KHATULISTIWA SMART : JOURNAL OF ARTIFICIAL INTELLIGENCE
e-ISSN : 3089-9710
p-ISSN : XXXX-XXXX

Khatulistiwa : Journal of Artificial Intelligence
Vol. 3 No. 1 (2026) : 43 - 55
Available online at : https://ejournal.narasikhatulistiwa.org/index.php/kjarti

Shallot Leaf Disease Detection Using Convolutional Neural Network and Support Vector Machine

Rania Majdoubi [1], Beny Yusman[2]
[1] Ibn Tofail University, Morocco, [2] Universitas Hafshawaty Zainul Hasan, Indonesia
majdoubi.rania@uit.ac.ma [1], beny.univer@gmail.com [2]

Doi

Submitted: 23 May 2026	Revision: 22 Jun 2026	Accepted: 25 Jun 2026	Published: 30 Jun 2026

Abstract Shallot is one of the important horticultural commodities in Indonesia with high economic value and increasing market demand. However, shallot productivity often decreases due to plant diseases, especially purple blotch and moler disease. Manual disease identification through visual observation still has limitations because it depends on farmers' experience and may lead to diagnostic errors. This study aims to compare the performance of Convolutional Neural Network and Support Vector Machine methods in detecting shallot leaf diseases based on leaf images and to implement the best-performing model into a Flask-based web application. The dataset used in this study consists of 520 shallot leaf images divided into two classes: purple blotch and moler. The data were collected from primary sources through direct image acquisition in agricultural fields and secondary sources from Kaggle and Roboflow. The research stages include literature study, data collection, preprocessing, image resizing, data annotation, dataset splitting, Convolutional Neural Network model training, Support Vector Machine training, model evaluation, and Flask web implementation. The results show that the Convolutional Neural Network model achieved a training accuracy of 98% and a validation accuracy of 98%, while the Support Vector Machine model achieved an accuracy of 95%. These results indicate that both methods can effectively detect shallot leaf diseases. The Flask web implementation can also assist users in detecting diseases more practically and quickly.

Keywords *Convolutional Neural Network; Disease Detection; Leaf Image; Shallot; Support Vector Machine*

1. INTRODUCTION

Indonesia is an agricultural country where a large proportion of the population works in the agricultural and plantation sectors. Therefore, agriculture plays an important role in supporting the national economy. One horticultural commodity with high economic value is shallot. Shallot is a leading vegetable commodity widely needed by the community, and its demand continues to increase over time. This commodity has also become a government concern because of its important contribution to food needs and the community economy (Djamaluddin et al., 2022). Shallot is a seasonal plant that grows in clusters. It has a pseudo-stem, short fibrous roots, and a shallow root system, making it vulnerable to drought conditions. In addition to having high economic value, shallot also has promising development prospects because market demand continues to increase. However, the increasing demand for shallot is not always balanced by stable productivity. One of the main causes of decreased shallot productivity is pest and disease attacks, which can lead to crop failure and reduced harvest quality (Djamaluddin et al., 2022).

Diseases in shallot plants can be caused by pathogenic fungi, bacteria, and environmental factors. Diseases caused by fungi remain a major problem because their spread can occur rapidly through soil and air. Several diseases that commonly attack shallot plants are purple blotch and moler disease. Purple blotch is generally characterized by the appearance of spots on the leaves, which may expand and cause the leaves to turn yellow or dry. This disease is often associated with fungal infection on plant leaves. Meanwhile, moler disease is caused by the fungus Fusarium oxysporum f. sp. cepae, which attacks the roots and bulbs, with symptoms such as twisted leaves, drooping leaves, yellowing, and premature plant death (Sari & Inayah, 2020; Nurcahyanti & Sholeh, 2023).

Identification of shallot plant diseases is generally still carried out visually by farmers. However, this method has limitations because symptoms of several diseases may appear similar, especially on the leaves. This condition can cause subjectivity and inaccuracy in determining the type of disease. Misidentification can also result in the use of inappropriate pesticides or control methods, making disease management less effective and potentially causing new problems. Therefore, a technology-based approach is needed to assist the process of detecting shallot plant diseases more quickly, objectively, and accurately (Darwis, 2023; Zalvadila, 2023).

