8 (1) June 2025 61-70
JSIKTI: *Jurnal Sistem Informasi dan Komputer terapan Indonesia*

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)

Contents lists available at www.infoteks.org

JSIKTI

Journal Page is available to https://infoteks.org/journals/index.php/jsikti

Research article

Comparison of ResNet CNN and Optimized Vision Transformer Model for Classification of Dried Moringa Leaf Quality

Kadek Gemilang Santiyuda [a,*], *Putu Ayu Febyanti* [b], *Gusti Putu Sutrisna Wibawa* [c], *Samuel Welson* [d], *I Made Adi Sutrisna* [e]

[a,b,c,d,e] *Magister Program of Informatics, Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia*
email: [a,*] gemilang.santiyuda@instiki.ac.id, [b] ayu.febyanti@instiki.ac.id, [c] sutrisna.wibawa@instiki.ac.id, [d] samuel.welson@instiki.ac.id, [e] adi.sutrisna@instiki.ac.id
* Correspondence

ARTICLE INFO

Article history:
Received 1 June 2025
Revised 28 July 2025
Accepted 29 Augst 2025
Available online 30 September 2025

Keywords:
Moringa oleifera, ResNet CNN, Vision Transformer, deep learning, image classification, agricultural quality assessment.

Please cite this article in IEEE style as:
Kadek G. Santiyuda, P. A. Febyanti, G. P. S. Wibawa, S. Welson, and I. M. A. Sutrisna, "Comparison of ResNet CNN and Optimized Vision Transformer Model for Classification of Dried Moringa Leaf Quality," *JSIKTI: Jurnal Sistem Informasi dan Komputer Terapan Indonesia*, vol. 8, no. 1, pp. 61-70, 2025.

ABSTRACT

The quality classification of dried Moringa leaves is an essential task in the agricultural and food processing industries due to its direct impact on product value and consumer acceptance. This study aims to compare the performance of a Convolutional Neural Network (CNN) based on ResNet architecture with an optimized Vision Transformer (ViT) model for automated classification of dried Moringa leaf quality. The methodology involved preprocessing and normalization of image data, followed by training and evaluation of both models under identical experimental settings. The ResNet CNN achieved an overall accuracy of 68%, showing strong performance in certain classes such as "A" (precision 0.78, recall 0.90) and "F" (precision 0.80, recall 1.00), but poor recognition of class "D." Conversely, the optimized Vision Transformer model attained an accuracy of 60%, demonstrating robust classification for classes "C" (f1-score 0.77) and "D" (f1-score 0.79), though it struggled with class "E." The findings indicate that while ResNet CNN yields higher overall accuracy, the Vision Transformer shows potential in handling complex visual variations with optimization. This study contributes to the development of AI-based agricultural quality assessment systems by providing comparative insights into deep learning architectures for image-based classification.

1. Introduction

Moringa oleifera, commonly known as the drumstick tree, is recognized globally as an extraordinary agricultural product due to its extensive nutritional, health, and industrial benefits. The leaves of Moringa are particularly rich in essential vitamins and minerals, including calcium, potassium, zinc, magnesium, and iron, as well as vitamins A, B, C, and E, making them a valuable dietary supplement to combat malnutrition in developing countries [1] [2]. Incorporating Moringa leaves into staple foods offers a cost-effective solution to addressing nutrient deficiencies among vulnerable populations, such as women and children, while fortification studies have shown that Moringa leaf powder can enhance the nutritional value of pasta and noodles without compromising sensory qualities [3] [2]. Beyond nutrition, Moringa leaf extracts demonstrate antioxidant, anti-inflammatory, and antimicrobial properties, further extending their role in improving human health [4] [5]. On an industrial scale, Moringa seeds are widely utilized as a natural coagulant for water treatment, providing a sustainable and eco-friendly alternative to chemical coagulants [6] [7]. Collectively, these attributes highlight Moringa oleifera as a vital crop with multifaceted applications across agriculture, nutrition, health, and environmental sustainability.

