JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 9, No. 1, June 2026, pp. 37~47
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v9i1.5421

A Comparative Analysis of Deep Learning Architectures for The Classification of Madura Sliced Tobacco

Muhammad Thahiruddin[1*], Siti Khotijah[2], Adib El Farras[3], Afton Izzul Hasan[5], Ali Ya'lu S[5]
[1,2,3,4,5,]Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Annuqayah, Indonesia

Article Info

Article history:

Received 03 04, 2026
Revised 04 23, 2026
Accepted 05 26, 2026

Keywords:

Deep Learning
Tobacco Classification
Madura Sliced Tobacco
Computer Vision
Class Imbalance

ABSTRACT

Manual tobacco grading, a critical process that determines market value, is inherently subjective, labor-intensive, and prone to inconsistency. This study investigates the application of deep learning techniques to automate the quality classification of Madura sliced tobacco, a high-value agricultural commodity in Indonesia. A novel, high-quality dataset consisting of 1,065 high-resolution images was developed under standardized lighting and controlled environmental conditions. All images were carefully annotated by a professional tobacco sorter with more than five years of experience and categorized into four quality grades: Grade A (premium), Grade B (medium), Grade C (lower), and Grade X (waste). This research conducts a comparative evaluation of four pre-trained deep learning architectures—MobileNetV3-Small, ResNet18, MobileViTV2, and EfficientNet-B0—which were fine-tuned for the classification task. Model performance was evaluated using a 5-fold cross-validation approach to assess their ability to accurately classify tobacco quality levels. The experimental results show that the lightweight MobileNetV3-Small model achieved the best performance with a mean test accuracy of 56.87% ± 2.71%. Further error analysis indicates that all models performed relatively well on the majority classes but encountered significant difficulties in identifying minority classes due to severe class imbalance within the dataset. These findings demonstrate the feasibility of applying lightweight deep learning models for automated tobacco quality classification. The proposed approach has practical potential to improve the objectivity, consistency, and efficiency of quality assessment processes within the agricultural supply chain, particularly in the tobacco industry. In addition, this study provides an important benchmark for future research and highlights class imbalance as a key challenge that must be addressed to develop reliable and deployable automated grading systems.

Corresponding Author:

Muhammad Thahiruddin
Program Studi Matematika,Fakultas Sains dan Teknologi,
Universitas Annuqayah,Indonesia
Email: muhammad.thahiruddin@ua.ac.id

1. Introduction

The global agricultural sector relies heavily on standardized quality assessment to facilitate trade and ensure product consistency. In the tobacco industry, the process of quality grading is of paramount economic importance, directly influencing the pricing, processing, and final quality of consumer products. Tobacco leaves exhibit significant diversity based on genetic variety, cultivation practices, and post-harvest curing methods, making accurate classification a cornerstone of the supply chain. This classification, or grading, determines the leaf's suitability for different blends and products, thereby establishing its market value. This research focuses specifically on Madura sliced tobacco (tembakau rajangan Madura), a distinct and culturally significant variety from Indonesia, whose economic viability is intrinsically linked to the precision of its quality assessment.

Traditionally, the task of grading tobacco has been the exclusive domain of highly skilled human experts, known as graders. This manual process involves a sophisticated sensory evaluation of the tobacco's physical attributes, including color, texture, shape, and aroma. However, this reliance on human expertise introduces significant challenges. The process is inherently subjective and can be influenced by a range of external factors, such as ambient lighting, as well as internal factors like grader fatigue and personal bias [1]. This subjectivity can lead to inconsistencies in quality assessment, creating inefficiencies and potential disputes within the market. Moreover, the manual process is labor-intensive, costly, and does not scale effectively to meet the demands of large-scale industrial production. These limitations create a compelling need for an objective, consistent, and automated alternative.

The ongoing technological revolution in agriculture, often termed "smart farming" or "precision agriculture," is increasingly leveraging data-driven approaches to optimize efficiency and productivity [2]. Among these technologies, computer vision, particularly when powered by deep learning algorithms, has emerged as a transformative force [3]. Deep learning models, especially Convolutional Neural Networks (CNNs), have demonstrated remarkable success in a variety of agricultural applications, including crop health monitoring, weed detection, and product quality grading [4]. A key advantage of deep learning is its ability to automatically learn and extract hierarchical features directly from raw image data, obviating the need for the complex and often subjective process of manual feature engineering that characterized earlier machine vision systems [5],[6]. This capability makes deep learning exceptionally well-suited for nuanced visual classification tasks like tobacco grading.

A primary obstacle to the widespread adoption of deep learning in specialized agricultural domains is the frequent scarcity of large, meticulously annotated datasets. The process of collecting and labeling thousands of agricultural images requires significant time, resources, and domain-specific expertise.

