Automatic Tomato Plant Leaf Disease Classification using Multi-Kernel Support Vector Machine

Jayanthi M.G, Shashikumar D. R.

Abstract: *In agriculture the major problem is leaf disease identifying these disease in early stage increases the yield. To reduce the loss identifying the various disease is very important. In this work , an efficient technique for identifying unhealthy tomato leaves using a machine learning algorithm is proposed. Support Vector Machines (SVM) is the methodology of machine learning , and have been successfully applied to a number of applications to identify region of interest, classify the region. The proposed algorithm has three main staggers, namely pre-processing, feature extraction and classification. In pre-processing, the images are converted to RGB and the average filter is used to eliminate the noise in the input image. After the pre-processing stage, features such as texture, color and shape are extracted from each image. Then, the extracted features are presented to the classifier to classify an input tomato leaf as a healthy or unhealthy image. For classification, in this paper, a multi-kernel support vector machine (MKSVM) is used. The performance of the proposed method is analysed on the basis of different metrics, such as accuracy, sensitivity and specificity. The images used in the test are collected from the plant village. The proposed method implemented in MATLAB.*

Key words: - Tomato, Leaf disease, plant, median filter, texture, colour, shape, multi-kernel support vector machine.

I. INTRODUCTION

Agriculture has a main role in all aspect of human life, including clothing, food, medicine and employment, all over the world. It is a primary source of food for every country which is directly related to its economy as well. In the main categories of agriculture, tomato is the world's most being used vegetable, representing 16% of whole production of crops [1]. Tomato, a typical financial harvest, is one of the most famous vegetables, possesses a huge extent of vegetable development in our every day lives. In all actuality, nonetheless, diseases have been considered as the significant issues that genuinely influence the yield and

Jayanthi M G*, Computer Science and Engineering, Cambridge Institute of technology,Visvesvaraya Technological University, Bangalore, India. Email: jayanthi.cse@cambridge.edu.in

Dr. Shashikumar D. R., Computer Science and Engineering, Cambridge Institute of technology,Visvesvaraya Technological University, Bangalore, India. Email: hod.cse@cambridge.edu.in

nature of tomato. Most tomato diseases can be perceived by the side effects that show up on leaves. Tomato disease acknowledgment dependent on diseased leaf pictures has become an inexorably significant and testing research subject in the network [2]. The main point of this problem is that almost every year a huge amount of tomato crops is affected by different diseases. Due to improper care and lack of timely detection of diseases, theses crops are facing damage of quality, quantity and productivity as well. Diseases demolish the whole crops and affect the adjacent crops as well [3]. It is found that the crops damaged by diseases result thousand billion dollars loss every year [4]. Plants diseases are very hard to control facing the challenge to detect the disease at right time in terms of as early as possible to reduce the loss and getting as increased production as possible. If a plant influenced by any disease is distinguished by existing strategies expends additional time, work and cost. What's more ID of plant disease wrongly will prompts gigantic loss of time, cash, work, yield and item quality. Consequently for beneficial development the state of plant ought to be identified precisely. All in all this sort of disease is distinguished through experienced individuals however because of condition changes the manual forecast turns out to be excessively troublesome. To avoid the problem, automatic detection is proposed. Simultaneously, in certain nations, ranchers don't have legitimate offices or even thought that they can contact to specialists. Because of which counseling specialists even cost high just as tedious as well. In such conditions, the recommended strategy ends up being gainful in observing huge fields of crops [5]. Automatic location of the diseases by simply observing the side effects on the plant leaves makes it simpler just as less expensive. The automatic conclusion framework dependent on plant disease highlights decreases the reliance on specialists in the zone concerned [6,7]. This likewise bolsters machine vision to give picture based automatic procedure control, review, and robot direction. It very well may be accomplished utilizing picture preparing procedures which have a fine potential for early recognition of diseases [8]. Nowadays, lot of machine learning algorithms are introduced to identify the diseases in plants namely, k-nearest neibour classifier [9], naïve bias classifier [10], artificial neural network classifier [11], support vector machine [12] and convolution neural network [13]. Albeit many crop diseased leaf picture handling and grouping strategies have been proposed, there is still a lot of space for the improvement of diseased leaf characterization in term of computational expense

and arrangement precision [14,15].The main objective of proposed methodology is to classify a tomato leaf as normal or abnormal. For classification, in this paper, MKSVM classifier is utilized. The rest of the paper is organized as follows; the chapter 2 describes related research topics and the proposed methodology is presented in chapter 3. The experimental result is presented in section 4 and conclusion part is presented in section 5.

