RESEARCH ARTICLE

SPECIAL ISSUE

International Research Journal on Advanced Science Hub
2582-4376
Vol. 05, Issue 05S May
www.rspsciencehub.com

Check for updates

RSP Science Hub

http://dx.doi.org/10.47392/irjash.2023.S070

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

Anthracnose Disease Detection in Cashew Leaf Using Machine Learning Technique Based on Contour Detection and Principal Component Analysis

Sudha P [1], Kumaran P [2]

[1]Research Scholar, Department of Computer Science and Engineering, National Institute of Technology Puducherry, India
[2]Assistant Professor, Department of Computer Science and Engineering, National Institute of Technology Puducherry, India

Email: viswasudha@gmail.com

Article History

Received: 28 February 2023
Accepted: 21 March 2023

***Keywords*:**

Cashew Leaf;
PCA;
Contour Detection;
Machine Learning Technique

Abstract

Detecting and classifying leaf diseases in cashew crops is critical for farmers to find pest and disease infections. Cashew leaf diseases can reduce productivity if not detected early. Creating an automated method utilizing image processing for leaf disease identification decreases time and expense and primarily contributes to a rise in cashew nut yield. For image segmentation, canny edge detection and an active contour model are utilized. A feature extraction method, Principal Component Analysis (PCA), is applied when the contour has been applied. After the features have been extracted, they are submitted for categorization. This study analyzed several classifiers' accuracy, precision, and recall values. These classifiers included Random Forest, SVM, KNN, and Naive Bayes. This research tries to answer whether a machine learning classifier provides the best results when the diseased area is divided using the canny edge detection and contour detection technique.

1. Introduction

The growth and profitability of India's agriculture sector are the primary drivers of the country's economy. The technique of identifying each leaf disease in agricultural applications is the most difficult. Image processing techniques are becoming increasingly popular as a beneficial tool for raising agricultural efficiency, and farming practices, boosting procedure precision and quality whilst simultaneously minimizing the amount of human monitoring performed by farmers.

Studies on leaf-based classification and the detection of disease leaf images are also vital for identifying plant diseases. Detecting edge bases and effectively classifying data can be challenging because of noise and classification errors. To get better classification performance, we will first need to construct an accurate model (Greeshma, Balakrishnan, et al. Tulshan, Raul, et al.). Diseases that affect crop leaves can range in size, shape, and appearance. Some diseases have the same color but different shapes, while others have different shapes but the same color. Yet others have different colors but the same form. Machine learning methods are frequently utilized to recognize the photos of the afflicted leaves. In this study, various machine learning algorithms that are used to detect whether a plant is afflicted with a disease are described. These algorithms determine whether a plant has a disease. This was accomplished through a series of phases, including image acquisition, feature extraction, sickness classification, and the results displayed (Varshney et al.).

This work determines the best machine learning classifier for partitioning the sick region using canny edge and contour detection. Subsequent feature extraction using PCA is applied, then the data is categorized. The rest of this study is organized as 2 Literature Review, 3Proposed Model, 4 Result and Discussion,5 Conclusion, and Future Scope.

2. Literature Review

Tulshan, A. S., et al. (2019) the collected data is preprocessed, segmented, and then feature extraction is applied to the data and classified using K Nearest Neighbor (KNN). Predicting leaf diseases in plants using the proposed implementation has shown an accuracy of 98.56 percent. Further data on a leaf disease in a plant are also presented, including the Accuracy, Affected Area, Sensitivity, Disease Name, and Elapsed Time (Tulshan, Raul, et al.).

Zamani, A. S., et al. (2022) describe a paradigm for identifying leaf sickness. This framework can take as input a picture of a leaf. To begin with, noise is removed from leaf pictures during preprocessing. To eliminate ambient noise, use the mean filter. The image quality is improved via histogram equalization. In photography, segmentation splits a single image into many parts or segments. It helps to define the limits of the image. The K-Means method is employed to segment the image. The principal component analysis is used to carry out feature extraction. The next step is categorizing the photos using random forest, RBF-SVM, ID3, and SVM (Zamani et al.).

Syed-Ab-Rahman et al. (2022) In order to detect plant diseases and classify citrus diseases from leaf photos, this work employs a dual-stage deep Classifier. The proposed model has two primary steps: (i) identifying unhealthy areas using a region proposal network, and (ii) assigning the most likely target area to the appropriate disease class via a classifier. In terms of detection, the proposed model achieves an accuracy of 94.37 percent, with an average precision of 95.8 percent (Farhana Syed-Ab-Rahman, Hesamian, and Prasad).

