2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)

Tea Leaf Disease Classification and Tea Bud Identification

S. Jayanthy
Department of ECE
Sri Ramakrishna Engineering College
(Affiliated to Anna University)
jayanthy.s@srec.ac.in

V. M. Sathyendraa
Department of ECE
Sri Ramakrishna Engineering College
(Affiliated to Anna University)
sathyendraa.1802206@srec.ac.in

K. P. Sumedh
Department of ECE
Sri Ramakrishna Engineering College
(Affiliated to Anna University)
sumedh.1802229@srec.ac.in

S. Suresh
Department of ECE
Sri Ramakrishna Engineering College
(Affiliated to Anna University)
suresh.1802231@srec.ac.in

2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) | 978-1-6654-5635-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICERECT56837.2022.10059683

***Abstract*—Tea is one of the popular and widely cultivated plantation crops in Tamil Nadu. Their productions are affected by different diseases that affect them. The quality of the tea leaves and bud needs to be monitored to increase the profit. In this paper a model is implemented using deep learning and machinelearning techniques to detect the common diseases that affect thetea leaves and to classify the tea bud whether they are in the right stage to be plucked or not which signifies the high-quality and low-quality bud respectively. In this project two algorithms namely CNN and SVM are implemented using real time dataset. Since deep learning requires a large dataset for training. 1050 images for tea leaves disease detection and 262 images for tea bud classification are collected in real time and augmentation on these images is carried out to increase the number of input images. In CNN Dense-Net Architecture 121 and 201 are implemented for disease and bud classification which with deep convolution layers gives a high accuracy of 96.703% and 96.923% respectively. In SVM the same convolution model is built and converted into an SVM kernel which gives better results than regular implementation with an accuracy as 65.934% and 73.077% for disease and bud classification respectively. Evaluation metrics such as Accuracy, Loss, Precision, Recall, F1 Score, Mean Square Error, Mean Absolute Error, Specificity and Sensitivity of the model were analyzed**

Keywords— deep learning, machine learning, CNN, SVM, F1 Score, specificity, sensitivity.

I. INTRODUCTION (*HEADING 1*)

Tea is one of the important economic crops which has medical and health care functions like improving the human immunity power. It should always be in good quality for consumption. Tea crop productivity may be significantly reduced as a result of plant disease and these diseases reduce the quality of tea. Early disease detection can reduce the amount of loss. When it comes to tea plants, professionals can spot diseases by looking at the leaves. There by, offering professionals to treat disease Identification could be quite expensive. With the rapid development of technologies in agriculture, machine learning methods can be applied to identify whether the tea leaves have been diseased or not and also to find whether the tea bud has attained the perfect stage to be plucked or not. The yield and quality of tea leaves can be increased by monitoring the health of tea leaves[2]

The objective of this project is to identify whether the tea leaves are diseased or not and also to find whether the tea bud has attained the stage to be plucked or not. This can be used to find the high-quality tea leaves and it helps to deliver quality defined tea product to the market. The Convolutional Neural Network (CNN and Support Vector Machines (SVM) have been proposed for tea disease detection and tea bud categorization. Deep learning is currently the state-of-the-art in computer vision for object recognition. A total of 1050 photos of tea leaves are gathered, including a healthy class and seven frequent ailments in India. About 262 tea bud images were collected in real time of two classes namely in the right time to be plucked state and in cannot be used state. These images were augmented and given to the algorithm to make prediction.

II. RELATED WORK

India's Economy is one of the fastest growing economies in the world and Tea production has a great economic importance around the world. Their productions are heavily affected and destroyed by different diseases like Blister blight, red rust etc. and not plucking the tea leaves at right time of harvest. This leads to a huge loss to the tea producing companies and estate authorities. The tea leaf disease classification is common problem and many private and government organizations have provided feasible solutions to overcome the disease detection. Whereas the identifying the stage of the tea bud to classify them as high to low quality is a problem that is still in early research stage. There still isn't a viable way to measure the size and color of the bud in an effective manner to classify them.

