INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 16-05-2024

Accepted: 07-08-2024

Published: 24-08-2024

Citation: Gaba I, Ramamurthy B (2024) A Robust Solution for Tea Leaf Classification using the SMOTE and Soft Voting Classifier. Indian Journal of Science and Technology 17(34): 3502-3511. https://doi.org/10.17485/IJST/v17i34.1667

* **Corresponding author**.

ira.gaba@res.christuniversity.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

A Robust Solution for Tea Leaf Classification using the SMOTE and Soft Voting Classifier

Ira Gaba[1]*, B Ramamurthy[1]

1 Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

Abstract

Background/Objectives: The purpose of the proposed work is to identify how a tea leaf is conceptualized. This further enables the categorization of leaves into healthy and unhealthy groups, assisting both the vendor and customer in obtaining the right quality for their requirements. The primary cause of the low production of tea in various parts of the country is the late detection of diseased leaves. Literature suggests that the current AI techniques can detect but have a high level of similarities. Hence, the objective is to identify diseased leaf images at early stage. **Methods**: Earlier studies have used mostly CNN for classification which is more time consuming, and they create their own parameters. In this study, we extracted the textual features of images by using GLCM, and further classification was implemented with the help of SMOTE and a soft voting classifier. **Findings**: The proposed method resulted in optimal timeline without using CNN for classification. The accuracy is better when compared to current methods. **Novelty**: This work presents a novel approach in calculating textual features of leaves that can be further used for classification or tea-leaf grading purposes.

Keywords: Image processing; Feature extraction; GLCM; SMOTE; Voting classifier

1 Introduction

India's economy is growing mostly due to its agricultural sector. In addition to being exported outside India, tea also plays a significant part in that. It is necessary to identify the illness at an early stage of crop planting as it is a critical step in the manufacturing process of the finished product. Before the growth reaches the final stage of crop planting it is a critical step in the manufacturing process of the finished product. Before the growth reaches the final stage, the farmers should be aware of the illness and will be able to treat it on time. Both the farmer and the crop production will profit from it. Should the illness proliferate throughout the crop, it would swiftly harm the entire output, leading to subpar yield [1,2]. Based on the category of the leaves, farmers may be assured of a competitive edge for relatively excellent consistency. Occasionally, certain diseases are not obvious to the unaided eye, therefore automated crop analysis is crucial and a popular study issue for identifying diseases on leaves. The primary idea of computer vision and machine learning is applied to the identification and detection of leaves. Farmers can benefit from the application designed to identify diseases. However, occasionally an error might occur as a result of left noise in the image. numerous methods are being used in the same way. We can extract features for our work using the GLCM, and we have also integrated SMOTE with a voting classifier for classification [3].

Additionally, the paper is structured as follows: section II provides a background study, section III delves into the methodology and proposed approach of the work, section IV shows the results and finally, section V concludes the paper.

The main novelty of this work is the combination of the classification, the SMOTE, and the soft voting classifier. The GLCM is used for the textual parameter of leaves. It will help in the identification of the healthy leaf at the early stage. Furthermore, if the diseased leaves are detected at the early stage, this may also enable prompt resolution of the relevant problems and preservation of the corpse. The identification and prediction of the quality of tea leaves has been a subject of research in the field of agriculture. The main aim is to identify healthy leaves which can be used in my future work to identify the quality of the leaves. The proposed model's primary goal is to detect diseased leaves early on, which will aid in high-quality production. In future work, it will be used to distinguish between leaf quality categories, such as the best, medium, and worst quality, and prices for various classes can be estimated using this information. The primary goal of this effort is to distinguish between leaves so that they can be categorized in the future. This work differs from previous research since it uses a voting classifier in addition to SMOTE for classification. The vote classifier yields an effective accuracy result by averaging various classifiers and different algorithms. The novelty of this work is that texture features are extracted by using the GLCM, and for voting classifier five algorithms are used to obtain best result.

1.1 Background of the Study

In this section, we will be able to discuss some of the works which are already carried out for the detection. Mostly all of them have classified with implementing the CNN only. The features were extracted manually and separated from the dataset.

Hu Gensheng (2019) [1], was able to identify the diseased leaf, but they used manual extraction only for the diseased images, basically, they extracted the images by visual prediction they further used machine learning methods for classification and the accuracy was around 90. However, the main drawback of this paper was the feature extraction part.

S.Gayathri (2020) [2], was able to predict only the diseased leaf by using the LeNet which is a defined CNN model. The accuracy was almost 85% on average. They have just taken the diseased leaf into consideration.