To overcome this challenge, the technique of transfer learning has become a standard and foundational methodology in the field [7],[8],[9]. Transfer learning involves taking a model that has been pre-trained on a massive, general-purpose image dataset, such as ImageNet, and fine-tuning it for a new, specific task (Guo et al., 2024). This approach leverages the rich feature representations learned from millions of diverse images, allowing the model to achieve high performance on the target task with a much smaller, domain-specific dataset. This not only accelerates the training process but also significantly improves model generalization, making it an ideal strategy for applications like tobacco classification.

The selection of deep learning models for this study was strategic, designed to evaluate a spectrum of architectures representing different trade-offs between computational complexity and performance. The chosen models include:

1. ResNet18: A well-established and powerful CNN architecture renowned for its use of residual connections, which effectively mitigate the vanishing gradient problem in deep networks. It serves as a robust and widely used baseline in agricultural classification tasks [10].
2. MobileNetV3-Small and EfficientNet-B0: These models represent the state-of-the-art in lightweight, efficient architectures. They are specifically designed for high performance on resource-constrained platforms, such as mobile devices, making them ideal candidates for practical, in-field agricultural applications [11].
3. MobileViTV2: A modern hybrid architecture that synergistically combines the local feature extraction capabilities of CNNs with the global context modeling strengths of Vision Transformers (ViTs). It represents the cutting edge of efficient models for mobile vision tasks [12].

This strategic selection allows for an investigation that goes beyond simply identifying a functional model, to instead probe which architectural philosophy is most suitable for a fine-grained agricultural classification task where practical deployment is a key consideration [13].

Therefore, the primary objective of this study is to conduct a rigorous comparative analysis of these four diverse deep learning architectures for the automated classification of Madura sliced tobacco. This analysis is performed on a novel, high-quality image dataset created under controlled conditions and labeled by a domain expert, with the ultimate goal of identifying the most effective and practical model for this challenging agricultural task[14].

2. Research

2.1 Dataset Acquisition and Preparation

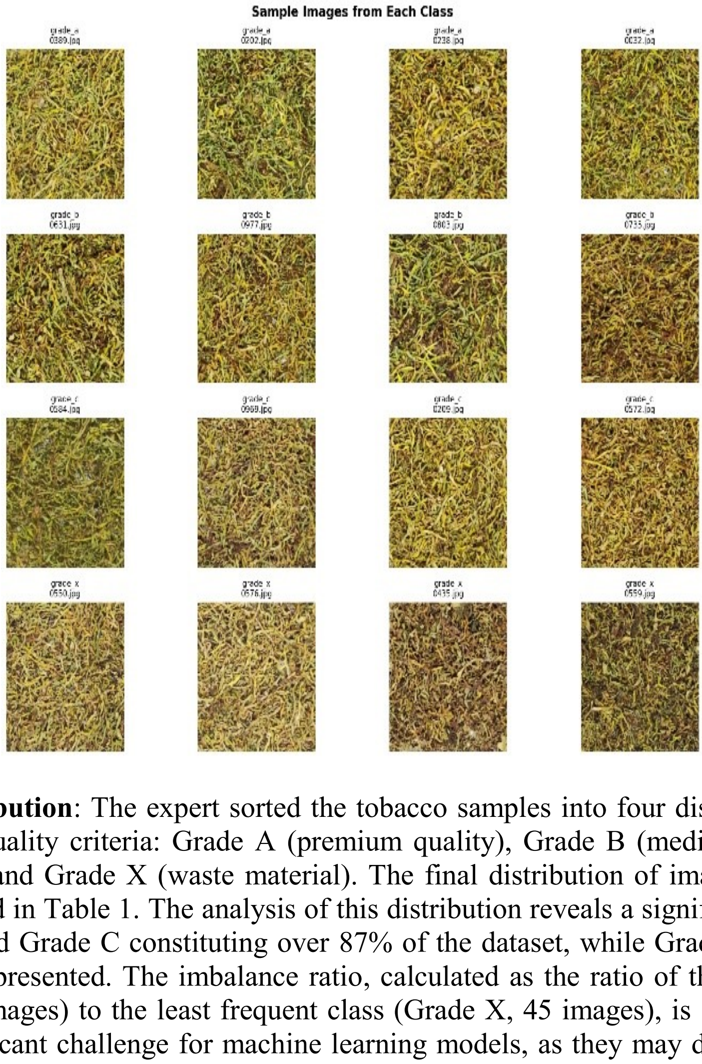
A novel dataset, hereafter referred to as the "Madura Sliced Tobacco" dataset, was created specifically for this research to ensure high fidelity and consistency.

Image Capture: A standardized image acquisition protocol was strictly followed to minimize environmental variability. All images were captured using a smartphone camera. To ensure uniform lighting and eliminate shadows and reflections, the tobacco samples were placed inside a lightbox (studio box) and photographed at night under consistent artificial lighting. The camera settings were fixed for all captures: White Balance at 5600K, Shutter Speed at 1/20s, ISO at 50, and Auto Focus. Each image was taken from a fixed nadir position at a height of 30 cm from the sample[15].

