ANSF APPLIED AND NATURAL SCIENCE FOUNDATION

Journal of Applied and Natural Science
15(2), 852 - 859 (2023)
ISSN : 0974-9411 (Print), 2231-5209 (Online)
journals.ansfoundation.org

Research Article

Performance analysis of support vector machine for early identification of citrus diseases

Utpal Barman*
Faculty of Computer Technology, Assam down town University, Guwahati (Assam), India
Ridip Dev Choudhury
Department of Computer Science, Krishna Kanta Handiqui State Open University, Guwahati (Assam), India
Bishwajeet Bal
Faculty of Computer Technology, Assam down town University, Guwahati (Assam), India
Anirban Bhattacharjee
Faculty of Computer Technology, Assam down town University, Guwahati (Assam), India
Samujjal Talukdar
Faculty of Computer Technology, Assam down town University, Guwahati (Assam), India
Nilanjan Nath
Faculty of Computer Technology, Assam down town University, Guwahati (Assam), India
Parikhit Basnet
Faculty of Computer Technology, Assam down town University, Guwahati (Assam), India
Abhay Bhowmick
Faculty of Computer Technology, Assam down town University, Guwahati (Assam), India

*Corresponding author: E-mail: utpalbelsor@gmail.com

Article Info
https://doi.org/10.31018/jans.v15i2.4630
Received: April 14, 2023
Revised: June 6, 2023
Accepted: June 11, 2023

How to Cite
Barman, U. *et al.* (2023). Performance analysis of support vector machine for early identification of citrus diseases. *Journal of Applied and Natural Science*, 15(2), 852 - 859. https://doi.org/10.31018/jans.v15i2.4630

Abstract
Early citrus disease detection is necessary for optimum citrus productivity. But detecting a citrus disease at an early stage requires expert views or laboratory tests. But getting an expert view of all time is impossible for rural farmers. The present study aimed to create a low-cost, intelligent, affordable citrus disease classification system. This study offered a Support Vector Machine (SVM) based smart classification method for categorizing various citrus diseases. Citrus photos were subjected to a variety of image processing techniques to categorize the diseases using SVM and the kernel. Prior to classification, the images were segmented and the hue channel threshold value was used to differentiate the diseased area from the remaining portion of the image. The segmented image's color and grey domains were used to extract 13 different texture and color features. This study outlined three different SVM kernel types- Linear, Gaussian, and Polynomial, and evaluated their accuracy and confusion matrix performances. The Radial Based Function with a polynomial kernel derived from the SVM outperformed the SVM's linear and Gaussian kernel.

Keywords: Citrus, Kernal, Machine learning, Plant disease, Support vector machine

INTRODUCTION

The leaves and fruits of a plant are an important part of a plant. Crop fruit and leaf surfaces are the major areas for plant diseases. Plant infections always reduce crop production and lower the fruit's quality. An expert can frequently find the plant's sickness in the leaf or fruit regions. But from a farmer's perspective, it is difficult to identify the plant problem. Farmers can occasionally identify a plant's disease with their unaided eyes. Looking at the plant leaves and fruits, they are unsure about the disease's class. Thus, they frequently struggle with the issue of spotting plant illness. Experts can identify a plant's sickness using a laboratory technique or their knowledge of the subject. The laboratory procedures need a lot of labour and time. Yet, computer automation enables humans to develop the difficult problem's most error-free solution(Barman *et al.*, 2020; Barman and Choudhury, 2021). Therefore, creating a digital system to identify plant diseases using automation and computers is crucial. The Support Vector Machine (SVM) with the kernel is used in this study to categorize the various citrus diseases using a digital image-based disease classification method.

Recent research has concentrated on computerized systems for identifying and diagnosing plant diseases. Sharif *et al.* (2018) reported a method for diagnosing citrus illness that uses multiclass SVM and better-weighted segmentation. They used this technique on three different datasets, including the Plant Village, the Citrus Disease Photo Gallery Dataset(Citrus Dataset, 2021 (n.d.). www.idtools.org), and the Database of gathered photographs. They had a classification accuracy of 97% on the dataset from the citrus disease photo gallery dataset, 89% on the total dataset, and 90.4% on our local dataset.