The development of artificial intelligence provides great opportunities in agriculture, especially in plant disease detection based on leaf images. Convolutional Neural Network (CNN) is widely used in image processing because it can automatically extract visual patterns through convolutional layers. CNN is effective in recognizing objects, textures, and patterns in images, making it useful for plant disease classification. In addition to CNN, Support Vector Machine (SVM) is also widely used in classification because it can separate data based on the best hyperplane in the feature space. SVM can be used independently or combined with features extracted from CNN to improve classification performance (Felix et al., 2019; Sulistiyana & Anardani, 2023).

Several previous studies have applied machine learning and deep learning methods for plant disease detection. Sulistiyana and Anardani (2023) developed a corn plant disease detection application using CNN and SVM, achieving 98% accuracy for CNN and 87% accuracy for SVM. Zalvadila (2023) used SVM and CNN methods to classify shallot plant diseases. In addition, Darwis (2023) applied K-Nearest Neighbor and CNN for shallot plant disease classification. These studies show that image-based methods have significant potential to support automatic plant disease detection.

Nevertheless, comparative research on CNN and SVM for detecting purple blotch and moler disease in shallot plants still needs further development, especially through the implementation of an easy-to-use web application for farmers. Therefore, this study aims to compare the performance of CNN and SVM methods in detecting shallot leaf diseases based on leaf images and to implement the detection results into a Flask-based web application. This study is expected to help farmers recognize early symptoms of shallot diseases so that control measures can be carried out more accurately, quickly, and effectively.

2. RESEARCH METHODS

2.1. Research Design

This study uses an experimental method with machine learning and deep learning approaches to detect shallot leaf diseases. The methods compared in this study are Convolutional Neural Network (CNN) and Support Vector Machine (SVM). CNN is used to recognize visual patterns in shallot leaf images, while SVM is used as a classification method based on extracted features. The research stages include literature study, data collection, data preprocessing, CNN implementation, SVM implementation, model evaluation, and Flask-based web application implementation. The research framework is shown in Figure 1.

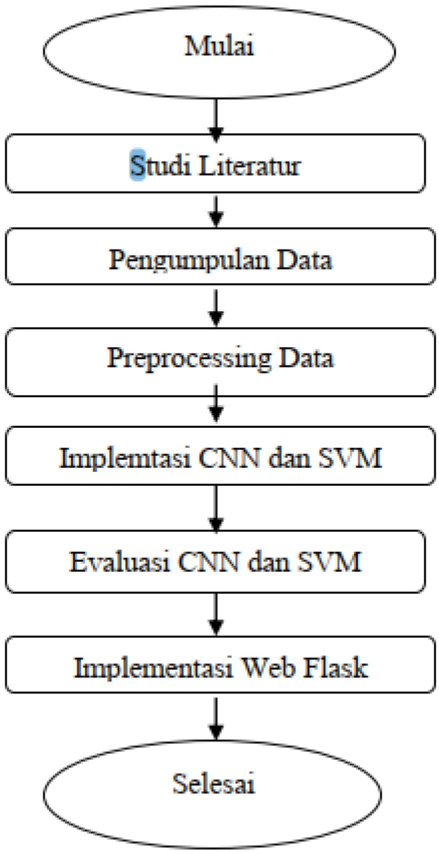

Figure 1. Model Training Framework

The initial stage of the study was conducted by identifying the problem, namely the need for a shallot leaf disease detection system. The next stage was a literature study to obtain theoretical foundations and references from previous studies. After that, the shallot leaf image dataset was collected, followed by preprocessing, CNN and SVM model training, model performance evaluation, and model implementation into a Flask web application.

2.2. Literature Study

The literature study was conducted by reviewing various scientific sources, such as journals, articles, research reports, and official sources relevant to shallot disease detection. The literature used in this study is related to shallot diseases, purple blotch, moler disease, CNN, SVM, image processing, and web-based application implementation. Previous studies served as the basis for determining the methods used in this research. Several studies used as references include Sulistiyana and Anardani's (2023) study on corn plant disease detection using CNN and SVM, Sari and Inayah's (2020) study on disease inventory in shallot varieties, and Nurcahyanti and Sholeh's (2023) study on the development of moler disease in shallot production centers.

2.3. Dataset

The dataset used in this study consists of shallot leaf images divided into two classes, namely purple blotch and moler. The data were obtained from two sources: primary data and secondary data. Primary data are shallot leaf images taken directly from agricultural fields using a smartphone camera or digital camera. Secondary data were obtained from open sources, namely Kaggle and Roboflow, using the keyword "shallot leaf disease." The total dataset consists of 520 images. The dataset includes 262 images of purple blotch and 258 images of moler disease. The details of the dataset sources are shown in Table 1.