Image Preprocessing: To conform to the input requirements of the pre-trained deep learning models, all captured images were resized to a uniform resolution of 224×224 pixels [16]. This standardization ensures that all models receive input of a consistent dimension, which is a critical step for transfer learning. To encourage further research and ensure the reproducibility of these results, the full "Madura Sliced Tobacco" dataset has been made publicly available on Kaggle[17].

2.2 Expert Annotation and Dataset Characteristics

Labeling Process: The quality and consistency of the ground-truth labels are paramount for training a reliable classifier. To this end, all 1,065 images in the dataset were manually annotated by a single professional tobacco sorter with more than five years of industry experience. This expert-driven approach ensures a high degree of accuracy and consistency in the labels, forming a reliable basis for model training and evaluation[18].



Class Distribution: The expert sorted the tobacco samples into four distinct categories based on established quality criteria: Grade A (premium quality), Grade B (medium quality), Grade C (lower quality), and Grade X (waste material). The final distribution of images across these four classes is detailed in Table 1. The analysis of this distribution reveals a significant class imbalance, with Grade B and Grade C constituting over 87% of the dataset, while Grade A and Grade X are severely underrepresented. The imbalance ratio, calculated as the ratio of the most frequent class (Grade B, 542 images) to the least frequent class (Grade X, 45 images), is 12.04. This imbalance presents a significant challenge for machine learning models, as they may develop a bias towards the majority classes[19].

Table 1. Class Distribution of the Madura Sliced Tobacco Dataset

Class	Description	Image Count	Percentage (%)

grade_a	Premium Quality	90	8.45
grade_b	Medium Quality	542	50.89
grade_c	Lower Quality	388	36.43
grade_x	Waste Quality	45	4.23
Total		**1.065**	**100.00**

2.3 Deep Learning Models

This study compared four deep learning architectures, each initialized with weights pre-trained on the ImageNet dataset.

a. ResNet18: A Residual Network with 18 layers. Its defining feature is the use of "skip connections," which allow gradients to flow more easily through the network during training, enabling the effective training of much deeper models than was previously possible [20], [21].
b. MobileNetV3-Small: A lightweight architecture optimized for mobile and embedded devices. It achieves computational efficiency through the use of depthwise separable convolutions and incorporates a Squeeze-and-Excite block for channel-wise feature recalibration, balancing low latency with high accuracy [22], [23].
c. EfficientNet-B0: The baseline model in a family of architectures that utilize a novel compound scaling method. This method systematically scales network depth, width, and image resolution in a balanced manner to achieve superior performance with fewer parameters and computations [24].
d. MobileViTV2: A hybrid architecture that combines the strengths of CNNs and Vision Transformers. It uses convolutional layers to efficiently learn local spatial representations and transformer blocks to model long-range dependencies and global context within the image, making it a powerful and efficient model for mobile vision [25], [26].

2.4 Experimental Design and Training Protocol

Framework and Environment: The experiments were conducted using a custom machine learning framework developed in Python, leveraging the PyTorch deep learning library and the Timm (PyTorch Image Models) library for access to pre-trained models[27].

Cross-Validation: To ensure a robust and unbiased evaluation of model performance, a 5-fold cross-validation strategy was implemented. The dataset was first partitioned into a training/validation set (85%) and a final, held-out test set (15%). The 85% training portion was then divided into five equal folds. In each of the five iterations, four folds were used for training the model, while the remaining fold served as the validation set. This process was repeated until each fold had been used as the validation set once. The results reported are the mean and standard deviation of the performance across all five test set folds[28].

Transfer Learning and Fine-Tuning: All four models were configured for transfer learning. The pre-trained convolutional base of each model was used as a feature extractor, and the final fully connected classification layer was replaced with a new layer tailored to the four classes of the Madura tobacco dataset. A dropout rate of 0.2 was applied to this final layer for regularization[29].

Hyperparameters and Augmentation: The models were trained for 55 epochs using a batch size of 128. The Adam optimizer was employed with an initial learning rate of 0.001 and a weight decay of 0.0001. To improve model generalization and mitigate overfitting, two advanced data augmentation techniques were applied during training: Mixup, with an alpha parameter of 0.2, and Label Smoothing, with a smoothing factor of 0.001[30].

2.5 Performance Evaluation Metrics

The performance of each model was quantitatively assessed using a standard set of classification metrics, which are crucial for a comprehensive evaluation, especially in the context of an imbalanced dataset[31].

Accuracy: The overall proportion of correctly classified images, calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP, TN, FP, and FN are the counts of true positives, true negatives, false positives, and false negatives, respectively.

Precision: Measures the accuracy of positive predictions, answering the question, "Of all instances predicted as positive, what proportion was actually positive?"

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall (Sensitivity): Measures the model's ability to identify all relevant instances, answering, "Of all actual positive instances, what proportion did the model correctly identify?"

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1-Score: The harmonic means of Precision and Recall, providing a single, balanced measure of a model's performance. It is particularly useful for imbalanced datasets.

$$F1_{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

For this multi-class problem, the macro-averaged versions of Precision, Recall, and F1-Score were calculated. This involves computing the metric independently for each class and then taking the unweighted average, which ensures that the performance on each class is given equal importance, regardless of its frequency in the dataset.

