Explainable Deep Learning Models for Nutmeg Seed Image Quality Assessment

Manuel Soares Dos Reis Pacheco [1,*], Hadiyanto Hadiyanto [1], and Ridwan Sanjaya [2]

[1] Doctoral Program of Information Systems, School of Postgraduate Studies, Universitas Diponegoro, Semarang 50241, Indonesia
[2] Department of Information Systems, Soegijapranata Catholic University, Semarang 50234, Indonesia
Email: soares.students@undip.ac.id (M.S.D.R.P); hadiyanto@che.undip.ac.id (H.H.); ridwan@unika.ac.id (R.S.)
*Corresponding author

***Abstract*—Nutmeg seed quality plays a key role in determining the economic value and export potential of Indonesian spice products. Manual assessment of quality is often subjective, time-consuming, and depends on the evaluator's experience. This study introduces an explainable deep learning method designed to automatically and transparently evaluate nutmeg image quality. The model employs the EfficientNet-B0 architecture trained on 300 nutmeg images categorized into three quality levels: good, moderate, and rotten. To enhance interpretability, Explainable Artificial Intelligence (XAI) techniques—specifically Gradient-Weighted Class Activation Mapping (Grad-CAM) and local Interpretable Model-Agnostic Explanations (LIME)—were applied to highlight the image regions most influential in the model's decisions. Model performance was evaluated using accuracy, precision, recall, F1-score, and mean Average Precision (mAP). The EfficientNet-B0 model, coupled with XAI methods, achieved an overall accuracy of 78% and a mAP of 70.29%, with Grad-CAM and LIME visualizations consistently highlighting the key visual features that determine nutmeg quality. These findings demonstrate that integrating deep learning with XAI can produce an objective, efficient, and dependable quality assessment system, which offers potential applications to other agricultural products.**

***Keywords*—explainable deep learning, EfficientNet-B0, Gradient-Weighted Class Activation Mapping (Grad-CAM), Interpretable Model-Agnostic Explanations (LIME), nutmeg seed quality, artificial intelligence, smart agriculture**

I. INTRODUCTION

Nutmeg is a leading spice commodity in Indonesia with significant economic importance, particularly in the Maluku Islands, which are recognized as the global hub of nutmeg cultivation [1, 2]. The quality of nutmeg seeds greatly influences their market value, both for export and for derivative sectors such as essential oils, pharmaceuticals, and food processing [3, 4]. The evaluation of nutmeg seed quality is predominantly performed manually by experts, who visually inspect color, surface texture, cracks, and mold [5]. These traditional methods are often subjective, slow, and less reliable when used at large-scale operations.

The development of deep learning technology, especially Convolutional Neural Networks (CNNs), has demonstrated remarkable performance across a range of image classification tasks, including in agriculture and crop processing. Early models such as LeNet-5, AlexNet, and VGGNet laid the foundation in building image-based classification systems due to their ability to learn visual patterns from image data automatically [6, 7]. More recent architectures—including ResNet, DenseNet, and EfficientNet—have successfully enhanced accuracy and efficiency through residual learning and compound scaling methods [8−10]. Despite their outstanding performance, many of these models are criticized for their "black box" nature, as they often produce accurate predictions without explaining the reasoning behind those decisions. This raises concerns regarding interpretability and user trust, especially in determining the quality of agricultural products, which requires transparency in decision-making.

To address these issues, research in Explainable Artificial Intelligence (XAI) has advanced, focusing on clarifying and visualizing the decision-making processes of AI models [11, 12]. Approaches such as Gradient-Weighted Class Activation Mapping (Grad-CAM), Layer-Wise Relevance Propagation (LRP), and Shapley Additive Explanations (SHAP) help identify image regions that most influence classification results [13, 14]. When applied to nutmeg seed images, these techniques can identify visual cues—such as color uniformity, surface cracks, and mold stains—which serve as key indicators of seed quality. This interpretability not only enhances the model's transparency but also offers valuable insights for farmers, industry professionals, and quality certification bodies to refine inspection standards and minimize human bias.

This study seeks to design and evaluate an explainable deep learning model for the purpose of assessing the quality of nutmeg seed. The integration of a high-performance CNN architecture with XAI visualization

Manuscript received November 5, 2025; revised December 9, 2025; accepted January 7, 2026; published May 22, 2026.

techniques is expected to yield accurate, interpretable classification.

The proposed framework aims to not only automatically evaluates quality but also offers clear explanations of the classification process. The outcomes of this research are expected to support the creation of intelligent quality inspection systems, advance agricultural digitalization, and bolster the long-term competitiveness of Indonesia's spice exports.

II. Related Works

Research on image-based assessment of agricultural commodity quality has advanced considerably over the past decade, driven by the increasing use of computer vision and deep learning methods to replace subjective and imprecise manual techniques [15, 16]. Previous studies have demonstrated the effectiveness of CNNs in accurately classifying the quality of various agricultural products such as coffee beans, fruits, and spices. Wang *et al.* [17] developed an automated system for evaluating coffee bean quality that achieved an accuracy of 90%. Meanwhile, Pardede *et al.* [18] used the VGG16 model with transfer learning to classify fruit ripeness and obtained a similar accuracy rate by applying Dropout 0.5 as a regularization technique. Similar research by Dai *et al.* [19] on pepper leaf images using the GoogLeNet architecture resulted in higher accuracy and better computational efficiency through inception structure compression and the use of Spatial Pyramid Pooling (SPP) to merge local and global features. Trained on 9,183 images of six types of pepper leaf diseases, the model reached an accuracy of 97.87%, outperforming GoogLeNet-V1 and V3 by 6%. Bilal *et al.* [20] developed an automated CNN-based mango variety classification system using data augmentation techniques, which resulted in 85% training accuracy and 65% validation accuracy, and prediction confidence exceeding 90%. Complementing these efforts, Nithya *et al.* [21] created a CNN-based computer vision system to automatically evaluate mango fruit quality, with an accuracy of 98%. A comparable approach was proposed by Sudjud *et al.* [22], who suggested using the YOLOv9 model to detect cocoa pod rot disease, a major challenge in the cocoa industry. The model was trained on a two-class dataset—healthy and infected—for 10 to 50 epochs and showed notable performance gains as the number of epochs increased. By the 50th epoch, the model achieved precision and recall values above 99%.

