Jurnal ELTIKOM :
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 9, No. 2, December 2025, page. 176-183
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v9i2.1566
Available online at http://eltikom.poliban.ac.id

EXPLAINABLE AI-DRIVEN CONVOLUTION NEURAL NETWORK FOR QUALITY GRADING OF SOYBEAN SEEDS

Valencia Sefiana Putri, Setio Basuki*
Faculty of Engineering, Universitas Muhammadiyah Malang, Malang, Indonesia
e-mail: valenciasfn@webmail.umm.ac.id, setio_basuki@umm.ac.id

Received: 17 May 2025 – Revised: 11 August 2025 – Accepted: 12 August 2025

ABSTRACT

This study developed a soybean seed grading system based on Explainable Artificial Intelligence (XAI). Traditional soybean quality assessment is time-consuming, and limited research has applied explainable AI methods to the grading process. To address these issues, this study employed classification and XAI methods through several stages. First, it examined five main categories of soybean seed characteristics: broken, immature, intact, skin-damaged, and spotted. Second, it used the Soybean Seeds Dataset containing 5,513 images. Third, data preprocessing was carried out, including image normalization and data division for training and testing. Finally, a Convolutional Neural Network (CNN) model based on the VGG-16 architecture was used for classification experiments. Three XAI methods, namely Shapley Additive Explanations (SHAP), Local Interpretable Model Agnostic Explanations (LIME), and Layerwise Relevance Propagation (LRP), were applied to evaluate model performance and interpretability. The VGG-16 model achieved an accuracy of 91%, with precision, recall, and F1-score values of 0.91, 0.91, and 0.90, respectively. The interpretability analysis using SHAP, LIME, and LRP showed that the model consistently identified key features such as seed shape and surface texture, demonstrating that the system is transparent and reliable in determining soybean seed quality.

Keywords: *convolutional neural network (CNN), explainable artificial intelligence (XAI), image classification, soybean seeds, VGG-16.*

I. INTRODUCTION

INDONESIA is known as a country rich in diversity and natural resources, one of which is soybean. Soybean is one of the most important plant-based food commodities in Indonesia, holding significant value in both the agricultural sector and the food industry [1]. The National Food Agency (Bapanas) projected that Indonesia would produce between 346,821 and 355,087 tons of soybeans in 2023 [2]. Soybeans are high in protein and are used in a wide range of foods, from traditional fermented products such as miso, natto, and tempeh, to non-fermented products like tofu, soy milk, and edamame, as well as modern food products that use soy protein as a base ingredient [3].

Currently, many time-consuming traditional procedures are still used in the quality evaluation or grading of soybean seeds, where visual quality assessment must be carried out by qualified professionals [4]. The nutritional content, physical condition, and structural integrity of soybean seeds are among the factors that influence their quality. A major challenge in grading soybean seeds is their relatively small size and the large quantity processed in each batch. These characteristics distinguish soybeans from other agricultural products, such as fruits, and require more efficient algorithms to achieve high classification accuracy and speed. Consequently, computer vision and machine learning technologies enable automatic, objective, and accurate soybean grading. Explainable Artificial Intelligence (XAI) has emerged as a new approach to address the complexity and opacity of deep learning models, aiming to produce results that are easier to interpret and understand [5].

Lin et al. [5] successfully developed a deep learning model that could classify soybeans into various quality categories with an accuracy of 95.63%. Zhang et al. [6] used the VGG, AlexNet, DenseNet, and ResNet architectures to classify soybean seeds, with DenseNet121 achieving the highest accuracy of 98.48%. Furthermore, several studies related to the XAI approach have been conducted in the medical

