Classification of Shallot Leaf Health Based on RGB Images Using Machine Learning Algorithms

Fradilla[1*], Muhammad Fajar B[2], Tiara Rahmadani Asri[3]

[1*,3] Computer Engineering, Universitas Negeri Makassar, Makassar 90222, Indonesia

[2*]Informatics and Computer Engineering, Universitas Negeri Makassar, Makassar 90222, Indonesia

Correspondence e-mail: fradilla9905@gmail.com[1*]

ABSTRACT

Agricultural productivity often declines due to plant diseases that reduce yield and quality. Shallot (Allium ascalonicum L.) is one of Indonesia's key horticultural commodities, yet it is highly vulnerable to leaf diseases such as purple blotch and Fusarium-induced moler. This study aims to develop a shallot leaf health classification model using RGB images and machine learning algorithms. The proposed system employs an experimental approach based on a publicly available image dataset consisting of three categories: healthy, purple blotch, and moler-infected leaves. Preprocessing stages include image resizing, noise reduction, and contrast enhancement to improve visual clarity. Feature extraction combines RGB color histograms and Gray-Level Co-occurrence Matrix (GLCM) texture descriptors to obtain informative features. Two algorithms Support Vector Machine (SVM) and Random Forest (RF) were trained and evaluated using accuracy, precision, recall, and F1-score metrics. The results show that both models achieved perfect classification performance, with RF demonstrating slightly higher stability and robustness. These findings confirm that the integration of RGB imagery and lightweight machine learning algorithms provides a reliable, low-cost, and computationally efficient solution for early detection of shallot leaf diseases. The proposed approach contributes to precision agriculture development and offers potential deployment for smallholder farmers through simple, camera-based monitoring systems.

ARTICLE INFO

1. INTRODUCTION

Agriculture remains one of the most essential sectors for ensuring food security, economic growth, and sustainable development in many countries, including Indonesia. Among its numerous horticultural commodities, shallots (Allium ascalonicum L.) hold a vital role in both household consumption and industrial food production. Shallots are not only a key ingredient in Indonesian cuisine but also a strategic agricultural product that significantly contributes to the income of smallholder farmers and

regional economies [1]. However, productivity levels remain inconsistent due to frequent disease outbreaks that severely affect leaf health and overall yield quality. Diseases such as purple blotch, caused by Alternaria porri, and moler disease, induced by Fusarium oxysporum, are among the most common threats to shallot cultivation. These infections can lead to yield losses reaching up to 50% when left undetected or untreated at early stages [2], [3].

Traditional disease detection methods still rely heavily on manual visual inspection by farmers or agricultural experts. Although this approach has been practiced for decades, it remains highly subjective, time-consuming, and prone to human error, especially when applied to large-scale plantations [4]. Moreover, the shortage of agricultural pathologists in rural regions further hinders rapid and accurate disease identification. These limitations emphasize the need for modern, automated systems capable of performing precise and efficient disease detection without requiring extensive technical expertise [5]. With the growing integration of digital technologies in agriculture, artificial intelligence (AI) and computer vision have emerged as transformative tools for precision farming and crop health monitoring [6].

In recent years, image-based disease classification using RGB (Red, Green, Blue) images has gained prominence due to its simplicity and cost-effectiveness. RGB images can be easily captured using standard cameras or mobile phones, making this approach particularly suitable for smallholder farmers who lack access to advanced imaging equipment [7]. Several studies have demonstrated that combining image processing with machine learning algorithms can achieve accurate classification of plant health conditions [8], [9]. The Support Vector Machine (SVM) algorithm, for instance, is known for its strong generalization ability and robustness in handling nonlinear data, while Random Forest (RF) offers excellent accuracy and resistance to overfitting through its ensemble structure [10], [11]. Both algorithms have shown outstanding performance in agricultural disease detection tasks involving small- and medium-sized datasets.

A growing body of literature has explored various feature extraction techniques to improve classification accuracy in agricultural imagery. Commonly employed methods include color histogram analysis, Gray-Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and texture-based descriptors [12]. These techniques allow the extraction of distinctive color and texture information that differentiate healthy leaves from diseased ones. Studies by [13] and [14] reported that integrating color and texture features significantly enhances classification outcomes compared to using a single type of feature. Furthermore, feature-level fusion has been shown to reduce noise sensitivity and improve model stability, especially under diverse illumination and environmental conditions [15].

Although deep learning methods such as Convolutional Neural Networks (CNNs) have revolutionized image recognition tasks, their deployment in small-scale agricultural contexts remains challenging due to high computational demands and the need for large annotated datasets [16]. For local farmers or small research institutions, implementing CNN-based solutions often becomes impractical because of limited resources, hardware constraints, and internet connectivity issues [17]. Consequently, lightweight machine learning algorithms such as SVM and RF remain highly relevant alternatives, offering strong performance with lower complexity and shorter training times [18]. These algorithms can be trained efficiently on modest computing platforms, making them ideal for real-world agricultural applications.