Although these studies have been successful in enhancing model accuracy and efficiency, most still emphasize predictive performance without sufficiently addressing model interpretability [23]. In agricultural settings, transparency and explainability of the reasons behind model decisions are essential, especially for non-technical users like farmers, quality inspectors, and industry stakeholders [24].

The inability of a model to explain its classification results often raises doubts about the reliability of automated systems in the field. To overcome this challenge, XAI approaches have recently been applied in research to enhance understanding of the model's decision-making process. Methods such as Grad-CAM and LIME (local interpretable model-agnostic explanations) are able to highlight the image regions that exert the greatest influence on classification results, providing an intuitive visual explanation for why the model recognizes a particular object or category [25]. These methods are crucial for boosting user confidence in the deep learning system used to assess the quality of agricultural commodities [26].

III. Research Method

This study developed an explainable deep learning model to classify images into three quality categories—good, moderate, and rotten—while providing a clear visual explanation of its decision-making process. The proposed architecture combined the transfer learning capabilities of modern convolutional networks with XAI methods to ensure that the system focuses not only on accuracy but also on interpretability for non-technical users.

The model was based on the EfficientNet-B0 architecture, which was selected for its efficiency and strong visual feature representation with relatively low computational complexity [27].

EfficientNet-B0 was chosen over VGG16, ResNet50, and MobileNet due to its superior balance of accuracy, computational efficiency, and interpretability. VGG16 and similar deep models were prone to overfitting on small datasets like the 300 images used in the study and required substantial computational resources. ResNet50 offered strong feature extraction but introduced higher model complexity and longer training time, making it less suitable for lightweight agricultural applications. MobileNet was efficient but produced less stable and less interpretable Grad-CAM activation maps in preliminary tests. In contrast, EfficientNet-B0's compound scaling enabled effective feature learning on texture-rich nutmeg images while maintaining parameter efficiency, providing the best trade-off between performance, generalization, and explainability for this task.

The architecture leveraged compound scaling to balance network depth, width, and resolution, leading to optimal performance with fewer parameters [28]. The 224×224×3 input image was processed through a series of convolutional and pooling layers to extract hierarchical features that represented the nutmeg's surface pattern, color variation, and texture. These features were subsequently flattened and passed to a fully connected layer with a ReLU activation function, along with a Dropout of 0.5 to prevent overfitting. A SoftMax activation function was utilized for the output layer to generate classification probabilities for three nutmeg quality classes.

Model training was performed using the Adam optimizer with an adaptive learning rate of 0.0001 and a categorical cross-entropy loss function. The model was trained for 50 epochs with a batch size of 32 and included an early stopping mechanism that halted training when validation accuracy showed no significant improvement. To enhance generalization, data augmentation—including rotation, horizontal flipping, brightness adjustments, and random zooming—was applied throughout the training process.

The initial phase of the research involved collecting and labeling nutmeg seed images into visual quality categories. Each image was taken with a high-resolution digital camera under consistent lighting from a diffuse LED source to reduce shadows and reflections. The nutmeg seeds were placed on a neutral background to ensure consistent color and facilitate replication of the study. The pre-processing phase included resizing, normalizing pixel values to the [0,1] range, and data augmentation to accurately represent the variations in real-world nutmeg seed conditions.

EfficientNet-B0 served as the convolutional backbone for feature extraction. The initial layers of the network extracted low-level visual details such as edges and color gradients. In contrast, deeper layers captured high-level semantic features related to the texture and surface defects of nutmeg seeds. These feature representations were then used by the classifier head to produce final quality predictions. In addition to the classification component, the model was equipped with an XAI module employing two main approaches: Grad-CAM and LIME. Grad-CAM highlighted important regions of an image that influenced classification results by generating a heatmap from convolutional activation weights. Meanwhile, LIME provided a local interpretation of the prediction by identifying the pixels that most contributed to the model's decision. The combination of these two methods allowed users, including farmers and quality assessors, to understand the visual factors behind the classification decisions, thereby enhancing transparency and trust in the system.

Model evaluation was conducted using both quantitative and qualitative methods. Quantitatively, performance was evaluated through accuracy, precision, recall, F1-score, and ROC-AUC metrics to determine the model's ability to differentiate and its overall reliability. Qualitatively, Grad-CAM and LIME visualizations were examined to confirm that the model's focus areas matched relevant visual features, such as surface cracks, discoloration, or texture irregularities. This comprehensive evaluation ensured that the model not only achieved high predictive accuracy but also maintained interpretability consistent with human visual reasoning, making it suitable for nutmeg seed quality assessment. Fig. 1 illustrates the research process flow.

A. Data Collection

The nutmeg seed image dataset was collected via direct photography using a high-resolution digital camera against a plain white background. Lighting conditions were carefully controlled to ensure image consistency. Images were taken from a top-down perspective at a fixed distance and angle, with a diffuse LED light source to minimize shadows and reflections on the nutmeg surface. The images were then categorized into three quality classes—good, moderate, and rotten—based on visual indicators such as surface coloration, textural patterns, and the presence of minor defects. Labeling was performed semi-automatically through the use of visual feature recognition from a pre-trained CNN model as the initial basis for class grouping. The method utilized the pre-trained model's ability to extract color and texture features, reducing the need for manual labeling. All images were stored in 24-bit JPEG format and standardized to a resolution of 224×224 pixels, which made them compatible with the EfficientNet-B0 architecture's input requirements during training. The final dataset was divided into training (70%), validation (15%), testing (15%) subsets to ensure balanced model evaluation and prevent training bias.