Prabu, M., et al. (2022) suggested a mango leaf disease classification structure. The framework comprises data preparation, feature selection, learning and classification, and performance evaluation. We chose 380 healthy and sick photos (Mango Anthracnose, Bacterial black spot, and Sooty mold). Data augmentation methods reduce overfitting and improve generalization. Next, a crossover-based levy flight distribution convolutional neural network improves feature selection. The pre-trained MobileNetV2 model is employed for learning, and at the end, a support vector machine classifies diseases (Prabu and Chelliah).

Kumar, V. V., et al. (2022) described, identified, and quantified paddy plant crop diseases such as brown spots, bacterial blight, and leaf blasts. Risk analysis of paddy crop leaf images detects and recognizes. Deep convolutional neural networks (DCNNs) and fuzzy logic are used in our Deep Convolutional Neuro-Fuzzy Method (DCNFM). Fuzzy logic and DCNNs help the synthesis extract critical features from unstructured input (V. V. Kumar et al.).

Al-gaashani, M. S., et al.(2022) proposed transfer learning and feature concatenation are used to classify tomato leaf diseases. Kernel principal component analysis concatenates and reduces the dimensionality of MobileNetV2 and NASNetMobile's pre-trained kernels (weights). They then integrate these features into a standard learning algorithm. Concatenated features improve classifier performance according to experiments. Multinomial logistic regression outperformed random forest, support vector machine, and multinomial regression with an average accuracy of 97% (Al-gaashani et al.).

Ali, S., Hassan, et al. (Ali et al.) This research presented potato disease detection using PCA-LDA classification and Feature Fusion (FF-PCA-LDA). RGB images yield bespoke hybrid and deep characteristics. TL-ResNet50 extracts deep features. Fused hybrid and deep features are handcrafted. After fusing picture features, PCA selects the most discriminant properties for LDA (Kruse et al. B. Sharma, V. K. Sharma, and S. ; M. Kumar Fraiwan et al.) model development and provides 98.20% accuracy.

Caglayan, A. et al. (Caglayan, Guclu, and Can) of this paper, show how images of leaves can be used to identify plants. Classification algorithms like k-Nearest Neighbor, Support Vector Machines, Naive Bayes, and Random Forest [17] use the shape and color of leaf images to determine what kind of plant it is. The method shown here is tested on 1897 images of leaves and 32 different kinds of leaves. The results showed that the Random Forest method could help people recognize plants up to 96% more often when both shape and color features are used.

Gulhane, V. A. et al. (Gulhane and Kolekar) uses the Nearest Neighbor Classifier (NNC) and Principal Component Analysis (PCA) to diagnose illnesses on cotton leaves. The statistical information for the Green (G) channel of an RGB image can be examined once PCA/KNN multi-variable algorithms have been implemented. As diseases or elemental deficiencies are reflected accurately by the green channel, it is considered for reliable feature gathering. PCA/KNN-based classifiers have been observed to have a 95% accuracy rate in classification.

Sujith, A. et al. (Sujith and Aji) The suggested approach provides an ideal feature set developed via features extraction methods using a Local Binary Pattern (LBP), Gray Level Co-Occurrence Matrix (GLCM) (Fraiwan et al.), and a Histogram of Oriented Gradients (HOG). Neighborhood Component Analysis optimizes the combined feature vector (NCA). The classification performance and computational efficiency have been improved using feature selection, and dimensionality reduction approaches. The experiment's proposed method has an average classification accuracy of 97.63% in 291.24 seconds of calculation time, using three plant datasets: Flavia, D-Leaf, and Swedish Leaves.

3. Proposed Model

This section describes the proposed model of the study. Here, we begin with dataset collection, which contains 100 healthy and 100 infected leaves. After data augmentation, we have 850 healthy and 850 diseased leaves in our Cashew Crop Diseased Database (CCDDB). dataset. Data transformations such as scaling, rotation, and flipping are carried out in Pre-processing.

Canny edge detection is carried out for segmentation. Image segmentation, feature extraction, and classification of diseased cashew leaves are all shown in Fig. 1 of the proposed model.

Edge detection: A crucial task in pattern recognition is edge detection. Identifying borders across regions with various attributes, such as intensity or texture, can be characterized as it.

Canny edge detection: The canny edge detector is a multi-stage technique for image edge detection. The 1986 publication "A computational method to edge detection" by John F. Canny introduced it.