Feng Jiang (2020) [3], was able to detect the four disease categories of the rice leaf. They have depicted the classification with the help of CNN and SVM. However, they only implemented the detection by using the color parameters. They combined CNN with SVM for the classification stage and the accuracy was more the 90%. According to the author, this work still needs to be carried out with other deep learning methods to increase accuracy.

Serosh Karim Noon (2020) [4], they were able to depict the stress of various leaves. They have taken the dataset from the PlantVillage. And for the classification, CNN is carried out only. To increase the size of the dataset image augmentation is done. Accuracy was above 90% with different models like AlexNet, GoogleNet, and Inception. They have used the pre-defined model for the classification. According to this work, more research needs to be done if there is more than one stress on a leaf. Different leaf stress can act differently.

Mohit Agarwal (2020) [5], was able to depict the diseased images of the tomato crop leaves. The classification was also implemented with CNN. In this work, the accuracy of proposed was compared with the work carried out by the other authors based on VGGNet, Inception, and MobileNet. This work can also be applicable for the CT scan images for detection. However, the feature extraction is only extracted with the help of HSV and LBP features.

Hu Gensheng (2019) [6], was able to depict the diseases like red leaf spot, tea leaf blight, and tea red scab, with the help of CNN models only. The features were not extracted instead segmentation is carried out. Based on the segmentation, further, C-DCGAN is performed to depict the disease spot. The accuracy of this work was almost 90%. The implementation was done on the diseased leaf only.

Lawrence C. Ngugi (2020) [7], was able to predict the leaf pest and detect the disease with image processing. However, the recognition is done with the help of IPTs. To calculate the accuracy CNN is implemented. The comparison is done based on 7 performance metrics. In this, they have used the pre-defined methods only.

Yuzhen Wei (2019) [8], in this work main focus is to determine the moisture content of tea leaves. The visual inspection method in a production plant essentially coincides with the scheme proposed in this research and is a successful solution to the problems caused by the differences between the different sides of the tea leaves.

Mehmet Metin Ozguven (2019) [9], in this study, an Updated Faster R-CNN architecture was developed by changing the parameters of a CNN model and a Faster R-CNN architecture for automatic detection of leaf spot disease. The proposed approach yielded better outcomes for relevant parameters. Totally 155 images were used for model training and testing. An accuracy of 95% was achieved with the present model in disease detection and classification.

Yunyun Sun (2018)[10], in this proposed work a new algorithm is developed with the combination of SLIC and SVM. The combination method of SLIC and SVM can accurately extract the saliency map of tea plant leaf disease, which is conducive to the accurate identification of tea leaf disease and lays an effective foundation for rapid and accurate detection and disease prevention.

1.2 Research Gap

There are a few research papers reviewed in this section, which will provide ideas for new work for researchers.

1. Deep learning has focused on the many applications of image and pattern recognition but practically it becomes very difficult to identify the diseased leaf at the initial stage. So is more attention required for this work.
2. A leaf can be affected by more than one disease. Some diseases can occur at the initial stage of growth some might get affected in the later stage of growth of leaf production. But no author had talked about that in detail in the literature. This can also become one of the future research areas.
3. In most of the journals, the author had only identified the diseased leaf, but in the future, we can also work on the quality of the leaf.
4. The implementation of the deep learning networks may differ from plant to plant. With CNN the accuracy rate may differ and with SVM it might differ. So it is important to find the best fit of the combination for the classifiers, which can provide more direction for future research.

1.3 Limitations of existing work

There are few limitations have been observed from the previous work carried out.

1. The existing work only uses CNN or SVM for classification and in most of the work, the accuracy rate is less.
2. The main challenge is the need for high-end processing hardware, which in turn is more expensive.
3. If the dataset is big, it is more time-consuming.
4. The dataset has to be of specific properties.
5. The existing work is implemented to differentiate between healthy and diseased leaves only.

2 Material and methods

2.1 Methodology

The following are the step-by-step procedures used for the detection of the diseases in the Tea leaves.

- **Image pre-processing:** The data was collected, and the initial step was to do the image processing. The main step of image processing is to remove the unwanted noise from the datasets. In this work, various filters were implemented on the dataset to remove the noise[4,5].
- **Feature Extraction:** The second step for the proposed work is to extract the features for the images of dataset. Feature extraction is implemented to extract the important features from the images and provides the required information needed for the classification. It can be based on various parameters such as the color, texture and shape of the leaves. However, the texture method is the one usually used[6].
- **Feature Selection:** The feature selection is implemented after the feature extraction, basically reducing a greater number of parameters. There are various types of selection methods that can be used. Like feature importance, information gain, etc.[7].
- **Classification:** The last step is classification; it will tell us whether the input images are healthy or unhealthy images. The main idea behind the use of feature selection is to identify the classification of the input images based on the parameters being extracted in the feature selection method[8,11]. The following figure depict the layout of proposed work. It basically shows the step wise flow of the work carried out in proposed model.