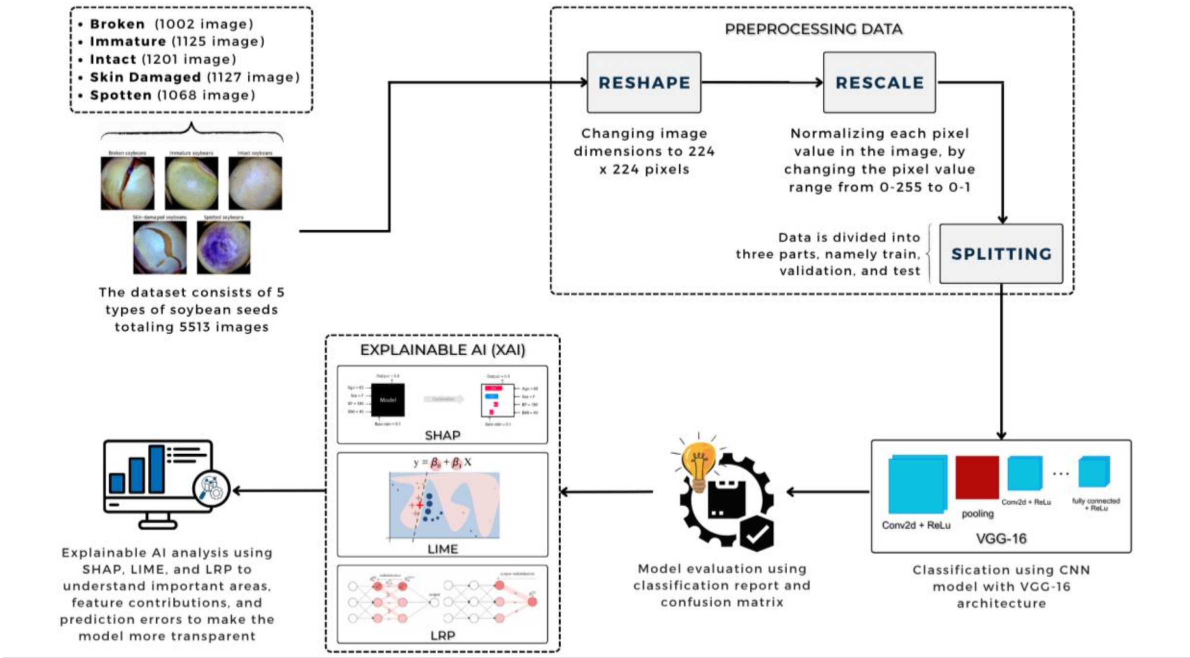


Figure 1. The architecture of the soybean seed grading system using image classification and XAI

field [7]-[10]. Compared to models trained on the original dataset, Bento et al. [11] reported a significant improvement, with a 20% increase in the F1 score when using XAI techniques such as Layerwise Relevance Propagation (LRP).

Through a review of existing literature, it was found that few studies have applied XAI methods to soybean grading, despite the importance of quality standards in the agricultural and food industries. The main issues can be summarized as follows. First, previous studies have primarily focused on traditional approaches without considering model interpretability. Second, no research has specifically employed explainable AI techniques such as SHAP, LIME, or LRP to evaluate soybean grading models. To address these gaps, this study applies Explainable Artificial Intelligence (XAI) to identify the key elements influencing model decisions and to achieve more accurate soybean classification using a deep learning-based model.

Specifically, this paper seeks to answer two research questions: (i) how can a VGG-16-based CNN model be developed for accurate and interpretable soybean classification? and (ii) which XAI method (SHAP, LIME, or LRP) is most effective in visualizing model decisions for quality control applications? The main contributions of this study are as follows: (i) developing a soybean seed dataset consisting of five classes, namely broken, immature, intact, skin-damaged, and spotted, totaling 5,513 instances; (ii) conducting classification experiments using the VGG-16 model that achieved an accuracy rate of 91% across the five categories; (iii) finding that the intact class achieved the highest classification performance, while the immature class was frequently misclassified as intact; (iv) demonstrating that SHAP, LIME, and LRP generated key visual explanations that contributed to the model's decisions and provided insights from different interpretability perspectives; and (v) revealing that among the three XAI methods, LIME produced the most comprehensible visualization, as it highlighted more segmented image areas considered important by the model, making interpretation easier.

II. RESEARCH METHOD

The soybean grading system was implemented through several stages, including image data collection, data preprocessing, classification, model evaluation, and explainable AI analysis. The proposed system architecture is illustrated in Figure 1.