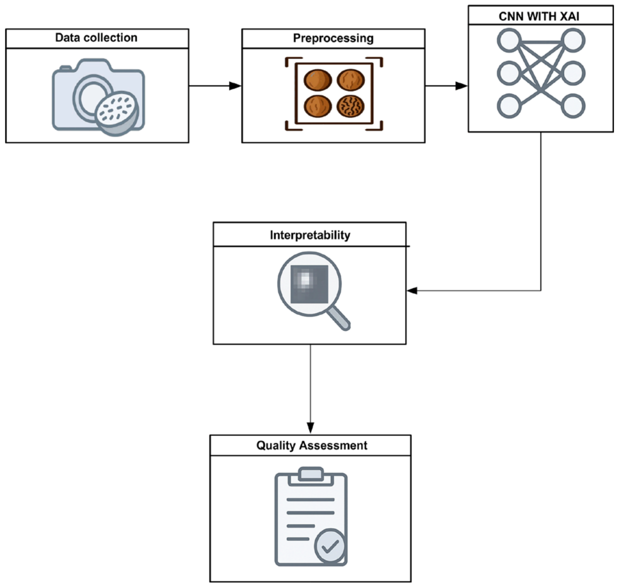

Fig. 1. Research stages flow.

To achieve adequate sample diversity, the dataset was constructed using nutmeg seeds collected from multiple batches and different physical specimens to capture natural variations in texture, color, and surface defects. Although the dataset consisted of 300 images, care was taken to include samples displaying a wide range of quality conditions. The final dataset comprised 100 images per class to minimize class imbalance. To further address the limited dataset size and lower the risk of overfitting, extensive data augmentation techniques—including rotation, flipping, brightness adjustment, and zooming—were implemented. Additionally, regularization strategies such as a dropout rate of 0.5, a validation split of 15%, and early stopping were employed to boost model generalization and mitigate overfitting. These measures collectively strengthened the representativeness of the dataset and ensured more reliable and robust model performance.

B. Data Preprocessing

A preprocessing stage was carried out to make sure that all images shared a consistent format and characteristics prior to being introduced into the model training process. Each collected nutmeg image was resized to 224×224 pixels to match the input size of the pre-trained EfficientNet-B0 architecture. The resolution was chosen to balance computational efficiency with the ability to capture visual features.

To bolster generalization, various random data augmentations were applied, including rotation (±30°), horizontal flipping, zooming, and brightness adjustment, with the aim of increasing image variation without adding more data, while simulating real-world conditions like changes in orientation and lighting.

The dataset was also shuffled prior to training to maintain an even class distribution within each batch, preventing the model from becoming biased toward any particular class.

C. CNN and Explainable Artificial Intelligence (XAI)

The proposed model integrates a Convolutional Neural Network (CNN) with Explainable Artificial Intelligence (XAI) techniques for nutmeg seed quality classification. The CNN component performs feature extraction and classification, while XAI methods are applied to analyze the regions contributing to each prediction. The architecture is based on EfficientNet-B0, which employs compound scaling to balance network depth, width, and input resolution, enabling efficient feature learning with a reduced number of parameters. Early convolutional layers extract low-level features such as edges and color gradients, whereas deeper layers capture higher-level semantic information related to surface texture and defects. The network incorporates mobile inverted bottleneck convolution (MBConv) blocks with Squeeze-and-Excitation (SE) mechanisms to enhance discriminative features [29]. Global average pooling is applied followed by a dropout layer (0.5) to reduce overfitting, and a Softmax layer produces class probabilities for three quality categories. The model is trained using the Adam optimizer with a learning rate of 0.0001 and a categorical cross-entropy loss function for 50 epochs with early stopping and ReduceLROnPlateau strategies [30]. Grad-CAM and LIME are subsequently applied to generate visual explanations of the trained model's predictions.

D. Interpretability

In addition to presenting visual explanations, further clarification was necessary to show how these XAI techniques reinforced the interpretability of the model. Grad-CAM provided a global view of the model's attention by highlighting feature regions that contributed most strongly to classification decisions, enabling users to verify whether the network focused on meaningful characteristics such as discoloration, cracks, or texture defects. LIME complemented this by contributing localized, instance-level explanations that revealed how specific superpixels exerted influence on the predicted class. Together, these methods demonstrated that the model relied on semantically relevant nutmeg features rather than background artifacts, thus increasing user confidence in the decision process. This deeper examination of XAI outputs also highlighted the practical value of explainability for farmers, inspectors, and industry stakeholders who require transparent and trustworthy quality assessment tools.

Interpretability was pivotal to this research as it visually explained the basis for the deep learning model's decisions. To accomplish this, two main approaches from XAI were used: Grad-CAM and LIME. The Grad-CAM method created a heatmap from the final convolutional layer to highlight image areas with the most substantial influence on the classification results. This visualization showed that the model focused on the surface of nutmeg seeds, which displayed distinctive color and texture differences. High-quality images usually show activation concentrated in smooth, evenly colored areas, while low-quality images indicate activation spread across cracked or dark areas. This suggests that the model was learning to recognize visual patterns that corresponded semantically to nutmeg quality features.

The LIME approach complemented these results by adding local interpretations of individual predictions. LIME identified image segments that contributed most to the model's decisions, either positively or negatively. The results demonstrated that brightly colored, smoothly textured areas increased the model's confidence in the good class, while dark areas represented degraded nutmeg images.

While Grad-CAM and LIME offered meaningful visual explanations, their reliability in fine-grained nutmeg seed quality assessment remained limited. Grad-CAM may be sensitive to the choice of convolutional layer and generate coarse activation maps that fail to capture subtle texture variations distinguishing closely related classes such as good and moderate. LIME, despite offering more localized insights, depended heavily on superpixel segmentation, which might lead to inconsistent explanations across visually similar samples. These constraints indicated that XAI results should be interpreted with caution in high intra-class similarity settings. Nevertheless, the complementary nature of Grad-CAM and LIME helped compensate for their individual weaknesses and offered a more balanced interpretability framework for understanding the model's decision-making process.

