EdgeTech Insights Journal
Vol.1, 1 (Juni, 2025), pp. 34-48
ISSN: xxxxx EISSN: xxxxx, DOI: xxxxx

IMAGE-BASED CHILI LEAF DISEASE DETECTION USING THE K-NEAREST NEIGHBORS ALGORITHM

Cristina Pangaribuan[1], Vanessa Hardini[2]
[1]Lancang Kuning University; cristinapangaribuan54@e-mail.com
[2]Lancang Kuning University ; hardinivanesa@e-mail.com

ARTICLE INFO

Keywords:

Chili leaf disease;
digital image processing;
image classification;
k-nearest neighbors;
feature extraction

Article history:

Received 2025-05-16
Revised 2025-05-24
Accepted 2025-05-28

ABSTRACT

Chili pepper is one of the important horticultural commodities that plays a significant role in food security and the national economy, particularly in Indonesia. However, chili production is often disrupted by plant diseases, especially those affecting the leaves. Diseases such as Cercospora, Murda Complex, Powdery Mildew, and nutrient deficiencies can significantly reduce both the quality and quantity of crop yields. Early detection of these diseases is crucial, as initial symptoms are often difficult for farmers to identify visually, making a fast, accurate, and objective disease detection system essential. This study aims to classify chili leaf diseases based on digital images using the K-Nearest Neighbors (KNN) algorithm by utilizing color and texture features. The dataset used was obtained from the public "Chili Leaf Disease Detection" dataset available on the Kaggle platform, consisting of 250 chili leaf images in .jpg format divided into five classes: Cercospora, Healthy, Murda Complex, Low Nutrient, and Powdery Mildew. The data were split into 70% training data and 30% testing data. The research stages include image preprocessing, which involves resizing images to 256 × 256 pixels, converting color space from RGB to HSV, leaf segmentation using the Saturation channel with a threshold of S > 0.2, and median filtering to reduce noise. Feature extraction was performed using HSV color histograms and Local Binary Pattern (LBP) to represent color and texture characteristics. Classification was carried out using the KNN algorithm with K values of 3, 5, and 7. The results show that the best classification accuracy of 85.33% was achieved at K = 5, indicating that this method is effective for chili leaf disease classification based on digital images.

Corresponding Author:
Cristina Pangaribuan
Lancang Kuning University; cristinapangaribuan54@e-mail.com

1. INTRODUCTION

As a highly important sector in many countries, including Indonesia, agriculture plays a major role in meeting the food needs of the population (Afriyanti et al., 2023). One of the most widely cultivated horticultural commodities with high economic value is chili pepper(Bangun, 2021). In addition to being a primary ingredient in various cuisines, chili pepper is also an important export commodity that contributes to the national economy(Shaker et al., 2021). However, during the cultivation process, chili plants often face various challenges, one of which is disease infestation(Iqbal et al., 2023). Diseases such as leaf spot, root rot, and wilt can threaten crop yields and reduce fruit quality(Tripathi, Maurya, Pandey, & Behera, 2024). The impact of these diseases not only shortens the lifespan of the plants but also leads to a significant decline in productivity(Mahanta, Dange, Trivedi, & Nandeha, 2024). Since the early symptoms of chili plant diseases are difficult to detect visually by farmers, fast, accurate, and efficient technology is required to enable early disease detection in order to prevent greater losses(Tiwari et al., 2025).

Advances in digital image processing technology provide significant opportunities for the automatic detection of plant diseases(Ruby et al., 2024). Various classification methods have been developed to analyze images of plant leaves, one of which is the K-Nearest Neighbors (KNN) algorithm(Jia & Liao, 2023). KNN is a distance-based classification method that determines the class of an object based on its proximity to other objects in the feature space(Toennies, 2024). In the context of agriculture, particularly chili plants, KNN can be used to classify leaf conditions as either healthy or infected with specific diseases(Patil, Patil, & Lad, 2022). By utilizing this algorithm, the disease detection process can be performed more quickly and objectively, thereby assisting farmers in making appropriate treatment decisions(Baurai et al., 2024).

This study aims to apply the K-Nearest Neighbors (KNN) algorithm to detect diseases in chili plants using digital leaf images(Pulungan, Furqan, & Rifki, 2024). Digital imaging enables the identification of visual characteristics of leaves, such as color changes, texture variations, and spot patterns, which serve as indicators of plant diseases(Harshitha, Nagaraja, & Pruthiraja, 2024). By recognizing these features, the types of diseases affecting chili plants can be identified at an early stage(Ayeshmi et al., 2025). Early detection is crucial, as it allows disease control measures to be implemented more promptly before the infection spreads widely(Kunta, Park, Braswell, da Graça, & Edwards, 2021). Furthermore, the application of an image-based detection system has the potential to reduce farmers' dependence on excessive pesticide use, thereby lowering crop maintenance costs and contributing to environmental sustainability(Mohammed, Nayyef, & Tuama, 2025).

The K-Nearest Neighbors (KNN) algorithm has been widely applied in previous studies to detect diseases in various crops, such as citrus and tomato plants, and has demonstrated reasonably good performance(Raut & Kasat, 2024). However, research that specifically examines the application of KNN to chili plants remains relatively limited. Chili plants have distinct leaf characteristics and present specific challenges in image processing(Aishwarya & Reddy, 2024). Several previous studies have also reported limitations related to inconsistent image quality, variations in lighting conditions, and limited amounts of training data(Pratama, Rasywir, Suyanti, Siswanto, & Fachruddin, 2025). Therefore, this study focuses on chili plants to address the existing research gap and to overcome common challenges encountered in image-based plant disease detection(Lokeswari, Pramesti, & Fakhrurroja, 2025).

