



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2024, 2024:19

<https://doi.org/10.28919/cmbn/8404>

ISSN: 2052-2541

## CORN STALK DISEASE CLASSIFICATION USING RANDOM FOREST COMBINATION OF EXTRACTION FEATURES

NACHNUL ANSORI<sup>1,\*</sup>, AERI RACHMAD<sup>2</sup>, EKA MALA SARI ROCHMAN<sup>2</sup>, HERMAWAN BIN FAUZAN<sup>2</sup>,  
YULI PANCA ASMARA<sup>3</sup>

<sup>1</sup>Industrial Engineering, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

<sup>2</sup>Department of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

<sup>3</sup>Faculty of Engineering and Quantity Surveying, INTI International University, Negeri Sembilan 71800, Malaysia

Copyright © 2024 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract:** Agriculture is one of the essential sectors for human livelihood sustainability. The primary crop cultivated worldwide is corn. Unfortunately, it is often susceptible to various diseases that can threaten crop yields and food availability. One vulnerable part of the corn plant to bacterial and viral infections is the corn stalk. The corn stalk disease is a critical issue that can impact the growth and yield of the crop. It serves as the primary support system for the plant and, is crucial for maintaining the stability and productivity of corn plants. Therefore, a preventive effort to maintain plant health and enhance agricultural productivity for initial detection is essential. Technologies in data mining for digital image classification are implemented. To classify corn stalk diseases, this study suggests machine learning strategies such as Random Forest, Support Vector Machine, and K-Nearest Neighbor (K-NN). Furthermore, a mix of LBP (Nearby Double Example) and HSV (Tone Immersion Worth) highlight extraction is utilized in this examination. A dataset of digital images of corn plants containing 750 records with 5 classes is assessed. Results show

---

\*Corresponding author

E-mail address: [nachnul@trunojoyo.ac.id](mailto:nachnul@trunojoyo.ac.id)

Received December 17, 2023

the highest accuracy that the Random Forest algorithm has 82%, and AUC is 96.2%. F1-Score, Precision, and the Recall are 82%.

**Keywords:** corn stalks disease; food productivity; classification; random forest; support vector machine; K-Nearest Neighbor (K-NN).

**2020 AMS Subject Classification:** 92C80.

## 1. INTRODUCTION

Agriculture is one of the crucial sectors in the sustainability of human life [1]. Meanwhile, corn is one of the major crops cultivated worldwide [2]. In Indonesia, corn is one of the staple foods widely consumed, instead of rice, as a source of carbohydrates [3]. However, corn is often susceptible to various diseases that can threaten crop yields and food availability. One vulnerable part of the corn plant to bacterial and viral infections is the corn stalk. Diseases affecting the corn stalk are a serious issue that can impact the growth and yield of this crop. The corn stalk, serving as the main support system for the plant, is crucial for maintaining the stability and productivity of corn crops.

Diseases of corn can be caused by various factors, including pathogens such as fungi, bacteria, or viruses [4] and [5]. The symptoms of diseases in corn can vary, ranging from lesions, tissue damage, and to the death of stem sections. These diseases can lead to reduced growth, decreased crop yields, and even significant economic losses for farmers. In addition, environmental factors such as humidity, temperature, and plant density can influence the development of diseases in corn stalks [6]. Therefore, preliminary detection is an initial preventive measure for maintaining plant health and enhancing agricultural productivity.

Being the preliminary disease detection implemented, technology such as data mining in digital image classification plays a crucial role. The classification of digital images for diseases in corn by Machine Learning technology is an innovative approach with significant potential to address challenges in modern agriculture. Similar to its application in the 2018 research on butterfly image classification using HSV and Local Binary Pattern as feature extraction methods

with the GLCM algorithm, the study revealed that the combination of these two extraction methods resulted in an accuracy rate of 72% [7].