E. Quality Assessment

The quality assessment phase evaluated model performance both quantitatively and qualitatively and ensured that the classification system not only achieved high accuracy but also delivered stable, consistent, and visually explainable results.

Quantitative evaluation involved calculating key performance metrics, which included accuracy to describes the overall accuracy of the classification against the test data, precision to determine the proportion of correct predictions out of all positive predictions made by the model, and recall to measure the model's ability to identify all true positive samples. The F1-score was used to harmonize precision and recall to address class imbalance. In addition, mean Average Precision (mAP) measured the overall performance of multi-class classification, taking into account the average detection accuracy for each nutmeg quality class.

A confusion matrix further analyzed how predictions were distributed across classes and identified potential bias toward a specific class. The ROC curve and Area Under the Curve (AUC) were employed to examine the model's ability to distinguish among different quality categories.

Qualitative analysis was conducted using Grad-CAM and LIME to verify that the model's focus areas matched relevant visual features, such as color, texture, and surface pattern of nutmeg seeds. Therefore, classification results were evaluated not only by accuracy but also by the model's understanding of the visual traits that truly indicate nutmeg seed quality.

IV. Results and Discussion

This section outlines the experimental design, technical setup, and testing procedures to evaluate the performance of the explainable deep learning model in nutmeg seed quality classification. The main goal of this experiment was to determine how effectively the EfficientNet-B0-based model, combined with Grad-CAM and LIME, accurately and visually identifies the levels of nutmeg seed quality.

A. Experimental Setup

The dataset included 300 images of nutmeg seeds captured in real time with a high-resolution digital camera. Images were taken under controlled lighting conditions against a white background to ensure consistent color and minimize shadows. Each image was manually labeled according to its visual surface condition into three quality categories: good, moderate, and rotten. The labeling considered factors such as seed coat color, texture uniformity, and visible physical defects.

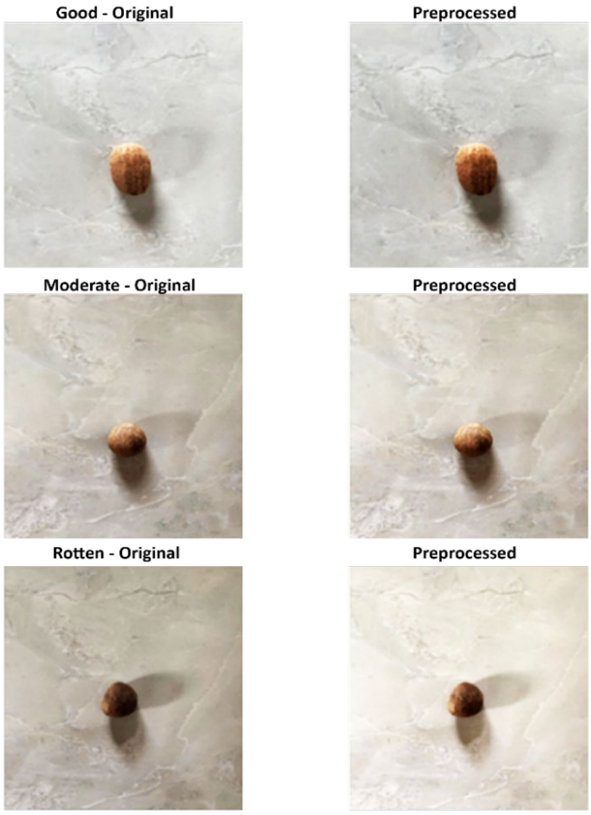

Fig. 2. Nutmeg seed image preprocessing.

The results of the nutmeg seed image preprocessing are presented in Fig. 2. The examples represent three quality categories in its original and preprocessed form, using the preprocess_input() function of the EfficientNet-B0 architecture. The visualization indicates that preprocessing normalizes color intensity and inter-pixel contrast, producing a more uniform image without altering the nutmeg seed's primary structure or texture.

Fig. 3 displays the outcomes of data augmentation. Rotation of up to ±30°, horizontal flipping, and variations in zoom and brightness successfully created significant visual diversity without distorting the shape of the nutmeg. This augmentation artificially boosted the diversity of the training data, enabling the model to recognize the characteristics of nutmeg seeds across different orientations and lighting conditions.

The pre-processing and data augmentation process confirmed that all nutmeg seed images were successfully normalized to 224×224 pixels, aligning with the input requirements of the EfficientNet-B0 architecture. These steps enhanced the images by balancing contrast and lighting without altering key visual features such as shape, texture, and color. As shown in Fig. 2, distinctions among the good, moderate, and rotten quality classes are clearly visible. Images in the good category feature bright, uniform surface colors and smooth textures, while those in the moderate category exhibit uneven brownish color patterns with slight texture irregularities. The rotten class appears darker, with an uneven surface and signs of physical deterioration.

This preprocessing not only enhances overall image clarity but also normalizes pixel intensities, allowing the model to extract relevant features reliably. This guarantees that each image fed into the model maintains a consistent level of representation quality, reducing potential bias caused by variations in lighting or background contrast.

To expand the data range and strengthen the model's generalization, random data augmentation techniques were applied, including ±30° rotation, horizontal flipping, zooming, positional translation, and contrast adjustment. Fig. 3 shows an example of the augmentation results and that the transformations create a variety of orientations and lighting conditions without altering the nutmeg's main shape.

Through the use of the standardized, preprocessed, and augmented datasets, the model was trained to recognize three nutmeg quality categories using the EfficientNet-B0 architecture. The training process lasted for 50 epochs. Table I presents the results of training the nutmeg image quality model.