Variations in plant conditions, environmental factors, and image quality represent major challenges in detecting diseases in chili plants(Araujo, Malemathh, & Sundaram, 2022). Differences in lighting conditions, image acquisition angles, and leaf conditions can significantly affect feature extraction results and the accuracy of classification models(Rzanny, Seeland, Wäldchen, & Mäder, 2017). To address these issues, this study seeks to optimize the K-Nearest Neighbors (KNN) algorithm by applying more effective image pre-processing techniques. Pre-processing steps such as color normalization, contrast enhancement, and noise reduction are expected to improve image quality prior to the classification process(Weng, Liu, & Guo, 2025). In addition, the use of a more representative dataset is anticipated to further enhance the performance and accuracy of the disease detection system(Pakutharivu, Sasirekha, Devaraj, & Gopi, 2023).

This study plays an important role in supporting efforts to improve chili agricultural productivity in Indonesia(Purboseno, Dharmawati, & Rahayu, 2025). By providing a fast and accurate disease detection system, farmers can take timely and appropriate control measures, thereby minimizing losses caused by disease outbreaks(Mallesh et al., 2023). Furthermore, this research is relevant to strengthening both national and global food security(Korolev, Khorev, Salikov, & Kolomyceva, 2023). The primary objective of this study is to identify diseases affecting chili leaves using the K-Nearest Neighbors (KNN) algorithm and to evaluate its level of effectiveness(Vedanty, Kesiman, Sunarya, & Indradewi, 2023). This research is expected to address gaps in the existing literature on chili plant disease detection and to contribute by implementing more effective image processing techniques to enhance the accuracy of plant disease diagnosis(Rajendrakumar, Rajashekarappa, & Parvati, 2025).

2. METHODS

The K-Nearest Neighbors (KNN) algorithm is employed in the quantitative experimental approach of this study to detect and classify diseases in chili leaves using digital images(Pulungan et al., 2024). This approach is selected because it is simple, efficient, and capable of objectively evaluating classification performance by analyzing texture and color features(Shaparia, Patel, & Shah, 2017).

This research was conducted through a systematic and structured workflow, beginning with the collection of a chili leaf image dataset, followed by image pre-processing, feature extraction, classification using the K-Nearest Neighbors (KNN) algorithm, and concluding with the evaluation of system performance(Pulungan et al., 2024).

2.1 Data Collection

The public dataset entitled “Chili Leaf Disease Detection”, obtained from the Kaggle platform, was used as the data source in this study(Rajendrakumar et al., 2025). This dataset was selected because it provides multiple categories of chili leaf diseases and offers sufficient image quality to support analysis based on digital image processing techniques(Aishwarya & Reddy, 2024).

This study begins with the use of a chili leaf image dataset as the input data(Suwarningsih et al., 2024). The acquired images are then processed during the pre-processing stage to enhance image quality and to separate the leaf objects from the background(Govindarajan & S, 2023). Feature extraction is subsequently performed using three scenarios: HSV color features, Local Binary Pattern (LBP) texture features, and a combination of HSV and LBP features(Nasir, Suciati, & Wijaya, 2017). The extracted features are then utilized in the classification stage using the K-Nearest Neighbors (KNN) algorithm(Turkoglu & Hanbay, 2019).The final stage involves performance evaluation, which aims to assess the effectiveness of the system in classifying chili leaf diseases based on the prediction results obtained(Mashuri, Sunyoto, & Kusnawi, 2024).

This dataset consists of 250 chili leaf images in .jpg format, which are categorized into the following five main classes:

(a) Cercospora (b) Healthy (c) Murda complex (d) Nutrional (e)Powdery
mildew

Each class contains 50 images, of which 35 are used for training and 15 for testing. Accordingly, the dataset is composed of 175 training images (70%) and 75 testing images (30%), resulting in a total of 250 images used in this study.

Table 1. Distribution of Training and Testing Data for Chili Leaf Disease Classes

No	CLASS	Training Data	Testing Data	Total
1	Cercospora	35	15	50
2	Healthy	35	15	50
3	Murda Complex	35	15	50
4	Nutritional Deficiency	35	15	50
5	Powdery Mildew	35	15	50
Total	-	**175 (70%)**	**75 (30%)**	**250 (100%)**

2.2 Pre-processing

The pre-processing stage aims to improve the quality of chili leaf images and to prepare them prior to the feature extraction process(Rajendrakumar et al., 2025). Pre-processing is essential for reducing noise, standardizing image size, and ensuring that the extracted features accurately represent the chili leaf objects(Aminuddin, Joret, Zulkifli, Morsin, & Tukiran, 2023).
The pre-processing steps applied in this study are as follows:

2.2.1 Image Resizing

To reduce computational load and standardize image dimensions, all chili leaf images were resized to 256 × 256 pixels. This size normalization ensures that each image can be processed consistently during the feature extraction and classification stages.

2.2.2 Color Space Conversion (RGB to HSV)

The chili leaf images, originally represented in the RGB color space, were converted to the HSV (Hue, Saturation, Value) color space using the cv2.cvtColor() function. The HSV color space was selected because it is more robust to variations in lighting conditions than the RGB color space, allowing for a more stable representation of the color information of chili leaves.

2.2.3 Leaf Segmentation

The segmentation process was performed to separate the chili leaf objects from the image background. Segmentation was carried out by utilizing the Saturation (S) channel in the HSV color space, applying a threshold of $S > 0.2$, which was empirically determined based on visual observation to ensure optimal leaf area segmentation.

