Identification of Rice Disease Types Based on Digital Images Leaves Using Algorithm Support Vector Machine (SVM)

K Saputra[1] and Z Zuriati[1]

[1]Informatics Management, Politeknik Negeri Lampung,

 Jl. Soekarno Hatta No.10, Rajabasa Raya, Rajabasa, Bandar Lampung, Lampung

*e-mail: zuriati_mi@polinela.ac.id

Abstract. Diseases that commonly attack rice plant are: bacterial leaf blight, brown spot, and leaf smut. The disease can lead to a significant decrease in the quality and quantity of agricultural products so that it can cause economic loss. Farmers usually find out that the rice plant has been affected by the disease when it already shows severe symptoms and has spread to various parts of the plant so it is too late to control. Another thing that causes the failure of disease management in rice plant is the lack of knowledge of farmers about the symptoms of the disease which causes the provision of inappropriate treatment actions caused by errors in identifying the type of disease that appears with almost the same physical symptoms. Therefore, we need a solution in the form of a model that is able to identify the types of disease in rice plant so that farmers including the general public can detect disease that attack rice plant accurately and quickly. This study aims to develop a model for identification of rice plant disease based on plant leaf images. The research stages are: collecting rice leaf image data, sharing data for training and testing data using cross validation techniques, implementing the SVM algorithm and evaluating the model using a confusion matrix and calculating the accuracy, precision and recall of the SVM algorithm. The results of the evaluation of this identification model indicate that SVM can identify leaf disease of rice plant with an accuracy value of 0.90 or 90%, a precision of 0.902 or 90% and a recall of 0.900 or 90%.

1.   Introduction

Rice (*Oryza sativa. L*) is one of the oldest foodstuffs in the world that is still consumed today, is the main source of carbohydrates for the world community including Indonesia, therefore rice is a very strategic trading commodity. Indonesia ranks as the fourth largest rice producer in the world by producing 35.4 million metric tons per year. The Central Statistics Agency (BPS, 2021) in the Publication of Indonesian Statistics 2021 stated that the rice harvest area in 2021 reached around 10.41 million hectares, a decrease of 245.47 thousand hectares or 2.30 percent compared to the rice harvest area in 2020 which was 10.66 million hectares. Meanwhile, rice production in 2021, which was 54.42 million tons of GKG, decreased by 23.91 thousand tons or 0.43 percent compared to rice production in 2020 which was 54.65 million tons of GKG. Meanwhile, rice production in 2021 for food consumption of the population reached 31.3 million tons, a decrease of 140.73 thousand tons or 0.45 percent compared to rice production in 2020 which was 31.50 million tons.

The Indonesian government is trying to maintain and even increase the amount of national rice production. Various efforts have been made by the government, including through improving the package of cultivation technology and post-harvest handling, increasing the area of rice fields, rehabilitating land and creating new paddy fields, as well as dealing with pests and diseases of rice plants. Diseases that attack rice plants can cause crop failure, so there can be a decrease in the amount of rice production. Plant disease is a condition in which plant cells and tissues cannot function normally, which is caused by the presence of continuous disruption by pathogenic genes (biotics) or environmental (abiotic) factors. Plant disease occurs when one or more of the physiological functions of a plant become abnormal due to the presence of pathogenic disorders or certain environmental conditions. Some types of diseases that affect the leaves of rice plants proposed by the ministry of agriculture in the recommendation guidelines for the control of pests for plant organisms of cereal plants (2013) are: *bacterial leaf blight, brown spot* and *leaf smut.* The disease can result in death and a

significant decrease in the quality and quantity of agricultural products so that it can economically cause losses to farmers. In addition to conducting early identification, farmers must know how to prevent and treat each disease that attacks rice plants by knowing the factors or symptoms that arise as early as possible. Farmers generally know the disease that attacks rice leaves when they have shown severe symptoms and have spread to various areas so that they are too late to be controlled, causing crop failure. Another thing that causes the failure to deal with rice diseases is that rice diseases generally have almost the same physical symptoms. Therefore, a step is needed for early identification of diseases that attack rice plants so that preventive measures can be taken [1].