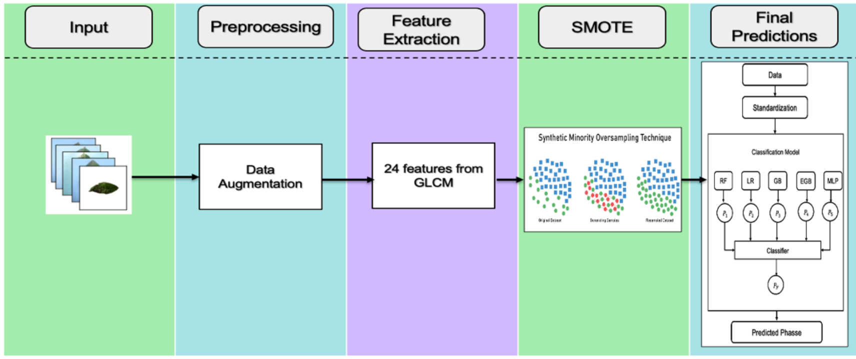

Fig 1. Layout of proposed model

2.2 Datasets

The datasets of the tea leaves images have been obtained from the Kaggle source and some of the images are real-time images too. All images in the dataset consist of healthy and unhealthy leaves. The main idea of the proposed model is to identify whether the input image belongs to the healthy or unhealthy leaves. The dataset consists of 1988 images of both healthy and unhealthy tea leaves. These images were used for the identification and classification of the two classes of healthy and unhealthy images [9,10]. The Tea leaves datasets utilized in this. Initially, the datasets were divided into training and test data sets through the application of the k-fold cross-validation technique. Subsequently, classification algorithms including SMOTE, and Soft Voting Classifier were employed to develop respective classification models using the provided training data. These classification models were then evaluated using the test data to determine their accuracy. Finally, a comparison was made between the accuracy rates of each implemented classification model, leading to a conclusive result. The objective of this study is to identify a tea leaf that is conceptualized. This further enables the categorization of leaves into several groups, assisting both the vendor and customer in obtaining the right quality for their requirements. Based on the category of the leaves, farmers may be assured of a competitive edge for relatively excellent quality and consistency. Furthermore, if the diseased leaves are detected at the early stage, this may also enable prompt resolution of the relevant problems and preservation of the corpse. The discipline of recognition is seeing rapid advancements in AI approaches, which when combined effectively may not only solve problems but also improve recognition accuracy.

2.3 Training and Testing

The dataset is the one that is further split into the testing and the training set. It is carried out by splitting the above images into two sets 70% in the training set and 30% of the testing set. Python was used for splitting the dataset. [8,11]

2.4 Proposed Approach

When The image acts as an input to the proposed system, the very first step is to conduct the image pre-processing on the image basically to remove the unwanted noise from it. The following figure shows a few images of both healthy and unhealthy leaves. After the pre-processing stage and next comes the feature extraction and the feature selection method. This part plays a very important role as the defiltration is carried out based on the features extracted from the image [12,13]. In the proposed model we have used the GLCM for the texture feature of the image.

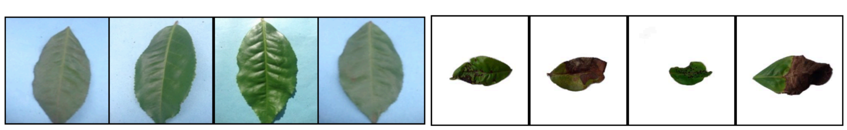

Fig 2. The above are the few images of the health and unhealthy tea leaves

Further the images are classified with the help of the SMOTE Classifier and the voting classifier.

2.4.1 GLCM:

The Grey Level Co-occurrence Matrices (GLCM) are used to calculate the texture parameters of the image. The GLCM consists of total 24 parameters such as Contrast, autocorrelation, joint average, cluster prominence, cluster shade, cluster tendency, correlation, difference average, difference entropy, difference variance, joint energy, joint entropy, informational measure of correlation (IMC) 1, informational measure of correlation (IMC) 2, inverse difference moment, maximal correlation coefficient, inverse difference moment normalized, inverse difference, inverse difference normalized, inverse variance, maximum probability, sum average, sum entropy, sum of squares. The following table shows the parameters being calculated by the GLCM. Table 1 represents the parameters and method to calculate those[14–16]. The following table will show the parameters that are calculated for all the 1988 leaf images collected, both healthy and unhealthy leaves.

