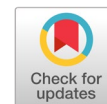


Broccoli leaf diseases classification using support vector machine with particle swarm optimization based on feature selection



Yulio Ferdinand^{a,1}, Wikky Fawwaz Al Maki^{a,2,*}

^a Telkom University, Jl. Telekomunikasi No. 1, Bandung 40257, Indonesia

¹ albertusyulio@student.telkomuniversity.ac.id; ² wikkyfawwaz@telkomuniversity.ac.id

* corresponding author

ARTICLE INFO

Article history

Selected paper from The 2022 5th International Symposium on Advanced Intelligent Informatics (SAIN'22), Yogyakarta (Virtually), September 14, 2022, <http://sain.ijain.org/2022/>. Peer-reviewed by SAIN'22 Scientific Committee and Editorial Team of IJAIN journal

Received July 20, 2022

Revised September 4, 2022

Accepted November 30, 2022

Available online November 30, 2022

Keywords

Broccoli leaf
Classification
Image processing
SVM

ABSTRACT

Broccoli is a plant that has many benefits. The flower parts of broccoli contain protein, calcium, vitamin A, vitamin C, and many more. However, in its cultivation, broccoli plants have obstacles such as the presence of pests and diseases that can affect production of broccoli. To avoid this, the authors build a model to identify diseases in broccoli through leaf images with a size of 128x128 pixels. The model is constructed to classify healthy leaves, and disease leaves using the image processing method that uses machine learning stages. There are several stages, including K-Means segmentation, colour feature extraction, and classification using SVM (Support Vector Machine) with RBF kernel and PSO (Particle Swarm Optimization) for reduce dimensionality data. The model that has been built compares the SVM model and the SVM-PSO model. It produces good accuracy in the training of 97.63% and testing accuracy of 94.48% for SVM-PSO and 85.82% for training, and 86.25% for testing in the SVM model. Therefore, this proposed model can produce good results in categorizing healthy and diseased leaves in broccoli.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Broccoli (*Brassica oleracea* var. *Italica*) is a family of Brassicaceae plants and vegetables. Broccoli comes from the Latin brachium, which means branch and arm. Broccoli has been a vegetable consumed by the Greeks for approximately 2500 years. Countries that have become one of the highest-producing broccoli are China, followed by India, the United States, Spain, Mexico, and Italy. In recent decades, broccoli consumption has increased due to concerns about vitamins, minerals, and low calories [1]. Broccoli is one of the green plants that are often consumed by humans. Broccoli contains many nutrients, including minerals, and it also contains vitamins A, C, E, and K. Broccoli provides a large number of secondary metabolites such as phenolic compounds (especially flavonoid derivatives and hydroxycinnamic acid) and glucosinolates [2]. Therefore, broccoli cultivation is harvested quickly, and this plant's need for exports and imports is also increasing. In 2014, the united states were the most used vegetable, with the utilization of fresh broccoli per capita to about 6.7 lbs, or approximately 3.0 kg (USDA-RS 2016) [3].

In their cultivation, farmers have problems with disease problems in broccoli. Diseases commonly found in broccoli are usually pests, viruses, and diseases caused by a lack of knowledge. Some of the diseases found in broccoli include Alternaria Spot, Black Rot, Downy Mildew, White Rust, Damping Off, and many more types [4]. One of these problems. It can be overcome by using techniques from

image processing. The method proposed in this study uses machine learning techniques with stages such as preprocessing, segmentation, color feature extraction, and classification techniques [5]. In this study, the authors did not collect all types of diseases on broccoli leaves. The authors limited the types of diseases in data collection because the datasets were still rare and had never been studied before on these leaves. Therefore, the author only limits it to 3 classes: *Alternaria* spot, Black rot, and Health leaf. Each type of leaf will be explained in the materials and methods section for characters. It hopes that this model can help farmers and the general public distinguish between healthy broccoli leaves and not by sharing this dataset and algorithm with the public source.

The study of PSO-SVM was also obtained from Maen Takruri et al. The authors used a histopathological dataset of skin cancer (melanoma). The method used in this research is to use the SVM and also PSO. The image is converted to a grayscale image histogram and segmented using edge detection. The SVM-PSO model is implemented to find the best feature and values of the SVM parameters, namely C and gamma. PSO will work to find the best performance in classifying. The results obtained produce an accuracy of 87.13%, the authors also compare the method of Wavelet Packet Transform and Sequential Forward Selection WPT-SFS-SVM produces an accuracy of 77.4% and SFS-SVM of 75.00% [6]. The following study, by Nianyin Zeng et al., is to diagnose Alzheimer's Disease using a total dataset of 361 images. The authors used principal component analysis and SVM to create a model. To increase the accuracy, the author uses a switching delayed particle swarm optimization (SDPSO) algorithm that results in an accuracy of 79.54% on PSO-SVM, and an increase in SDPSO-SVM-PCA results in an accuracy of 85.71% [7]. In this connection, we employ the KMeans algorithm for the segmentation process. In this paper, we also conducted a study on the color difference of each leaf. In broccoli leaf diseases classification, no researcher has implemented SVM with PSO-based feature selection before. Also, another study that implements SVM and PSO employs grayscale images. In our study, this approach leads to low accuracy in the classification process. By converting the color space of the input images to HSV and employing the KMeans algorithm as the segmentation method, we achieve a high accuracy.