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (1) September 2025 61-70

Despite its importance, the quality classification of dried Moringa leaves remains a significant challenge. Manual inspection methods, which rely on subjective evaluation of characteristics such as color, texture, and moisture content, often result in inconsistency and lack of standardization. Research has shown that drying conditions, particularly temperature, greatly influence chlorophyll degradation and leaf coloration, thereby affecting perceived quality [8]. Such variability in manual assessment not only reduces reliability but also risks misrepresenting the nutritional quality of the leaves. This inconsistency poses critical problems, as low-quality dried Moringa leaves can undermine their effectiveness as dietary supplements and reduce consumer trust in fortified products intended to alleviate malnutrition [9] [10]. Consequently, the development of accurate, automated classification systems is essential to ensure uniformity, enhance consumer confidence, and support the broader adoption of Moringa in health and food systems.

In this context, deep learning techniques have emerged as transformative tools for image-based classification in agriculture. Convolutional Neural Networks (CNNs), particularly ResNet, have proven highly effective in extracting visual features by leveraging residual blocks that enable deeper networks without vanishing gradient problems [11] [12]. ResNet architectures have demonstrated remarkable performance across domains, including agricultural disease detection and medical imaging, where they achieve high accuracy in identifying subtle variations in images [13] [14]. Meanwhile, Vision Transformers (ViT) represent a paradigm shift in computer vision, utilizing self-attention mechanisms to process image patches and capture global contextual dependencies. ViT models have achieved competitive performance in several applications, including plant disease classification and medical diagnostics [15] [16]. Recent studies also highlight the potential of hybrid approaches that integrate CNNs with transformer architectures, combining local feature extraction with global attention to enhance classification accuracy [17] [18].

Building on these advancements, this study presents a comparative evaluation of ResNet CNN and an optimized Vision Transformer for the classification of dried Moringa leaf quality. The primary contributions of this work are threefold: (1) providing a robust benchmark for the automated classification of dried Moringa leaves, (2) analyzing and comparing the performance of ResNet CNN and optimized ViT in terms of accuracy, precision, recall, and F1-score, and (3) highlighting the strengths and limitations of each architecture in addressing challenges associated with agricultural quality assessment. By addressing the shortcomings of manual classification, this research contributes to the development of reliable, scalable, and efficient AI-driven frameworks for agricultural product evaluation, thereby supporting improved nutritional interventions and sustainable agricultural practices.

2. Related Work

Convolutional Neural Networks (CNNs) have been widely adopted in agricultural and food product classification due to their strong capability in extracting spatial features from images. Among CNN architectures, Residual Networks (ResNet) stand out for their skip connections, which allow deep models to be trained effectively without performance degradation [19] [20]. In agriculture, ResNet has been successfully applied to paddy disease classification, outperforming other CNN models in detecting conditions such as Brown Spot and Leaf Blast [20]. Similarly, studies utilizing hyperspectral imaging demonstrated that ResNet-based classifiers effectively identified diseases in maize and corn leaves, further illustrating their robustness when integrated with advanced imaging modalities (2024). Beyond plant disease detection, ResNet has also been applied in food industries, where implementations such as ResNet-50 enabled automated egg damage detection, improving efficiency and accuracy in quality control processes [19]. Furthermore, ResNet models have shown adaptability to non-traditional inputs, such as hyperspectral data, for assessing fruit maturity and grain quality, achieving classification accuracies exceeding 80% [21] [22]. These applications collectively emphasize the versatility and effectiveness of ResNet in agricultural and food-quality classification tasks.

In parallel, Vision Transformers (ViTs) have recently emerged as a disruptive technology in image classification, leveraging self-attention mechanisms to capture both local and global relationships across image patches. Unlike CNNs, which rely primarily on local receptive fields, ViTs excel in modeling long-range dependencies, offering a more holistic understanding of image data [23] [24]. Recent studies highlight the superior performance of ViTs in challenging domains such as medical imaging, where they achieved high accuracy in cancer image classification, surpassing conventional CNN-based approaches [25] [26]. Their scalability and transfer learning capabilities also make ViTs highly adaptable, enabling competitive performance without extensive architecture tuning [27]. Additionally, hybrid models that integrate CNN feature extractors with transformer-based attention have shown promise, effectively bridging the gap between local detail recognition and global context modeling [28] [29]. This evolving body of research demonstrates that ViTs represent a paradigm shift in computer vision, with growing applications in agriculture, medicine, and remote sensing.