Table 1. Distribution of Shallot Leaf Disease Dataset

No	Class	Data Source	Number of Data
1	Purple Blotch	Kaggle	104
2	Purple Blotch	Primary Agricultural Data	158
3	Moler	Kaggle	104
4	Moler	Roboflow	154
Total			**520**

The purple blotch dataset consists of 262 images, while the moler dataset consists of 258 images. The difference in the number of images occurred because collecting moler images was more difficult than collecting purple blotch images. However, the dataset distribution is still relatively balanced for model training and evaluation.

2.4. Data Preprocessing

Preprocessing was carried out to prepare raw data so that it could be used in the model training process. The preprocessing stages include image resizing, data annotation, and dataset splitting. First, image resizing was performed to standardize the image size. Image size standardization is necessary so that all images have uniform dimensions as model input. The initial image size obtained from the annotation process was 250 pixels, which was then resized to 170 pixels. This size was selected because it can still preserve the visual details of diseases on shallot leaves while remaining efficient for computation.

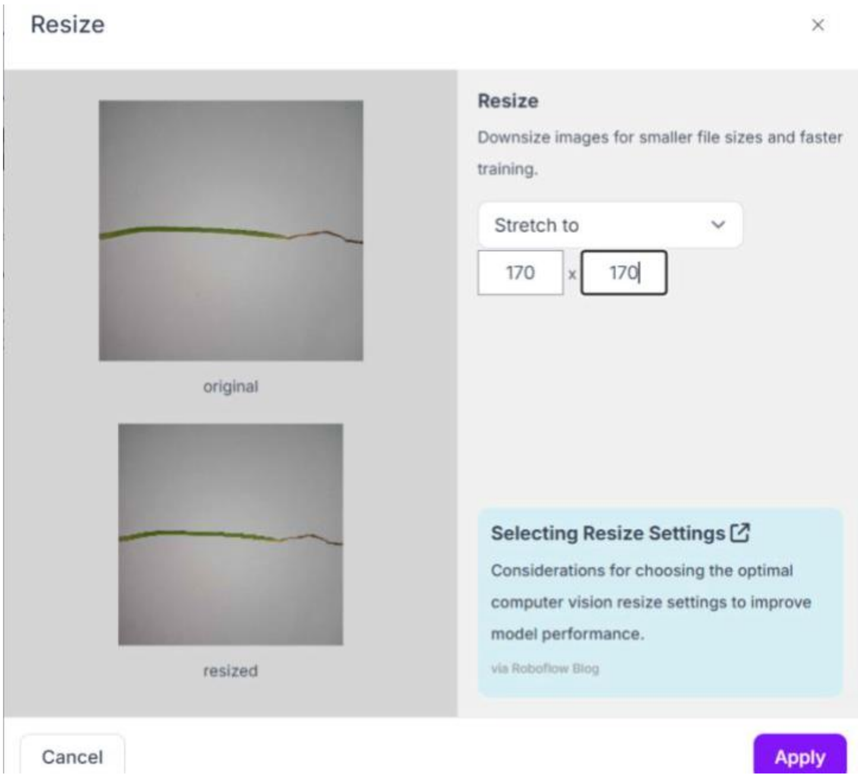

Figure 2. Image Resizing Process

Second, data annotation was performed to mark disease objects in leaf images. The annotation process was carried out using Roboflow. At this stage, each image was labeled according to its disease class, namely purple blotch or moler. Annotation is needed so that the model can better recognize the visual patterns of each class.