The segmentation results were subsequently refined using morphological operations, namely bwareaopen to remove small-sized noise and imfill to fill holes within the leaf regions. This step aims to obtain clean and intact leaf areas so that the feature extraction process is not adversely affected.

2.2.4 Median Filtering

To reduce residual noise, the segmented chili leaf images were processed using a median filter with a kernel size of 3×3. This method effectively smooths the image while preserving important edge information on the leaf surface. Through this pre-processing stage, clean chili leaf images focused on the leaf objects were obtained, making them suitable for subsequent color and texture feature extraction.

2.3 Feature Extraction

The feature extraction stage aims to obtain important characteristics from chili leaf images that can serve as the basis for disease classification(Loti, Noor, & Chang, 2021). In this study, the extracted features include color features and texture features, as these two types of features are capable of representing visual changes in chili leaves caused by disease infection(Rahadiyan, Hartati, Wahyono, & Nugroho, 2022).

2.3.1 Color Features (HSV Statistics)

Color features were extracted from chili leaf images that had been converted to the HSV (Hue, Saturation, and Value) color space. For each HSV component, a color histogram was computed to represent the distribution of color intensities on the surface of the chili leaves. Statistical measures, namely the mean and standard deviation, were then extracted from these histograms and used as color features.Statistical HSV-based color features were employed because they are effective in distinguishing between healthy leaves and disease-infected leaves based on color variations, such as changes in greenness, yellowing, and the appearance of spots on the leaf surface.

2.3.2 Texture Features (Local Binary Pattern / LBP)

In addition to color features, texture features were also extracted using the Local Binary Pattern (LBP) method. This method operates by comparing the intensity value of a central pixel with those of its surrounding pixels, thereby generating a binary pattern that represents local texture information on the leaf surface.

LBP was selected because it is capable of capturing texture characteristics such as speckles, spots, and irregular patterns that commonly appear on disease-infected chili leaves.

2.3.3 Feature Vector Construction

The extracted color and texture features were subsequently combined into a single feature vector. This feature vector was used as the input for the classification stage employing the K-Nearest Neighbors (KNN) algorithm. The integration of color and texture features is expected to enhance the system's ability to more accurately distinguish between different classes of chili leaf diseases.

2.4 Classification

The classification process in this study was carried out using the K-Nearest Neighbors (KNN) algorithm, which is a supervised learning method that determines the class of a data sample based on its distance-based similarity to training data(Vega-Huerta et al., 2025). The KNN algorithm operates by measuring the proximity between testing data and training data within the feature space to identify the most appropriate class(Acito, 2023).

The classification stages using the KNN algorithm in this study consist of the following steps:

1. Selection of the k Value
   The value of k represents the number of nearest neighbors used in the classification process. In this study, several k values, namely k = 3, 5, and 7, were evaluated to obtain optimal classification performance.
2. Euclidean Distance Calculation
   The distance between the feature vector of a testing image and the feature vectors of the training images was calculated using the Euclidean distance formula as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

where:

- *x*_{*i*} represents the *i*-th feature value of the testing data,
- *y*_{*i*} represents the *i*-th feature value of the training data, and
- *n* denotes the number of features used.

3. Selection of Nearest Neighbors
   After the distance values were computed, the training data were sorted in ascending order based on the shortest distance. Subsequently, the k nearest neighbors with the smallest distances to the testing data were selected.
4. Image Class Determination
   The class of the testing image was determined based on the majority class among the selected k nearest neighbors. The class with the highest frequency of occurrence was assigned as the classification result.
5. KNN Implementasi
   The implementation of the KNN algorithm was carried out using the Scikit-learn library, specifically the KNeighborsClassifier class. The fit() function was used to train the model with the training data, while the predict() function was employed to predict the class labels of the testing images.

2.5 Evaluation

Evaluation was conducted to assess the performance of the K-Nearest Neighbors (KNN) algorithm in classifying chili leaf images into five disease classes. Several evaluation metrics were employed in this study, as described below:

1. Confusion Matrix
   The confusion matrix was used to display the number of correct and incorrect predictions for each class. Since the dataset consists of five classes, a 5 × 5 confusion matrix was employed.
2. Accuracy
   Accuracy was used to measure the overall classification performance, defined as the ratio between the number of correct predictions and the total number of testing samples. In multi-class classification, the accuracy value is calculated based on the total number of correctly classified samples across all classes relative to the entire testing dataset.

$$Accuracy = \frac{\sum_i^n = 1\, TP_i}{N}$$

3. Precision, Recall, dan F1-score.
   Precision indicates the accuracy of the model's predictions for a particular class, recall reflects the model's ability to correctly identify all samples belonging to that class, and the F1-score represents the harmonic mean of precision and recall.
4. Results Analysis
   The evaluation was conducted under three feature scenarios, namely the use of color features (HSV), texture features (LBP), and a combination of both (HSV + LBP). Each scenario was tested using variations of k = 3, 5, and 7, and the best performance was determined based on the highest average values of the evaluation metrics.

3. FINDINGS AND DISCUSSION

In this study, the classification of chili leaf diseases was performed by applying the K-Nearest Neighbors (KNN) algorithm, which utilizes a combination of color features in the HSV color space and texture features based on Local Binary Pattern (LBP). The image data were obtained from the public Chili Leaf Disease Detection dataset available on the Kaggle platform, which contains chili leaf images in .jpg format categorized into five disease and leaf condition classes.

The dataset was divided into two subsets, namely training data and testing data, with a proportion of 70% for training and 30% for testing. For each class, 35 images were used as training data and 15 images were used as testing data, resulting in a total of 250 images analyzed in this study.

