Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A COMPUTATIONAL MODEL FOR TEA LEAF PRICE PREDICTION BASED ON QUALITY FACTORS USING HYBRID MACHINE LEARNING TECHNIQUES

IRA GABA[1], B RAMAMURTHY[2]

[1]PhD Research Scholar, Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, India.
[2]Associate Professor, Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, India.
E-mail: [1]gabaira@res.christuniversity.in, [2]ramamurthy.b@christuniversity.in

ABSTRACT

This document reflects the effort made to calculate and identify the grade of the tea leaves based on the assessment of the leaves' size and color. The leaves were classified based on their severity with the help of HSV. The leaves were further classified using the k prototypes clustering once their length and width were established. The leaves were then further categorized in line with that. Light, medium, and dark are the three-color categories into which it belongs. The leaves were further sorted according to their quality so that the farmer could sell the produce at a better price. With the machine learning method for the categorization part, we were able to show its values. All of the healthy leaves were considered in a different dataset, and the images were obtained using the feature selection method. The length and width of each individual leaf, along with its color and shape, were then measured using those leaves. We were able to differentiate between the various leaf grades based on the findings. The healthy leaves were separated from the diseased leaves using the textual features. Additionally, we were able to use the other criteria to obtain higher-grade leaves.

Keywords: *Image Pre-Processing, Feature Selection, Classification, HSV, Color Parameters, K-Prototypes Clustering.*

1. INTRODUCTION

India's economic growth depends on tea leaves. Demand has increased as a result of the long-standing expansion of the tea-drinking culture worldwide. Therefore, it is essential to increase tea production while preserving the quality of the original tea leaf manufacture. This study's main objective is to categorize leaves according to their quality, identifying those that are the best, average, or lowest quality. The farmer should be able to determine the leaf's stage before picking it. He should be able to identify sick leaves early on and stop the entire output. The main objective of this activity is to assess the leaf's quality and assign grades based on that assessment.

Farmers used to accomplish this by hand, sorting the leaves according to color and examining their texture. However, this approach proved unfeasible as depending exclusively on human ability has drawbacks and could lead to the selection of subpar leaves. As a result, advanced artificial technologies may be used to highlight the need for a more accurate, scalable, and effective method of determining and evaluating quality. To differentiate between healthy and unhealthy leaves, we employed hybrid CNN and MLP in the previous study [1]. The healthy leaves in this study were subsequently separated into three categories: high quality, moderate quality, and poorest quality based on HSV and leaf size measures. I was able to ascertain the leaf's length, width, and general contour for the size dimension by employing canny edge detection.

This work's main objective is to show how minor changes can fortify the entire model and produce better results than current approaches. Additionally, the variety of feature selection information from photographs can improve overall performance. Convolutional neural networks (CNN) and multi-layer perceptron are two extremely helpful deep learning techniques for recognizing difficult patterns in images (MLP). A simple neural network uses many layers and neurons for different functions. These basic methods are used to identify and classify photographs as well as to classify other diseases. However, the classifier has an impact on the accuracy rate. In order to distinguish between the various leaf quality levels, we also computed the leaf's dimensions and applied HSV[11]. This will help classify the leaf into several groups.

The main contribution of this work includes:

- Creating a framework for identifying the quality of tea leaf.
- Performing the required pre-processing techniques
- The features are extracted by using feature extraction methods like GLCM, GLZSM, GLDM, GLRLM and NGTDM.
- Later the parameters are reduced by using feature selection method like information gain, feature importance, Anova, etc.
- The calculations are carried out with classifiers of the proposed model.
- The healthy leaves images are further examined with HSV and size(length and width) dimensions.
- Finally the leaves are categorised into three best, average and worst category with the help of K prototype clustering.

The next sections are managed as follows: The findings of literature review are shown in Section2. In Section 3, the methods implemented are explained in details. The Section 4, shows and discuss the results of the proposed work. Section 5 and Section 6 include the conclusion and future development of the work.

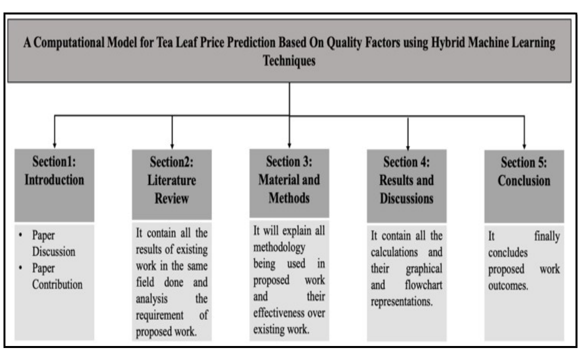


Figure 1: Work flow of the paper.

2. LIETURATURE REVIEW

The survey was done to find about the work being done till date for finding quality of the leaf, which indeed helps farmers to sell their production on a genuine cost, it will further help them in profit. Many research have been concentrated on machine learning and image processing to obtain quality of the leaves.

The researcher [1] was able to detect the diseased leaf, but they only retrieved the diseased photos manually. Essentially, they extracted the images by visual prediction, and then they classified them using machine learning techniques, with an accuracy of about 90. However, the feature extraction portion of this research was its worst flaw.

Using the LeNet, a well-known CNN model, the study in this article [2] was only able to predict the sick leaf. On average, the accuracy was over 85%. The sick leaf has just been taken into account.

The work to illustrate the stress of different leaves is highlighted in this study [3]. The PlantVillage provided them with the dataset. CNN is used exclusively for the classification. Image augmentation is used to expand the dataset's size. With many models, including AlexNet, GoogleNet, and Inception, accuracy was higher than 90%. For the classification, they have employed the pre-established model. This work suggests that if a leaf is under several stresses, further investigation is necessary. The behaviour of various leaf stresses can vary.

The four types of rice leaf diseases were identified by the researcher [4]. CNN and SVM have been used to illustrate the categorization. But they just used the color characteristics to carry out the detection. For the classification stage, they coupled CNN and SVM, and the accuracy exceeded 90%. To improve accuracy, the author claims that additional deep learning techniques must be used in this work.

Using CNN models alone, the authors of this work [5] were able to illustrate diseases such as tea leaf blight, red leaf spot, and tea red scab. Instead of extracting the features, segmentation was done. C-DCGAN is then used to illustrate the illness location based on the segmentation. This job was nearly 90% accurate. Only the diseased leaf was subjected to the implementation.