II. LITERATURE SURVEY

Lot of researchers had developed tomato leaf disease classification using different techniques. Among them some of the works are analysed here; Jinzhu et al. [17] have explored the chance of segregating tomato yellow leaf curl disease by a hyperspectral imaging strategy. A hyperspecral imaging framework gathered hyperspectral pictures of both sound and tainted tomato leaves. The reflectance spectra, first subsidiary reflectance spectra and supreme reflectance distinction spectra in the frequency scope of 500–1000 nm of both foundation and the leaf territory were broke down to choose touchy frequencies and band proportions. 853 nm was chosen to make a cover picture for foundation division, while 720 nm from the reflectance spectra, four pinnacles (560, 575, 712, and 729 nm) from the principal subordinate spectra and, four frequencies with higher qualities (586, 720 nm) and lower esteems (690, 840 nm) in the total distinction spectra were chosen as a lot of delicate frequencies. This exploration likewise showed that multispectral pictures at 560, 575 and 720 nm have a potential for distinguishing tomato yellow leaf curl infection contamination in field applications. In [16], Mohammed et al., have presented a convolution neural network (CNN) based tomato leaf disease classification. Here, the dataset contains 14828 images of tomato leaves infected by nine diseases. To avoid the hand crated features, CNN used for this paper.

Furthermore, Ramesh [18] explained the identification and classification of rice leaf diseases using optimal DNA. They achieved the plant disease classification using four stages. Initially, in image calculation, the plant leaves are captured from the area. After that, they convert the RGB image into HSV. Then, they cluster the disease area with the help of the K-clustering algorithm. Finally, the classification was carried out with the help of ODNN. For the purpose, the data were collected from the rural areas of Tirunelveli district of Tamil Nadu, Ayikudi and Panboli. In total, 650 images were analyzed, including 95 normal images, 125 bacterial blight images, 170 blast images, 110 envelope rotation images, and 150 brown images.

in [19], detailed investigation of various image procedures to identify the disease in rice plant is presented. Essential hues are RGB picture used to recognize the disease in division. In Such strategies used to recognize the disease from the early side effect of the yield misfortune. This original copy would assist the scientists with understanding rice disease ID utilizing PC vision. Finally, this paper has the conversation of various specialist's advantages and disadvantages of all investigations identified with plant disease distinguishing proof. In advanced picture preparing strategies, they are utilized for improvement of the picture. GLCM and SURF highlights are utilized for include

extraction. Edge recognition and FCM is utilized for division. ANN is utilized for classification. In [20], J. Praveen Kumar, and S. Domnic have shown a strategy for leaf district extraction from plant pictures and checking the quantity of leaves. This introduced strategy has three stages. The initial step includes another factual based strategy for picture upgrade. The subsequent advance includes in the extraction of leaf district in plant picture utilizing a diagram based strategy. The third step includes in including the quantity of leaves in the plant picture by applying Circular Hough Transform (CHT). A robotized method for crop disease recognizable proof on different leaf test pictures comparing to various crop species utilizing Local Binary Patterns (LBPs) for include extraction was introduced by X.E. Pantazi et al [21]. This uses a committed One Class Classifier for each plant wellbeing condition including, sound, fleece mold, fine buildup and dark spoil. The calculations prepared on vine leaves have been tried in an assortment of crops accomplishing an extremely high speculation conduct when tried in different crops.

