Identifying Issues and Proposing Solutions to Improve Sir Lankan Tea Cultivation

Heshani Weerasinghe Department of Information Technology Sri Lanka Institute of Information Technology Malabe, Sri Lanka Isuri Tissera Department of Information Technology Sri Lanka Institute of Information Technology Malabe, Sri Lanka Kithmi Welivitigoda Department of Information Technology Sri Lanka Institute of Information Technology Malabe, Sri Lanka

Amandi Maheshavi Department of Information Technology Sri Lanka Institute of Information Technology Malabe, Sri Lanka

ABSTRACT

Tea production is one of the most significant parts of Sri Lanka's economy. Sri Lankan tea holds a unique position in the global market. Traditional and non-standardized approaches affect tea cultivation manually monitoring tea leaf diseases is time-consuming. This study aims to develop a mobile application that uses image processing and machine learning to diagnose tea leaf diseases based on visual indications and provide treatment recommendations. With the advancements of technology, this farming process can be performed with the deep learning model. This research applies Convolutional Neural Networks (CNN) to Identify the tea leaves disease spread with an accuracy of 99%, Pest disease Identification 90% and tea leaves classification 96%respectively. This system predicts the yield of tea leaves cultivation with an accuracy of 100%. This developed system which aids tea cultivation and production industry is connected through the mobile application.

Keywords

Yield Prediction, Machine Learning, Convolutional Neural Network (CNN), Image Processing, Quality, Support vector machine (SVM), Image processing

1. INTRODUCTION

The tea industry is one of Sri Lanka's most significant economic sectors and Sri Lanka has a higher percentage of the global market share in the tea market. Every year more than 300,000 MT is produced by Sri Lanka and nearly 290,000 of MT of tea is exported to the global market. In 2019, 300,134 MT of was tea produced and 292,657 MT of tea was exported [1] Currently, more than 3 billion cups of tea are consumed every day worldwide. This popularity is attributed to its health benefits, which include the prevention of breast cancer, skin cancer, colon cancer, neurodegenerative complication, prostate cancer, and many others. Tea is also attributed to the prevention of diabetes and boosting metabolism [2]

Although Tea production has a huge impact on the Sri Lankan economy, tea cultivation and tea production are hindered in many ways. For example, Tea Cultivation is home to a wide variety of insect pests and also various types of leaf diseases. Since these pests and diseases are unable to be identified in the early stages and are not provided with solutions, they have an impact on tea cultivation's overall productivity. Currently, a manual inspection using their own eyes is the only way to diagnose these diseases, which necessitates a large number of people and laboratories with expensive equipment [3].

Another issue commonly seen in Tea Production is related to the quality of the tea leaves plucked which may hinder the quality of tea produced. Currently, about 20% of plucked tea leaves are not of high quality. Furthermore, the plucked tea leaves could be damaged and matured. Some tea leaves may have withered due to changes in climate during the transportation of leaves from the plantation to the factory. Therefore, identifying and classifying tea leaves that are high quality, matured, and withered is important to achieve high-quality tea and it will improve the market of Sri Lankan tea.

From the perspective of tea plantation owners, it would be beneficial if the yield of the production could be predicted. However, there are several environmental factors that affect the growth of tea crops in cultivation. With the change of such factors, the production yield of plants may change. The current method of predicting yield of tea crops is done manually, which necessitates a lot of manpower and paperwork.

According to the research done earlier though there were features to detect pest and diseases, there was no feature to locate the area where the diseases were spread. And also, for predicting the harvest, in earlier research, there was no feature to predict the harvest using the environmental factors which are temperature, humidity, soil-water availability and rainfall. In the existing research, classification was done to identify the tea bud. This research has increased the quality of tea bud leaves as well. as the classification type and enables the user to identify mature and wilted leaves.

The main objective of this project is to implement an Intelligence System and a mobile App capable of tea leaves infected with diseases by identifying symptoms, identifying and classifying the tea leaves based on the tea bud, shape, freshness, maturity level, and determining high-quality tea and based on the humidity level, predict the production yield of tea crops recommendation according to the environmental factors on the cultivation and detect pest damage on tea plants, identify the insects that are developing, and locate the pest spread areas.