One way that can be used to help the initial identification process of diseases in rice plants is to use data mining algorithms through design and build or model engineering that can help identify diseases in rice plants in detail and thoroughly. This can be done by utilizing and processing the leaf image of the rice plant affected by the disease, then comparing the leaf image with the condition of the rice plant which will be identified as the type of disease. The shape of the spotted spots, color, and texture are important parameters that are commonly used to identify the type of disease in rice leaves [2],[3]. Leaves are the part of rice plants that are often attacked by diseases, but leaves are also the most easily observed part of rice plants to identify the type of disease in rice [4]. Therefore, through the right method and by observing the image on the leaves, the engineering model that is built can group or identify the type of rice plant disease.

The research objective is to implement one of the data mining algorithms namely support vector machine (SVM) to identify types of rice plant diseases based on rice leaf images. This study carried out pattern mapping of leaf images affected by disease, mapping was carried out for three (3) categories or types of rice plant diseases, namely: *bacterial leaf blight, brown spot* and *leaf smut*. The SVM algorithm has been proven to be able to recognize and identify images well and with a high degree of accuracy [5],[6]. Several studies have applied the SVM algorithm, including: for the identification of rice quality [4], prediction of student graduation [7],[8], promotion strategy [9], classification of child growth and development [10], and classification of film opinion data [11]. The benefit of research is as an effort to early detect diseases in rice plants appropriately and quickly, so that farmers or the community can carry out treatment to overcome diseases in rice plants. The stages of the study are: data collection of rice plant leaf image data, data sharing for training and testing data with *cross validation* techniques, implementation of the SVM algorithm and model evaluation using a confusion matrix and calculating the accuracy of the SVM algorithm.

Research on diseases in rice plants has been widely carried out by applying various data mining algorithms, including: a combination of *certainly factor* and *forward chaining* algorithms [12], *case-based reasoning* (CBR) algorithms [13], *adaptive neuro fuzzy inference system* (ANFIS) algorithms with an accuracy of 98.5% [14]. Meanwhile [3] it uses several *machine learning algorithms*, namely: *k-nearest neighbor* (KNN), *decision tree* J48, Naïve Bayes, and regression logistic. Meanwhile, [15] classified rice diseases based on leaf imagery using the Resnet101 trained model with 100% accuracy performance and a loss value of 5.61%. The study using the *algorithm convolutional neural network (CNN)* was carried out by [16], [17].

Previous researchers who have applied the SVM algorithm for various cases, including: for the identification of rice quality [4], prediction of student graduation [7], [8], promotion strategy [9], classification of child growth and development [10], and classification of film opinion data [11]. Research applying the SVM method for the detection of rice disease is [18].

2. Methods

The research stage carried out starts from data collection, then applies *cross validation* techniques for determining training data and testing data used for the implementation of the SVM algorithm. The model evaluation technique uses a confusion matrix to then calculate the accuracy of the SVM algorithm. In Figure 1 is presented a chart of the stages of the study.
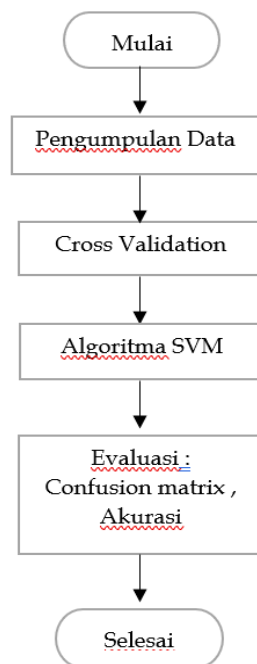
Figure 1. Stages of research

The stages of the research for the development of this monitoring system are:

3.1 Data collection

The initial stage of the study is the data collection stage. The data used is in the form of digital image data of rice plant leaves. The leaf data collected consists of three types of diseases, namely *bacterial leaf blight, brown spot,* and *leaf smut.* For each type of disease collected 40 pieces of imagery, so the total data amounted to 120 data.