Images of the tomato crop leaves that were infected might be shown in this study [6]. CNN was also used to implement the classification. The accuracy of the suggested method was compared in this work with the work of other authors using VGGNet, Inception, and MobileNet. The CT scan images for detection may also benefit from this study. Nevertheless, only HSV and LBP characteristics are used to extract the features.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

There are numerous study gaps pertaining solely to the use of CNN for image classification, aside from looking at the literature review and the limitations mentioned in the papers that were provided. Based on the texture and color of the leaf, the researchers in the majority of the aforementioned studies were able to differentiate between healthy and diseased leaves. However, more criteria are required to assess the leaf's quality and grading. Therefore, additional characteristics that can yield detailed features to aid in classification need to be examined and analysed.Despite their effectiveness, the CNN and deep learning models utilized in this study still require labelled data for observation. On the other hand, the feature selection and classification procedure is more transparent and comprehensible when simply texture features are extracted. This will reduce the possibility that deep learning models would capture unnecessary and irrelevant features. Thus, additional traits must be categorized in order to determine the leaf's quality as well.

2.1 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 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 CNN+MLP 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.

2.2 Limitations of existing work

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

1. The existing work only uses CNN+MLP 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.

The existing work is implemented to differentiate between healthy and diseased leaves only.

3. MATERIAL AND METHODS

Finding the leaves' quality and classifying them into grades based on that quality is the primary goal of this study. Image processing, feature extraction, feature selection, classification, and leaf grading constitute the initial phase. The grading of leaves based on quality is determined using the healthy leaves that were acquired in the earlier procedure. The photos gathered from UPASI make up the dataset. The flowchart that explains the methodology employed in this investigation is shown in the following figure. The techniques for feature selection and extraction are used to extract useful elements from photos of leaves[2,3].

Using the textural characteristics of the leaf photos, these techniques are utilized to differentiate between healthy and unhealthy leaves. We work on the leaves' length, width, and shape after we have healthy leaves. K-prototype clustering was later employed for classification, and the leaves were categorized according on their quality.

3.1 Dataset

Dataset used for this study consist of tea leaves images obtained from UPASI Research Centre and real time images, which consist of healthy and unhealthy tea leaves. It is shown in figure 2. The images are collected from UPASI research Centre and from Kaggle. The images consists of benchmark images and real time images.

The figure 3 shows the flow chart of proposed research work. It basically consists of phase 1(shown in figure 3) and phase 2. Phase 1 will help us to extract healthy leaves and diseased leaves separately by implementing the textual feature extraction.

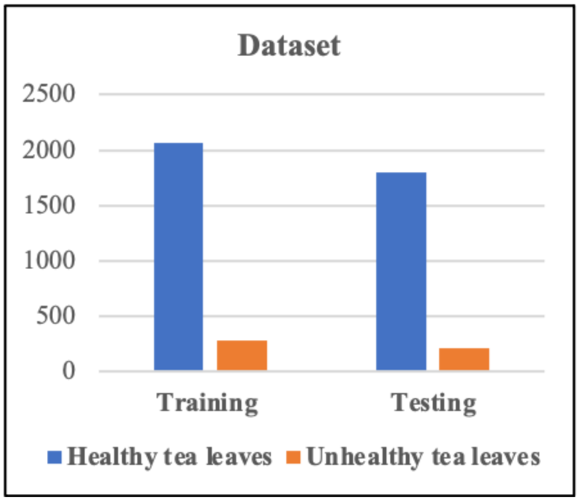


Figure 2: Dataset of leaves images.

And in phase 2 we are able to used only the healthy leaves extracted from phase 1 to calculate the quality of leaves and based on that we were able to categorized it into grade 1, grade 2 and grade 3. And in last we can obtain the pricing of it.

3.2 Image Processing

This process involves improvisation done to the leaves image to obtain a better accuracy. Pre-processing generally results in improving the quality of image by using specific techniques and removing the unwanted part or by eliminating the unwanted part by using filter.

3.3 Feature Extraction

It transforms the original features into a group of features with logical significance. In this study, textual features are extracted from the original image based on the radiomic parameters [12]. Then further feature selection method is implemented on them to get better number of parameters. As per the study there were 75 parameters extracted initially, which are moved forward for selection filter methods. The table 1 will show all the parameters selected.

The following are the methods used for extracting textual features.

3.3.1.GLCM: Gray Level Co-occurrence Matrix is a statistical technique for analyzing texture properties in image processing is the matrix. The spatial relationships between pairs of gray-level values in an image are described by the GLCM matrix. The number of times certain gray-level value combinations appear in neighboring pixels is counted to produce the matrix. Twenty-four statistical features, including contrast, homogeneity, and energy, were extracted from the images using GLCM.

3.3.2. GLSZM: Gray Level Size Zone Matrix is another statistical method used for texture analysis. The matrix is created by counting the number of homogeneous zones with a specific size and gray-level value. GLSZM was used to extract 16 features such as the zone size variability, the zone percentage, and the gray-level non-uniformity.

3.3.3. GLRLM: Gray Level Run Length Matrix is a texture analysis method that characterizes the length and frequency of runs of pixels with the same gray-level value. The matrix is created by counting the number of runs with a specific length and gray-level value in different directions. GLRLM was used to extract 16 features such as the short-run emphasis, the long-run emphasis, and the run-length non-uniformity.

3.3.4. NGTDM: Neighborhood Gray-Tone Difference Matrix is another statistical method used for texture analysis. The matrix is created by counting the number of times that a pixel has a specific difference in gray-level value with its surrounding neighbors. NGTDM was used to extract 5 features such as the coarseness, the contrast, and the complexity of an image.

5. GLDM: Gray Level Dependence Matrix is another statistical method used for texture analysis. The matrix is created by measuring the frequency of pixels with a specific gray-level value that have at least one neighboring pixel with a certain gray-level distance and value. GLDM can be used to extract 14 features such as coarseness, contrast, and busyness from an image.

The following table shows all the parameters extracted by feature extraction methods.

3.3 Feature Selection

When creating a predictive model, this method is used to lower the number of input variables. The relationship between the input and target variables determines which features are more effective. The feature selection method uses a variety of filtering techniques. We used Random Forest, Information Gain, Feature Importance, and Correlation with Target (ANOVA) for this investigation [4,5].

Out of the 75 features that were previously retrieved, we were able to identify the 29 most useful features by applying the feature selection techniques. The quality of the leaves was further assessed by using these characteristics to differentiate between healthy and unhealthy leaves. These methods are explained below:

Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

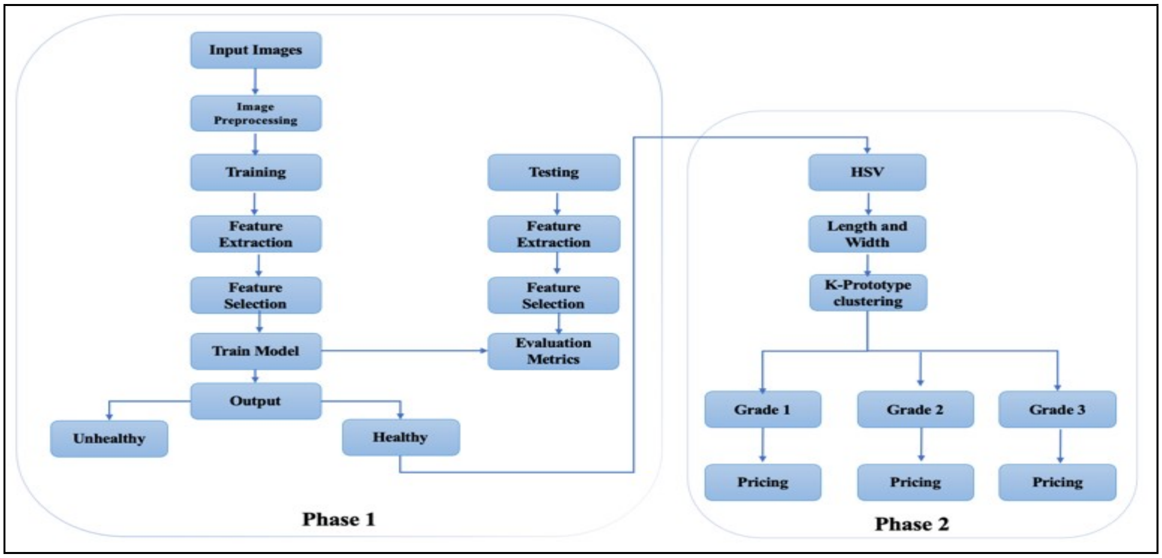


Figure 3: Workflow of proposed framework.

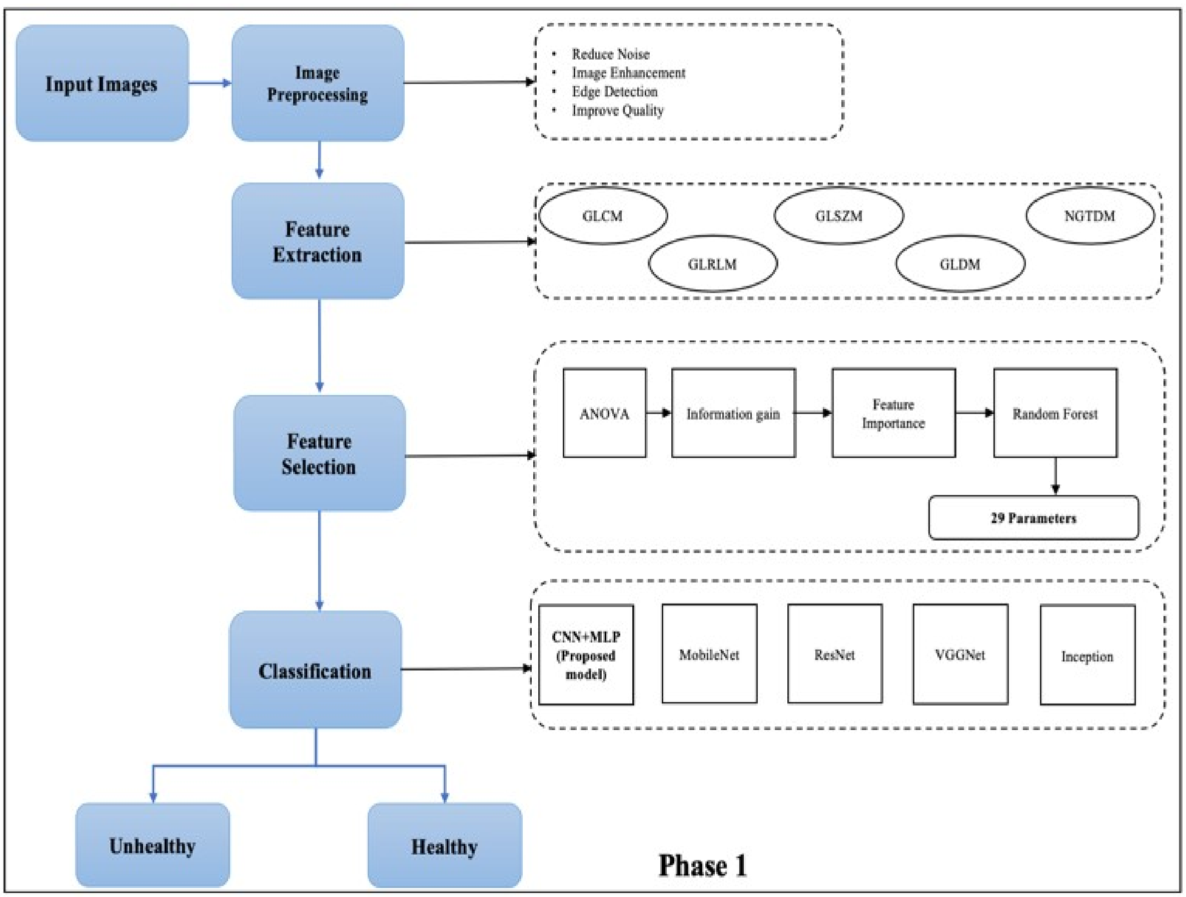


Figure 4: Workflow of Phase 1.

3.3.1. Correlation with target (ANOVA) - It examine the amount of variation within each sample, relative to the amount of variation between the samples

3.3.2. Information Gain - It is defined as the amount of information provided by the feature for identifying the target value and measures reduction in the entropy values.

3.3.3. Feature Importance - It calculates the score for all input features in a model to establish the importance of each feature in the decision-making process.

3.3.4. Random Forest - It calculate the importance of the features, which helps us neglecting the less useful. In this, features are firstly sorted accordingly to their importance score, and unimportant features are eliminated. For example, the important features are arranged by using F1 score in decreasing order, and the less important features are eliminated under the given threshold.[6]

After executing the feature selection we are able to reduce the number of parameter from 75 to 29 parameter for textual features of the leaves which will further used to calculate it accuracy. The following table shows the parameters which are selected after using feature selection methods.

The above figure is obtain by implementing the selection methods on all the parameters obtained from feature extraction phase. The values which were higher than the threshold value are selected by it. Therefore we were able to obtain 29 parameters in total.