2. LITERATURE REVIEW

The study by Karunasena et al. [4] Identifies the buds of plucked tea leaves using machine learning and image processing Techniques. By identifying these bud leaves, can identify high-quality tea leaves. This research focused on identifying whether plucked tea leaves have buds or not. In this study, the researcher used a cascade classifier, Histogram of Oriented Gradients (HOG) for feature extraction and as the final stage, using a support vector machine classifier for the tea bud leaves detection. For the training of the cascade classifier, the author used 150 positive samples and 300 negative samples. For the testing process used 40 tea leaves images for testing samples, which contain four length ranges of, tea buds, and selected 10 samples for each tea bud length range. Through this research, they achieved only 55% accuracy in identifying buds. In their model accuracy for average length buds10-30 mm) reached high accuracy and small (0-10 mm) and big (30-40 mm) buds were shown very little accuracy.

In the research by Betty S. et al. [5] the study's main goal was to see if regression models could predict tea yield responses to variations in maximum, minimum, and precipitation temperatures. The following specific objectives were employed to attain the main objective, using scatter diagrams, correlation analysis, and trend analysis, determine the statistical relationship between maximum, minimum, and temperature, and create multiple linear models to forecast tea yield based on climate data in the study area, use a contingency table to check the model's performance.

The research aimed to estimate tea yield using a regression model based on variations in maximum temperature, minimum temperature, and precipitation over the study area. The quality of climatic data was determined using the single mass curve technique. The variables under inquiry were statistically described, including mean and skewness. The statistical relationship between the variables under investigation was determined using correlation analysis. The data collected as a result of this regression was utilized to verify and analyze the model. The statistical behavior of the various variants of the regression model was used to determine the "best fit" model [5].

In the research by R. D. Baruah, S. Roy, R. M. Bhagat, and L.

N. Sethi et al. [6] To estimate tea yield analysis, data mining techniques were applied. Tea production proved to be highly influenced by climate factors. Total rainfall, sunshine length, and the difference between mean maximum and mean minimum temperatures all have a substantial impact on tea output in Assam's four tea-growing regions. These parameters have a statistically significant impact when compared to all other parameters that were initially evaluated in the study but were subsequently removed one by one based on their pvalues (value of Extra close brace or missing open brace means it has no significant impact on the dependent variable). The results of regression on the selected factors revealed a substantial link between observed and forecasted production values. Multiple regression results suggest that the constructed model may be used to accurately predict tea production in a certain region. Accurate climatic parameter estimates would lead to accurate production forecasts in the future. As a result, this model will be a valuable tool for the tea sector in making optimal management decisions ahead of time to maximize plantation profits.

In the research by K. A. Reddy, N. V. M. C. Reddy, and S. Sujatha.et al. [7] In pest identification and classification applications, image processing has a greater impact. These innovations will reduce the amount of time it takes to detect pests as well as the amount of manpower required. The purpose of this paper is to discuss various methods and algorithms for pest detection and classification using a computer vision approach. For the detection of pests, an image processing-based arrangement is proposed and evaluated. This paper presents a useful and accurate framework for detecting the affected image. K-Means clustering is used in the proposed method because it has a high level of precision when compared to other approaches and takes less time to process [7].

In the research by Shaik Sameer; B.D. Niharika; S. Vasavi; M. Rohith; VR Abhishek. et al. [8] To detect pests, an automatic pest detection system is used in this paper. To extract the pests from the captured image, many image processing algorithms are used to accustom the notice. A median filter was used to remove the noise and grain caused by a completely different lighting situation. The mechanism in use extracts the depicted pests in the image in a straightforward manner, scanning the image both vertically and horizontally to determine every coordinate and save the items in the image. The results in this paper are promising, but many improvements on each item and method are being monitored and administered to meet the requirements and needs of a fully automated gadfly detection, extraction, and identification system of pests. In the future, we'll focus on improving the accuracy of Image enhancement methods under a variety of lighting conditions [8].