This paper discuss the solution of identifying the common tea leaf diseases and classifying the tea bud at the right time for plucking using machine learning and deep learning techniques. This model has been incorporated both using CNN and SVMand advanced machine learning techniques to achieve that. For this study, real-time images of tea leaves andtea buds around 1050 and 262 images have been collected respectively from Tamil Nadu Tea Plantation Corporation Limited (TANTEA), Kothagiri.

III. LITERATURE SURVEY

A CNN model with 1 input layer, 4 convolutional layers, and 2 fully connected layers was suggested by R.S. Latha et al in 2021 [1]. The picture is delivered to the input layer. The output layer categorises the supplied picture into 8 classes: normal leaf, Algal leaf spot, Gray blight, White spot, Brown blight, Red scab, Bud blight, and Grey blight. The input picture in the dataset is largely what the convolution layers extract information from. While accounting for diverse tea leaf diseases, the proposed model successfully diagnoses the ailment with a 94.45% accuracy rate.

Jing Chen and Junying Jia (2020) [3] created a deep convolutional neural network called Leaf Net that is capable of identifying the seven types of diseases from photographs of tea leaf disease in order to provide timely and accurate diagnosis services in China's distant and mountainous tea estates. For comparison analysis, the traditional machine learning approach is utilised concurrently. It develops a bag of visual words model to express the image based on the DSIFT descriptor after extracting the image's dense scale-invariant feature transform. The multilayer perceptron and support vector machines, with accuracy rates of 60.91 and 70.94 percent, were used to identify illnesses of the tea leaf.

A machine learning method for identifying tea bud leaves was proposed by G.M.K.B. Karunasena and H.D.N.S. Priyankara (2020) [3]. For the automated tea leaf grading apparatuses, the tea bud's identification is essential. The identification of the tea bud leaf takes cognitive practise because there are currently no procedures to separate the tea bud leaf from the main tea leaf. Experimental information acquired by conducting experiments with MATLAB software is used to support the usefulness of the recommended approach. Results show that the suggested technique offers an overall accuracy of 55% for identification.

Iromi R Paranavithana and Viraj R Kalansuriya (2021) [5] proposed an approach on CNN to develop a model that identifies and predicts the appropriate tea buds for the plucking.The optimal trained model was estimated using test data. The experimental results show 70.15 accuracy for 10000 image samples, while Support Vector Machine(SVM) and Inception V3 gives accuracy of 65.86 and68.70 independently.

IV. WORKFLOW

A. Work flow for tea leaf disease classification

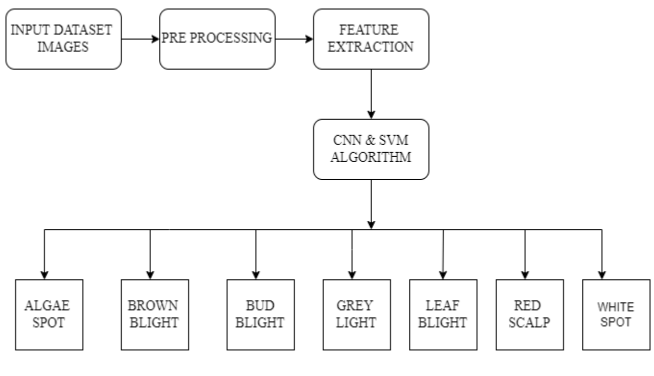

Fig.1. Work flow for tea leaf disease classification

The Figure 1 shows the block diagram which represents tea leaf disease classification. The real time dataset images were given to the CNN with dense net architecture and SVM algorithm. The input images were pre-processed and given as input to the respective algorithms. The feature of the images was extracted and predicted which class the disease belongs to with the help of SVM and CNN algorithm with dense net architecture.

The model was built in the way such that the base layer of the model consists of dense net layer. In tea leaf disease classification, densenet121 layer was used whereas in bud identification densemet201 layer was used. The input shape and the weights are initialized. The dense layer is adeeply connected neuralnetwork layer in which each neuron in the dense layer receives input from all neurons in the previous layer. DenseNet-121 has 1 7x7, Convolution 58 3x3 Convolution, 61 1x1 Convolution, 4 Avg-Pool and 1 Fully Connected Layer. Dropout layer is added immediately after the dense net layer which prevent the overfitting. The flatten layer is next to dropout layer helps to flatten the images.