A review study on citrus plants disease identification and categorization using techniques for image processing was forwarded by Iqbal *et al.* (2018). They discussed several citrus diseases, diverse image-processing methods, picture segmentation, feature extraction, and possible measures for categorizing citrus diseases. After segmenting the leaves picture using K-Mean (Devi *et al.,* 2018) reported a plant disease detection and classification system utilizing an SVM classifier. For the classification, they had an accuracy rate of 98.3%. Using the expectation-maximization approach, Zhang *et al.* (2019) presented the classification system for diseases of cucumber leaves. They used superpixel clustering to segment the photos. A cost-SVM-based method for detecting citrus greening disease was presented by Deng *et al.* (2016). After taking them under the visible spectrum, they used the Hue Saturation Value (HSV) histogram and the Gray histogram to extract the features of the photographs. Compared to the other SVM kernels, they had the highest accuracy (87.52%) using the RBF kernel.

A disease detection retrieval system for the soybean plant employing low-level image features was examined by Patil and Kumar (2017). They analyzed the color feature using the HSV histogram, the shape feature using the scale-invariant feature transform, and the texture feature using the local grey Gabor pattern. For three distinct kinds of soybean disease, they integrated all three techniques and attained retrieval efficiencies of 96%, 68%, and 76%, respectively. A deep convolution network-based plant leaf disease classification system was presented by Sladojevic *et al.* (2016). The created model is capable of identifying 13 various plant diseases kinds. During separate class tests, they received 91% and 98% accuracy scores. Khirade and Patil (2015) demonstrated a method for detecting plant diseases using an image processing technique. For picture segmentation, they used the K mean approach, for feature extraction, they used a grey-level co-occurrence matrix; and for disease classification, they used an artificial neural network.

Citrus scab and citrus anthracnose disease classification using SVM with the aid of RBF and polynomial kernel was presented by Gavhale *et al.* (2014). Using Radial Based Function (RBF) based SVM, they reached a maximum acceptance rate of 96%. Arivazhagan *et al.* (2013) demonstrated a system for detecting leaf illness using a textural feature. They used the green pixel masking method to segment the previously processed images, and the Gray Level Co-occurrence Matrix (GLCM) was used to determine the characteristics of the images. The classification of the many diseases affecting the various species makes use of artificial neural networks, which have an accuracy rate of 87.66% for identifying leaf disorders. Asraf *et al.* (2012) used SVM with the linear kernel to classify palm leaf disease and achieved 95% accuracy. Al-Hiary *et al.* (2011) gave a disease detection and classification system employing K Mean clustering, green pixel masking for picture segmentation, a co-occurrence approach for texture feature analysis, and an artificial neural network. They have an illness detection accuracy of 83% and a disease classification system accuracy of 94%. Zhang and Meng (2011) presented an autonomous Citrus leaf canker detection system employing an upgraded AdaBoost Algorithm and lesson descriptor. The classification system for plant diseases based on Artificial Neural Network (ANN) and SVM was presented by Pujari *et al.* (2016). They calculated the input photos' colour and texture attributes, and their categorization accuracy peaked at 92%.

The yield and quality of citrus fruits are constantly declining due to several illnesses that affect citrus. A few different types of research have been identified in citrus, although several image processing and machine learning approaches have been created to detect and classify the many plant diseases. From the research above, an image's color and textural properties can be used to identify and categorize various plant diseases. In the present study, citrus leaves and fruit photos were taken from a public dataset, namely, Citrus ID (Citrus Dataset, 2021) database. Then, the disease section of the images was separated from the other parts of the images using hue-based image segmentation techniques. Finally, the study aimed to categorize citrus diseases using SVM, and the effectiveness of SVM was examined using Linear, RBF, and Polynomial kernels.

MATERIALS AND METHODS

This study proposed a citrus disease classification system using HSV thresholding-based segmentation, GLCM-based feature extraction, and SVM for citrus disease classification. Fig. 1 represents the block diagram of the proposed method.

Dataset for citrus images

Citrus leaves and fruits were gathered from a public dataset called Citrus ID (Citrus Dataset, 2021) dataset (Fig. 2). As seen in Fig. 2, the dataset image was made up of 96 dpi fruit and leaf images with a dimension of 100x150. Alternaria, Anthracnose, CCDV disease, Chimera, Citrus Blackspot, Canker, Scab, Citrus Subburn, Melanose, Leprosy, and a few more groups of citrus diseases were among those present in the dataset (Citrus Dataset, 2021 (n.d.). www.idtools.org).