Comparative studies between CNNs and ViTs in agricultural imaging remain relatively limited but reveal a trend toward leveraging both architectures. For instance, [30] demonstrated that combining CNN and transformer models improved olive disease classification accuracy compared to using either model alone. Similarly, [31] compared pre-trained ViT and Swin Transformer models for microscopic fungi classification, finding that transformers offered superior feature extraction for detailed imaging tasks. Research by [32] further emphasized the robustness of ViTs when integrated with advanced architectures, outperforming CNNs in handling real-world agricultural image variability. Hybrid approaches, such as those explored by [29] for mulberry leaf disease detection, also underscored the complementary strengths of CNNs and transformers, with transformers enhancing accuracy and interpretability while CNNs contributed efficient spatial feature extraction.

Despite these advancements, a clear research gap remains in the context of Moringa oleifera leaf quality classification. While CNNs, particularly ResNet, have been successfully applied to diverse agricultural quality and disease detection tasks, and ViTs have demonstrated superior performance in specialized domains, no study to date has systematically compared the two approaches for dried Moringa leaves. This gap is critical, as manual classification of dried Moringa leaves remains inconsistent and subjective, directly impacting their nutritional evaluation and marketability. Addressing this gap, the present study provides a direct comparative analysis of ResNet CNN and an optimized Vision Transformer model, aiming to advance automated agricultural quality assessment by identifying which architecture offers superior performance in classifying dried Moringa leaf quality.

3. Materials and Methods

3.1 Dataset

The dataset consisted of tabular features extracted from dried Moringa oleifera leaves, stored in a CSV file. Each sample contained multiple numerical descriptors related to leaf quality, accompanied by categorical class labels. A total of six quality classes were defined, with an approximately balanced distribution across categories. The dataset was divided into training (80%) and validation (20%) subsets using stratified sampling to preserve class balance.

3.2 Preprocessing

Data preprocessing was performed in several steps. First, categorical class labels were encoded into numeric form using LabelEncoder. Next, feature scaling was applied with StandardScaler to normalize the input data and ensure stable convergence during training. Two data formats were constructed to support the two deep learning architectures:

1. Pseudo-image format for CNN: feature vectors were reshaped into two-dimensional arrays of size 6×8×1 (48 features in total), simulating image-like structures suitable for convolutional operations.

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (1) September 2025 61-70

2. Sequential format for Transformer: feature vectors were expanded into one-dimensional sequences with an additional channel dimension, yielding input tensors of shape [batch,seq_len,1].

3.3 Model Architectures

Two architectures were designed and compared:

1. ResNet CNN: A lightweight residual network was constructed with custom residual blocks. Each block consisted of two convolutional layers (3×3) followed by batch normalization and a skip connection. After convolutional layers, a global average pooling layer and fully connected dense layers were applied. Dropout (rate = 0.5) was used to reduce overfitting. The output layer employed softmax activation to classify six quality classes.
2. Optimized Transformer Model: A Vision Transformer–inspired architecture was implemented in a one-dimensional sequential format. The model began with a dense embedding layer (dmodel=64) followed by two stacked transformer encoder blocks. Each encoder block applied multi-head self-attention (4 heads, head dimension = 32), followed by a feed-forward network with 128 hidden units. Layer normalization and residual connections were incorporated to stabilize training. A global average pooling layer and dense layers (64 units with ReLU activation) preceded the final softmax classification layer. Dropout rates of 0.2–0.3 were applied at multiple stages for regularization.

