

Leaf and skin disease detection using image processing

Manjunatha Badiger^{a,*}, Varuna Kumara^b, Sachin C N Shetty^c, Sudhir Poojary^d

^a Department of ECE, Sahyadri College of Engineering & Management, India

^b Department of ECE, Moodlakte Institute of Technology, India

^c Department of ECE, Sahyadri College of Engineering & Management, India

^d Department of ECE, Moodlakte Institute of Technology, India

ARTICLE INFO

Keywords:

Algorithm
Classification
Feature extraction
Plant leaf diseases
Segmentation
Training

ABSTRACT

Agricultural production is something on which the economy significantly relies. Leaf diseases in agriculture are the key issue for every nation, as the food demand is expanding at a rapid speed due to a rise in population. Skin disorders are usually seen in animals and humans, it is a particular sort of illness caused by germs or infection. Early and accurate identification and diagnosis of leaf and skin diseases are vital to keeping them from spreading. Image processing techniques can be used for disease detection which involves mathematical equations and mathematical transformations. For humans eyes image is a mixture of RGB colour, because of these colours we can extract some of the features from the image, but modern computer stores image in a mathematical format which means computer sees the image as numbers, hence after evaluating the image as a number arrays or matrix we will perform various transforms on them, these transforms will extract specific details from the picture, before transforming the image must go under various operation like feature adjustment which is also carried out mathematically. The project is implemented using K-Means Clustering and Support Vector Machine Algorithm in MATLAB through which we can detect and distinguish different types of leaf and skin diseases.

Introduction

Image is a mixture of RGB colours for normal eyes, but images are real numbers inside a matrix for digital devices such as computers and cameras. Since modern images exhibit digital and numerical characteristics, we can apply some mathematical transformation and by adjusting their parameters we can extract some of the hidden details on them [1]. Based on those extracted features, we can do further research like calculating a man's age by processing face and extracting skin data, etc. Images are regarded as numbers within the matrix, and the magnitude of these numbers determines the colour of that pixel in the real world. To extract a little information, the precise numbers within that matrix should be highlighted by using a few algorithms and mathematically changing the matrix. However, it is not possible to detect the nature of the sickness by extracting the features; instead, the analysis must be done on the collected data, which takes time and involves a human operation. Advancements in image processing technology and algorithms can be used in medical science to determine the various diseases and their stages. Some of the diseases can be visually inspected and some of them require both visual inspection and medical tastings for confirmation [2]. As computers can adjust the parameters of images, they can accurately differentiate some parts on images as abnormal based on training data and hence this saves time. The key feature of AI is that

after training this will learn on its own hence for new data types this will offer a dynamic response. And, in today's world, AI is employed everywhere, and in the future, all AI-based systems may be integrated as one, making the interface simple.

Initially, Disease Detection System (DDS) is trained by providing numerical data of different disease types. In the disease detection mode, DDS will constantly compare sample data with training data and provide an output based on less error between the two data types. In other words, it will provide a result based on a higher likelihood of closed approximation. Indian agriculture has a rich history, and it contributes to the country's economic strength. However, as new types of diseases are emerging and tests take a long time and are expensive, many farmers are reluctant to do these tests, and as a result, they may suffer losses [3]. Similarly, skin illnesses are spreading at an alarming rate. Most of us will neglect minor skin diseases due to our hectic lifestyles and the initial cost of the test. Later, these minor diseases will cause substantial issues, therefore we may use image processing-based detectors to determine the type of the sickness in the primary stage and take measures or therapy on primary days to avoid health concerns.

Literature survey

Sachin D. Khirade and A.B. Patil [4] proposed a method in which leaf disease can be detected by providing green leaves, the method is based

* Corresponding author.

E-mail address: badiger_manju@yahoo.com (M. Badiger).

on K-Means clustering and some auxiliary algorithms the detection steps involves image acquisition after that pre-processing stage takes over and does some basic transforms after that image segmentation takes place which is carried out by K-Means cluster algorithm then classification is carried out by ANN block also this paper gives some mathematics behind mathematical transforms.

The paper proposed by Leon Bottou and Yoshua Bengio [5] gives all details about popular K-Means clustering especially this paper described as gradient descent algorithm and/or this can also be described by improving mathematics of EM algorithm. Also, they proved that quantization error is minimum, the reason for this is due to the usage of a very fast Newton algorithm.

The paper proposed by P.K. Agarwal and C.M. Procopiuc [6] proposes an algorithm for solving the K-Centre problem in various system-related matrixes, they showed that the algorithm can also be extended to other matrixes and by solving them we can get a solution to discreet K-Centre problems. also, they described a simple $(1 + l)$ approximation algorithm for the K-Centre problem.

