Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12195 p-ISSN : 2541-044X

Comparison of Tomato Leaf Disease Classification Accuracy Using Support Vector Machine and K-Nearest Neighbor Methods

P.P.P.A.N.W. Fikrul Ilmi R.H. Zer[1)*], Fazli Nugraha Tambunan[2)], Rika Rosnelly[3)], Wanayumini[4)]
[1)2)3)4)]Potensi Utama Univercity, Medan, Indonesia
[1)]fikrulilmizer@gmail.com, [2)]fazlinugraha313@gmail.com, [3)]rikarosnelly@gmail.com, [4)]wanayumini@gmail.com

Submitted : Feb 9, 2023 | **Accepted** : Mar 10, 23 | **Published** : Apr 7, 2023

Abstract: Tomato Leaf Disease is one of the common things for farmers in growing tomatoes. Tomatoes are one of the popular crops that can grow in low and high areas but are susceptible to disease. For this reason, farmers take precautions by looking at the characteristics and texture of tomato leaves. However, this requires more time and money and a long process. One of the efforts that can be made is to classify tomato leaf diseases. This research aims to classify using the Support Vector Machine and K-Nearest Neighbor methods. The dataset used is tomato leaf image data with 4 classes of leaves affected by disease and 1 healthy leaf. We evaluate and analyze all models using 5-Fold, 10-Fold, and 20-Fold Cross Validation with accuracy, precision, and recall for the best accuracy. The best results of this study are accuracy in the SVM method of 0.953 or 95.3%, Precision of 0.953 or 95.3%, and Recall of 0.953 or 95.3% with 10-Fold Cross-Validation. Compared to the K-NN method, it only obtained an accuracy of 0.907 or 90.7%, a Precision of 0.908 or 90.8%, and a Recall of 0.907 or 90.7% with 10-Fold Cross-Validation.
Keywords: Support Vector Machine, K-Nearest Neighbor, Tomato, Tomato Leaf Disease, Classification

INTRODUCTION

Tomatoes are one type of plant that produces healthy fruit. Tomatoes or Solanum Lycopersicum grow in various mediums and land elevations (Gemilang & Lubis, 2022). Tomato plants are widely consumed by the community and are one of the vegetable commodities that have increased from year to year. Tomato plants are vulnerable when attacked by disease, coupled with a lack of care that causes the quality of tomatoes to be poor (Khultsum & Subekti, 2021). Disease attacks on tomato plants can be recognized visually through the leaves of the plant because they have unique texture and color characteristics. Disease attacks on tomatoes are caused by fungi, bacteria, viruses, pests, and insects. Tomato plants that are attacked by disease experience changes in the shape of the characteristics and texture of the leaves of tomato plants (Putri, 2021). When farmers detect diseases on tomato leaves, they have leaf shape characteristics that are similar to the plant disease. So farmers are sometimes mistaken in using drugs when controlling disease in tomato plants.

Along with technological developments, one of them in agriculture is trying to develop techniques for overcoming problems in plants. In overcoming problems in tomato plants, one of detecting leaf disease in tomatoes is by using classification on tomato leaf images (Ashok et al., 2020). Pattern recognition or characteristics of the characteristics and texture of tomato leaves is an important step in

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12195 p-ISSN : 2541-044X

identifying tomato plant diseases correctly and efficiently. With a technique that can recognize tomato leaf disease, control efforts can be made so that tomato leaf disease can be overcome.

Classification techniques now use sophisticated technology to find solutions to every problem, namely with Artificial Intelligence (AI). Classification methods that will be used are Support Vector Machine and K-Nearest Neighbor methods. The Support Vector Machine method is a method that works by identifying the boundary between two classes with the maximum distance from the closest data. This method is attractive for analyzing gene expression, able to handle large data sets and feature spaces (Rizal et al., 2019). The K-Nearest Neighbor method is the oldest and most popular NN-based method (Prahudaya & Harjoko, 2017). This method looks for groups of k objects in the training data that are closest (similar) to objects in the testing data used (Tangguh Admojo & Ahsanawati, 2020).