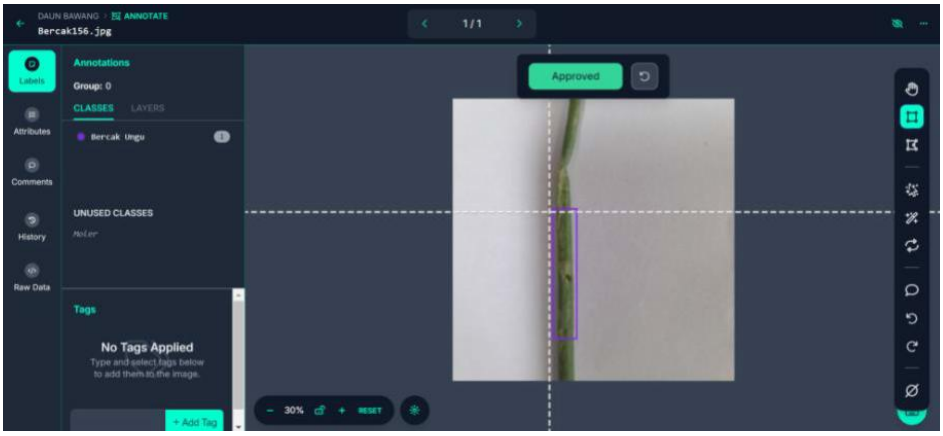

Figure 3. Data Annotation Process Using Roboflow

Third, the dataset was split into training data and validation data with an 80:20 ratio. Training data were used to train the model to recognize disease patterns, while validation data were used to evaluate the model's performance on data not used during the training process.

Table 1. **Distribution of Shallot Leaf Disease Dataset**

Class	Training Data	Validation Data	Total
Purple Blotch	209	53	262
Moler	206	52	258
Total	**415**	**105**	**520**

2.5. Convolutional Neural Network Implementation

Convolutional Neural Network was used to recognize visual patterns and features in shallot leaf images. CNN works through several layers, such as convolutional layers, pooling layers, and dense layers. The convolutional layers extract important features from images, such as edges, colors, textures, and disease spot patterns on leaves. Pooling layers reduce feature dimensions, while dense layers perform the final classification. The CNN model in this study was built using Python on Google Colaboratory with the TensorFlow framework. The model was designed for binary classification, namely distinguishing between purple blotch and moler disease. The model was compiled using the Adam optimizer and binary cross-entropy loss function. The model summary showed that the CNN architecture had a total of 19,034,177 trainable parameters.

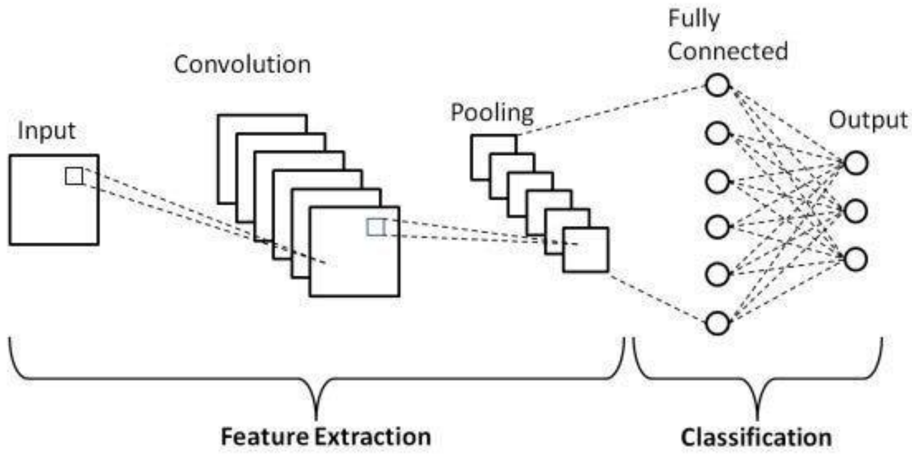

Figure 4. CNN Model Architecture

The CNN model was trained using training data for 20 epochs. The training results were then saved in .h5 format so that they could be reused in the Flask web implementation stage.

2.6. Support Vector Machine Implementation

Support Vector Machine was used as a classification model to distinguish between purple blotch and moler disease. In this study, SVM was trained using features extracted from the CNN model. These features are numerical representations of leaf images used as input for the SVM model. The SVM model was built using the Scikit-learn library with a linear kernel. The linear kernel was used because it can separate data based on a hyperplane in the feature space. After the model was trained using training data, it was evaluated using validation data. The SVM model was then saved in .pkl format using the joblib library so that it could be reused without retraining.

2.7. Model Evaluation

Model evaluation was conducted to measure the performance of CNN and SVM in detecting shallot leaf diseases. The evaluation metrics used include accuracy, precision, recall, F1-score, and confusion matrix. These metrics were used to determine how well the model performed classification. The evaluation formulas used are as follows:

$$\text{Accuracy} = (TP + TN) / \text{Total Data}$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

The confusion matrix was used to display the number of correctly and incorrectly classified data in each class. Through the confusion matrix, prediction errors in each class can be analyzed in more detail.

