International Journal on Emerging Technologies **11**(3): 707-714(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Morphological based Optimized Random Forest classification for Indian Oxygen Plants

Shilpi Aggarwal[1], Rosy Madaan[2] and Madhulika Bhatia[3]
[1]Assistant Professor, Faculty of Engineering and Technology,
Manav Rachna International Institute of Research and Studies, Faridabad (Haryana), India.
[2]Associate Professor, Faculty of Engineering and Technology,
Manav Rachna International Institute of Research and Studies, Faridabad (Haryana), India.
[3]Associate Professor, Department of Computer Science Engineering,
Amity University, Noida (Uttar Pradesh), India.

(Corresponding author: Shilpi Aggarwal)
(Received 19 March 2020, Revised 09 May 2020, Accepted 11 May 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Pollution is growing rapidly in the environment. There are number of factors like burning of forest, burning of fossil fuels, traffic pollution etc. Various efforts are being planned to maintain the environmental balance. We have 391,000 species of vascular plants in the world which are presently investigated by science, so there is a need to classify these species. In the present scenario, an optimized algorithm is required to classify the plants. In this study, the chosen dataset of Indian plants species belongs to five non identical categories which are rich in oxygen namely *Ocimum tenuiflorum*, *Sansevieria trifasciata*, *Chlorophytum comosum*, *Azadirachta indica*, *Aloe vera*. These samples were initially pre-processed by image processing in MATLAB 2019a. From the sample images dataset, morphological features like texture, shape, color and corner were extracted from the processed input image samples by using grab cut method, Gray co- level co-occurrence matrix (GLCM). Further the machine learning techniques i.e. Support vector machine (SVM) classifier, Random Forest and MLP Classifiers are applied on the features extracted from the processed samples. For optimizing the classification results, one hot encoding has also applied on the dataset. As a result, the prediction model gets the accuracy of 96 percent in case of SVM, 96.6 percent in case of Random forest and 86.6 percent in case of Multilayer Perceptron respectively. Among the three classifiers schemes random Forest gives the best results of the classification.

Keywords: Classification, oxygen**,** Plants, Prediction, Random Forest, SVM,

Abbreviations: SVM*, Support Vector Machine*; K-NN,K-Nearest Neighbor; MMC, Move Median Center; GLCM, Gray-Level Co-Occurrence; MLP, Multilayer Perceptron; PCA, Principal Component Analysis; ROC, Receiver Operating Characteristics; TPR, True Positive Rate; FPR, False Positive Rate.

I. INTRODUCTION

According to the biological department of the world, there are total 391,000 species of plants exists in the world's nature and out of that 21 percent of the plant species are on the way of extinction. Due to lack of awareness of plant knowledge, plant species are becoming rare and many of them are about to exterminated [1]. As the deforestation is going in a seamless way so it is important to sustain the list of the plant species that are supportive in preserving our environment from the uneven environmental circumstances like pollution. Plants play a vital role in maintaining the ecological balance of the environment. There are various types of plants like herbal, medicinal, antioxidants etc. As huge number of plant species is in hand of nature so there is a need to classify them. Classification is accomplished by human beings through analyzing the plants with the support of their sense organs. But as far as electronic machines are concerned, we have to provide an algorithm or a procedure as a result of which, identification [2] of the plants happens. There are many traditional classification approaches like K-NN, Naïve Bayes, SVM or their hybrid combinations can be used to give better accurate results [3]. Computer vision techniques can also be used for identifying plants [4]. Pal and Mitra (1992) proposed the Multilayer Perceptron model with back propagation which was capable to classify fuzzy sets which is implemented on the speech recognition system [5]. Amin and Khan (2013) proposed a recognition scheme called Distributed Hierarchical Graph Neuron (DHGN) and applied the scheme on K-NN for the purpose of classification. He got the average recall accuracy of 71.5% [6].

In 1986, the researcher Guyer recognized the plant images by using Bayes classifier for the first time [7]. The researcher Qingfeng work on 6 species of plants and calculated aspect ratio, leaf dent, leaf vein, edges [8-10] and invariant moment to identify plant [11]. Further, Stephen used probabilistic neural network to classify 32 types of plants. Many features such as aspect ratio (ratio between length and width of leaf), ratio of perimeter to diameter of leaf, and vein features were used to characterize the leaf with accuracy of 90.3% [12]. Wang *et al.,* (2008) proposed an efficient classification framework in which they used Watershed segmentation method combined with pre-segmentation and morphological operation to segment leaf images with complicated background based on the prior shape information. Authors extracted seven Hu geometric moments and sixteen Zernike moments as shape features from segmented binary images after leafstalk removal. They applied MCH classifier was used to classify twenty plant leaves with an average classification rate of up to 92.6% [13].