Citrus Black Spot, Citrus Scab, and Citrus Leprosis were three separate kinds of citrus diseases classified using SVM. There are 101 photos of citrus leaves and fruit in the Citrus Black Spot Class. Leprosis class only had 43 photos of citrus, compared to Citrus Scab's 78 varied citrus images.

Image pre-processing and segmentation

The process of improving the quality of an image by size adjustment, contrast improvement, noise reduction, etc., is image pre-processing. The contrast of the photos needs to be improved because it varies from image to image in the dataset. With the help of MatLab 2015's built-in function (stretch), the citrus image contrasts were improved. The contrast enhancement of citrus scab disease is shown in Fig. 3.

The segmentation of the citrus image is the next process. Segmentation in computer vision refers to dividing a picture into smaller pieces that can be more easily understood and analyzed (Sunny and Gandhi, 2018; Zhang *et al.*, 2019). The present study created a mask

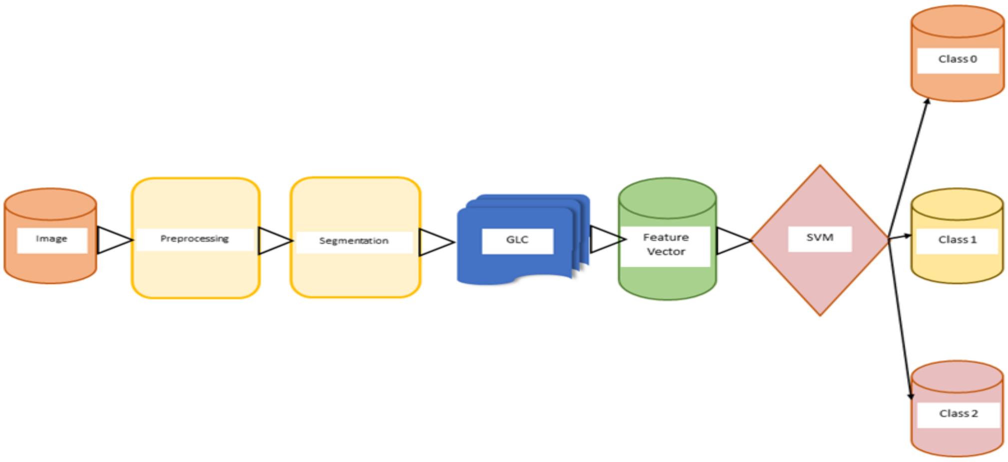

Fig. 1. *Block diagram of the citrus disease classification system*

Fig. 2. *Different sample images of citrus disease (Citrus Dataset, 2021)*

that segmented the images using the input image's color threshold as a reference point. The input images' Red, Green, and Blue (RGB) colors were converted into Hue Saturation Value (HSV) color images during the process, and the color histogram was used to determine the threshold for each HSV color space channel. In the experiment, the threshold values for the hue, saturation, and value channels ranged from 0.056 to 0.136 for hue, 0.128 to 1.000 for saturation, and 0.087 to 0.913 for value. The segmented images of the citrus leprosy disease are shown in Fig. 4.

Feature extraction for images

In the realm of image processing, visual feature extraction is crucial. Even though each citrus image has a unique form, dimension, color, and texture (Khirade and Patil, 2015) extracting various citrus image characteristics was essential. While categorizing a citrus image, colors and textures were crucial. Table 1 illustrates the different colors and textures of citrus.

Enhanced segmented images were taken into account for classification in the present work. For feature extrac-