The paper proposed by Ahmed S. Abljaleb [7] has the main goal of extending entropy-based thresholding technique for 2D histogram, also study of individual pixels' grey level value and neighbourhoods average value is carried out. This proves that the threshold is a vector that has two access points, the first one is pixels' grey level and the second one is its average value. Also, this proposed method works very efficiently when the noise level is small.

The paper proposed by Nidhal K. El Abbadi, Nizar Saadi Dahir, Muhsin A. AL-Dhalimi, and Hind Restom [8], proposed a system for detecting skin disease using ANN. The key to detecting disease is colour of skin and GLCM and based on training data and colour detection they successfully detected psoriasis skin disease and the algorithm omitted healthy skin.

The paper proposed by Dr. S.Arivazhagan, Mrs. R.Newlin Shebiah, Ms. K.Divya, Ms. M.P.Subadevi [9] implemented a human skin disease detection system that works on the principle of an automated system based on texture analysis. the skin element melanin and haemoglobin distributions are differentiated by independent component analysis which is based on skin colour, the grey level run-length matrixes are used to derive texture features.

The paper proposed by S. Kolkur1, D. Kalbande, P. Shimpi, C. Batpat2, and J. Jata-kia [10] implemented RGB colour based human skin disease detection system, here three primary colours RGB, Hue, saturation, value, Chrominance, and Luminance are the base for detection, their main goal is to detect pixel of given image very efficiently. The algorithm is designed to consider both primary colour and combinational range which naturally increases accuracy in recognizing the affected area.

Shima Ramesh and Ramachandra Hebbar [11] proposed that Plant infections are a substantial danger to sustenance security, but their rapid distinguishing verification remains problematic in many parts of the globe because of the non-presence of the fundamental foundation. The advancement of accurate algorithms in the area of leaf-based image categorization has demonstrated outstanding results. This article makes use of Random Forest in differentiating between healthy and sick leaves using the data sets provided. Their suggested work comprises multiple stages of implementation like dataset generation, feature extraction, training the classifier, and classification. The produced datasets of infected and healthy leaves are jointly trained under Random Forest to categorize the infected and healthy images. For extracting characteristics of an image, they utilize a Histogram of an oriented Gradient (HOG). Applying machine learning to train the massive data sets accessible publicly offers a clear technique to identify the illness existing in plants at a gigantic scale.

Monzurul Islam and Monzurul Islam [19] Modern phenotyping and plant disease detection give hopeful steps towards food security and sustainable agriculture. In particular, image and computer vision-based phenotyping gives the capacity to analyse quantitative plant physiology.

On the contrary, manual interpretation demands a significant amount of effort, knowledge in plant diseases, and also requires excessive processing time [20]. In this study, we describe a technique that mixes image processing and machine learning to enable identifying illnesses from leaf pictures. This automated technique identifies illnesses on potato plants using a publicly accessible plant picture collection named 'Plant Village' [21]. Our segmentation technique and application of support vector machine exhibit illness classification over 300 images with an accuracy of 95%. Thus, the suggested technique gives way toward automated plant diseases identification on a huge scale.

Proposed methodology

The basic control flow diagram of the disease detection system is shown in Fig 1. The system is mainly used to detect the disease of leaves and skins. The flowchart of the disease detection system comprises of

Input image

For performing image processing the presence of the image is the primary requirement, an appropriate image with appropriate size and resolutions must be provided, the image can be loaded from the root folder, or for live image processing we can interface an external camera to our system after capturing the image can be loaded to the system as the primary input. If required we can use an online image library as an image source here the images are stored in cloud storage by accessing them, we can directly load the image, without the image the system decides input as a null matrix.

Pre-processing

The Disease Detection System converts the input image into a two-dimensional matrix expressed in RGB format. Based on the magnitude of