This research compares the Support Vector Machine and K-Nearest Neighbor methods by finding the best accuracy value (Fawzy et al., 2020). We will compare the accuracy values of the two methods to conclude the best method for classifying tomato leaf diseases. Previously, this research was based on previous research with the same background, namely tomato leaf disease. Previous research used the Convolutional Neural Network method with a resulting accuracy of 98% with a dataset of 10 categories of tomato leaf disease (Ashok et al., 2020). So we conducted research with different methods to get comparative results.

LITERATURE REVIEW

Previous related research with the object under study and the method used. One of the studies related to the Support Vector Machine (SVM) method is face classification research using the Support Vector Machine method. The dataset used is a face image with a tilt angle of the subject's face with an image size of 640 x 480 pixels. The number of samples used is 200 samples with 60% as training data and 40% as testing data. The results of this study obtained very good accuracy with an average true detection of 90% and false detection of 10% in classifying faces (Rizal et al., 2019).

In contrast to previous research that used the Support Vector Machine method in classifying pneumonia. Where the study used a dataset of 5853 lung X-ray images divided into 2 types of X-ray images, namely normal lungs and pneumonia lungs. The classification method used in the study was Support Vector Machine (SVM) and Gray Level Co-Occurrence (GLCM) for the extraction method. The results of the study showed that the best accuracy obtained was 62.66% (Wati et al., 2020).

This is also different from previous research that uses the K-Nearest Neighbor method in classifying guava quality using the K-Nearest Neighbor method based on color and texture features. The research dataset uses 80 guava sample images with images taken from two sides of the guava. The classification is divided into 4 quality classes, namely superclass, class A, class B, and outside quality. From the test results obtained classification with the K-Nearest Neighbor method obtained the best accuracy at k = 3 with an accuracy of 91.25% (Prahudaya & Harjoko, 2017).

And there are other studies that use a comparison of the Support Vector Machine and K-Nearest Neighbor methods in classifying breast cancer abnormalities. The research dataset uses Mammography Image Analysis Society (MIAS) data. The results showed that the accuracy of the SVM method was better than the K-Nearest Neighbor method with an accuracy of 93.88% (Harefa & Pratiwi, 2016). And there are also other studies in the comparison of Support Vector Machine and K-Nearest Neighbor methods, namely in classifying liver disease. The research dataset was 583 sample data with 20% of the data as testing data and 80% of the data as training data. The results of this study obtained an accuracy of 82.90% for the SVM method and 72.46% for the K-NN method (Assegie, 2021).

There are other related studies that use the object of tomato leaf disease but use methods that are different from our research. The study used the Convolutional Neural Network method with a dataset of 10,000 training data with 50 images from each class of 10 classes with 1 class of healthy tomato leaves. The research results were obtained with an average accuracy of 91.2% (Agarwal et al., 2020). Based on previous related research, we conducted tests to compare the accuracy results using the Support Vector Machine and K-Nearest Neighbor methods with tomato leaf disease data. We used these two methods because they can classify images with good accuracy.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12195
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

In this research, we use a dataset obtained from the website www.kaggle.com with data on tomato leaf diseases. We use image data of tomato leaf diseases with 5 types, namely early blight, healthy, late blight, mosaic virus, and yellow leaf curl virus. This research was conducted using Support Vector Machine and K-Nearest Neighbor methods. The following is an example of a sample image of tomato leaf disease in Figure 1:

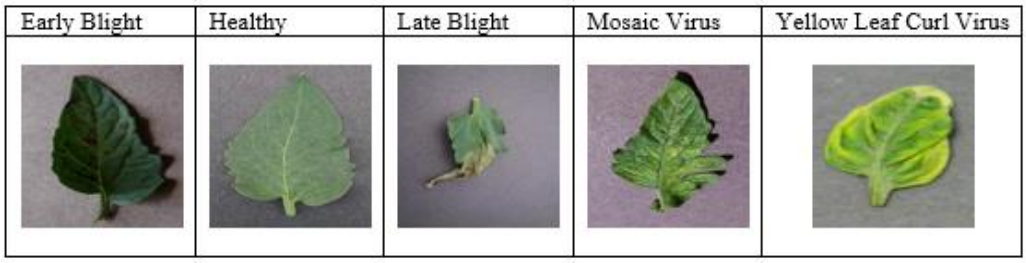

Figure 1. Sample data of tomato leaf diseases

The tools we use in this research are Tools Orange for testing classification data using the Support Vector Machine and K-Nearest Neighbor methods. The stages of this research can be seen in Figure 2:

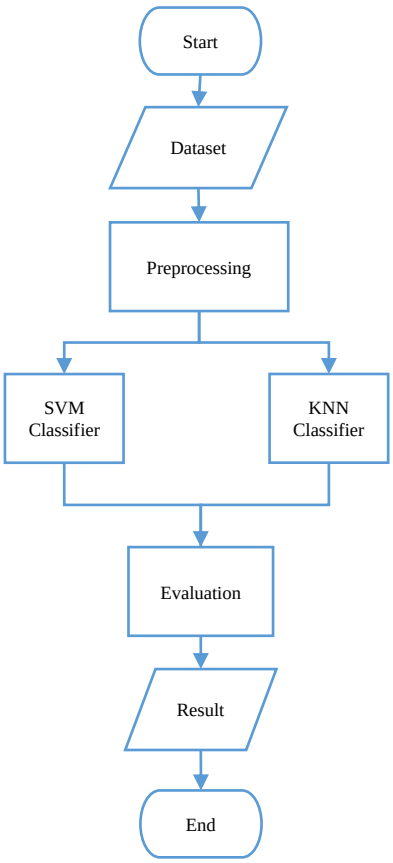

Figure 2. Research Flow with Flowchart

We can see Figure 2, explains the flow of research conducted. Processing stages in the figure explain the process of preparing data processing by dividing data into training data and testing data. Then from the data that has been prepared, classification is carried out using the Support Vector Machine and K-Nearest Neighbor methods. After that, evaluation is carried out by testing based on 5-Fold, 10-Fold, and 20-Fold Cross Validation and getting accuracy. The last stage is the results using both methods, with the best accuracy and validation.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12195
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Support Vector Machine is a classification method in the form of a feature vector to obtain testing predictions (Neneng et al., 2021). The mapped vectors will be calculated where the furthest distance will be used as a hyperplane as a class separator. The Hyperlane equation is a classification equation that can be seen in equation (1) with the classification parameters w and b as the weight and bias values in equations (2) and (3).

$$f_{svm}(x) = w.x + b \quad (1)$$

$$w = \sum_{i=1}^{n} a_i y_i x_i \quad (2)$$

$$b = -\frac{1}{2}(w.x^{+} + w.x^{-}) \quad (3)$$

K-Nearest Neighbor is a method that classifies objects based on training data that is closest to the object. The best k value in this method depends on the data where a high k value will reduce the effect of noise on the classification problem. For classification problems, it is used based on the closest training data or what is called the nearest neighbor. The Euclidean Distance method is the most widely used distance calculation formula for researchers in the K-Nearest Neighbor method.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

RESULT

The steps in this research using the Support Vector Machine and K-Nearest Neighbor methods can be seen as follows:

1. Training process, where this stage will be carried out training data that has been prepared using Support Vector Machine and K-Nearest Neighbor.
2. Testing Process, where this stage will be tested from data that has previously been trained. The prepared testing data will be tested for classification.
3. Evaluation, where this stage will display a comparison of the accuracy of the two methods used by changing Cross-Validation.

This research uses 5000 data for 5 types of tomato leaf characteristics and shapes. With the division of each type of tomato leaf totaling 1000 images. The training data used for each type is 800 data and the testing data used for each type is 200 data so the total training data is 4000 data and the testing data is 1000 data. Here are sample images for each type of tomato leaf shape:

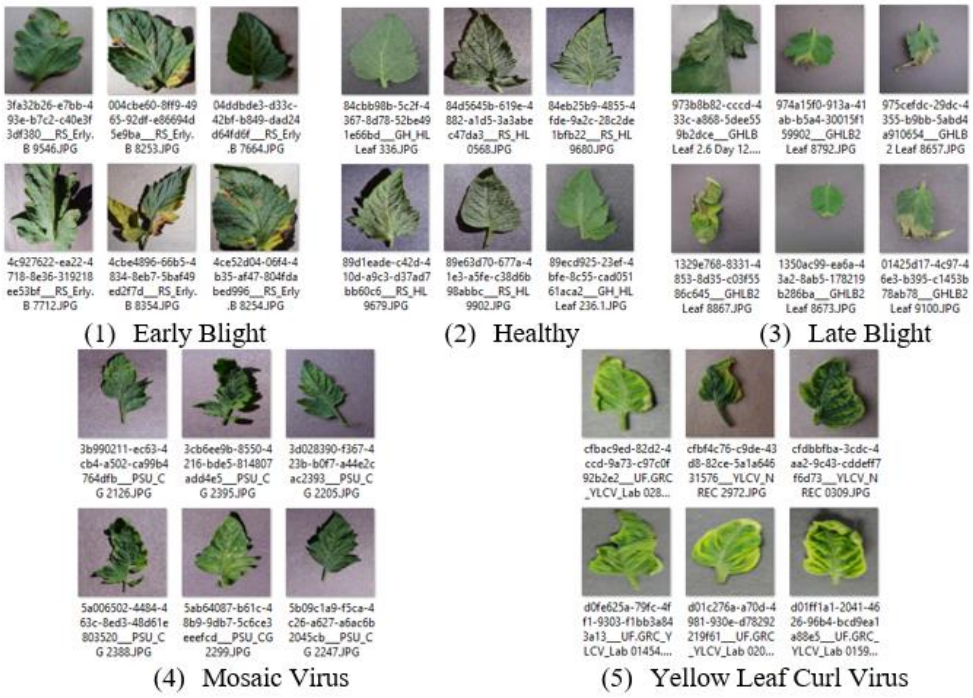

Figure 3. Image Samples and image division

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12195
e-ISSN : 2541-2019
p-ISSN : 2541-044X

This research conducts a comparison of tests to provide the most accurate and effective results to be applied. We use 3 types of Cross Validation to get accurate results, consisting of 5-Fold, 10-Fold, and 20-Fold Cross-Validation.

a. 5-Fold Cross Validation

The results of testing with 5-Fold using the Support Vector Machine and K-Nearest Neighbor methods to classify tomato leaf diseases can be seen as follows:

Table 1. K-NN Confusion Matrix Results with 5-Fold

		Predicted					
Actual		Early Blight	Late Blight	Yellow Leaf Curl Virus	Mosaic Virus	Healthy	Σ
	Early Blight	851	56	12	39	42	1000
	Late Blight	157	775	9	26	33	1000
	Yellow Leaf Curl Virus	22	7	941	20	10	1000
	Mosaic Virus	9	5	1	980	5	1000
	Healthy	7	3	2	12	976	1000
	Σ	1046	846	965	1077	1066	5000

Table 2. SVM Confusion Matrix Results with 5-Fold

		Predicted					
Actual		Early Blight	Late Blight	Yellow Leaf Curl Virus	Mosaic Virus	Healthy	Σ
	Early Blight	866	108	6	11	9	1000
	Late Blight	152	842	1	2	3	1000
	Yellow Leaf Curl Virus	15	3	979	2	1	1000
	Mosaic Virus	3	0	0	997	0	1000
	Healthy	5	2	0	1	992	1000
	Σ	1041	955	986	1013	1005	5000

Based on Table 1 and Table 2, the results of the Confusion Matrix of the Support Vector Machine and K-Nearest Neighbor methods obtained classification accuracy, Precision, and Recall which can be seen in Table 3.

Table 3. Classification Results with 5-Fold

Method	AUC	CA	F1	Precision	Recall
kNN	0,982	0,905	0,904	0,906	0,905
SVM	0,994	0,935	0,935	0,936	0,935

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12195
e-ISSN : 2541-2019
p-ISSN : 2541-044X

b. 10-Fold Cross Validation
The results of testing with 10-Fold using the Support Vector Machine and K-Nearest Neighbor methods to classify tomato leaf diseases can be seen as follows:

Table 4. K-NN Confusion Matrix Results with 10-Fold

		Predicted					
Actual		Early Blight	Late Blight	Yellow Leaf Curl Virus	Mosaic Virus	Healthy	Σ
	Early Blight	853	56	10	37	44	1000
	Late Blight	154	783	8	25	30	1000
	Yellow Leaf Curl Virus	20	8	942	19	11	1000
	Mosaic Virus	11	5	1	980	3	1000
	Healthy	8	4	2	9	977	1000
	Σ	1046	856	963	1070	1065	5000