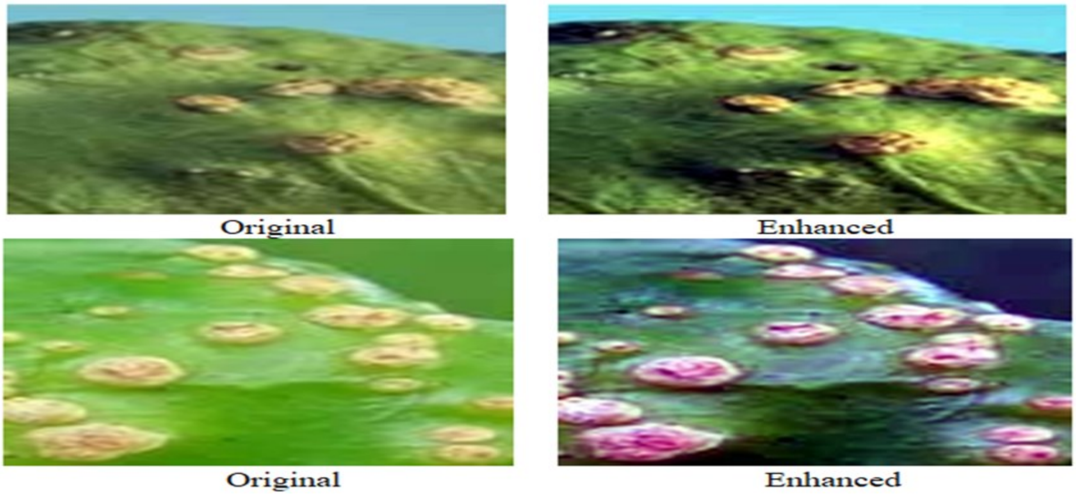

Fig. 3. *Contrast enhancement of citrus scab disease*

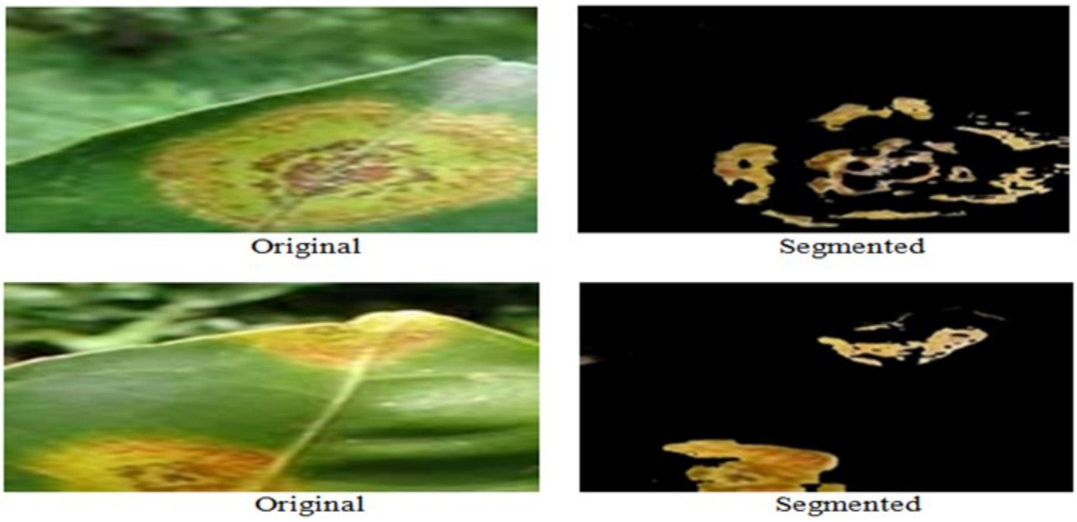

Fig. 4. *Segmented images of citrus leprosies disease*

Table 1. Different color and texture features of the citrus image

Color feature	Texture feature	Texture feature
Color Histogram	Entropy	Skewness
Color Mean	Root Mean Square	Inverse Difference Moment
Color Moment	Variance	Contrast
Standard Deviation	Smoothness	Correlation
	Kurtosis	Energy
	Dissimilarity	Homogeneity