2.8. Flask Web Implementation

After the CNN and SVM models were trained and evaluated, the next stage was implementing the model into a Flask-based web application. This web application functions to receive shallot leaf image input from users, process the image, and display the disease prediction result. Users can upload an image through the web page, and the system will display whether the image belongs to the purple blotch class or the moler class.

3. RESULTS

3.1. Data Collection Results

The dataset used in this study consists of 520 shallot leaf images. The data consist of 262 purple blotch images and 258 moler images. The images were obtained from primary and secondary sources. Primary data were collected directly from agricultural fields using a camera, while secondary data were obtained from Kaggle and Roboflow.

Figure 5. Example of Moler Disease Image

Figure 6. Example of Purple Blotch Disease Image

The collected dataset was then stored in folders according to each class. Dataset grouping was conducted to facilitate preprocessing, model training, and model evaluation.

3.2. Data Preprocessing Results

In the preprocessing stage, all shallot leaf images were processed through several stages, namely image resizing, data annotation, and dataset splitting. The resizing process was conducted to standardize image size according to the model input requirements. The annotation process was performed using Roboflow to mark disease objects on the leaves. After that, the dataset was divided into training data and validation data with an 80:20 ratio. The dataset splitting results show that the purple blotch class had 209 training data and 53 validation data, while the moler class had 206 training data and 52 validation data. The total training data consisted of 415 images, and the total validation data consisted of 105 images.

3.3. CNN Model Training Results

The CNN model was trained using training data for 20 epochs. During the training process, the model learned the visual patterns of each disease class. The training results show that model accuracy increased as the number of epochs increased, while the loss value tended to decrease. This indicates that the model was able to learn from the training data.

Figure 7. CNN Model Accuracy and Loss Graph

The final CNN training results are shown in Table 3.

Table 3. Table 3. CNN Model Training Results

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
20	0.0374	0.9867	0.0447	0.9810

Based on Table 3, the CNN model achieved a training accuracy of 0.9867 or 98%, while the validation accuracy was 0.9810 or 98%. The training loss of 0.0374 and validation loss of 0.0447 indicate that the model had a low error rate on both training and validation data.

3.4. SVM Model Training and Evaluation Results

The SVM model was trained using features extracted from the CNN model. After the training process was completed, the model was tested using validation data. The evaluation results show that the SVM model achieved an accuracy of 95%.

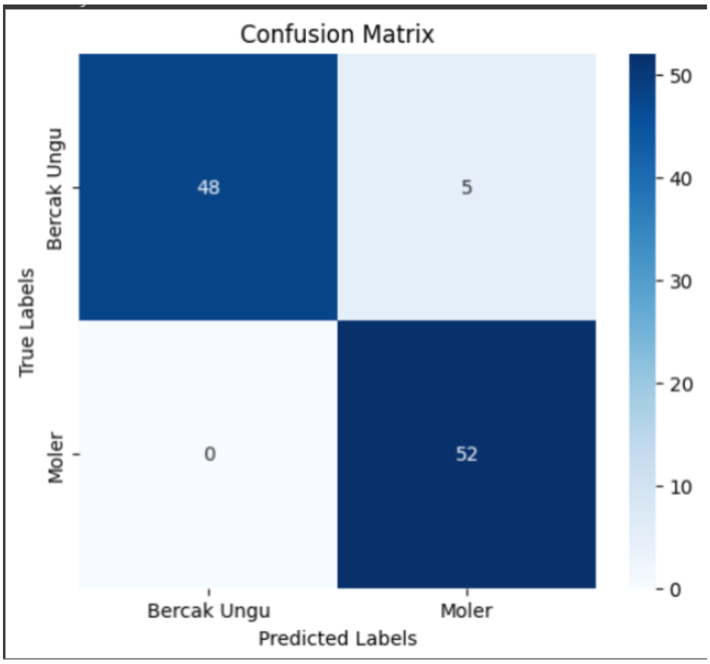

Figure 8. SVM Model Confusion Matrix

Table 4. SVM Model Confusion Matrix

Class	TP	FP	FN	Total
Purple Blotch	48	0	5	53
Moler	52	5	0	57

Based on Table 4, out of 53 purple blotch data, 48 data were correctly predicted as purple blotch, while 5 data were incorrectly predicted as moler. In the moler class, 52 data were correctly predicted as moler, and no moler data were incorrectly predicted as purple blotch. Out of 105 validation data, 100 data were correctly classified, resulting in an overall SVM model accuracy of 95%. The precision, recall, and F1-score results for each class are shown in Table 5.