In the article by Karan C. et al. [9] Crop prediction analysis considers several soil parameters such as nitrogen, fertility, pH, Phosphate, Potassium, and some atmospheric parameters such as sunshine, rainfall, and humidity to predict the best crop. The factors listed above have a direct impact on crop yield. As the agriculture system works with a huge amount of data derived from a variety of sources, complexity increases. A variety of techniques have been used to predict agricultural yields using various algorithms. The suggested approach

considers data on environmental factors, soil factors, weather, soil fertility, and previous year productivity to recommend the most lucrative crops that may be grown in the given environment. The ANN approach generates a list of all conceivable crops, allowing the farmer to select the most profitable one. In the foreseeable future, the integration of artificial intelligence and agriculture will benefit the majority of farmers. In comparison to other solutions, Neural Network is the best answer for agriculture problems (such as crop yield prediction).

3. METHODOLOGY

This research aims to address issues that tea cultivators face due to the manual farming practices followed by them. Thereby, a study was conducted aiming to develop smart tea cultivation solutions for tea cultivators in Sri Lanka to solve the issues.

A . Data Gathering for dataset creation

The Research Group initially carried out a survey among a group of tea planters from different areas of the country and identified essential data such as the ideal environmental conditions required for tea plantations in Sri Lanka. Currently, there is no standard database of images of unique tea diseases in Sri Lanka or images of tea plants of their life cycle.

Thereby the research team selected 3 farms and collected a total of 625 images of 8 diseases on tea plants and a total of 680 from different categories in tea leaves.

B. System Overview

The system consists of 04 functions: Classify Tea Bud Leaf Function, Tea Leaves Disease Detection and recommending solution function, Yield prediction function, and pest disease detection function. The farmer provides the required inputs to the respective functionalities to ease the farming methodology. The required outputs to the planters are provided to him through a mobile application. The overview of the STC Functionality is illustrated in figure 1.

Internet

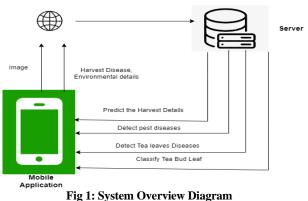


Fig 1: System Overview Diagram

This system is divided into four different segments in order to provide solutions to improve Sri Lankan tea cultivation. Each Segment describe below

3.1 Disease Detection and recommending solutions

A dataset containing 320 images in total was collected of 03 different types of diseases which are likely to grow in tea cultivation. The images collected to create the dataset were captured using a phone camera consisting a of resolution of 4160*3120 pixels. Table 1 shows 3 main identified tea leaves diseases found in Sri Lanka

Table 1: Diseases on Tea Leaves



Red Blister Blight Red Rust Tea White Disease Blister Blight

Convolutional Neural Network, a deep learning algorithm, and an image processing technique were used to diagnose Red Blister Blight, Red Rust Tea Disease, and White Blister Blight. A total of 320 images were used to identify the diseases of the tea leaves. The images which include their symptoms are taken in 3 tea farms. The dataset carries 100 images as 'Red Blister

Blight', 110 images as 'Red Rust Tea disease, and 110 images as 'White Blister Blight' The images which need to be checked for diseases will be captured through the mobile applications. The best advantage of using Mobile Net V2 architecture is that it performs faster than a consistent convolutional and is more suitable for mobile applications. The architecture began with 3x3 convolutional kernels and then progressed to 16 depth-wise separable convolutional blocks to offer a mobile model that is effective.

The dataset used an average pooling layer with size 4x4 and 1240 neurons to reduce dimensions and spatial variance. To achieve more successful accuracy, the dataset was trained to 100 epochs. Mainly the disease is segmented according to an object in the image, depending on color. The images come in 3 types. Color (RGB: Red, Green, Blue), grey, and binary. The input images are processed through several convolutional layers and eventually a fully connected layer that displays the classification results as illustrated shown in figure 2.