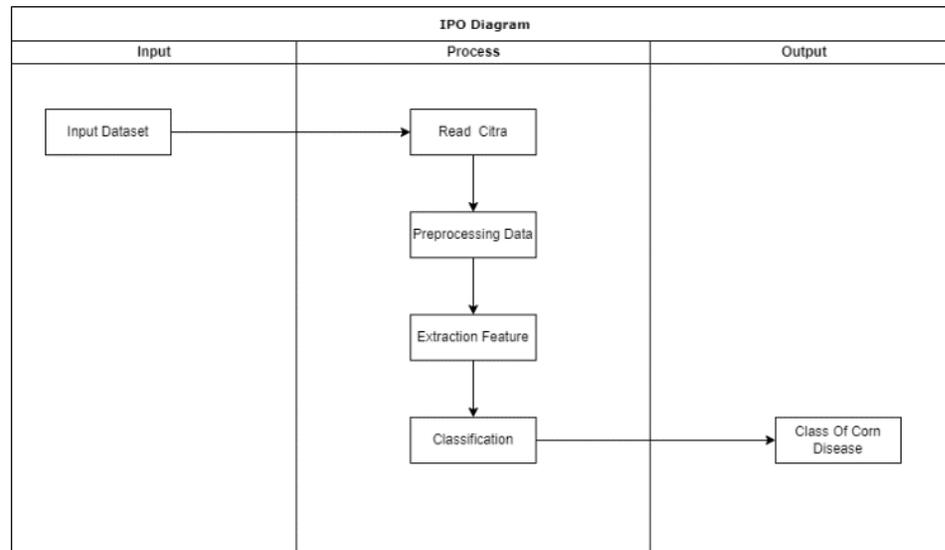
Furthermore, in the year 2020, a study also utilized the Local Binary Patterns method for feature extraction to classify digital brain tumor images using multiple algorithms, including K-Nearest Neighbor, Neural Networks, Random Forest, and LDA (Linear Discriminant Analysis). Based on the results of that research, the K-Nearest Neighbor method demonstrated the most optimal accuracy, achieving a classification accuracy of 95.56% [8]. In addition, in the year 2022, a study was conducted on the classification of kinnow orange images using the Support Vector Machine method, with feature extraction performed using the Local Binary Patterns method. This research yielded excellent accuracy, reaching 94.67% [9].

Therefore, based on previous issues and research, the researchers are interested in examining several machine learning algorithms, namely K-Nearest Neighbor, Support Vector Machine, and Random Forest, as methods for classification using the Local Binary Pattern (LBP) and Hue Saturation Value (HSV) methods for feature extraction. These methods are capable of modeling a system to classify diseases in corn with high accuracy through digital image analysis. Consequently, this system will assist farmers and agricultural experts in detecting diseases in corn more quickly and efficiently, allowing for prompt preventive or treatment measures. The research utilizes a dataset consisting of 750 digital image records categorized into 5 classes.

## **2. PRELIMINARIES**

This study focuses on classifying corn stalk diseases based on image data. Classification is one aspect of the field of data mining. Data mining is a process that involves the extraction, identification, and analysis of various pieces of information to identify patterns in the data using mathematical, artificial intelligence, statistical, and machine learning methods [10][11]. The term data mining is also often referred to as Knowledge Discovery from Data (KDD). In terms of its functions, data mining can be categorized into several categories, such as description, prediction, estimation, classification, clustering, and association [12].

In this research, the classification of digital images involves several stages. They are involved in data collection, feature extraction, and classification, culminating in the evaluation process to assess the performance of the developed classification model. Figure 1 shows the IPO diagram that represent the steps in the classification process.



**Fig. 1** IPO Diagram

Base on Figure 1, the formation of the classification system consists of three parts. They are, the input phase, the core phase, and the output phase. The detailed components are:

- a. Input Data Process, this is the initial step in the data mining process, which involves importing the dataset to be classified.
- b. Preprocessing Data, the process involves separating objects from the background of the image by performing several steps, including resizing, contrast stretching, and then proceeding with segmentation using thresholding.
- c. Feature Extraction, the feature extraction process is performed to obtain a set of features that can represent the unique characteristics of each image. In this research, feature extraction methods used are HSV for color feature extraction and LBP for texture feature extraction.
- d. Validation Process, involves splitting the data with 80% training data and 20% testing data ratio. Then, the 80% of training data is augmented and further divided into training and validation data. This research uses a k-fold value of 5.

- e. Classification, the learning process is carried out to obtain a classification model using several different approaches, namely K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and Random Forest. The accuracy results from these various classification models will then be analyzed based on the evaluation results obtained from each algorithm to determine the most optimal classification outcome.
- f. Output, the result of the classification process will categorize the attributes of corn plant data into their respective target classes based on the model created using the proposed methods in this research.

### 2.1. Data Input

The data used in this classification process consists of image data from corn plants, comprising 750 data records with 5 target classes. The details of the target class can be seen in Table 1.

**Table 1** Dataset Class

No	Class	Total
1	Healthy	90
2	Erwina Carotovora	102
3	Pythium	131
4	Stenocarpella	210
5	Gibberella	217
<b>Total</b>		<b>750</b>

### 2.2. Feature Extraction

Feature extraction in digital images is the process of identifying and extracting important information or specific characteristics from a digital image [13] and [14]. It is to reduce the complexity of data in the image and provide the image in a simpler and more meaningful form that can be used for further analysis, such as object recognition, classification, or pattern detection. In this process, features such as edges, textures, colors, and shapes can be extracted from the image

to assist machines or computers in understanding and making decisions based on the information contained in the image. Two feature extraction methods that are used in this study Local Binary Pattern (LBP) and Hue Saturation Value (HSV).