A. Image Data Collection

This study used image data from the soybean seed dataset provided by Jiangsu University of Science and Technology and Nanjing Agricultural University, which was obtained from the Mendeley Data

Figure 2. Images of Soybean Seeds [12]

platform [12]. Each of the five categories in the dataset (broken, immature, intact, skin-damaged, and spotted) contains more than 1,000 soybean seed images, totaling 5,513 images of uniform size (277 × 277 pixels). Figure 2 displays sample images of soybean seeds from each class.

B. Data Preprocessing

The data preprocessing stage is essential for improving data quality and preventing suboptimal results. This stage consists of three main steps: (a) reshaping, which adjusts image dimensions to the size required by the model (224 × 224 pixels); (b) rescaling, which normalizes each pixel value in the image from 0 to 255; and (c) splitting, which divides the dataset into multiple parts so that training and testing data can effectively evaluate the model's performance [13]. In this study, the dataset was divided into three subsets: training, validation, and testing, with proportions of 70%, 10%, and 20%, respectively.

C. Classification stage

This study employed a Convolutional Neural Network (CNN) model based on the VGG-16 architecture, a deep learning approach for the segmentation and classification of soybean seeds aimed at solving the automatic sorting problem. The model was developed by Karen Simonyan and Andrew Zisserman at the University of Oxford in 2014 [14]. VGG-16 is a deep learning architecture consisting of 16 trainable layers, including 13 convolutional layers, 3 fully connected layers, and 5 max-pooling layers that reduce the spatial dimensions of each convolutional output. The input to VGG-16 is an RGB image with dimensions of 224 × 224 × 3 pixels, and the output represents the probability of each category. Each convolutional layer uses a 3 × 3 kernel with a ReLU activation function to enhance non-linearity without significantly increasing the number of parameters. A max-pooling layer with a 2 × 2 kernel is applied after each pooling step to progressively reduce the image's spatial dimensions.

One of the advantages of VGG-16 is its ability to support transfer learning. The model was pre-trained using the ImageNet dataset, which contains over one million images across 1,000 classes. The pre-trained weight sets can be customized for specific tasks through a fine-tuning process. In this study, the VGG-16 architecture was modified by replacing the final fully connected layer to correspond to five classes: broken, immature, intact, skin-damaged, and spotted. The data composition for training, validation, and testing was set at 70% (3,856 instances), 10% (551 instances), and 20% (1,106 instances), respectively, ensuring a balanced dataset distribution. The VGG-16 implementation involved using the pre-trained model available in the PyTorch framework and fine-tuning it on the soybean seed dataset. Model performance was evaluated using common image classification metrics, namely accuracy, precision, recall, and F1-score.

D. Explainable AI analysis

In this study, Explainable Artificial Intelligence (XAI) was applied to describe the model's decision-making process and improve the transparency of the assessment procedure, allowing users to understand how each feature contributes to the prediction results [15]. SHapley Additive exPlanations (SHAP) interprets machine learning model outputs using a game theory-based approach that assigns a Shapley value to each feature to measure its contribution to the prediction [16]. Introduced by Lundberg and Lee in 2017 [17], SHAP provides a comprehensive framework for explaining model predictions using Shapley values from coalition theory. This method calculates the contribution of each feature to the prediction compared with a baseline (or excluded) value, as represented in equation (1), where $v(S)$ denotes the predicted value when only the active features in subset S are considered. Because of its model-agnostic nature, SHAP provides consistent explanations both locally (for individual instances)