III. PROPOSED METHODOLOGY

To improve the productivity of tomato's, early detection of disease in leaf is proposed. The proposed method consists of three modules such as Preprocessing, Feature Extraction and Classification. First is the pre-processing, in which we extract the green band of input images using the RGB conversion. And then use the median filter to remove the noise in those extracted green band images. The second stage is feature extraction, is to extract some features such as shape, texture, color. After the feature extraction, the extracted features are given to Multi-Kernel Support Vector Machine (MKSVM) classifier to classify an image as a normal or abnormal. The overall diagram of the proposed method is given below figure 1

3.1 Pre-processing

Pre-processing is the primary step of image processing techniques. The aim of pre-processing is to improve the quality of the input image and remove impurities and noise in the image. Pre-processing consists of two sub-tasks, such as the green band and the mean filter. Firstly, extract only the green band images from the input images using RGB conversion. Second, we remove the noise in the extracted Green band image using a median filter.

i) RGB Transformation

In the RGB model, each color represents the basic color elements of red, green, and blue, using an 8-bit monochrome standard. The corresponding RGB color image is 24 bit / pixel - 8 bit for each color band (red, green and blue). In the RGB model, each color represents the basic color elements of red, green, and blue, using an 8-bit monochrome standard. RGB color images have 24 bit / pixel sizes. Each 8 bit represents each color band (red, green and blue). Colors, red, green, blue values are given in terms of RGB conversion by,
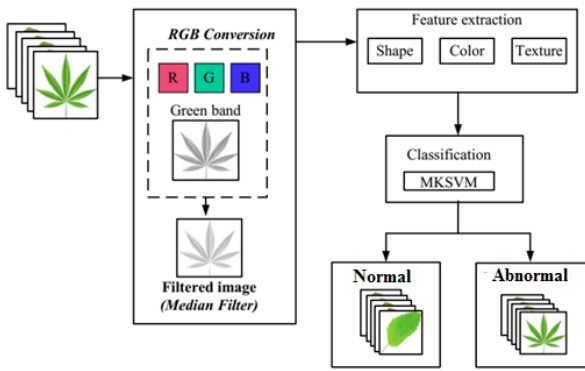
561

Figure 1: Overall diagram of the proposed method

$$r = \frac{R^R}{(R^R + G^G + B^B)} \quad (1)$$

$$g = \frac{G^G}{(R^R + G^G + B^B)} \quad (2)$$

$$b = \frac{B^B}{(R^R + G^G + B^B)} \quad (3)$$

The assumption is that R^R, G^G, B^B have been normalized so that they are in the range [0, 1]. Here, we extract the only the green band images from the RGB conversion. And a simple transformation to create a feature representing green band from the RBG band is given below,

Green Band Image $= 2G^G - B^B - R^R$ (4)

Where, R^R, G^G and B^B represent the intensity levels recorded by the red, green and blue leaf images of digital camera. This change uses the high contrast between the input color images. This variation is due to the inherent difference between the green leaf and the background elements, so that the camera of accidental radiation and red, green and blue detectors must be relatively independent of the calibration. The obtained green band will resist leaf reflection until a portion of the radiation is in contact with the chlorophyll. We use the median filter to remove noise from the extracted green band image.

3.2. Feature Extraction

The output of median filter is analyzed by utilizing several features. In the proposed algorithm three various types of features namely, shape, texture and color features. The shape feature includes area, perimeter, eccentricity and color feature consists of mean and Standard deviation.

1. Shape

This is one of the best features to distinguish the medicinal leaf. Utilizing exceptionally generated algorithms to recognize the medicinal variety of leaf shape is conceivable through image processing. This will likewise incorporate the edge, tip and base features of the leaf. The shapes of the leaf have different statistical features like area, perimeter and eccentricity.

a. Area

The estimation of leaf region is east to evaluate, just checking the quantity of pixels of binary esteem 1 on smoothed leaf picture. Leaf area is meant as A.

Area = Area of pixel * Total no of pixels present in a leaf. (5)

b. Perimeter

The leaf perimeter is denoted as P, leaf perimeter is evaluated by counting the quantity of pixels involving leaf edge.

c. Eccentricity

The binary value 0 and 1 for leaf characterizes, Scalar that determines the eccentricity of the ellipse that has indistinguishable second-minutes from the locale. The eccentricity is the proportion of the distance between the foci of the ellipse and its major axis length. The value is somewhere in the range of 0 and 1. This property is suitable for 2-D input matrices.

$$E = 2 \times \frac{\sqrt{\left(\frac{max}{2}\right)^2 - \left(\frac{min}{2}\right)^2}}{max} \quad (6)$$

Where, E- Eccentricity, max- Major axis, min- Minor axis.

