DEVELOPMENT OF AN AIoT-BASED COFFEE BEAN CLASSIFICATION AND SORTING SYSTEM USING A VISION TRANSFORMER

Dody Pintarko[1]*, Basuki Rahmat[2], Faisal Muttaqin[3].
Program Studi Magister Teknologi Informasi, Fakultas Ilmu Komputer,
Universitas Pembangunan Nasional "Veteran" Jawa Timur
E-mail: 24066020008@student.upnjatim.ac.id

Received : 03 July 2026	Accepted: 08 July 2026
Revised : 05 July 2026	Published: 19 July 2026

Abstract

Manual coffee bean sorting is highly prone to subjectivity, inconsistency, and low operational efficiency. This study aims to develop an automated classification and sorting system based on the Artificial Intelligence of Things (AIoT). The method integrates a Vision Transformer (ViT) model, TensorFlow Lite, Firebase, and an ESP32 microcontroller within a Mobile–Cloud–Edge Computing architecture. The ViT model was trained on four coffee roast levels to perform real-time inference on Android devices linked to physical sorting actuators. Experimental results showed that the ViT model achieved a 96.87% classification accuracy, while the automated physical sorting mechanism achieved 95.83% accuracy with an average response time of 462 ms. In conclusion, the integration of Vision Transformer and AIoT provides a fast and reliable post-harvest automation solution tailored for smart agricultural applications.

Keywords: AIoT, Vision Transformer, Coffee Bean Classification, TensorFlow Lite, Smart Agriculture.

INTRODUCTION

Coffee bean quality is a key factor in determining the economic value and competitiveness of coffee products. In the post-harvest stage, sorting based on roasting level plays a crucial role in maintaining product quality uniformity. However, this process is still largely performed manually through visual observation, making the results highly dependent on operator experience. This situation has the potential to cause inconsistencies, increase sorting errors, and reduce efficiency as production volumes increase. Therefore, an automated classification system is needed that can produce a faster, more objective, and consistent sorting process through the use of computer vision and deep learning technologies. Various studies have applied Convolutional Neural Networks (CNN), ResNet, and YOLO to classify coffee bean quality and roasting levels with a fairly good level of accuracy. However, convolution-based models still focus on local feature extraction, making them less than optimal in distinguishing classes with very similar visual characteristics, such as Light Roast and Medium Roast. In contrast, the Vision Transformer (ViT) utilizes a self-attention mechanism to build a global image representation, thereby capturing relationships between image parts more comprehensively. These advantages make the Vision Transformer potentially improve classification accuracy for objects with relatively small visual differences. On the other hand, the implementation of Artificial Intelligence of Things (AIoT) in the agricultural sector is generally still oriented towards monitoring and data collection functions, while the integration between artificial intelligence-based classification, cloud communication, and physical device control in one automated system is still relatively limited.

Based on these conditions, several research gaps remain. First, the application of Vision Transformer to coffee bean roasting classification has not been widely explored compared to CNN-based approaches. Second, most AIoT research has not integrated mobile AI-based inference, cloud data synchronization, and edge device actuation in a unified Mobile–Cloud–Edge Computing architecture. Third, previous research has generally focused on improving model accuracy, while comprehensive evaluation of system implementation, including computational efficiency, response time, and automatic sorting performance, is rarely reported. This condition indicates the need to develop a system that not only has high classification performance but is also capable of real-time implementation in an operational environment. Based on these research gaps, this study develops an Artificial Intelligence of Things (AIoT)-based coffee bean classification and sorting system using Vision Transformer in a Mobile–Cloud–Edge Computing architecture. The system integrates TensorFlow Lite on Android devices for real-time inference, Firebase Realtime Database as a data synchronization medium, and ESP32 as a controller for the automatic sorting mechanism. The main contributions of this research include the implementation of an end-to-end AIoT system that connects image acquisition, classification, data communication, and device actuation in a single integrated workflow, as well as a comprehensive evaluation of classification accuracy, computational efficiency, response time, and automatic sorting performance. Thus, this research not only offers improved classification model performance but also contributes to the implementation of a more integrated AIoT system to support the development of smart agriculture technology.