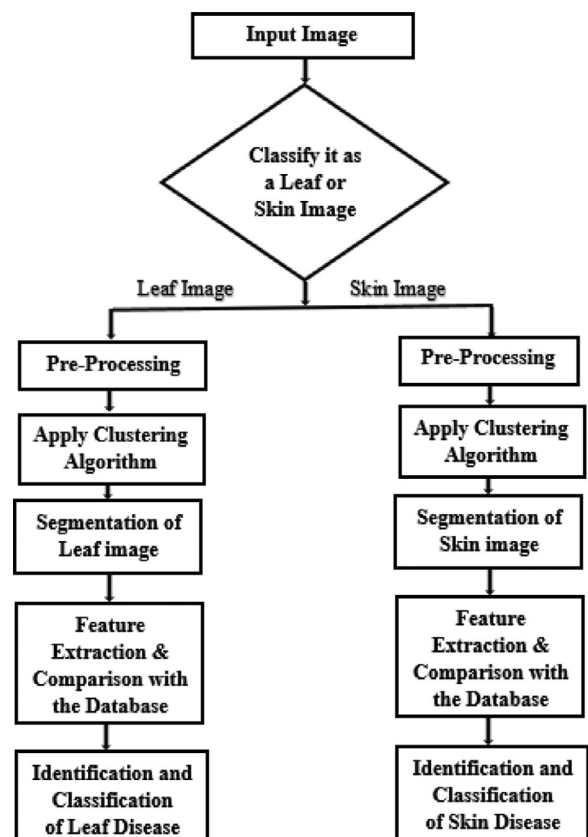


Fig. 1. Flow chart of Disease Detection System.

the numbers inside the matrix, it is possible to determine the prevailing colour of the input image as well as to detect and classify the image type as skin or leaf. If $R > G$, the image is of the skin type; otherwise, it is of the leaf type. Then input image is converted to LAB colour space from standard RGB format. All the images are finally reduced to a consistent size.

Clustering and segmentation of image

The input image has all of the required information for the processing. But the underlying difficulty is that the impacted area may be anywhere on the image. To detect the damaged area K-means algorithm is used which splits the whole image into tiny sections and then does image processing on each component. If any unaffected areas are discovered it excludes them from consideration. If it detects the impacted areas, it reserves them for further analysis. The K-means algorithm is an iterative technique that segments the dataset into K pre-defined discrete non-overlapping subgroups.

Extraction of feature & comparison with database

This is one of the most important steps in the flow of disease detection. Feature extraction refers to the process of turning raw image data into numerical features. This lowers the number of resources required to represent the data and may be processed while maintaining the information of the original data set to offer superior outcomes. The image components like colour, intensity, etc., are adjusted such that the hidden features of the diseases get highlighted. These extracted features will be helpful to detect disease very quickly and efficiently when compared with the test database of skin or leaf disease.

Identification and classification of disease

An image of type either diseased leaf or diseased skin is given as input to the disease detection system. SVM Classifier is used for classification purposes. If the input image is of skin disease, the system will classify it into Melanocytic naevus, Basal-cell carcinoma, and Actinic keratosis. If the input image is of leaf disease, the system will identify and classify it to Alternaria Alternata, Anthracnose, Bacterial Blight, Cercospora Leaf Spot, and Healthy Leaf.

Algorithm description

Algorithms are sets of instructions that accomplish tasks like arithmetic, data processing, automated reasoning, and automated decision-making. The majority of machine learning solutions are created and deployed using off-the-shelf machine learning algorithms with small tweaks. Some of the algorithms utilised in the DSS are listed below.

K - Means Clustering

It is a centroid based iterative algorithm, where K represents the number of clusters and at the beginning, for $K = N$ it creates a centroid C_N and the algorithm starts at some random point C_m , where m is a random number that lies between zero and N . This is used as a prime technique in our proposed design and this algorithm method was first suggested by Hugo Steinhuas in the year 1956 but the technique was modified and shaped by Stuart Lloyd in the year 1956. This method is widely used in image processing and data mining. In this algorithm smaller the index of the cluster better is the speed and performance. To reduce this index error criterion and square errors are used as the base of this algorithm. At the beginning of the process, it will select some points on the interesting area this point of interest indicates the focal point of the initial cluster. After this it will continue its process concerning remaining part to their focal point if distance vector is minimum

then its output will be indicated initial classification, else again modification to the initial stage is made by recalculating focal points of each cluster this process will take place till we get a satisfying result, this method is generally simple and reliable [12]. This K-cluster algorithm heavily depends upon initial points, since points of selection are random each time the produced outcome will be different. For extremum, this algorithm uses gradient method or in other words, this algorithm also depends upon target function, the gradient method's direction search always travels along the axis which is nothing but the direction of energy decrease. In other words, if a selection of the initial point is not proper then the whole process or algorithm becomes the local minimum point [13]. This algorithm uses an objective function which is given by

$$J = \sum_{i=1}^c \sum_{j=1}^k (\|x_i - v_j\|)^2 \quad (1)$$

Where

x_i, v_j is the Euclidean distance between x_i and v_j .

c = Number of clusters

k = Number of data points

The algorithm flow is described in the following steps:

Initialize the centroid and randomly select c cluster centre then calculate $x_i - v_j$ for different values of i & j .

Now assign a data point to a cluster centre such that the distance between the data point and that cluster is minimum compared to other cluster centroids.

For assigned data point recalculate centre using the equation:

$$v_j = \frac{1}{c_i} \sum_{x_i \in v_j} x_i \quad (2)$$

Where c_i is the number of data points in the i^{th} cluster.