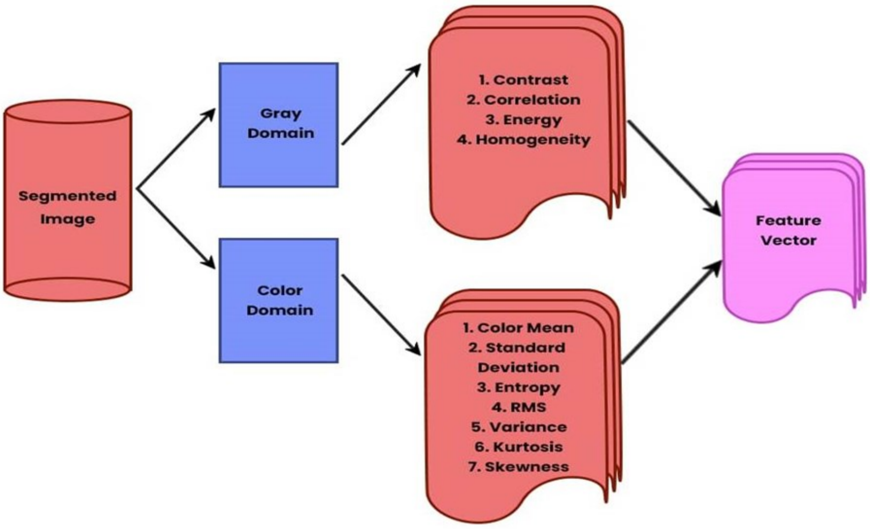

Fig. 5. *Feature extraction of citrus images*

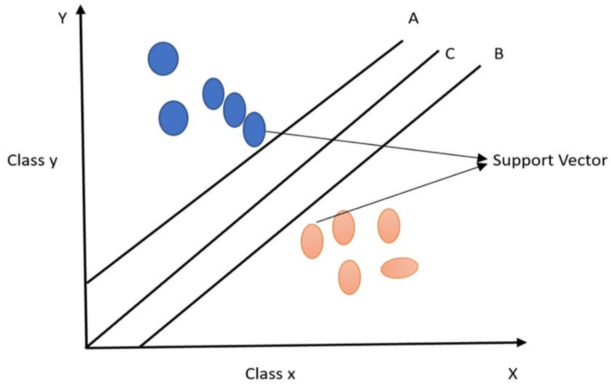

Fig. 6. *SVM for binary classification*

Table 2. Label vector of different citrus disease class

Vector position	Class labelled	Assigned result
1-43	0	Leprosis
44-144	1	Citrus Black Spot
145-222	2	Citrus Scab

tion, both color and texture features were chosen. The colors and textures of an image were identified using the GLCM (Barman *et al., 2023;* Sladojevic *et al.*, 2016; Sunny and Gandhi, 2018; S. Zhang *et al.*, 2019). The frequency with which one gray-level value appears horizontally next to another gray-level value was used to calculate GLCM. 18 different statistics of an image were provided by a single grey-level co-occurrence matrix(Poojary and Shabari, 2018; Zhang and Meng, 2011). However, the segmented image in the grayscale domain was used to construct four statistics of the GLCM features. These were contrast, correlation, energy, and homogeneity. Again, the color was determined using nine additional features, including Mean (M), Standard Deviation (SD), Entropy(E), RMS, Variance (V), Smoothness(S), Kurtosis (K), Skewness (SKW), and Inverse Difference Movement (IDM).

A single image was used to determine a total of 13 characteristics. The final feature vector of the three classes was 222x13 in size, while the size of the feature vector for a single image was 1x13. As fewer photos were in each class in the dataset, fivefold cross-validation was used to assess the SVM's accuracy. Fig.5 shows the flow chart for feature extraction.

Disease classification using SVM

A binary supervised classification algorithm is a support vector machine. A hyperplane was used in support vector machines to divide classes. Support vectors were the values closest to this hyperplane (Poojary and Shabari, 2018). However, it was later expanded to provide multi-class classification. The support vector machine for binary classification is depicted in Fig. 6.

Thirteen distinct traits were included in nonseparable form in the feature vector of the suggested technique. For dealing with nonseparable data, SVM was used with a kernel. SVM uses linear, polynomial, RBF, and sigmoid kernels. In this study, the accuracy of SVM was assessed using linear, polynomial, and RBF kernels. Using Matlab2015a, the SVM was implemented individually with the kernels. Each citrus class has a label displayed in Table 2.

RESULTS AND DISCUSSION

In Matlab 2015a, SVM was implemented using linear, Gaussian, and polynomial kernels. Apart from the linear kernel, the RBF and polynomial SVM kernels were useful for non-separable data. The kernel scale parameter was used to train the SVM model. The model's kernel scale parameter was 1. There were fewer photos in the dataset for citrus illness. Fig. 7 shows the scatter plot of the feature together with the appropriate contrast and correlation aspects of the image. Leprosy, citrus black spot, and citrus scab were each represented by the colors red, green, and blue, respectively. It demonstrated that the features of the two classes may be linearly separated, but the features of the three classes cannot.

The model's accuracy was assessed using five-fold cross-validation. The data was divided into five equal folds by the five-fold, and SVM was trained for each fold using data from outside the fold. All of the data that was within the fold was used for testing. The confusion matrix was used to assess how well the model was working. The confusion matrix is a table that displays the SVM's performance and contains correctly and incorrectly classified data. Fig. 8 depicts the SVM's confusion matrix.