### **A. Local Binary Pattern (LBP)**

Local Binary Pattern (LBP) is a method used in image processing to extract features that are valuable for texture analysis [15] and [16]. This method operates by comparing the pixel values around each pixel in the image to the value of the central pixel (the pixel being analyzed) and then generates a binary representation based on these comparisons. The main benefit of the LBP method is its ability to extract robust texture features that are resistant to variations in lighting [17]. This makes it highly valuable in applications such as face recognition, object detection, and texture analysis in medical images.

The steps of LBP extraction method are [18]:

1. Input the image data.
2. Convert the input image data into grayscale images.
3. Set the neighbor value to 8, and the radius value to 1.
4. Having neighbor pixel values, calculate the threshold value by comparing the intensity value of the central pixel (the pixel being processed) with the intensity value of each neighboring pixel. If the value of a neighboring pixel is greater than or equal to the central pixel's value, assign a value of 1; otherwise, assign a value of 0. The result of this thresholding calculation will produce a binary number that represents the neighbor pattern. This stage is also referred to as the binarization step.
5. The result of the previous step is the Local Binary Pattern (LBP) value for a pixel, which will be converted into a decimal number.
6. By calculating the LBP values for each pixel in the image, the LBP histogram as the output is gained.

### **B. Hue Saturation Value (HSV)**

The HSV is a method to extract features from digital images by utilizing color information

within the image. This involves converting the image into the HSV color space, which has three main components: Hue (color), Saturation (color intensity), and Value (intensity value) [19]. The detailed explanations are:

1. Hue, this component describes the type of color in the image, such as red, blue, or green.
2. Saturation, this component reflects how vividly the colors appear in the image and can be thought of as the strength or purity of the color.
3. Value, this component represents the level of brightness or darkness in the image. In Value, texture details in the image become more pronounced. By analyzing the variations in intensity values across the entire image, we can extract texture features such as patterns, structures, and contrast.

### 2.3 The K-nearest Neighbor (K-NN)

The K-nearest neighbor (K-NN) algorithm is one of the methods in machine learning used for classification and regression problems [20]. The fundamental idea of this algorithm is that an object tends to be similar to its nearest neighbors. Therefore, when using this algorithm to determine the category or value that best fits an object, it relies on the category or value of the most similar objects in its vicinity. The steps involved in applying this algorithm to classify data are as follows [21]:

1. The first step is to choose the value of K, which is the number of nearest neighbors to be used for making predictions. The choice of K can influence the prediction outcomes. A small K value tends to be more sensitive to noise, while a larger K value tends to result in more general predictions.
2. The second step is to calculate distances. The K-NN algorithm uses distance metrics like Euclidean, Manhattan, or others. This is done for each data point in the dataset concerning the data point to predict. This research employs distance measurement using the Euclidean method, which can be interpreted with the following equation:

$$dist(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

$d(x, y)$ : Euclidean distance

$X$  : Data 1

$Y$  : Data 2

$i$  : Attribute  $i$ -th

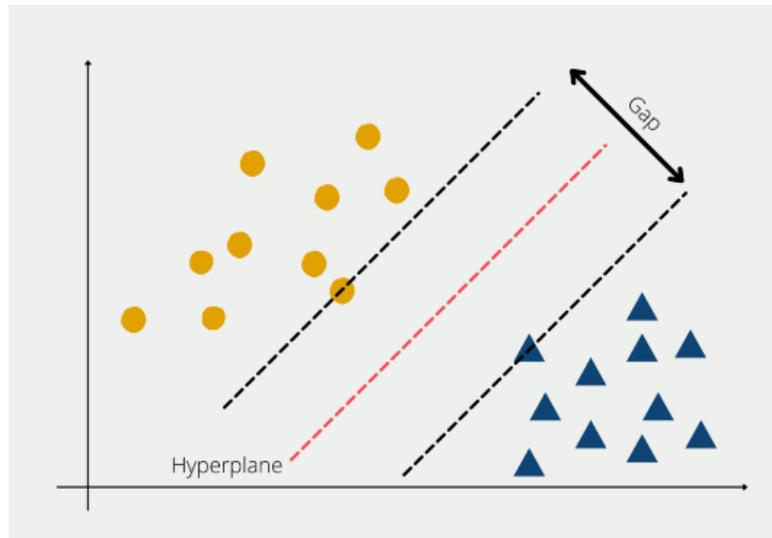
$n$  : Number of attribute

3. Calculating the distances between the data points to predict, sort the distance values from largest to smallest. Then, selection the  $K$  data points with the closest distances to the data point can be predicted. These data points are referred to as neighbors.
4. The fourth step is to collect the categories or class labels from the neighbors' data based on their  $K$  values.
5. The final step in the  $K$ -Nearest Neighbor ( $K$ -NN) algorithm is to determine the majority category among the nearest neighbors to be used in predicting the category for a new data point, known as the test data.