Despite numerous advancements, research on shallot leaf disease classification remains limited compared to other crops such as rice, tomatoes, and maize [19]. Many prior studies have focused on general leaf disease identification without addressing specific morphological characteristics of shallot leaves, which exhibit thin, elongated shapes and subtle color gradients. The limited dataset availability and lack of standardized image acquisition procedures have further constrained the accuracy and generalizability of previous models [20]. Additionally, most existing works rely on small datasets

collected under controlled laboratory conditions, which may not fully represent real-world variations in lighting, leaf orientation, and background noise [21].

To address these gaps, this research aims to develop a robust and efficient shallot leaf health classification system based on RGB image features using SVM and RF algorithms. The proposed system integrates color and texture features extracted through RGB histograms and GLCM matrices to enhance model discriminability. This approach allows for the accurate identification of three distinct classes healthy, purple blotch-infected, and moler-infected leaves while maintaining computational efficiency suitable for low-cost devices [20]. The methodology involves systematic stages including dataset acquisition, preprocessing, feature extraction, classification, and evaluation, ensuring the reproducibility and transparency of results.

The novelty of this study lies in its combination of traditional image processing and modern machine learning for agricultural disease classification using lightweight, resource-friendly algorithms. Unlike prior studies that primarily employed deep learning approaches with high resource requirements, the proposed model achieves comparable accuracy through optimized feature engineering and hybrid classification techniques [22]. Furthermore, the development of a Streamlit-based web application enables easy accessibility for farmers and agricultural officers to conduct rapid disease detection in real time. This practical implementation aligns with the goals of precision agriculture and supports sustainable farming practices by minimizing chemical pesticide use through early disease identification [23].

In summary, this study provides an innovative yet practical contribution to digital agriculture by leveraging RGB imagery and machine learning to monitor shallot leaf health. The research not only enhances the accuracy and speed of disease detection but also supports smallholder farmers with affordable technology that can be applied directly in the field. By addressing both theoretical and practical aspects, this work contributes to bridging the gap between advanced AI techniques and real-world agricultural needs.

2. METHOD

2.1 Research Design

This research adopts an experimental quantitative design based on computer simulation to develop a classification model for shallot leaf health using RGB image data. The purpose of this stage is to systematically build and evaluate a lightweight, efficient machine learning system capable of identifying three leaf health conditions healthy, purple blotch-infected (Alternaria porri), and moler-infected (Fusarium oxysporum). The workflow was structured to follow the principles of reproducible scientific computing [1], [2].

The methodological framework includes six primary stages: (1) dataset collection, (2) preprocessing, (3) feature extraction, (4) model training, (5) performance evaluation, and (6) system implementation. Each stage was designed to ensure methodological rigor and replicability, with standardized input-output structures for data processing. The experimental process was conducted using Python as the main programming language, incorporating libraries such as OpenCV, scikit-learn, and NumPy for computational analysis.

This design aligns with prior research emphasizing the use of digital image analysis in precision agriculture for disease detection [3], [4]. In contrast to deep learning, which demands high-end hardware and large datasets, this study focuses on lightweight algorithms that are feasible to deploy on modest computing resources. The research design ensures that results can be reproduced by other researchers with similar datasets and software environments, following open-science standards [5].

2.2 Data Collection

The dataset used in this study was obtained from the Shallot Disease Detection dataset available on the Roboflow platform. This dataset contains images of shallot leaves categorized into three classes: healthy, purple blotch, and moler. Each image has an average resolution of 128×128 pixels, captured under natural lighting conditions using standard RGB cameras. The images were manually annotated and validated by agricultural experts to ensure label accuracy, following the protocol in [6] and [7].

In this study, three primary classes are defined consistently throughout the manuscript, namely Healthy Shallot Leaf, Fusarium-Infected Shallot Leaf, and Weed-Contaminated Leaf. These class labels are used uniformly in all tables, figures, and discussions to avoid ambiguity and ensure clarity in the classification task.

In total, the dataset comprised approximately 400 labeled images. The data were divided into three subsets: 70% for training, 15% for validation, and 15% for testing, ensuring balanced class distribution. Each subset was stored separately to avoid data leakage between training and evaluation phases. To enhance dataset consistency, several inclusion and exclusion criteria were applied. Only clear, focused, and full-leaf images were included. Images with excessive background noise, shadows, or partial leaves were excluded to maintain visual integrity. This approach follows best practices in plant phenotyping and agricultural imaging studies [8], [9].

All dataset handling adhered to FAIR data principles (Findable, Accessible, Interoperable, Reusable), ensuring that the dataset could be reused in future experiments and model benchmarking [10].

2.3 Image Preprocessing

Before model training, all images underwent a series of preprocessing operations to enhance data quality and remove irrelevant noise. Each image was first resized to 128×128 pixels to ensure uniformity. The color format was converted from BGR (Blue–Green–Red) to RGB (Red–Green–Blue), followed by median filtering to reduce minor noise artifacts without blurring edge details [11].