The author uses SVM because one of the effective machine learning techniques that are easy to implement, able to work with small datasets, and provides an explorable technique known as the kernel with several hyperparameters used, so it can solve complex problems [8]. Panigrahi et al [9] developed a model to classify maize plants using 3823 images. The proposed approach consists of image preprocessing, image segmentation, feature extraction and image classification using SVM. In the classification task that employs a large number of input image, it is important to reduce features. Reduced features will decrease execution time and increase accuracy. However, this approach has not been implemented by Panigrahi et al. In this study, we apply feature selection so that the data dimensionality can be reduced. Furthermore, the execution time will be reduced and the proposed model can achieve a high accuracy. For feature selection, the author uses PSO because it is one of the techniques for feature selection and can be implemented quickly because quite a few parameters are implemented for an effective and efficient global search technique [10]. However, many studies have been on classifying images using various machine learning methods with different images, either with single datasets or with multiple datasets class. The author also relies on references in comparisons made by researchers, such as image size, classification methods, segmentation methods, and techniques, to find the best parameters. The main contribution of this research is to develop machine learning-based methods for real-world applications. This method is applied to classify broccoli leaf diseases based on leaf characteristics. Another contribution is helping a farmer identify broccoli disease based on leaf characteristics. The novelty in our study is a method of reducing features obtained by applying color histograms on leaf images. This method can decrease execution time since the number of processed data decreases. Also, the accuracy of the classification becomes higher compared to that of the model in which the PSO is not employed as the feature selection method. In addition, color space conversion and segmentation algorithms employed in our study yield a classification accuracy of 94.48%.

This research is organized as follows. Section 2 describes the flow of modeling a classification algorithm. Section 3 provides information on the results of the research that has been carried out,

containing the results of accuracy, time, limitations, and comparisons. Section 4 presents the conclusions and future research.

2. Method

In this study, the author proposes categorizing healthy and diseased leaves on broccoli leaves. Constructing a classification model using machine learning methods includes several stages: dataset search, preprocessing dataset, image segmentation, color feature extraction, data split, classification, and performance measurements. The flowchart picture can be shown in Fig. 1. We selected this model since it has many parameters that can be investigated and algorithms that can be implemented. In addition, it provides a good accuracy.

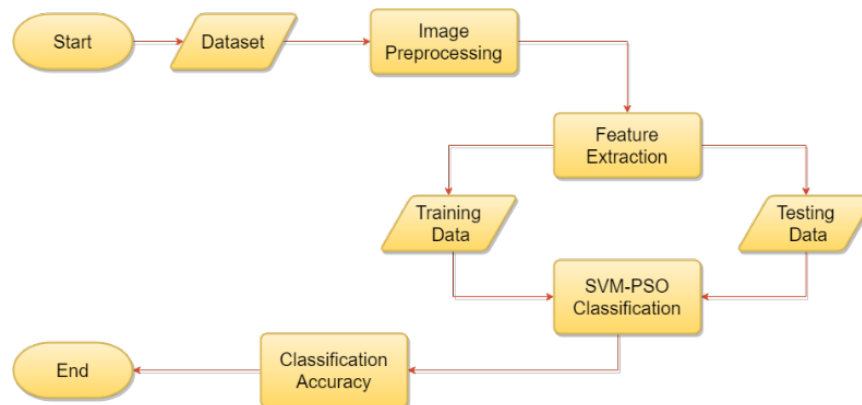


Fig. 1. Flowchart of a classification model

2.1. Dataset

The dataset used is an image of diseased broccoli leaves. The authors divide them into several classes: Black Rot, Alternaria Spot, and Healthy Leaf. Black rot is caused by the bacterium *Xanthomonas campestris*. In mature broccoli plants, the typical symptom of attack is the presence of yellow spots that resemble the letter Vshaped along the edge of the leaf towards the center of the leaf. In severe attacks, all the leaves turn yellow and fall off easily [11]. Alternaria Spot is caused by the fungus *Alternaria brassicae* or *Alternaria brassicicola*. Pathogens are transmitted through the air. The color of the spots is brown with concentric circles. A gray to black mass of conidia is seen above the spot. The spots produced by *A. brassicicola* are darker [12]. Healthy leaf is a curly and pretty decorative with full green color without any spots and color differences on the leaves [13]. The total dataset used is 2000 images from each class. The total is more than 6000 images. The dataset is collected through plantations and then collected and photographed using a camera phone with 12 megapixels resolution with a white background. The image dataset of each class can be shown in Fig. 2.