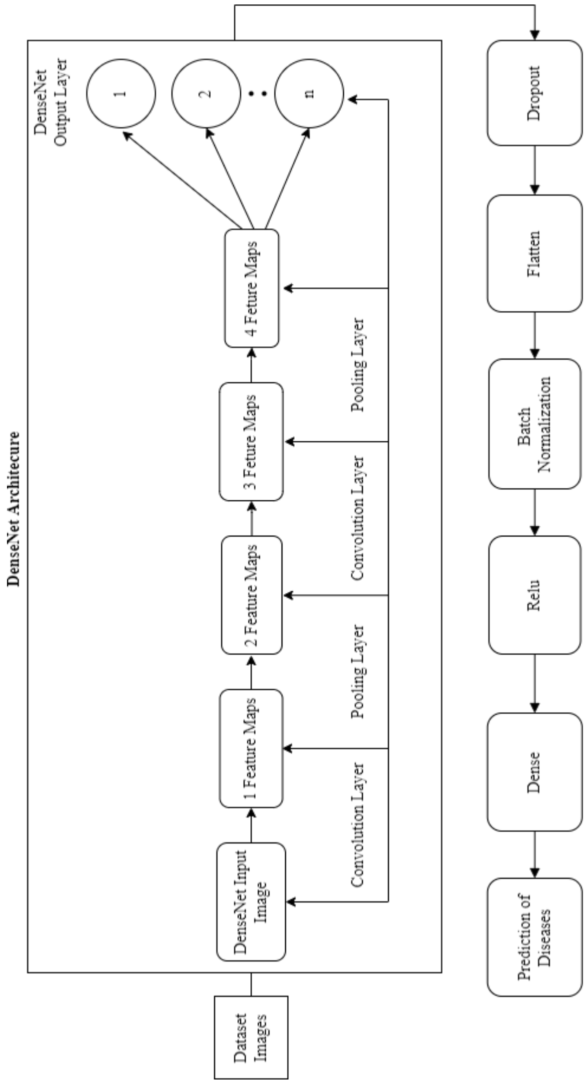

Fig.2. Work flow for tea leaf disease classification

After the flatten layer, batch normalization layer was added which dynamically normalize the inputs on a per mini-batch basis followed by a dense layer. Kernel Initializer “he uniform” is used to assign the weight statistically to start the neural network. The output from the dense layer was again batch normalized and given to the “Re-Lu”activation function, which helps to set all the negative values to 0. After this completion again added the dropout layer, dense layer, batch normalization, activation function layer followed by a final dropout layer.

After this final dropout layer, final output denselayer was added which has the output of 7 nodes and having the activation function of sigmoid which produces 1 if the prediction is close to input value else the value will be set as 0. Since the tea leaf disease classification is a multiclass classification, the loss function is set to “categorial cross entropy”. Fortea bud classification, the loss function is set to “Binary cross entropy” because it has only 2 classes to classify. For both the tea leaf disease classification and tea bud identification, Adam optimizer was used because regardless of the dataset number increases it uses less memory, whereas the learning rate was set to 0.001 and 0.00001 respectively.