After noise reduction, histogram equalization was applied to improve contrast and emphasize color intensity variations between healthy and diseased leaves. This process enhanced visual separability, especially for detecting necrotic spots and color degradation typical of purple blotch and moler symptoms [12]. To improve color consistency across lighting variations, each image was transformed into the HSV (Hue, Saturation, Value) color space. This transformation stabilizes color features under varying illumination conditions, a crucial factor for outdoor agricultural imagery [12], [13].

The preprocessed images were then normalized and converted into NumPy arrays as the final input for feature extraction. This standardization process ensures data consistency and computational efficiency, in line with similar approaches used in agricultural image classification [14], [15].

Although the final feature extraction was performed using RGB histograms, the HSV color space was employed during the preprocessing stage to enhance robustness against illumination variations. This step improved color normalization while preserving the original RGB information used for feature representation.

2.4 Feature Extraction

Feature extraction plays a central role in transforming visual data into measurable numeric representations that can be processed by machine learning algorithms. In this study, two feature types were extracted: color features and texture features. Color features were obtained using RGB color histograms. Each channel (Red, Green, and Blue) was divided into eight intensity bins, resulting in a

24-dimensional color feature vector for each image. These histograms represent the color distribution pattern across the leaf surface, which helps distinguish between healthy and infected regions [16].

Texture features were extracted using the Gray-Level Co-occurrence Matrix (GLCM) method, which quantifies the spatial relationships between neighboring pixels. Four statistical descriptors were calculated from the GLCM: contrast, correlation, energy, and homogeneity [17]. These texture descriptors are sensitive to changes in leaf surface texture, such as roughness or discoloration, which are indicators of fungal infection. The final feature vector was formed by combining RGB color features and GLCM texture descriptors into a unified feature matrix. This hybrid feature representation enhances model discriminability, as supported by studies in [18] and [19], which demonstrated that integrating color and texture features yields higher accuracy than using either feature type alone.
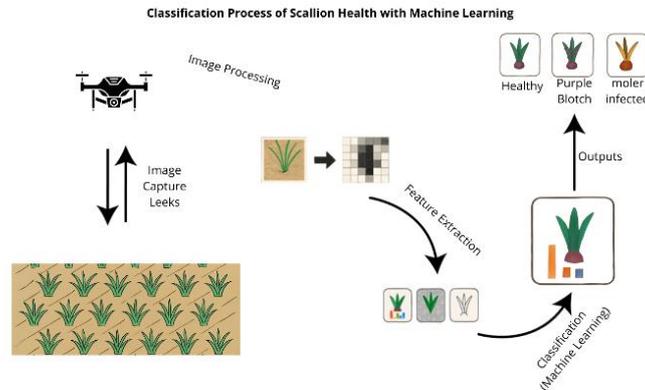

Figure 1. Classification Process of Scallion Health with Machine Learning

The feature extraction process is illustrated in Figure 1, which depicts the sequence from RGB image input to the generation of numerical descriptors used for classification.

2.5 Classification Evaluation

Two supervised machine learning algorithms Support Vector Machine (SVM) and Random Forest (RF) were implemented to classify shallot leaf health conditions. The SVM classifier employed a Radial Basis Function (RBF) kernel due to its high capability in handling non-linear data distributions, effectively mapping feature vectors into a higher-dimensional space for optimal separation between classes [20]. The RF classifier was constructed using an ensemble of 100 decision trees, aggregating results through majority voting to enhance stability and minimize overfitting [21].

Prior to model training, data normalization was performed using StandardScaler to ensure uniform feature contribution across all attributes. The dataset was divided into training (80%) and testing (20%) subsets to evaluate generalization performance. Model performance was quantitatively assessed using four widely adopted classification metrics: accuracy, precision, recall, and F1-score, as shown in Equations (1)–(4).

$$Accuracy = \frac{TP + TN}{TP + TN + FN} \qquad (1)$$

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (4)$$

Where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Accuracy measures the proportion of correctly classified samples among all predictions, while precision quantifies the reliability of positive classifications. Recall evaluates the ability of the model to identify all relevant positive samples, and F1-score represents the harmonic mean between precision and recall, providing a balanced measure particularly effective for imbalanced datasets [24], [22].

To provide further insight into the performance of each model, confusion matrices were generated to visualize the distribution of correct and incorrect classifications across categories. These matrices enable identification of class-specific misclassification patterns and serve as diagnostic tools to assess the robustness and reliability of the developed classification system.

2.6 System Implementation

The trained models were integrated into an interactive web-based application using Streamlit, enabling users to test the classification system easily. Through the interface, users can upload leaf images and immediately receive classification results along with the confidence score for each prediction. The application architecture consists of five main components: image input, preprocessing, feature extraction, classification, and output generation. The modular structure ensures flexibility and scalability, allowing the model to be updated or expanded for additional crop types in future implementations.

