id/index.php/msc/article/view/4305>.
- [15] V. Choudhary, P. Guha, G. Pau, R. K. Dhanaraj, and S. Mishra, "Automatic Classification of Cowpea Leaves Using Deep Convolutional Neural Network," *Smart Agricultural Technology*, p. 100209, 2023, doi: [10.1016/j.atech.2023.100209](https://doi.org/10.1016/j.atech.2023.100209).
- [16] Windarto, A.P., Herawan, T., Alkhairi, P. (2024). Early Detection of Breast Cancer Based on Patient Symptom Data Using Naive Bayes Algorithm on Genomic Data. In: Farhaoui, Y., Hussain, A., Saba, T., Taherdoost, H., Verma, A. (eds) Artificial Intelligence, Data Science and Applications. ICAISE 2023. Lecture Notes in Networks and Systems, vol. 837. Springer, Cham. https://doi.org/10.1007/978-3-031-48465-0_64.
- [17] P. Alkhairi, A. P. Windarto, and M. M. Efendi, "Optimasi LSTM Mengurangi Overfitting untuk Klasifikasi Teks Menggunakan Kumpulan Data Ulasan Film Kaggle IMDB," vol. 6, no. 2, pp. 1142–1150, 2024, doi: [10.47065/bits.v6i2.5850](https://doi.org/10.47065/bits.v6i2.5850).
- [18] W. Al-Khater, "Using 3D-VGG-16 and 3D-Resnet-18 deep learning models and FABEMD techniques in the detection of malware," *Alexandria Engineering Journal*, vol. 89, pp. 39–52, 2024, doi: <https://doi.org/10.1016/j.aej.2023.12.061>.
- [19] T. A. T. K. Azaharan, "Investigation of VGG-16, ResNet-50 and AlexNet Performance for Brain Tumor Detection," *International journal of online and biomedical engineering*, vol. 19, no. 8, pp. 97–109, 2023, doi: <https://doi.org/10.3991/ijoe.v19i08.38619>.
- [20] T. Tang, "Integration of VGGNet-Driven Precipitation Similarity Analysis and Hydrological Modeling for Enhanced Streamflow Forecasting," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 14543–14555, 2024, doi: [10.1109/JSTARS.2024.3441550](https://doi.org/10.1109/JSTARS.2024.3441550).
- [21] J. Shedbalkar, "Deep transfer learning model for brain tumor segmentation and classification using UNet and chopped VGGNet," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 3, pp. 1405–1415, 2024, doi: [10.11591/ijeecs.v33.i3.pp1405-1415](https://doi.org/10.11591/ijeecs.v33.i3.pp1405-1415).
- [22] A. Sharma, "Multilevel detection and classification of diseased plant leaf images using high-resolution superlet transform and E-ResNet," *International Journal of Information Technology (Singapore)*, vol. 16, no. 5, pp. 3135–3147, 2024, doi: [10.1007/s41870-024-01822-0](https://doi.org/10.1007/s41870-024-01822-0).
- [23] H. Sahli, "ResNet-SVM: Fusion based glioblastoma tumor segmentation and classification," *Journal of X-Ray Science and Technology*, vol. 31, no. 1, pp. 27–48, 2023, doi: <https://doi.org/10.3233/XST-221240>.
- [24] S. Shivadekar, "Deep Learning Based Image Classification of Lungs Radiography for Detecting COVID-19 using a Deep CNN and ResNet 50," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 1, pp. 241–250, 2023, doi: [10.1016/j.bspc.2022.103977](https://doi.org/10.1016/j.bspc.2022.103977).
- [25] S. Wang, "Multi-sensor fusion rolling bearing intelligent fault diagnosis based on VMD and ultra-lightweight GoogLeNet in industrial environments," *Digital Signal Processing: A Review Journal*, vol. 145, 2024, doi: [10.1016/j.dsp.2023.104306](https://doi.org/10.1016/j.dsp.2023.104306).
- [26] P. Samundeeswari, "An Efficient Fully Automated Lung Cancer Classification Model Using GoogLeNet Classifier," *Journal of Circuits, Systems and Computers*, vol. 32, no. 14, 2023, doi: [10.1142/S0218126623502468](https://doi.org/10.1142/S0218126623502468).
- [27] F. B. N. Barber, "Human exons and introns classification using pre-trained Resnet-50 and GoogleNet models and 13-layers CNN model," *Journal of Genetic Engineering and Biotechnology*, vol. 22, no. 1, 2024, doi: [10.1016/j.jgeb.2024.100359](https://doi.org/10.1016/j.jgeb.2024.100359).
- [28] W. Luo, "Identification of the Reflected Signal in the COSMIC-2 Occultation Signal Using the Improved GoogLeNet Deep Learning Model," *Chinese Journal of Atmospheric Sciences*, vol. 47, no. 3, pp. 631–641, 2023, doi: [10.3878/j.issn.1006-9895.2202.21096](https://doi.org/10.3878/j.issn.1006-9895.2202.21096).
- [29] F. Ederbali, "Mobilenet, inception ResNet and GoogleNet for epilepsy detection using spectrogram images," *Indonesian*

- Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, pp. 870–877, 2024, doi: [10.11591/ijeecs.v34.i2.pp870-877](https://doi.org/10.11591/ijeecs.v34.i2.pp870-877).
- [30] X. Huang, “Improved Model Based on GoogLeNet and Residual Neural Network ResNet,” *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 16, no. 1, 2022, doi: [10.4018/IJCINL.313442](https://doi.org/10.4018/IJCINL.313442).