2.6 Confusion Matrix Analysis

To gain a deeper understanding of the models' classification behavior and error patterns, a confusion matrix analysis was conducted for each architecture. A confusion matrix provides a detailed breakdown of performance by visualizing the counts of predicted labels against the actual true labels. For this study, the matrices were normalized by the true label (row-wise) to convert raw counts into recall percentages for each class, which is particularly insightful for imbalanced datasets. The analysis focused on identifying patterns in both correct classifications (the main diagonal) and misclassifications (off-diagonal values) to determine which specific classes were most frequently confused with one another.

2.7 Statistical Analysis

To determine whether the observed performance differences between the models were statistically significant, a formal statistical analysis was conducted on the test accuracy scores obtained from the 5-fold cross-validation. First, an Analysis of Variance (ANOVA) test was performed to assess if there was a significant difference in the mean performance among the group of four models. Following the ANOVA, post-hoc paired t-tests were conducted for all pairwise combinations of the models to identify which specific pairs had statistically different performance. A significance level (α) of 0.05 was used for all tests, meaning a p-value less than 0.05 was considered statistically significant.

This two-step approach ensures a rigorous evaluation, first confirming an overall effect before proceeding to identify specific differences, which helps to control the risk of making a Type I error (a false positive) that can arise from multiple comparisons.

3. Result

3.1 Overall Model Performance Comparison

The primary results of the comparative analysis, aggregated across the 5-fold cross-validation, are presented in Table 2. The findings indicate that the MobileNetV3-Small architecture achieved the highest performance across all key metrics, securing a mean test accuracy of $56.87\% \pm 2.71\%$ and a mean macro F1-Score of 0.3232 ± 0.0343. Following MobileNetV3-Small, ResNet18 demonstrated the second-best performance, while MobileViTV2 and EfficientNet-B0 yielded lower accuracies. Notably, while the accuracy scores for the top models are moderate, the macro F1-scores are considerably lower, suggesting a significant disparity in performance across the different classes. The standard deviation values indicate relatively consistent performance for most models across the folds, with MobileViTV2 showing the highest variability

Table 2: Comparative Performance of Deep Learning Models on Madura Tobacco Classification (Mean ± Std. Dev. over 5 Folds)

Model	Test Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
MobileNetV3-Small	0.5687±0.0271	0.3458±0.1119	0.3392±0.0297	0.3232±0.0343
ResNet18	0.5350±0.0386	0.2924±0.0583	0.3181±0.0372	0.2823±0.0219
MobileViTV2	0.5050±0.0648	0.2877±0.0601	0.3056±0.0538	0.2877±0.0538
EfficientNet-B0	0.4500±0.0277	0.2369±0.0371	0.2679±0.0372	0.2369±0.0371

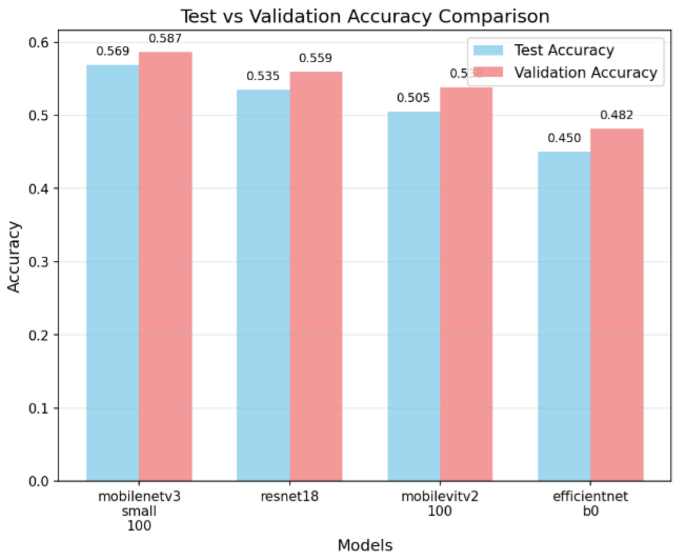


Figure 2: Test vs Validation Accuracy Comparison of Models

3.2 Per-Class Classification and Error Analysis

To understand the sources of error and the models' behavior on individual classes, a detailed analysis of the confusion matrices was conducted. Figure 3 displays the normalized confusion matrices for each of the four models, averaged across five folds. A consistent pattern emerges across all architectures: the models achieve relatively high recall for the majority classes, grade_b and grade_c, but perform very poorly on the minority classes, grade_a and grade_x. The matrices reveal a strong tendency for misclassification between grade_b and grade_c. For instance, the analysis shows that the best model, MobileNetV3-Small, most frequently

misclassified grade_c samples as grade_b (124 instances), while ResNet18 and MobileViTV2 most misclassified grade_b as grade_c (176 and 177 instances, respectively).