Each confusion matrix (Barman *et al.*, 2023) contains the value of the True Positive Rate(TPR) and False Negative Rate(FNR). TPR means the truly classified citrus rate and FNR means the falsely classified citrus rate. This ratio defined the percentage of correctly classified data to not classified data. The linear kernel of SVM classified 10 images as Class 0, whereas 5 and 28 images were misclassified as Class 1 and Class 2, respectively. So, the TPR for Class 0 was 23.3% and the FNR was 76.7%. The TPR of Class 0 was 44.2% in the Gaussian kernel, whereas the TPR of Class 0 was 62.8% and 60.5% for the polynomial kernel with order 2 and 3, respectively. The Gaussian kernel reported 91.1% and 84.6% TPR for Class 1 and Class 2, respectively. Doh *et al.* (2019) reported a 91% TPR rate for citrus fruit disease classification and Deng *et al.* (2016) reported 88% accuracy for citrus greening disease classification using cost SVM.

The overall accuracy, error, TPR, and FNR values of three classes of citrus disease concerning the kernel are represented in Table 3.

Sunny and Gandhi (2018) reported one vs. all methods to evaluate the SVM in their study with adaptive histogram equalization enhancement techniques. The one vs all SVM model was used in the present study for each of the three kernels. For SVM's three-class classification, the Gaussian kernel and Polynomial kernel of degree 3 produced the highest accuracy (79.7%). All three SVM kernels showed lower model accuracy for Class 0 (Leprosy). However, the Polynomial kernel with degree 3 responded well to Class 0 compared to the other SVM kernels. For all three SVM kernels, the true positive rate of the other two classes (Class 1 and Class 2) was favorable.

The results of the present study were compared to some of the earlier research in Table 4. To detect citrus diseases, the SVM was compared to its many forms of the kernel. The degree 2 polynomial kernels showed the best accuracy. The raw citrus images were segmented using a threshold-based segmentation algorithm, which was demonstrated.

Conclusion

This study demonstrated a categorization method that combined machine learning with image processing. The model's overall accuracy for the three classes was

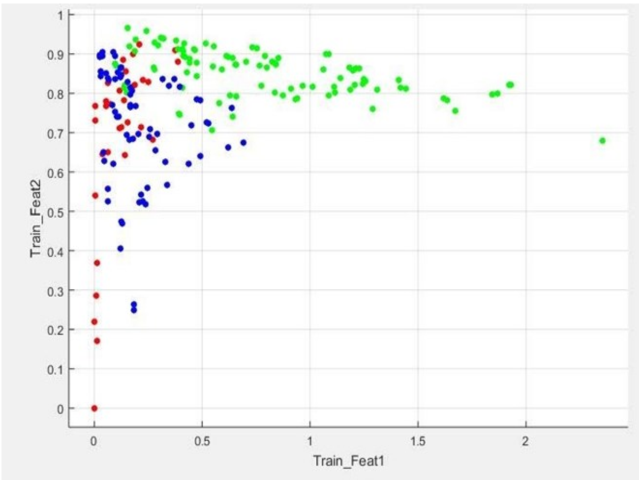

Fig. 7. *Scatter plot of the features of the citrus dataset*

Table 3. Result analysis of citrus disease class

Kernel	Class	TRP	FNR	Accuracy	Error
Linear	0	23.3	76.7	78.8	21.2
	1	89.1	10.1		
	2	96.2	3.8		
Gaussain RBF	0	44.2	55.8	79.7	20.3
	1	91.1	8.9		
	2	84.6	15.4		
Polynomial with degree 2	0	62.8	37.2	79.3	20.7
	1	90.1	9.9		
	2	75.6	24.4		
Polynomial with degree 3	0	60.5	39.5	79.7	20.3
	1	89.1	10.9		
	2	76.9	23.1		