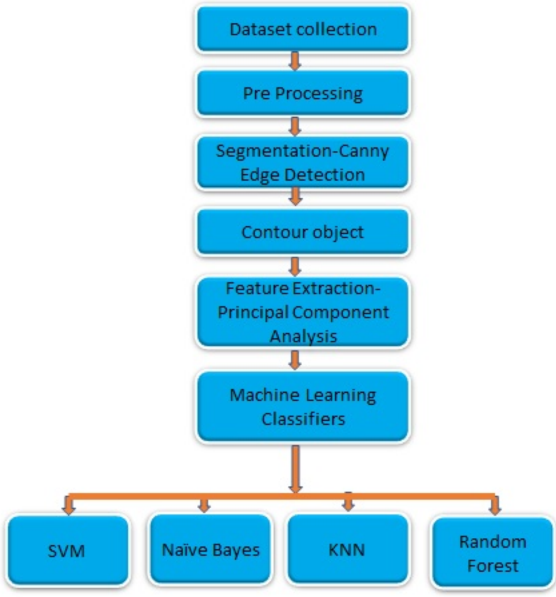

FIGURE 1. Proposed Model for diseased cashew leaf classification

Finding edges in an image is, in a nutshell, the technique of edge detection. An edge is often a sharp change in color from one image pixel to the next, like from black to white. It is one of the most common ways to find edges. It is also popular because it gives good results.

There are four stages involved in the Canny edge detection algorithm:

- Noise reduction accomplished by blurring the image with a Gaussian function.
- Doing calculations on the image's intensity gradients.
- The smoothing out of the edges.
- Utilizing a method known as hysteresis thresholding

First, we import our dataset, then convert it to grayscale, and last, we use the cv2.GaussianBlur filter to blur the image and get rid of the noise. After that, we use the cv2.canny function to implement the Canny edge detector. This function calls for six parameters, three of which are necessary. In our instance, we made use of only the mandatory parameters. The image containing the edges we wish to detect passed in as the first argument. The hysteresis technique uses two thresholds; the second and third arguments are those thresholds.

Figure 2(a) illustrates an example of a leaf image used as input for the proposed model. Before carrying out the canny edge detection procedure, the input leaf needs to be transformed into a grayscale

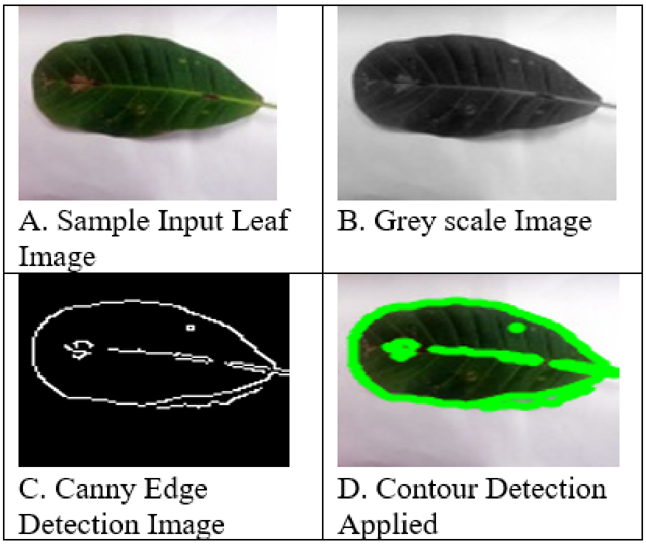

FIGURE 2. **Experiment on Sample image**

image, as demonstrated in figure 2. (b). Fig. 2(c) displays the canny edge detection, and the contour detection applied to the original leaf is displayed in Fig. 2(d). We can see that the algorithm prioritized the edges it considered most significant. Experiment with varying the threshold values to observe how that affects the detection of edges.

Contour Detection Technique: ”contour” refers to a curve connecting all points on a shape’s perimeter. In binary images, contours can be recognized with great accuracy. As a result, each image must be converted to grayscale and then have a threshold set to it. cv2.findContours function takes three arguments: the source image, the contour retrieval mode, and the contour approximation technique. To determine the shapes of the objects, we employed the binary image produced by the Canny edge detector. Hierarchy is kept in RETR TREE. The function’s output includes photos, contours, and hierarchy. All image contours are included in the output.

Principal Component Analysis: PCA reduces the number of variables while keeping the underlying structure and patterns intact in high-dimensional datasets. The purpose is to identify and isolate the most salient aspects of the data, which will then be expressed as a collection of summary indexes known as principal components. By compressing the data into fewer dimensions that serve as feature summaries.