METHOD

This study uses a Research and Development (R&D) approach to develop a coffee bean classification and sorting system based on Artificial Intelligence of Things (AIoT). The research stages include dataset preparation, Vision Transformer (ViT) model development, optimization using TensorFlow Lite, integration in Mobile–Cloud–Edge Computing architecture, and system performance evaluation. The dataset consists of 6,000 images representing four roasting levels (Green Bean, Light Roast, Medium Roast, and Dark Roast). All images were processed through resizing to 224 × 224 pixels, normalization, and data augmentation, then divided into training data (70%), validation (15%), and testing (15%). The model training configuration is presented in Table 1.

Table 1. Research Configuration

Parameter	Mark
Dataset	6,000 images
Class	Green, Light, Medium, Dark Roast
Image Size	224 × 224 pixels
Data Sharing	70 % : 15% : 15%
Model	Vision Transformer (ViT-Base)
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Epoch	100
Platform	Android + TensorFlow Lite
Cloud	Firebase Realtime Database
Edge	ESP32 + Servo + Load Cell

All training configurations were selected based on preliminary experimental results to achieve a balance between classification accuracy and computational efficiency in the TensorFlow Lite implementation. The Vision Transformer model was chosen because it is capable of constructing a global image representation through self-attention, making it more effective at distinguishing roast levels with similar visual characteristics compared to CNN-based approaches.

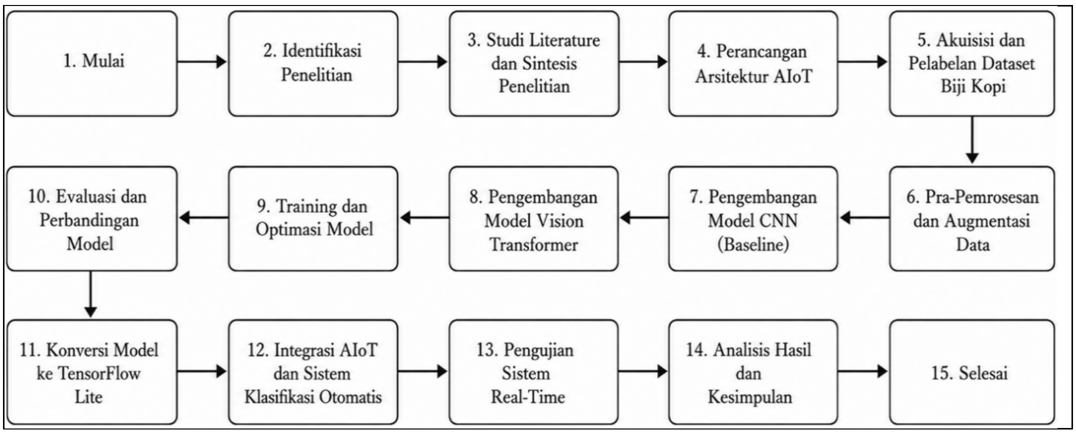

Figure 1: Research Framework

Figure 1 shows the research stages, starting with problem identification, dataset compilation, model development, optimization, AIoT system implementation, and performance evaluation. This flow ensures that each stage is carried out systematically, ensuring that the resulting model not only has good classification performance but is also ready for implementation in an operational environment.