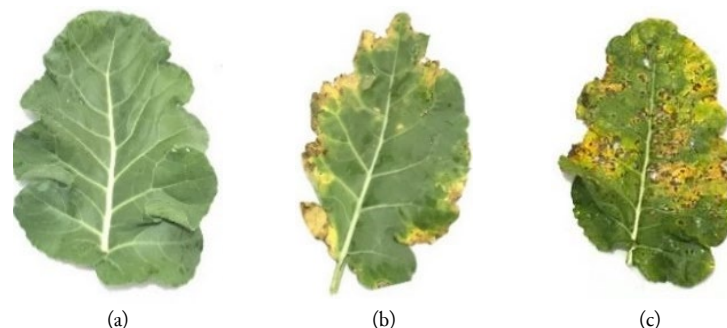


Fig. 2. The Dataset of Broccoli Leaves (a) Health Leaf (b) Black Rot (c) Alternaria Spot

All datasets are divided into three classes, split into training and testing data, with a ratio of 70:30 [14]. Each class is divided as follows. Table 1 shows the number of each class.

Table 1. Number of datasets

Class	Images
Health Leaf	2052
Black Rot	2054
Alternaria Spot	2050
Total	6156

2.2. Preprocessing

After collecting the dataset, the next step is image preprocessing. Image preprocessing also improves image quality, containing unwanted distortions and amplifying some image features. Image preprocessing techniques are also helpful for changing image size and shape, noise filtering, and image conversion [15]. This study is used to resize an image and convert it to HSV, and the size of pixels is converted to 128x128. Fig. 3 shows the image before and after conversion to HSV color.

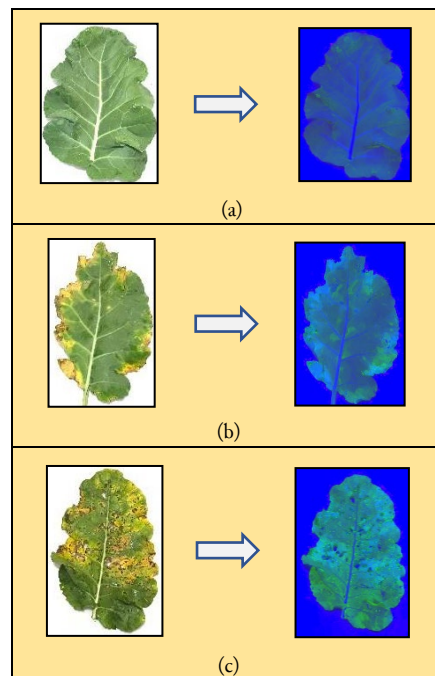


Fig. 3. The color of broccoli leaves (a) Health Leaf HSV Color (b) Black Rot HSV Color (c) Alternaria Spot HSV Color

2.3. Segmentation

Image segmentation is a fundamental goal in image processing. At this stage, it is used as a process to divide the partition into several segments. By segmenting the image, it becomes simpler and easier to analyze. This step is commonly used for object detection, image recognition, medical imaging, and others. Image segmentation can also be used in single or multi-class. For multi-class, this process will work by grouping pixels of the same attribute and creating pixel masks [16]. The algorithm for image segmentation is based on two fundamental properties, namely discontinuity and similarity. The first is based on different intensities, such as edges in the image. Then the second approach is based on pre-determined criteria for partitioning the image [17]. This segmentation stage uses the basics of K-Means. The algorithm is based on classical distance, and the similarity will be assessed from a distance obtained. The equation of K-Means is expressed in (1) [18].

$$fit = \sum_{i=1}^k \sum_{X_j \in S_i} d(C_i, X_j) \quad (1)$$

where $d(a, b)$ represents the square of the Euclidean distance between a and b , the parameter value used in this model is $k=9$, and the iterations number is set to 10. The process carried out on the K-Means algorithm includes: Clusters are randomly selected, then will be calculated from the center to cluster distance from each category, and classify the category as the center of the closest cluster, then recalculate the center of the cluster from each separated class, and do iterations by repeating the previous step until the center can reach the threshold [19].

2.4. Feature Extraction

In this study, feature extraction is implemented based on color. Color is a visual attribute commonly used in computer vision, especially image processing. One of the methods used in this study is the use of a color histogram, which effectively represents features. Color selection in the image can be made before displaying a histogram. In image processing, one of which is the color selection is done on RGB color because the resulting representation is concise and computationally low. Then, there is HSV, where this color is another option from RGB. This color can imitate the human visual system and is usually used in feature extraction [20]. HSV (Hue, Saturation, and Value) is a different model from RGB (Red, Green, and Blue). This method can eliminate false data caused by light, hue, or shadow colors, and saturation is used to describe the amount of gray level and brightness [21]. Fig. 4 shows a cylindrical view of the HSV color model [22].

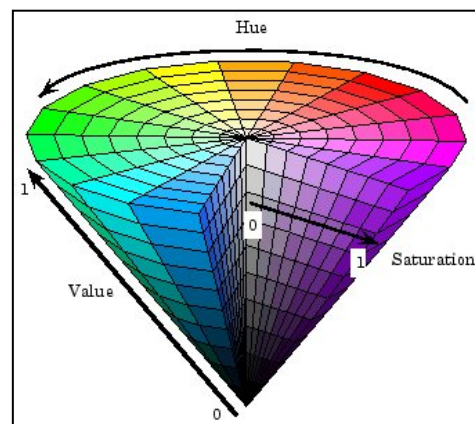


Fig. 4. HSV cylinder color

Fig. 5 shows the plot results of each example class of broccoli leaf images. The author displays the plot based on HSV color (Hue, Saturation, and Value) with blue line color Hue, green line color saturation, and red line color value. To calculate histograms, several parameters are used. The parameters begin with three channels in the form of hue, saturation, and value. Then, a mask parameter is set to 'none' to obtain the histogram result with the complete image. The histSize parameter represents the number of bins and their size range, which is usually 0 to 256 [23].