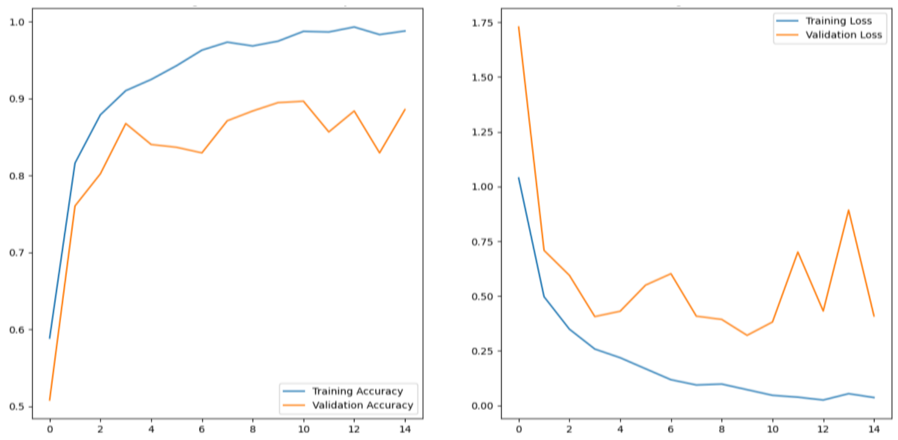


Figure 3. Training and Validation Process on the VGG-16

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)] \tag{1}$$

$$\xi(x) = \arg min_{\{g\, \in G\}}\, L(f, g, \pi_x) + \Omega(g) \tag{2}$$

$$R_i = \sum_j \frac{a_i w_{ij}}{\sum_{i'} a_{i'} w_{i'j} + \epsilon} R_j \tag{3}$$

and globally (for the entire model) while maintaining efficiency, symmetry, and additivity. In this study, SHAP was used to analyze visual contributions in the VGG-16-based soybean seed classification model, implemented through the SHAP library by visualizing heatmaps for interpretability.

Local Interpretable Model-Agnostic Explanations (LIME) were introduced by Ribeiro et al. in 2016 [18]. This method explains classifier predictions by constructing locally interpretable models around specific predictions. LIME generates a local dataset with variations of the original data and then builds a simple interpretable model, such as linear regression, to explain the contribution of features in a data sample through perturbations. To obtain a human-understandable representation, LIME determines whether a continuous path or "superpixel" that provides the best representation of the output class exists, using a binary vector x∈ {0, 1} [19]. operates on a single data input at the patch level, where small input variations can influence the output. Equation (2) defines the method, where $\xi(x)$ represents the local explanation for instance x, f is a complex model, g is an interpretable model, L measures prediction error, π_x assigns weights to surrounding data x, and $\Omega(g)$ maintains the simplicity of g.

Layer-Wise Relevance Propagation (LRP) is a method used to interpret Multi-Layer Perceptron (MLP) models [20]. It explains the predictions of deep learning models, particularly neural networks, by assigning a relevance score to each neuron in all layers, indicating each neuron's contribution to identifying the most influential features in the model's decision-making process [21]. LRP propagates relevance scores backward from the model's prediction output to its inputs, following a layer-wise path. This method adheres to the principle of continuous relevance conservation at each layer and produces interpretations in the form of heatmaps. Together, these three methods aim to enhance the interpretability of the model, making the prediction results easier to understand. The mathematical formulations are presented in (1)–(3).

III. RESULT AND DISCUSSION

This section presents the experimental results for classification and explainable AI analysis. The main performance indicator used in the evaluation is the accuracy. In addition, F1, recall, and precision values were used to assess the overall performance of the model. Three XAI methods were analyzed to enhance result interpretability, namely SHAP, LIME, and LRP. Each method demonstrated that the model's classification relies heavily on visual characteristics such as the size, shape, and texture of soybean seeds.

TABLE 1
CLASSIFICATION PERFORMANCE OF THE VGG-16 MODEL ON THE FIVE CLASSES

Class	Precision	Recall	F1-Score	Accuracy
Broken soybeans	0.89	0.87	0.88	
Immature soybeans	0.92	0.89	0.90	
Intact soybeans	0.88	0.99	0.93	0.91
Skin-damaged soybeans	0.89	0.92	0.91	
Spotted soybeans	0.96	0.84	0.90	