Now distance value for the recalculated centre and the cluster can be calculated. If the distance is very minimum or it is zero, stop the iteration otherwise repeat the procedure from step 2.

Otsu threshold algorithm

The working of this algorithm is mainly dependant on a set threshold value, if a certain part of the image falls below the threshold value, then those parts are represented by binary zero, and the parts which are above thresholds are represented by binary one, thus this algorithm is used to transform greyscale images into a binary image [14]. Here each pixel will hold some numerical values from 0 to 256, the value of these pixels indicates the intensity of that pixel, so to highlight or remove that pixel we set threshold values. Once the threshold values are set, the intensity of that pixel may be modified based on that value, which aids in the removal of some of the portions of the input picture. This approach is a prominent technique in image processing because of this feature. This approach is usually inferred when the picture is converted to a greyscale image or when the image is converted to a binary image. The threshold value will be determined based on the region of interest and it will now be extremely easy to separate the undesirable area from the interested area.

The Otsu algorithm is as follows:

Since the output is a two-state binary distinguish pixel into two clusters.

Now calculate the mean values of both clusters and square the subtraction values of them.

If m and n hold the pixel values of the individual cluster, then multiply both m and n .

Boundary and spot detection algorithm

Boundary detection is a prominent approach in image processing that may be applied in the identification of an item, type, and segmentation in an image. The boundary or edge may be determined by detecting a change in pixel values across a range of integers. This may be utilized as a pre-processing technique in disease diagnosis since the

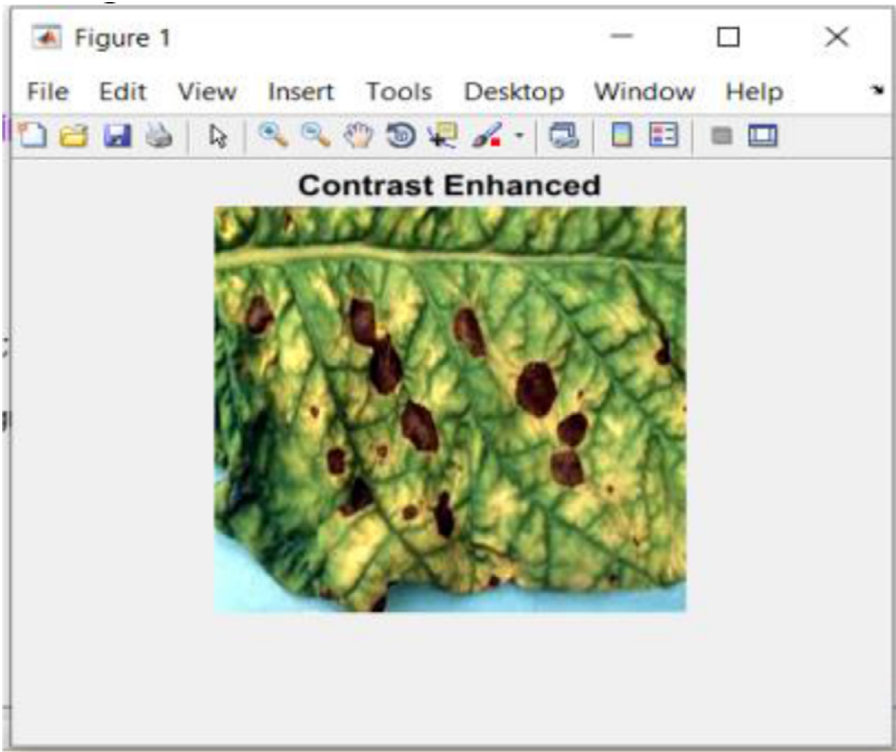


Fig. 2. The enhanced version of the input.

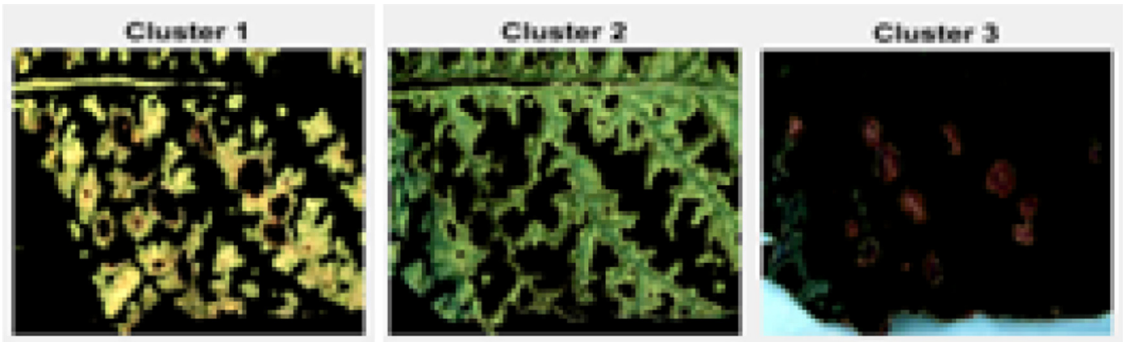


Fig. 3. After K-Means cluster.