Classification [27-29] can be done on the basis of shape features extracted by using graph based methods [15], Move Median Center (MMC) classifier [16]. Feature extraction can also be done by using semi supervised locally linear embedding technique [17], combined with K-NN classification. Chaki *et al*., (2018) used hierarchical approaches to classify plants based on hue, shape and texture and proposed a feature based shape selection technique for the choice of shape features [18]. The combination of shape [18], texture, color [19, 20] was also used in automatic classification [21, 22]. Palanisamy *et al.,* (2019) used K-means clustering to classify 70 leaf images on the basis of colors. In this, seven different neural networks were used for the classification. Out of seven classifiers, regression and Radial Baisneural networks found to have a better performance [23]. Saleem *et al.,* (2019) proposed an algorithm which is tested on Flavia and self collected dataset of 625 leaves. Different classifiers were tested on the dataset and among all KNN performed better accuracy [24].

A very renowned technique called digital image processing [25, 26] is used to process the images which is used to extract the useful and valuable features. An image is said to be a 2-D function f(x,y), where x and y are the two distinguish coordinates. The amplitude of function f(x,y) is called as a gray level, intensity. When the intensities and coordinate values combined, it forms a digital image. Image processing has enormous implementations in distinct areas such as computer vision, remote monitoring, microscopic images, medical images [27-29], astronomy, patter recognition etc.

The different researchers applied the classification techniques on the various predefined leaf datasets like Flavia, Imageclef, Plantscan etc. and this work is accomplished on the self collected dataset. The samples are rich in oxygen that absorbs various contaminates present in the air. Discussion regarding the dataset is done in the materials and methods section. We have shown classification by optimizing the classifiers according to the algorithms defined in Table 1 and 2.

Initially, there is background study and introduction to classification and then the various classification schemes applied by various researchers. Next, we move forward with the materials and methods required for the work of plant classification. Further, the result corresponds to classification. Finally, the last section concludes the paper.

This paper contributes to the work of classification of the oxygen releasing plants in order to combat the environmental imbalance. The plant species (*Ocimum tenuiflorum*, *Sansevieria trifasciata*, *Chlorophytum comosum*, *Azadirachta indica*, *Aloe vera*) are preprocessed with the help of image processing techniques then different features were shown like shape, texture, corners and maximum color values which vary from plant to plant. Different machine learning [30, 31] classification models are applied to classify the samples belongs to different classes.

II. MATERIALS AND METHODS

A. Image Dataset

There are total 150 RGB image samples of oxygen rich plants of 5 distinct species used in this study are collected from the city of Faridabad (Haryana, India) as shown in Table 1. The choice of these samples from the Indian species was because of ease in availability, antioxidant properties of holy basil, ability of *Sansevieria trifasciata* to absorb carbon monoxide from the environment. *Chlorophytum comosum* also purifies the air and absorb carbon monoxide and formaldehyde and xylene. *Azadirachta indica* is the medicinal plant that is used in curing many problems like loss of appetite, gingivitis, eye disorders and liver problems, and Aloe Vera is having antioxidant and antibacterial properties. These samples were observed for continuous 15 days and images were captured with compact digital still camera having 12.1 Megapixel camera having resolution of 4000 × 3000 pixels in the period of 10th April 2018 to 26th April 2018.

All the samples of plant images were first converted into the gray-scale images then they undergone the process of segmentation which was done by Grabcut algorithm [32]. Further the features like texture, shape, corners, color were extracted from the segmented image. All the progressive results are shown in Fig. 1 (Column 1 to 4).

Table 1: Samples species.

S. No.	Scientific Names	Common Names
1.	*Ocimum tenuiflorum*	Holy Basil
2.	*Sansevieria trifasciata*	Snake
3.	*Chlorophytum comosum*	Spider
4.	*Azadirachta indica*	Neem
5.	*Aloe vera*	Indian Aloe

