Automated Defect Classification of Coffee Beans Using Deep-Stacking Ensemble Learning

Porntida Kaewkamol[a], Sujitra Arwatchananukul[b,d*],
Rattapon Saengrayap[c,d], and Phasit Charoenkwan[a]
Received: January 17, 2025; Revised: February 3, 2025;
Accepted: March 4, 2025; Published Online: April 3, 2025

Abstract

Quality control in coffee production is essential for maintaining product standards and preserving market value. A key practice in this process is to identify and remove defective beans which ensures high-quality standards and enhances consumer experience. However, traditional methods of classifying coffee bean defect often rely on manual inspection which is labour-intensive, time-consuming and subject to human errors. As such, adopting image classification for coffee bean defects could improve accuracy and boosts operational efficiency. This study explores the effectiveness of stacking-based deep learning ensemble method for coffee bean defect classification. The methodology involves a performance study as a baseline approach from fourteen traditional machine learning algorithms, including Support Vector Machines (SVM) and Random Forest (RF), along with ten different feature extraction techniques, such as FOS and GLDS. Besides, twenty well-known deep learning architectures including ResNet50, ConvNeXt and EfficientNet were compared to fourteen lightweight models such as TinyNet and MobileNet. Additionally, the performance of stacking-based deep learning models is also analysed to optimise coffee bean defect classification. The results indicate that ConvNeXt achieved the highest testing accuracy at 72.94% across all DL architectures. Additionally, the stacking approach significantly improves classification performance as it achieved an accuracy improvement from 72.94% to 87.64%. This study contributes to a comprehensive benchmarking to evaluate a diverse range of machine learning and deep learning algorithms. It also highlights the effectiveness of the stacking ensemble model to enhance accuracy in coffee bean defect classification.

Keywords: Stacking-based Ensemble Learning, Deep Learning Algorithm, Machine Learning Algorithm, Image Classification, Model Benchmarking

[a] College of Arts, Media and Technology, Chiang Mai University
[b] School of Applied Digital Technology, Mae Fah Luang University
[c] School of Agro-Industry, Mae Fah Luang University
[d] Integrated AgriTech Ecosystem Research Group (IATE), Mae Fah Luang University
* Corresponding Author Email: sujitra.arw@mfu.ac.th

Introduction

Coffee production is a key contributor to economies and communities globally. Quality control in coffee is crucial for preserving its market value and ensuring consumer satisfaction since coffee is graded based on its quality. Defective beans such as discoloration, physical damage, or fungal contamination can negatively affect the flavour, aroma, and overall quality of coffee. Traditionally, the process to classify and detect coffee bean defects has relied on manual inspection which is labour-intensive, time-consuming and subject to human errors. The challenges of coffee bean defect classification highlight the need for more efficient methods to enhance its accuracy and scalability.

The advent of machine learning (ML) and deep learning (DL) has introduced automated solutions for image classification tasks in quality control. ML involves algorithms such as Support Vector Machines (SVM), Random Forest (RF), and Naïve Bayes (NB) that rely on feature extraction and statistical learning. DL, on the other hand, employs neural networks capable of automatically extracting features from raw data which leads to superior performance in complex datasets. Additionally, stacking is an ensemble learning method that combines the predictions of multiple models to improve overall performance. As such, stacking outputs into meta-models allows for significant performance improvements in multi-class image classification tasks.

This study aims to benchmark a variety of M: and DL architectures on a dataset of coffee bean defects. First, this study evaluates performance from fourteen traditional machine learning algorithms, including Support Vector Machines (SVM) and Random Forest (RF), along with ten different feature extraction techniques, such as FOS and GLRLM to establish a baseline for model performance. Second, this study compares between twenty well-known deep learning architectures including TinyNet and MobileNet to fourteen lightweight models such as ConvNeXt and ResNet50 to explore their potential for enhancing classification accuracy and achieve better trade-offs between accuracy and efficiency, especially in resource-constrained scenarios such as small farms or mobile devices. Third, this study evaluates the performance of a stacked deep learning model where features from the highest accuracy DL models are used to train traditional ML models to improve classification accuracy.