Table 5. Precision, Recall, and F1-Score Results of the SVM Model

Class	Precision	Recall	F1-Score
Purple Blotch	1.00	0.906	0.951
Moler	0.912	1.00	0.954
Average	**0.956**	**0.953**	**0.953**

Based on Table 5, the purple blotch class achieved a precision of 1.00, recall of 0.906, and F1-score of 0.951. Meanwhile, the moler class achieved a precision of 0.912, recall of 1.00, and F1-score of 0.954. These results show that the SVM model had good classification performance for both shallot leaf disease classes.

3.5. Comparison of CNN and SVM Models

The comparison of CNN and SVM evaluation results is shown in Table 6.

Table 6. Accuracy Comparison of CNN and SVM

Model	Accuracy
Convolutional Neural Network	98%
Support Vector Machine	95%

Based on Table 6, CNN achieved higher accuracy than SVM. CNN achieved an accuracy of 98%, while SVM achieved an accuracy of 95%. This indicates that CNN has better capability in recognizing visual patterns in shallot leaf images. However, SVM still produced very good results, especially when using features extracted from CNN. The higher performance of CNN indicates that deep learning methods can automatically extract visual features without requiring manual feature extraction. Meanwhile, SVM has advantages in simpler and more efficient classification. Therefore, both methods can be used in shallot leaf disease detection systems, although CNN performed better based on accuracy.

3.6. Flask Web Implementation Results

The trained model was then implemented into a Flask-based web application. This application has the main function of receiving shallot leaf image input from users, processing the image, and displaying the disease prediction result. The main page of the application provides an image upload feature, and the system automatically displays the detection result.

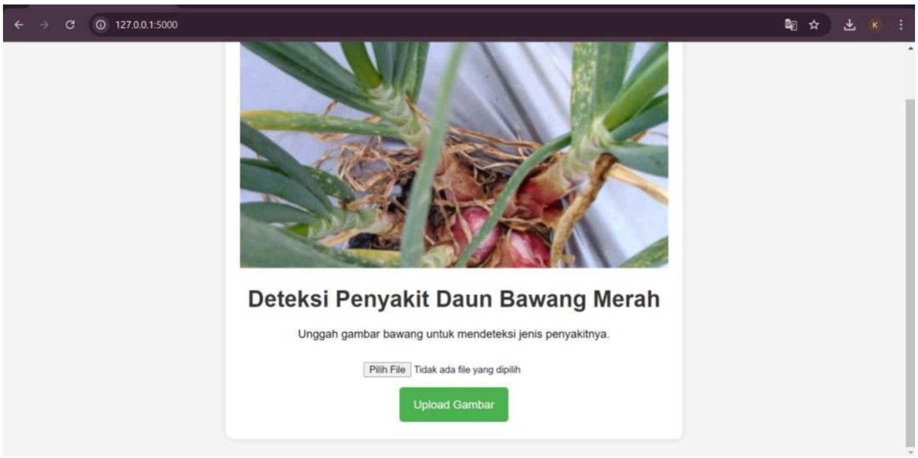

Figure 9. Shallot Leaf Disease Detection Application Interface

In application testing, users can upload shallot leaf images through the upload button. After the image is processed by the system, the application displays the disease detection result. If the image is detected as purple blotch, the system displays information that the leaf is infected with purple blotch. If the image is detected as moler, the system displays information that the leaf is infected with moler disease.

