PLANT LEAF DISEASE DETECTION USING MULTICLASS SVM

Suriya D[1] ,M.Parvathy[2], S.Asha[3]

Assistant Professor [1],Prof & Head [2] ,Assistant Professor [3]

Department of Computer Science and Engineering,

Sethu Institute of Technology

Abstract— In India, the country where the main source of income is from agriculture. Farmers grow a variety of crops based on their requirement. Since the plants suffer from the disease, the production of crop decreases due to infections caused by several types of diseases on its leaf, fruit, and stem. Leaf diseases are mainly caused by bacteria, fungi, virus etc. Diseases are often difficult to control. Diagnosis of the disease should be done accurately and proper actions should be taken at the appropriate time. Image Processing is the trending technique in detection and classification of plant leaf disease. This work describes how to automatically detect leaf diseases. The given system will provide a fast, spontaneous, precise and very economical method in detecting and classifying leaf diseases. This paper is envisioned to assist in the detecting and classifying leaf diseases using Multiclass SVM cascaded classification technique. First, the affected region is discovered using segmentation by K-means clustering, then features (color and texture) are extracted. Lastly, classification technique is applied in detecting the type of leaf disease. The proposed system effectively detects and also classify the disease.

Keywords—Multicast SVM, Radial Basis Function Neural Network (RBFNN)

1. INTRODUCTION

India is a fast-developing country and agriculture is the back bone for the country's development in its early stage. However, agricultural field faces lots of hurdles including huge loss in the crop production. However, disease prediction using classification algorithms appears to be a difficult task as the accuracy varies for different input data. In this paper, several research contributions related to various plant leaf diseases detection using different classification algorithms are reviewed and compared. The existing method encompasses human involvement for classification and identification of diseases. This procedure is time-consuming and costly. Automatic segmentation of disease from plant leaf images using soft computing approach can be reasonably useful than the existing one. System named as Bacterial foraging optimization based Radial Basis Function Neural Network (BRBFNN) for identification and classification of plant leaf diseases automatically. For assigning optimal weight to Radial Basis Function Neural Network (RBFNN) we use Bacterial foraging optimization (BFO) that further increases the speed and accuracy of the network to identify and classify the regions infected of different diseases on the plant leaves. The proposed method attains higher accuracy in identification and classification of diseases. Computers have evolved to be a vital device in a number of applications like defense, medical, agriculture, engineering etc. with its ability to process multimedia information like images captured from some computing devices. An image contains important information that can be retrieved by using some computational method. Image segmentation is a task for partitioning an image into smaller parts that are more meaningful. Interestingly, it can be stated as identification and classification of some region of interest. The segmentation is performed based on some common properties of the objects present in an image like color, texture and, shape etc. Image segmentation is a preprocessing step for image processing generally performed by using two methods (i) Traditional method and (ii) Soft computing method.

LITERATURE REVIEW

In [1] SherlyPuspha Annabel et al.,In this system, Plants are considered to be important as they are the source of energy supply to mankind. Plant diseases can affect the leaf any time between sowing and harvesting which leads to huge loss on the production of crop and economical value of market. Therefore, leaf disease detection plays a vital role in agricultural field. However, it requires huge manpower, more processing time and extensive knowledge about plant diseases. Hence, machine learning is applied to detect diseases in plant leaves as it analyses the data from different aspects, and classifies it into one of the predefined set of classes. The morphological features and properties like color, intensity and dimensions of the plant leaves are taken into consideration for classification. This paper presents an overview on various types of plant diseases and different classification techniques in machine learning that are used for identifying diseases in different plant leaves.

In [2]Rajinikanthet al.,In this system,In ophthalmology, substantial advancement can be found in assessment and

evaluation of the abnormality in retinal anatomical structures, such as optic nerve, disc, and vasculature. Most of the retinal abnormality assessments can be done using the imaging procedures, in which the retinal parts are recorded using a dedicated imaging device called the Fundus Camera (FC) and these images are called the Fundus Camera Images (FCI). In this work, FCI assessment procedure is proposed using the Spider Monkey Optimization Algorithm. The SMOA-assisted Shannon's Entropythresholding is initially executed to enhance the retinal sections of FCI. Then an Active Contour segmentation procedure is implemented to extract the optic disc/optic cup. Finally, a relative investigation between the extracted optic disc/optic cup and the expert provided disc/cup section is carried out to compute the Image Similarity Parameters (ISP). In this work, the benchmark FCI dataset, called the Rim-One is adopted for the investigation. During this study, Rim-One FCI dataset with the optic disc and stereo image (dual image) are considered for the examination. The performance of the SMOA is then assessed with other heuristic approaches, such as Particle Swarm Optimization, Bacterial Foraging Optimization, and Firefly Algorithm approaches. The experimental investigation confirms that all these heuristic approaches offer approximately a similar result on the considered Rim-One FCI dataset.