B. Work flow for tea bud identification

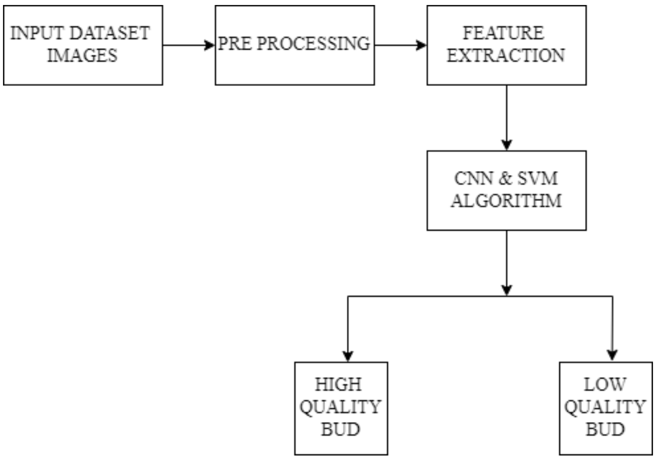

Fig.3. Work flow for tea bud classification

The Figure 3 shows the block diagram which represents the tea bud classification. The real time dataset images were given to the Convolution Neural Network with dense net architecture and SVM algorithm. The input images were pre-processed and given as input to the respective algorithms. The feature of the images was extracted and predicts whether the tea sprout was at the right time to harvest or not.

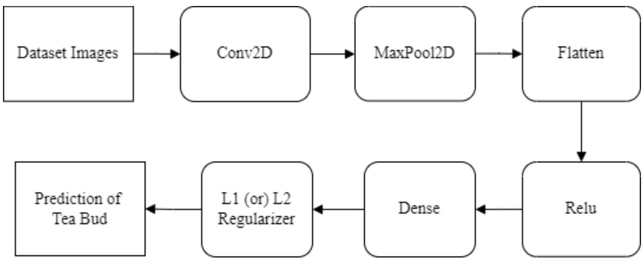

Fig.4. Work flow for tea bud classification

The Figure 4 represents the block diagram for tea bud classification. The real time dataset images were given to the CNN with dense net architecture and SVM algorithm. The input images were pre-processed and given as input to the respective algorithms.

In this algorithm, the input images were given to the conv2D layer which has a filter size of 32. The strides and the kernel size were set. The conv2D layer consist of Re-Lu activation function set the negative values to 0. Only the values greater than 0 are taken to the next layer. After the conv2D layer, Maxpool2D layer is added. Max Pooling layer downscale the Image by extracting most important feature and Remove Invariances. After the first Conv2D layer and first Maxpool2D layer, the second Conv2D and Maxpool2D layers are added. Then flatten layer is added which helps to flatten the images and then to dense layer. In this dense layer, the activation function is used to enhance the output prediction. For tea disease classification and tea bud identification l2 kernel regularizer with a learning rate of 0.01 and tea bud l1 kernel regularizer with a learning rate of 0.01 were used respectively. Adam optimizer with hinge loss function and square hinge loss function is used for tea bud identification and tea leaf disease classification respectively

V. Experimental Results

In this chapter the results obtained by implementing CNN and SVM Algorithms on the two datasets for tea leaf disease detection and tea bud classifications. For tea disease, images of 1050 in 7 classes namely Algal leaf spot, Gray blight, White spot, Brown blight, Red scab, Bud blight, and Grey blight are pre-processed and given as input and prediction of the images is made of these classes. Likewise, for bud classification the images of 262 in 2 classes of High Quality and Low Quality which are pre-processed and given as input and prediction of the images is made of these classes. The implementation, summary,results and outputs of these algorithms is inferred and displayed in this chapter. The algorithms are implemented in the Jupyter Notebook platform using Anaconda Navigator.

Fig.5. Sample images of tea leaf disease data set

The figure 5 shows the sample images of common diseases that affect tea leaves from the tea leaf disease dataset. The dataset collected at real time consists of 1050 images which were further increased by using various image processing techniques to improve the number of images in training and test datasets.

The figure 6 and 7 shows the sample images from the tea bud classification dataset consisting of 262 images which were collected in real time from Tamil Nadu Tea Plantation Corporation Limited (TANTEA), Kothagiri. Image processing, image augmentation and image segmentation were carried out to increase the number of input images and to provide quality images to the algorithm.

Fig.6.High quality tea sprouts.

The figure 6 show the good quality buds which represent the right stage to be plucked.

Fig.7. Low quality tea sprouts.

The figure 7 show the bad quality buds which cannot be used for tea making process.