Table 1. Parameters calculated with the help of GLCM

S.No.	Parameter	S.No.	Parameter
1.	Autocorrelation	13.	IMC 1
2.	Joint Average	14.	IMC 2
3.	Cluster Prominence	15.	IDM
4.	Cluster Shade	16.	MCC
5.	Cluster Tendency	17.	IDMN
6.	Contrast	18.	Inverse Difference
7.	Correlation	19.	IDN
8.	Difference Average	20.	Inverse Variance
9.	Difference Entropy	21.	Maximum Probability
10.	Difference Variance	22.	Sum Average
11.	Joint Energy	23.	Sum Entropy
12.	Joint Entropy	24.	Sum of Squares

2.4.2 Voting Classifier:

A voting classifier is a machine learning model that uses many models to activate and predict a result based on the output's maximum probability of the selected class. The computation is done using the highest possible number of votes. There are two types of voting: Hard voting and soft voting. In the proposed model we have used soft voting. In soft voting, the outcome of the class is computed on the average probability of the class. For instance, class A = (0.33, 0.45, 0.55) and class B = (0.22, 0.36, 0.45), so the average of class A is 0.44 and of class B is 0.34. Therefore, class A has the highest probability averaged by each classifier[17,18]. The main idea of using the voting classifier is to find average value of all the classifier methods. In the proposed model we have taken the following voting classifiers Random Forest classifier, logistic regression, Gradient boosting classifier, Extreme gradient boosting, and Multilayer perceptron.

1. **Random Forest Classifier:** An example of an ensemble classifier is this one. To improve the accuracy of the predictions, decision tree models are used. It creates a large number of trees and applies the bootstrap approach to each tree from the training data set. Every tree in the forest receives procedure input before casting a single vote for the class to which it belongs. Ultimately, the class with the highest number of votes is chosen by the RF.

2. **Logistic Regression:** Logistic regression employs statistical techniques to infer binary outcomes (y = 0 or 1). One type of linear learning algorithm is logistic regression. Logistic regression forecasts are expressed in terms of the likelihood that an event will occur. Each data point is mapped by the LR algorithm using the sigmoid function. The S-shaped curve is produced by the usual logistic function. The sigmoid function is displayed in following equation.

Sigmoid equation = 1/(1+e^(-x))

3. **Gradient Boosting Classifier:** Regression and classification tasks are handled by the Gradient Boosting classifier. A well-known decision tree implementation that beats random forests for a given set of data is the gradient boosting machine. Similar to other boosting techniques, they are constructed in a step-wise manner, but they are more adaptable because they may be used for any differentiable loss function. As a result, we fit each base model in the ensemble to the negative gradient of the loss function, forming a gradient descent.

4. **Extreme Gradient Boosting Classifier:** Machine learning models can be trained efficiently and scalable with the help of the XGBoost classifier. It is an ensemble learning technique that generates a stronger prediction by aggregating the predictions of several weak models. Extreme Gradient Boosting, or XGBoost, is a machine learning method that has gained popularity and

widespread usage because it can handle enormous datasets and attain state-of-the-art performance in many machine learning tasks, including regression and classification. XGBoost's effective handling of missing values is one of its primary characteristics, enabling it to handle real-world data with missing values without requiring a lot of pre-processing. Furthermore, XGBoost comes with built-in support for parallel processing, which enables training models on big datasets quickly.

5. **Multilayer Perceptron:** A multi-layer perceptron (MLP) is a type of artificial neural network consisting of multiple layers of neurons. The neurons in the MLP typically use nonlinear activation functions, allowing the network to learn complex patterns in data. MLPs are significant in machine learning because they can learn nonlinear relationships in data, making them powerful models for tasks such as classification, regression, and pattern recognition.

There above five classifiers are used for voting classifier to obtain the better accuracy value for proposed model.

2.4.3 SMOTE:

Synthetic Minority Over-sampling Technique is an over-sampling technique that is used to find the accuracy of the dataset. This algorithm is used to solve the problem of overfitting problem by random oversampling. The Receiver operating characteristic (ROC) curve[16] is a standard technique for the performance of true positive and false positive. The performance measures algorithms are calculated by the confusion matrix.