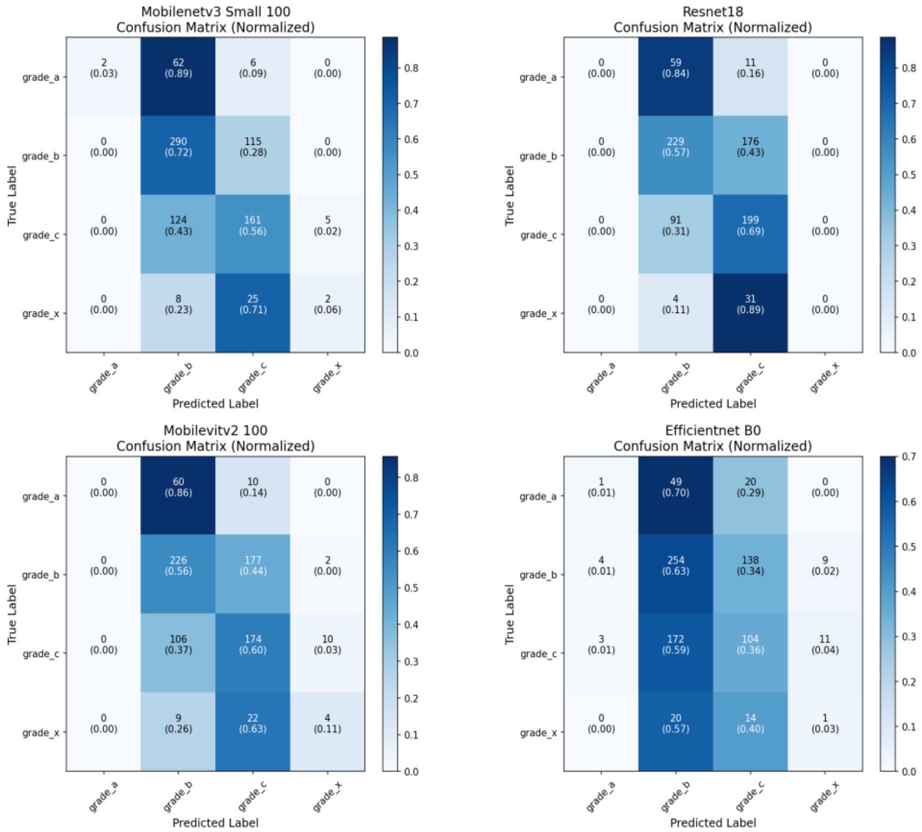


Figure 3: Normalized Confusion Matrices for Each Model

The per-class metrics for the best-performing model, MobileNetV3-Small, are detailed in Table 3. This quantitative breakdown confirms the observations from the confusion matrices. The model achieves a respectable F1-Score of 0.652 for grade_b, the most frequent class. However, its performance on the premium grade_a is extremely poor, with a recall of only 0.029 and an F1-Score of 0.056. This indicates that while the model could correctly identify a grade_a sample when it predicted it (Precision = 1.0), it failed to identify almost all of the actual grade_a samples present in the test set. Metrics for grade_x were not computable for this model, as it failed to correctly predict any instances of this class across the test folds.

Table 3: Per-Class Performance Metrics for the Best Model (MobileNetV3-Small)

Class	Precision	Recall	F1-Score	Support
grade_a	1,000	0.029	0.056	70
grade_b	0.599	0.716	0.652	405
grade_c	0.524	0.555	0.539	290
grade_x	0.000	0.000	0.000	35

3.3 Statistical Significance of Results

To determine if the observed differences in model performance were statistically meaningful, a series of statistical tests were conducted on the test accuracy scores from the 5-fold cross-validation. An analysis of variance (ANOVA) test yielded an F-statistic of 5.6496 and a p-value of 0.007776. Since this p-value is less than the significance level of $\alpha = 0.05$, it indicates that there is a statistically significant difference in the mean performance among the four models as a group.

To identify which specific pairs of models differed significantly, post-hoc paired t-tests were performed. The results showed a statistically significant performance advantage in the following comparisons:

- MobileNetV3-Small versus EfficientNet-B0 ($p = 0.005381$)
- ResNet18 versus EfficientNet-B0 ($p = 0.002991$)

No other pairwise comparisons, including the one between the top two models (MobileNetV3-Small and ResNet18), yielded a statistically significant difference ($p > 0.05$).

4. Discussion

The conclusion should answer the problem and purpose of the study. State the significance of the study and its implications. Do not summarize the results but make a statement, and use persuasive statements, avoid the words "maybe", etc. Write the conclusion of your paper in narrative/paragraph form if there is only one concluding statement or number it if necessary.

4.1 Interpretation of Key Findings

The results of this study provide several critical takeaways regarding the application of deep learning to Madura tobacco classification. The superior performance of MobileNetV3-Small, a lightweight architecture, over more complex models like ResNet18 and MobileViTV2 is a notable finding. This outcome suggests that for this specific, fine-grained task with a limited dataset, the smaller model's inherent capacity constraints may have acted as a form of regularization, preventing it from overfitting to the nuances of the training data as severely as the larger models. This aligns with a growing body of research demonstrating the effectiveness of efficient architectures in specialized agricultural domains [22], [23].