2. Color

Color is a good identification parameter when there are many colors in the leaves of different species. Color moments refer to color features to represent the color picture. Features that can be added are mean and standard deviation. As for the RGB color space, the features can be expanded from red, green and blue on each plane. The mean of the color feature is given below

$$\mu = \frac{1}{DM} \sum_{i=1}^{D} \sum_{j=1}^{M} Rn_{ij} \quad (7)$$

The standard deviation of color feature is shown in below,

$$\sigma = \sqrt{\frac{1}{DM} \sum_{i=1}^{D} \sum_{j=1}^{M} (Rn_{ij} - \mu)^2} \quad (8)$$

Where, D and M is the imension of the image. Rn(ij) is the values of the color on column i and row j. The difference between the green average and the red average in the process, the difference between the green average and the blue average, the difference between the green standard deviation and the red standard deviation, the difference between the green standard deviation and the blue standard deviation of the three color planes. Implement software.

3. Texture

Leaf image system provides information about the spatial arrangement of color or intensity in a leaf image. The diversity of the techniques available to the system. There are many different types of system features that we have taken GLCM features. There are 22 types of features in GLCM. Inverse differential moment, inverse difference normalization, variance, entropy, energy, joint mean, joint entropy, squares or variance, integrity, correlation, autocorrelation, cluster shadow, maximum probability, difference variance, difference entropy, cluster significance, directionality, unity , Measurement of contact information.

562

IV. MULTI-KERNEL SVM FOR DISEASE CLASSIFICATION

In this section, using the MKSVM classifies the tomato leaf as a normal or abnormal. The multiple kernel function is viewed as a fruitful technique to plot an optimal kernel. The optimal kernel can basically be gotten as a curved mix of more than one kernel. In this paper, linear kernel function, and Quadratic kernel function is used to consolidate in SVM process. SVMs are paired linear arrangement strategies which channel for the hyperplane (in the hyperspace of characteristics) that isolates two classes in a readiness set. This hyperplane is found by expanding the accumulated edge, which is the parcel from the hyperplane to the closest focuses, exhibited support vectors. Consider the training data set $T = \{a_i, b_i\}_{i=1}^{m}$, where $a_i \in R^n$ is the ith input feature vectors, $y_i \in \{+1, -1\}$ is the class label of a_i, m is the total number of training data. the classification function of SVM is calculated as follows;

total number of training data. the classification function of SVM is calculated as follows;

$$F(a) = w^T \phi(a) + x \qquad (9)$$

Where;

$\phi(a) \rightarrow$ Non-linear mapping function which

$x \rightarrow$ Bias

$w \rightarrow$ Weight vector

Here, $\phi(a)$ is used for map the input feature vector into a higher dimensional feature space, $x \ and \ w$ are used to determine the position of the separating hyper-plane. The linear decision function based SVM classification can be written as follows;
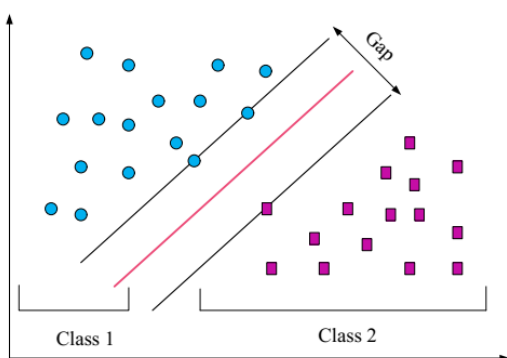

Figure 2: Basic Diagram of Support Vector Machine (SVM)

$$F(x) = Sgn\left[\sum_{i=1}^{n} \alpha_i \ b_i \left(\phi(a_i).\phi(a_j)\right) + x\right] \qquad (10)$$

The above function only used for linear decision. To apply the SVM classifier to non-linear decision function, kernel function is included. In this paper, multi-kernel is used for classification function. In this manner, we can map the input data into high dimensional feature space. The proposed non-linear decision function can be written as follows;

$$F(x) = Sgn\left[\sum_{i=1}^{n} \alpha_i \ K\left(a_i, a_j\right) + x\right] \qquad (11)$$

Here, $K(a, a_i)$ represent the kernel function. In this paper, multi kernel is utilized and the hyper-plane separation is performed based on the multi-kernel functions. In this paper, liner kernel and quadratic kernels are used.