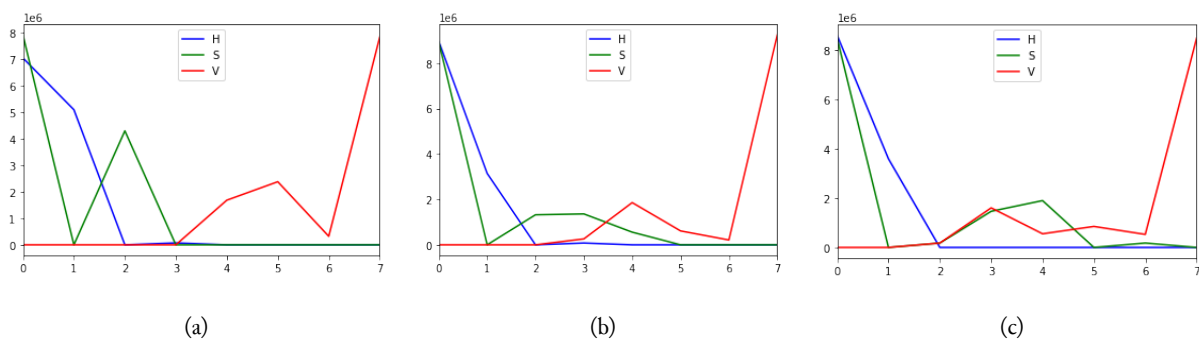


Fig. 5. Color histogram plot (a) Health Leaf (b) Black Rot (c) Alternaria Spot

2.5. Feature Selection

Optimization is needed from a built model to increase accuracy in classification. In this study, to increase the accuracy, the author uses the feature selection method, where feature selection is a process for selecting similar or relevant features. The workings of feature selection are to reduce some data that is not informative, then with a few select features can speed up the training process with a faster time. Fig. 6 shows the architecture of PSO feature selection. PSO is described as a flock where it is simulated with a flock of birds. This algorithm is usually used to solve optimization problems for multi-objective optimization, nonlinear programming, combinatorial optimization problems, and others [24]. The flow of feature selection can be shown in Fig. 6.

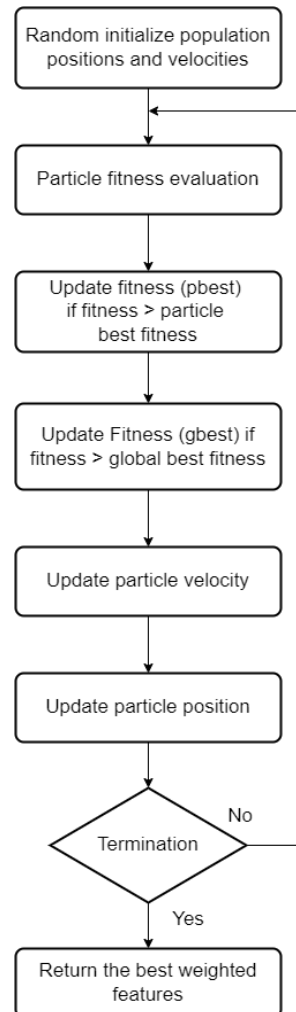


Fig. 6. Feature subset selection flow

Based on Fig. 6, PSO starts with random initialize for population positions and velocities, where the solution vector will represent subset features $x = [x_1, x_2, \dots, x_d]$; $x_i \in [0,1]$ the total features number in dataset written with d . Next, threshold to determine which feature to choose, is expressed in (2).

$$x_i = \begin{cases} 1, & \text{if } x_i > 0.05 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The optimized function is the accuracy of the classification, which is penalized by the number of selected features. Here is the solution expressed in (3).

$$f(x) = \alpha \times (1 - P) + (1 - \alpha) \times \frac{N_{\text{selected}}}{N_{\text{features}}} \quad (3)$$

where α is the parameter that determines the trade-off between classifier performance accuracy P and the size of a selected feature $N_{features}$ with respect to $N_{selected}$. The parameter value used in this model is $\alpha = 0,99$ and threshold value = $0,05$.

2.6. Support Vector Machine (SVM)

The next stage is classification. In this research, the identification of plant diseases through leaf imagery uses the SVM. SVM is a binary statistical classification technique that has two parts. The first part is to find the optimal hyperplane between two samples using quadratic or linear programming. Then the second part is to map, commonly called a kernel, whose way of working is to convert the information space into a feature space of higher dimensions, making linear discrimination better [25]. The hyperplane of SVM to separate the training samples shows in (4). This proposed method uses several hyperparameters on the SVM, namely the RBF kernel with the value of C set to 100, and the gamma is scale.

$$f(x) = \text{sign}(w \cdot x) + b \quad (4)$$

where w is the vector's weight, and b is a threshold value. And b for minimized to maximize margin $w \in f$ in (5), which can be expressed as a quadratic problem in (6) [26].

$$y_i(w \cdot x_i) + b \geq 1 \quad (5)$$

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (6)$$

2.7. Confusion Matrix

After all, models have been successfully executed. The results will be evaluated with a confusion matrix. The confusion matrix table used can be shown in Table 2 [27].