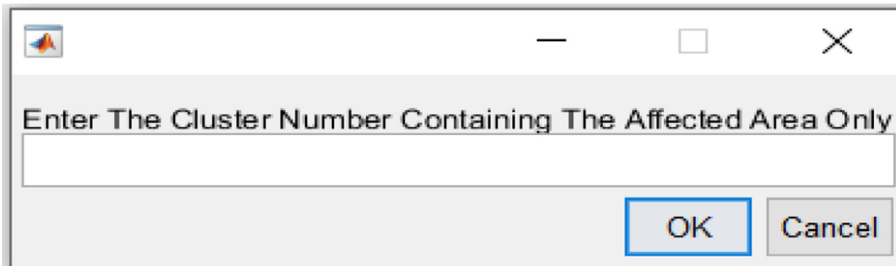


Fig. 4. Entry of afflicted area's cluster number.

impacted portion separation is the initial stage in disease identification. Hence by identifying the edge and by deleting the remainder of the component, we may sequentially extract that section. The impacted region detection is based on the change of colour of a target item [15]. Even though it is a pre-processing tool plenty of calculations have to be done. This algorithm converts the provided colour picture into three sections called Hue, Intensity, and saturation (HIS) model. With the aid of this enhanced image, the algorithm quickly detects the border and spot of a specified colour for the provided image. Hence, it's possible to determine the damaged region. Also, this method helps to separate the

background from the target item hence it changes the raw image into the processable image.

GLCM algorithm

Any RGB image can be resolved into HSI (Hue, Saturation and Intensity) model, we can extract specific features by adjusting the RGB values. The grey level Co-occurrence matrices simply known as GLCM is one of the widely used algorithms in image processing to extract the textural aspects of the input image [16]. This is one of the widely used methods

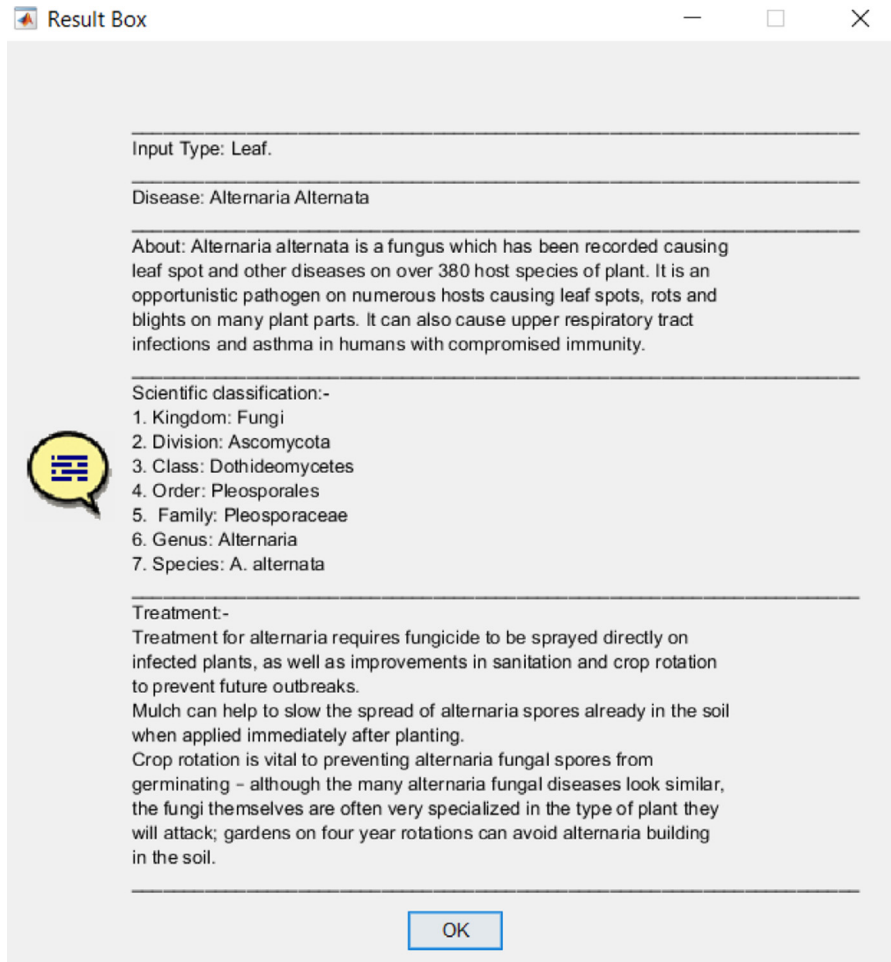


Fig. 5. The result window.