A. Kernel functions

In SVM, kernels can be utilized to grow the information into a virtual space where it may be less mentioning to isolate them. The rule perfect situation of kernel functions is that the essential task should have been depicted in the new virtual space is the interior item.

Linear kernel function: The two classes are linearly separable, which proposes that it is conceivable to organize no short of what one hyperplane depicted by a vector w with a bias, which can separate the classes with zero mistake. The straight kernel function is tended to as takes after:

$$M_{k1} \Rightarrow K(c, c_i) = c.c_i \qquad (12)$$

$K(c, c_i)$ Represents the inner product of two vectors induced in the feature space by the input vector c and input pattern c_i pertaining to the ith pattern.

Quadratic Programming: Quadratic Kernel Functions in SVM Function It is a kernel matrix that feels complete and shapes a wide dimensional space in this way

$$M_{k2} \Rightarrow K(b - b_i) \times (1 + (b - b_i)) \qquad (13)$$

In these two kernel functions were combined and the average of two functions is used in this research work to separate the data's of same classification. The hybrid concept of two proposed kernel function is described as below:

$$M_K(b, b_i) = \frac{\left(Linear_k(b, b_i) + Quad_k(b, b_i)\right)}{2} \qquad (14)$$

Where $M_K(b, b_i)$ in the above equation symbolizes the integrated kernel function representation of the linear and quadratic kernels to classify the normal or abnormal image.

V. RESULT AND DISCUSSION

This section presents information and discussion on pruning leaf detection and segmentation using adaptively regularized multi-kernel-based fuzzy-means and optimal BNN classifiers. To implement the proposed technique, we used MATLAB version (7.12). This proposed mechanism has an Intel Core i5 processor speed of 1.6 GHz and 4GB of RAM on Windows machine.

The proposed system has been tested on Tomao leaf data sets collected from the Plant Village (https://www.plantvillage.org/en/ plant_images). The sample images for the test are listed below;

Figure 3: Experimental used sample images

A. Evaluation metrics

The system performance is analysed using evaluation metrics such as sensitivity, specification and accuracy, which are depicted below.

Sensitivity: The ratio of many true positives to the sum of true positive and false negatives is called sensitivity.

$$Sensitivity = \frac{No.of\ (TP)}{No.of\ (TP)\ +\ No.of\ (FN)} \times 100 \tag{15}$$

Specificity: Is the ratio of many true negatives to the sum of true negative and false positives.

$$Specificity = \frac{No.of\ (TN)}{No.of\ (TN)\ +\ No.of\ (FP)} \times 100 \tag{16}$$

$$(24)$$

Accuracy: Accuracy is calculated by measures of sensitivity and specificity. This is indicated by,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{17}$$

Table 1: Results Comparisons

Parameters	MKSVM	KNN	SVM
Accuracy	97.34	87.3	90
Specificity	99.02	96	97
Sensitivity	97.34	89	91.04

B. Experimental results

The main objective of this paper is Automatic tomato plant leaf disease classification using multi-kernel support vector machine. In this paper, linear and quadratic kernels are used for multi-kernel process. To prove the effectiveness of proposed method, the proposed MKSVM based leaf disease classification compared with SVM based classification and KNN based classification. The comparison results are shown in Table 1.
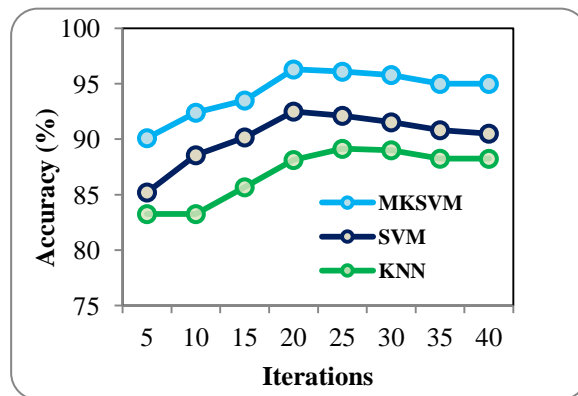