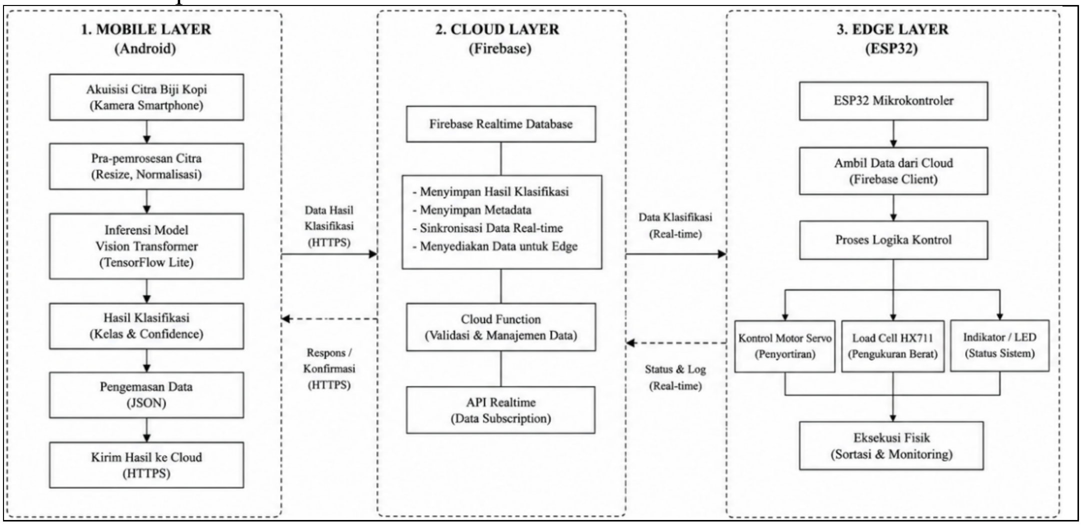

Figure 2: Mobile – Cloud – Edge Computing System Architecture

Figure 2 shows the integration of the mobile layer, cloud layer, and edge layer in the Mobile–Cloud–Edge Computing architecture. The Android device serves as the mobile layer for image acquisition and inference using TensorFlow Lite. The classification results are then sent to the Firebase Realtime Database as the cloud layer for real-time data synchronization, while the ESP32 as the edge layer receives the inference results and controls the load cell and servo motor in the automatic sorting process. This division of functions allows the computation process, data communication, and device actuation to take place efficiently with low latency.

The evaluation was conducted on two aspects: classification model performance and AIoT system implementation. Model performance was measured using Accuracy, Precision, Recall, and F1-Score to evaluate classification capabilities for each roasting category. Meanwhile, system implementation was evaluated based on model size, inference time, data synchronization, ESP32 response, end-to-end processing time, sorting accuracy, and robustness under various lighting conditions and object positions. This evaluation approach provides a comprehensive overview of the system's capabilities, both in terms of artificial intelligence and AIoT implementation in the automated sorting process.

RESULTS AND DISCUSSION

This research successfully implemented a coffee bean classification and sorting system based on Artificial Intelligence of Things (AIoT) through the integration of Vision Transformer, TensorFlow Lite, Firebase Realtime Database, and ESP32 in a Mobile–Cloud–Edge Computing architecture. The system performs inference directly on an Android device, sends the classification results to Firebase, and then controls the sorting mechanism via ESP32 in real-time. This integration forms an end-to-end automation process that connects image acquisition, classification, data communication, and device actuation so that all stages can take place automatically without operator intervention.

Table 2: Comparison of Model Performance

Model	Accuracy	Precision	Recall	F1-Score
CNN	93.24	92.76	92.14	92.45
Vision Transformer	96.87	96.31	96.05	96.18

Based on Table 2, the Vision Transformer improves accuracy by 3.63% compared to CNN and achieves the best scores across all evaluation metrics. This improvement indicates that the self-attention mechanism is capable of constructing a global image representation, thus more effectively distinguishing roast levels with similar color and texture characteristics. This capability reduces classification errors in classes with subtle visual differences, particularly Light Roast and Medium Roast. These results align with research by Hassan (2024) and Takahashi et al. (2024) which demonstrate that the Vision Transformer has superior feature representation capabilities compared to convolution-based models in complex image classification tasks.

Optimization using TensorFlow Lite resulted in a more efficient model for mobile devices. The model size was reduced from 86 MB to 21 MB, while the inference time decreased from 420 ms to 118 ms, with a decrease in accuracy of only 0.55%. These results demonstrate that the optimization can maintain classification performance while increasing computational efficiency. This allows the inference process to run directly on Android devices, reducing reliance on cloud computing, reducing communication latency, and improving system responsiveness in mobile edge AI implementations.