Table 1: Parameters of Radiomic features

GLCM	**GLSZM**	**GLRLM**	**GLDM**	**NGTDM**
Autocorrelation	SAE	SRE	SDE	Coarseness
Joint Average	LAE	LRE	LDE	Contrast
Cluster Prominence	GLN	GLN	GLN	Busyness
Cluster Shade	GLNN	GLNN	DN	Complexity
Cluster Tendency	SZN	RLN	DNN	Strength
Contrast	SZNN	RLNN	GLV	-
Correlation	Zone Percentage	Run Percentage	Dependence Variance	-
Difference Average	GLV	GLV	Dependence Entropy	-
Difference Entropy	Zone Variance	Run Variance	LGLE	-
Difference Variance	Zone Entropy	Run Entropy	HGLE	-
Joint Energy	LGLZE	LGLRE	SDLGLE	-
Joint Entropy	HGLZE	HGLRE	SDHGLE	-
IMC 1	SALGLE	SRLGLE	LDLGLE	-
IMC 2	SAHGLE	SRHGLE	LDHGLE	-
IDM	LRLGLE	LRLGLE	-	-
MCC	LRHGLE	LRHGLE	-	-
IDMN	-	-	-	-
IDN	-	-	-	-
Sum Average	-	-	-	-
Inverse Difference	-	-	-	-
Inverse Variance	-	-	-	-
Maximum Probability	-	-	-	-
Sum Entropy	-	-	-	-
Sum of Squares	-	-	-	-

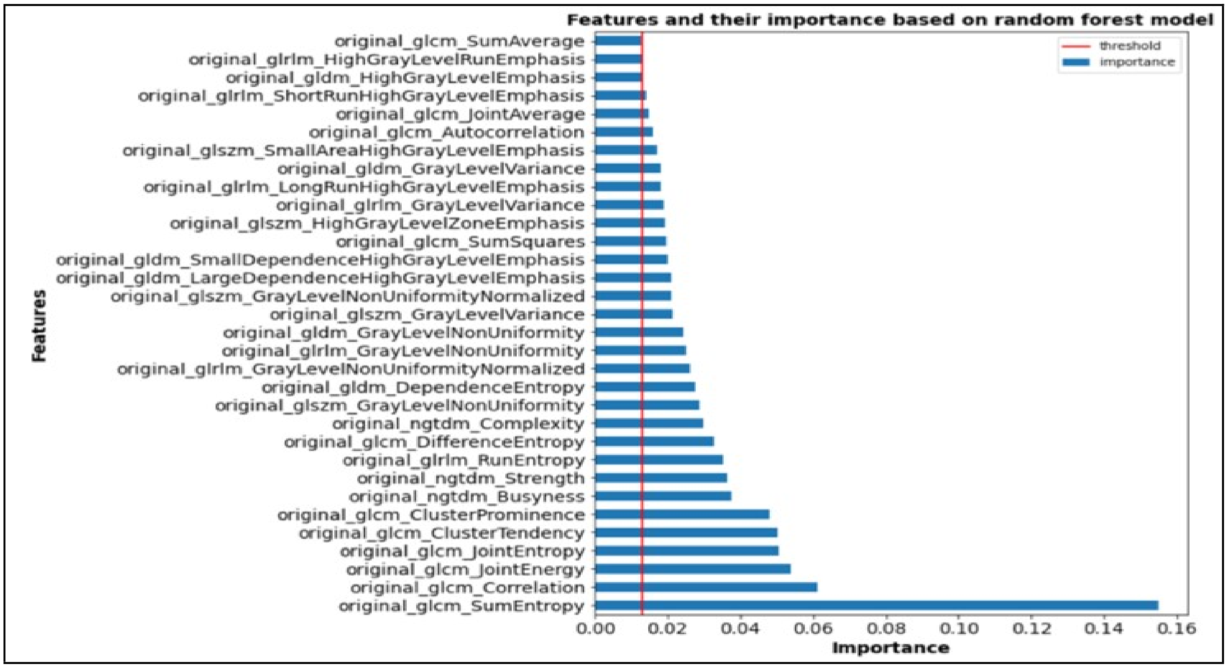


Figure 5: Parameters extracted after implying feature selection methods.

Table 2: Selected Parameters after implying Selection methods.

GLCM	GLSZM	GLRLM	NGTDM	GLDM
Sum Squares	SAHGLE	GLNN	Strength	LDHGLE
Joint Energy	HGLZE	RE	Busyness	DE
Cluster Tendency	GLN	GLN	Complexity	GLV
Joint Entropy	GLV	GLV	-	SDHGLE
Cluster Prominence	GLNN	SRHGLE	-	GLN
Correlation	-	LRHGLE	-	-
SE	-	-	-	-
Autocorrection	-	-	-	-
Difference Entropy	-	-	-	-
Joint Average	-	-	-	-

3.4 Classification: For this work, the CNN and MLP hybrid model will be used. This results in both healthy and unhealthy leaves, with the healthy leaves being used for quality control and further grading. The Rectified Linear Unit (ReLU) is the activation function for each neuron in the CNN's input and hidden layers.

The ReLU activation function's primary benefits are:

1. Deep learning and convolutional layers: It is the most widely used activation function for deep learning and convolutional layer training.
2. Computational simplicity: Only a max() function is needed to implement the rectifier function.
3. Representational sparsity: The rectifier function's ability to provide a real zero value is a significant advantage.
4. Linear behavior: When a neural network exhibits linear or nearly linear behavior, optimization becomes simpler.

The following figure shows the working of it.

3.4.1 Convolutional Layer: The fundamental layers of the CNN model extract features using a set of filters (kernels) of present sizes.

The results are added together to form a single output value, which is then placed in the proper

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

location inside the output matrix, also known as a feature map. Multiple feature maps are created by representing different elements of the input image using multiple filters[7].

3.4.2. Max Pooling: Convolutional neural networks (CNNs), which are neural network models used for computer vision or image classification applications, use the max pooling layer. This layer, as shown in figure 7, which comes after the convolution layer, is intended to preserve the most significant features while reducing the spatial dimensions of the feature maps that the convolution layer produces.

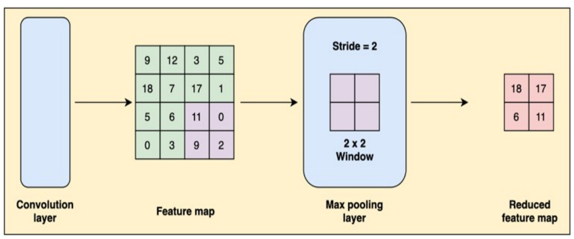


Figure 6: Max pooling for CNN.

3.4.3. Sigmoid: A smooth, continuously differentiable function that has traditionally been significant in the creation of neural networks is the sigmoid activation function, which is commonly denoted as σ(x). The mathematical version of the sigmoid activation function is as follows:

$$f(x) = 1 / (1 + e^{-x}).$$

The following figure shows the graphical representation of it:

3.4.4. Multilayer Perceptron (MLP): In the proposed work the numerical values obtained from the parameters calculated after feature selection are passed to MLP for better optimisation.

We created a hybrid CNN and MLP model for our research in the suggested work. This design was chosen because of its ability to predict textual properties. Our suggested model combines a multilayer perceptron (MLP) with a convolutional neural network (CNN). MLP was utilized to manage the numerical values that were produced following the feature selection process, while CNN was utilized to extract features from the input images.[8,9] The model's performance was enhanced by modifying several parameters, primarily the batch size, epochs, number of neurons, and hidden layers. Initially, hidden layers and the number of neurons were chosen at random; however, the grid search approach was eventually used to identify the ideal parameters.