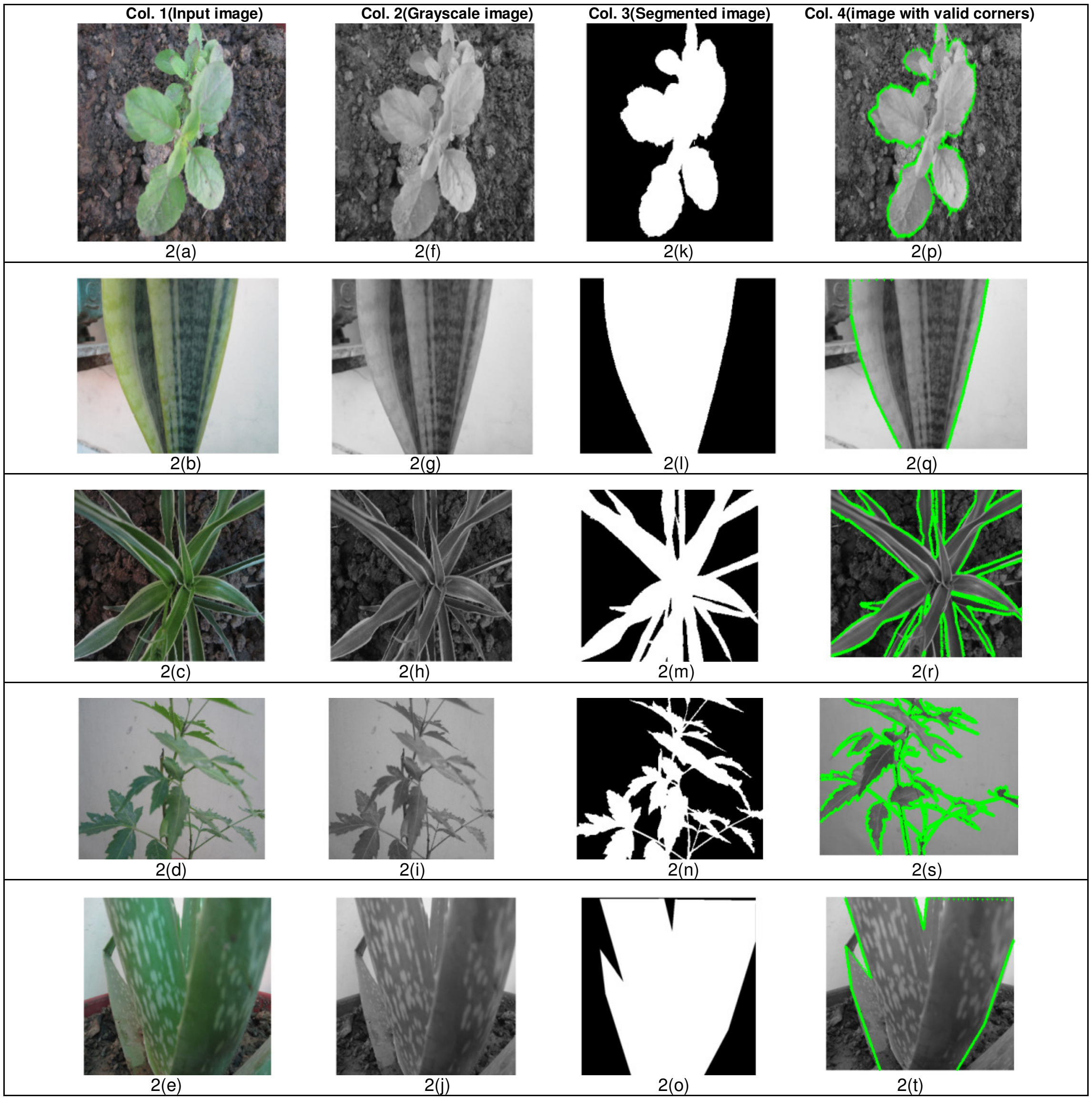


Fig. 1. Sample image belongs to different classes of species used in this research. Columns 1(a-e) shows the sample images of the plants. Column 2(f-j) shows the gray-scale images of the samples. Column 3(k-o) represents the segmented image. Column 4(p-t) shows the images of highlighting the corner feature.

B. Co-occurrence Matrix

Texture refers to the variation in the intensities of the gray levels of an image. A statistical method of analyzing texture of the image by considering the spatial relationship of the pixels is termed to as gray level co-occurrence matrix. GLCM is created by implementing a graycomatrix function [33]. This function evaluates the frequent occurring of a pixel i having gray level intensity value with a spatial association to a pixel having value j. In our study, we have created multiple i.e. 16 GLCM's and then several statistics like contrast, correlation, energy, homogeneity were calculated. Then further, average is taken.

C. Shape Dimensions

Morphological features like area, perimeter, minor axis length, major axis length were calculated by using the algorithm used [34].

D. Maximum color values

We have taken the maximum value of the red, green and blue values from the image samples.

E. One Hot encoding

Encoding is one of the procedures of preprocessing the data. Basically, to encode the data in which a computer machine can accept. There are number of encoding schemes like label encoding, frequency encoding etc. Out of all, one hot encoding is worn in this work. One hot encoding is a scheme which permutes the categorical data into binary values with one 1's and other 0's. It assists to train the data effectively. After encoding it will be facile for the machine learning algorithms to do classification. In our dataset one hot encoding is put in on the target values.

F. Feature Extraction

There are total 13 features extracted from the samples. Corners and valid corners features were found using Harris-Stevens algorithm [35]. By applying the algorithm on the threshold binary image some of the important feature points were detected called the corners and valid corners. Shape features were extracted by using the algorithm used in [34]. Texture features were pulling out by using the gray co-occurrence matrix and color max values were filtered out from the samples.