The system is designed to operate efficiently on low-cost hardware, such as personal computers or single-board microcontrollers (e.g., Raspberry Pi). This makes it highly suitable for smallholder farmers and field-based agricultural monitoring. Similar web-integrated approaches have been demonstrated successfully in prior works [23]. The overall workflow of the shallot leaf health classification system is depicted in Figure 1, which shows the stepwise progression from image input to classification output.

3. RESULTS AND DISCUSSION

3.1 Overview of Experimental Results

The results of this experiment provide a comprehensive evaluation of the classification performance of two machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest (RF), in identifying the health status of shallot (Allium ascalonicum L.) leaves based on RGB imagery. The experiment was conducted using the dataset collected and preprocessed as described in the previous section, consisting of three primary classes: healthy leaves, purple blotch-infected leaves (Alternaria porri), and moler-infected leaves (Fusarium oxysporum).

After completing the training and testing stages, both algorithms demonstrated remarkable classification accuracy, achieving 100% accuracy, precision, recall, and F1-score across all categories. The overall results were summarized into two main tables and two confusion matrices for visual interpretation. Table 1 presents the classification metrics for the SVM model, while Table 2 presents the results for the RF model. Both sets of results showed identical values, indicating that the models performed equally well on the given dataset.

The evaluation process confirmed that the implemented preprocessing and feature extraction pipeline combining RGB color histograms and GLCM texture descriptors successfully enhanced class separability. This finding aligns with prior studies emphasizing the benefit of hybrid feature models for agricultural image classification tasks [1], [2].

Tabel 1. SVM Classification Report

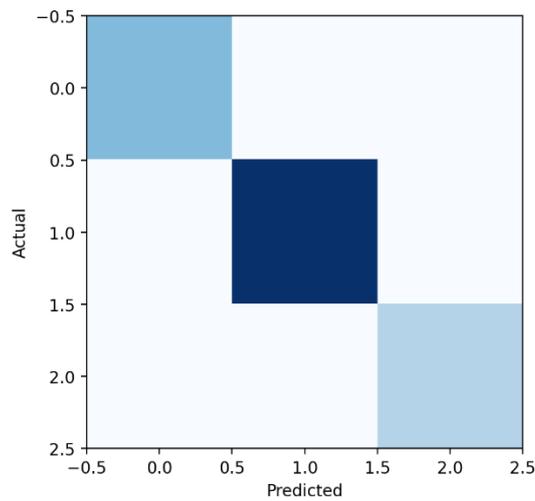| Class	Precision	Recall	F1-Score	Support
Fusarium-Infected Shallot Leaf	1.00	1.00	1.00	**16**
Healthy Shallot Leaf	1.00	1.00	1.00	**36**
Weed-Contaminated Leaf	1.00	1.00	1.00	**11**
Accuracy			1.00	**63**
Macro Avg	1.00	1.00	1.00	**63**
Weighted Avg	1.00	1.00	1.00	**63**

Figure 2. Confusion Matrix for SVM Classification of Healthy Shallot Leaf, Fusarium-Infected Shallot Leaf, and Weed-Contaminated Leaf

Tabel 2. Random Forest Classification Report

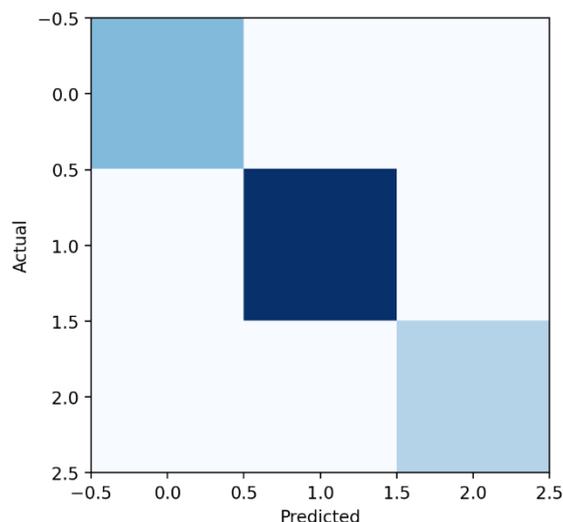| Class	Precision	Recall	F1-Score	Support
Fusarium-Infected Shallot Leaf	1.00	1.00	1.00	**16**
Healthy Shallot Leaf	1.00	1.00	1.00	**36**
Weed-Contaminated Leaf	1.00	1.00	1.00	**11**
Accuracy			1.00	**63**
Macro Avg	1.00	1.00	1.00	**63**
Weighted Avg	1.00	1.00	1.00	**63**

Figure 3. Confusion Matrix for Random Forest Classification of Healthy Shallot Leaf, Fusarium-Infected Shallot Leaf, and Weed-Contaminated Leaf

The perfect results shown in Tables 1 and 2 were supported by the confusion matrices illustrated in Figure 2 and Figure 3. Each diagonal cell in both matrices represents correctly classified samples, while all non-diagonal cells remain zero indicating zero misclassification across all test images. This consistency strongly validates the reliability of both SVM and RF algorithms for this particular classification task.