3.2. Cross validation

Cross validation is a common method used to evaluate *classifier* performance. In the cross validation approach, each record is used several times in the same amount for training data and for testing data. This method partitions the data into two subsets of the same size. Select one as the training data and the other for the testing data, then the exchange of functions from the subset is carried out so that the subset that was previously the training data becomes the testing data and vice versa. *K-fold cross-validation* generalizes this approach by segmenting data into k-sized partitions. During the process, one of the partitions is selected for training, while the rest is for testing. This procedure is repeated k times in such a way that each partition is used for testing exactly once. *K-fold cross-validation is* used in order to find the best parameters of one model. For the use of the best number of folds for validity tests, it is recommended to use *10-fold cross validation* in the model, since 10 folds is the right amount to get the best estimate [13].

3.3 SVM algorithm

The next stage is to implement the SVM algorithm using the Orange tool version 3.3. SVM is a data mining method that is included in the supervised learning class. The main goal of this technique is to search for the best hyperplane of several existing hyperplanes. The SVM algorithm seeks to obtain the optimal separator function to be able to separate two sets of data from two different classes. In the SVM technique, the best hyperplane is a hyperplane located in the middle between two data sets of two different classes. Looking for the best hyperplane is to maximize margins. Margin is the perpendicular distance between the hyperplane and the nearest object. The nearest object is also called

the support vector. Hyperplane searches will be performed in multidimensional workspaces to maximize the distance between classes [19].

3.4. Evaluation

How to measure the performance of the system is done using a confusion matrix using a table that records the results of the SVM algorithm identification work. *The confusion matrix* contains information that compares the results of the classification carried out by the system with the results of the supposed classification. Some of the studies that apply *confusion matrix* for evaluation are [6],[8],[20]. The matric confusion consists of True positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). The TP (True Positive) value is obtained from the number of predictions that are true and the number of correct original values for each class tested. The FN (False Negative) value is obtained from the number of false predictions and the sum of the correct native values for each class tested. The FP (False Positive) value is obtained from the number of predictions that are true and the number of false original values for each class tested. The TN (True Negative) value is obtained from the number of incorrect predictions and the number of false native values for each class tested. The confusion matrix table is presented in Table 1 below:

Table 1. Confusion matrix

Label Current	Label Predictive	
	1	2
1	True Positive	False Negative
2	False Positive	True Positive

Accuracy is defined as the degree of correlation between the predicted value and the actual value, so that it can be used to calculate the accuracy of the system to classify data correctly. Precision is the level of accuracy between the data expected by the user and the answers given by the system, or the accuracy of the truth value in the classification according to the actual truth value. Meanwhile, recall is the level of success of the system in calculating how much the truth value (positive) of the dataset that is indeed of true value (positive) appears. Here are the mathematical equations for calculating accuracy, precision, and recall (nest).

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

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

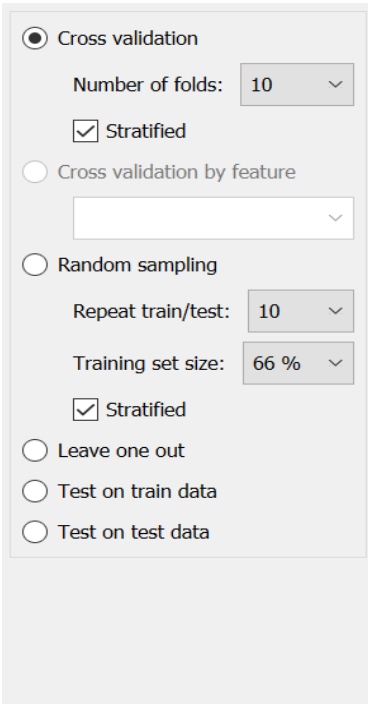$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

3. Results and Discussion

The data used is digital image data of rice plant leaves collected from the website: https://www.kaggle.com. There are 3 groups of image data: *bacterial leaf blight, brown spot,* and *leaf smut* with a total of 40 images each, so that the total data is 120 images. Table 2 presents an example of a digital image of the leaves of the rice plant used for this study.