The algorithm 1 used for extracting features is given in Table 2.

Algorithm 2 is shown in Table 3 which is used to implement the Random Forest and MLP Classification Model on the features extracted by algorithm 1. This algorithm is implanted by using python platform.

G. Feature Selection

The features like shape (Area, Perimeter, Minor axis length, Major axis length), texture (Contrast, Homogeneity, Energy, and Correlation), corners (corners and valid corners) and color (maximum values of R, G, B values) are filtered out to provide the best working of the classification model as shown in Fig. 2. So out of total 13 features only 10 are selected and applied on SVM Classification Model. In case of Random Forest Classifier and MLP Classifier, all the features are selected .

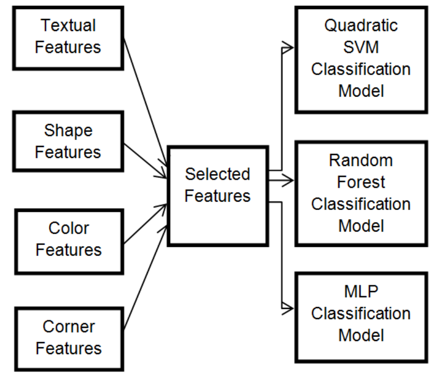


Fig. 2. Feature Selection and reduction hierarchy.

Table 2: Algorithm to calculate the features like color, texture, shape, corners.

Algorithm-1. Calculate Features like texture, shape, color, corners	
Input: I is the input sample RGB image Output: Contrast, Homogeneity, Energy, Correlation, Area, Perimeter, Minor axis length, Major axis length, redmax, greenmax, bluemax, Corners, Valid Corners.	
1.	Start
2.	While (Image Counter<=0) Then
3.	Set I← Input RGB image
4.	Set I_G ←Gray-scale image(I)
5.	Set I_B← Grab Cut (I_G)
6	Set [glcm] ←graycomatrix(I_G,'Offset',[0 1; 0 2; 0 3; 0 4;... -1 1; -2 2; -3 3; -4 4;... -1 0;-2 0;-3 0;-4 0;... -1 -1;-2 -2;-3 -3;-4 -4]);
7.	Set[contrast, homogeneity, correlation,energy] ←graycoprops (glcm)
8.	Set [corners] ←detect Harris Features (I_B)
9.	Set[features1,valid_corners] ←extract Features(Bw, corners)
10.	Set [Area, Perimeter, Minor axis length, Major axis length] ←regionprops (I_B)
11.	Set [redmax, greenmax, bluemax] ←max(I)
12.	End while loop
13.	Features are selected from above extracted features.
14.	Apply the selected features on the SVM classification model $\text{class}(z) = sign\left(z'\beta\hat{} + b\hat{}\right) = \text{sign}(f\hat{}(z))$
15.	Evaluate the model.
16.	End

Table 3: Algorithm for Random Forest and MLP Classifier.

Algorithm 2. Test the dataset on different Machine Learning Models.	
Input: Features extracted in the Algorithm1. Output: Accuracy.	
1.	Start
2.	Load the Required Packages.
3.	Load the dataset.
4.	Apply the Hot Encoding on target values.
5.	Dividing the dataset into independent and dependent attributes.
6	Separate the training and test data.
7.	Apply the Random Forest Classifier and MLP Classifier model on the training data.
8.	Evaluate the Models by calculating its accuracy and confusion matrix.
9.	End

H. Evaluation of diagnostic model

As there are number of classification models available in the history of classification, but in this study we have used SVM quadratic classification model, Random Forest Classifier and MLP Classifier. The total samples are categories into 5 distinguished classes C1, C2, C3, C4, and C5. SVM gives the appropriate results for the set of samples considered in this paper for the classification.

I. SVM Classifier

Support vector machine is the supervised machine learning algorithm. It makes a hyper-plane among the classes so that the margin between the classes and the plane is the maximum. Support vectors are the small ratio of the training objects that are used to calculate the maximum distance to the hyper-plane. In our work, we have used this model for training the dataset and for classifying the samples in 5 different classes.

J. Random Forest Classifier

Random forest classifier [36] which comprises of is the classifier that comprises of tree structured classifier $c(x, \theta_k), k = 1, \ldots$ where, θ_k are autonomous indistinguishable scattered random vectors and each tree selects the most prominent class at input x. When there are group of classifiers $c_1(x), c_2(x), c_3(x), c_4(x) \ldots\ldots c_k(x)$, given with the training set chosen at random W, Z, define the margin function as

$$mg(W,Z) = av_k I\,(c_k(Z) = W) - \max_{j \neq W} av_k\, I(c_k(Z) = j)$$

Where, $I(.)$ is the indicator function. The margin evaluates the range of average number of selections at W, Z for the correct class oversteps the average selection of any other class. The huge margin leads to additional effectiveness in the classification. One advantage of using Random forest is that it doesn't overfits, when the count of trees increases and disadvantage is that it gives rise to a certain estimate of generalization error.