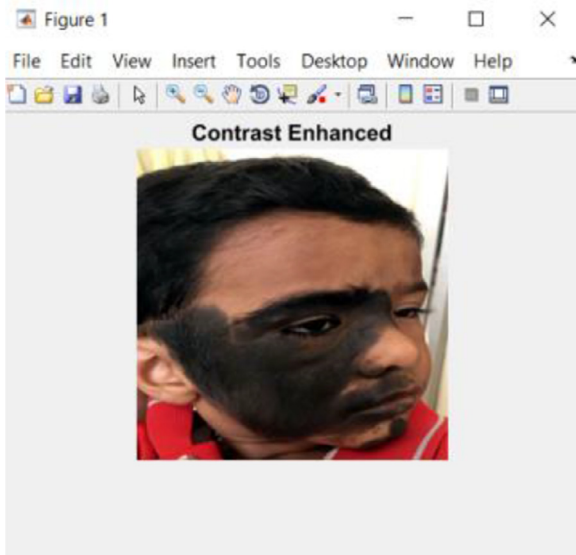


Fig. 6. The enhanced version of the input.

in the analysis of image texture. This also provides multidimensional co-occurrence matrices that are often utilised in object identification and comparison. Hence this is a beneficial approach when clustering is essential. This is mathematically described in the following equations:

$$H = \{\theta; B(G \ 360 - \theta; B)G\} \quad (3)$$

Where

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-B)(R-G)]}{\left[(R-G)^2 + (R-B)(G-B)^{\frac{1}{2}} \right]} \right\} \quad (4)$$

The Saturation(S) and Intensity(I) value can be given by

$$S = \frac{3}{[R + G + B]} \min [R, G, B] \quad (5)$$

$$I = \frac{1}{3} [R + G + B] \quad (6)$$

Hence HSI is a function of RGB this method adjusts RGB values to obtain the desired output.

SVM algorithm

Support Vector Machine (SVM) is a supervised machine learning approach that may be utilized for both classification and regression issues. However, it is usually employed in categorization difficulties. In the SVM method, we represent each data item as a point in n-dimensional space with the value of each feature being the value of a given coordinate. Then, we accomplish classification by selecting the hyper-plane that separates the two classes extremely effectively. Support Vectors are just the coordinates of each observation. The SVM classifier segregates the two classes very efficiently SVMs are Effective in high dimensional scenarios and also It is memory efficient since it employs a subset of training points in the decision function called support vectors.



Fig. 7. After K-Means cluster.

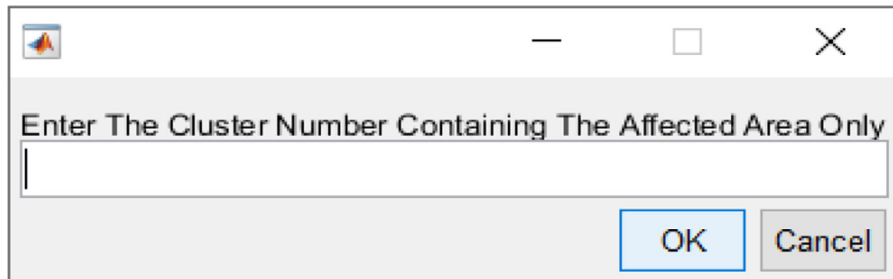


Fig. 8. Entry of afflicted area's cluster number.