Table I summarizes the classification performance for each nutmeg quality category. Overall, the precision, recall, and F1-score values indicate that the model achieves consistent classification performance across classes, with some variation. The good category obtained the highest precision (83.33%) but a relatively lower recall (66.77%), indicating that a portion of high-quality samples was misclassified into adjacent classes. In contrast, the moderate category achieved the highest recall (86.66%) and a relatively high precision (76.42%), suggesting effective identification of samples in this class. The rotten category shows a balanced performance with a precision of 75.01% and a recall of 80.34%, indicating reliable detection of visibly degraded nutmeg seeds. The overall mean average precision (mAP) across all classes is approximately 70.29%, reflecting the model's ability to learn discriminative visual features while leaving room for further improvement through model refinement.

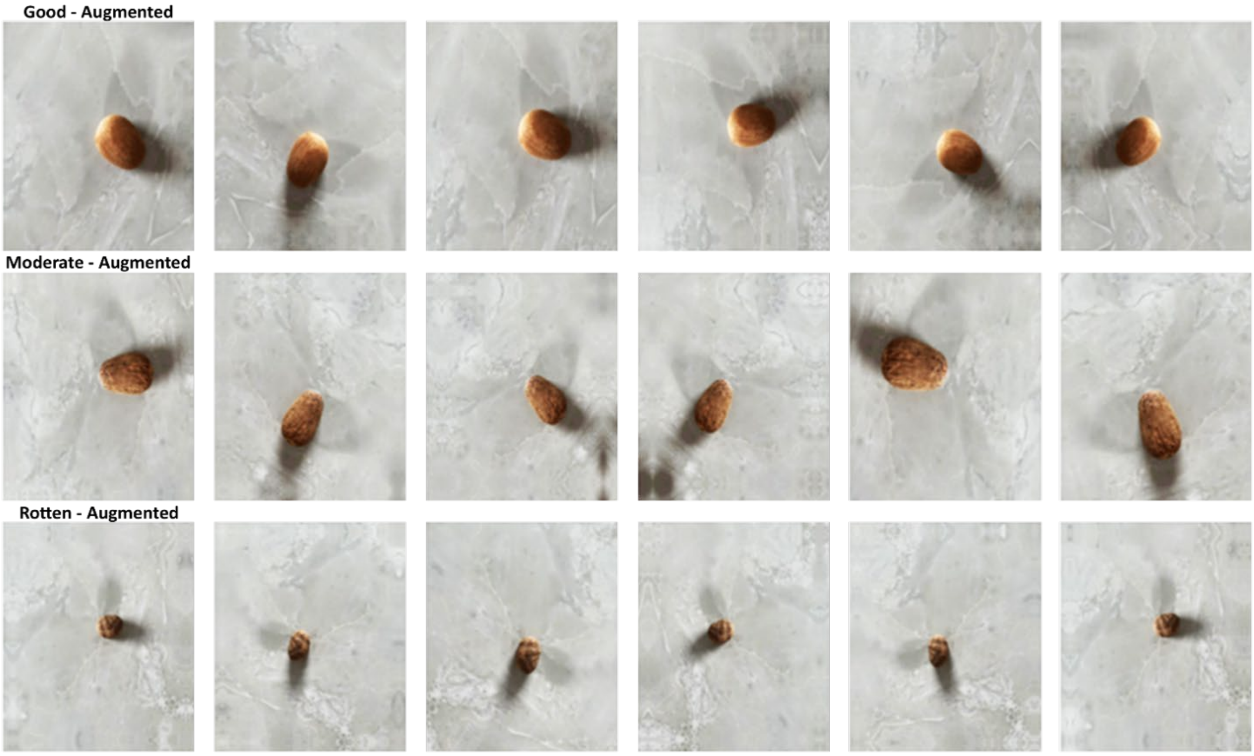

Fig. 3. Nutmeg seed IMAGE augmentation.

TABLE I. MODEL TRAINING RESULTS

Quality Category	Precision (%)	Recall (%)	F1-Score (%)	Accuracy per Class (%)	mAP	Support (Image)
Good	83.33	66.77	74.07	77.21	67.73	100
Moderate	76.42	86.66	81.25	81.33	69.24	100
Rotten	75.01	80.34	77.42	80.07	73.91	100

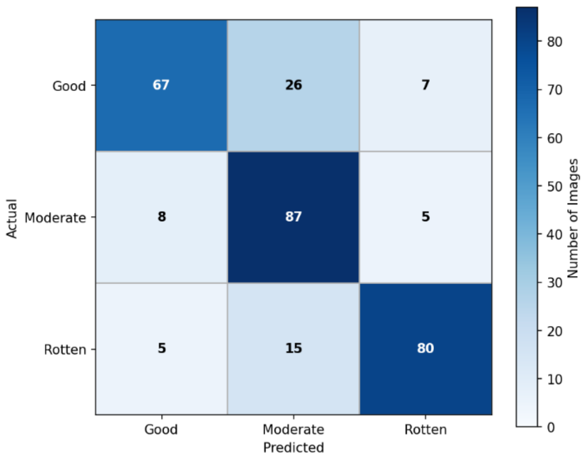

Fig. 4. Confusion matrix.

To better understand the classification pattern, the training results are visualized as a confusion matrix in Fig. 4.

To further clarify how the model arrives at its predictions, this study employed an XAI framework using two main methods: Grad-CAM and LIME. The Grad-CAM visualization in Fig 5 shows red areas as parts of the image with the greatest influence on the model's decisions, while blue areas signify lower contribution. For the good category, the model focused on smooth, uniformly colored seed surfaces, which are a sign of high quality. In the moderate class, Grad-CAM highlighted scattered areas across the surface with uneven brownish hues, indicating concerns. As for the rotten category, the model primarily focused on dark areas and surface spots on nutmeg seeds, which pinpoint notable physical damage and color changes.