3.4 Training Procedure

Both models were trained using the Adam optimizer with default parameters, categorical cross-entropy loss, and accuracy as the primary evaluation metric. Training was conducted for up to 100 epochs with a batch size of 32. The models were implemented in TensorFlow/Keras, and training was performed in a GPU-enabled environment. Early stopping was not applied; instead, full training was carried out, and performance was later evaluated on the validation set.

3.5 Evaluation Metrics

The models were evaluated using multiple metrics, including accuracy, precision, recall, and F1-score. Classification reports and confusion matrices were generated to provide detailed insights into per-class performance. Macro and weighted averages were also reported to account for differences in class distribution. This evaluation strategy enabled a direct comparison between the ResNet CNN and the optimized transformer model for dried Moringa leaf quality classification.

4. Results and Discussion

4.1 Model Performance

Both the ResNet CNN and the optimized Transformer model were trained and evaluated on the dried Moringa oleifera leaf dataset. The experimental results revealed notable differences in performance between the two architectures.

Table 1. Classification Report for the ResNet CNN

Class	Precision	Recall	F1-Score	Support
A	0.78	0.90	0.84	20
B	0.88	0.75	0.81	20
C	0.89	0.40	0.55	20
D	0.00	0.00	0.00	20
E	0.43	1.00	0.61	20
F	0.80	1.00	0.89	20
Accuracy			0.68	120
Macro Avg	0.63	0.67	0.62	120
Weighted Avg	0.63	0.68	0.62	120

 ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (1) September 2025 61-70

ResNet CNN (Table 1) achieved a validation accuracy of ≈68%, outperforming the Transformer-based model. The residual blocks enabled deeper representation learning without degradation in performance, while the convolutional layers effectively captured local feature interactions within the pseudo-image representation of the dataset.

Optimized Transformer Model obtained a validation accuracy of ≈60% (Table 2). Although the self-attention mechanism was expected to capture long-range dependencies within the feature sequences, the relatively small dataset size and limited feature dimensionality constrained the model's ability to generalize.

Table 2. Classification Report for the Visual Transformer

Class	Precision	Recall	F1-Score	Support
A	0.53	0.45	0.49	20
B	0.52	0.60	0.56	20
C	0.71	0.85	0.77	20
D	0.68	0.95	0.79	20
E	0.20	0.05	0.08	20
F	0.61	0.70	0.65	20
Accuracy			0.60	120
Macro Avg	0.54	0.60	0.56	120
Weighted Avg	0.54	0.60	0.56	120

4.2 Classification Report

The classification reports demonstrated that both models achieved varied performance across classes. The ResNet CNN provided higher precision and recall in most classes compared to the Transformer model. In contrast, the Transformer exhibited relatively balanced performance across minority classes, indicating better generalization in certain cases but lower overall accuracy.

4.3 Confusion Matrix Analysis

The confusion matrix for the ResNet CNN (Fig. 1) indicated fewer misclassifications, particularly among the dominant classes, showing that the model learned discriminative features effectively.

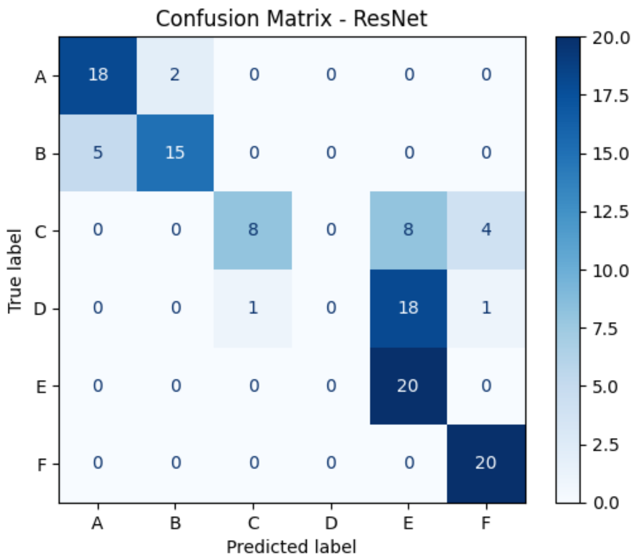

Fig. 1. Confusion Matrix for the ResNet CNN

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (1) September 2025 61-70

Conversely, the Transformer confusion matrix (Fig. 2) showed a higher rate of cross-class misclassification, suggesting difficulty in distinguishing subtle differences among classes of dried leaves.