Despite advancements, the application of stacking deep learning for coffee bean defect classification remains underexplored. Previous studies often focus on standalone models which leaves a gap in utilizing ensemble strategies to combine their strengths. This study addresses these gaps by implementing a stacking-based framework to leverage probability outputs from well-known deep learning models to train traditional ML algorithms. The proposed method aims to enhance classification accuracy and robustness for quality control in the coffee industry.

Literature review

Machine Learning (ML)

Machine Learning (ML) is a subset of artificial intelligence that enables systems to learn patterns from data and predict outcomes, classify items, or make decisions (Zhang, Chan, & Mahadevan, 2022).. Common applications of ML span various domains, e.g. image classification, natural language processing, and predictive analytics (Motta et al., 2025).

The application of ML in coffee classification has been extensively studied across various aspects in past research including quality control and disease detection. Regarding quality control, previous study suggests that ML techniques could enhance the efficiency and accuracy of coffee bean classification (Pragathi & Jacob, 2022).

Another study highlights the use of ML to classify specialty coffee based on maturity, roasting, defects, sensory attributes, and diseases. The featured ML models include Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). This study underlines that ML enhances efficiency in coffee production. (Motta et al., 2025).

Besides, the past study combines single coffee bean mass spectrometry (MS) with a one-hidden-layer neural network. and SHAP feature extraction to improve machine learning results which achieved an accuracy of 99.58% for coffee bean classification (Tsai et al., 2023).

Another related research focused on using ML combined with projection pursuit to improve quality control in specialty coffee sensory data. The study demonstrated that hierarchical clustering with the Ward method and Minkowski distance achieved the best results with a validation error rate of 1.44%. (Ossani, et al., 2020).

The past research also emphasizes the role of ML in manual coffee sorting. The study combines ML techniques and dielectric spectroscopy to classify coffee cherry maturity stages The results showed that Support Vector Machine (SVM) achieved the best classification results with a training accuracy of 90.7% Velásquez et al., 2021).

Similarly, past research studied on applying ML for classifying roasted Arabic coffee into three roasting levels: light, medium, and dark. This classification was achieved using color information, chemical composition, and antioxidant properties with ML classifiers such as Random Forests (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN). The results revealed that Random Forest (RF) achieved the highest accuracy with color or chemical composition data (Alamri et al., 2023).

Additionally, ML has played an essential role in detecting coffee plant diseases. XGBoost and SHAP techniques have been used to classify infected coffee cherries with high accuracy at 98.56%. This emphasizes the potential of image-based ML for early disease detection in coffee plant (Selvanarayanan, Rajendran, & Alotaibi, 2023).

Deep Learning (DL)

Deep learning (DL) focuses on training artificial neural networks (ANNs) with many layers to recognize patterns, classify data, and make predictions (Hassan, 2024). In contrast to traditional machine learning approaches that require manual feature extraction, deep learning automates this process and enables the model to learn hierarchical representations of data directly from raw inputs such as images or text (Ramamurthy et al., 2023).

DL has emerged as a powerful tool to address challenges in agricultural industry (Jantawong, 2024). Especially in coffee production industry such as quality classification and disease detection. In past research, coffee bean classification has advanced through the application of YOLOv8 for real-time automated grading which reduces processing times to seconds while maintaining high accuracy (Thai et al. 2024). Similarly, the use of EfficientNet-B0 and ResNet-34 for classifying coffee beans into nine grades achieved 100% accuracy under challenging lighting and background conditions. (Balakrishnan et al., 2024).

Other pre-trained DL models were also explored_in previous research to evaluate accuracy in classifying coffee bean types. AlexNet, LeNet, HRNet, GoogleNet, MobileNetV2, ResNet-50, VGG, EfficientNet, Darknet, and DenseNet were applied to a coffee dataset comprising 4,800 images. The results indicated that VGG achieved the highest accuracy at 100% for coffee classification into four roasting levels (Hassan, 2024).

Besides, DL models have also been used for early detection of coffee leaf diseases. Past research combined InceptionV3 and DenseNet121 as a hybrid DL model to detect coffee leaf diseases. The results highlighted that the model achieved an accuracy of 99% (Singh & Kumar, 2024).

Similarly, Thakur et al. (2024) developed the SUNet model which integrated U-Net and SegNet architectures with VGG16-based encoder for precise segmentation and classification in early detection of coffee leaf diseases. The results demonstrated that it achieved an accuracy of 98.45%.