Fig. 5 also illustrates that the EfficientNet-B0 model not only detects global features, such as seed shape and size, but also learns local features, including surface texture and color, that distinguish quality classes. These visualizations confirm that the classification process relies on semantically relevant visual patterns, rather than merely differences in background or lighting.

To complement the Grad-CAM visualization, LIME provided explanations of individual predictions for each test image. The LIME image in Fig. 6 features two panels: the original image on the left and the LIME interpretation on the right. The right panel displays a yellow boundary line outlining a superpixel, highlighting the image area most influential in the model's decision to classify nutmeg as good. The enclosed region indicates the part of the nutmeg surface with a smooth texture and uniform color hue, which act as key indicators of high quality. These suggest that the model performs more functions than detecting the overall shape of the nutmeg; it also focuses on important semantically relevant visual features. In other words, the model relies on visual cues that align with human perception of high-quality nutmeg. This LIME approach reinforces previous Grad-CAM interpretations, confirming that the model's predictions are based on true, empirically verifiable features rather than background or lighting artifacts.

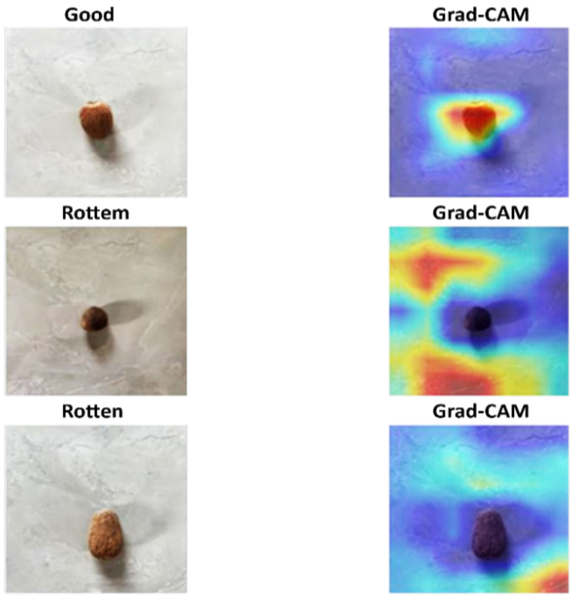

Fig. 5. Heatmap visualization OCF nutmeg seed image.

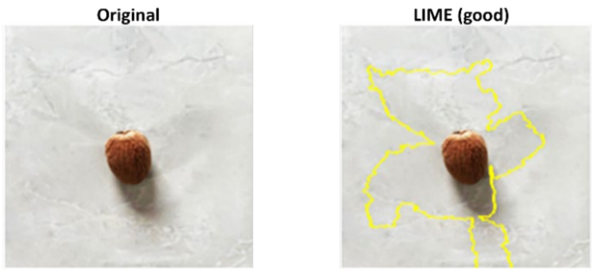

Fig. 6. Lime visualization of nutmeg seed quality image.

B. Discussion

However, the model struggles with images that have similar visual features across different classes, especially between the good and moderate categories. This is because the texture and color patterns on the nutmeg seed surface are relatively similar, making it hard for the model to tell them apart. Such challenges are common in agricultural product quality classification systems, in which class boundaries are often subjective and influenced by lighting conditions and surface wear. This comparison with previous studies is presented in Table II.

Based on the experimental results, the EfficientNet-B0 model performed reliably in classifying three nutmeg seed quality categories. The overall accuracy of 78% and the average mAP of 70.29% demonstrate the model's capacity to consistently identify differences in the visual features of nutmeg seeds, despite the relatively small dataset. This performance suggests that combining the EfficientNet-B0 architecture, data augmentation, and transfer learning effectively balances computational efficiency with classification accuracy.

Based on Table II, most previous studies employed convolutional neural network (CNN)–based approaches for quality classification or disease detection in agricultural products, using architectures with varying levels of complexity. The VGG16 model achieved 90% accuracy through transfer learning and regularization strategies such as Dropout; however, this architecture involves a large number of parameters and requires substantial computational resources [18]. Higher accuracies ranging from 95% to 96.5% for mango variety classification and defect detection were reported, supported by extensive data augmentation and large-scale datasets, but without incorporating interpretability mechanisms [20, 21]. Similarly, a deep CNN achieved 93.4% accuracy for coffee bean quality assessment, although the reliance on complex color and texture features increased computational overhead [17]. The DGLNet model reported the highest accuracy (99.82%) for rice leaf disease classification, benefiting from a lightweight architecture with a small memory footprint (13.5 MB) suitable for field deployment [31].

Compared to these studies, the EfficientNet-B0 model proposed in this research prioritizes a balanced trade-off between performance, computational efficiency, and explainability. Although the achieved accuracy of 78% and overall Mean Average Precision (mAP) of 70.29% do not surpass the highest accuracies reported in large-scale studies, they are considered appropriate for fine-grained nutmeg seed quality assessment, which involves subtle texture variations and color gradients between adjacent classes. Such distinctions are challenging even for human inspectors and are commonly associated with lower performance on small, visually complex datasets. Traditional machine learning approaches generally struggle under these conditions due to their limited ability to capture hierarchical feature representations. In contrast, EfficientNet-B0 leverages multi-level feature extraction to model nuanced visual cues more effectively, while the integration of Grad-CAM and LIME enables visual inspection of the decision process, providing interpretability that is largely absent from earlier studies. Therefore, the obtained performance reflects a realistic and meaningful improvement over conventional baselines for this domain.