In the confusion matrix, TN is the number of negative correctly classified, FP is the number of negative incorrectly classified as positive, FN is the number of positive incorrectly classified as negative, and TP is the number of positive correctly classified[19].

	Predicted Negative	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

Fig 3. Confusion Matrix

The performance is measured by calculating the accuracy, recall, F1 score, and precision for the overall. The accuracy is calculated by the following equation.

Accuracy = (TP + TN)/(TP + FP + TN + FN)

The precision is telling us the predicted cases turned out to be correct and positive.

Precision = TP/(TP + FP)

Recall is the actual positive cases that were able to be predicted correctly.

Recall = TP/(TP + FN)

F1- score captures both precision and recall.

F1 score = 2/(1/Recall + 1/Precision)

The aforementioned values for the proposed model's accuracy, precision, recall, and f1-score are displayed in Figure 6. And for us to receive the result, the ROC is being used. To create the consecutive points on the ROC curve, the majority class is first over-sampled and then under-sampled at progressively greater degrees. The under-sampling quality is the same as that of simple under-sampling. Thus, starting with varying degrees of minority over-sampling, each corresponding point on each ROC curve is generated. Every point on the ROC curve represents the outcome of a classifier that was either taught using simple under-sampling, a classifier learned for a specific combination of under-sampling and SMOTE, or a classifier learned with simple under-sampling, a classifier learned using a different prior for the minority class, or a classifier learned using some loss ratio is represented by each point on the ROC curve[20–22].

3 Results

This section will include all the results of the proposed work being carried out to identify whether the image of the leaf belongs to the diseased section or the healthy category of the leaves. The **Supplementary table A** shows the values of all the parameters of GLCM for the health and the diseased leaves.

The **Supplementary table A** shows all the values of the GLCM parameters which will further be used for the classification so that to categorize the leaf into healthy or unhealthy leaf, these values are used by the voting and the SMOTE classifier. The table shows the calculated values for a total of eight images. It stores the values in the form of CSV files. Further, the prediction is calculated with the help of the confusion matrix. In this proposed method, the accuracy obtained is up to 92%.

3.1 Textual parameters calculated with the help of GLCM

The GLCM calculated textual parameters of image which is further carried out for the classification. In proposed model the main agenda is to concentrate on the textual features of leaf.

Supplementary table A depicts the various values calculated for healthy and diseased leaves, only few examples are shown. There are various parameters like contrast, homogeneity, brightness, cluster tendency, correlation are few parameters which play vital role in obtaining the accuracy in classification. The GLCM plays vital role in obtaining scalable and effective accuracy for proposed model as compared to existing models. All the values are saved in CSV files.

3.2 Confusion matrix calculated for Voting classifier

This section depicts the confusion matrix of all algorithms implemented in proposed model. Figure 4 **(a-e)** depicts confusion matrix for various classifiers used predicted by the proposed soft voting classifier. The algorithms used above are Radom Forest, Logistic Regression, Gradient Boosting, Extreme Gradient Boosting and Multilayer perceptron (MLP).