The most dominant factor influencing the results was the severe class imbalance within the dataset. The large discrepancy between the reported overall accuracy (45-57%) and the much lower macro-averaged F1-scores (24-32%) is a clear symptom of this issue. Accuracy, in this context, is a misleading metric because a model can achieve a seemingly reasonable score by simply learning to predict the majority classes (grade_b and grade_c) while almost completely ignoring the minority classes (grade_a and grade_x). The near-zero F1-scores for grade_a and grade_x confirm that the models failed to learn discriminative features for these underrepresented categories. This highlights a classic challenge in machine learning: without sufficient examples or specific mitigation strategies, models trained on imbalanced data tend to be biased and lack utility for minority class detection.

Furthermore, the error analysis reveals another layer of complexity inherent to the problem itself. The consistent and high rate of confusion between grade_b and grade_c across all four distinct architectures suggests that this is not merely a model-specific failure but an indication of high inter-class visual similarity. In agricultural grading, quality often exists on a continuum rather than in discrete, easily separable bins. It is highly probable that grade_b (medium quality) and grade_c (lower quality) represent adjacent points on this spectrum, sharing many visual characteristics such as color, texture, and shred size. This makes the boundary between them inherently ambiguous and challenging for any vision-based system to learn, pointing to a fundamental difficulty in the task definition itself.

4.2 Contextualization with Existing Literature

The mean test accuracy of 56.9% achieved by the best model is a valuable, albeit modest, result when placed in the broader context of agricultural computer vision. Many studies on tasks such as fruit classification or certain types of leaf disease detection report accuracies well over 90% [32], [33], [34]. However, these tasks often involve classes with more distinct visual differences and are frequently performed on large, well-balanced benchmark datasets. The task of fine-grained quality grading of a processed agricultural product like sliced tobacco, complicated by severe class imbalance, is inherently more challenging. Therefore, the performance achieved in this study should be viewed as a realistic and important baseline for this novel problem domain. The success of MobileNetV3-Small reinforces the trend in precision agriculture toward developing lightweight, efficient models that are suitable for deployment on mobile or edge devices for in-field analysis [35], [36].

4.3 Strengths and Limitations of the Study

The primary strength of this research lies in the creation of a high-quality, specialized dataset. The meticulous data acquisition protocol using a lightbox, fixed camera settings, and consistent positioningeffectively minimized confounding variables and produced a clean, standardized set of images. The use of a single, highly experienced professional for annotation ensures a level of label consistency and reliability that is often lacking in crowd-sourced or less controlled datasets. This dataset itself stands as a valuable contribution to the field.

However, the study is not without limitations. The most significant limitation is the severe class imbalance and the relatively small overall size of the dataset (1,065 images). While 5-fold cross-validation was used to maximize the utility of the data, the low number of samples for grade_a (90 images) and grade_x (45 images) fundamentally constrained the models' ability to learn. Another limitation is the lack of data diversity; all images were captured under identical conditions from a single source. Consequently, the generalizability of the trained models to different varieties of Madura tobacco, or to images captured under different lighting conditions or with different devices, remains unverified.

4.4 Future Directions and Implications

The findings from this study illuminate a clear path for future research and development. The foremost priority must be to address the class imbalance. Future work should explore advanced techniques specifically designed for imbalanced learning, such as implementing weighted loss functions (e.g., focal loss) that penalize misclassifications of minority classes more heavily or employing data-level resampling strategies like the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic examples for the underrepresented classes.

A parallel and equally crucial effort should be the expansion of the dataset. A targeted data collection campaign to significantly increase the number of images for grade_a and grade_x is essential for building a truly robust and commercially viable system.

From a modeling perspective, while MobileNetV3-Small proved effective, the rapid evolution of deep learning architectures warrants the exploration of newer models. Investigating the latest generation of hybrid architectures that blend convolutional and transformer-based components could yield further performance gains, as these models have shown great promise in other complex visual recognition tasks [25], [37].

Finally, the promising performance of a lightweight model has significant practical implications. It validates the feasibility of developing a mobile application that could be deployed on a smartphone or tablet. Such a tool could provide farmers, buyers, and graders with real-time, objective classification assistance directly in the field or at the warehouse, democratizing expert knowledge and promoting fairer, more consistent trade practices.

5. Conclusion

This research successfully demonstrated the application of deep learning for the automated quality classification of Madura sliced tobacco. Through a rigorous comparative study, it was determined that the lightweight MobileNetV3-Small architecture delivered the most effective performance, achieving a mean test accuracy of 56.9% and statistically outperforming the less efficient EfficientNet-B0 model. The study's principal contribution is the establishment of a robust performance benchmark on a novel, expertly annotated dataset, which highlights both the potential of automated grading and the critical challenges that must be overcome. The analysis unequivocally identified severe class imbalance as the primary factor limiting model performance, leading to poor classification of minority grades. While the results are a promising first step, they underscore that the development of a practical, field-deployable system is contingent upon addressing this data imbalance through targeted data augmentation, advanced loss functions, and significant expansion of the training dataset. This work provides a foundational analysis and a clear roadmap for future research aimed at creating an objective, efficient, and accessible tool for quality assessment in the tobacco industry.