In [3]SKumbhar et al.,In this system, Agriculture is one of the important professions in many countries including India. As most part of the Indian financial system is dependent on agriculture production, the keen attention to the concern of food production is necessary. The taxonomy and identification of crop infection got much importance in technical as well as economic in the Agricultural Industry. While keeping track of diseases in plants with the help of specialists can be very costly in agriculture region. There is a need for a system which can automatically detect the diseases as it can bring revolution in monitoring large fields of crop and then plant leaves can be taken cure as soon as possible after detection of disease. The aim of the proposed system is to develop an application which recognizes cotton leaf diseases. For availing this user need to upload the image and then with the help of image processing we can get a digitized color image of a diseased leaf and then we can proceed with applying CNN to predict cotton leaf disease.

In [4]Gittaly et al.,In this system,The natural products are inexpensive, non-toxic, and have fewer side effects. Thus, their demand especially herbs based medical products, health products, nutritional supplements, cosmetics etc. are increasing. The quality of leafs defines the degree of excellence or a state of being free from defects, deficits, and substantial variations. Also, the diseases in leafs possess threats to the economic, and production status in the agricultural industry worldwide. The identification of disease in leafs using digital image processing, decreases the dependency on the farmers for the protection of agricultural products. So, the leaf disease detection and classification is the motivation of the proposed work. In this paper, a novel fuzzy set extended form neutrosophic logic based segmentation technique is used to evaluate the region of interest. The segmented neutrosophic image is distinguished by three membership elements: true, false and intermediate region. Based on segmented regions, new feature subset using texture, color, histogram and diseases sequence region are evaluated to identify leaf as diseased or healthy. Also, 9 different classifiers are used to monitor and demonstrate the discrimination power of combined feature effectiveness, where random forest dominates the other techniques. The proposed system is validated with 400 cases (200 healthy, 200 diseased). The proposed technique could be used as an effective tool for disease identification in leafs. A new feature set is promising and 98.4% classification accuracy is achieved.

In [5]Mohammadet al.,In this system, Continuous droughts and water scarcity have led to the need for optimal exploitation of dams' reservoirs. Thus, the new meta-heuristic algorithm, spider monkey, is suggested for complex modelling of the multi-reservoir system in Iran with the aim of decreasing irrigation deficiencies. Golestan and Voshmgir dams' operations are optimized with the spider monkey algorithm. The algorithm based on the exchange of information between local and global leaders with the other monkeys which improves the convergence speed. Average deficiencies for Golestan dam is computed as 2.1 and 1.9 MCM by spider monkey algorithm while it is respectively computed as 6.7, 16.4, 11.1, 4.1, 14.6, 19 MCM by particle swarm algorithm, harmony search algorithm, imperialist competitive algorithm, water cycle algorithm, genetic algorithm, and standards operation policy method. Also, the computation time of the spider monkey algorithm is 50 and 47 s for the Golestan and Voshmgir dams while the genetic algorithm optimizes the problem in 172.6 s and 112 s and the particle swarm algorithm needs 117.4 s and 100 s for the Golestan and Voshmgir, respectively. Also, root means square error and mean absolute error between demand and released water for the spider monkey algorithm has the least values among the applied evolutionary algorithms. Thus, the spider monkey algorithm is suggested as an appropriate method for optimizing the operation policy for the dam and reservoir systems.