The performance of the CNN and SVM model with different activation functions implemented on the tea leaf disease dataset and tea bud dataset was analyzed. In this analysis individual activation functions such as Re-Lu, Sigmoid, SoftMax and Linear were implemented along with these the combinations of these activation functions such as Re-Lu with Sigmoid, Re-Lu with SoftMax, Linear with Sigmoid, Linear with SoftMax and Re-Lu with Linear were also implemented.

By comparing the results of different combinations of activation functions enable the way to perform comparative analysis which in turn gives best possible outcome. These were performed for both datasets and both algorithms to maximize accuracy. The accuracy of the model with each activation on the datasets is shown below.

TABLE I
ACCURACY TABLE

ACTIVATION FUNCTIONS	CNN DISEASE	SVM DISEASE	CNN BUD	SVM BUD
RE-LU	19.780%	19.780%	72.302%	53.846%
SIGMOID	87.312%	17.582%	95.385%	53.846%
SOFT-MAX	69.231%	17.582%	46.154%	53.846%
LINEAR	7.692%	56.044%	**96.923%**	53.846%
RE-LU SIGMOID	**96.703%**	10.989%	84.615%	53.846%
RE-LU SOFTMAX	75.824%	**65.934%**	46.154%	46.154%
LINEAR SIGMOID	-	28.571%	-	46.154%
LINEAR SOFTMAX	-	17.582%	-	46.154%
RE-LU LINEAR	-	-	61.538%	**73.077%**

Table I represents the accuracy of CNN and SVM model for different activation functions. The combination of activation functions which were used, and the results were analyzed to determine the optimal solution.

From the Table I it was observed that CNN model with dense net architecture having "Relu-Sigmoid" activation function predicts with highest accuracy of 96.703% on comparing with other activation functions. Also, it was clear that Linear activation function is unfit for this tea leaf disease classification using CNN, whereas the similar dense net architecture having "Linear" activation function predicts with highest accuracy of 96.923% for tea bud identification. It is also observed the Re-Lu alone as the activation function produced poor results but when combined with sigmoid function at the output dense layer produced the best result. These accuracies of the CNN models were optimized after repetitive iterations and incremental training and testing of the algorithm on the datasets. The other parameters such as optimizers, learning rates, epochs etc., were also altered accordingly to produce these results.

From this comparison, it was observed that in the SVM model having Re-Lu activation function with Softmax activation function in the output dense layer predicts with the highest accuracy of 65.934% for disease classification and model having Re-Lu activation function with Linear activation function at the output dense layer predicts with highest accuracy of 73.077% for bud identification.

The other evaluation metrics other than accuracy such as Precision, Recall, F1 Score, Loss, Mean Square Error, Mean Absolute Error were analyzed, and the results are tabulated and given below in Table II.

TABLE II
EVALUATION METRICS

EVALUATION METRICS	RE-LU WITH SIGMOID FOR CNN DISEASE	RE-LU WITH SOFTMAX FOR SVM DISEASE	LINEAR FOR CNN BUD	RE-LU WITH LINEAR FOR SVM BUD
PRECISION	0.972	0.655	0.916	0.867
RECALL	0.967	0.626	0.912	0.866
F1 SCORE	0.967	0.579	0.912	0.866
LOSS	0.127	0.042	0.051	0.008
MEAN SQUARE ERROR	0.099	5.330	0.088	0.134
MEAN ABSOLUTE ERROR	0.055	1.220	0.088	0.134

Table II shows the comparisons of the number of evaluation metrics of the two algorithms on both the datasets with different activation functions allowed to provide the optimal combination of parameters to give the best result. These metrics are coherent and the activation which provided the highest accuracy is taken and the corresponding metrics are given here. The table II provides enough evaluation metrics which provide better understanding of the performance of themodels.

Along with these the metrics such as sensitivity and specificity are also analysed for the CNN and SVM algorithms on the tea bud identification dataset.

Sensitivity is a measure of how well a machine learning model can detect positive instances. Sensitivity is used to assess the performance of a model because it allows to seehow many positive instances the model has correctly identified.

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

The sensitivity of the CNN model was found to be 0.5732.