Using the grid search approach, the following parameters were optimized: Batch Size =50, Epochs = 100, Learning Rate = 0.001. After being concatenated, both outputs are regarded as a single input. Two further dense layers with four neurons were added after the freshly obtained single input was counted as an initial input. The CNN and MLP models were concatenated using the Keras functional API, which offers the possibility of creating models with numerous inputs and outputs.

The following figure will show output from classification of hybrid model CNN and MLP is shown below. After that, suggested work is compared to the current model in order to verify its accuracy.

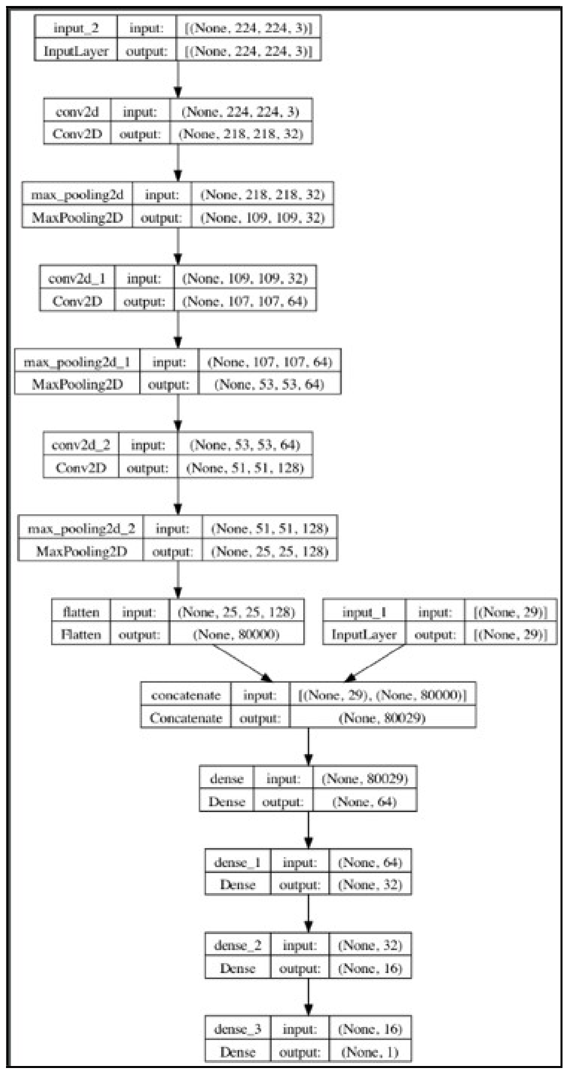


Figure 7: Graphical representation of CNN and MLP.

3.5 Quality Check based on Color, Length and Width parameters:

The primary goal of this study is to determine the quality of the leaves. Based on the information above, we were able to determine the textual features of the leaves and obtain healthy leaves. The problem statement seeks to ascertain the quality of the leaves by measuring their length, width, and size. In order to identify various quality grade levels, this phase attempts to investigate the implications of using both HSV and physical factors in the categorization process.[12]

The following are the steps we can follow to calculate the HSV.

1. Normalize RGB values
2. Calculate the Value(V): It represents color brightness and normalized R,G and B values.

$$V = \max(R,G,B)$$

3. Calculate the Saturation(S): it shows intensity of color. It is calculated as follows

$$S = \frac{V - \min(R,G,B)}{V}$$

If the maximum value (V) is 0, then the Saturation is also).

4. Calculate the Hue (H): it is calculated based on the differences between the normalized R,G, and B values. It also shows the color type.

- If V = R; $H = 60 \times \left(\frac{G-B}{V-\min(R,G,B)}\right)$
- If V = G; $H = 60 \times \left(2 + \frac{B-R}{V-\min(R,G,B)}\right)$

If V = B; $H = 60 \times \left(4 + \frac{R-G}{V-\min(R,G,B)}\right)$

Based on the Hue Saturation Value scale value, the color is separated into three folders. The color range was fixed, and the color saturation was determined using a histogram. If the color range is equal to or greater than 200, it falls into category 1, if it is between 170 and 200, it falls into category 2, and if it is equal to 170 or less than 170, it falls into category 3 shown in figure 14. The generated photographs are separated into three folders, as seen below, while the graph representation is displayed in the following part.

We can determine the length and width of the healthy tea leaves after computing the color parameter. With the aid of a clever edge detection filter, the leaf's boundary is determined. The leaf's length and width are then computed and saved in a CSV file.

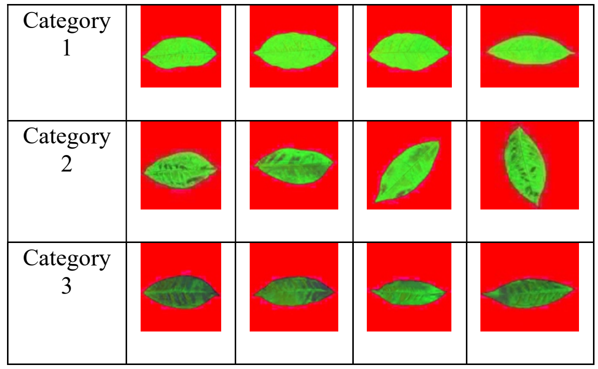


Figure 8: Representation of tea leaves by HSV.

Additionally, we compared the results using K-Means and K-Prototype clustering for the grading section. Additionally, the K-Prototype clustering produces superior results. The K-Means method repeatedly divides observations into a predetermined number of non-overlapping clusters. Each data point is first randomly assigned to a cluster using the K-Prototypes technique. It then determines the mode (for categorical features) and centroid (for numerical values) for every cluster. Then, using the combined dissimilarity metric, each data point is reassigned to the cluster whose centroid/mode is closest to it. Until the clusters stop changing or the maximum number of iterations is achieved, the centroid/mode computations and data point assignments are carried out again.[10]

One of the most used clustering algorithms is K-Means. It is quick, has a solid sklearn implementation, and is simple to comprehend intuitively.[13,14] Working with mixed data types is an advantage of K-Prototypes, a less well-known brother. In addition to measuring the distance between categorical features using the number of matching categories, it also uses Euclidean distance (similar to K-means) to assess the distance between numerical features.

We will use the Silhouette Score and Davis Bouldin Index to compare the K-means and K-prototype

- **Silhouette Score**

The calculation of Silhouette score can be done by using the following formula:

Silhouette_score=(p−q)/max(p,q)

Here, p = mean distance to the points in the nearest cluster And, q = mean intra-cluster distance to all the points.

Note: Higher the silhouette score better is cluster.

- **Davis-Bouldin Index**

Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

We can calculate DB index with the help of following formula

$$DB=\frac{1}{n}\sum_{i=1}^{n}\max_{j\neq i}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right)$$

Here, n = number of clusters
σ_i = average distance of all points in cluster i from the cluster centroid c_i
Note: Less the DB index, better the clustering model is.

Table 3: Calculation for comparison.