2. PROBLEM STATEMENT

In this work identification and classification of plant leaf disease is performed by using Bacterial foraging optimization based Radial Basis Function Neural Network (BRBFNN). The feature extraction process is carried out by seeding and grouping the points having similarity in some manner using region growing approach the training of the RBFNN is performed by using bacterial foraging optimization that proofs to be an efficient and powerful tool for initializing the weight of RBFNN and training the network that can correctly identify different affected regions on plant leaf image. With the help of BFO, the existing algorithm achieves higher convergence ratio and accuracy. The methodology of the existing work is given by Fig. 1
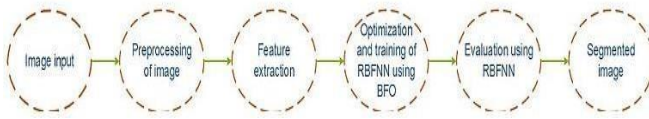
PROPOSED WORK

In this section, we first propose an algorithm to select the optimal features, respectively, for each application. Then, we build the binary sub-classifiers respectively for those applications by using those selected features. Finally, based on Theorem 1 and Lemma 1, we present an algorithm for making an optimal cascade of those sub-classifiers. Our main objective here is to improve classification accuracy by addressing the issues of both discriminator bias and class imbalance in traffic classification
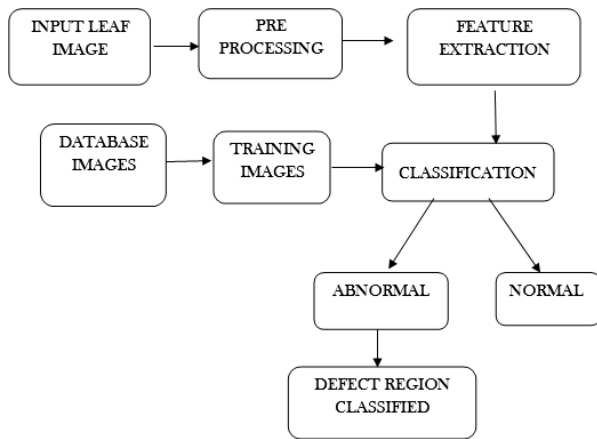
MODULE DESIGN

Database

The image database has been widely used to test neural-network-based leaf-detection systems in complex backgrounds. The database consists of three sets of gray-level images. These images are scanned photographs, newspaper pictures; files collected from the World Wide Web, and digitized television shots. The first two sets contain 169 leaves in 42 images and 183 plants in 65 images respectively.

Pre-process

Rgb2gray converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance new map = **rgb2gray**(map) returns a gray scale colormap equivalent to map . Class Support. If the input is an RGB image, it can be of class uint8 , uint16 , or double .

Image Segmentation

During image segmentation, the given image is separated into a homogeneous region based on certain features. Larger data sets are put together into clusters of smaller and similar data sets using clustering method. In this proposed work, K-means clustering algorithm is used in segmenting the given image into three sets as a cluster that contains the diseased part of the leaf. Since we have to consider all of the colors for segmentation, intensities are kept aside for a while and only color information is taken into consideration. The RGB image is transformed into LAB form (L-luminous, a*b-chromous). Of the three dimensional LAB, only last two are considered and stored as AB. As the image is converted from RGB to LAB, only the "a" component i.e. the color component is extracted. Properties and process of K-Means Algorithm are as follows:

Properties

i.K number of the cluster should be present always.

ii In each given cluster, at least one item should be present.

iii Overlapping of clusters should never happen.

iv Every participant of the single cluster should be close to its own cluster than any other cluster

Process

1. The given data set should be divided into K number of clusters and data points need to be assigned to each of these clusters randomly.

2. For each data point, the distance from data point to each cluster is computed using Euclidean distance The Euclidean distance is nothing but the distance between two-pixel points and is given as follows: Euclidean Distance=$\sqrt{((x1-x2)^2 + (y1-y2)^2)}$ where (x1, y1) & (x2, y2) are two pixel points (or two data points).