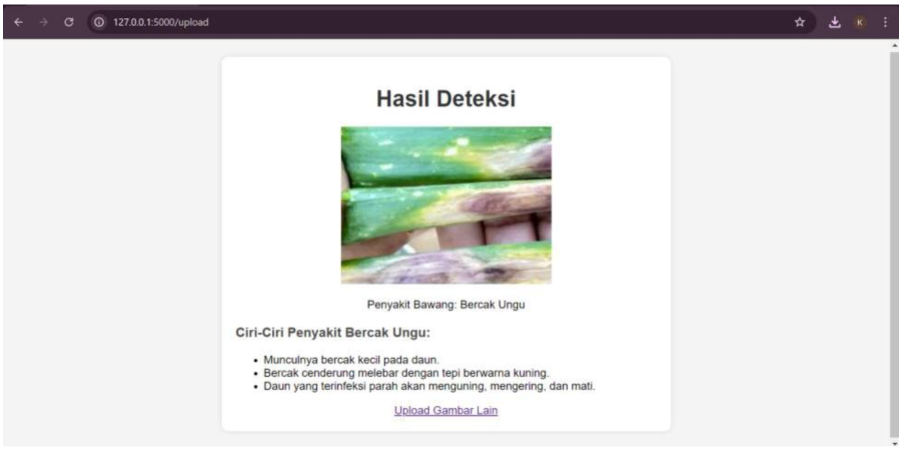

Figure 10. Purple Blotch Disease Detection Result

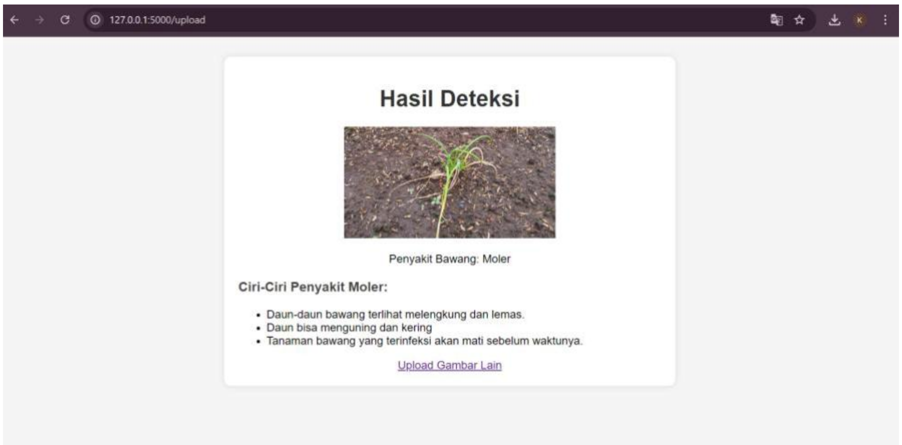

Figure 11. Moler Disease Detection Result

The implementation results show that the Flask web application can be used to assist the shallot leaf disease detection process practically. This system can serve as an initial tool for farmers to recognize diseases based on leaf images before further control actions are taken.

4. CONCLUSIONS

Based on the results of this study, it can be concluded that Convolutional Neural Network and Support Vector Machine methods can be used to detect shallot leaf diseases based on leaf images. This study focused on two disease classes, namely purple blotch and moler, with a total dataset of 520 images. The dataset was obtained from primary sources through direct image acquisition in agricultural fields and secondary sources from Kaggle and Roboflow. The testing results show that the CNN model achieved an accuracy of 98%, while the SVM model achieved an accuracy of 95%. Therefore, CNN performed better in recognizing visual disease patterns on shallot leaves. However, SVM also showed very good performance, especially when using features extracted from CNN. Based on the SVM confusion matrix, out of 105 validation data, 100 data were correctly classified and 5 data were misclassified, indicating that the model had good generalization ability. The Flask web implementation shows that the model can be applied to a web-based system to detect shallot leaf diseases more practically. This application allows users to upload shallot leaf images and obtain detection results in the form of purple blotch or moler classes. Practically, this system can help farmers recognize early symptoms of disease so that control measures can be carried out more quickly and accurately. This study still has limitations, particularly in the relatively limited dataset size and the inclusion of only two disease classes. Therefore, future research is recommended to increase the number of data, expand disease classes, apply data augmentation, and test other deep learning architectures such as MobileNet, ResNet, EfficientNet, or YOLO so that the detection system can achieve better accuracy and generalization ability.

5. REFERENCES

Darwis, H. (2023). K-Nearest Neighbor dan Convolutional Neural Network pada klasifikasi penyakit tanaman bawang merah. Techno.Com, 22(3). https://doi.org/10.33633/tc.v22i3.8533

Djamaluddin, R. R., Sukmawaty, E., Masriany, M., & Hafsan, H. (2022). Identifikasi gejala penyakit dan cendawan patogen tanaman bawang merah (Allium ascalonicum) di Kecamatan Buntu Batu Kabupaten Enrekang. Teknosains: Media Informasi Sains dan Teknologi, 16(1), 81-92. https://doi.org/10.24252/teknosains.v16i1.26027

Fakhrunnia, B. R. R., & Sesoca, J. (2023). Pemanfaatan metode Multiclass Support Vector Machine dalam klasifikasi penyakit daun kacang tanah. SMARTICS Journal, 9(2), 62-70. https://doi.org/10.21067/smartics.v9i2.9077