Methods	K-Means	K-Prototype
Silhouette Score	0.17	0.19
Davis-Boulding Index	1.9	1.8

Based on the above comparisons we were able to finalize K-Prototype clustering for our proposed work. The performance metrics were calculated further for better enhancement such as accuracy, precision, recall and F1 score.
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).

Where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.
The aforementioned values for the proposed model's accuracy, precision, recall, and f1-score [15,16] are displayed in the following figure.
The further results are shown in the below section. All the results and quality and grading are performed and shown below.

4.RESULT AND DISCUSSIONS

Results from the proposed model are conducted and reported in this section and a formal analysis is presented, which further highlights the comparisons of existing models and proposed work of CNN and MLP using the performance measures, followed by the quality check and the grading of the healthy leaf and finally price prediction.

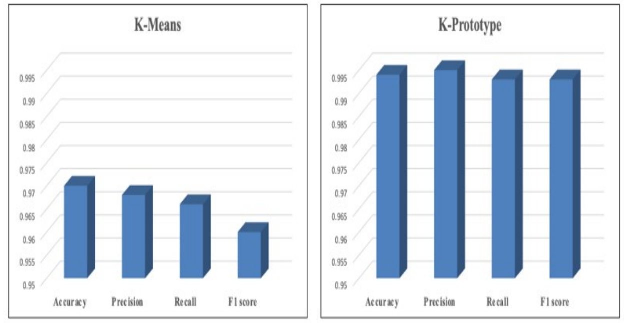


Figure 9: Performance matrices calculation for K-Means and K-prototype.

4.1. Accuracy and loss for CNN and MLP proposed model: The accuracy of the model is calculated to obtain the healthy leaf. The accuracy and loss validation are shown below. It is generally assumed that accuracy and loss are inversely related, meaning that lower loss values correspond to higher accuracy values. The main idea behind this work is that the proposed method of CNN and MLP provide the better accuracy as compare to the individual work carried out by the CNN and MLP.

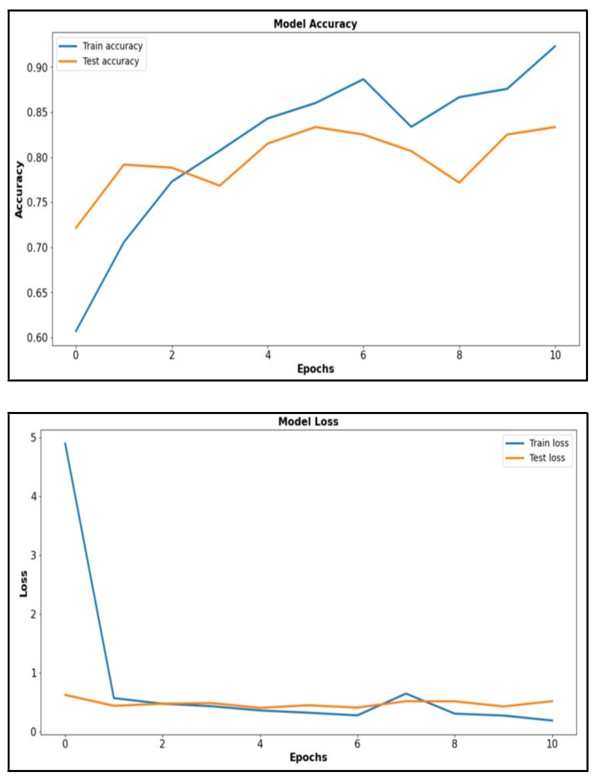
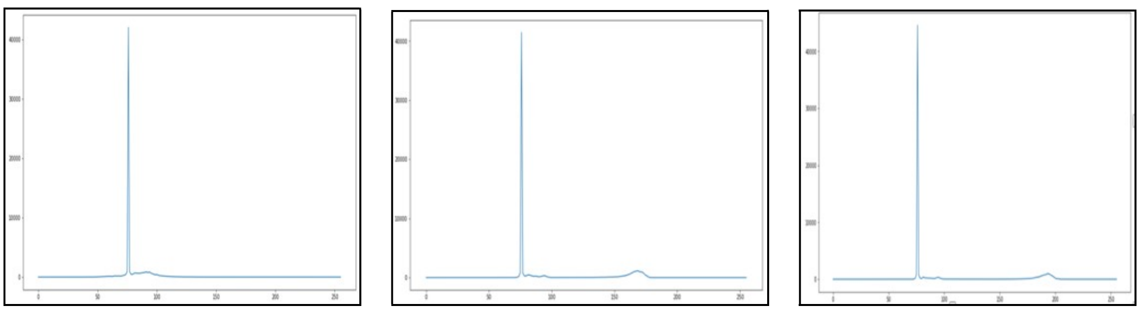


Figure 10: (a) Accuracy and (b)loss validation graph.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

S.No.	Model	Training Model				Testing Model			
		Accuracy	F1 Score	Precision	Recall	Accuracy	F1 Score	Precision	Recall
1.	**CNN+MLP (Proposed Model)**	**0.9**	**0.9**	**0.91**	**0.9**	**0.88**	**0.88**	**0.89**	**0.88**
2.	MobileNet	0.88	0.88	0.87	0.88	0.72	0.72	0.74	0.72
3.	VGG- 16	0.88	0.88	0.89	0.88	0.71	0.71	0.72	0.71
4.	Inception	0.89	0.88	0.89	0.88	0.71	0.71	0.72	0.71
5.	ResNet	0.56	0.56	0.55	0.56	0.48	0.48	0.50	0.49

Figure 11: Comparison between existing models and proposed model.



(a) (b) (c)

Figure 12: Graphical representation of HSV values for leaves belonging to different categories (a) Category 1 (b)Category2 (c)Category3.

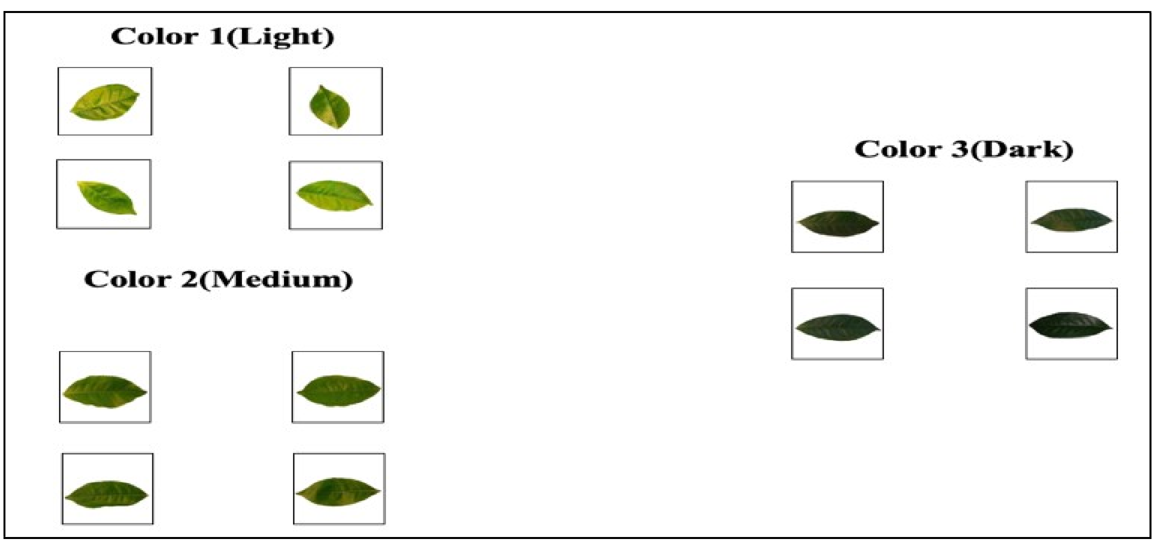


Figure 13: Category based on color.

Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10
© Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

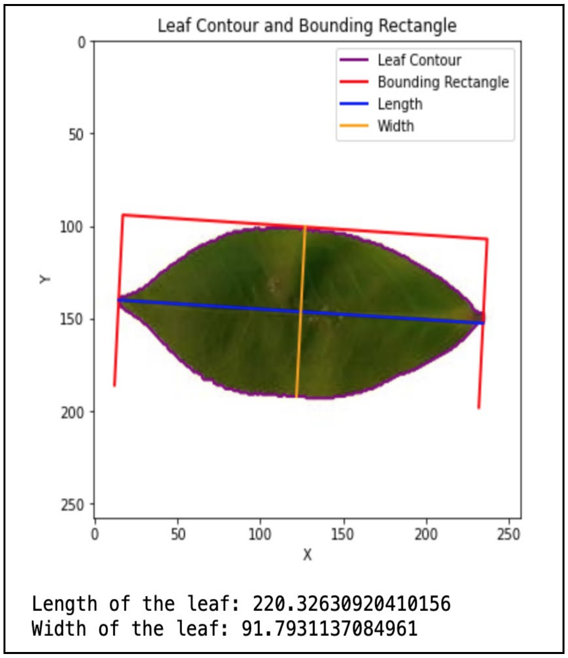


Figure 14: length and width calculation.

The graph is basically representing the accuracy of trained and test model and loss accuracy of the test and trained model for the CNN+MLP(Figure 10). The graph is drawn with the help of epochs and the accuracy and the epochs and the loss.

The above figure 11, shows the different histogram which shows the range of color according to which the categorization is carried out on the basis of color. Further the length and width are calculated with the help of canny edge detection and contour of the leaves image(Figure 14). The following figure shows the calculation of one leaf.[20] We were able to store the values of all the leaves in the CSV file.

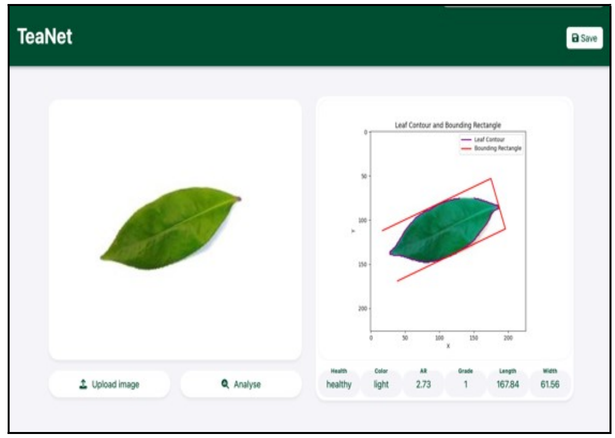
4.2 Quality Check

The healthy leaves obtain after CNN and MLP classification are further sent for quality check based on Color, length and width.

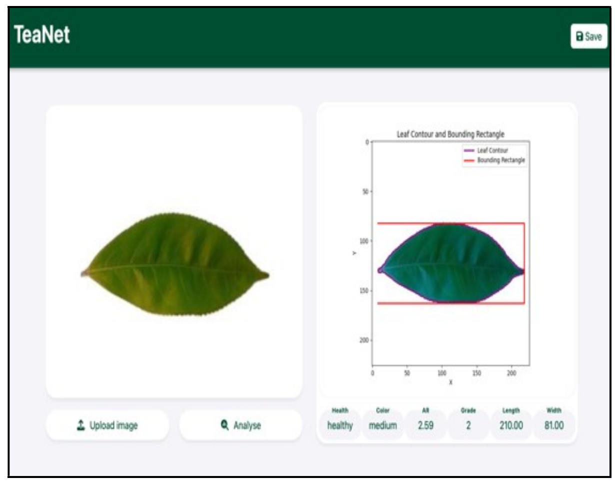
We will perform the HSV on leaves for color extraction and following are the images obtained based on following histogram values.[19]

4.3 Grading and Pricing.

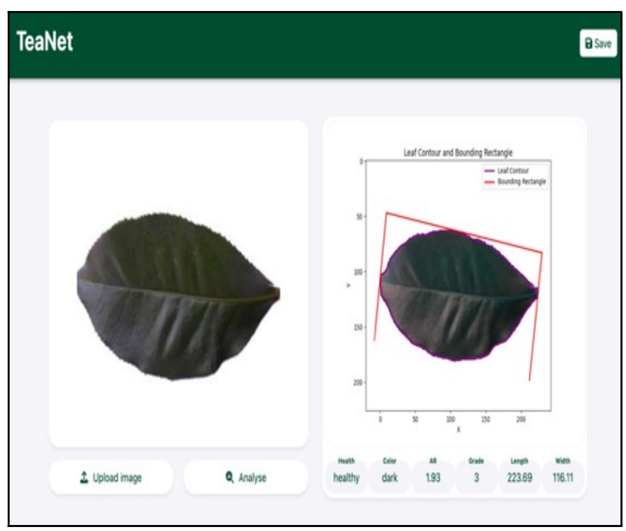
The last stage of the proposed model is to categorized the leaves into best quality, average quality and worst quality and lastly according to categorization we were able to price the leaf accordingly.



(a)



(b)



(c)

Figure 15: Represents the grade, length, width and color category of the leaves.

The following table shows the comparision of the existing models with the proposed model of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CNN+MLP to check the accuracy of thesystem in differentiating between healthy and diseased leaves.

Table 4: Accuracy compared with pretrained models in research work with proposed work.