Table 2. Confusion Matrix

		Predicted Class	
		Positive	Negative
Actual Class	True	TP	FN
	False	FP	TN

There are four types of assessments: first, true positive, which means that both the actual and predicted classes are correct. Second, true negative, which means that the predictive model is the real value but the actual is negative. Third, false negative, meaning that the predicted model is incorrect but the actual value is positive. Lastly, false positive, meaning that the predicted model is incorrect but the actual value is negative. The most common indicator include precision, recall, f1-score, and accuracy [28]. Indicator calculations can be displayed in (7)-(10) [29].

$$\text{Accuracy} = \frac{TP+TN}{N} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$F1 - \text{Score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

3. Results and Discussion

The results obtained with several stages have been an experiment. The model has been run on a device with specifications Windows 10 (64-bit) and an Intel Core i5-7200U CPU. This model uses a total dataset of broccoli leaves with three types of classes, namely Alternaria Spot, Black Rot, and Health

Leaf, with a total of 6256 images and then split into 70% training and 30% testing, each image measuring 128x128 pixels. The initial stage of preprocessing the data was done using the K-Means segmentation method using the parameter $k=9$. The dataset then goes through the feature extraction stage to determine each class's characteristics based on the image's color by changing the image to HSV color, then to improve the results of a good hyperplane using PSO feature selection. The model and there are two comparisons. The first is carried out using only the SVM classification method using RBF Kernel with $C=100$, and the gamma is scale. The second experiment uses the PSO feature selection and SVM classifier using the same number of datasets and parameters from SVM. The time complexity of this algorithm is $O(n^3)$ as the standard of SVM Classifier, this is because the size of the large dataset and the multi-class dataset causes high time complexity [30]. For the PSO technique, the time complexity is $O(n^2t)$ when passing through a population of size n and one outer loop for t cycles [31]. Then the model is rebuilt by adding feature selection to compare the final results. The results obtained from the training stage produce the highest train accuracy on SVM PSO. Table 3 proves that the results of the training model on various types of broccoli leaves are a total of 4309 data. After the training, the accuracy value with SVM is quite good and increases by 11.81% using PSO. The details of train accuracy can be shown in Table 3.

Table 3. The results of training accuracy

Algorithm	Training
SVM	85.82%
SVM + PSO	97.63%

In this study, we demonstrate the effect of PSO, i.e. increasing the accuracy of the model that employs SVM as the classification method and decreasing the execution time. Followed by the results of the accuracy, which can be shown in Table 4. The highest accuracy results were obtained using the PSO feature selection compared to the SVM classifier. The most outstanding accuracy yields 94.48% for SVM PSO and 86.25% for SVM classifier. The calculation indicator and the average of each class is calculated, and then the results are obtained in the form of a graph shown in Fig. 7. It can be seen that all indicators include precision, recall, F1-Score, and accuracy. The accuracy of the SVM PSO is higher than using only the SVM classifier.

Table 4. The results of testing accuracy

Algorithm	Class	Precision	Recall	F1-Score	Accuracy	Processing Time (s)
SVM	Alternaria Spot	94%	83%	88%	86.25%	3.691
	Black Rot	91%	79%	80%		
	Health Leaf	85%	97%	90%		
SVM + PSO	Alternaria Spot	95%	94%	94%	94.48%	1.412
	Black Rot	92%	92%	92%		
	Health Leaf	96%	98%	97%		

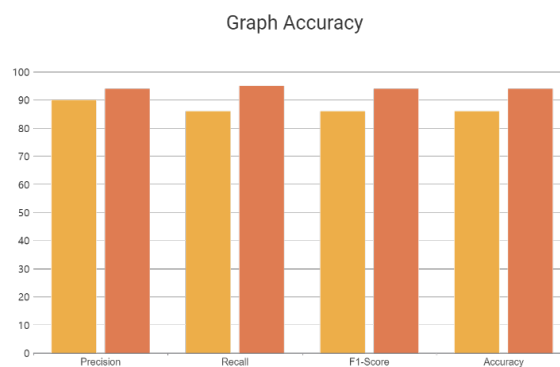


Fig. 7. Average accuracy indicator

It was proven that the results increased by approximately 8.23% on the broccoli leaf image. It also proves that the SVM with feature selection PSO is faster in running the model, achieving a difference of 2.279s. PSO that is employed as the feature selection algorithm can reduce the execution time. This model can improve the accuracy of classification. Also, by employing PSO as a feature selection method, we can reduce unnecessary features for the classification process. This approach will reduce execution time by computer. In addition, since we reduce features for the classification, the data dimensionality is also reduced. Table 5 shows a state of the art comparison. From Table 5, it has been proved that the use of PSO combined with SVM can produce a good accuracy.