Table 5. SVM Confusion Matrix Results with 10-Fold

		Predicted					
Actual		Early Blight	Late Blight	Yellow Leaf Curl Virus	Mosaic Virus	Healthy	Σ
	Early Blight	907	69	5	11	8	1000
	Late Blight	106	885	2	3	4	1000
	Yellow Leaf Curl Virus	11	3	983	2	1	1000
	Mosaic Virus	3	1	0	996	0	1000
	Healthy	4	2	0	2	992	1000
	Σ	1031	960	990	1014	1005	5000

Based on Table 4 and Table 5 the results of the Confusion Matrix of the Support Vector Machine and K-Nearest Neighbor methods obtained classification accuracy, Precision, and Recall which can be seen in Table 6.

Table 6. Classification Results with 5-Fold

Method	AUC	CA	F1	Precision	Recall
kNN	0,984	0,907	0,906	0,908	0,907
SVM	0,996	0,953	0,953	0,953	0,953

c. 20-Fold Cross Validation
The results of testing with 20-Fold using the Support Vector Machine and K-Nearest Neighbor methods to classify tomato leaf diseases can be seen as follows:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12195
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 7. K-NN Confusion Matrix Results with 20-Fold

		Predicted					
Actual		Early Blight	Late Blight	Yellow Leaf Curl Virus	Mosaic Virus	Healthy	Σ
	Early Blight	856	55	8	37	44	1000
	Late Blight	152	789	8	24	27	1000
	Yellow Leaf Curl Virus	20	6	944	19	11	1000
	Mosaic Virus	10	5	1	982	2	1000
	Healthy	8	4	2	10	976	1000
	Σ	1046	859	963	1072	1060	5000

Table 8. SVM Confusion Matrix Results with 20-Fold

		Predicted					
Actual		Early Blight	Late Blight	Yellow Leaf Curl Virus	Mosaic Virus	Healthy	Σ
	Early Blight	888	90	6	10	6	1000
	Late Blight	145	846	2	4	3	1000
	Yellow Leaf Curl Virus	12	3	980	4	1	1000
	Mosaic Virus	2	1	0	997	0	1000
	Healthy	3	1	0	2	994	1000
	Σ	1050	941	988	1017	1004	5000

Based on Table 7 and Table 8 the results of the Confusion Matrix of the Support Vector Machine and K-Nearest Neighbor methods obtained classification accuracy, Precision, and Recall which can be seen in Table 9.

Table 9. Classification Results with 20-Fold

Method	AUC	CA	F1	Precision	Recall
kNN	0,984	0,909	0,909	0,911	0,909
SVM	0,995	0,941	0,941	0,941	0,941

Based on the results of the 5-Fold, 10-Fold, and 20-Fold Cross Validation described earlier, we display a classification comparison table of the results of all tests which can be seen in table 10 below.

Table 10. Classification Comparison Results

K-Fold	KNN			SVM		
	CA	PRECISION	RECALL	CA	PRECISION	RECALL
5 fold	0,905	0,906	0,905	0,935	0,936	0,935
10 fold	0,907	0,908	0,907	0,953	0,953	0,953
20 fold	0,909	0,911	0,909	0,941	0,941	0,941

Table 10 explains the comparison results of classification testing using the Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) methods to conclude which method has good accuracy.

*name of corresponding author

DISCUSSIONS

Based on the results of testing the classification of tomato leaf disease datasets in table 10, the classification results obtained by the Support Vector Machine (SVM) method are better than the K-Nearest Neighbor (K-NN) method. This is evidenced by the results of Classification Accuracy both with 5-Fold, 10-Fold, and 20-Fold Cross Validation testing. For 5-Fold Cross Validation testing, SVM gets an accuracy of 0.935 or 93.5%, while K-NN gets an accuracy of 0.905 or 90.5%. For 10-Fold testing, SVM gets an accuracy of 0.953 or 95.3%, while K-NN gets an accuracy of 0.907 or 90.7%. For 20-Fold testing, SVM gets an accuracy of 0.941 or 94.1%, while K-NN gets an accuracy of 0.909 or 90.9%. For further research, it is recommended to add types of tomato leaf diseases in order to get better classification results and accuracy.