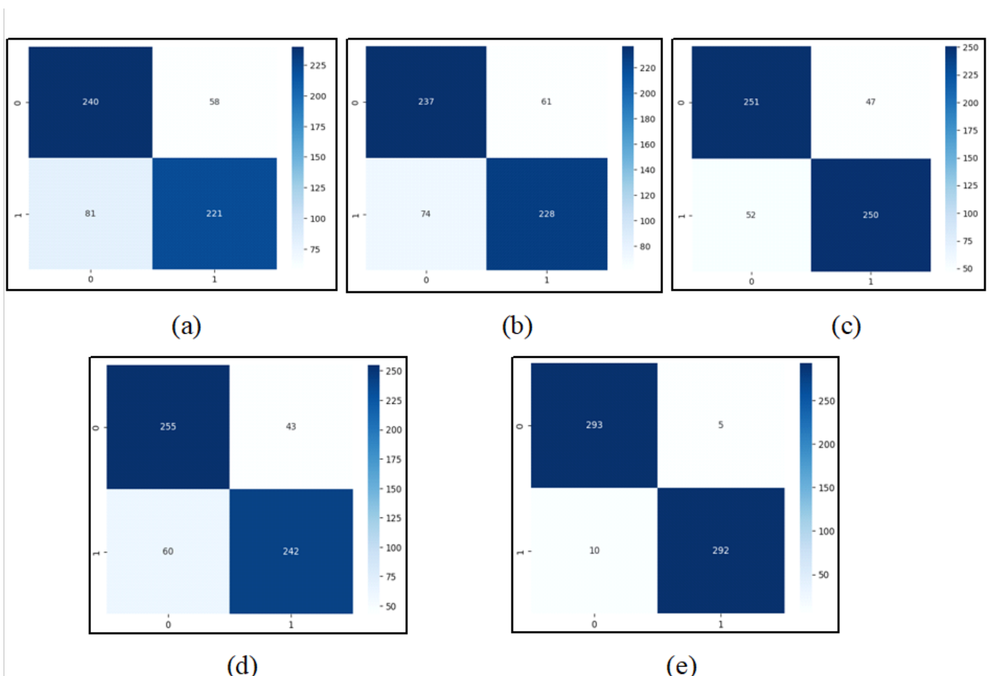

Fig 4. Confusion Matrix (a) Random Forest (b) Logistic Regression (c) Gradient Boosting (d) Extreme Gradient Boosting (e) Multilayer Perceptron

Further, for the classification the SMOTE was performed and the ROC was calculated which is depicted below. The graph's apparent ROC curve for voting classifier is close to 1 suggests that the proposed model has improved discriminative power and can successfully distinguish between positive and negative instances, which is extremely desirable. Figure 5 gives comparative Receiver Operating Chara (ROC).

Further, for the accuracy it is shown based on the confusion matrix. The following figure will depict the value of the same. It employed the classification's performance. The accuracy of pre-trained models in previous research work with the proposed work in this paper. The results from the proposed models are discussed in this section and a comparative analysis is shown, which spotlight the differences between our proposed model and existing models. In Table 2 the accuracy, precision, recall, F1

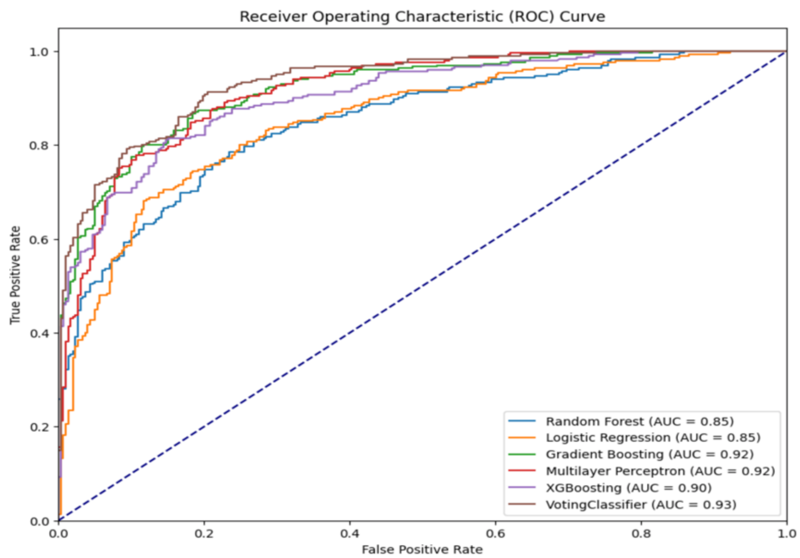

Fig 5. ROC curve comparison

score of the different models are compared with the existing models. It is observed that GLCM with SMOTE classifier stands out an accurate model as compared to others.

It can be drawn from the table that proposed model performs better than transfer learning approaches. Therefore, it can be implemented that proposed model is more computed and effective in determining capabilities.

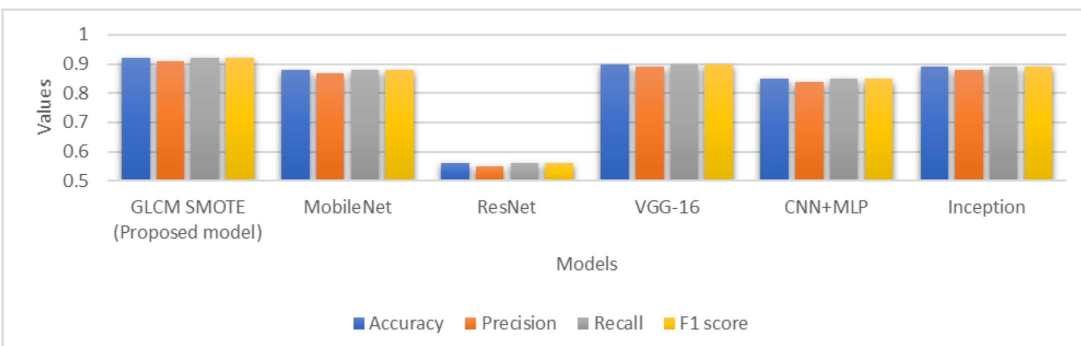

Fig 6. Comparison of proposed model with transfer learning models

The below table further shows the comparison of methods used in literature review with proposed model.