Table 5. State-of-the-art

Author	Method	Accuracy
Takruri <i>et al.</i> [6]	SVM+PSO. The image is converted to a grayscale image histogram and segmented using edge detection	87.13%
Zeng <i>et al.</i> [7]	SDPSO+SVM model with PCA. The image is converted to grayscale value	85.71%
Proposed Model	SVM+PSO. The image is converted to HSV and using KMeans segmentation	97.63%

Since the study of broccoli leaf diseases classification has not been performed by the other researchers, the author compare the results of classification using SVM with the result of classification using SVM+PSO with segmentation algorithm and color space conversion to demonstrate the effect of feature selection performed by PSO. Table 6 shows an extra set of experiments by compare the color space namely RGB, HSV, and YCrCb.

Table 6. An experiment with color space

Dataset	Color Space	Accuracy
SVM	RGB	78.34%
	HSV	86.25%
	YCrCb	80.29%
SVM+PSO	RGB	92.80%
	HSV	97.63%
	YCrCb	80.67%

Some images do not match the appropriate labels between classes. From the confusion matrix image, the healthy leaf class is most recognized by the model, then the Alternaria spot class, and the most incorrect class is black rot. More details of the confusion matrix can be shown in Fig. 8. For healthy broccoli leaves, the confusion matrix shows promising results for the SVM classification model for healthy leaves resulting in 596 of the 615 testing data. In contrast, for the SVM PSO model, results that 603 of the testing data can be recognized. Then, the black rot class resulted from a total of 616 classes. The SVM model recognized as many as 486 testing images, and the SVM PSO well recognized 565 testing images. And lastly, for the Alternaria spot class from a total of 616 class images, the model can recognize 511 images for SVM and 577 for SVM PSO. From the whole process of data testing, healthy leaves tend to be easily recognized because the color of healthy leaves has a consistent color compared to diseased leaves. Our method has not been tested on imbalanced data. This scenario should be investigated further.

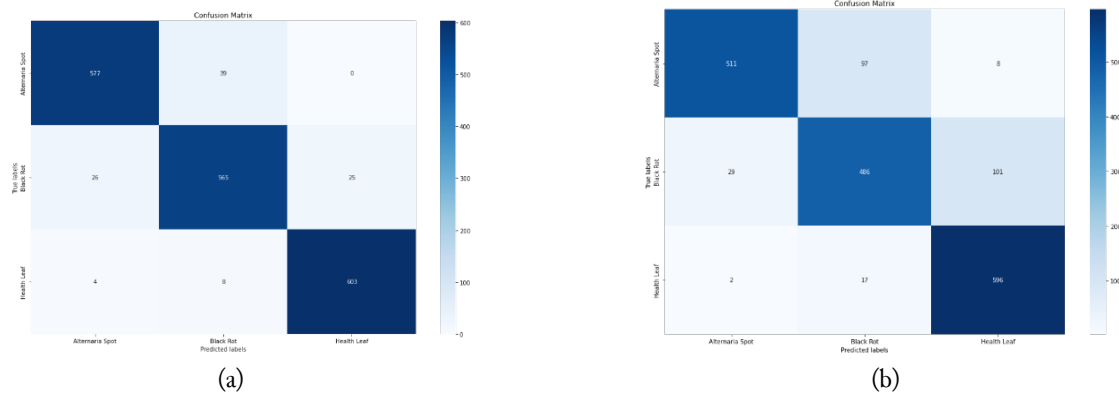


Fig. 8. The confusion matrix (a) SVM (b) SVM PSO

4. Conclusion

Based on the results, the author concludes that the model has been built by exploring for classification on the broccoli leaf image and by optimizing to increase accuracy has succeeded with 94,48% test accuracy, it can reduce the complexity and make train algorithm faster than only using standard classifier method. This paper still has limitations, i.e., the collection of datasets is still small. In the future, we will investigate the influence of various color spaces and various segmentation methods to the accuracy of the model. In addition, we will investigate the performance of the other feature selection methods in increasing the accuracy of our model. Also, we will employ deep learning methods for broccoli leaf diseases classification.

Acknowledgment

The author is very grateful to have completed this research and is very grateful to the Telkom University campus, which has provided support and funding in carrying out this research.

Declarations

Author contribution. Yulio Ferdinand organized data collection, article compilation, data analysis and interpretation. Wikky Fawaz Al Maki conceived Conception or design of the work, data analysis and interpretation, and critical revision of the article

Funding statement. This research is supported by Telkom University

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paperr.