The  $K$ -Nearest Neighbor method is widely used in classifying data since it is relatively easy to understand and implement. Moreover, it can handle noisy data or outliers quite well because it primarily considers the nearest neighbors. However, this method can be computationally intensive when applied to large datasets because it requires distance calculations between every data point in the dataset.

## 2.4 Support Vector Machine

Support Vector Machine (SVM) is one of the machine learning algorithms used for classification problems and can also be applied to regression problems. SVM operates by finding the best separator (hyperplane) between two different classes of data. The objective is to locate the hyperplane with the largest margin between the two classes, thus minimizing the risk of classification errors [22] and [23]. Here is an illustration of the SVM algorithm, as shown in Figure 2.



**Fig. 2** Support Vector Machine

Figure 2 illustrates two groups of data distinguished by different patterns. As seen in the image, these two groups are separated by a dashed red line referred to as a hyperplane. In the context of this algorithm, the hyperplane will be adjusted in such a way as to be the most efficient in separating the two different data groups. The more efficient the resulting hyperplane, the greater its ability to reduce the classification error rate in the system.

SVM can use kernel functions to transform data into higher dimensions. This allows SVM to handle non-linear data [24]. There are several kernel functions provided by this algorithm, including:

- a. Kernel Linier

$$K(x_i, x) = x_i^T x \quad (2)$$

- b. Kernel Polynomial

$$K(x_i, x) = (\gamma x_i^T + r)^p, \gamma > 0 \quad (3)$$

- c. Kernel Radial Basis Function

$$K(x_i, x) = \exp(-\gamma \|x - x_i\|^2) \quad (4)$$

- d. Sigmoid Kernel

$$K(x_i, x) = \tanh(\gamma x_i^T + r) \quad (5)$$

There are steps in performing classification using the Support Vector Machine (SVM) algorithm as follow:

1. The first step is to initialize all alpha values to zero and the bias to zero.
2. Next, calculate the Hessian matrix. The Hessian matrix is the product of the kernel function and the y values. The y values are the vector values of 1 and -1. The Hessian matrix can be calculated using the following equation.

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (6)$$

Note:

$D_{ij}$  = The element of the Hessian matrix at the i-th row j column.

$\lambda$  = the theoretical limits that will be derived

$y_i$  = the class of i data

$y_j$  = the class of j data

3. The third step involves calculating the error value using equation 7. Then, calculate the delta alpha values to update the alpha values using equation 8. Subsequently, refine the alpha values with the new alpha values using equation 9 as follows.

$$E_i = \sum_{i=1}^1 a_i D_{ij} \quad (7)$$

$$\delta a_i = \min\{\max[\gamma(1 - E_i), -a_i], C - a_i\} \quad (8)$$

$$a_i = a_i \delta a_i \quad (9)$$

Note:

$E_i$  = error score

$\alpha_i$  = alpha

$\delta \alpha_i$  = delta alpha

$C$  = Constanta

4. The fourth step is to calculate the new bias value using equation 10.

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (10)$$

5. Next, in a high-dimensional space where the number of features is very large, which makes it suitable for many applications, including image processing and text recognition, involves computing the result of the dot product between the training data and the testing data.

The final step is to determine the class of the test data using equation 11.

$$f(x) = \sum_{i=1}^l a_i y_i K(x_i, x) + b \quad (11)$$

Currently, the Support Vector Machine (SVM) algorithm is widely used for data classification. It is because the consistently of providing reliable results in various classification tasks. This can be relied upon for various types of problems and can handle non-linear data effectively using kernel functions like polynomial or Gaussian kernels, allowing SVM to model complex relationships between features and targets. This method also exhibits good efficiency in high-dimensional spaces where the number of features is large, making it suitable for a wide range of applications, including image processing and text recognition.

One of the weaknesses of this method is its difficulty in handling imbalanced data, which may result in a model biased toward the majority class. Therefore, it may require additional methods to balance the class distribution.

## 2.5 Random Forest

Random Forest is one of the machine learning algorithms that can be used for classification and regression tasks [25]. This algorithm is a type of ensemble algorithm, which means it combines the results of several simpler machine learning models called decision trees, where each tree provides a classification estimate (referred to as a vote), and all the votes from each tree are combined to choose the most frequently occurring classification to make stronger and more stable predictions [26]. Decision trees are a crucial component of Random Forest. Here are the steps involved in performing classification using this algorithm [27]:

1. The initial step is to take random samples from the training dataset with replacement. This means that some rows of data can be selected more than once, and some may not be selected at all, creating various datasets for use in building the model.
2. Afterward, the algorithm will build many different decision trees using the datasets sampled in the first step. Decision trees are machine learning models that make decisions based on a set of rules. Each tree will 'learn' to classify or predict data based on rules derived from a subset of the dataset.