K. MLP Classifier

The Multilayer Perceptron [37] is widespread model which is used for classification and regression problems. MLP is based on feed forward neural network which involve of various layers of modes having unidirectional networks often trained by the mode namely back propagation. The total count of inputs in the networks is

$$y_j^{d+1} = \sum_i x_i^d\, z_{ji}^d - \theta_j^{d+1}$$

Where, x_i^d is the state of ith neuron in the previous dth layer, z_{ji}^d is the weight of the connection among the two adjacent layers of the network and θ_j^{d+1} is the threshold of jth neuron in the layer d+1.

The output neuron can be represented in a monotonic form of non linear function of its input which is given by

$$x_i^d = \frac{1}{1 + e^{x_j^d}}$$

During the learning procedure of MLP, the input vector x is n dimensional and output vector is also m dimensional. Therefore, the learning becomes slow. The least mean square can be represented as

$$E(w) = \frac{1}{2}\sum_{j,c}(x_{j,c}^d(w) - d_{j,c})^2$$

Where, $x_{j,c}^d(w)$ is the state retrieved for resultant node j in layer D in input output case c, and $d_{j,c}$ is the desired ratio.

Gradient decent can be used to minimize the error. From equations 1,2,3 we get

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial y_j} x_j{}^d(1 - x_j{}^d)x_i^{d-1}$$

For the output layer (d=D), we substitute in equation 4

$$\frac{\partial E}{\partial_{yj}} = y_j^D - d_j$$

The basic idea is first use the forward pass then backward pass, to compute the activity levels of all the neurons in the network. At last, use back propagation to allow weight updating until the input layer is reached.

III. RESULTS AND DISCUSSION

A. Classification

The researcher Salman A. *et al.,* (2017) uses canny edge detection scheme to extract features of leaf samples and classify them with SVM classifier and obtained 85% of accuracy [38]. Caglayan *et al*., (2013) applied various classification algorithms like K-NN, Naïve Bayes, Random Forest and SVM. He applied these models on the shape and color features of the leaves and obtained the accuracy of 96% in case of random forest classifier [39].

In our work, we have implemented the Quadratic SVM classification model, Random Forest Classifier, MLP Classifier on the features like color, texture, shape, corners to classify 150 samples of plant images. Quadratic SVM classification model is implemented using MATLAB 2019a .The total image set was divided into two parts. Training set and the test set.

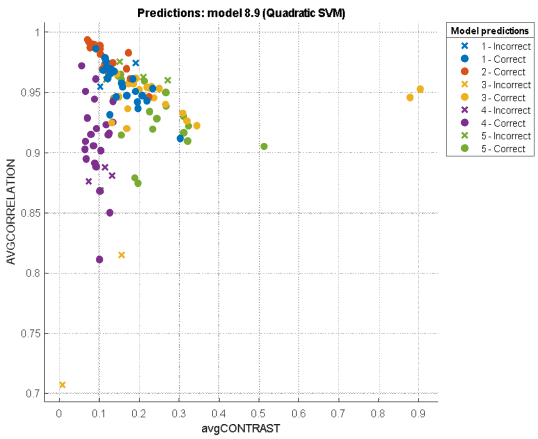
Training set consists of 130 samples and remaining was considered as test set. All the results of the classification model are shown in Table 4.

The model is trained at the prediction speed of 2500 obs/sec (approx.) and took the training time of 1.675 seconds. The model is set to be quadratic SVM with kernel function as quadratic and box constraint level is set to 4. PCA was disabled while training of data. The 5 cross validation was used while training. By applying all the parameters the model got the accuracy of 96%.

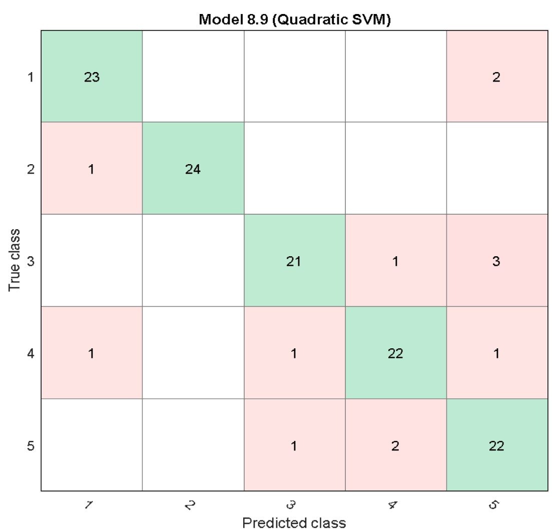
Table 4: Classification Results.