Additionally, another previous work proposed a CoffeeNet framework which combined ResNet-50 with a spatial-channel attention mechanism to detect coffee leaf disease. This model Achieved an accuracy of 98.54% (Nawaz et al., 2024).

EfficientNetB0 has also been widely utilised for early detection of coffee leaf diseases. Ramamurthy et al. (2023) demonstrated the integration of EfficientNetB0 with the Ghost module in detecting coffee diseases. The model achieved an overall 84% classification accuracy.

Bhoomika and Verma (2024) further studied the performance of EfficientNetB and its adaptability in resource-constrained settings such as farming application. The results showed that the model achieved the overall accuracy of 96% to classify diseases in coffee leaves.

ML, DL and Stacking ensemble learning

ML and DL approaches have been widely applied for image classification tasks. Common ML techniques such as Support Vector Machines (SVM), Decision Trees, and K-nearest neighbors (KNN) typically require feature extraction to achieve optimal performance.. In contrast, DL methods, particularly Convolutional Neural Networks (CNNs), automatically extract and process features in the datasets (Karypidis et al., 2022) Generally, traditional ML models achieve superior performance on small-scale datasets while DL models are highly effective in managing large-scale datasets to achieve accuracy. (Wang et al., 2021).

Regarding coffee industry, previous research mentioned that ML techniques have been used for disease detection, crop seeds classification and quality assessment (Kumar et al., 2024). Similarly, DL models such as AlexNet and VGG16 have also been applied for crop disease detection and achieved superior accuracy compared to classical methods (Rangarajan et al., 2021).

However, ML and DL techniques can be integrated into an ensemble model to enhance predictive accuracy and robustness. Stacking is the ensemble learning technique that combines predictions from multiple base models using a meta-learner. In other words, stacking ensemble learning integrates predictions from multiple base models such as ML and DL to improve predictive performance (Khan et al., 2022). In previous research, hybrid models using CNNs and traditional ML classifiers as SVMs have been shown to enhance performance of crop seeds classification (Kumar et al., 2024).

Methodology

Dataset Preparation

The image dataset in this study comprises seventeen categories of coffee bean defects, for example, broken, dry cherry, full black, fungus damage and withered. The 979 original images were augmented by using rotation techniques at multiple angles as 45, 90, 135, 180, 225, and 270 to expand to a total of 6,853 images. All images were also cropped to 500x500 pixel. The dataset was then split into training sets (70%), validation sets (20%) and testing sets (10%). (Arwatchananukul et al., 2024).

Baseline Performance Study

Fourteen traditional ML algorithms and ten feature extraction techniques were selected to establish a baseline for a benchmark. Each model was evaluated on both training and testing sets using metrics such as accuracy, precision, recall, and F1-score. Table 1 and Table 2 show ML algorithms and feature extraction techniques in this study.

Table 1: ML algorithms in this study

Title	ML Algorithm
SVM	Support Vector Machine
NB	Naive Bayes
1NN:	1-Nearest Neighbor
DT:	Decision Tree
Logistic	Logistic Regression
PLS	Partial Least Squares
LinearSVC	Linear Support Vector Classification
LDA	Linear Discriminant Analysis
RF:	Random Forest
E-Tree:	Extra Trees
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
ADA	AdaBoost
MLF	Multi-Layer Perceptron (MLP)

Table 2: Feature extraction techniques in this study

Title	Feature Extraction
FOS	Feature of Similarity
GLDS	Gray-Level Dependence Matrix
GLRLM	Gray-Level Run Length Matrix
GLSZM:	Gray-Level Size Zone Matrix
Haralick	Haralick Features
LTE	Local Ternary Patterns
NGTDM	Neighboring Gray Tone Difference Matrix
DWT:	Discrete Wavelet Transform

Table 2: (Continued)

Title	Feature Extraction
GT:	Gray Tone (Gray-Level)
Zernike Moments	Zernike Moments:

Evaluating DL Architectures

In this study, DL architectures were categorized into two groups as well-known and lightweight models. Well-known models demonstrate optimal performance for unconstrained environments while lightweight models emphasize deployment feasibility on resource-constrained devices.