Classifiers: This study has 4 different classifiers: Random Forest, SVM, KNN, and Naive Bayes. Support vector machines, sometimes known as SVMs, are a group of supervised learning methods that can be used for detecting outliers, as well as for classification and regression. Effective in environments with a high number of dimensions. Even when the number of dimensions exceeds the number of samples, the method is still effective. The RF depends on many self-learning decision trees forming a ”Forest.” Compared to a single DT, an ensemble of several decision trees (also known as an ensemble) can reach a sound and reliable conclusion more than a single DT alone. The K-Nearest Neighbors algorithm assumes that the new case or data is comparable to existing cases and places the new example into the category that is the most similar to the categories that are already accessible. The K-NN algorithm remembers all accessible data and determines how to categorize a new point depending on its similarity to the stored data. This indicates that when fresh data becomes available, it may be quickly sorted into a well-suited category using the K-NN method. The Nave Bayes classifier is a supervised machine learning algorithm. It also belongs to the family of generative learning algorithms, which models the input distribution of a particular class or category.

4. Result and Discussion

This section describes the result and discussion of this study. This study has 4 different classifiers: Random Forest, SVM, KNN, and Naive Bayes. Out of which Naïve byes outperforms other classifiers.

During Contour detection, the convergence criteria, which are utilized in the energy reduction procedure, determine the accuracy of the results. A higher level of precision calls for more stringent convergence criteria, leading to longer computation durations, due to this reason, the classifier’s performance is not satisfactory. In decreasing the energy throughout their outlines, they frequently fail to notice minute details because of the importance of this goal.

From Table 1, we can identify the accuracy of Naïve Bayes is comparatively reasonable.

The accuracy graph of four machine learning classifiers is shown in Fig. 4, where Naive Bayes provides the best results.

5. Conclusion and Future Scope

This study investigates which machine learning classifier performs well when the canny edge detection

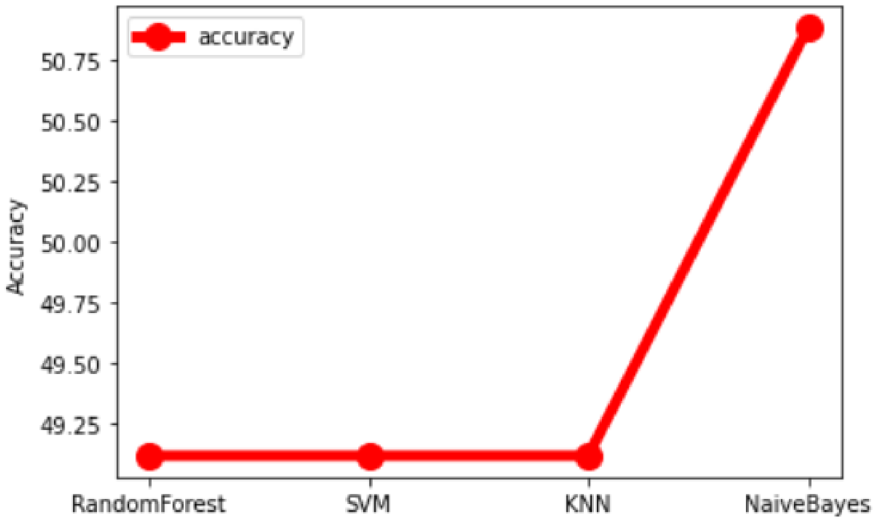

FIGURE 3. **Accuracy Graph of Different Machine Learning Classifiers**

TABLE 1. **Metrics of different Machine Learning Classifier**

	RandomForest	SVM	KNN	Naïve Bayes
Accuracy %	49.12	49.12	49.12	50.88
Precision%	100.0	24.13	24.13	25.89
Recall %	49.12	49.12	49.12	50.88