Plant species	No. of samples trained (130)	No. of samples test(25)	Total no. of Samples correctly classified after Training	Total no. of Samples correctly classified after Testing
OcimumTenuiflorum	25	5	23	4
SansevieriaTrifasciata	25	5	24	5
Chlorophytum Comosum	25	5	21	5
AzadirachtaIndica	25	5	22	5
Aloe Vera	25	5	22	5
Overall Accuracy			89.6%	96%

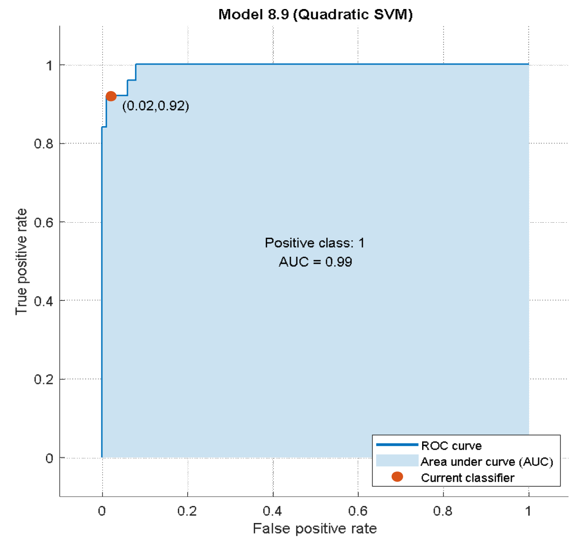
As in Fig. 3 (a) shows the scatter plot of the 130 samples out of whole dataset as the prediction model. Filled circles and cross (x) of dissimilar colors shows correct and incorrect predictions of the different classes (C1, C2, C3, C4, C5). This plot is drawn against the two important texture features one is contrast and another is correlation.



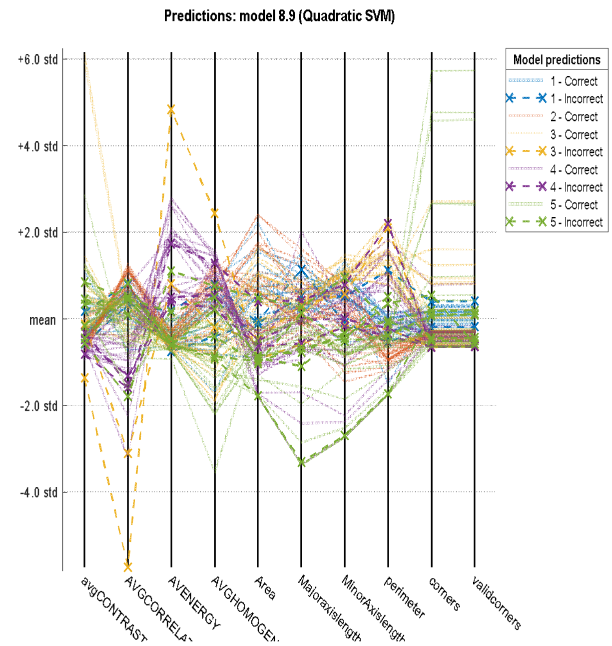
(a) Scatter plot of the training dataset out of 130 samples



(b) Confusion matrix of the classified samples.



(c) ROC curve between the TPR and FPR.



(d) Correct and incorrect predictions of samples into 5 classes.

Fig. 3. Results of the SVM Quadratic Classification Model.

Fig. 3(b) shows the confusion matrix that illustrates the correct number of predictions of each and every class diagonally.

Model is evaluated by using the ROC curve. Measuring the area under the ROC curve is important. Basically, the ROC curve is the ratio of true positive rate and the false negative rate as shown in Fig. 3(c).

In this study of the plants classification, the TPR is true positive rate and are defined as correctly or truly classified samples and FPR is false positive rate which means the ratio between the number of plants that are predicted as not belong to that class correctly and the total number of samples who are not having that class in actual. *Correct and incorrect* predictions can be shown by parallel coordinates plot as shown in Fig. 3(d).

In order to do Random Forest Classification, Algorithm 2 is followed. The features extracted from the input sample images of distinguished species are used as the dataset. It is divided in the 80% and 20% training and test data. In this case, the configuration of the model is set to be in effective manner to achieve the best accuracy. The number of trees in the forest is taken as 1000. The Gini criterion is opted for the quality of spilt for the information gain. The maximum depth is set to 10. Random state is chosen to be 1 for the best division of features at each node. The confusion matrix obtained in Fig 4. When the Random Forest Classifier with such configuration is applied on the dataset, the accuracy obtained is 96.6%.

	C1	C2	C3	C4	C5
C1	6	0	0	0	0
C2	0	7	0	0	0
C3	0	0	8	0	0
C4	1	0	0	5	0
C5	0	0	0	0	3

Fig. 4. Confusion Matrix Obtained by Random Forest Classifier.