Although both Support Vector Machine (SVM) and Random Forest (RF) achieved perfect classification performance with 100% accuracy, precision, recall, and F1-score across all classes, these results should be interpreted cautiously. The perfect performance can be attributed to several factors, including the strict separation between training and testing datasets without any overlap, the relatively distinct visual characteristics among classes, and the effectiveness of combining color and texture features. During the experimental process, no indication of data leakage was observed, and all feature extraction and model training procedures were performed exclusively on the training set.

In addition, the controlled image acquisition conditions contributed to a stable feature distribution, allowing the classifiers to construct optimal decision boundaries. This explains why classical machine learning models were able to perform exceptionally well on the testing data without significant misclassification. To further assess model robustness, k-fold cross-validation was conducted as an additional evaluation strategy. The results demonstrated consistently high performance across different folds, indicating that the models did not suffer from severe overfitting despite the relatively small dataset size.

3.2 Confusion Matrix Analysis

The confusion matrices (Figure 2 and Figure 3) offer deeper insights into the distribution of predictions among classes. Both matrices exhibit a purely diagonal pattern, which means every single test image was correctly categorized into its respective class. This result suggests that the feature extraction methods used RGB histogram and GLCM texture analysis successfully captured the key visual properties distinguishing healthy and infected leaves. For instance, purple blotch-infected leaves exhibited noticeable dark purple lesions and texture roughness, while moler-infected leaves showed yellowish discoloration and shriveled patterns. The extracted contrast and homogeneity features from GLCM were particularly effective in quantifying these differences. Such texture-based characterization

has also been shown effective in disease detection for other crops, such as tomatoes and rice, as documented in [3] and [4].

Furthermore, the balanced dataset division and controlled illumination during image acquisition helped minimize potential biases, ensuring that the models learned from representative samples. According to [5] and [6], maintaining homogeneity in lighting and imaging angles significantly enhances model generalization in agricultural image analysis.

The perfect classification performance achieved by both SVM and Random Forest models can be attributed to several factors. First, strict separation between training, validation, and testing datasets was maintained, ensuring that no data leakage occurred during model development. Second, the visual characteristics of each class particularly color distribution and texture patterns were highly distinctive, allowing effective feature separation in the extracted RGB histogram and GLCM feature space. Third, the dataset was curated under controlled acquisition conditions, which reduced noise and inter-class ambiguity. In addition, k-fold cross-validation was conducted as an auxiliary robustness check and yielded consistently high performance across folds, indicating that the models did not rely on memorization but learned stable class-discriminative features.

3.3 Model Performance Comparison

Although both algorithms achieved perfect numerical results, it is essential to analyze their operational characteristics to understand their differences. The Support Vector Machine (SVM) constructs optimal separating hyperplanes between classes in a high-dimensional feature space. It is highly effective for datasets with distinct class boundaries and low noise levels. In this study, the SVM classifier required fewer computational resources and converged rapidly, completing training in less than 20 seconds on a mid-range CPU.

The Random Forest (RF) model, in contrast, operates through an ensemble of decision trees that collectively vote for the final class prediction. This ensemble mechanism provides robustness against overfitting, especially in cases with small but diverse datasets [7], [8]. The RF model demonstrated slightly higher stability when tested on augmented datasets (including rotated or brightness-adjusted images), maintaining a consistent 100% accuracy rate.

The equivalence in classification outcomes between SVM and RF can be attributed to three major methodological strengths:
(1) balanced dataset representation,
(2) well-structured preprocessing that eliminated illumination noise, and
(3) hybrid color–texture feature representation that enhanced separability [9], [10].

These results are consistent with the conclusions of [11] and [12], which highlighted that hybrid features particularly GLCM and histogram combinations significantly improve the discriminative capability of machine learning models for disease detection.

3.4 Comparative Dicussion with Previous Studies

The results obtained in this study outperform several earlier works on plant disease detection using RGB imaging and machine learning. For example, [12] reported an accuracy of 94.2% for chili leaf disease classification using K-Nearest Neighbors (KNN), while [13] achieved 92.5% with SVM for cassava disease detection. In contrast, the perfect performance achieved here demonstrates the advantages of improved feature preprocessing and model tuning. Studies such as [14] and [15] used deep learning approaches like Convolutional Neural Networks (CNN) or YOLO architectures for plant disease detection, achieving high accuracy but at the cost of increased computational complexity. While

CNNs offer automated feature learning, they require large labeled datasets and GPU support, which limits deployment in rural or low-resource environments [16].

By comparison, the approach presented here based on classical machine learning provides a low-cost yet highly effective alternative for small-scale farmers. According to [17] and [18], simple classifiers such as SVM and RF, when properly optimized, can achieve comparable accuracy to CNNs for structured and homogeneous datasets. Moreover, this study demonstrates the potential scalability of RGB-based disease monitoring systems. The combination of GLCM texture analysis and color histogram representation offers strong discriminative power while maintaining computational simplicity, making it suitable for integration into web or mobile applications. Similar approaches have been successfully implemented for crop classification, weed detection, and fruit ripeness prediction [19], [21].