Table 2. Values of Accuracy being calculated from Previous Work compared with Proposed Work

Study	Year	Objective	Methods	Accuracy
Hu Gensheng et al.[1]	2019	Diseased leaf	CNN	89%
S.Gayathri et al.[2]	2020	Tea disease leaf	LeNet	85%
Feng Jiang et al.[3]	2020	Rice disease leaf	CNN and SVM	90%
Serosh Karim Noon[4]	2020	Different type of leaves	AlexNet, GoogleNet, Inception	90%
Mohit Aggarwal[5]	2020	Tomato disease leaf	CNN and VGGNet	87%
Hu Gensheng et al.[6]	2019	Tea disease leaf	C-DCGAN	90%
Lawrence C. Ngugi et al.[7]	2020	Leaf pest and disease	CNN	87%

Continued on next page

Table 2 continued

Yuzhen Wei et al. [8]	2019	Moisture content of tea leaves	CNN	85%
Proposed work	**2024**	**Tea healthy and disease leaves**	**SMOTE + voting classifier**	**91.3%**

The validation for proposed model is carried out with help of K-fold cross validation. It basically depicts the best accuracy percentage of the method. Sometimes it is important to observe perfect accuracy rate of the proposed method. The following figure depict cross validation and also the mean of it. This will help in obtaining the best accuracy rate.

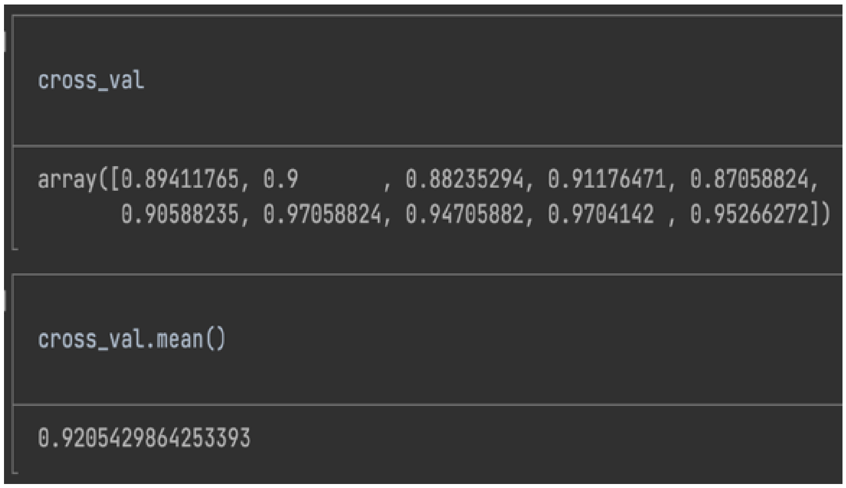

Fig 7. K-fold cross validation

4 Conclusion

The proposed work will be able to depict the tea leaves under the two categories: healthy or diseased. The proposed work was able to depict 92% of the work. Most of the previous studies have used the color features of the leaf for feature extraction, whereas, in this proposed work, the textual features were calculated with the help of GLCM. The novelty in this work is in using the SMOTE with a Voting classifier for classification. This work can further be used to find the quality of the leaf for grading purpose.