Febrywinata, E. (2024). Pengenalan dan klasifikasi jenis buah menggunakan metode CNN secara sederhana dengan menggunakan Google Colab. Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika, 2(4), 185-193. https://doi.org/10.61132/merkurius.v2i4.162

Felix, F., Faisal, S., Butarbutar, T. F., & Sirait, P. (2019). Implementasi CNN dan SVM untuk identifikasi penyakit tomat via daun. Jurnal SIFO Mikroskil, 20(2), 117-134. https://doi.org/10.55601/jsm.v20i2.670

Fikriah, F. K., Sulthan, M. B., Mujahidah, N., & Roziqin, M. K. (2022). Naïve Bayes untuk klasifikasi penyakit daun bawang merah berdasarkan ekstraksi fitur Gray Level Cooccurrence Matrix (GLCM). Jurnal Komtika: Komputasi dan Informatika, 6(2), 133-141. https://doi.org/10.31603/komtika.v6i2.7925

Hikmahwati, H., Auliah, M. R., Ramlah, R., & Fitrianti, F. (2020). Identifikasi cendawan penyebab penyakit moler pada tanaman bawang merah (Allium ascalonicum L.) di Kabupaten Enrekang. AGROVITAL: Jurnal Ilmu Pertanian, 5(2), 83-86. https://doi.org/10.35329/agrovital.v5i2.1745

Misdiyanto, M., Suhandini, Y., & Aprilia, I. (2020). Identifikasi jenis-jenis burung lovebird menggunakan pengolahan citra digital dengan metode K-Means Clustering. J-SAKTI: Jurnal Sains Komputer dan Informatika, 4(2), 445-456.

Negoro, N. K., Utami, E., & Yaqin, A. (2023). Klasifikasi deteksi penggunaan masker menggunakan metode Convolutional Neural Network. JIPI: Jurnal Ilmiah Penelitian dan Pembelajaran Informatika, 8(2), 664-674. https://doi.org/10.29100/jipi.v8i2.3748

Nurcahyanti, S. D., & Sholeh, M. I. (2023). Perkembangan penyakit moler (Fusarium oxysporum f. sp. cepae) pada sentra produksi bawang merah di Kabupaten Probolinggo.

Berkala Ilmiah Pertanian, 6(2), 56-62. https://doi.org/10.19184/bip.v6i2.35392

Nuryanto, N., Pamungkas, D. P., & Widodo, D. W. (2024). Penerapan metode Convolutional Neural Network pada tanaman bawang merah. In Prosiding SEMNAS INOTEK: Seminar Nasional Inovasi Teknologi (Vol. 8, No. 3, pp. 1337-1345).

Prasetyo, A. R., & Aditya, B. (2023). Analisis perbandingan algoritma Support Vector Machine (SVM) dan Convolutional Neural Network (CNN) untuk sistem deteksi katarak. Jurnal Ilmiah Teknik Mesin, Elektro dan Komputer, 3(1), 1-10. https://doi.org/10.51903/juritek.v3i1.604

Salamiah, S., Pramudi, M. I., Rosa, H. O., & Fitriyanti, D. (2022). Demplot penerapan eco farming untuk mengendalikan hama dan penyakit tanaman bawang merah di lahan gambut. ALAMTANA: Jurnal Pengabdian Masyarakat UNW Mataram, 3(2), 78-86.

Sari, W., & Inayah, S. A. (2020). Inventarisasi penyakit pada dua varietas lokal bawang merah (Allium ascalonicum L.) Bima Brebes dan Trisula. Jurnal Pro-Stek, 2(2). https://doi.org/10.35194/prs.v2i2.1166

Sulistiyana, F., & Anardani, S. (2023). Aplikasi deteksi penyakit tanaman jagung dengan metode CNN dan SVM. In Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SENATIK) (Vol. 6, No. 1, pp. 423-432).

Yana, Y. E., & Nafi'iyah, N. (2021). Klasifikasi jenis pisang berdasarkan fitur warna, tekstur, bentuk citra menggunakan SVM dan KNN. Research Journal of Computer, Information System & Technology Management, 4(1), 28. https://doi.org/10.25273/research.v4i1.6687

Zalvadila, A. (2023). Klasifikasi penyakit tanaman bawang merah menggunakan metode SVM dan CNN. Jurnal Informatika: Jurnal Pengembangan IT, 8(3), 255-260. https://doi.org/10.30591/jpit.v8i3.5341