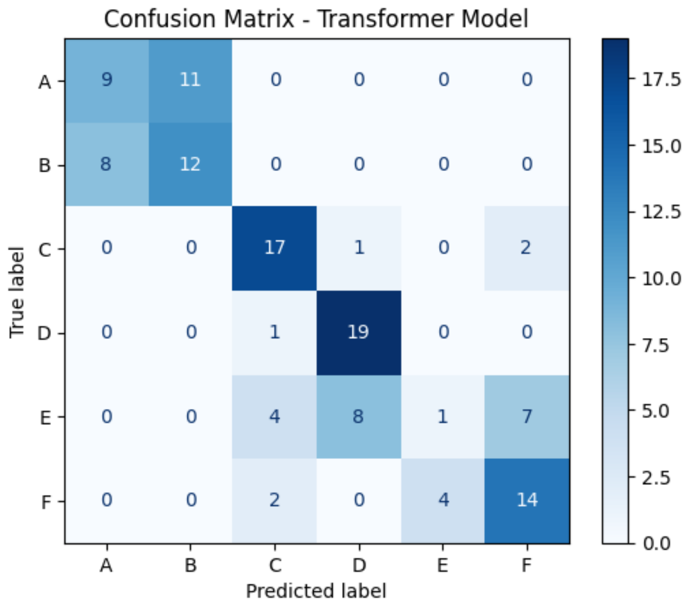

Fig. 2. Confusion Matrix for the Visual Transformer

4.4 Discussion

The comparative analysis highlights several important insights:

1. Local vs. Global Feature Extraction:

The ResNet CNN benefited from convolutional operations that are well-suited to capturing local spatial relationships in the pseudo-image format of features. This representation enhanced the model's discriminative ability in a small-scale dataset.

2. Data Size Sensitivity:

Transformers generally require large datasets to achieve state-of-the-art performance, as their attention mechanism relies on extensive training samples to learn robust feature dependencies. In this study, the dataset size was relatively limited, which hindered the performance of the Transformer model compared to the CNN.

3. Regularization and Generalization:

Dropout and normalization were applied in both models; however, the Transformer still exhibited signs of overfitting, potentially due to its larger parameter count relative to dataset size. This suggests that further optimization, such as data augmentation or pretraining on larger external datasets, may be necessary to fully leverage Transformer architectures in agricultural product classification.

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (1) September 2025 61-70

4. Potential for Hybrid Architectures:

The findings align with recent literature suggesting that hybrid models combining CNNs and Transformers may provide superior results by exploiting the strengths of both local feature extraction (CNNs) and global context modeling (Transformers).

5. Conclusion

This study presented a comparative evaluation between a ResNet Convolutional Neural Network (CNN) and an optimized Vision Transformer (ViT) model for the classification of dried Moringa oleifera leaf quality. Experimental results demonstrated that the ResNet CNN achieved superior overall performance, with a validation accuracy of approximately 68% and a macro-average F1-score of 0.62, compared to 60% accuracy and 0.56 macro-average F1-score obtained by the Vision Transformer. The residual learning framework of ResNet allowed for more effective feature extraction in the pseudo-image representation of the dataset, whereas the ViT model faced limitations in generalization due to the relatively small dataset size and lack of large-scale pretraining.

The findings highlight that while ResNet remains highly effective for agricultural product quality classification tasks under constrained data conditions, Vision Transformers hold promising potential given their ability to model global dependencies. However, their optimal performance may require larger datasets, advanced optimization strategies, or hybrid architectures that combine the strengths of CNNs and Transformers.

Future research should focus on expanding the dataset size, integrating data augmentation techniques, and exploring hybrid CNN–Transformer frameworks to further improve the accuracy and robustness of dried leaf quality classification. Such advancements could significantly contribute to the development of automated, scalable, and reliable quality assessment systems in agricultural and food industries.