Twenty well-known DL models in five architectures were evaluated for their standalone performances. They were also compared to fourteen lightweight DL models in seven architectures. Additionally, various pre-trained model configurations from ImageNet were also utilized for transfer learning to accelerate convergence and improve generalization, for example, pre-training dataset in1k: Indicates the model is pretrained on ImageNet-1K (1,000 classes). Table 3 and Table 4 show well-known and lightweight DL architectures in this study.

Table 3: Well-known DL architectures in this study

Architecture	Example Model	Pre-training
VGG	VGG16_bn.tv_in1k VGG19_bn.tv_in1k	ImageNet-1K
ResNet	Resnet50.a1_in1k Resnet101 a1h_in1k	ImageNet-1K/12K
EfficientNet	Efficientnetv2_rw_t.ra2_in1k	ImageNet-1K
ConvNeXt	ConvNext_tiny.fb_in1k	ImageNet-12K/22K
RegNet	RegNetx_032.tv2_in1k RegNetx_080.tv2_in1k	ImageNet-1K

Table 4: Lightweight DL architectures in this study

Architecture	Example Model	Pre-training
MobileNet	MobileNetv3_small_050.lamb_in1k	ImageNet-1K
TinyNet	TinyNet_e.in1k TinyNet_d.in1k	ImageNet-1K
LCNet	LCNet_050.ra2_in1k LCNet_075.ra2_in1k	ImageNet-1K
MNASNet	MNASnet_small.lamb_in1k	ImageNet-1K
RegNetX RegNetY	RegNetx_002.pycls_in1k,	ImageNet-1K
ResNet Variants	ResNet10t.c3_in1k	ImageNet-1K
RepGhostNet	RepGhostNet_050.in1k	ImageNet-1K

Each model underwent hyperparameter tuning including learning rate adjustments and optimization strategies. An optimal learning rate was also determined to ensure efficient training using the Learning Rate Finder method. This process ensures optimal learning dynamics for all benchmarked models especially ConvNeXt and ResNet.

Establishing Stacking-Based Models

In this step, a stacking-based ensemble approach was analysed to leverage the best-performing model. This process uses the probabilistic outputs from the best DL models as input features for training traditional ML algorithms such as SVM, RF.

In other words, each DL model outputs a probability distribution over the 17 defect classes for every input images. These 17 probabilities are treated as features representing the image. The 17-class probability vectors are combined into a feature matrix and are used as input to traditional ML algorithms. The traditional ML models serve as base learners in the stacking ensemble. A meta-learner e.g., Logistic Regression or GBM is trained to combine the predictions from the base learners to generate the final class label.

Results and Discussion

Performance of Traditional ML Baseline

The results from traditional ML models were analyzed to assess the performance of coffee bean defect classification. Traditional ML models, including Support Vector Machine (SVM), Random Forest, and others, were benchmarked against the dataset using metrics as Accuracy, Precision, Recall, F1-score.

The results indicated that SVM model with Haralick feature extraction yielded the best accuracy performance as depicted in Fig 1. Accuracy for traditional ML models demonstrated that some models struggled with the complexity of the task. The haralick feature extraction method provided the best results across multiple models in both training and testing phases.

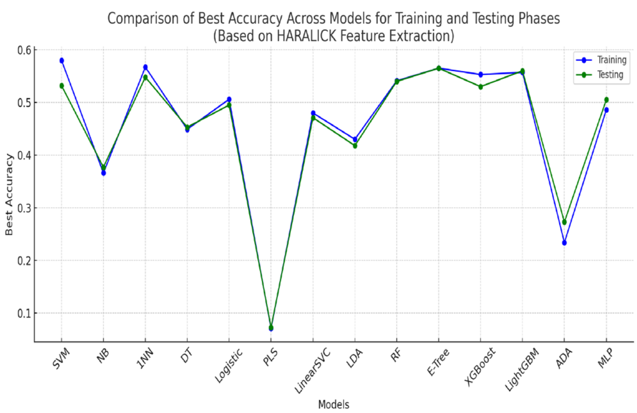

Fig 1. Accuracy across ML models with Haralick feature extraction

Performance of DL Architectures

The performance of DL models for coffee bean defect classification was conducted across two categories: well-known and lightweight models. The results highlight a clear distinction between the performance capabilities of these groups.