A series of experiments was conducted in a sequential manner, starting with image pre-processing, followed by feature extraction, classification using the KNN algorithm with variations of k = 3, 5, and 7, and model performance evaluation. To obtain the optimal configuration for multi-class classification, a comparative analysis was performed across three feature schemes: HSV color features, LBP texture features, and a combination of HSV and LBP features. The results of this comparison were used to determine the approach that provides the most optimal classification performance in detecting chili leaf diseases.

1. Preprocessing

a. HSV Color Space Conversion

During the pre-processing stage, chili leaf images originally represented in the RGB color space were converted into the HSV (Hue, Saturation, Value) color space using a color space conversion function. This transformation was performed to separate color information from illumination intensity, allowing the color characteristics of the leaves to be analyzed more effectively. The Hue channel represents the type of color tone, while the Saturation channel describes the level of color intensity on the leaf surface.

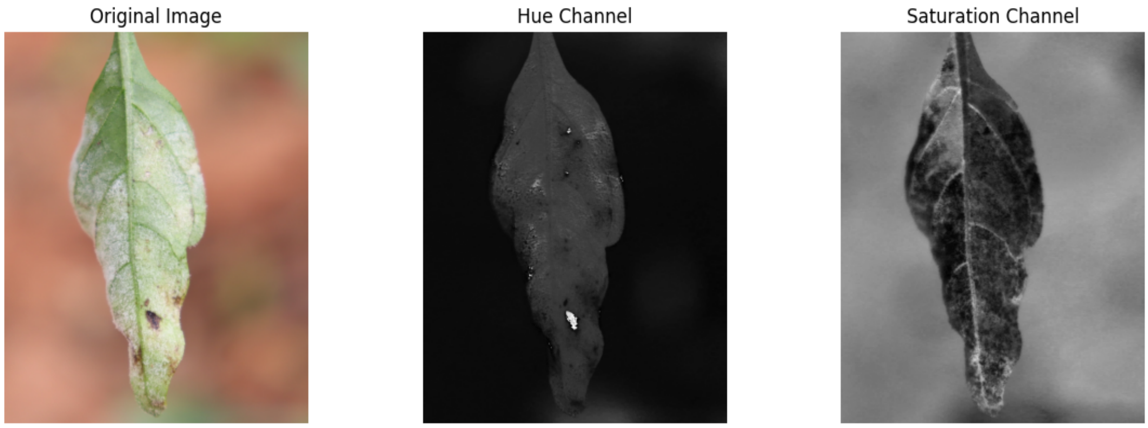

Figure 6: Converted image

Figure 6 illustrates the results of converting a chili leaf image affected by Powdery Mildew from the RGB color space to the HSV color space, which includes the original image, the Hue channel, and the Saturation channel. In the original RGB image, the chili leaf exhibits grayish-white spots on the leaf surface, which are characteristic symptoms of Powdery Mildew, while the background appears relatively uniform and lacks prominent patterns. After conversion to the HSV color space, the Hue channel shows the distribution of color tones across the leaf surface. Although the overall shape of the leaf remains recognizable, the variation in hue appears less distinct due to low contrast, resulting in less prominent color detail on the leaf surface. In contrast, the Saturation channel displays the leaf area with higher contrast compared to the backgound. The higher saturation values of the leaf object make its shape and texture more clearly visible, while the background exhibits lower saturation levels. This visualization demonstrates that the Saturation channel is more effective in highlighting the chili leaf object than the Hue channel. Therefore, the Saturation channel was used as the basis for the segmentation stage to more accurately separate the leaf object from the background.

b. Segmentation

The segmentation stage aims to separate the chili leaf object from the image background by utilizing the Saturation channel in the HSV color space. In this step, a threshold value of $S > 0.2$ was applied to select regions with relatively high color saturation, which were assumed to correspond to the leaf area. This approach is effective because the leaf object generally exhibits higher saturation values than the background.

After the thresholding process, the segmentation results were refined using morphological operations. The *bwareaopen* operation was employed to remove noise or small unwanted objects, while the *imfill* operation was applied to fill holes within the leaf regions. The output of this stage is a clear binary mask, which was then used to separate the leaf object from the background of the original image.

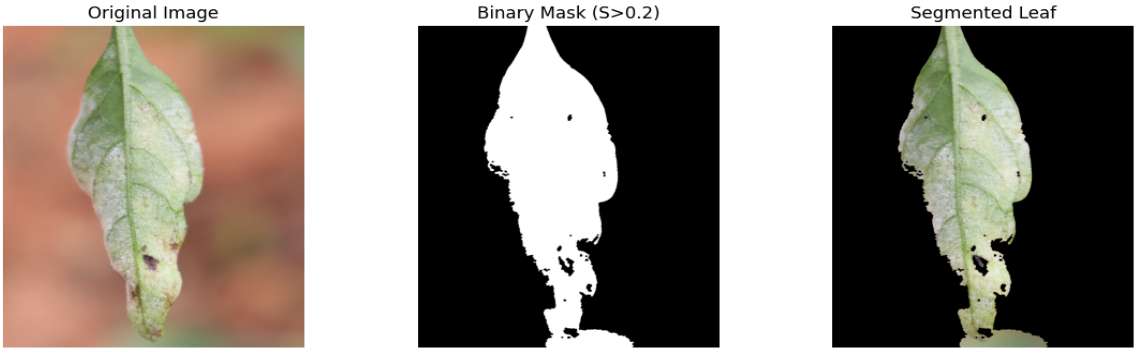

Figure 7: Segmentated image

Figure 7 shows the stages of chili leaf image segmentation. The left panel presents the original chili leaf image, the middle panel displays the binary mask obtained from thresholding (S > 0.2), where the leaf area is represented in white and the background in black. The right panel shows the final segmented leaf image, in which only the leaf region is retained with its color and texture details still visible, while the background has been completely removed. This segmentation process enables accurate separation of the leaf object, thereby supporting the subsequent stages of feature extraction and chili leaf disease classification.

c. Median Filtering

At this stage, the segmented chili leaf images were processed using median filtering to reduce noise without degrading the edges or the main structural features of the leaves. A 3 × 3 median filter was applied separately to the Hue and Saturation channels. This approach was selected because it is effective in suppressing small spot-like noise while preserving important details on the leaf surface.