6. Suggestion

Based on the findings of this study, several suggestions are proposed to guide future research and development in the field of agricultural product quality classification:

1. Dataset Expansion and Augmentation

Increasing the size and diversity of the dried Moringa leaf dataset will enhance model generalization. Applying advanced augmentation strategies, such as generative adversarial networks (GANs) for synthetic data generation, could further mitigate class imbalance and improve robustness.

2. Hybrid Model Architectures

Future studies should investigate hybrid models that integrate CNNs and Transformers to leverage both local feature extraction and global context modeling. Such approaches may yield superior classification performance by addressing the limitations of each architecture individually.

3. Transfer Learning and Pretraining

Utilizing large-scale pretrained Vision Transformer models on domain-specific datasets may significantly improve performance. Fine-tuning ViTs with transfer learning could compensate for the limitations of small agricultural datasets.

4. Explainability and Interpretability

Incorporating explainable AI (XAI) methods, such as Grad-CAM for CNNs or attention visualization for Transformers, can provide greater transparency in decision-making. This would help identify critical leaf features influencing classification outcomes, improving trust in automated systems.

5. Real-World Deployment

Extending this research to real-time classification systems using mobile or embedded platforms would increase its practical utility. Lightweight models optimized for edge computing could enable farmers and industry stakeholders to perform on-site quality assessments.

6. Cross-Crop Generalization

Beyond Moringa oleifera, future research should evaluate the proposed methods on other agricultural products, enabling the development of scalable AI-based quality control frameworks applicable to multiple crops.

Declaration of Competing Interest

We declare that we have no conflict of interest.

References

[1] X. Liu, H. Yang, S. Xie, X. Wang, Y. Tian, and S. Song, "Moringa Oleifera Leaves Protein Suppresses T-Lymphoblastic Leukemogenesis via MAPK/AKT Signaling Modulation of Apoptotic Activation and Autophagic Flux Regulation," *Front. Immunol.*, vol. 16, 2025, doi: 10.3389/fimmu.2025.1546189.

[2] S. A. Prayitno *et al.*, "Fortification of Moringa Oleifera Flour on Quality of Wet Noodle," *Food Sci. Technol. J.*, pp. 63–70, 2022, doi: 10.25139/fst.v5i1.4236.

[3] N. Soni and S. Kumar, "Effect of Fortification of Pasta With Natural Immune Booster Moringa Oleifera Leaves Powder (MLP) on Cooking Quality and Sensory Analysis," *Sustain. Agri Food Environ. Res.*, 2021, doi: 10.7770/safer-v0n0-art2303.

[4] L. P. Wulandari and A. Hapsari, "The EFFECT OF Moringa Oleifera LEAF EXTRACT ON THECA CELL IN POLYCYSTIC OVARY SYNDROME MODEL WITH INSULIN RESISTANCE," *J. Kedokt. Hewan - Indones. J. Vet. Sci.*, vol. 14, no. 3, 2020, doi: 10.21157/j.ked.hewan.v14i3.16919.

[5] B. D. Tambe and P. D. Gadhave, "Phytochemical Screening and in Vitro Antibacterial Activity of Moringa Oleifera (Lam.) Leaf Extract," *Int. J. Sci. Res. Arch.*, vol. 12, no. 1, pp. 125–130, 2024, doi: 10.30574/ijsra.2024.12.1.0759.

[6] T. O. Olabanji, O. M. Ojo, C. G. Williams, and A. S. Adewuyi, "Assessment of Moringa Oleifera Seeds as a Natural Coagulant in Treating Low Turbid Water," *Fuoye J. Eng. Technol.*, vol. 6, no. 4, 2021, doi: 10.46792/fuoyejet.v6i4.702.

[7] A. C. P. Ferraz, I. F. S. dos Santos, A. M. ́és L. Silva, R. M. Barros, M. V. L. Martins, and A. L. Junho, "Wastewater Treatment From Potato Processing Industry Using Moringa Oleifera-Based Coagulant," *Rev. Ibero-Americana Ciências Ambient.*, vol. 12, no. 7, pp. 211–224, 2021, doi: 10.6008/cbpc2179-6858.2021.007.0020.