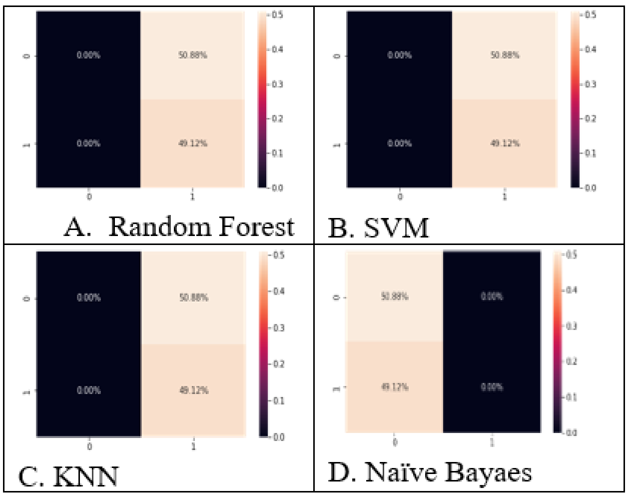

FIGURE 4. **Confusion Matrix of Different Machine Learning Classifiers**

coupled with the contour detection algorithm is performed to partition the diseased region. After that, the PCA method is used for feature extraction, and the data is then classified. When performing canny edge detection on images, they become less distinct due to the gaussian smoothing, which has the same effect on the edges. When there is a significant difference in brightness level between the foreground items and the image's background, the contour detection technique performs remarkably well. It frequently becomes gets trapped in a local minimum. When working with images of a large size, this method performs more slowly. In this study, Naïve Bayes performs better than other classifiers. In the future, we will establish an effective segmentation and feature extraction method to classify diseased cashew leaf images.

Authors' Note

The authors declare that there is no conflict of interest regarding the publication of this article.

The authors confirmed that the paper was free of plagiarism.

References

Ali, Safdar, et al. "FF-PCA-LDA: Intelligent Feature Fusion Based PCA-LDA Classification System for Plant Leaf Diseases". *Applied Sciences* 12.7 (): 3514–3514.

Caglayan, Ali, Oguzhan Guclu, and Ahmet Burak Can. "A Plant Recognition Approach Using Shape and Color Features in Leaf Images". *Image Analysis and Processing – ICIAP 2013*. Springer Berlin Heidelberg, 2013. 161–170.

Farhana Syed-Ab-Rahman, Sharifah, Mohammad Hesam Hesamian, and Mukesh Prasad. “Citrus disease detection and classification using end-to-end anchor-based deep learning model”. *Applied Intelligence* 52.1 (2022): 927–938.

Fraiwan, Luay, et al. “Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers”. *Biocybernetics and Biomedical Engineering* 41.1 (2021): 1–14.

Al-gaashani, Mehdhar S A M, et al. “Tomato leaf disease classification by exploiting transfer learning and feature concatenation”. *IET Image Processing* 16.3 (2022): 913–925.

Greeshma, M O and Sasikala, P, S G Balakrishnan, et al. “” *Leaf Disease Classification Based on Edge Detection Using Training Neural Network. Solid State Technology* 63 (2020): 20765–20776.

Gulhane, Viraj A and Maheshkumar H Kolekar. “Diagnosis of diseases on cotton leaves using principal component analysis classifier”. *2014 Annual IEEE India Conference (INDICON)* (2014): 1–5.

Kruse, Ole Mathis Opstad, et al. “Pixel classification methods for identifying and quantifying leaf surface injury from digital images”. *Computers and Electronics in Agriculture* 108 (2014): 155–165.

Kumar, V Vinoth, et al. “Paddy Plant Disease Recognition, Risk Analysis, and Classification Using Deep Convolution Neuro-Fuzzy Network”. *Journal of Mobile Multimedia* (2022).

Prabu, M and Balika J Chelliah. “Mango leaf disease identification and classification using a CNN architecture optimized by crossover-based levy flight distribution algorithm”. *Neural Computing and Applications* 34.9 (2022): 7311–7324.

Sharma, Basudev, Vivek Kumar Sharma, and Sandeep ; M Kumar. “Sigmoidal Spider Monkey Optimization Algorithm”. *Advances in Intelligent Systems and Computing* 1053 (2020): 109–117.

Sujith, A and S Aji. “An Optimal Feature Set With LBP for Leaf Image Classification”. *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (2020): 220–225.

Tulshan, Amrita S, Nataasha Raul, et al. “Plant Leaf Disease Detection using Machine Learning”. *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (2019): 1–6.

Varshney, Divyanshu, et al. “Machine Learning Techniques for Plant Disease Detection”. *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2021. 1574–1581.

Zamani, Abu Sarwar, et al. “Performance of Machine Learning and Image Processing in Plant Leaf Disease Detection”. *Journal of Food Quality* 2022 (2022): 1–7.

Embargo period: The article has no embargo period.

To cite this Article: , Sudha P, and Kumaran P . “**Anthracnose Disease Detection in Cashew Leaf Using Machine Learning Technique Based on Contour Detection and Principal Component Analysis**.” International Research Journal on Advanced Science Hub 05.05S May (2023): 520–525. http://dx.doi.org/10.47392/irjash.2023.S070