In case of MLP Classifier, the configuration of the classifier is set up by fixing the values of the attributes in the function. The solver is used for weight optimization. Its value is lbfgs which is used as an optimizer that is present in the family of quasi- newton methods. Alpha parameter value is set to 1e-5.The attribute hidden layer sizes is set to (150, 10) which means the i^{th} element represents the count of neurons in the ith hidden layer. Random state is equal to 1 which means best division of features. After all the settings the accuracy obtained is 86.6%.

	C1	C2	C3	C4	C5
C1	6	0	0	0	0
C2	0	7	0	0	0
C3	3	0	5	0	0
C4	1	0	0	5	0
C5	0	0	0	0	3

Fig. 5. Confusion Matrix of MLP Classifier.

IV. CONCLUSION

The air pollution leads to worse the environmental conditions. This causes number of skin and respiratory diseases like atopic dermatitis, eczema, psoriasis or acne, skin cancer, asthma, lung cancer etc. Through this paper, a tiny step has been forwarded to classify the plant samples that are rich in oxygen. In this paper, sequential steps have been followed according to the algorithm. Initially, all the input plant images which belong to five distinguish species (*Ocimum tenuiflorum*, *Sansevieria trifasciata*, *Chlorophytum comosum*, *Azadirachta indica*) were pre-processed by existing techniques of image processing in MATLAB 2019a.Secondly, distinguish relevant features were extracted from the sample images by sing GLCM and grab cut method. Thirdly, the machine learning approach of classification that is SVM, random Forest and MLP are applied on the extracted features. This gives the accuracy of 96.3 percent, 96 and 93.3 percent respectively. Out of three, Random Forest Classifier performs the best. This means optimized random forest classifier classified the sample correctly with 96.6 percent accuracy.

V. FUTURE SCOPE

To classify the oxygen rich plants is really important to control the environmental imbalance. In future, the dataset can be extended to work on. This model can be applied on other plants datasets to check the performance of the model. Other classification models like Gradient Forest Classification algorithm can also be applied on the dataset used in this research for the better accuracy.

Conflict of Interest. No Conflict of Interest

REFERENCES

[1]. Valliammal, N., & Geethalakshmi, S. N. (2009). Analysis of the classification techniques for plant identification through leaf recognition. *Int. J. of Data Mining Knowledge Engineering*, *1*(5), 239-243.

[2]. Pandey, M., Bhatia, M., & Bansal, A. (2016). IRIS based human identification: analogizing and exploiting PSNR and MSE techniques using MATLAB. In *2016 international conference on innovation and challenges in cyber security (ICICCS-INBUSH)*, 231-235.

[3]. Aggarwal, S., & Bhatia, M. (2019). Anatomy of Leaf Classification Techniques. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 510-513.

[4]. Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering*, *25*(2), 507-543.

[5]. Pal, S. K., & Mitra, S. (1992). Multilayer perceptron, fuzzy sets, classification. *Transactions on neural networks, 3*(5), 683-997.

[6]. Amin, A. H. M., & Khan, A. I. (2013). One-shot classification of 2-D leaf shapes using distributed hierarchical graph neuron (DHGN) scheme with k-NN classifier. *Procedia Computer Science*, *24*, 84-96.

[7]. Guyer, D. E., Miles, G. E., Schreiber, M. M., Mitchell, O. R., & Vanderbilt, V. C. (1986). Machine vision and image processing for plant identification. *Transactions of the ASAE*, *29*(6), 1500-1507.

[8]. Bansal, A. (2013). Implementing edge detection for detecting neurons from brain to identify emotions. *International Journal of Computer Applications*.

[9]. Yadav, D., Gupta, P., Kaur, G., Singh, J., Gandhi, M., & Singh, A. (2013). Implementing edge detection for medical diagnosis of a bone in Matlab. In *2013 5th International Conference and Computational Intelligence and Communication Networks* . 270-274.

[10]. Patel, A. K., Khandelwal, A. (2019) A Review of Techniques used for Edge Detection in Image. International *Journal of Electrical, Electronics and Computer Engineering, 8*(2), 6-9.

[11]. Wu, Q., Zhou, C., & Wang, C. (2006). Feature extraction and automatic recognition of plant leaf using artificial neural network. *Advances in Artificial Intelligence*, *3*, 5-12.

[12]. Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *2007 IEEE international symposium on signal processing and Information Technology*, 11-16.