3. Next, after several decision trees are built, the algorithm will combine the results from each tree to make the final prediction. In the case of classification, the results from each tree will be calculated, and the prediction with the majority vote will be chosen as the outcome. To determine the majority voting in a tree, the first step is to calculate the entropy value using the equations (14) and (13).

$$\begin{aligned} & \text{entropy}(S) \\ &= \sum_{i=1}^c p_i \log_2 p_i \end{aligned} \tag{13}$$

Note:

S : dataset

c : the number of classes

$p_i$  : the probability frequency of class- $i$  in the dataset

$$\text{entropy}(T, X) = \sum P(c)E(c) \tag{14}$$

Note:

(T, X): T and X feature

$P(c)$  : the probability of feature class

$E(c)$  : entropy result from feature class

In Random Forest, each decision tree is constructed independently with different subsets of data and features, and then their results are combined to create a stronger and more stable model for classification or regression. This is one of the reasons why Random Forest is often used in various machine learning applications. However, on very large datasets, this method can be computationally time-consuming because the algorithm constructs many decision trees, each of which takes time to build.

## 2.6 K-Fold Cross-Validation

K-Fold Cross-Validation is a commonly used evaluation technique in machine learning to assess how well a model can generalize to unseen data [28]. This method divides the dataset into

several subsets (usually K subsets) of equal or nearly equal size and then performs multiple iterations to train and test the model [29].

K-Fold Cross-Validation offers several advantages that make it one of the most useful evaluation techniques in machine learning. One of them is its efficient use of data. By splitting the dataset into multiple subsets and iterating K times, we can ensure that each data point is used for both training and validation. This helps improve the accuracy of model evaluation because the model is assessed on various data points.

## 2.7 Confusion Matrix

Confusion matrix is a tool used for evaluating the performance of a classification model in machine learning. It helps us understand how well our model succeeds or fails in correctly classifying data. With information from the confusion matrix, we can calculate various performance evaluation metrics for the model, such as accuracy, precision, recall, F1-score, and AUC [30]. Table 2 shows for the confusion matrix that is used as a reference for calculations.

**Table 2** Confusion Matrix

		Actual Value	
		True	False
Predicted Value	True	TP (True Positive)	FP (False Positive)
	False	FN (False Negative)	TN (True Negative)

Based on the confusion matrix table, the values of accuracy, precision, recall, F1-score, and AUC can be calculated using the equations below [31].

1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

2. Precision

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (16)$$

3. Recall

$$Recall = \frac{TN}{TN + FN} \times 100\% \quad (17)$$

4. F1 Score

$$F - measure = 2 \times \frac{Recall \times Precision}{Recall + Precision} \times 100\% \quad (18)$$

5. AUC

$$AUC = \frac{1}{2} \times \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (19)$$

Note:

TP : The number of cases where the model correctly predicts the positive class and the actual result is also positive.

TN : The number of cases where the model correctly predicts the negative class and the actual result is also negative.

FN : The number of cases where the model predicts the negative class, but the actual result is positive.

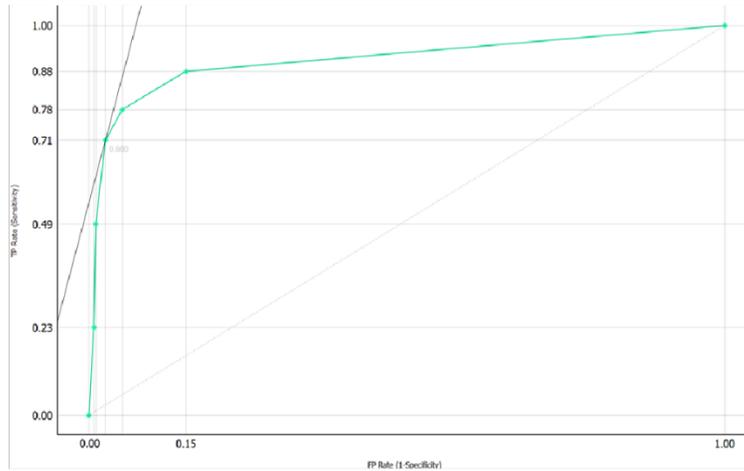
FP : The number of cases where the model predicts the positive class, but the actual result is negative.

### 3. MAIN RESULTS

In most cases, the hardware for the implementation of this research is a laptop running Windows 10 with 64 bits. The definite particulars of the PC include: 1.6-1.8 GHz Intel Core i5 processor with 8 GB of RAM.