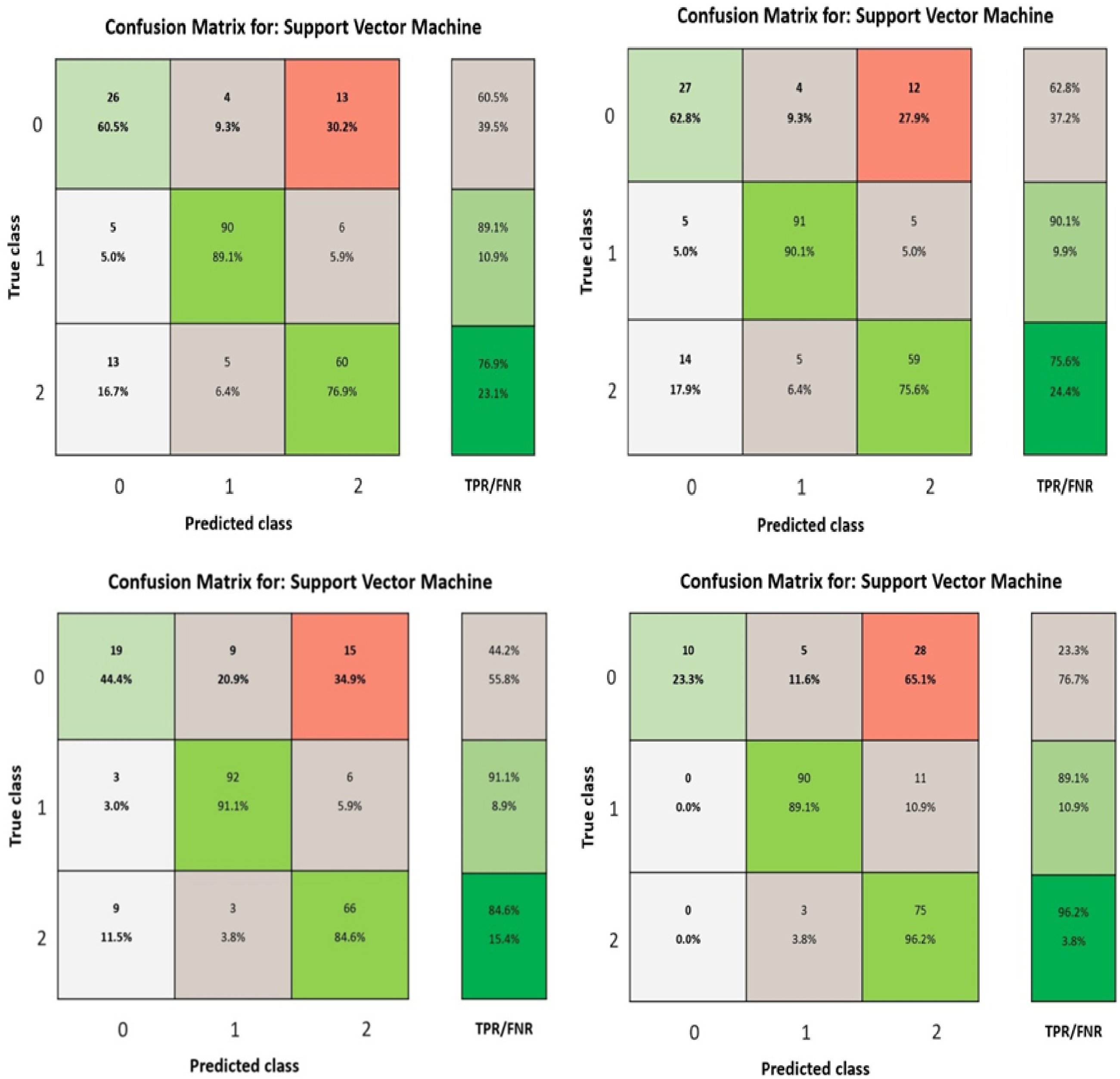

Fig. 8. *Confusion matrix of SVM with linear kernel for citrus disease classification*

Table 4. Comparison of the present study with the previous study for citrus disease detection

Paper	Dataset	Segmentation	Feature	Algorithm	Accuracy
Dhiman et al., (2022)	Citrus Fruit	NA	Deep Feature	VGG	96%
Barman & Choudhury, (2021)	Own Citrus Dataset	NA	GLCM	DNN	99%
Deng et al., (2016)	Citrus Greening	NA	Color Feature With PCA	C SVC	88.20%
Dananjayan et al., (2022)	CCL'20 Citrus dataset	NA	Deep Feature	YOLO	96%
Sharif *et al.*, (2018)	Citrus ID	NA	Hybrid Feature	Multi-SVM	89%
Proposed System	Citrus ID	Yes	Color and GLCM	SVM	79.7%

79.7%. Following pre-processing, the citrus image was segmented using HSV color histogram masking and thresholding. GLCM is used to compute citrus pictures' features. The GLCM approach effectively removed the image's color and texture components. SVM was used for the classification. Except for citrus leprosy, SVM with RBF provided good accuracy for citrus scab and citrus black spot illnesses. Future research is needed to classify other citrus diseases in conjunction with leprosy and attempt to increase the leprosy disease's accuracy by expanding the sample size.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

1. Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M. & Alrahamneh, Z. (2011). Fast and accurate detection and classification of plant diseases. *International Journal of Computer Applications*, *17*(1), 31–38.
2. Arivazhagan, S., Shebiah, R. N., Ananthi, S. & Varthini, S. V. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal*, *15*(1), 211–217.
3. Asraf, H. M., Nooritawati, M. T. & Rizam, M. S. (2012). A comparative study in kernel-based support vector machine of oil palm leaves nutrient disease. *Procedia Engineering*, *41*, 1353–1359.
4. Barman, U. & Choudhury, R. D. (2021). Smartphone assist deep neural network to detect the citrus diseases in agri-informatics. *Global Transitions Proceedings*.
5. Barman, U., Choudhury, R. D., Sahu, D. & Barman, G. G. (2020). Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease. *Computers and Electronics in Agriculture*, *177*, 105661.
6. Barman, U., Choudhury, R. D., Saud, A., Dey, S., Pratim, M. B. & Gunjan, B. G. (2018). Estimation of chlorophyll using image processing. *Int J Recent Sci Res*, *9*(3), 24850–24853.
7. Barman, U., Pathak, C. & Mazumder, N. K. (2023). Comparative assessment of Pest damage identification of coconut plant using damage texture and color analysis. *Multimedia Tools and Applications*, 1–23.
8. Dananjayan, S., Tang, Y., Zhuang, J., Hou, C. & Luo, S. (2022). Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images. *Computers and Electronics in Agriculture*, *193*, 106658.
9. Deng, X., Lan, Y., Hong, T. & Chen, J. (2016). Citrus greening detection using visible spectrum imaging and C-SVC. *Computers and Electronics in Agriculture*, *130*, 177–183.
10. Devi, T. G., Neelamegam, P. & Srinivasan, A. (2018). Plant leaf disease detection using K means segmentation. *Int. J. Pure Appl. Math*, *119*(15), 3477–3483.
11. Dhiman, P., Kukreja, V., Manoharan, P., Kaur, A., Kamruzzaman, M. M., Dhaou, I. B., & Iwendi, C. (2022). A novel deep learning model for detection of severity level of the disease in citrus fruits. *Electronics*, *11*(3), 495.
12. Doh, B., Zhang, D., Shen, Y., Hussain, F. Doh, R. F., & Ayepah, K. (2019). Automatic citrus fruit disease detection by phenotyping using machine learning. *2019 25th International Conference on Automation and Computing (ICAC)*, 1–5.
13. Gavhale, K. R., Gawande, U. & Hajari, K. O. (2014). Unhealthy region of citrus leaf detection using image processing techniques. *International Conference for Convergence for Technology-2014*, 1–6.
14. Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H. ur Rehman, M. H., & Javed, K. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Computers and Electronics in Agriculture*, *153*, 12–32.
15. Khirade, S. D. & Patil, A. B. (2015). Plant disease detection using image processing. *2015 International Conference on Computing Communication Control and Automation*, 768–771.
16. Patil, J. K. & Kumar, R. (2017). Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features. *Engineering in Agriculture, Environment and Food*, *10*(2), 69–78.
17. Poojary, H. & Shabari, S. B. (2018). A survey on plant disease detection using support vector machine. *2018 International Conference on Control, Power, Communication and Computing Technologies (ICCPCCT)*, 292–295.
18. Pujari, D., Yakkundimath, R. & Byadgi, A. S. (2016). SVM and ANN based classification of plant diseases using feature reduction technique. *IJIMAI*, *3*(7), 6–14.
19. Sharif, M., Khan, M. A., Iqbal, Z., Azam, M. F., Lali, M. I. U. & Javed, M. Y. (2018). Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Computers and Electronics in Agriculture*, *150*, 220–234.
20. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, *2016*.
21. Sunny, S. & Gandhi, M. I. (2018). An efficient citrus canker detection method based on contrast limited adaptive histogram equalization enhancement. *Int. J. Appl. Eng. Res*, *13*(1), 809–815.
22. Zhang, M. & Meng, Q. (2011). Automatic citrus canker detection from leaf images captured in field. *Pattern Recognition Letters*, *32*(15), 2036–2046.
23. Zhang, S., You, Z. & Wu, X. (2019). Plant disease leaf image segmentation based on superpixel clustering and EM algorithm. *Neural Computing and Applications*, *31*, 1225–1232.