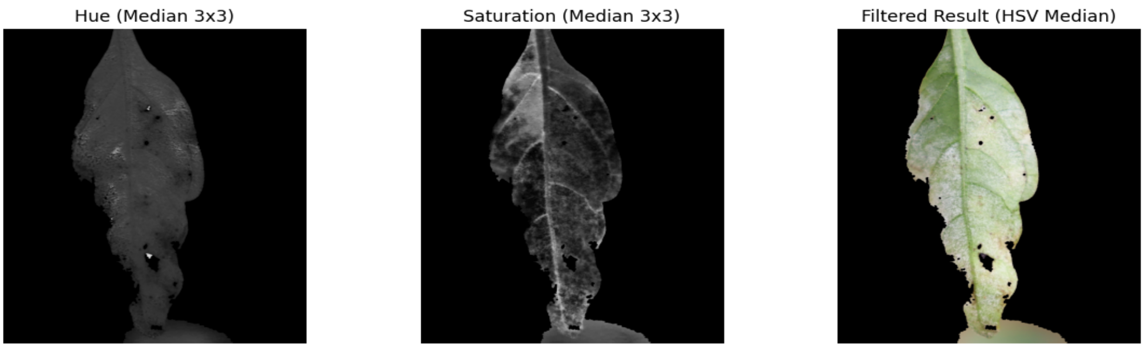

Figure 8: Median Filtering image

The application of median filtering to the Hue channel resulted in more uniform color intensity, while the Saturation channel exhibited improved clarity of leaf structures, including veins and surface texture. After the filtering process was completed, the filtered HSV channels were recombined to produce the final image (Filtered Result – HSV Median).

As shown in Figure 8, the final image appears cleaner with reduced noise, while the main color characteristics and shape of the chili leaf are well preserved. This median filtering process produces smoother and more stable images, making them more suitable for subsequent stages of feature extraction and chili leaf disease classification.

2. Feature Extraction

During the feature extraction stage, the pre-processed chili leaf images were analyzed to obtain color and texture features. Color features were extracted using the HSV (Hue, Saturation, and Value) color space to represent the color distribution on the leaf surface, while texture features were extracted using the Local Binary Pattern (LBP) method to describe local texture patterns of the chili leaves. The combination of these two types of features was employed to enhance the model's ability to distinguish between different classes of chili leaf diseases.

Feature extraction was performed on the testing images, which consist of five classes: Cercospora, Healthy, Murda Complex, Nutritional Deficiency, and Powdery Mildew. The HSV color feature values presented in Table 1 represent the average feature values for each disease class.

Table 2. HSV Color Feature Values of Chili Leaves

Class	Mean Hue (H)	Mean Saturation (S)	Mean Value (V)	Std. Dev. Saturation (S)	Std. Dev. Value (V)
Cercospora	0.214	0.198	0.176	0.155	0.132
Healthy	0.308	0.286	0.264	0.238	0.211
Murda Complex	0.261	0.244	0.219	0.191	0.165
Nutritional Deficiency	0.231	0.213	0.192	0.170	0.149
Powdery Mildew	0.275	0.251	0.228	0.201	0.176

Variations in color distribution among chili leaf classes can be observed from the HSV color feature values presented in Table 2, where healthy leaves generally exhibit higher and more stable Hue and Saturation values compared to disease-infected leaves. In contrast, leaves affected by Powdery Mildew show a decrease in Saturation values due to the presence of grayish-white spots on the leaf surface. These differences indicate that HSV color features are capable of effectively representing visual changes in chili leaves and are therefore suitable to be used as a basis for chili leaf disease classification.

Table 3. LBP Texture Feature Values of Chili Leaves

Class	LBP 1	LBP 2	LBP 3	LBP 4	LBP 5
Cercospora	0.142	0.168	0.193	0.214	0.236
Healthy	0.118	0.145	0.172	0.196	0.219
Murda Complex	0.161	0.187	0.214	0.238	0.261

Nutritional Deficiency	0.154	0.179	0.203	0.226	0.249
Powdery Mildew	0.172	0.198	0.224	0.248	0.271

Nilai The LBP texture feature values presented in Table 3 represent the average LBP histogram values for each chili leaf disease class. Healthy chili leaves tend to exhibit more uniform texture patterns, resulting in more stable LBP values, whereas leaves affected by various diseases show more irregular textures due to changes in leaf tissue structure, such as spots, tissue damage, and surface irregularities. These differences in texture characteristics indicate that LBP features are effective in capturing variations in the surface conditions of chili leaves and contribute to the disease classification process. Higher LBP histogram values observed in disease classes reflect increased variability in local texture patterns, which corresponds to surface irregularities caused by spots, fungal infections, and tissue damage.