Well-known models achieved superior accuracy and generalization. ConvNeXt which was the best performance model achieved a test accuracy of 72.94%. This result aligns with findings in prior works and further validates the generalization capability of well-known DL models (Balakrishnan et al., 2024; Singh & Kumar, 2024; Nawaz et al., 2024).

Besides, the relationship between learning rate and training loss of the ConvNeXt final model can be shown in Fig 2. The learning rate finder curve was analyzed to identify two critical points: Valley Point indicates the learning rate where loss begins to decrease. Slide Point suggests the learning rate with the steepest loss decline.

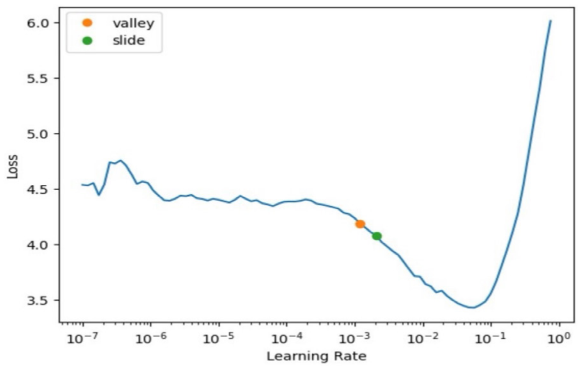

Fig 2. Learning rate

Additionally, the strong results of ConvNeXt also reinforce the value of pre-trained architectures for classification tasks. (Hassan et al.,2024). The performances of well-known DL architectures are summarised in Fig 3.

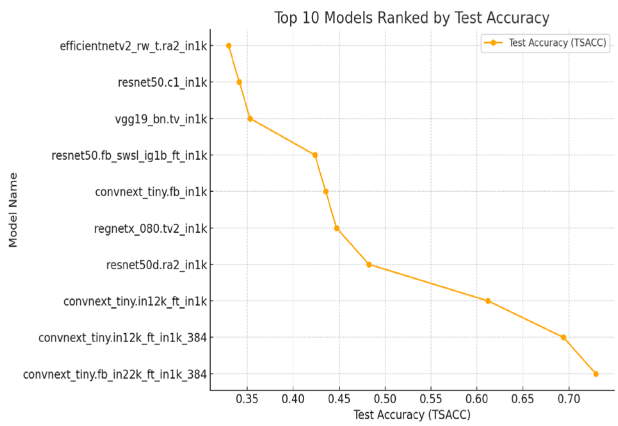

Fig 3. Performances of well-known DL architectures

In contrast, lightweight models achieved moderate performance with LCNet achieved the highest test accuracy at 29.41%. These models showed limitations in achieving high accuracies on the dataset which might be due to their constrained capacity. In summary, well-known DL models outperformed lightweight models across accuracy metric. These results highlight that while lightweight models are optimized for resource constraints, they require further tuning strategies to handle complex datasets effectively Fig 4 demonstrated test accuracy distribution by model category.

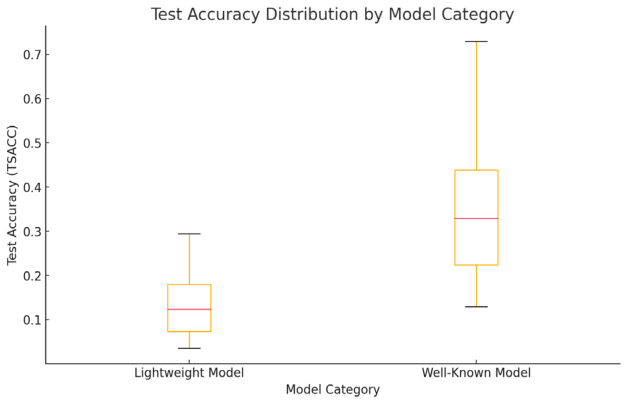

Fig 4. Test accuracy distribution by model category.

Performance of Stacking-based Models

The stacking ensemble leveraged the probabilistic outputs from the best performance DL models (ConvNeXt) and trained the traditional ML models such as SVM and LR. This approach effectively combined the feature extraction capabilities of DL models with the predictive robustness of traditional ML algorithms. The results indicated a marked improvement with accuracies improved from 72.94% to 87.64% as illustrated in Fig 5.