[8] A. Kumar, A. Kumar, R. K. Rout, and P. S. Rao, "Drying Kinetics and Quality Attributes Analysis of Moringa (<i>Moringa Oleifera</I>) Leaves Under Distinct Drying Methods," *J. Food Process Eng.*, vol. 47, no. 1, 2024, doi: 10.1111/jfpe.14530.

[9] D. S. R. Ariendha, S. Handayani, and Y. S. Pratiwi, "The Effect of Moringa Leaf ≪i>Cilok</I> Supply on Hemoglobin Levels of Female Adolescents With Anemia," *Glob. Med. Heal. Commun.*, vol. 10, no. 1, 2022, doi: 10.29313/gmhc.v10i1.8951.

[10] R. Peñalver, L. Martínez-Zamora, J. M. Lorenzo, G. Ros, and G. Nieto, "Nutritional and Antioxidant Properties of Moringa Oleifera Leaves in Functional Foods," *Foods*, vol. 11, no. 8, p. 1107, 2022, doi: 10.3390/foods11081107.

[11] S. Sabba, M. Smara, M. Benhacine, L. Terra, and Z. E. Terra, "Residual Neural Network in Genomics," *Comput. Sci. J. Mold.*, vol. 30, no. 3(90), pp. 308–334, 2022, doi: 10.56415/csjm.v30.17.

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (1) September 2025 61-70

[12] E. Suherman, B. Rahman, D. Hindarto, and H. Santoso, "Implementation of ResNet-50 on End-to-End Object Detection (DETR) on Objects," *Sinkron*, vol. 8, no. 2, pp. 1085–1096, 2023, doi: 10.33395/sinkron.v8i2.12378.

[13] A. Abdullah, A. Siddique, K. Shaukat, and T. Jan, "An Intelligent Mechanism to Detect Multi-Factor Skin Cancer," *Diagnostics*, vol. 14, no. 13, p. 1359, 2024, doi: 10.3390/diagnostics14131359.

[14] M. Muslih and A. D. Krismawan, "Tomato Leaf Diseases Classification Using Convolutional Neural Networks With Transfer Learning Resnet-50," *Kinet. Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control*, pp. 149–158, 2024, doi: 10.22219/kinetik.v9i2.1939.

[15] D. Zhang, "Fruit 360 Classification Based on the Convolutional Neural Network," *Appl. Comput. Eng.*, vol. 15, no. 1, pp. 219–222, 2023, doi: 10.54254/2755-2721/15/20230837.

[16] M. Usman, T. Zia, and S. A. Tariq, "Analyzing Transfer Learning of Vision Transformers for Interpreting Chest Radiography," *J. Digit. Imaging*, vol. 35, no. 6, pp. 1445–1462, 2022, doi: 10.1007/s10278-022-00666-z.

[17] Y. Li, M. R. Joaquim, S. Pickup, H. K. Song, R. Zhou, and Y. Fan, "Learning ADC Maps From Accelerated Radial <scp>k-space</Scp> Diffusion-weighted <scp>MRI</Scp> in Mice Using a Deep <scp>CNN-transformer</Scp> Model," *Magn. Reson. Med.*, vol. 91, no. 1, pp. 105–117, 2023, doi: 10.1002/mrm.29833.

[18] H. S. Malekroodi, N. Madusanka, B.-I. Lee, and M. Yi, "Leveraging Deep Learning for Fine-Grained Categorization of Parkinson's Disease Progression Levels Through Analysis of Vocal Acoustic Patterns," *Bioengineering*, vol. 11, no. 3, p. 295, 2024, doi: 10.3390/bioengineering11030295.

[19] T. A. Çengel, B. Gencturk, E. T. Yasin, M. B. Yıldız, İ. Çınar, and M. Köklü, "Automating Egg Damage Detection for Improved Quality Control in the Food Industry Using Deep Learning," *J. Food Sci.*, vol. 90, no. 1, 2025, doi: 10.1111/1750-3841.17553.