References

1) Gensheng H, Xiaowei Y, Yan Z, Mingzhu W. Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustainable Computing: Informatics and Systems*. 2019;24. Available from: https://doi.org/10.1016/j.suscom.2019.100353.
2) Gayathri S, Wise DCJW, Shamini P, Muthukumaram N. Image Analysis and Detection of Tea Leaf Disease using Deep Learning. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE. 2020. Available from: https://doi.org/10.1109/ICESC48915.2020.9155850.
3) Jiang F, Lu Y, Chen Y, Cai D, Li G. Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Computers and Electronics in Agriculture*. 2020;179. Available from: https://doi.org/10.1016/j.compag.2020.105824.
4) Noon SK, Amjad M, Qureshi MA, Mannan A. Use of deep learning techniques for identification of plant leaf stresses: A review. *Sustainable Computing: Informatics and Systems*. 2020;28. Available from: https://doi.org/10.1016/j.suscom.2020.100443.
5) Aggarwal M, Gupta SK, Biswas KK. Development of Efficient CNN model for Tomato crop disease identification. *Sustainable Computing: Informatics and Systems*. 2020;28. Available from: https://doi.org/10.1016/j.suscom.2020.100407.
6) Gensheng H, Haoyu W, Yan Z, Mingzhu W. A low shot learning method for tea leaf's disease identification. *Computers and Electronics in Agriculture*. 2019;163. Available from: https://doi.org/10.1016/j.compag.2019.104852.
7) Ngugi LC, Abelwahab M, Abo-Zahhad M. Recent advances in image processing techniques for automated leaf pest and disease recognition – A review. *Information Processing in Agriculture*. 2021;8(1):27–51. Available from: https://doi.org/10.1016/j.inpa.2020.04.004.
8) Weia Y, Wub F, Xuc J, Shaa J, Zhaod Z, Hea Y, et al. Visual detection of the moisture content of tea leaves with hyperspectral imaging technology. *Journal of Food Engineering*. 2019;248:89–96. Available from: https://doi.org/10.1016/j.jfoodeng.2019.01.004.
9) Ozguven MM, Adem K. Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms. *Physica A: Statistical Mechanics and its Applications*. 2019;535. Available from: https://doi.org/10.1016/j.physa.2019.122537.
10) Sun Y, Zhang ZJL, Dong W, Rao Y. SLIC_SVM based leaf diseases saliency map extraction of tea plant. *Computers and Electronics in Agriculture*. 2019;157:102–109. Available from: https://doi.org/10.1016/j.compag.2018.12.042.

11) Myburgh JC, Mouton C, Davel MH. Tracking Translation Invariance in CNNs. In: Southern African Conference for Artificial Intelligence Research (SACAIR 2021);vol. 1342 of Communications in Computer and Information Science. Springer. 2020;p. 282–295. Available from: https://doi.org/10.1007/978-3-030-66151-9_18.
12) Bradley AP. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*. 1997;30(7):1145–1159. Available from: https://doi.org/10.1016/S0031-3203(96)00142-2.
13) Chen YT, Chen SF. Localizing plucking points of tea leaves using deep convolutional neural networks. *Computers and Electronics in Agriculture*. 2020;171. Available from: https://doi.org/10.1016/j.compag.2020.105298.
14) Mishraa P, Asaari ANMSM, Lian G, Redfern S. Fusing spectral and textural information in near-infrared hyperspectral imaging to improve green tea classification modelling. *Journal of Food Engineering*. 2019;249:40–47. Available from: https://doi.org/10.1016/j.jfoodeng.2019.01.009.
15) Barbedo JGA. Plant disease identification from individual lesions and spots using deep learning. *Biosystems Engineering*. 2019;180:96–107. Available from: https://doi.org/10.1016/j.biosystemseng.2019.02.002.
16) Esgario JG, Krohling RA, Ventura JA. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*. 2020;169. Available from: https://doi.org/10.1016/j.compag.2019.105162.
17) Sowmya BJ, Shetty C, Seema S, Srinivasa KG. Chapter 7 - Utility system for premature plant disease detection using machine learning. In: Hybrid Computational Intelligence: Challenges and Applications. Elsevier. 2020;p. 149–172. Available from: https://doi.org/10.1016/B978-0-12-818699-2.00008-1.
18) Wadood SA, Boli G, Xiaowen Z, Hussain I, Yimin W. Recent development in the application of analytical techniques for the traceability and authenticity of food of plant origin. *Microchemical Journal*. 2020;152. Available from: https://doi.org/10.1016/j.microc.2019.104295.
19) Pawara P, Okafor E, Groefsma M, He S, Schomaker LRB, Wiering MA. One-vs-One classification for deep neural networks. *Pattern Recognition*. 2020;108:1–13. Available from: https://doi.org/10.1016/j.patcog.2020.107528.
20) Chen J, Liu Q, Gao L. Visual Tea Leaf Disease Recognition Using a Convolutional Neural Network Model. *Symmetry* . 2019;11(3):1–13. Available from: https://doi.org/10.3390/sym11030343.
21) Xin Z, Jun S, Xiaohong W, Bing L, Ning Y, Chunxia D. Research on moldy tea feature classification based on WKNN algorithm and NIR hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2019;206:378–378. Available from: https://doi.org/10.1016/j.saa.2018.07.049.
22) Tetila EC, Machado BB, Astolfi G, de Souza Belete NA, Amorim WP, Roel AR, et al. Detection and classification of soybean pests using deep learning with UAV images. *Computers and Electronics in Agriculture*. 2020;179. Available from: https://doi.org/10.1016/j.compag.2020.105836.