Table 3. Summary of AIoT Implementation Performance

Parameter	Mark
TensorFlow Lite Model Size	21 MB
Inference Time	118 ms
Firebase Sync	82 ms
ESP32 Response	41 ms
Servo Actuation	126 ms
End-to-End Time	462 ms
Sorting Accuracy	95.83%
Throughput	320 seeds/hour

Table 3 shows that the integration of TensorFlow Lite, Firebase Realtime Database, and ESP32 produces an AIoT system with low response time and stable sorting performance. The AIoT implementation shows that the system successfully sorted 1,150 out of 1,200 coffee beans with 95.83% accuracy, a 98.67% servo actuation success rate, a throughput of 320 beans per hour, and an average end-to-end response time of 462 ms. This success is influenced not only by the accuracy of the classification model, but also by data synchronization via Firebase and the ESP32's actuation response, which is able to maintain real-time coordination between devices. The division of functions between the mobile layer, cloud layer, and edge layer also makes the computation and communication processes more efficient, so the system remains responsive when implemented in an operational environment. Despite providing good results, this study still has several limitations. The dataset used only covers four roasting level categories, and testing was conducted on a single Android device configuration and ESP32. Furthermore, the testing was conducted in a laboratory environment, making it less representative of industrial conditions. Therefore, further research needs to evaluate the system on different coffee varieties, more diverse lighting conditions, and devices with different specifications to test the system's generalizability and scalability.

Overall, this research successfully addresses the research gap through the application of Vision Transformer to roasting level classification and the implementation of an end-to-end AIoT system that integrates mobile device inference, cloud synchronization, and edge device control in a single Mobile–Cloud–Edge Computing architecture. The research contribution is demonstrated not only by the improvement in classification performance, but also by the successful integration of artificial intelligence with data communication and real-time automatic sorting mechanisms. These results indicate that the proposed approach has the potential to be applied to coffee post-harvest automation and can serve as a reference for the development of intelligent classification systems in various smart agriculture applications.

CONCLUSION

This research successfully developed a coffee bean classification and sorting system based on Artificial Intelligence of Things (AIoT) using Vision Transformer in Mobile–Cloud–Edge Computing architecture. The integration of TensorFlow Lite, Firebase Realtime Database, and ESP32 enables the classification process, data exchange, and sorting to take place automatically and in real-time. The results show that the proposed approach not only improves classification performance compared to CNN-based models but also results in efficient implementation on mobile devices. These findings prove that the combination of Vision Transformer and mobile edge AI is capable of supporting an accurate and responsive classification system for post-harvest automation needs.

The main contribution of this research lies in the development of an end-to-end AIoT system that integrates inference on mobile devices, cloud communication, and edge device control in a single unified platform. This approach demonstrates that the Mobile–Cloud–Edge Computing architecture not only improves processing efficiency and system response speed, but also provides a reliable integration mechanism between artificial intelligence and physical devices. Thus, this research provides a scientific contribution in the form of the implementation of an intelligent classification system that can be used as a reference for the development of AIoT applications in the agricultural product sorting process and other automation systems that require real-time decision making.

Despite yielding satisfactory results, this study still has limitations in the use of four roasting level categories, one hardware configuration, and testing conducted in a laboratory environment. Therefore, further research can be directed at developing a lighter Vision Transformer model, testing on various coffee varieties and industrial operating conditions, and integrating multimodal sensors to improve the system's robustness, scalability, and readiness for production-scale implementation.

Thank-you note

The author would like to express his gratitude to the Master of Information Technology Study Program, Faculty of Computer Science, National Development University "Veteran" East Java for the academic support and research facilities provided during the implementation of this research. The author also expresses his gratitude to the supervisors for their valuable guidance, input, and guidance, which enabled this research to be successfully completed.

Author Contribution

Dody Pintarko was responsible for research design, system development, software and hardware implementation, data collection, analysis, and manuscript preparation. Basuki Rahmat was responsible for developing the research methodology, model validation, and research supervision. Faisal Muttaqin contributed to system evaluation, experimental analysis, and manuscript revision.

Conflict of Interest

The author declares that there is no conflict of interest related to the research, writing, or publication of this article.

REFERENCES

Albanese, A., Nardello, M., & Brunelli, D. (2021). *Automated Pest Detection with DNN on the Edge for Precision Agriculture*. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 11(4). https://doi.org/10.1109/ JETCAS.2021.3101740.