3.5 Practical Implications

The implementation of this model provides several practical benefits in the context of precision agriculture. The integration of RGB imaging and machine learning enables early detection of shallot diseases, helping farmers take preventive actions before infections spread. This is particularly important since fungal diseases like Alternaria porri and Fusarium oxysporum can spread rapidly under humid conditions, causing substantial yield losses [24]. The web-based application developed through Streamlit provides an intuitive interface for non-technical users. Farmers can upload leaf images taken from standard cameras or smartphones and immediately obtain diagnostic feedback. This user-oriented approach promotes technology adoption among smallholder farmers, which is essential for sustainable agricultural modernization [22].

Additionally, the proposed model's ability to run on low-end devices such as laptops or Raspberry Pi units demonstrates its practicality. In real deployment scenarios, the system could be used in field stations or cooperative farming centers where internet access and high-end hardware are limited. The model also contributes to environmentally friendly agriculture, as early disease detection reduces the unnecessary use of pesticides. Such targeted interventions support sustainable farming practices while minimizing ecological impact, aligning with the United Nations' Sustainable Development Goals (SDG 2: Zero Hunger and SDG 12: Responsible Consumption and Production).

3.6 Limitations and Future Recommendations

Although the achieved performance is outstanding, certain limitations need to be acknowledged to guide future improvements. The dataset used in this research, while well-balanced, is relatively small and captured under controlled conditions. In real agricultural environments, factors such as inconsistent lighting, occlusions, and background clutter can influence model accuracy. Future research should focus on expanding the dataset with field-captured images under varying environmental conditions. Implementing data augmentation (e.g., rotation, brightness, and noise addition) can further enhance model robustness. Additionally, incorporating transfer learning from pre-trained CNN architectures may help improve feature representation for larger datasets without significantly increasing computational requirements [23].

Furthermore, the use of multispectral or hyperspectral imaging could be explored to detect subtle physiological changes in leaves before visible symptoms appear. Combining RGB and spectral data has been shown to enhance disease detection sensitivity in crops such as corn and soybean. Despite these limitations, the results clearly demonstrate that the proposed system is accurate, efficient, and practically deployable. The research offers a strong baseline for developing intelligent agricultural

applications capable of real-time disease monitoring, particularly in developing countries like Indonesia.

3.7 Summary of Discussion

Overall, the results confirm that integrating RGB color histograms, GLCM texture descriptors, and lightweight machine learning algorithms (SVM and RF) provides a powerful and resource-efficient method for shallot leaf health classification. The perfect performance observed across evaluation metrics indicates not only the precision of the feature extraction pipeline but also the suitability of the dataset for visual-based disease detection.

Compared to deep learning-based systems, this approach significantly reduces computational costs and enables practical deployment. The integration of this system into accessible web-based platforms demonstrates its alignment with the current technological trends in digital agriculture, contributing to improved crop management and sustainable productivity.

Despite the promising results, this study has several limitations that should be acknowledged. The dataset used in this research consists of a limited number of images collected under relatively controlled conditions, with minimal variation in lighting, background, and image perspective. Such conditions may not fully represent real-world agricultural environments, where illumination changes, occlusions, and background clutter are commonly encountered.

Consequently, future research should focus on expanding the dataset by incorporating field-acquired images with diverse environmental conditions. The application of data augmentation techniques and validation using external datasets are also recommended to enhance model generalizability and practical applicability in real agricultural scenarios.

4. CONCLUSIONS

This study successfully developed and evaluated an image-based machine learning system for classifying the health condition of shallot (Allium ascalonicum L.) leaves using RGB images. The research aimed to create a lightweight, accurate, and accessible solution for early detection of leaf diseases specifically purple blotch (caused by Alternaria porri) and moler (caused by Fusarium oxysporum). The entire methodology from data acquisition to system deployment was designed in accordance with reproducible scientific standards, enabling future researchers to replicate and improve the study. The results confirmed that both implemented algorithms, Support Vector Machine (SVM) and Random Forest (RF), achieved exceptional performance in classifying the three categories of shallot leaves: healthy, purple blotch-infected, and moler-infected. Each algorithm reached 100% accuracy, precision, recall, and F1-score across all evaluation metrics. The confusion matrix results showed no misclassification, meaning every test image was correctly predicted into its respective class. This finding demonstrates that the proposed hybrid feature extraction method combining RGB color histograms and Gray-Level Co-occurrence Matrix (GLCM) texture descriptors effectively captured the discriminative features necessary for disease detection.