- [31] X. Lin, “Application of Data Mining Technology with Improved Clustering Algorithm in Library Personalized Book Recommendation System,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 11, pp. 494–504, 2023, doi: [10.14569/IJACSA.2023.0141151](https://doi.org/10.14569/IJACSA.2023.0141151).
- [32] H. J. Kweon, “Influence of Summer Pruning Time on Shoot Growth and Fruit Quality of ‘Fuji’/M.9 Apple Tree Damaged by the Low Air Temperature at Flowering Period,” *Korean Journal of Environmental Agriculture*, vol. 41, no. 4, pp. 328–334, 2022, doi: <https://doi.org/10.5338/KJEA.2022.41.4.38>.
- [33] H. Wang, “GAN-STD: small target detection based on generative adversarial network,” *Journal of Real-Time Image Processing*, vol. 21, no. 3, 2024, doi: [10.1007/s11554-024-01446-4](https://doi.org/10.1007/s11554-024-01446-4).
- [34] H. C. Chang, “Classification of the implant-ridge relationship utilizing the MobileNet architecture,” *Journal of Dental Sciences*, vol. 19, no. 1, pp. 411–418, 2024, doi: [10.1016/j.jds.2023.08.002](https://doi.org/10.1016/j.jds.2023.08.002).
- [35] E. M. Senan, F. W. Alsaade, M. I. A. Al-Mashhadani, T. H. H. Aldhyani, and M. H. Al-Adhaileh, “Classification of histopathological images for early detection of breast cancer using deep learning,” *Journal of Applied Science and Engineering (Taiwan)*, vol. 24, no. 3, pp. 323–329, 2021, doi: [10.6180/jase.202106_24\(3\).0007](https://doi.org/10.6180/jase.202106_24(3).0007).
- [36] A. Choudhury, “PSSO: Political Squirrel Search Optimizer-Driven Deep Learning for Severity Level Detection and Classification of Lung Cancer,” *International Journal of Information Technology and Decision Making*, 2023, doi: [10.1142/S0219622023500189](https://doi.org/10.1142/S0219622023500189).
- [37] P. Yadlapalli, “Intelligent classification of lung malignancies using deep learning techniques,” *International Journal of Intelligent Computing and Cybernetics*, vol. 15, no. 3, pp. 345–362, 2022, doi: [10.1108/IJICC-07-2021-0147](https://doi.org/10.1108/IJICC-07-2021-0147).
- [38] M. A. Lubis, D. G. S. Saragih, I. D. Anastasia, A. P. Windarto, and P. Alkhairi, “Application of the ANN Algorithm to Predict Access to Drinkable Water in North Sumatra Regency/City,” *International Journal of Informatics and Data Science*, vol. 1, no. 1, pp. 18–25, 2023, doi: <https://doi.org/10.64366/ijids.v1i1.16>.
- [39] M. Mingzhu, “The value of MobileNet in classification of bedside chest radiograph,” *Chinese Journal of Radiology (China)*, vol. 57, no. 12, pp. 1325–1330, 2023, doi: <https://www.mdpi.com/2076-3417/11/6/2751>.
- [40] R. Chen, “Rapid Detection of Multi-QR Codes Based on Multistage Stepwise Discrimination and a Compressed MobileNet,” *IEEE Internet of Things Journal*, vol. 10, no. 18, pp. 15966–15979, 2023, doi: [10.1109/JIOT.2023.3268636](https://doi.org/10.1109/JIOT.2023.3268636).
- [41] O. Bouazizi, “Road Object Detection using SSD-MobileNet Algorithm: Case Study for Real-Time ADAS Applications,” *Journal of Robotics and Control (JRC)*, vol. 5, no. 2, pp. 551–560, 2024, doi: [10.18196/jrc.v5i2.21145](https://doi.org/10.18196/jrc.v5i2.21145).
- [42] A. Zhang, “2DDSRU-MobileNet: an end-to-end cloud-noiserobust lightweight convolution neural network,” *Journal of Applied Remote Sensing*, vol. 18, no. 2, p. 24511, 2024, doi: [10.1117/1.JRS.18.024511](https://doi.org/10.1117/1.JRS.18.024511).
- [43] F. Valeri, “UNet and MobileNet CNN-based model observers for CT protocol optimization: comparative performance evaluation by means of phantom CT images,” *Journal of Medical Imaging*, vol. 10, 2023, doi: [10.1117/1.JMI.10.S1.S11904](https://doi.org/10.1117/1.JMI.10.S1.S11904).
- [44] C. Mai, “MobileNet-Based IoT Malware Detection with Opcode Features,” *Journal of Communications and Information Networks*, vol. 8, no. 3, pp. 221–230, 2023, doi: [10.23919/JCIN.2023.10272350](https://doi.org/10.23919/JCIN.2023.10272350).
- [45] B. Ruhani, “Hydrogen production via renewable-based energy system: Thermoeconomic assessment and Long Short-Term Memory (LSTM) optimization approach,” *International Journal of Hydrogen Energy*, vol. 52, pp. 505–519, 2024, doi: [10.1016/j.ijhydene.2023.03.456](https://doi.org/10.1016/j.ijhydene.2023.03.456).
- [46] S. Liu, “New onset delirium prediction using machine learning and long short-Term memory (LSTM) in electronic health record,” *Journal of the American Medical Informatics Association*, vol. 30, no. 1, pp. 120–131, 2023, doi: [10.1093/jamia/ocac210](https://doi.org/10.1093/jamia/ocac210).