The results of stacking-based DL models in this study aligns and strengthen the results from previous research that stacking ensemble learning can improve predictive performance (Khan et al., 2022; Kumar et al., 2024).

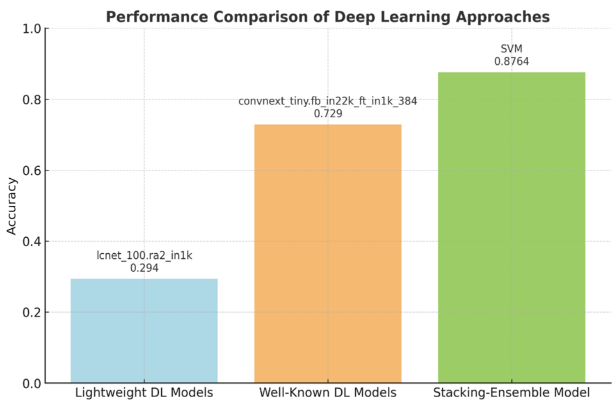

Fig 5: Performance of DL appraoches

Besides, the comparison with standalone models further underscores the advantages of stacking-ensemble models. Traditional ML models, when trained independently, achieved limited accuracy due to their reliance on handcrafted features and inability to capture complex patterns. In contrast, stacking ensembles benefited from the deep learning model's ability to extract high-level and discriminative features. As such, it can enhance overall classification performance.

Conclusion

This study investigates the effectiveness of stacking-based ensemble learning for coffee bean defect classification. The study evaluates the performance of fourteen ML algorithms using ten feature extraction techniques. This study also compares twenty well-known DL architectures, e.g., ConvNeXt, ResNet50 against fourteen lightweight models, e.g., TinyNet, MobileNet and integrates probabilistic outputs from the best-performing DL models as ConvNeXt into the meta-models as Support Vector Machines (SVM).

According to the results, this study demonstrates the feasibility of integrating DL and ML models in a stacking ensemble framework to address the challenges of coffee bean defect classification. The results indicate that the stacking ensemble method outperformed both standalone ML and DL models by combining their strengths and can significantly improve accuracy and robustness. The integration of probabilistic outputs from the well-known DL model (ConvNeXt,) with the traditional ML meta-models (SVM) resulted in an accuracy improvement from 72.94% to 87.64%.

However, this study acknowledges certain areas where further improvement is needed The dependence on resource-intensive DL models such as ConvNeXt highlights the need for optimizing resource efficiency for real-time applications. Moreover, further exploration of additional feature extraction techniques, meta-model architectures, and hybrid frameworks could uncover new opportunities for performance enhancement.

.

Reference

Alamri, E. S., Altarawneh, G. A., Bayomy, H. M., & Hassanat, A. B. (2023). Machine learning classification of roasted Arabic coffee: Integrating color, chemical compositions, and antioxidants. Sustainability, 15(11561). https://doi.org/10.3390/su151511561

Arwatchananukul, S., Xu, D., Charoenkwan, P., Moon, S. A., & Saengraya, R. (2024). Implementing a deep learning model for defect classification in Thai Arabica green coffee beans. Smart Agricultural Technology, 9, 100680. https://doi.org/10.1016/j.atech.2024.100680

Balakrishnan, B. N., Jayakumari, A. N. K. M., Stephen, S. A., Venkitesan, P., & Raghavendra, V. (2024). Coffee bean graded based on deep net models. International Journal of Electrical and Computer Engineering (IJECE), 14(3), 3084–3093. https://doi.org/10.11591/ijece.v14i3.pp3084-3093

Bhoomika, & Verma, G. (2024). Enhancing coffee plant disease identification with EfficientNetB0 and deep learning. Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC 2024). IEEE. https://doi.org/10.1109/I-SMAC61858.2024.10714676

Hassan, E. (2024). Enhancing coffee bean classification: A comparative analysis of pre-trained deep learning models. Neural Computing and Applications, 36(9023–9052). https://doi.org/10.1007/s00521-024-09623-z

Jantawong, S. (2024). Image detection of GT pesticide test kit results in organic agricultural products using the CiRA-CORE deep learning model. *International Journal of Science and Innovative Technology, 7*(2)