3. Classification

During the classification stage, the K-Nearest Neighbors (KNN) algorithm was used to categorize chili leaf images into five classes: *Cercospora*, *Healthy*, *Murda Complex*, *Nutritional Deficiency*, and *Powdery Mildew*. The dataset was divided into 35 training samples and 15 testing samples for each class, resulting in a total of 175 training data and 75 testing data.

The classification process was conducted using variations of K = 3, 5, and 7 with Euclidean distance as the distance metric. The class of each testing image was determined based on the majority class among the K nearest neighbors. To analyze the effect of different feature types on classification performance, three feature scenarios were employed: HSV color features, LBP texture features, and a combination of HSV and LBP features.

Table 4. KNN Classification Results Using Different Feature Types

Jenis Fitur	**K = 3**	**K = 5**	**K = 7**
HSV	56	59	57
LBP	51	53	52
HSV + LBP	61	64	62

The values presented in Table 4 indicate the number of testing images that were correctly classified out of a total of 75 testing samples for each combination of feature type and K value.

Based on Table 4, the performance of the KNN algorithm is influenced by both the type of features used and the value of K. The use of HSV color features resulted in a higher number of correctly classified samples compared to LBP texture features, indicating that color variations in chili leaves are an important indicator for distinguishing different disease classes.

The best performance was achieved when color and texture features were combined (HSV + LBP), with the highest number of correct classifications obtained at K = 5. This finding indicates that the integration of color and texture information is able to represent the characteristics of chili leaves more comprehensively, thereby improving both the accuracy and stability of the classification process. Overall, the results of this study demonstrate that appropriate feature selection and the determination of an optimal K value play a crucial role in enhancing the performance of the KNN algorithm for image-based chili leaf disease classification.

4. Evaluation

The evaluation stage was conducted to assess the performance of the K-Nearest Neighbors (KNN) algorithm in classifying chili leaf diseases based on the features used. The evaluation was performed on 75 testing samples derived from five chili leaf classes: *Cercospora*, *Healthy*, *Murda Complex*, *Nutritional Deficiency*, and *Powdery Mildew*. The predicted results were compared with the ground-truth labels to determine the classification accuracy.

Model performance was assessed using accuracy, which was calculated as the ratio of correctly classified samples to the total number of testing samples. The evaluation was carried out using three variations of K values, namely $K = 3$, $K = 5$, and $K = 7$, and three feature scenarios: HSV color features, LBP texture features, and a combination of HSV and LBP features.

Table 5. KNN Accuracy Evaluation Results

Nilai K	HSV	LBP	HSV + LBP
K = 3	74.67	68.00	81.33
K = 5	78.67	70.67	85.33
K = 7	76.00	69.33	82.67

Based on the evaluation results presented in Table 5, it can be observed that the combination of HSV and LBP features consistently produces the highest accuracy compared to the use of a single feature type. This finding indicates that color and texture information complement each other in representing the characteristics of chili leaves.

HSV color features demonstrate better performance than LBP texture features because most chili leaf diseases exhibit noticeable color changes on the leaf surface. In contrast, LBP features yield lower accuracy due to similarities in texture patterns among several disease classes. Among the tested K values, $K = 5$ provides the most stable and optimal results and was therefore selected as the best K value in this study.

4. CONCLUSION

This study demonstrates that the K-Nearest Neighbors (KNN) algorithm can be effectively applied for image-based classification of chili leaf diseases by utilizing a combination of HSV color features and Local Binary Pattern (LBP) texture features. The dataset used in this study consists of 250 chili leaf images divided into five classes: Cercospora, Healthy, Murda Complex, Nutritional Deficiency, and Powdery Mildew. The data were proportionally split into 175 training images (70%) and 75 testing images (30%), ensuring a balanced distribution across all classes.

The experimental results indicate that the combination of color and texture features provides better classification performance than the use of a single feature type. HSV features effectively represent color changes in leaves caused by disease, while LBP features capture variations in surface texture patterns. The integration of these two feature types results in a more comprehensive visual representation, thereby improving classification accuracy.

Furthermore, the evaluation of different K values in the KNN algorithm shows that K = 5 yields the most stable and optimal performance compared to other values. A small

K value tends to be more sensitive to noise, whereas a large K value may reduce the model's ability to discriminate between disease classes.

Although the obtained results are satisfactory, some misclassifications still occur, particularly among disease classes with similar color and texture characteristics. Therefore, future research may be enhanced by incorporating more discriminative features, applying alternative classification methods, or using larger and more diverse datasets to improve the accuracy and reliability of chili leaf disease detection systems.

REFERENCES

Acito, F. (2023). *k Nearest Neighbors*. https://doi.org/10.1007/978-3-031-45630-5_10

Afriyanti, G., Mariya, A., Natalia, C., Nispuana, S., Farhan Wijaya, M., & Phalepi, M. Y. (2023). The Role Of The Agricultural Sector On Economic Growth In Indonesia. *Indonesian Journal of Multidisciplinary Sciences (IJoMS)*, *2*, 167–179.