References

- [1] G. S. Nagraj, A. Chouksey, S. Jaiswal, and A. K. Jaiswal, "Broccoli," in *Nutritional Composition and Antioxidant Properties of Fruits and Vegetables*, Elsevier, 2020, pp. 5–17, doi: [10.1016/B978-0-12-812780-3.00001-5](https://doi.org/10.1016/B978-0-12-812780-3.00001-5).
- [2] I. Hassini, J. J. Rios, P. Garcia-Ibañez, N. Baenas, M. Carvajal, and D. A. Moreno, "Comparative effect of elicitors on the physiology and secondary metabolites in broccoli plants," *J. Plant Physiol.*, vol. 239, pp. 1–9, Aug. 2019, doi: [10.1016/j.jplph.2019.05.008](https://doi.org/10.1016/j.jplph.2019.05.008).
- [3] P. Bhattacharjee and R. S. Singhal, "Broccoli and Cauliflower," in *Handbook of Vegetables and Vegetable Processing*, vol. 2–2, Chichester, UK: John Wiley & Sons, Ltd, 2018, pp. 535–558, doi: [10.1002/9781119098935.ch23](https://doi.org/10.1002/9781119098935.ch23)
- [4] O. I. A. LAfi, H. A. El-Hamarnah, N. J. H. Al-Saloul, H. I. A. Radwan, and S. S. Abu-Naser, "A Proposed Expert System for Broccoli Diseases Diagnosis," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 6, pp. 43–51, 2022, Accessed: Jan. 04, 2020. [Online]. Available: <https://philpapers.org/rec/LAFAPE>.

- [5] N. M. Patil and M. U. Nemade, "Content-Based Audio Classification and Retrieval Using Segmentation, Feature Extraction and Neural Network Approach," *Adv. Intell. Syst. Comput.*, vol. 924, pp. 263–281, 2019, doi: [10.1007/978-981-13-6861-5_23/COVER](https://doi.org/10.1007/978-981-13-6861-5_23/COVER).
- [6] M. Takruri, M. K. A. Mahmoud, and A. Al-Jumaily, "PSO-SVM hybrid system for melanoma detection from histo-pathological images," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 2941–2949, Aug. 2019, doi: [10.11591/IJECE.V9I4.PP2941-2949](https://doi.org/10.11591/IJECE.V9I4.PP2941-2949).
- [7] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, and Y. Li, "A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 320, pp. 195–202, Dec. 2018, doi: [10.1016/J.NEUCOM.2018.09.001](https://doi.org/10.1016/J.NEUCOM.2018.09.001).
- [8] I. Izonin, A. Trostianchyn, Z. Duriagina, R. Tkachenko, T. Tepla, and N. Lotoshynska, "The combined use of the wiener polynomial and SVM for material classification task in medical implants production," *Int. J. Intell. Syst. Appl.*, vol. 10, no. 9, pp. 40–47, Sep. 2018, doi: [10.5815/IJISA.2018.09.05](https://doi.org/10.5815/IJISA.2018.09.05).
- [9] K. P. Panigrahi, H. Das, A. K. Sahoo, and S. C. Moharana, "Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms," *Adv. Intell. Syst. Comput.*, vol. 1119, pp. 659–669, 2020, doi: [10.1007/978-981-15-2414-1_66/COVER](https://doi.org/10.1007/978-981-15-2414-1_66/COVER).
- [10] D. K. Choubey, P. Kumar, S. Tripathi, and S. Kumar, "Performance evaluation of classification methods with PCA and PSO for diabetes," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 9, no. 1, pp. 1–30, Dec. 2020, doi: [10.1007/S13721-019-0210-8/METRICS](https://doi.org/10.1007/S13721-019-0210-8/METRICS).
- [11] C. Vega-álvarez, M. Francisco, and P. Soengas, "Black Rot Disease Decreases Young Brassica oleracea Plants' Biomass but Has No Effect in Adult Plants," *Agron. 2021, Vol. 11, Page 569*, vol. 11, no. 3, p. 569, Mar. 2021, doi: [10.3390/AGRONOMY11030569](https://doi.org/10.3390/AGRONOMY11030569).
- [12] S. De Britto, D. Kapera, and S. Jogaiyah, "First Report of Leaf Spot Disease Caused by *Alternaria brassicicola* on Broccoli in Papua New Guinea," <https://doi.org/10.1094/PDIS-04-20-0721-PDN>, vol. 104, no. 11, p. 3073, Sep. 2020, doi: [10.1094/PDIS-04-20-0721-PDN](https://doi.org/10.1094/PDIS-04-20-0721-PDN).
- [13] "Retracted: Kale (*Brassica oleracea* var. *sabellica*) as miracle food with special reference to therapeutic and nutraceuticals perspective," *Food Sci. Nutr.*, vol. 10, no. 9, p. 3175, Sep. 2021, doi: [10.1002/FSN3.2476](https://doi.org/10.1002/FSN3.2476).
- [14] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math. Probl. Eng.*, vol. 2021, 2021, doi: [10.1155/2021/4832864](https://doi.org/10.1155/2021/4832864).
- [15] S. Krig, "Image Pre-Processing," in *Computer Vision Metrics*, Berkeley, CA: Apress, 2014, pp. 39–83, doi: [10.1007/978-1-4302-5930-5_2](https://doi.org/10.1007/978-1-4302-5930-5_2).
- [16] D. S. Manoharan, "Performance Analysis of Clustering Based Image Segmentation Techniques," *J. Innov. Image Process.*, vol. 2, no. 1, pp. 14–24, Mar. 2020, doi: [10.36548/jiip.2020.1.002](https://doi.org/10.36548/jiip.2020.1.002).
- [17] K. Jeevitha, A. Iyswariya, V. Ramkumar, S. Mahaboob Basha, V. Praveen Kumar, and R. M. K. Engineering, "A Review on Varius Segmentation Techniques in Image Processing," *Eur. J. Mol. Clin. Med.*, vol. 7, no. 4, pp. 1342–1348, 2020, Accessed: Jan. 04, 2020. [Online]. Available: https://ejmcm.com/article_1830_068d587434d5cb7da8434f84e1d29cd8.pdf.
- [18] R. C. Hrosik, E. Tuba, E. Dolicanin, R. Jovanovic, and M. Tuba, "Brain Image Segmentation Based on Firefly Algorithm Combined with K-means Clustering," *Stud. Informatics Control*, vol. 28, no. 2, pp. 167–176, Jul. 2019, doi: [10.24846/v28i2y201905](https://doi.org/10.24846/v28i2y201905).
- [19] P. Shan, "Image segmentation method based on K-mean algorithm," *Eurasip J. Image Video Process.*, vol. 2018, no. 1, pp. 1–9, Dec. 2018, doi: [10.1186/S13640-018-0322-6/TABLES/2](https://doi.org/10.1186/S13640-018-0322-6/TABLES/2).
- [20] H. Qazanfari, H. Hassanpour, and K. Qazanfari, "Content-Based Image Retrieval Using HSV Color Space Features," *Int. J. Comput. Inf. Eng.*, vol. 13, no. 10, pp. 537–545, 2019, doi: [10.5281/zenodo.3566277](https://doi.org/10.5281/zenodo.3566277).
- [21] T. Tiay, P. Benyaphaichit, and P. Riyamongkol, "Flower recognition system based on image processing," in *2014 Third ICT International Student Project Conference (ICT-ISPC)*, Mar. 2014, pp. 99–102, doi:

[10.1109/ICT-ISPC.2014.6923227](https://doi.org/10.1109/ICT-ISPC.2014.6923227).

- [22] A. M. Al-Hetar, M. A. Rassam, O. Shormani, A. S. A. Salem, and H. Al-Yousofi, "Color-based Object Categorization Model Using Fuzzy HSV Inference System," in *2019 First International Conference of Intelligent Computing and Engineering (ICOICE)*, Dec. 2019, pp. 1–6, doi: [10.1109/ICOICE48418.2019.9035173](https://doi.org/10.1109/ICOICE48418.2019.9035173).
- [23] S. Talla, P. Venigalla, A. Shaik, and M. Vuyyuru, "Multiclass Classification Using Random Forest Classifier," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 5, no. 2, pp. 493–496, Mar. 2019, doi: [10.32628/CSEIT183821](https://doi.org/10.32628/CSEIT183821).
- [24] M. A. El-Shorbagy and A. E. Hassanien, "Particle Swarm Optimization from Theory to Applications," *Int. J. Rough Sets Data Anal.*, vol. 5, no. 2, pp. 1–24, Apr. 2018, doi: [10.4018/IJRSDA.2018040101](https://doi.org/10.4018/IJRSDA.2018040101).
- [25] D. K. Jain *et al.*, "An approach for hyperspectral image classification by optimizing SVM using self organizing map," *J. Comput. Sci.*, vol. 25, pp. 252–259, Mar. 2018, doi: [10.1016/j.jocs.2017.07.016](https://doi.org/10.1016/j.jocs.2017.07.016).
- [26] M. Aminul Islam, M. Sayeed Iftexhar Yousuf, and M. M. Billah, "Automatic Plant Detection Using HOG and LBP Features With SVM," *Int. J. Comput.*, Accessed: Jan. 04, 2023. [Online]. Available: <http://ijcjournal.org/>.
- [27] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," vol. 49, no. 9, pp. 2080–2093, May 2019, doi: [10.1080/03610926.2019.1568485](https://doi.org/10.1080/03610926.2019.1568485).
- [28] M. Tripathi, "Analysis of Convolutional Neural Network based Image Classification Techniques," *J. Innov. Image Process.*, vol. 3, no. 2, pp. 100–117, Jun. 2021, doi: [10.36548/jiip.2021.2.003](https://doi.org/10.36548/jiip.2021.2.003).
- [29] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," in *Computer Science & Information Technology*, Jun. 2020, pp. 01–14, doi: [10.5121/csit.2020.100801](https://doi.org/10.5121/csit.2020.100801).
- [30] G. N. Kouziokas, "SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting," *Appl. Soft Comput.*, vol. 93, p. 106410, Aug. 2020, doi: [10.1016/j.asoc.2020.106410](https://doi.org/10.1016/j.asoc.2020.106410).
- [31] A. Khan, H. Hizam, N. I. bin Abdul Wahab, and M. Lutfi Othman, "Optimal power flow using hybrid firefly and particle swarm optimization algorithm," *PLoS One*, vol. 15, no. 8, p. e0235668, Aug. 2020, doi: [10.1371/journal.pone.0235668](https://doi.org/10.1371/journal.pone.0235668).