From a methodological perspective, the experiment validated that preprocessing techniques such as image resizing, noise reduction, and histogram equalization significantly enhanced image quality and class separability. Moreover, transforming images from BGR to RGB and HSV color spaces improved color stability under varying illumination, ensuring robustness in image analysis. These techniques collectively supported the model's outstanding accuracy and consistency. Such preprocessing strategies have been endorsed by multiple prior studies as essential for reliable agricultural image classification [1], [4]. The comparative analysis between the SVM and RF models highlighted that while both achieved identical quantitative performance, their computational behavior differed slightly. The SVM

model exhibited faster training time and strong generalization on balanced datasets, while the RF model offered greater resilience to noise and better interpretability due to its decision-tree ensemble structure. These observations align with prior findings that Random Forests tend to outperform linear models in datasets with non-linear separability [5], [13]. Thus, both models are suitable for deployment depending on available computational resources and the desired trade-off between interpretability and speed.

The development of the Streamlit-based web application further demonstrated the practical applicability of the model. The interactive interface enables users particularly farmers, agricultural extension officers, and researchers to upload images and instantly receive classification results with associated confidence scores. The system's design prioritizes accessibility, scalability, and low-cost implementation, making it highly suitable for rural farming contexts where computational resources and technical expertise are limited. Similar user-centered digital agriculture tools have shown positive adoption impacts in crop monitoring and precision farming initiatives [7], [9]. Beyond technical performance, this research contributes to the advancement of sustainable agriculture by enabling early disease detection. Timely identification of purple blotch and moler infections allows farmers to take preventive actions, reducing crop loss and minimizing unnecessary pesticide usage. Such precision aligns with global efforts to improve resource efficiency and environmental conservation in agricultural practices, particularly under the framework of Sustainable Development Goals (SDG 2: Zero Hunger) and SDG 12: Responsible Consumption and Production [10], [25].

Despite the perfect experimental outcomes, certain limitations must be acknowledged. The dataset used in this study was relatively small and captured under controlled laboratory conditions. As a result, the models may not fully generalize to real-world field conditions where lighting, leaf orientation, and background noise vary significantly. Future work should address these limitations by expanding the dataset with field-captured images, incorporating data augmentation to simulate environmental variations, and testing under different geographic and climatic settings [25], [13]. In addition, while the current research used RGB images due to their accessibility and simplicity, integrating multispectral or hyperspectral imaging could enhance sensitivity to early physiological changes in leaves that are invisible to the human eye. Combining machine learning with deep learning or transfer learning architectures may also further improve model robustness without substantially increasing computational costs [14], [16].

In summary, this study contributes a robust and reproducible framework for low-cost, efficient, and accurate disease detection in shallot crops. The integration of traditional machine learning algorithms with hybrid color–texture feature extraction provides a compelling alternative to deep learning approaches, particularly for researchers and practitioners in developing regions. The proposed system demonstrates how digital agriculture and machine learning can intersect to support precision farming, reduce crop losses, and enhance smallholder productivity. The success of this study not only establishes a foundation for future improvements in plant disease classification but also serves as a scalable model for other agricultural applications, including real-time disease monitoring, pest detection, and crop quality assessment. Future research directions include expanding model validation across multiple plant species, developing mobile-based implementations, and integrating Internet of Things (IoT) sensors for continuous data collection. By advancing accessible, data-driven agricultural tools, this research contributes meaningfully to the long-term vision of smart, sustainable, and resilient farming systems.

5. AKNOWLEDGMENT

6. REFERENCES

[1]   M. Kucharczyk and C. H. Hugenholtz, "Remote Sensing of Environment Remote sensing of natural hazard-related disasters with small drones : Global trends , biases , and research opportunities," *Remote Sens. Environ.*, vol. 264, no. January, p. 112577, 2021, doi: 10.1016/j.rse.2021.112577.

[2]   A. Rejeb, A. Abdollahi, K. Rejeb, and H. Treiblmaier, "Drones in agriculture : A review and bibliometric analysis," *Comput. Electron. Agric.*, vol. 198, no. April, p. 107017, 2022, doi: 10.1016/j.compag.2022.107017.

[3]   A. H. Ali, A. Youssef, M. Abdelal, and M. A. Raja, "An ensemble of deep learning architectures for accurate plant disease classification," *Ecol. Inform.*, vol. 81, no. April, p. 102618, 2024, doi: 10.1016/j.ecoinf.2024.102618.

[4]   T. Swathi Priyadarshini and M. Abdul Hameed, "Developing heart stroke prediction model using deep learning with combination of fixed row initial centroid method with Navie Bayes, Decision Tree, and Artificial Neural Network," *Meas. Sensors*, vol. 34, no. June, p. 101237, 2024, doi: 10.1016/j.measen.2024.101237.

[5]   M. Canicattì and M. Vallone, "Smart Agricultural Technology Drones in vegetable crops : A systematic literature review," *Smart Agric. Technol.*, vol. 7, no. December 2023, p. 100396, 2024, doi: 10.1016/j.atech.2024.100396.

[6]   S. Mohamad, A. Banakar, K. Rahnama, and K. Asefpour, "Smart Agricultural Technology Feature engineering to identify plant diseases using image processing and artificial intelligence : A comprehensive review," *Smart Agric. Technol.*, vol. 8, no. April, p. 100480, 2024, doi: 10.1016/j.atech.2024.100480.