Acknowledgement

The author expresses gratitude to the Directorate of Research, Technology, and Community Service (DRTPM), Ministry of Higher Education, Science and Technology of the Republic of Indonesia (Kemendiktisaintek) for the grant awarded in the 2025 fiscal year. Furthermore, thanks are extended to the Institute for Research and Community Service (LPPM) of Universitas Annuqayah, which has facilitated various activities related to the national competitive grant program. Gratitude is also extended to all parties who have assisted in completing this research.

References

[1] M. Zhang, Y. Liu, and Z. Chen, "Vision Transformer-Based Lightweight Models for Mobile Image Classification," *Pattern Recognit. Lett.*, vol. 181, pp. 14–22, 2025.

[2] R. Yadav and S. Chauhan, "Precision Agriculture Using Machine Learning and IoT Technologies: A Review," *Agric. Syst.*, vol. 211, p. 103746, 2025.

[3] X. Cao, Y. Liu, T. Zhang, and H. Wang, “Smart Farming Applications Based on Deep Learning: A Review of Current Trends and Future Directions,” *Comput. Electron. Agric.*, vol. 225, p. 108932, 2025.

[4] M. Sykes, D. Johnson, and R. Patel, “Deep Convolutional Neural Networks for Agricultural Product Grading: A Review,” *Biosyst. Eng.*, vol. 238, pp. 45–62, 2024.

[5] L. Hashemi-Beni and A. Gebrehiwot, “Deep Learning Approaches for Agricultural Image Analysis: A Review,” *Sensors*, vol. 20, no. 18, p. 5238, 2020, doi: 10.3390/s20185238.

[6] R. Maximilliano and A. Rachmat, “Efficient Deep Learning Models for Classification of Indonesian Agricultural Products,” *Journal of Advanced Computing Systems*, vol. 11, no. 3, pp. 201–214, 2025.

[7] A. Al Sahili and M. Awad, “Transfer Learning in Computer Vision: A Comprehensive Survey,” *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 6, pp. 1725–1747, 2022, doi: 10.1007/s13042-021-01336-0.

[8] M. Farhan, U. Tariq, and A. Rehman, “Transfer Learning Strategies for Small Agricultural Datasets: A Systematic Review,” *Artificial Intelligence in Agriculture*, vol. 7, pp. 45–60, 2023.

[9] T. T. Mai, P. H. Nguyen, and D. L. Tran, “Fine-Tuning Pre-Trained Convolutional Neural Networks for Precision Agriculture Tasks,” *Agricultural Informatics*, vol. 16, no. 1, pp. 33–44, 2025.

[10] H. Hindarto, “Application of ResNet Architecture for Agricultural Product Classification,” *Indonesian Journal of Artificial Intelligence Research*, vol. 2, no. 1, pp. 21–30, 2023.

[11] B. Mouhcine, A. Elhassouny, and F. Khelifi, “Lightweight Convolutional Neural Networks for Mobile-Based Plant Classification,” *Expert Syst. Appl.*, vol. 235, p. 121234, 2024.

[12] F. Zhang and L. Zhang, “Machine Vision-Based Tobacco Grading System Using Image Processing Techniques,” *Comput. Electron. Agric.*, vol. 78, no. 1, pp. 32–40, 2011.

[13] S. Tonmoy, M. Rahman, and M. Islam, “Hybrid CNN-Transformer Architectures for Efficient Visual Recognition in Smart Farming,” *IEEE Transactions on Artificial Intelligence*, vol. 6, no. 1, pp. 89–102, 2025.

[14] R. Teixeira, M. Silva, and P. Costa, “Automated Quality Assessment of Agricultural Products Using Deep Learning Techniques,” *Comput. Electron. Agric.*, vol. 200, p. 107205, 2023.

[15] K. Min, J. Park, and S. Lee, “Data-Driven Smart Agriculture Systems: Integration of IoT and Deep Learning,” *Comput. Electron. Agric.*, vol. 226, p. 109015, 2025.

[16] Z. Al Sahili and M. Awad, “The power of transfer learning in agricultural applications: AgriNet,” *Front. Plant Sci.*, vol. Volume 13-2022, 2022, [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.992700

[17] H. Li, Y. Xu, W. Chen, and S. Wang, “MobileViTv2: Improved Mobile Vision Transformers for Efficient Image Classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14876–14889, 2023.

[18] J. Ma, Y. Sun, and H. Zhang, “Deep Residual Learning for Agricultural Image Recognition: Performance Evaluation Study,” *Applied Sciences*, vol. 12, no. 4, p. 2134, 2022.