Table 2. Digital image of rice plant leaves

Types of Disease	Sample Imagery Data	Amount of Data

Bacterial Leaf Blight		40
Brown Spot		40
Leaf Smut		40

After the data is collected, data sharing is carried out using cross validation techniques. Figure 2 shows the details of using *10-folds cross validation.*

Figure 2. *10-folds cross validation*

The SVM algorithm applied is linear kernel. In Figure 3 is presented the selection of parameters used for the SVM algorithm. The parameters used in the experiment are Cost (C) = 1.00, Regression loss epsilon (ε) = 0.10. For optimization, a repeat of 100 times is carried out, with a numerical tolerance value of 0.001.
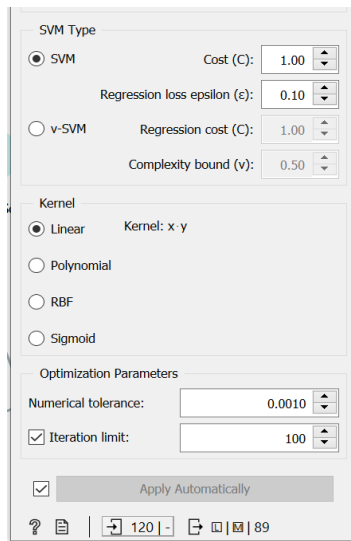
Figure 3. Parameters of the SVM algorithm

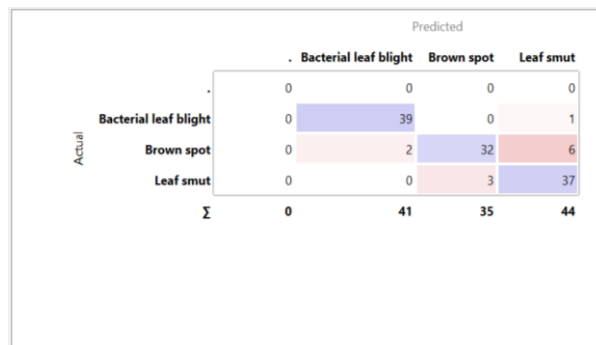The results of the confusion matrix obtained are presented in Figure 4.

Figure 4. Confusion matrix

Based on the confusion matrix in Figure 4, the values of accuracy (CA), precision, and recall can be calculated. The calculation results are presented in the following Figure 5.
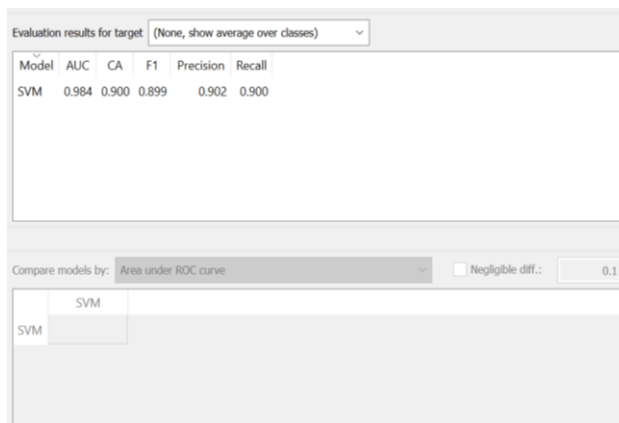

Figure 5. Algorithm evaluation

In Figure 5, the accuracy calculation result is 0.900 or 90%, the precision calculation result is 0.902 or 90%, and the recall calculation result is 0.900 or 90%.

4. Conclusions

The SVM algorithm is proven to be able to classify rice plant diseases with an accuracy of 0.900 or 90%, precision is 0.902 or 90%, and recall is 0.900 or 90%.