3. The data point which is nearer to the cluster to which it belongs to should be left as it is.

4. The data point which is not close to the cluster to which it belongs to should be then shifted to the nearby cluster.

5. Repeat all the above steps for entire data points.

6. Once the clusters are constant, clustering process needs to be stopped.

Feature Extraction

From the input images, the features are to be extracted. To do so instead of choosing the whole set of pixels we can choose only which are necessary and sufficient to describe the whole of the segment. The segmented image is first selected by manual interference. The affected area of the image can be found from calculating the area connecting the components. First, the connected components with 6 neighborhood pixels are found. Later the basic region properties of the input binary image are found. The interest here is only with the area. The affected area is found out. The percent area covered in this segment says about the quality of the result. The histogram of an entity or image provides information about the frequency of occurrence of certain value in the whole of the data/image. It is an important tool for frequency analysis. The co-occurrence takes this analysis to next level wherein the intensity occurrences of two pixels together are noted in the matrix, making the co-occurrence a tremendous tool for analysis. From gray-co-matrix, the features such as Contrast, Correlation, Energy, Homogeneity' are extracted. The following table lists the formulas of the features

Feature selection

We use feature selection to optimize classification accuracy of each sub-classifier. Algorithm 1 details the process of feature selection. Algorithm is performed on the training dataset Ds that contains two kinds of classes. The samples of the given application are attributed to one class. The rest samples are attributed to the other class. We propose a wrapper method to select the optimal features for a given application. Specifically, the algorithm selects the features that make C4.5 decision tree bin tree achieve the highest AUC value for the given application. We apply backward search method in building feature subset feature subset of the feature set original set, and then train the binary classifier bin tree with feature subset. If feature subset makes bin tree achieve the higher AUC value, the features in feature subset are inserted into final feature. It is worth noting that we call weka library to calculate AUC value. If the times for searching features is larger than the given threshold value maxtimes, the algorithm returns final feature as the final feature subset Using the statistical MATLAB commands the other properties are found out. Those are Mean Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM. Mean: Average or

$$Mean = \left(\frac{1}{N}\right) Xi$$

mean value of the array. Mean is given by
Where Xi->pixel intensity, N->a total number of pixels of an image.

Standard Deviation: Standard deviation is computed using the below formula:

$$Standard\ Deviation = (1/N)\ \Sigma\ (Xi - \mu)\ ½$$

Where μ-> mean.

Entropy: Entropy is a statistical measure of randomness that is used to characterize the texture of the input image. Entropy is defined as

$$Entropy = -sum\ (p.*log2\ (p))$$

Where p-> histogram counts
Variance: Variance is computed using

$$Variance = (1/N)\ \Sigma\ (Xi - \mu)^2$$

Variability is measured using variance. Skewness: The image surface is judged with the Skewness

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - x')^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - x')^2\right)^{\frac{3}{2}}}$$

The same feature set is used for training the SVM as well to identify the class of the input image. Training 1. Start with images of which classes are known for sure. 2. Find the property set or feature set for each of them and then label suitable. 3. Take the next image as input and find features of this one as new input. 4. Implement the binary SVM to multi class SVM procedure. 5. Train SVM using kernel function of choice. The output will contain the SVM structure and information of support vectors, bias value etc. Find the class

of the input image. 7. Depending on the outcome species, the label to the next image is given. Add the features set to the database. 8. Steps 3 to 7 are repeated for all the images that are to be used as a database. 9. Testing procedure consists of steps 3 to 6 of the training procedure. The outcome species is the class of the input image. 10. To find the accuracy of the system or the SVM, in this case, random set of inputs are chosen for training and testing from the database. Two different sets for train and test are generated. The steps for training and testing are same, however, followed by the test is performed.

Multiclass SVM classification

The binary classifier which makes use of the hyper-plane which is also called as the decision boundary between two of the classes is called as Support Vector machine (SVM). Some of the problems of pattern recognition like texture classification make use of SVM. Mapping of nonlinear input data to the linear data provides good classification in high dimensional space in SVM. The marginal distance is maximized between different classes by SVM. Different kernels are used to divide the classes. SVM is basically binary classifier which determines the hyper plane in dividing two classes. The boundary is maximized between the hyper plane and the two classes. The samples that are nearest to the margin will be selected in determining the hyper plane are called as support vectors.
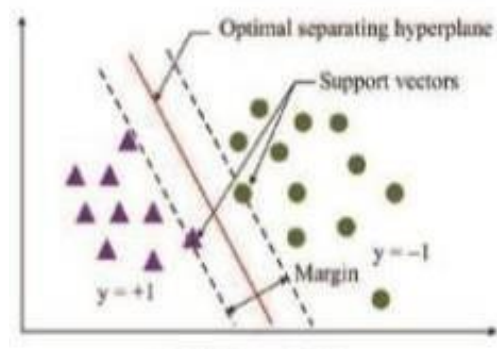