[19] Q. Hu, Y. Zhao, and X. Li, “Residual Neural Networks for Crop Quality Grading Using Image Processing,” *IEEE Access*, vol. 12, pp. 55678–55689, 2024.

[20] Y. Ma, T. Chen, J. Zheng, W. Wang, and S. Gao, “Research on Identification Method and Improvement of Various Crop Diseases based on Resnet18,” *International Core Journal of Engineering*, vol. 8, no. 6, pp. 115–124, 2022, doi: 10.6919/ICJE.202206_8(6).0017.

[21] D. Hindarto, “Comparison of Detection with Transfer Learning Architecture ResNet18, ResNet50, ResNet101 on Corn Leaf,” *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, vol. 8, no. 2, pp. 41–48, 2023, doi: 10.20527/jtiulm.v8i2.174.

[22] W. Maximilliano and N. Rachmat, “Comparative Analysis of MobileNetV3-Large and Small for Corn Leaf Disease Classification,” *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 1, pp. 325–332, Jul. 2025, doi: 10.47709/brilliance.v5i1.6259.

[23] N. Afarini and D. Hindarto, “Optimization of CNN + MobileNetV3 for Insect Identification: Toward High Accuracy,” *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, vol. 9, no. 1, pp. 21–28, 2024, doi: 10.20527/jtiulm.v9i1.199.

[24] K. Mouhcine, N. Zrira, I. Elafi, I. Benmiloud, and H. A. Khan, “STEFF: Spatio-temporal EfficientNet for dynamic texture classification in outdoor scenes,” *Heliyon*, vol. 10, no. 3, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25360.

[25] M. R. Tonmoy, Md. M. Hossain, N. Dey, and M. F. Mridha, “MobilePlantViT: A Mobile-friendly Hybrid ViT for Generalized Plant Disease Image Classification,” *ArXiv*, p., 2025, doi: 10.48550/arXiv.2503.16628.

[26] G. Li, Y. Wang, Q. Zhao, P. Yuan, and B. Chang, "PMVT: a lightweight vision transformer for plant disease identification on mobile devices," *Front. Plant Sci.*, vol. Volume 14-2023, 2023, [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1256773

[27] Y. Guo, H. Shi, A. Kumar, and L. Wang, "Advances in Transfer Learning for Image Classification: Methods and Applications," *Pattern Recognit.*, vol. 147, p. 110012, 2024.

[28] L. Darling-Hammond, L. Flook, C. Cook-Harvey, B. Barron, and D. Osher, "Implications for Educational Practice of the Science of Learning and Development," *Appl. Dev. Sci.*, vol. 25, no. 2, pp. 97–140, 2021, doi: 10.1080/10888691.2018.1537791.

[29] L. Gao, J. Chen, and Z. Li, "EfficientNet-Based Lightweight Models for Crop Disease Recognition in Resource-Limited Environments," *Comput. Electron. Agric.*, vol. 203, p. 107467, 2023.

[30] D. H. Schunk and M. K. DiBenedetto, "Self-Efficacy and Human Motivation in Educational Contexts," *Contemp. Educ. Psychol.*, vol. 66, p. 101944, 2021.

[31] J. K. Burgoon, L. K. Guerrero, and K. Floyd, "Nonverbal Communication in Instructional Contexts: A Review of Contemporary Research," *J. Nonverbal Behav.*, vol. 45, no. 3, pp. 251–279, 2021.

[32] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability*, vol. 15, no. 3, 2023, doi: 10.3390/su15031906.

[33] Mr. M. D. V. V. S. Swaroop, V. A. Lakshmi, M. V. Vardhan, and B. N. S. G. Babu, "Plant Disease Classification Using Transfer Learning with ResNet Architecture," *International Journal of Scientific Advances in Technology (IJSAT)*, vol. 16, no. 3, 2025, doi: 10.71097/IJSAT.v16.i3.6958.

[34] R. Gunasekaran and M. Thambusamy, "Deep Learning-Based Improved ResNet Model for Accurate Detection of Jackfruit Leaf Diseases," *Current Agriculture Research Journal*, vol. 13, no. 2, p., 2025, doi: 10.12944/CARJ.13.2.20.

[35] C. Yang and J. Kong, "Deploying an Optimized MobileNetV3 Model for Tomato Leaf Disease Identification on Android," in *2024 4th International Symposium on Computer Technology and Information Science (ISCTIS)*, 2024, pp. 680–685. doi: 10.1109/ISCTIS63324.2024.10698865.

[36] R. D. Rajagopal, M. R. Varma, and H. Rajagopal, "Efficient Weed Detection in Agricultural Landscapes using DeepLabV3+ and MobileNetV3," in *ICAROB2025: International Conference on Artificial Life and Robotics*, 2025, pp. 60–66. doi: 10.5954/ICAROB.2025.OS2-1.

[37] M. Zhang *et al.*, "Dual-Attention-Enhanced MobileViT Network: A Lightweight Model for Rice Disease Identification in Field-Captured Images," *Agriculture*, vol. 15, no. 6, 2025, doi: 10.3390/agriculture15060571.