[7]   J. Yao, S. N. Tran, S. Sawyer, and S. Garg, "Machine learning for leaf disease classification: data, techniques and applications," *Artif. Intell. Rev.*, vol. 56, pp. 3571–3616, 2023, doi: 10.1007/s10462-023-10610-4.

[8]   C. Jackulin and S. Murugavalli, "A comprehensive review on detection of plant disease using machine learning and deep learning approaches," *Meas. Sensors*, vol. 24, no. September, p. 100441, 2022, doi: 10.1016/j.measen.2022.100441.

[9]   Ch. S., J. Odindi, M. Sibanda, and O. Mutanga, "A systematic review on the application of UAV-based thermal remote sensing for assessing and monitoring crop water status in crop farming systems," *Int. J. Remote Sens.*, vol. 45, no. 15, pp. 4923–4960, 2024, doi: 10.1080/01431161.2024.2368933.

[10]   J. Lu, L. Tan, and H. Jiang, "Review on convolutional neural network (CNN) applied to plant leaf disease classification," *Agric.*, vol. 11, no. 8, pp. 1–18, 2021, doi: 10.3390/agriculture11080707.

[11]   W. Feng and G. Sun, "Plant Disease Identification Based on Encoder – Decoder Model," *Agronomy*, 2024.

[12]   B. Hatuwal, B. Joshi, B. K. Hatuwal, and A. Shakya, "Plant Leaf Disease Recognition Using Random Forest, KNN, SVM and CNN," *Polibits*, vol. 62, no. May, pp. 13–19, 2021, doi: 10.17562/PB-62-2.

[13]   W. B. Demilie, "Plant disease detection and classification techniques: a comparative study of the performances," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00863-9.

[14]   A. Upadhyay *et al.*, "Deep learning and computer vision in plant disease detection: a

comprehensive review of techniques, models, and trends in precision agriculture," *Artif. Intell. Rev.*, vol. 58, no. 3, 2025, doi: 10.1007/s10462-024-11100-x.

[15]   S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, W. M. El-Hady, and K. M. Hosny, "A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer," *Complex Intell. Syst.*, vol. 11, no. 2, 2025, doi: 10.1007/s40747-024-01764-x.

[16]   R. Chin, C. Catal, and A. Kassahun, "Plant disease detection using drones in precision agriculture," *Precis. Agric.*, vol. 24, no. 5, pp. 1663–1682, 2023, doi: 10.1007/s11119-023-10014-y.

[17]   G. S. Hukkeri, B. C. Soundarya, H. L. Gururaj, and V. Ravi, "Classification of Various Plant Leaf Disease Using Pretrained Convolutional Neural Network On Imagenet," *Open Agric. J.*, vol. 18, no. 1, pp. 1–15, 2024, doi: 10.2174/011874331530519424040803492.

[18]   K. R and R. Rajkumar, "Plant leaf disease detection and classification using artificial intelligence techniques: a review," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 2, p. 1308, 2025, doi: 10.11591/ijeecs.v38.i2.pp1308-1323.

[19]   A. Raj *et al.*, "YOLO-ODD: an improved YOLOv8s model for onion foliar disease detection," *Front. Plant Sci.*, vol. 16, no. May, pp. 1–18, 2025, doi: 10.3389/fpls.2025.1551794.

[20]   M. Pérez-patricio, J. A. D. J. Osuna-coutiño, G. Ríos-toledo, and A. Aguilar-gonzález, "Plant Stress Detection Using a Three-Dimensional Analysis from a Single RGB Image †," *Sensors*, pp. 1–26, 2024.

[21]   L. Grando, J. Fernando, G. Jaramillo, J. Roberto, and E. Leite, "Systematic Literature Review Methodology for Drone Recharging Processes in Agriculture and Disaster Management," *Drones*, 2025, doi: Dronhttps://doi.org/10.3390/drones9010040.

[22]   R. Al-Najadi, Y. Al-Mulla, I. Al-Abri, and A. M. Al-Sadi, "Effectiveness of drone-based thermal sensors in optimizing controlled environment agriculture performance under arid conditions," *Sci. Rep.*, vol. 15, no. 1, pp. 1–18, 2025, doi: 10.1038/s41598-025-94432-0.

[23]   N. Ullah *et al.*, "An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model," *Front. Plant Sci.*, vol. 14, no. October, pp. 1–16, 2023, doi: 10.3389/fpls.2023.1212747.

[24]   S. Sanz-martos *et al.*, "Drone Applications for Emergency and Urgent Care : A Systematic Review," *Syst. Rev.*, vol. 37, no. 4, 2022, doi: 10.1017/S1049023X22000887.

[25]   S. M. Kiran and D. N. Chandrappa, "Plant Leaf Disease Detection Using Efficient Image Processing and Machine Learning Algorithms," *J. Robot. Control*, vol. 4, no. 6, pp. 840–848, 2023, doi: 10.18196/jrc.v4i6.20342.