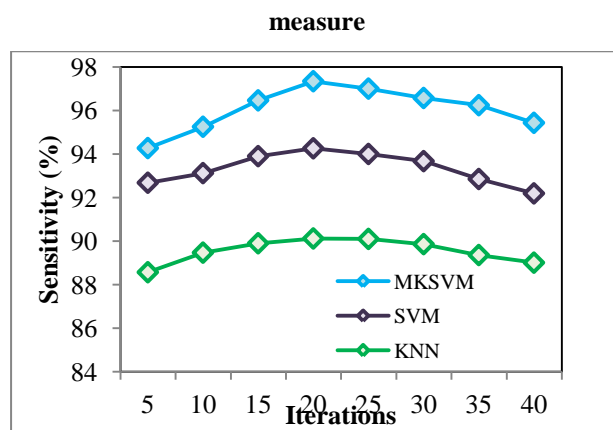**Figure 4: Comparative analysis based on accuracy measure**

Figure 5: Comparative analysis based on sensitivity measure
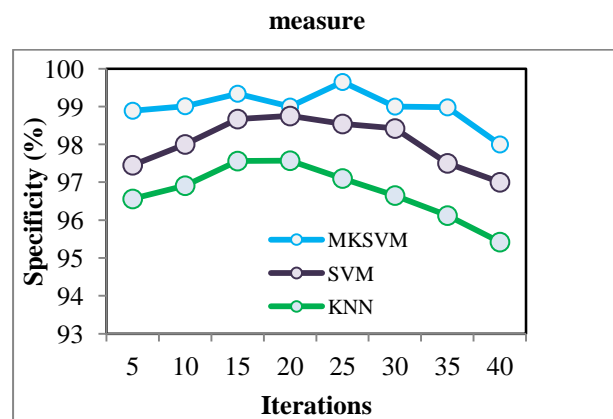

Figure 6: Comparative analysis based on specificity measure

VI. CONCLUSION

This work has developed an efficient automatic tomato leaf disease classification. For the classification process, the MKSVM classifier has been used. To improve the classification accuracy, linear and quadratic kernels are hybridized with SVM. The performance of the proposed method has been analysed on the basis of different metrics, such as accuracy, sensitivity and specificity. The proposed method has achieved a maximum accuracy of 97.34%.

The proposed method achieved a maximum sensitivity of 97.3%. The proposed method achieved a maximum sensitivity of 99.01%. In future work the same algorithm can be used for predicting other leaf diseases .The algorithm can used by using unsupervised methods to predict the images which is not trained by the model.