CONCLUSION

The conclusion of this research is that the Support Vector Machine and K-Nearest Neighbor methods can classify tomato leaf diseases. The results of the Support Vector Machine method are better than K-Nearest Neighbor, this is evidenced by the 10-Fold Support Vector Machine test being able to produce accuracy in the SVM method of 0.953 or 95.3%, Precision of 0.953 or 95.3%, and Recall of 0.953 or 95.3% with 10-Fold Cross-Validation. Compared to the K-NN method, it only obtained an accuracy of 0.907 or 90.7%, a Precision of 0.908 or 90.8%, and a Recall of 0.907 or 90.7% with 10-Fold Cross-Validation.

REFERENCES

Agarwal, M., Singh, A., Arjaria, S., Sinha, A., & Gupta, S. (2020). ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network. *ICCIDS 2019*, *2019*, 293–301. https://doi.org/10.1016/j.procs.2020.03.225

Ashok, S., Kishore, G., Rajesh, V., Suchitra, S., Gino Sophia, S. G., & Pavithra, B. (2020). Tomato Leaf Disease Detection Using Deep Learning Techniques. *ICCES*, *Icces*, 979–983. https://doi.org/10.1109/ICCES48766.2020.09137986

Assegie, T. A. (2021). Support Vector Machine And K-Nearest Neighbor Based Liver Disease Classification Model. *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *3*(1), 9–14. https://doi.org/10.35882/ijeeemi.v3i1.2

Fawzy, H., Rady, E. H. A., & Fattah, A. M. A. (2020). Comparison Between Support Vector Machines And K-Nearest Neighbor For Time Series Forecasting. *Journal of Mathematical and Computational Science*, *10*(6), 2342–2359. https://doi.org/10.28919/jmcs/4884

Gemilang, E. P., & Lubis, C. (2022). Klasifikasi Jenis Penyakit Pada Daun Tomat Dengan Menggunakan Convolutional Neural Network. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *10*(1). https://doi.org/10.24912/jiksi.v10i1.17839

Harefa, J., & Pratiwi, M. (2016). Comparison Classifier: Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) In Digital Mammogram Images. *Juisi*, *02*(02), 35–40. http://peipa.essex.ac.uk/pix/mias/.

Khultsum, U., & Subekti, A. (2021). Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat. *Jurnal Media Informatika Budidarma*, *5*(1), 186. https://doi.org/10.30865/mib.v5i1.2624

Neneng, N., Putri, N. U., & Susanto, E. R. (2021). Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern. *Cybernetics*, *4*(02), 93–100. https://doi.org/10.29406/cbn.v4i02.2324

Prahudaya, T. Y., & Harjoko, A. (2017). Metode Klasifikasi Mutu Jambu Biji Menggunakan Knn Berdasarkan Fitur Warna Dan Tekstur. *Jurnal Teknosains*, *6*(2), 113. https://doi.org/10.22146/teknosains.26972

Putri, A. W. (2021). Implementasi Artificial Neural Network (ANN) Backpropagation Untuk Klasifikasi Jenis Penyakit Pada Daun Tanaman Tomat. *MATHunesa: Jurnal Ilmiah Matematika*, *9*(2), 344–350. https://doi.org/10.26740/mathunesa.v9n2.p344-350

Rizal, R. A., Girsang, I. S., & Prasetiyo, S. A. (2019). Klasifikasi Wajah Menggunakan Support Vector Machine (SVM). *REMIK (Riset Dan E-Jurnal Manajemen Informatika Komputer)*, *3*(2), 1.

*name of corresponding author

https://doi.org/10.33395/remik.v3i2.10080
Tangguh Admojo, F., & Ahsanawati, A. (2020). Klasifikasi Aroma Alkohol Menggunakan Metode KNN. *IJODAS*, *1*(2), 34–38.
Wati, R. A., Irsyad, H., & Rivan, M. E. Al. (2020). Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine. *Jurnal Algoritme*, *1*(1), 21–32.

*name of corresponding author