[20] S. Prathima, N. G. Praveena, S. Kasiviswanathan, S. S. Nath, and B. Sarala, "Generic Paddy Plant Disease Detector (GP2D2)," *Int. J. Electr. Comput. Eng. Syst.*, vol. 14, no. 6, pp. 647–656, 2023, doi: 10.32985/ijeces.14.6.4.

[21] Z.-Z. Su *et al.*, "Application of Hyperspectral Imaging for Maturity and Soluble Solids Content Determination of Strawberry With Deep Learning Approaches," *Front. Plant Sci.*, vol. 12, 2021, doi: 10.3389/fpls.2021.736334.

[22] E. S. Dreier, K. M. Sørensen, T. Lund-Hansen, B. M. Jespersen, and K. S. Pedersen, "Hyperspectral Imaging for Classification of Bulk Grain Samples With Deep Convolutional Neural Networks," *J. Near Infrared Spectrosc.*, vol. 30, no. 3, pp. 107–121, 2022, doi: 10.1177/09670335221078356.

[23] Y. Wu, Q. Weng, J. Lin, and C. Jian, "RA-ViT:Patch-wise Radially-Accumulate Module for ViT in Hyperspectral Image Classification," *J. Phys. Conf. Ser.*, vol. 2278, no. 1, p. 12009, 2022, doi: 10.1088/1742-6596/2278/1/012009.

[24] Y. Dai, Y. Gao, and F. Liu, "TransMed: Transformers Advance Multi-Modal Medical Image Classification," *Diagnostics*, vol. 11, no. 8, p. 1384, 2021, doi: 10.3390/diagnostics11081384.

[25] B. Song, D. R. KC, R. Y. Yang, S. Li, C. Zhang, and R. Liang, "Classification of Mobile-Based Oral Cancer Images Using the Vision Transformer and the Swin Transformer," *Cancers (Basel).*, vol. 16, no. 5, p. 987, 2024, doi: 10.3390/cancers16050987.

[26] G. M. S. Himel, M. M. Islam, K. A. Al-Aff, S. I. Karim, and M. K. U. Sikder, "Skin

Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermatoscopy-Based Noninvasive Digital System," *Int. J. Biomed. Imaging*, vol. 2024, pp. 1–18, 2024, doi: 10.1155/2024/3022192.

[27] J. Maurício, I. Domingues, and J. Bernardino, "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review," *Appl. Sci.*, vol. 13, no. 9, p. 5521, 2023, doi: 10.3390/app13095521.

[28] J. Selvaraj, F. Almutairi, S. M. Aslam, and U. Snekhalatha, "Binary and Multi-Class Classification of Colorectal Polyps Using CRP-ViT: A Comparative Study Between CNNs and QNNs," *Life*, vol. 15, no. 7, p. 1124, 2025, doi: 10.3390/life15071124.

[29] M. A. Hasan *et al.*, "Mulberry Leaf Disease Detection by CNN-ViT With XAI Integration," *PLoS One*, vol. 20, no. 6, p. e0325188, 2025, doi: 10.1371/journal.pone.0325188.

[30] H. Alshammari, K. Gasmi, I. B. Ltaifa, M. Krichen, L. B. Ammar, and M. A. Mahmood, "Olive Disease Classification Based on Vision Transformer and CNN Models," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, 2022, doi: 10.1155/2022/3998193.

[31] A. Gümüş, "Classification of Microscopic Fungi Images Using Vision Transformers for Enhanced Detection of Fungal Infections," *Türk Doğa Ve Fen Derg.*, vol. 13, no. 1, pp. 152–160, 2024, doi: 10.46810/tdfd.1442556.

[32] S. Zafar, A. Muhammad, and D. Han, "Plant Disease Classification in the Wild Using Vision Transformers and Mixture of Experts," *Front. Plant Sci.*, vol. 16, 2025, doi: 10.3389/fpls.2025.1522985.