REFERENCES

1. Fuentes, A., S. Yoon, H. Youngki, Y. Lee, and D. S. Park. "Characteristics of Tomato Plant Diseases—A study for tomato plant disease identification." In Proc. Int. Symp. Inf. Technol. Converg, vol. 1, pp. 226-231. 2016.
2. Hidayatuloh, Akbar, M. Nursalman, and Eki Nugraha. "Identification of Tomato Plant Diseases by Leaf Image Using Squeezenet Model." In 2018 International Conference on Information Technology Systems and Innovation (ICITSI), pp. 199-204. IEEE, 2018.
3. P. S. R. Strange, "Plant Disease: A Threat to GLobal Food Security," Annual Review Phytopathol, vol. 43, pp. 83-116, 2005.
4. J. C. Lai, S. K. Li, B. Ming, N. Wang, K. R. Wang, R. Z. Xie, and S. J. Gao., "Advances in research on computer-vision diagnosis of cropdiseases. Scientia Agricultura Sinica",2009, vol. 42, pp.1215–1221
5. C. C. Tucker and S. Chakraborty. "Quantitative assessment of lesion characteristics and disease severity using digital image processing", J. Phytopathology, 1997, vol.145, pp.273–278.
6. S. Ahmad, J. F. Reid, M. R. Paulsen, and J. B. Sinclair, "Color classifier for symptomatic soybean seeds using image processing PlantDisease",1999, vol. 83, pp.320–327.
7. J. W. Olmstead, A. Gregory, and G. A. Lang., "Assessment of severity of powdery mildew infection of sweet cherry leaves by digital imageanalysi", Hortscience, 2001,vol. 36, pp.107– 111.
8. Kaur, L., & Laxmi, V. (2016). Detection of Unhealthy Region of plant leaves using Neural Network. Dis Manag, 1(05), 34-42
9. Hossain, Eftekhar, Md Farhad Hossain, and Mohammad Anisur Rahaman. "A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier." In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-6. IEEE, 2019.
10. Khirade, Sachin D., and A. B. Patil. "Plant disease detection using image processing." In 2015 International conference on computing communication control and automation, pp. 768-771. IEEE, 2015.
11. Ishak, Syafiqah, Mohd Hafiz Fazalul Rahiman, Siti Nurul Aqmariah, Mohd Kanafiahb, and Hashim Saadc. "Leaf disease classification using artificial neural network." Jurnal Teknologi 77, no. 17 (2015): 109-114.
12. Padol, Pranjali B., and Anjali A. Yadav. "SVM classifier based grape leaf disease detection." In 2016 Conference on advances in signal processing (CASP), pp. 175-179. IEEE, 2016.
13. Venkataramanan, Aravindhan, Deepak Kumar P. Honakeri, and Pooja Agarwal. "Plant Disease Detection and Classification Using Deep Neural Networks", International Journal on Computer Science and Engineering (IJCSE), vol.11,no.8, 2019
14. Singh, Vijai, and Ak K. Misra. "Detection of plant leaf diseases using image segmentation and soft computing techniques." Information processing in Agriculture 4, no. 1 (2017): 41-49.
15. Ghaiwat, Savita N., and Parul Arora. "Detection and classification of plant leaf diseases using image processing techniques: a review." International Journal of Recent Advances in Engineering & Technology 2, no. 3 (2014): 1-7.
16. Mohammed Brahimi,KamelBoukhalfa and Abdelouahab Moussaoui, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization", applied artificial intelligence, 2017.
17. Jinzhu Lu, Mingchuan Zhou, YingwangGao and Huanyu Jiang, "Using hyperspectral imaging to discriminate yellow leaf curl disease in tomato leaves", journal of Precision Agric, 2017.
18. Ramesh, S., and D. Vydeki. "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm", Information Processing in Agriculture (2019).
19. G. Jayanthi, K.S. Archana, A. Saritha, "Analysis of Automatic Rice Disease Classification Using Image Processing Techniques ", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019
20. Kumar, J.P. and Domnic, "Image based leaf segmentation and counting in rosette plants", Information Processing in Agriculture, 6(2), pp.233-246, 2019.

AUTHORS PROFILE

Mrs. Jayanthi M.G obtained her B.E degree in Computer Science from Madras University . Then she obtained his Master's degree in Computer Science and Engineering from UVCE,Bangalore University. She is pursuing Ph.D degree from Visvesvaraya Technological University(VTU). She has also obtained OCA professional qualifications. Currently, she is an associate professor Department of Computer Science, Cambridge Institute of Technology, Visvesvaraya Technological University (VTU). Her s specializations include Image Processing, Database Programming. Her current research interests are Agriculture Image Processing,

Dr. Dandinashivara Revanna Shashikumar received BE degree from Mysore University and ME degree from Bangalore University, Bangalore and Ph.D in Information and Communication Technology of Fakir Mohan University, Balasore, Orissa. He is currently working as Professor and HoD, Dept. of Computer Science, Cambridge Institute of Technology, Visvesvaraya Technological University (VTU). His research interests include Microprocessors, Pattern Recognition, and Biometrics, Computer Networks, Data mining and Data Warehouse He has published 20 research publications in referred National and International Journals. He is the reviewer for some of the International journals.