Aishwarya, M. P., & Reddy, A. P. (2024). Dataset of Chilli and Onion Plant Leaf Images for Classification and Detection. *Data in Brief*. https://doi.org/10.1016/j.dib.2024.110524

Aminuddin, N. F., Joret, A., Zulkifli, S. A., Morsin, M., & Tukiran, Z. (2023). Computational Approaches Based On Image Processingfor Automated Disease IdentificationOn Chili Leaf Images: A Review. *EMERGING ADVANCES IN INTEGRATED TECHNOLOGY*, *3*(2). https://doi.org/10.30880/emait.2022.03.02.002

Araujo, S. D. C. S., Malemathh, V. S., & Sundaram, K. M. (2022). Automated G-4 Disease Identification Agricultural Intelligent System: An impression and survey. *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, *2*, 37–43. https://doi.org/10.1109/iciptm54933.2022.9754145

Ayeshmi, S. R. M. A. M., Pranathi, J., Begum, D., Haritha, V., Khan, S., & Y, S. (2025). *Disease Detection In Chilli Plants By Using Deep Learning*. *5*(2). https://doi.org/10.33425/3066-1226.1103

Bangun, R. H. (2021). The determinants of production and feacibility of chili pepper in province of sumatera utara. *Agrisocionomics*, *5*(2), 8–18. https://doi.org/10.14710/agrisocionomics.v5i2.7154

Baurai, A., Rawat, P., Singh, A. N., Bhakuni, V. S., Ayush, & Manwal, M. (2024). *Leaf Disease Detection Using KNN*. 453–458. https://doi.org/10.1109/ictacs62700.2024.10840461

Govindarajan, S., & S, M. V. (2023). *A Sequential Method for Enhancing the Image Quality of Rice Leaf Diseases by Employing Various Preprocessing Techniques*. https://doi.org/10.1109/icsss58085.2023.10407605

Harshitha, H. S., Nagaraja, J., & Pruthiraja, D. (2024). *Plant Disease Detection Using Image Processing*. 1–6. https://doi.org/10.1109/icait61638.2024.10690485

Iqbal, M., Hussain, D., Javed, M. Y., Ellahi, F., Asim, M., Adeel, M., … Baqir, M. (2023). *In Vitro Efficacy Evaluation of Agrochemical Fungicides in the Mitigation of Colletotrichum Capsici Infection*. *2*(Issue 02), 128–135. https://doi.org/10.69501/0vk41476

Jia, Z., & Liao, S. (2023). *Leaf Recognition Using K-Nearest Neighbors Algorithm with Zernike Moments*. 665–669. https://doi.org/10.1109/icivc58118.2023.10270642

Korolev, M. I., Khorev, A., Salikov, Y., & Kolomyceva, O. (2023). *Development of scientific research in the field of food security at the global, national and regional levels*. https://doi.org/10.20914/2310-1202-2022-3-492-499

Kunta, M., Park, J.-W., Braswell, W. E., da Graça, J. V, & Edwards, P. (2021). *Modern Tools for Detection and Diagnosis of Plant Pathogens*. Springer, Singapore. https://doi.org/10.1007/978-981-15-6275-4_4

Lokeswari, N., Pramesti, D., & Fakhrurroja, H. (2025). *Enhancing the Accuracy of Chili Plant Disease Classification using Convolutional Neural Networks and Image Augmentation*. 123–128. https://doi.org/10.1109/iaict65714.2025.11100595

Loti, N. N. A., Noor, M. R. M., & Chang, S.-W. (2021). Integrated analysis of machine learning and deep learning in chili pest and disease identification. *Journal of the Science of Food and Agriculture*, *101*(9), 3582–3594. https://doi.org/10.1002/JSFA.10987

Mahanta, D., Dange, M. M., Trivedi, A., & Nandeha, N. (2024). Global Challenges Facing Plant Pathology: A Review on Multidisciplinary Approaches to Meet the Food Security. *Journal of Scientific Research and Reports*. https://doi.org/10.9734/jsrr/2024/v30i62106

Mallesh, A. S., Pamarthi, N., Murty, P. T. S., Sree, P. K., Daniya, T., & Maram, B. (2023). *Smart System for Early Detection of Agricultural Plant Diseases in the Vegetation Period*. https://doi.org/10.1109/idicaiei58380.2023.10406672

Mashuri, A. S., Sunyoto, A., & Kusnawi, K. (2024). Klasifikasi Penyakit Pada Daun Cabai Menggunakan Arsitektur VGG16. *Journal of Electrical Engineering and Computer*, *6*(2), 305–313. https://doi.org/10.33650/jeecom.v6i2.9116

Mohammed, A. S., Nayyef, Y. A., & Tuama, H. N. (2025). Real-Time Plant Disease Detection by AI. *European Journal of Ecology, Biology and Agriculture.*, *2*(4), 52–69. https://doi.org/10.59324/ejeba.2025.2(4).05

Nasir, M., Suciati, N., & Wijaya, A. Y. (2017). *Kombinasi Fitur Tekstur Local Binary Pattern yang Invariant Terhadap Rotasi dengan Fitur Warna Berbasis Ruang Warna HSV untuk Temu Kembali Citra Kain Tradisional*. *7*(1), 42–51. https://doi.org/10.35585/INSPIR.V7I1.2435

Pakutharivu, P., Sasirekha, D., Devaraj, V., & Gopi, R. S. (2023). Improving Plant Disease Detection Using Super-Resolution Generative Adversarial Networks and Enhanced Dataset Diversity. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. https://doi.org/10.37934/araset.35.2.144157

Patil, A. R., Patil, V. I., & Lad, K. B. (2022). *Feature Selection for Chili Leaf Disease Identification Using GLCM Algorithm*. https://doi.org/10.1007/978-981-16-3945-6_35

Pratama, Y., Rasywir, E., Suyanti, S., Siswanto, A., & Fachruddin, F. (2025). Enhancing Areca Nut Detection and Classification Using Faster R-CNN: Addressing Dataset Limitations with Haar-like Features, Integral Image, and Anchor Box Optimization. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *9*(3), 613–621. https://doi.org/10.29207/resti.v9i3.6496

Pulungan, M. R., Furqan, M., & Rifki, M. I. (2024). Klasifikasi penyakit pada daun cabai menggunakan gray level co-occurrence matrix dan k-nearest neighbor. *Syntax*, *5*(2), 549–554. https://doi.org/10.46576/syntax.v5i2.5386