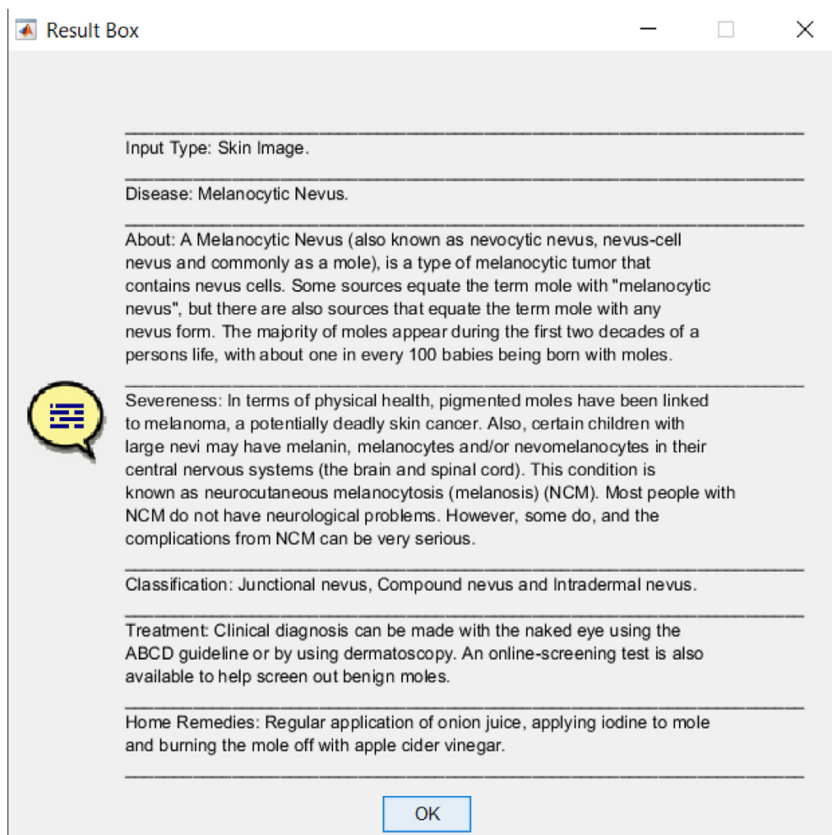


Fig. 9. The result window.

Experimental results

Different types of datasets related to leaf disease namely Alternaria, Anthracnose, bacterial blights, Cercospora, Leaf Spot and Healthy Leaf are considered for training and prediction purposes. Fig 2. shows the enhanced version of the input leaf image which is given as input to the

model. K-Means Clustering is used to determine the number of clusters as shown in Fig 3.

Once the clusters are obtained from the k-means algorithm, we should enter the cluster number of the impacted area in the pop-up window as shown in the below Fig 4.

Table 1

comparison of various detection techniques/ algorithms of leaf disease detection.

Reference	Applied technique	Disease	Accuracy
Asfarian et al., 2013 [17]	Texture Analysis and PNN	Rice	83%
Hu YH et al, 2016 [18]	Hyperspectral Imaging	Potato	95%
Monzurul Islam et al, 2017 [19]	RGB Imaging	Potato	95%
Proposed	K-Means & SVM	Tomato	96%

Table 2

Accuracy analysis of proposed Disease Detection System for different types of leaf and skin disease using SVM.

Disease type	Affected area	Accuracy
Bacterial Blight	15.0217%	96.3871%
Healthy Leaf	8.8002%	96.2%
Anthracoise	17.4244%	96.3%
Cercospora Leaf Spot	20.0391%	96.3871%
Alternaria Alternata	15.1245%	96.3871%
Melanocytic nevus	15.0015%	96.3871%
Melanoma	52.4064%	96.3871%
Basal-cell carcinoma	15.001%	96.3%
Dermatofibroma	15.0031%	96.2%
Actinic keratosis	32.5603%	96.3871%

After the impacted area number is entered a result window will be displayed which contains information like the type of the input image (leaf or skin), disease type and description, Scientific Classification and required treatments (Fig 5).

Similarly, we have considered different types of datasets related to skin disease namely Melanocytic naevus, Melanoma, Basal-cell carcinoma, Dermatofibroma and Actinic keratosis. Fig 6. shows the enhanced version of the input image.

Fig. 7 shows the K-Means Clustering of the image. Fig. 8 shows the entry of the afflicted area's cluster number and Fig. 9 depicts the Result window, which displays the kind of input, illness type and description, Severity, and needed treatments.

Table 1 provides the comparative analysis of different techniques/ algorithms of leaf disease detection. From the table above, we can simply determine that SVM performs better than other classifiers.

Table 2 shows the accuracy analysis of the disease detection system for different types of input images of either leaf or skin using SVM. The system provides almost 96% accuracy for all types of diseased input images.

Conclusion

The appropriate diseases detection and classification is very important to avoid some unexpected events. In this paper, we have discussed various functional blocks and algorithms which are mandatory for image processing. The project is implemented using MATLAB and successfully detected and classified skin and leaf diseases based on their physical appearances using the above-discussed algorithms. By varying the training data, we can extend the disease detection capability. The system de-

scribed above may be upgraded to a real-time video access system that enables uninterrupted plant care.