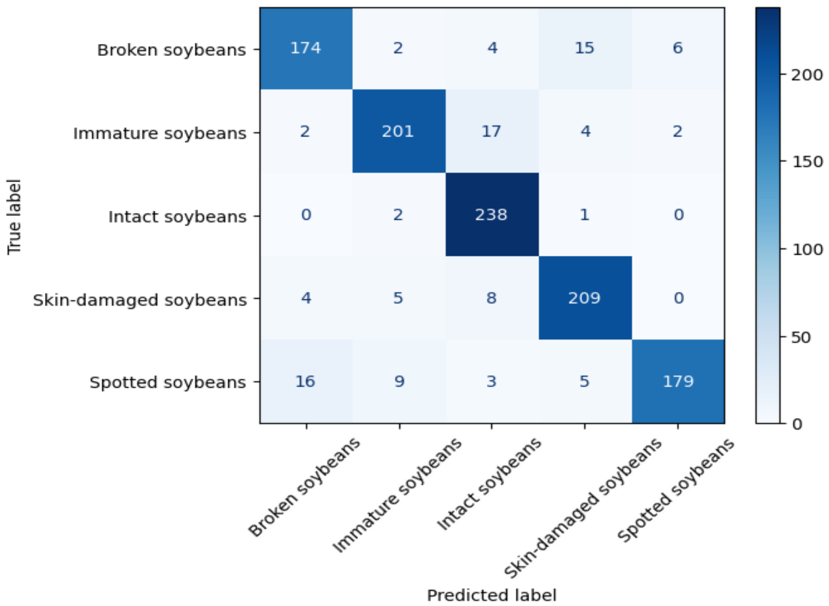


Figure 4. Confusion Matrix of the VGG-16 Model

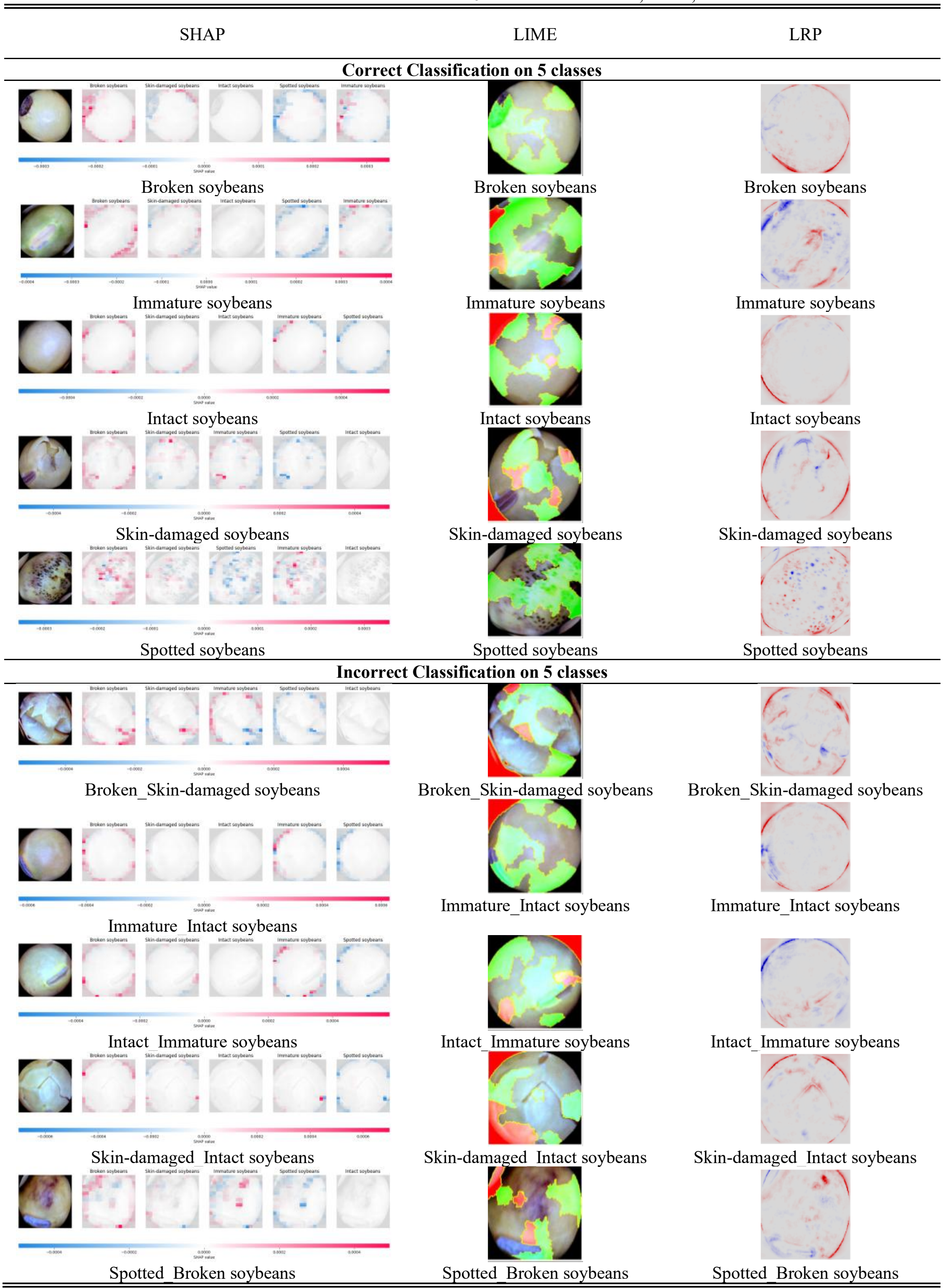
A. Classification Performance

The pretrained VGG-16 model was successfully evaluated from ImageNet with a fine-tuning approach on the last four layers. Stochastic Gradient Descent (SGD) optimization was applied for 20 training epochs with a learning rate of 0.001 and a momentum value of 0.9. As shown in Figure 3, validation accuracy fluctuated up to 0.89, while training accuracy gradually increased to 0.98. Training loss decreased to approximately 0.04, whereas validation loss rose from 0.32 to 0.89. The gap between validation and training losses indicates a possible overfitting issue. To address this, techniques such as data augmentation (to increase training data variability), regularization (for instance, dropout to reduce model complexity), and early stopping (to terminate training when validation performance ceases to improve) can be applied to enhance the model's generalization capability. The classification performance across all categories is summarized in Table 1. The model achieved a macro average precision of 91%, a macro average recall of 90%, a macro average F1-score of 90%, and an overall accuracy of 91%. Although the algorithm demonstrated strong predictive performance, as shown in the confusion matrix in Figure 4, some misclassifications were observed, including immature seeds predicted as intact (17%), spotted seeds predicted as broken (16%), and broken seeds predicted as skin-damaged (15%). For each soybean classification category, the model's high precision and recall indicate strong predictive reliability, and the high accuracy confirms the successful completion of the classification process.

B. Model Interpretation with XAI

Table 2 presents a comparison of the visualization results obtained from the three Explainable AI (XAI) methods. This visualization is divided into two main sections: correct and incorrect classifications across the five soybean seed classes. SHAP, LIME, and LRP were implemented to analyze the soybean classification process and interpret the decisions of the VGG-16-based CNN model using relevant Python libraries. Each method highlights the important areas of the soybean image that contribute to the classification of the five categories: broken, immature, intact, skin-damaged, and spotted soybeans.