Figure 1. Linear SVM

Figure 1 shows the concept of support vector machine. Multiclass classification is also possible either by using one-to-one or one-tomany. The highest output function will be determined as the winning class. Classification is performed by considering a larger number of support vectors of the training samples. The standard form of SVM was intended for two-class problems. However, in real life situations, it is often necessary to separate more than two classes at the same time. In this Section, we explore how SVM can be extended from binary problems to multi classification problems with k classes where k > 2. There are two approaches, namely the one-against-one approach and the one-against-all approach. In fact, multi-class SVM converts the data set to quite a few binary problems. For example, in one-to-one approach binary SVM is trained for every two classes of data to construct a decision function. Hence there are k (k−1)/2 decision

functions for the k-class problem. Suppose k = 15, 105 binary classifiers need to be trained. This suggests large training times. In the classification stage, a voting strategy is used where the testing point is designated to be in a class having the maximum number of votes. The voting approach is called the "Max Wins" strategy. In one-against-all approach, there will be one binary SVM for each of the class to isolate the members of one class from the other class.

CONCLUSION

The plant serves as the basic need for any living organisms. They are the most important and integral part of our surroundings. Just like a human or other living organism does plant do suffer from different kind of diseases. Such diseases are harmful to plant in a number of ways like can affect the growth of the plant, flowers, fruits, and leaves etc. due to which a plant may even die. So in this work, we have proposed a novel method named as Bacterial foraging optimization based Support vector machine with cascaded Network for identification and classification of plant leaf diseases. The results, when compared with other methods, show that the proposed method achieves higher performance both in terms of identification and classification of plant leaf diseases. The proposed method is also superior in terms of computational efficiency for identification and classification of disease.

REFERENCES

[1] S. Zhang, X. Wu, Z. You, and L. Zhang, ''Leaf image based cucumber disease recognition using sparse representation classification,'' Compute. Electron.Agricult., vol. 134, pp. 135–141, Mar. 2018, doi: 10.1016/j.compag.2018.01.014.

[2] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, ''Identification of rice diseases using deep convolutional neural networks,'' Neurocomputing, vol. 267,pp. 378–384, Dec. 2017, doi: 10.1016 /j.neucom. 2019. 06.023.

[3] T. Akram, S. R. Naqvi, S. A. Haider, and M. Kamran, ''Towardsrealtime crops surveillance for disease classification: Exploiting parallelism in computer vision,'' Comput. Electr. Eng., vol. 59, pp. 15–26, Apr. 2017,doi: 10.1016/j.compeleceng.2019.02.020.

[4] T. N. Tete and S. Kamlu, ''Plant disease detection using different algorithms,'' in Proc. Int. Conf. Res. Intell. Comput.Eng., vol. 10. 2021, pp. 103–106, doi: 10.15439/2017R24.

[5] V. Singh and A. K. Misra, ''Detection of plant leaf diseases using image segmentation and soft computing techniques,'' Inf. Process.Agricult.,vol. 4, pp. 41–49, Mar. 2020, doi: 10.1016 /j.inpa. 2016.10.005.

[6] S. Megha, C. R. Niveditha, N. SowmyaShree, and K. Vidhya, ''Image processing system for plant disease identification by using FCM clustering technique,'' Int. J. Adv. Res., Ideas Innov. Technol., vol. 3, no. 2,pp. 445–449, 2017.

[7] L. Yuan, Z. Bao, H. Zhang, Y. Zhang, and X. Liang, ''Habitat monitoring to evaluate crop disease and pest distributions based on multi-source satellite remote sensing imagery,'' Optik—Int. J. Light Electron Opt., vol. 145,pp. 66–73, Sep. 2021, doi: 10.1016/j.ijleo.2017.06.071.

[8] P. B. Padol and A. A. Yadav, ''SVM classifier based grape leaf disease detection,'' in Proc. Conf. Adv. Signal Process. (CASP), Jun. 2016, pp. 175–179, doi: 10.1109/CASP.2016.7746160.

[9] D. J. Pujari, R. Yakkundimath, and A. S. Byadgi, ''SVM and ANN based classification of plant diseases using feature reduction technique,'' Int. J. Interact. Multimedia Artif.Intell., vol. 3, no. 7, pp. 6–14, 2021,doi: 10.9781/ijimai.2016.371.