Study	Year	Objective	Methods	Accuracy
Sarangi et al. [1]	2020	Diseased leaf	CNN	89%
Myburgh al. [4]	2020	Tea disease leaf	LeNet	85%
Shuting Yang et al. [5]	2020	disease leaf	CNN- Rf	90%
Sonali Aggarwal [10]	2020	Different type of leaves	AlexNet, GoogleNet, Inception	90%
S.Ramesh [11]	2020	Paddy Leaf disease	Deep Learning	87%
Yefang et al. [18]	2020	Tea disease leaf	Spectroscopy	85%
Haridasan et al. [19]	2023	Paddy leaves	CNN	88%
Li Y et al. [20]	2024	Tea Leaf	MobileNetV3	90%
Proposed work	**2024**	**Tea healthy and disease leaves**	**CNN+MLP classifier**	**91.3%**

5. CONCLUSION

This paper proposes a conventional hybrid CNN-MLP approach based on a range of feature extraction and feature selection methods. The CNN-MLP hybrid approach outperforms the individual methods with an accuracy rate of up to 91.3%. The hybrid approach is more effective and reliable. This study's new approach decreased the number of parameters from 75 to 29 by using feature extraction and feature selection. To increase accuracy, these parameters were subsequently sent through the MLP architecture. The proposed work which was also illustrated in the preceding section is the most successful. With this technique, we were able to differentiate between healthy and diseased leaves.

The quality of the leaves in the following phase can be assessed using these healthy leaves. The leaf is scored differently depending on whether it is the highest, average, or lowest quality. The hybrid CNN+MLP and feature selection were the main innovations in the proposed work, which actually increased the accuracy of the job. We were able to classify it into three groups based on the color, length, and width calculations, and we were then able to determine its price. The healthy leaves obtain from phase 1 will be used for quality check. Quality check is carried out by calculating color, length width parameters of leaves. The leaves are categorized and pricing is obtained.

REFERENCES:

[1] Sarangi, Susanta; Sahidullah, Md; Saha, Goutam, Optimization of data-driven filterbank for automatic speaker verification, (September 2020), DOI:10.1016/j.dsp.2020.102795. S2CID 220 665533.

[2] Alpaydin, Ethem, Introduction to Machine Learning. London: The MIT Press. p. 110. ISBN 978-0-262-01243-0. Retrieved (4 February 2017).

[3] Hinton, Geoffrey E.; Srivastava, Nitish; Krizhevsky, Alex; Sutskever, Ilya; Salakhutdinov, Ruslan R, Improving neural networks by preventing co-adaptation of feature detectors. (2012).

[4] Myburgh, Johannes C.; Mouton, Coenraad; Davel, Marelie H Gerber, Aurona, Tracking Translation Invariance in CNN's. Artificial Intelligence Research. Springer International Publishing. 1342: 282–295, (2020), DOI :10.1007/978-3-030-66151-9_18.

[5] Shuting Yang, Lingjia Gu, Xiaofeng Li , Tao Jiang and Ruizhi Ren, "Crop classification method based on optimal feature selection and hybrid CNN-Rf networks for multi temporal remote sensing imagery" , Mdpi, 2020, *12*, 3119; doi:10.3390/rs12193119.

[6] Dave Gershgorn, The inside story of how AI got good enough to dominate Silicon Valley.(2018).

[7] Chellapilla, K.; Fogel, D.B."Evolving an expert checkers playing program without using human expertise". IEEE Transactions on Evolutionary Computation. 5 (4): 422–428.(2001), DOI:10.1109/4235.942536.

[8] Huang, Jie; Zhou, Wengang; Zhang, Qilin; Li, Houqiang; Li, Weiping. Video-based Sign Language Recognition without Temporal Segmentation, (2018).

[9] Le, T.-L., Duong, N.-D., Nguyen, V.-T., Vu, H., Hoang, V.-N., & Nguyen, T. T.-N. (2015). Complex background leafbased plant identification method based on interactive segmentation and kernel descriptor. In Proceedings of the 2nd international workshop on environmental multimedia retrieval (pp. 3–8).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[10] Sonali Agrawal, A.S.Jalal , Mhd. Aamir Khan, plamt identification using leaf image analysis, International Conference on Internet of Things and connected technologies (2018).

[11] S.Ramesh, D.Vydeki, "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm", in Information processing in Agriculture(2020), pp.249-260.

[12] Jharna Majumdar, Shilpa Ankalaki, Gagan Karthik, Thribhuvan Gupta, Srinath Ramachandran,Adithya v Amberkar, Darshan S Salimath, S Roshan Sameer Shreyan P Shetty, Texture feature for analysis of covid-19 x-ray images, journal of critical reviews issn- 2394-5125 vol 7, issue 15, 2020.

[13] Md Manjurul Ahsan, Tasfiq E. Alam, Theodore Trafalis and Pedro Huebner, "Deep MLP-CNN Model Using Mixed-Data to Distinguish between COVID-19 and Non-COVID-19 Patients", MDPI, September 2020.

[14] Mina Ibrahim Muhammad Mahmoud Reda M. Albadawy Hatem Abdulkader," Liver Multi-class Tumour Segmentation and Detection Based on Hyperion Pre-trained Models", International journal of Intelligent Engineering and systems,2022.

[15] S. Wankhade and V. S., "A novel hybrid deep learning method for early detection of lung cancer using neural networks", Healthc. Anal., Vol. 3, p. 100195, 2023, doi:10.1016/J.HEALTH.2023.100195.

[16] M. M. Emam, E. H. Houssein, N. A. Samee, M. A. Alohali, and M. E. Hosney, "Breast cancer diagnosis using optimized deep convolutional neural network based on transfer learning technique and improved Coati optimization algorithm", *Expert Syst. Appl.*, Vol. 255, p. 124581, 2024, doi: 10.1016/j.eswa.2024.124581.

[17] Dohyun Kwon, Jaemyung Ahn, Chang-Soo Kim, Dong ohk Kang and Jun-Young Paeng, "A deep learning model based on concatenation approach to predict the time to extract a mandibular third molar tooth", Kwon et al. BMC Oral Health (2022) 22:571, doi: 10.1186/s12903-022-02614-3.

[18] Yifeng Huanga, Wentao Dongb, Alireza Sanaeifarc, Xiaoming Wangb, Wei Luob, Baishao Zhanb, Xuemei Liub, Ruili Lid, Hailiang Zhangb,, Xiaoli Li, "Development of simple identification models for four main catechins and caffeine in fresh green tea based on visible and near-infrared spectroscopy", Computer and electronics in agriculture vol. 173, June 2020.

[19] Haridasan, A., Thomas, J. & Raj, E. Deep learning system for paddy plant disease detection and classification. Environ. Monit. Assess. 195, 120 (2023).

[20] Xiao, Zhisheng; Yan, Qing; Amit, Yali, An Out of Distribution Dectection score for Variational Auto-encoder.Advances in Neural Information Processing Systems. (2020).

[21] Li Y, Lu Y, Liu H, Bai J, Yang C, Yuan H, Li X and Xiao Q Tea leaf disease and insect identification based on improved MobileNetV3. Front. Plant Sci. 15:1459292. doi: 10.3389/fpls.2024.1459292 (2024).