TABLE 2
VISUALIZATION RESULTS OF THE VGG-16 MODEL USING SHAP, LIME, AND LRP



The evaluation of model interpretation compares the highlighted regions produced by each XAI method with the visual characteristics of each class. SHAP visualizations present feature contributions through a gradient heatmap, where red areas indicate positive contributions, particularly in surface texture, and blue areas represent negative contributions. LIME provides a more detailed local interpretation by segmenting the image into superpixels, identifying compact and clearly defined regions that influence classification. LRP visualizations also display pixel-level contributions, although they tend to be less focused and more abstract.

Based on the analysis, all three XAI methods offer a comprehensive understanding of the model's decision-making process. SHAP produces smooth heatmaps that integrate well with the original image, since it operates at the pixel level. In contrast, LIME uses superpixel segmentation that resembles puzzle pieces, forming distinct boundaries and shapes. LRP, on the other hand, generates more diffuse gradient maps. These results confirm the model's ability to recognize and utilize visual cues such as texture, shape, and contour to identify physical damage in soybean seeds.

Some predictions remain inaccurate based on the confusion matrix, particularly for classes such as intact and immature soybeans, as well as fractured and skin-damaged soybeans, which share similar shapes and textures. This indicates that the model has difficulty distinguishing visual characteristics among closely related categories. To address this issue, techniques such as early stopping, dropout, and data augmentation should be applied to increase data variability and improve the model's ability to differentiate class characteristics.