The results obtained in the classification process are applied using the K-Nearest Neighbor (K-NN) method with a k value of 5 on the corn stalk image dataset, which is then divided into training and testing data using cross-validation (k-fold) with a k value of 5. So, the accuracy graph of the classification process using the K-NN algorithm is displayed. Figure 3 shows, the K-NN algorithm at 70.8% that informs a good accuracy.

CORN STALK DISEASE CLASSIFICATION USING RANDOM FOREST COMBINATION

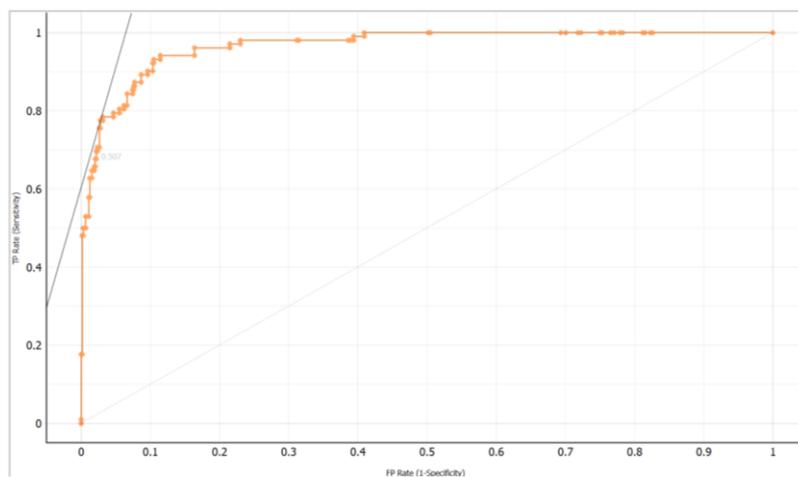


**Fig. 3** Accuracy Result of K-NN

In relevant to building-classification model for corn plant images using the SVM algorithm with specific parameters is shown in Table 3. The accuracy obtained from the corn plant image classification model using the Support Vector Machine method can be seen in Figure 4.

**Table 3** Parameter of SVM

Kernel	RBF (Radial Basis Function)
Coefficient Kernel	Scale ( $1 / (n\_features * X.var())$ )
Cost Value (C)	1.00
Epsilon	0.10
Iteration	100



**Fig. 4** Accuracy Result of SVM

Figure 4 shows the accuracy of the SVM method, it is 58.0%. These results are not very good, which may be due to data imbalance, so additional methods are needed to improve the accuracy.

In this section, model testing is done using the Random Forest algorithm with the following parameter values:

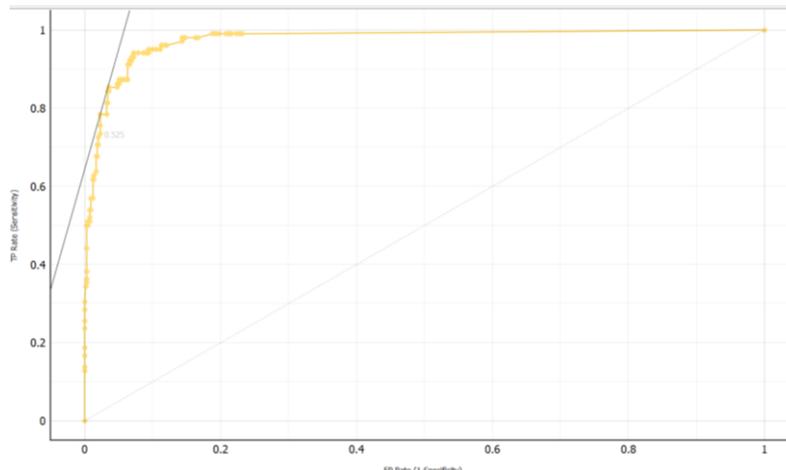
n\_estimator is 10

bootstrap is true

min\_samples\_split is 2, and

max\_depth is 7

By using the four specified parameters, the accuracy result of corn plant classification using the Random Forest method reaches 82.0%, as shown in the accuracy can be seen in Figure 5.



**Fig. 5** Accuracy Result of Random Forest

Figures 3, figure 4, and figure 5 show the accuracy of three proposed algorithms with varying outcomes. The comparison of the image classification model testing results for corn plants, as shown in Table 4.

**Table 4** Evaluation Result

Methods	AUC	Classification Accuracy	F1 Score	Precision	Recall
KNN	88.0 %	70.8 %	70.5 %	71.2 %	70.8 %
SVM	78.9 %	58.0 %	53.8 %	55.7 %	58.0 %
Random Forest	<b>96.2 %</b>	<b>82.0 %</b>	<b>82.0 %</b>	<b>82.0 %</b>	<b>82.0 %</b>

Table 4 shows the different accuracy among of three algorithms. Overall, the Random Forest method provides the best accuracy compared to the other two algorithms, with an accuracy of 82.2%, an AUC of 96.2%, and f1 score, precision, and recall all reaching 82.0%.