[13]. Wang, X. F., Huang, D. S., Du, J. X., Xu, H., & Heutte, L. (2008). Classification of plant leaf images with complicated background. *Applied mathematics and computation*, *205*(2), 916-926.
[14]. Bhatia, M., Bansal, A., & Yadav, D. (2017). A proposed quantitative approach to classify brain MRI. *International Journal of System Assurance Engineering and Management*, *8*(2), 577-584.
[15]. Zhao, A.,Tsygankov, D., & Qiu, P. (2017). Graph-based extraction of shape features for leaf classification. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 663-666.
[16]. Du, J. X., Wang, X. F., & Zhang, G. J. (2007). Leaf shape based plant species recognition. *Applied mathematics and computation*, *185*(2), 883-893.
[17]. Zhang, S., & Chau, K. W. (2009). Dimension reduction using semi-supervised locally linear embedding for plant leaf classification. In *International conference on intelligent computing*, 948-955.
[18]. Chaki, J., Parekh, R., & Bhattacharya, S. (2018). Plant leaf classification using multiple descriptors: A hierarchical approach. *Journal of King Saud University-Computer and Information Sciences*.
[19]. Munisami, T., Ramsurn, M., Kishnah, S., & Pudaruth, S. (2015). Plant leaf recognition using shape features and colour histogram with K-nearest neighbour classifiers. *Procedia Computer Science*, *58*, 740-747.
[20]. Rehman, N., Sinha, P. (2013). Colour and Texture Identification Using Image Segmentation. *International Journal on Emerging Technologies, 4*(2), 69-75
[21]. Jamil, N., Hussin, N. A. C., Nordin, S., & Awang, K. (2015). Automatic plant identification: is shape the key feature? *Procedia Computer Science*, *76*, 436-442.
[22]. Murat, M., Chang, S. W., Abu, A., Yap, H. J., & Yong, K. T. (2017). Automated classification of tropical shrub species: a hybrid of leaf shape and machine learning approach. *Peer J.*, *5*, e3792.
[23]. Palanisamy, T., Sadayan, G., & Pathinetampadiyan, N. (2019). Neural network–based leaf classification using machine learning. *Concurrency and Computation: Practice and Experience*, e5366.
[24]. Saleem, G., Akhtar, M., Ahmed, N., & Qureshi, W. S. (2019). Automated analysis of visual leaf shape features for plant classification. *Computers and Electronics in Agriculture*, *157*, 270-280.
[25]. Bhatia, M., & Mittal, M. (2017). Big data & deep data: minding the challenges. *Deep Learning for image processing Applications*, *31*, 177-193.
[26]. Madaan, A., Bhatia, M., & Hooda, M. (2018). Implementation of image compression and cryptography on fractal images. In *Advances in Data and Information Sciences,* 49-61.
[27]. Bhatia, M., Bansal, A., Yadav, D., & Gupta, P. (2015). Proposed algorithm to blotch grey matter from tumored and non tumored brain MRI images. *Indian Journal of science and Technology*, *8*(17), 1-10.
[28]. Pandey, M. (2016). An amalgamated strategy for iris recognition employing neural network and hamming distance. In *Information Systems Design and Intelligent Applications*, 739-747.
[29]. Pandey, M., Bhatia, M., & Bansal, A. (2016). An anatomization of noise removal techniques on medical images. In *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, 224-229.
[30]. Harris, C. G., & Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference* (15, No. 50, 10-5244).
[31]. Joshi, A., & Mehta, A. (2017). Comparative Analysis of Various Machine Learning Techniques for Diagnosis of Breast Cancer. *International Journal on Emerging Technologies* (Special Issue NCETST-2017), *8*(1), 522-526.
[32]. Rother, C., Kolmogorov, V., & Blake, A. (2004). “GrabCut” interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, *23*(3), 309-314.
[33]. Robert, M. H., Shanmugam, K., & Its’hak, D. I. N. S. T. E. I. N. (1973). Textural Features for Image Classification.
[34]. Aggarwal, S., Bhatia, M., & Pandey, H. M. (2020). A New Method to Classify Leaves Using Data Visualization: Spreading Awareness about Global Warming. In *Proceedings of ICETIT 2019*, 1129-1139.
[35]. Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning. *Neural and Statistical Classification*, *13*(1994), 1-298.
[36]. Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.
[37]. Suter, B. W. (1990). The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Transactions on Neural Networks*, *1*(4), 291.
[38]. Salman, A., Semwal, A., Bhatt, U., & Thakkar, V. M. (2017). Leaf classification and identification using Canny Edge Detector and SVM classifier. International Conference on Inventive Systems and Control (ICISC), Coimbatore. 1-4.
[39]. Caglayan, A., Guclu, O., & Can, A. B. (2013). A plant recognition approach using shape and color features in leaf images. In *International Conference on Image Analysis and Processing*, 161-170.

How to cite this article: Aggarwal, S., Madaan, R. and Bhatia, M. (2020). Morphological based Optimized Random Forest classification for Indian Oxygen Plants. *International Journal on Emerging Technologies*, *11*(3): 707–714.