5. Acknowledgments

6. References

[1] J. Kusanti, K. Penyakit, D. Padi, and A. Haris, "Klasifikasi Penyakit Daun Padi Berdasarkan Hasil Ekstraksi Fitur GLCM Interval 4 Sudut," *J. Inform. J. Pengemb. IT*, vol. 03, no. 01, pp. 1–6, 2018.

[2] Y. Zhou, Y. Wang, and Q. Yao, "Segmentation of rice disease spots based on improved BPNN," *IASP 10 - 2010 Int. Conf. Image Anal. Signal Process.*, pp. 575–578, 2010.

[3] K. Ahmed, T. R. Shahidi, S. M. Irfanul Alam, and S. Momen, "Rice leaf disease detection using machine learning techniques," *2019 Int. Conf. Sustain. Technol. Ind. 4.0, STI 2019*, no. May 2020, pp. 1–5, 2019.

[4] S. Saidah, M. B. Adinegara, R. Magdalena, and N. K. Caecar, "Identifikasi Kualitas Beras Menggunakan Metode k-Nearest Neighbor dan Support Vector Machine," *TELKA - Telekomun. Elektron. Komputasi dan Kontrol*, vol. 5, no. 2, pp. 114–121, 2019.

[5] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, 2019.

[6] S. Hartono, A. Perwitasari, and H. Sujaini, "Komparasi Algoritma Nonparametrik untuk Klasifikasi Citra Wajah Berdasarkan Suku di Indonesia," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 337, 2020

[7] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, 2021.

[8] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021.

[9] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi," *JUITA J. Inform.*, vol. 7, no. 2, p. 71, 2019.

[10] I. Monika Parapat and M. Tanzil Furqon, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3163–3169, 2018.

[11] Styawati, Andi Nurkholis, Zaenal Abidin, and Heni Sulistiani, "Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 904–910, 2021.

[12] M. Irsan, V. N. Pratama, and M. Fakih, "Sistem Pakar Identifikasi Penyakit Tanaman Padi Di Balai Penyuluhan Pertanian Sepatan Tangerang," *Konf. Nas. Sist. Inform.*, pp. 446–451, 2015.

[13] Minarni and I. Warman, "Sistem Pakar Identifikasi Penyakit Tanaman Padi Menggunakan Case-Based Reasoning," *Semin. Nas. Apl. Teknol. Inf.*, pp. 28–32, 2017.

[14] R. N. Whidhiasih and I. Ekawati, "Identifikasi Jenis Penyakit Daun Padi Menggunakan Adaptif Neuro Fuzzy Inferene System (Anfis) Berdasarkan Tekstur," *Pros. Semin. Nas. Energi Teknol.*, pp. 131–140, 2019.

[15] Ulfah Nur Oktaviana, Ricky Hendrawan, Alfian Dwi Khoirul Annas, and Galih Wasis Wicaksono, "Klasifikasi Penyakit Padi berdasarkan Citra Daun Menggunakan Model Terlatih Resnet101," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1216–1222, 2021.

[16] A. Sony, "Prediction of Rice Diseases Using Convolutional Neural Network (in Rstudio)," *Int. J. Innov. Sci. Res. Technol.*, vol. 4, no. 12, pp. 595–602, 2019.

[17]   fan fan, T. Roy, and K. Roy, "Classification and Detection Rice leaf Diseases Using Information and Communication Technology (ICT) Tools," *Int. J. Adv. Eng. Res. Sci.*, vol. 7, no. 6, pp. 460–470, 2020.

[18] A. Purnamawati, W. Nugroho, D. Putri, and W. F. Hidayat, "InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan Attribution-NonCommercial 4.0 International. Some rights reserved Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN," vol. 5, no. 1, 2020.

[19] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.

[20] U. Riyanto, "Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online," *JIKA (Jurnal Inform.*, vol. 2, no. 2, pp. 62–72, 2019.