Alzubi, AA, & Galyna, K. (2023). *Artificial Intelligence and Internet of Things for Sustainable Farming and Smart Agriculture*. IEEE Access, 11, 78686–78700. https://doi.org/10.1109/ACCESS.2023.3298215

Cao, Z., Sun, S., & Bao, X. (2025). *A Review of Computer Vision and Deep Learning Applications in Crop Growth Management*. Applied Sciences, 15(8438). https://doi.org/10.3390/app15158438

Chang, S.-J., & Huang, C.-Y. (2021). *Deep Learning Model for the Inspection of Coffee Bean Defects*. Applied Sciences, 11(8226). https://doi.org/10.3390/ app11178226

dos Santos, FFL, Rosas, JTF, Martins, RN, Araújo, GM, Viana, LA, & Gonçalves, JP (2020). *Quality Assessment of Coffee Beans through Computer Vision and Machine Learning Algorithms*. Coffee Science, 15, e151752. https://doi.org/10.25186/cs.v15i.1752

García, M., Candelo-Becerra, J.E., & Hoyos, F. (2019). *Quality and Defect Inspection of Green Coffee Beans Using a Computer Vision System*. Computers and Electronics in Agriculture, 162, 735–746.

Gope, L., Fukai, H., Ruhad, F. M., & Barman, S. (2024). *Comparative Analysis of YOLO Models for Green Coffee Bean Detection and Defect Classification*. Scientific Reports, 14, 28946. https://doi.org/10.1038/s41598-024-78598-7

Hassan. (2024). *Enhancing Coffee Bean Classification: A Comparative Analysis of Pre-Trained Deep Learning Models*. Neural Computing and Applications, 36(9), 9023–9052. https://doi.org/10.1007/s00521-024-09623-z

Korkmaz, A., Talan, T., Koşunalp, S., & Iliev, T. (2025). *Comparison of Deep Learning Models in Automatic Classification of Coffee Bean Species*. PeerJ Computer Science, 11, e2759. https://doi.org/10.7717/peerj-cs.2759

Mulyono, Apnitami, P., Wangi, IS, Wicaksono, KNP, & Apriono, C. (2022). *The Potential of Smart Farming IoT Implementation for Coffee Farming in Indonesia: A Systematic Review*. Green Intelligent Systems and Applications, 2(2), 53–70. https://doi.org/10.53623/gisa.v2i2.95

Nawaz, M., & Babar, MIK (2025). *IoT and AI for Smart Agriculture in Resource-Constrained Environments: Challenges, Opportunities and Solutions*. Discover Internet of Things, 5(24). https://doi.org/10.1007/s43926-025-00119-3

Oncu, E. (2025). *Deep Learning Framework for Coffee Quality Assessment via YOLOv8n*. IEEE Access, 13.

Rajak, P., Ganguly, A., Adhikary, S., & Bhattacharya, S. (2023). *Internet of Things and Smart Sensors in Agriculture: Scopes and Challenges*. Journal of Agriculture and Food Research, 14, 100776. https://doi.org/10.1016/ j.jafr .2023.100776

Runck, C., et al. (2024). *Real-Time Geoinformation Systems to Improve the Quality, Scalability, and Cost of IoT for Agri-Environment Research*. Computers and Electronics in Agriculture, 223, 108195.

Takahashi, S., Sakaguchi, Y., Kouno, N., et al. (2024). *Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review*. Journal of Medical Systems, 48(84). https://doi.org/10.1007/s10916-024-02105-8

Tariq, M.U., Saqib, S.M., Mazhar, T., Khan, M.A., Shahzad, T., & Hamam, H. (2025). *Edge-Enabled Smart Agriculture Framework: Integrating IoT, Lightweight Deep Learning, and Agentic AI for Context-Aware Farming*. Results in Engineering, 28, 107342. https://doi.org/10.1016/ j.rineng .2025.107342

Ye, B., Xue, R., & Xu, H. (2025). *ASD-YOLO: A Lightweight Network for Coffee Fruit Ripening Detection in Complex Scenarios*. Frontiers in Plant Science, 16, 1484784. https://doi.org/10.3389/fpls.2025.1484784

Zou, X., Liu, Z., Zhu, X., Zhang, W., Qian, Y., & Li, Y. (2023). *Application of Vision Technology and Artificial Intelligence in Smart Farming*. Agriculture, 13(2106).https://doi.org/10.3390/agriculture13112106ps://doi.org/ 10.3390/agriculture13112106