This interpretation has several limitations. First, model interpretability and accuracy (91%) are not always aligned. For example, LRP produces a more abstract heatmap than SHAP, even though both methods examine similar aspects such as surface texture. This highlights a trade-off between the practical usefulness of the interpretation and its level of detail. Second, real-world implementation requires integration with domain knowledge. For example, segmentation-based interpretation using LIME is more suitable for agricultural experts to identify physical damage, while heatmaps generated by SHAP are more appropriate for analyzing general patterns.

C. Hypothetical Deployment Scenarios

Based on a theoretical framework, the proposed system can be developed into two potential applications: (a) an automated sorting station in a seed processing plant equipped with industrial cameras and conveyor belts, and (b) a field diagnostic tool operated via mobile devices. In the first case, operators can perform real-time quality verification with the help of LIME visualization (Figure 5). In the second scenario, the system architecture must be optimized using quantization and pruning techniques to ensure compatibility with edge devices such as NVIDIA Jetson or smartphones.

The system also faces several implementation challenges: (i) Computational requirements: the VGG-16 architecture requires a GPU with a thermal design power (TDP) of at least 75W, which limits optimal use to specific hardware configurations; (ii) Environmental resilience: no tests have been conducted to assess the system's performance under variable lighting, high humidity, or exposure to dust particles; and (iii) System integration: an interface compatible with IoT sensors, including humidity and temperature sensors, is required. Potential solutions include developing hybrid models that combine CNN with lightweight algorithms, applying GAN-based data augmentation to simulate field conditions, and collaborating with industry partners to design integrated hardware systems.

D. Limitations

This study has several limitations: (i) as shown by the 9% gap between training and validation accuracy in Figure 3, the model does not perform optimally under varying lighting conditions and camera angles. This limitation arises because the dataset used consists of controlled soybean images with a neutral background and uniform lighting. (ii) To detect fine features such as surface texture, soybean images must have high resolution (224 × 224 pixels) for optimal model performance. (iii) The model's generalization capability is limited, as the texture patterns of skin damage identified through XAI-based visual characteristics may not function effectively when applied to other datasets without adjustment. The researchers recommend adopting lightweight architectures, performing cross-validation on other commodities, and applying data augmentation techniques that reflect realistic field conditions to obtain more robust results.

IV. CONCLUSION

The VGG16-based deep learning model developed to classify the quality of soybean seeds demonstrated strong performance with high accuracy and effectively recognized relevant visual characteristics during the classification process. The interpretability analysis using SHAP, LIME, and LRP revealed that each method offered distinct feature contributions. Among them, LIME provided the most comprehensible visualizations, as it produced segmented areas highlighting parts of the image considered important by the model, thus making interpretation easier. The findings of this study suggest several directions for future improvement. To minimize potential overfitting, researchers could explore more complex model architectures or employ data augmentation methods. Furthermore, future work may focus on (i) integrating the model into real-time systems, (ii) extending generalization to other seed types, and (iii) conducting user-based evaluations to support transparent and automated quality control in agriculture.

REFERENCES

[1] A. T. Kharisma and R. Efendi, "Perancangan mesin sortasi kedelai berdasarkan warna untuk menentukan kualitas benih," *Sultra J. Mech. Eng. SJME*, vol. 1, no. 1, pp. 33–47, Oct. 2022, doi: 10.54297/sjme.v1i1.307.

[2] "Badan Pangan Nasional - NFA Apresiasi Pengrajin Tahu Tempe dan Dorong Penguatan Ekosistem Kedelai Nasional - Blog." Accessed: Feb. 20, 2025. [Online]. Available: https://badanpangan.go.id/blog/post/nfa-apresiasi-pengrajin-tahu-tempe-dan-dorong-penguatan-ekosistem-kedelai-nasional

[3] M. Messina, "Perspective: Soybeans Can Help Address the Caloric and Protein Needs of a Growing Global Population," *Front. Nutr.*, vol. 9, p. 909464, May 2022, doi: 10.3389/fnut.2022.909464.