Khan, A. H., Iskandar, D. N. A., Mewada, H., & Sherazi, M. A. (2022). Ensemble learning of deep learning and traditional machine learning approaches for skin lesion segmentation and classification. Concurrency and Computation: Practice and Experience. https://doi.org/10.1002/cpe.6907

Karypidis, E., Mouslech, S. G., Skoulariki, K., & Gazis, A. (2022). Comparison analysis of traditional machine learning and deep learning techniques for data and image classification. WSEAS Transactions on Mathematics, 21(19). https://doi.org/10.37394/23206.2022.21.19

Kumar, V., Aydav, P. S. S., & Minz, S. (2024). Crop seeds classification using traditional machine learning and deep learning techniques: A comprehensive survey. SN Computer Science, 5(1031). https://doi.org/10.1007/s42979-024-03379-y

Motta, I. V. C., Vuillerme, N., Pham, H.-H., & de Figueiredo, F. A. P. (2025). Machine learning techniques for coffee classification: A comprehensive review of scientific research. Artificial Intelligence Review, 58(1), 15. https://doi.org/10.1007/s10462-024-11004-w

Nawaz, M., Nazir, T., Javed, A., Amin, S. T., Jeribi, F., & Tahir, A. (2024). CoffeeNet: A deep learning approach for coffee plant leaves diseases recognition. Expert Systems With Applications, 237, 121481. https://doi.org/10.1016/j.eswa.2023.121481

Ossani, P. C., Rossoni, D. F., Cirillo, M. Â., & Borém, F. M. (2020). Unsupervised classification of specialty coffees in homogeneous sensory attributes through machine learning. *Coffee Science, 15,* e151780. https://doi.org/10.25186/cs.v15i.1780

Pragathi, S. P., & Jacob, L. (2022). Review on image-based coffee bean quality classification: Machine learning approach. 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 706–711. https://doi.org/10.1109/ICAC3N56670.2022.10074124

Rangarajan, A. K., Purushothaman, R., Prabhakar, M., & Szczepa□ski, C. (2021). Crop identification and disease classification using traditional machine learning and deep learning approaches. Journal of Engineering Research, 11(1B), 228–252. https://doi.org/10.36909/jer.11941

Ramamurthy, K., Chattopadhyay, S., & Thekkath, R. D. (2023). A novel deep learning architecture for disease classification in Arabica coffee plants. Concurrency and Computation: Practice and Experience. https://doi.org/10.1002/cpe.7625

Selvanarayanan, R., Rajendran, S., & Alotaibi, Y. (2023). Early detection of Colletotrichum Kahawae disease in coffee cherry based on computer vision techniques. Computer Modeling in Engineering & Sciences. https://doi.org/10.32604/cmes.2023.044084

Singh, M. K., & Kumar, A. (2024). Coffee leaf disease classification by using a hybrid deep convolution neural network. SN Computer Science, 5(618). https://doi.org/10.1007/s42979-024-02960-9

Thai, H.-D., Ko, H.-J., & Huh, J.-H. (2024). Coffee bean defects automatic classification: Realtime application adopting deep learning. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3452552

Thakur, D., Gera, T., Aggarwal, A., Verma, M., Kaur, M., Singh, D., & Amoon, M. (2024). SUNet: Coffee leaf disease detection using hybrid deep learning model. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3476211

Tsai, J.-J., Chang, C.-C., Huang, D.-Y., Lin, T.-S., & Chen, Y.-C. (2023). Analysis and classification of coffee beans using single coffee bean mass spectrometry with machine learning strategy. Food Chemistry, 426, 136610. https://doi.org/10.1016/j.foodchem.2023.136610

Velásquez, S., Franco, A. P., Peña, N., Bohórquez, J. C., & Gutiérrez, N. (2021). Classification of the maturity stage of coffee cherries using comparative feature and machine learning. Coffee Science, 16, e161710. https://doi.org/10.25186/.v16i.1710

Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recognition Letters, 141, 61–67. https://doi.org/10.1016/j.patrec.2020.11.024

Zhang, X., Chan, F. T. S., & Mahadevan, S. (2022). Explainable machine learning in image classification models: An uncertainty quantification perspective. Knowledge-Based Systems, 243, 108418. https://doi.org/10.1016/j.knosys.2022.108418