#### **4. CONCLUSIONS**

KNN, SVM and Random Forest algorithms can classify the digital image data of corn plant diseases effectively. Random Forest is the best performance with the highest AUC of 96.2% The accuracy, f1 score, precision, and recall are 82.0%.

Based on the conclusion, further research is proposed, including:

- a) The next research is expected to involve the development of the system by incorporating data balancing methods such as SMOTE for imbalanced data.
- b) In addition to HSV and LBP feature extraction, further research could explore the use of other features that may be more relevant or informative for corn stalk disease recognition.
- c) Further research could optimize the algorithm parameter settings, such as Random Forest, to improve classification performance.

#### **ACKNOWLEDGEMENTS**

This study was funded by the University of Trunojoyo Madura in the scheme of Penelitian Mandiri-UTM in 2023 with contract number: 5684/UN46.4.1/ PT.01.03/2023.

#### **CONFLICT OF INTERESTS**

The authors declare that there is no conflict of interests.

#### **REFERENCES**

- [1] C.M. Viana, D. Freire, P. Abrantes, et al. Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review, *Sci. Total Environ.* 806 (2022), 150718.  
<https://doi.org/10.1016/j.scitotenv.2021.150718>.

- [2] S.S. Putro, M.A. Syakur, E.M.S. Rochman, et al. Comparison of backpropagation and ERNN methods in predicting corn production, *Commun. Math. Biol. Neurosci.* 2022 (2022), 10.  
<https://doi.org/10.28919/cmbn/7082>.
- [3] A. Rachmad, M. Fuad, E.M.S. Rochman, Convolutional neural network-based classification model of corn leaf disease, *Math. Model. Eng. Probl.* 10 (2023), 530-536. <https://doi.org/10.18280/mmep.100220>.
- [4] Y.M.A.Y. Bandara, T.T. Tesso, K. Zhang, et al. Charcoal rot and Fusarium stalk rot diseases influence sweet sorghum sugar attributes, *Ind. Crops Products.* 112 (2018), 188-195.  
<https://doi.org/10.1016/j.indcrop.2017.11.012>.
- [5] A. Rachmad, M. Syarief, S. Rifka, et al. Corn leaf disease classification using local binary patterns (LBP) feature extraction, *J. Phys.: Conf. Ser.* 2406 (2022), 012020. <https://doi.org/10.1088/1742-6596/2406/1/012020>.
- [6] A.C. Velásquez, C.D.M. Castroverde, S.Y. He, Plant–pathogen warfare under changing climate conditions, *Curr. Biol.* 28 (2018), R619-R634. <https://doi.org/10.1016/j.cub.2018.03.054>.
- [7] D.S.Y. Kartika, D. Herumurti, A. Yuniarti, Butterfly image classification using color quantization method on HSV color space and local binary pattern, in: *The 3rd International Seminar on Science and Technology, Indonesia, 2017*, 78-82.
- [8] K. Kaplan, Y. Kaya, M. Kuncan, et al. Brain tumor classification using modified local binary patterns (LBP) feature extraction methods, *Med. Hypoth.* 139 (2020), 109696. <https://doi.org/10.1016/j.mehy.2020.109696>.
- [9] S. Singh, K. Malik, Feature selection and classification improvement of Kinnow using SVM classifier, *Measurement: Sensors.* 24 (2022), 100518. <https://doi.org/10.1016/j.measen.2022.100518>.
- [10] S.S. Putro, F. Adiputra, E.M.S. Rochman, et al. Comparison of SAW and WP methods to determine the best agricultural land, *Commun. Math. Biol. Neurosci.* 2021 (2021), 49. <https://doi.org/10.28919/cmbn/5820>.
- [11] K. Darmaastawan, K.O. Saputra, N.M.A.E.D. Wirastuti, Market basket analysis using FP-growth association rule on textile industry, *Int. J. Eng. Emerging Technol.* 5 (2020), 24-30.
- [12] P. Guleria, M. Sood, Data mining in education: A review on the knowledge discovery perspective, *Int. J. Data Mining Knowl. Manage. Process.* 4 (2014), 47-60.
- [13] A. Rachmad, N. Chamidah, et al. Classification of mycobacterium tuberculosis based on color feature extraction using adaptive boosting method, *AIP Conf. Proc.* 2329 (2021), 050005. <https://doi.org/10.1063/5.0042283>.