Purboseno, S., Dharmawati, N. D., & Rahayu, E. (2025). Strategies for Low-Carbon Agricultural Development Through Smart Farming of Chili Peppers. *KnE Life Sciences*, *9*(1), 79–94. https://doi.org/10.18502/kls.v9i1.19352

Rahadiyan, D., Hartati, S., Wahyono, & Nugroho, A. P. (2022). Classification of Chili Plant Condition based on Color and Texture Features. *International Conference on Intelligent Computing*, 1–7. https://doi.org/10.1109/ICIC56845.2022.10006975

Rajendrakumar, S., Rajashekarappa, R., & Parvati, V. K. (2025). Identification of chilli leaf disease using contrast limited histogram equalisation and k-means clustering. *Indonesian Journal of Electrical Engineering and Computer Science*, *39*(2), 1100. https://doi.org/10.11591/ijeecs.v39.i2.pp1100-1108

Raut, S., & Kasat, N. (2024). *A Review: Citrus Disease Detection Using Machine Learning Approach*. 1–5. https://doi.org/10.1109/idicaiei61867.2024.10842864

Ruby, E. D. K., Amirthayogam, G., Sasi, G., Thangavel, C., Choubey, A., & Gopalakrishnan, S. (2024). Advanced Image Processing Techniques for Automated Detection of Healthy and Infected Leaves in Agricultural Systems. *Mesopotamian Journal of Computer Science*, *2024*, 62–70. https://doi.org/10.58496/mjcsc/2024/006

Rzanny, M., Seeland, M., Wäldchen, J., & Mäder, P. (2017). Acquiring and preprocessing leaf images for automated plant identification: understanding the tradeoff between effort and information gain. *Plant Methods*, *13*(1), 97. https://doi.org/10.1186/S13007-017-0245-8

Shaker, B. R. M., Kumar, J. H., Chaitanya, V., Sriranjitha, P., Kumar, K. R., & Rao, P. J. M. (2021). Economics of Chilli Cultivation in Khammam District of Telangana. *International Journal of Current Microbiology and Applied Sciences*, *10*(2), 893–901. https://doi.org/10.20546/IJCMAS.2021.1002.105

Shaparia, R., Patel, N. M., & Shah, Z. H. (2017). *Flower Classification using Texture and Color Features*. *2*, 113–118. https://doi.org/10.29007/6MT1

Suwarningsih, W., Evandri, E., Kirana, R., Khotimah, P. H., Riswantini, D., Nugraheni, E., … Ahmadi, N. R. (2024). Kumpulan data citra telepon pintar untuk identifikasi varietas cabai merah berbasis daun. *Baca: Jurnal Dokumentasi Dan Informasi*, 51–59. https://doi.org/10.55981/baca.2024.7786

Tiwari, R. G., Nyamasvisva, T. E., Ibrahim, N., Dixit, A., Trivedi, N. K., & Kumar, A. (2025). Hybrid Feature Fusion and Deep Learning for High-Accuracy Anthracnose Detection in Chili Plants. *Journal of Innovative Image Processing*, *7*(3), 602–621. https://doi.org/10.36548/jiip.2025.3.002

Toennies, K. D. (2024). *Distance-Based Classifiers*. https://doi.org/10.1007/978-981-99-7882-3_5

Tripathi, A., Maurya, S., Pandey, K. K., & Behera, T. K. (2024). Global Scenario of Vegetable Fungal Diseases. *Vegetable Science*, *51*(Special Is), 54–65. https://doi.org/10.61180/vegsci.2024.v51.spl.06

Turkoglu, M., & Hanbay, D. (2019). *Combination of Deep Features and KNN Algorithm for Classification of Leaf-Based Plant Species*. https://doi.org/10.1109/IDAP.2019.8875911

Turnandes, Y., & Irwanda, A. A. (2024). Analisis sentimen pada ulasan kegiatan seminar nasional sistem informasi dan teknologi komputer 2023 menggunakan natural language processing (NLP). Jurnal Karya Ilmiah Multidisiplin (JURKIM), 4(2), 101–109. https://doi.org/10.31849/jurkim.v4i2.22063

Vedanty, P. P., Kesiman, M. W. A., Sunarya, I. M. G., & Indradewi, I. G. A. A. D. (2023). *Identification of Leaf Diseases of Medicinal Plants Using K-Nearest Neighbor Based on Color, Texture, and Shape Features*. 1–6. https://doi.org/10.1109/icaicta59291.2023.10390034

Vega-Huerta, H., Pajuelo-León, J., la Cruz Vélez de Villa, P. E. De, Calderon-Vilca, H. D., Maquen-Niño, G. L. E., Rios-Castillo, M. E., … Benito-Pacheco, O. (2025). K-Nearest Neighbors Model to Optimize Data Classification According to the Water Quality Index of the Upper Basin of the City of Huarmey. *Applied Sciences*, *15*(18), 10202. https://doi.org/10.3390/app151810202

Weng, G., Liu, W., & Guo, L. (2025). *Improving Accuracy of Corn Leaf Disease Recognition Through Image Enhancement Techniques*. *2*(5), 1–12. https://doi.org/10.70393/6a6374616d.333136

Yunefri, Y., Sutejo, Y. E., Fadrial, K. A., Ramadhani, M., & Hardianto, R. (2022).

Implementation of object detection with you only look once algorithm in limited face-to-face times in pandemic. *Journal of Applied Engineering and Technological Science*, *4*(1), 400-404.