The sensitivity of the SVM model was found to be 0.5682

Specificity is a measure of how well the model can detect negative instances. The ability to correctly classify negative cases is called specificity which is a useful measure in skewed datasets.

$$Sensitivity = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

The specificity of the CNN model was found to be 0.2173.

The specificity of the SVM model was found to be 0.3714.

V. Conclusion and Future Scope

The goal of the proposed work is to classify and to improve the accuracy in predicting the tea disease and time of tea bud harvest. The model considers common tea leaf diseases, and it achieves the correct identification of them with accuracies of 96.703%, 65.934% using CNN and SVM Algorithm respectively. It also identifies whether the tea bud has acquired the precise time for harvesting or not with accuracies of 96.923%, 73.077% using CNN and SVM respectively. From the above results and the evaluation metrics it is analyzed that CNN model for both tea leaf disease classification and tea bud identification predict more accurately than the SVM algorithm.

In future the performance of the proposed idea can be further improved by varying the number of layers, the learning rate parameter, the optimizer, kernel initializer and regularizer used to get further better solution. This software model can be integrated with a drone like system as hardware and this system can automate monitoring and harvesting of the tea leaves. The leaves with diseases can be removed easily by the drones and the right leaves can be plucked and sent for tea making process. This can have a huge impact on not only in tea industries but in whole agricultural field since this removes manual effort which requires hard labour which in turn saves time and cost.

References

[1] R S Latha et el, “Automatic detection of tea leaf diseases using deep convolution neural network”, International conference on computer communication and informatics”, 2021.

[2] S.Jayanthy,E.Esakki Vigneswaran,A.Bharanidharan,” IOT Based Hexacopter For Mechanization And Automation In Tea Plantations” 201741000126 , 20.1.2017

[3] Jing Chen and Junying Jia, “Automatic recognition of tea diseases based on deep learning”, Licensee IntechOpen, 2020.

[4] G.M.K.B. Karunasena and H.D.N.S. Priyankara, “Tea bud leaf identification by using machine learning and image processing techniques”, International journal of Scientific and engineering research, 2020.

[5] Iromi R Paranavithana and Viraj R Kalansuriya, “Deep convolutional neural network for tea bud classification”, International Journal of computer science, 2021.

[6] Shaomin MU, Tingting SU, Xiaoxiao SUN, Yongyu XU, Zhihao CAO, “Image recognition of tea diseases based on convolutional neural network”, International conference in security, pattern analysis, cybernetics, 2018.

[7] A Yuan hong Zhong et al., “A convolutional neural network-based auto features extraction method for tea classification with electronic tongue”, MDPI, 2019.

[8] A.M.S.K.Abesinghe, K.S.P.Amaratunga, E.M.A.C.Ekanayake, M.A.R.Kavind and A. J.Mohotti, "Developing a selective tea plucking mechanism using image processing for a drone based tea harvesting machine", global scientific journals, 2020.

[9] P.Ouppaphan,“Corn disease identification from leaf images using convolutional neural networks”, 21st International Computer Science and Engineering Conference, 2017.

[10] B.C Karmokar, M.K. Siddiquee and M.S. Ullah, “Tea leaf diseases recognition using neural network ensemble”, International Journal of Computer Applications, 2015.

[11] Lin L, Mei H, Shi X, Xu Y and Zhou H, “The study and exploitation of diagnosed and controlled of tea disease’s expert system”, System Sciences and Comprehensive Studies in Agriculture, 2003.

[12] Mao H, Sun J, Wang P, Yang N, Yuan M and Zhang R, “Tea diseases detection based on fast infrared thermal image processing technology”, Journal of the Science of Food and Agriculture, 2019.

[13] Gong L, Huang Y, Liu .C and Zhao .Y, "A review of key techniques of vision-based control for harvesting robot", 2016.

[14] Abdur Razzak, Chakraborty. S, Hasan M. M, Hossain M. S, and Mou R. M, "M. Recognition and detection of tea leaf’s diseases using support vector machine", 2018.

[15] Dutta. M. K, Joshi. R. C. and Tiwari. V, "Dense convolutional neural networks based multiclass plant disease detection and classifcation using leaf images", 2021.