## CORN STALK DISEASE CLASSIFICATION USING RANDOM FOREST COMBINATION

- [14] N.E.R. Pah, S.A.S. Mola, A.Y. Mauko, Extraction of HSV color and shape characteristics of moment invariant for classification of red apples, *J-Icon*. 9 (2021), 142-153. <https://doi.org/10.35508/jicon.v9i2.5043>.
- [15] W. Zhou, S. Gao, L. Zhang, et al. Histogram of oriented gradients feature extraction from raw Bayer pattern images, *IEEE Trans. Circuits Syst. II*. 67 (2020), 946-950. <https://doi.org/10.1109/tcsii.2020.2980557>.
- [16] B. Kaddar, H. Fizazi, A.O. Boudraa, Texture features based on an efficient local binary pattern descriptor, *Computers Electric. Eng.* 70 (2018), 496-508. <https://doi.org/10.1016/j.compeleceng.2017.08.009>.
- [17] Y. El merabet, Y. Ruichek, A. El idrissi, Attractive-and-repulsive center-symmetric local binary patterns for texture classification, *Eng. Appl. Artif. Intell.* 78 (2019), 158-172. <https://doi.org/10.1016/j.engappai.2018.11.011>.
- [18] S.F. Ershad, Texture classification approach based on combination of edge & co-occurrence and local binary pattern, in: *Int'l Conf. IP, Comp. Vision, and Pattern Recognition*, (2012), 626-629.
- [19] A. Rachmad, N. Chamidah, R. Rulaningtyas. Mycobacterium tuberculosis identification based on colour feature extraction using expert system, *Ann. Biol.* 36 (2020), 196-202.
- [20] B. Gufron, E. Surarso, R. Gernowo, Implementation of the K-nearest neighbor method to determine the classification of the study program operational budget in higher education, in: *Proceeding of the 1st International Conference Health, Science And Technology (ICOHETECH)*, 2019, 201-204.
- [21] N.F. Munazhif, G. Yanris, M.N.S. Hasibuan, Implementation of the K-nearest neighbor (K-NN) method to determine outstanding student classes, *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 8 (2023), 719-732. <https://doi.org/10.33395/sinkron.v8i2.12227>.
- [22] A. Rachmad, N. Chamidah, R. Rulaningtyas, Mycobacterium tuberculosis images classification based on combining of convolutional neural network and support vector machine, *Commun. Math. Biol. Neurosci.* 2020 (2020), 85. <https://doi.org/10.28919/cmbn/5035>.
- [23] R. Rodríguez-Pérez, J. Bajorath, Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery, *J. Computer-Aided Mol. Design.* 36 (2022), 355-362. <https://doi.org/10.1007/s10822-022-00442-9>.
- [24] Z. Liu, H. Xu, Kernel parameter selection for support vector machine classification, *J. Algorithms Comput. Technol.* 8 (2014), 163-177. <https://doi.org/10.1260/1748-3018.8.2.163>.

- [25] S.S. Samy, S. Karthick, M. Ghosal, et al. Adoption of machine learning algorithm for predicting the length of stay of patients (construction workers) during COVID pandemic, *Int. J. Inf. Technol.* 15 (2023), 2613-2621. <https://doi.org/10.1007/s41870-023-01296-6>.
- [26] V.Y. Kulkarni, D.P.K. Sinha, Random forest classifier: A survey and future research directions, *Int. J. Adv. Comput.* 36(2013), 1144-1153.
- [27] G.S. Saragih, Z. Rustam, D. Aldila, et al. Ischemic stroke classification using random forests based on feature extraction of convolutional neural networks, *Int. J. Adv. Sci. Eng. Inform. Technol.* 10(2020), 2177-2182.
- [28] E.M.S. Rochman, Miswanto, H. Suprajitno, et al. Tuberculosis classification using random forest with K-prototype as a method to overcome missing value, *Commun. Math. Biol. Neurosci.* 2023 (2023), 81. <https://doi.org/10.28919/cmbn/7873>.
- [29] M. Yuvalı, B. Yaman, Ö. Tosun, Classification comparison of machine learning algorithms using two independent CAD datasets, *Mathematics.* 10 (2022), 311. <https://doi.org/10.3390/math10030311>.
- [30] A. Rachmad, F. Sonata, J. Hutagalung, et al. An automated system for osteoarthritis severity scoring using residual neural networks, *Math. Model. Eng. Probl.* 10 (2023), 1849-1856. <https://doi.org/10.18280/mmep.100538>.
- [31] E.M.S. Rochman, Miswanto, H. Suprajitno, et al. Utilizing LSTM and K-NN for anatomical localization of tuberculosis: A solution for incomplete data, *Math. Model. Eng. Probl.* 10 (2023), 1114-1124. <https://doi.org/10.18280/mmep.100403>.