TABLE II. COMPARATIVE ANALYSIS OF PREVIOUS STUDIES ON AGRICULTURAL PRODUCT IMAGE CLASSIFICATION

Author & Year	Object / Dataset	Model / Architecture	Augmentation / Technique	Accuracy (%)	mAP (%)	Remarks / Key Findings
[18]	Fruit ripeness (mango, apple, banana)	VGG16 (Transfer Learning)	Dropout 0.5, Batch Normalization	90.00	–	Transfer learning improves performance by 18.42% vs SVM color features.
[20]	Mango classification (6 varieties)	CNN + Linear Classifier	Random flipping, rotation, brightness, contrast	95.00	–	CNN achieves strong generalization; minor misclassification between similar classes.
[21]	Mango defect detection	Deep CNN	Image preprocessing, augmentation	96.50	–	Deep CNN effectively detects surface defects with high precision.
Wang *et al*. [17]	Coffee bean quality	CNN (Deep Learning)	Texture & color features (RGB, L*a*b*)	93.40	–	CNN outperformed ANN-based classifiers; requires high computation.
[31]	Rice leaf disease	DGLNet (Lightweight CNN)	Multi-scale architecture	99.82	13.5 (Model Size MB)	Highest accuracy among light CNN models; effective feature fusion.
Present Study	**Nutmeg seed quality (3 classes)**	**EfficientNet-B0**	**Data augmentation (rotation, flip, zoom)**	**78.00**	**70.29**	**Balanced accuracy and efficiency; added XAI (Grad-CAM, LIME) for interpretability.**

V. CONCLUSION

This research developed an explainable deep learning model based on the EfficientNet-B0 architecture to assess nutmeg quality by classifying images into three main categories: good, moderate, and rotten. The model was trained with 300 preprocessed and augmented images, and demonstrated satisfactory performance with an overall accuracy of 78% and a mean Average Precision (mAP) of 70.29%. The findings indicate the model's ability to extract and recognize key visual features such as color, texture, and surface patterns, which are essential indicators of nutmeg seed quality.

Analysis of the confusion matrix shows that most misclassifications occurred between the good and moderate classes, which share highly similar visual characteristics. However, the model consistently identified the rotten class, where surface cracks, dark spots, and texture degradation are more pronounced. This consistency suggests strong generalization in detecting visual signs of nutmeg quality deterioration. The integration of XAI using Grad-CAM and LIME further enhances transparency, revealing that the model's decision-making process is consistent with semantically relevant regions of the nutmeg surface. Thus, the approach is not only accurate but also interpretable, and therefore it is suitable for AI-driven quality inspection systems in the agricultural sector.

Future work should address several practical challenges to reinforce real-world applicability. First, scalability needs to be evaluated by testing the model on larger datasets and continuous inspection workflows commonly found in industrial processing lines. Second, robustness under varying illumination, background conditions, and camera settings should be further investigated, as such factors may significantly influence model performance outside controlled environments. Third, integration into industrial workflows shall require optimization for real-time inference on edge devices and compatibility with automated grading systems. Additional enhancements, such as fine-tuning deeper layers, exploring lightweight architectures like MobileNetV3 or EfficientNetV2, and leveraging IoT or Edge-AI platforms, may support deployment in low-power field environments and enable real-time nutmeg quality monitoring at production or packaging sites.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Manuel Soares Dos Reis Pacheco conceptualized the research, developed the methodology, and conducted the experiments; Hadiyanto contributed to data preparation, model training, and performance analysis, and assisted in writing and editing the manuscript; Ridwan Sanjaya supervised the research, validated the findings, and reviewed and edited the final manuscript; all authors have read and approved the final version of the manuscript.

REFERENCES

[1] H. Rehatta, J. A. Leatemia, and S. Laimeheriwa, "The impact of climate variables on nutmeg," in *Proc. IOP Conf. Ser. Earth Environ. Sci.*, 2021, vol. 797, no. 1. doi: 10.1088/1755-1315/797/1/012025

[2] W. Sulistiono, M. Assagaf, S. Darniadi, and A. Wahab, "Maintain and improve the quality of nutmeg seeds," in *Proc. IOP Conf.* Ser. *Earth Environ. Sci.*, 2022, vol. 1024, no. 1. doi: 10.1088/1755-1315/1024/1/012080

[3] S. S. A. Rawi, A. H. Ibrahim, H. J. Ahmed, and Z. O. Khudhur, "Therapeutic, and pharmacological prospects of nutmeg seed: A comprehensive review for novel drug potential insights," *Saudi Pharm. J.*, vol. 32, no. 6, 102067, 2024. doi: 10.1016/j.jsps.2024.102067

[4] M. A. Ibrahim, C. L. Cantrell, E. A. Jeliazkova, T. Astatkie, and V. D. Zheljazkov, "Utilization of nutmeg seed hydrodistillation time to produce essential oil fractions with varied compositions and pharmacological effects," *Moleculas*, vol. 25, no. 3, 2020. https://doi.org/10.3390/molecules25030565

[5] S. Sujianto *et al.*, "Assessing Indonesian nutmeg commodity trade competitiveness and developing sustainable strategies in the global market," *Agris On-line Pap. Econ. Informatics*, vol. 16, no. 3, pp. 121–138, 2024. doi: 10.7160/aol.2024.160309

[6] M. S. D. R. Pacheco, H. Hadiyanto, and R. Sanjaya, "Comparative analysis of LeNet-5, AlexNet, and VGGNet for nutmeg seed image quality classification," *Math. Model. Eng. Probl.*, vol. 12, no. 8, pp. 2793–2800, 2025. doi: 10.18280/mmep.120820

[7] X. Zhang, "The AlexNet, LeNet-5 and VGG NET applied to CIFAR-10," in *Proc. 2021 2nd International Conference on Big Data &* Artificial *Intelligence and Software Engineering (ICBASE)*, 2021, pp. 414–419. doi: 10.1109/ICBASE53849.2021.00083

[8] C. Lin, P. Yang, Q. Wang, Z. Qiu, W. Lv, and Z. Wang, "Efficient and accurate compound scaling for convolutional neural networks," *Neural Networks*, vol. 167, pp. 787–797, 2023. https://doi.org/10.1016/j.neunet.2023.08.053

[9] M. Alruwaili and M. Mohamed, "An integrated deep learning model with EfficientNet and ResNet for accurate multi-class skin disease classification," *Diagnostics*, vol. 15, no. 5, pp. 1–18, 2025. doi: 10.3390/diagnostics15050551

[10] X. Liu, "Comparison of four convolutional neural network-based algorithms for sports image classification," in *Proc. 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing*, 2024, pp. 78–186.