References

- [1] I.H. Sarker, Machine learning: algorithms, real-world applications and research directions, *SN Comput. Sci.* 2 (2021) 160, doi:10.1007/s42979-021-00592-x.
- [2] F. Andreotti, O. Carr, M.A.F. Pimentel, A. Mahdi, M. De Vos, Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG, 2017 *Comput. Cardiol. (CinC)* (2017) 1–4, doi:10.22489/CinC.2017.360-239.
- [3] L. Li, S. Zhang, B. Wang, Plant Disease Detection and Classification by Deep Learning—A Review, *IEEE Access* 9 (2021) 56683–56698, doi:10.1109/ACCESS.2021.3069646.
- [4] Sachin D. Khirade, A.B. Patil, Savitribai Phule Puneuniversity Pune, India: Plant Disease Detection Using Image Processing, *IEEE*, 2015.
- [5] L. Bottou, Y. Bengio, Convergence Properties of the k-means Algorithms, in: *Advances in Neural Information Processing Systems*, 7, MIT Press, 1995, pp. 585–592. G. Tesauro and D. Touretzky, eds.
- [6] P.K. Agarwal, C.M. Procopiuc, Exact and Approximation Algorithms for Clustering, in: *Proc. Ninth Ann. ACM-SIAM Symp. Discrete Algorithms*, Jan. 1998, pp. 658–667.
- [7] A.S. Abutaleb, Automatic thresholding of gray-level pictures using two-dimensional entropy, *Comput. Vis. Graph. Image Process.* 47 (1989) 22–32.
- [8] N. Abbadi, N. SaadiDahir, M. Dhalimi, H. Restom, Psoriasis Detection Using Skin Color and Texture Features, *Journal of Computer Science* 6 (6) (2010) 648–652 ISSN 1549–3636.
- [9] S. Arivazhagan, R. Shebiah, K. Divya, M. Subadevi, Skin disease classification by extracting independent components, *J. Emerg. Trends Comput. Inf. Sci.* 3 (10) (2012) 1379–1382.
- [10] S. Kolker, D. Kalbande, P. Shimpi, C. Bapat, J. Jatakia, Human skin detection using RGB HSV and YCbCr Color models, *Adv. Intell. Syst. Res.* 137 (2016) 324–332.
- [11] Ramesh, et al., Plant Disease Detection Using Machine Learning, in: 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), 2018, pp. 41–45, doi:10.1109/ICDI3C.2018.00017.
- [12] U. Rahamathunnisa, M.K. Nallakuruppan, A. Anith and S, K.S. Kumar, Vegetable Disease Detection Using K-Means Clustering And Svm, in: 6th International Conference on Advanced Computing & Communication Systems, 2020, pp. 1308–1311, doi:10.1109/ICACCS48705.2020.9074434.
- [13] M.B. Eisten, P.T. Spellman, P.O. Brown, D. Bostein, Cluster analysis and display of genome-wide expression pattern, *Proc. Nat. Acad. Sci.* 95 (1998) 14863–14868.
- [14] Z. Ye, L. Ma, W. Zhao, W. Liu, H. Chen, A Multi-level Thresholding Approach Based on Group Search Optimization Algorithm and Otsu, in: 2015 8th International Symposium on Computational Intelligence and Design, ISCID, 2015, pp. 275–278, doi:10.1109/ISCID.2015.26.
- [15] G. Shi, J. Suo, C. Liu, K. Wan, X. Lv, Moving target detection algorithm in image sequences based on edge detection and frame difference, in: 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference, 2017, pp. 740–744, doi:10.1109/ITOEC.2017.8122449.
- [16] Kumar Vaibhav, et al., K-mean clustering-based cooperative spectrum sensing in generalized K-μ fading channels, *Communication (NCC) 2016 Twenty Second National Conference*, 2016.
- [17] A. Asfarian, Y. Herdiyeni, A. Rauf, K.H. Mutaqin, Paddy diseases identification with texture analysis using fractal descriptors based on fourier spectrum, in: 2013 International Conference on Computer, Control, Informatics and Its Applications (IC3INA), 2013, pp. 77–81, doi:10.1109/IC3INA.2013.6819152.
- [18] Y.H. Hu, X.W. Ping, M.Z. Xu, W.X. Shan, Y. He, Detection of Late Blight Disease on Potato Leaves Using Hyperspectral Imaging Technique, *PubMed* 36 (2) (2016) 515–519.
- [19] M. Islam, Anh Dinh, K. Wahid, P. Bhowmik, Detection of potato diseases using image segmentation and multiclass support vector machine, in: 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering, CCECE, 2017, pp. 1–4, doi:10.1109/CCECE.2017.7946594.
- [20] Manjunatha Badiger, Jose Alex Mathew, Retrospective Review of Activation Functions in Artificial Neural Networks, in: V. Bindhu (Ed.), *Proceedings of Third International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering, 844, Third, Springer, Singapore, 2022, pp. 905–919. 981-16-8862-1_59, In press.
- [21] J.P. Nayak, PCB Fault detection using Image processing, *IOP Conference Series: Materials Science and Engineering*, 255, IOP Publishing Ltd, 2017, 012244.