[4] D. C. Santana et al., "Machine learning in the classification of asian rust severity in soybean using hyperspectral sensor," *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, vol. 313, p. 124113, May 2024, doi: 10.1016/j.saa.2024.124113.

[5] W. Lin, L. Shu, W. Zhong, W. Lu, D. Ma, and Y. Meng, "Online classification of soybean seeds based on deep learning," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106434, Aug. 2023, doi: 10.1016/j.engappai.2023.106434.

[6] N. Zhang, E. Zhang, and F. Li, "A Soybean Classification Method Based on Data Balance and Deep Learning," *Appl. Sci.*, vol. 13, no. 11, p. 6425, May 2023, doi: 10.3390/app13116425.

[7] R. Ghnemat, S. Alodibat, and Q. Abu Al-Haija, "Explainable Artificial Intelligence (XAI) for Deep Learning Based Medical Imaging Classification," *J. Imaging*, vol. 9, no. 9, p. 177, Aug. 2023, doi: 10.3390/jimaging9090177.

[8] D. Bhati, F. Neha, and M. Amiruzzaman, "A Survey on Explainable Artificial Intelligence (XAI) Techniques for Visualizing Deep Learning Models in Medical Imaging," *J. Imaging*, vol. 10, no. 10, p. 239, Sep. 2024, doi: 10.3390/jimaging10100239.

[9] Y. Yang, G. Mei, and F. Piccialli, "A Deep Learning Approach Considering Image Background for Pneumonia Identification Using Explainable AI (XAI)," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 21, no. 4, pp. 857–868, Jul. 2024, doi: 10.1109/TCBB.2022.3190265.

[10] M. Bhandari, T. B. Shahi, B. Siku, and A. Neupane, "Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI," *Comput. Biol. Med.*, vol. 150, p. 106156, Nov. 2022, doi: 10.1016/j.compbiomed.2022.106156.

[11] V. Bento, M. Kohler, P. Diaz, L. Mendoza, and M. A. Pacheco, "Improving deep learning performance by using Explainable Artificial Intelligence (XAI) approaches," *Discov. Artif. Intell.*, vol. 1, no. 1, p. 9, Dec. 2021, doi: 10.1007/s44163-021-00008-y.

[12] W. Lin, J. Fang, Q. Su, H. Liao, S. Liu, H. Yao, and P. Xu, "SMPTS: Segmentation Method for Physically Touching Soybean Images," in *Proc. 9th Int. Conf. Image, Vision and Computing (ICIVC)*, Jul. 2024, pp. 177–182.

[13] I. Muraina, "Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts," in *Proc. 7th Int. Mardin Artuklu Sci. Res. Conf.*, 2022.

[14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, *arXiv*. doi: 10.48550/ARXIV.1409.1556.

[15] W. Saeed and C. Omlin, "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities," *Knowl.-Based Syst.*, vol. 263, p. 110273, Mar. 2023, doi: 10.1016/j.knosys.2023.110273.

[16] Sangeeta Sharma, Karan Kaushik, Rachita Sharma, and Nikita Chaturvedi, "Explainable Artificial Intelligence (XAI)," *Int. J. Food Nutr. Sci.*, vol. 12, no. 1, Oct. 2023, doi: 10.48047/ijfans/v12/i1/271.

[17] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," 2017, *arXiv*. doi: 10.48550/ARXIV.1705.07874.

[18] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[19] A. Das and P. Rad, "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey," 2020, *arXiv*. doi: 10.48550/ARXIV.2006.11371.

[20] P. Sukpanichnant, A. Rago, P. Lertvittayakumjorn, and F. Toni, "LRP-based argumentative explanations for neural networks," in *Proc. 2nd Italian Workshop on Explainable Artificial Intelligence*, Jan. 2021, pp. 71–85.

[21] Y.-J. Jung, S.-H. Han, and H.-J. Choi, "Explaining CNN and RNN Using Selective Layer-Wise Relevance Propagation," *IEEE Access*, vol. 9, pp. 18670–18681, 2021, doi: 10.1109/ACCESS.2021.3051171.