[11] S. Ali *et al.*, "Explainable Artificial Intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence," *Inf. Fusion*, vol. 99, 101805, 2023. doi: 10.1016/j.inffus.2023.101805

[12] M. Jamil,"X-CoffeeNet: A novel framework for coffee bean image classification using Explainable Artificial Intelligence (XAI) and MobileNet," *Journal of Advances in Information Technology*, vol. 16, no. 10, pp. 1423–1429, 2025. doi: 10.12720/jait.16.10.1423-1429

[13] B. H. M. Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable Artificial Intelligence (XAI) in deep learning-based medical image analysis," *Med. Image Anal.*, vol. 79, 102470, 2022. doi: 10.1016/j.media.2022.102470

[14] J. Qian, H. Li, J. Wang, and L. He, "Recent advances in explainable artificial intelligence for magnetic resonance imaging," *Diagnostics*, vol. 13, no. 9, 2023. doi: 10.3390/diagnostics13091571

[15] M. R. Islam *et al.*, "Deep learning and computer vision techniques for enhanced quality control in manufacturing processes," *IEEE Access*, vol. 12, pp. 121449–121479, 2024. doi: 10.1109/ACCESS.2024.3453664

[16] N. Manakitsa, G. S. Maraslidis, L. Moysis, and G. F. Fragulis, "A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision," *Technologies*, vol. 12, no. 2, 2024, doi: 10.3390/technologies12020015.

[17] P. Wang, H. W. Tseng, T. C. Chen, and C. H. Hsia, 'Deep convolutional neural network for coffee bean inspection," *Sensors Mater.*, vol. 33, no. 7, pp. 2299–2310, 2021. doi: 10.18494/SAM.2021.3277

[18] J. Pardede, B. Sitohang, S. Akbar, and M. L. Khodra, "Implementation of transfer learning using VGG16 on fruit ripeness detection," *Int. J. Intell. Syst. Appl.*, vol. 13, no. 2, pp. 52–61, 2021. doi: 10.5815/ijisa.2021.02.04

[19] M. Dai *et al.*, "Pepper leaf disease recognition based on enhanced lightweight convolutional neural networks," *Front. Plant Sci.*, vol. 14, no. August, pp. 1–18, 2023, doi: 10.3389/fpls.2023.1230886.

[20] M. Bilal M, . S., S. H K, N. Kumar S, and V. Kumar Kaipa, "Automated mango classification using Convolutional Neural Networks (CNN)," *Int. J. Innov. Sci. Res. Technol.*, pp. 269–275, 2024. doi: 10.38124/ijisrt/ijisrt24sep163

[21] R. Nithya, B. Santhi, R. Manikandan, M. Rahimi, and A. H. Gandomi, "Computer vision system for mango fruit defect detection using deep convolutional neural network," *Foods*, vol. 11, no. 21, 2022. doi: 10.3390/foods11213483

[22] S. Sudjud, M. Jamil, and R. Melati, "Development of cocoa pod rot disease identification model using You Only Look Once (YOLO-v9)," *Math. Model. Eng. Probl.*, vol. 12, no. 4, pp. 1268–1274, 2025. doi: 10.18280/mmep.120418

[23] A. Cartolano, A. Cuzzocrea, and G. Pilato, "Analyzing and assessing explainable AI models for smart agriculture environments," *Multimed. Tools Appl.*, vol. 83, no. 12, pp. 37225–37246, 2024. doi: 10.1007/s11042-023-17978-z

[24] A. Kovari, "AI for decision support: Balancing accuracy, transparency, and trust across sectors," *Inf.*, vol. 15, no. 11, pp. 1–42, 2024. doi: 10.3390/info15110725

[25] H. Zhang and K. Ogasawara, "Grad-CAM-based explainable artificial intelligence related to medical text processing," *Bioengineering*, vol. 10, no. 9, 2023. doi: 10.3390/bioengineering10091070

[26] L. D. Quach, K. N. Quoc, A. N. Quynh, N. T. Nghe, and T. G. Nguyen, "Explainable deep learning models with gradient-weighted class activation mapping for smart agriculture," *IEEE Access*, vol. 11, pp. 83752–83762, 2023. doi: 10.1109/ACCESS.2023.3296792

[27] M. W. A. Khalid, A. Kunaefi, B. A. Nugroho, and A. Yusuf, "Convolutional Neural Network (CNN) efficientnet-B0 model architecture for paddy diseases classification," in *Proc. 2023 14th Int. Conf. Inf. Commun. Technol. Syst. ICTS 2023*, 2023, pp. 105–110. doi: 10.1109/ICTS58770.2023.10330828

[28] P. E. N. Taruno, G. S. Nugraha, R. Dwiyansaputra, and F. Bimantoro, "Monkeypox classification based on skin images using CNN: EfficientNet-B0," *E3S Web Conf.*, vol. 465, 2023. doi: 10.1051/e3sconf/202346502031

[29] J. Hu, L. Shen and G. Sun, "Squeeze-and-excitation networks," in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745

[30] A. A. Kababji, F. Bensaali, and S. P. Dakua, "Scheduling techniques for liver segmentation: Reducelronplateau vs onecyclelr," in *Proc. International Conference on Intelligent Systems and Pattern Recognition*, 2022, pp. 204–212.

[31] Y. Yang, G. Jiao, J. Liu, W. Zhao, and J. Zheng, "A lightweight rice disease identification network based on attention mechanism and dynamic convolution," *Ecol. Inform.*, vol. 78, 102320, 2023. https://doi.org/10.1016/j.ecoinf.2023.102320