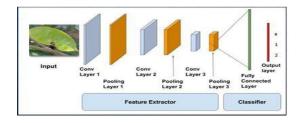


Fig 2: CNN Architecture

3.2 Tea leaves classification

A dataset containing 480 images in total was collected of 04 different types of tea leaves categories. Therefore, we captured the best quality tea leaves, God quality tea leaves, Matured tea leaves as well as Leaves that are withered in bad condition. To grow tea cultivation. The images collected to create the dataset were captured using a phone camera consisting of a resolution of 4160 * 3120 pixels.

Figure 3 shows the Standard quality levels of harvested tea leaves. [4]



Fig 3: Standard quality levels of harvested tea leaf

Images were captured from four different positions. A total of 680 images were used to identify the diseases of the mushroom. The dataset carries 250 images as 'Best', 230 images as 'Good', 100 images as 'Matured', and 100 images as 'Withered'. The image which needs to be checked for diseases will be captured through 25 mobile applications. The architecture began with 3×3 convolution kernels and then progressed to 16 depth-wise separable convolution blocks.

The input images are processed through several convolution layers, pooling layer. Figure 4 shows the summary of CNN architecture and its parameters.

	Output Shape	
conv2d (Conv2D)	(None, 198, 198, 16)	
max_pooling2d_1 (MaxPooling 2D)		
max_pooling2d_2 (MaxPooling 20)		
max_pooling2d_3 (MaxPooling 2D)		
<pre>max_pooling2d_4 (MaxPooling 2D)</pre>		

Fig 4: CNN Architecture

To achieve more successful accuracy, the Dataset was trained to 150 epochs. Furthermore, using a 'contour Area' can remove unwanted objects from the image.

3.3 Pest Diseases Identification and recommendation

A dataset of 05 different types of pest diseases that are likely to appear in tea cultivation was compiled, totaling 500 images. Using a phone camera with a resolution of 4160*3120 pixels, the images gathered to create the dataset were taken.

Low Country Live-Wood Termite, Faggot-Worm, Mealy Bug, Scarlet Mite, and Red Spider Mite were identified using Convolutional Neural Network, a deep learning algorithm, and an image processing technique.

The tea plant pest diseases were identified using a total of 500 images. The photographs of their symptoms were taken on four tea farms. Table 02 lists the five most common tea pest diseases that can be found in Sri Lanka.

Table 2: Pest Diseases on Tea Plants



The ability of Mobile Net V2 architecture to outperform a consistent convolutional network and be better suited for mobile applications is its greatest benefit. To provide an efficient mobile model, the architecture started with 3x3 convolutional kernels and then advanced to 16 depth-wise separable convolutional blocks. The input images are processed through several convolutional layers and eventually a fully connected layer that displays the classification results as illustrated shown in figure 5

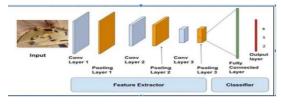


Fig 5: CNN Architecture

To reduce the dataset's dimensions and spatial variance, an average pooling layer with a 4x4 cell size and 1240 neurons were used. The dataset was trained to 100 epochs to improve accuracy. Generally speaking, the disease is divided into categories based on the color of an object in an image. Three types of images are available. Binary, gray, and RGB (Red, Green, and Blue) colors

3.4 Yield Prediction and recommending solutions

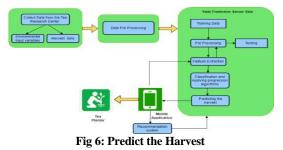
Tea plants grow well within specific environmental states such as temperature, humidity, rainfall, and soil-water availability. They grow and develop in a temperature that ranges from 26 to 32 $^{\circ}$ C, humidity within a range of 92 to 98 percent and rainfall level in a range of 2500-3000 mm, and soil water availability of 4.5-5.5 ph. as illustrated in Table 3

Table 3: Environmental factors Range

Environment Factor	Range
Temperature	26 - 32 °C
Humidity	92-95 %
Rainfall	2500 – 3000 mm
Soil-water Availability	4.5 – 5.5 ph

According to Figure 2, while creating a mobile-based solution for yield forecasting and making accurate tea crop suggestions using the collected data, the primary focus should be on how to collect information regarding environmental elements from the tea research center as data collection affects the outcome of the research.

The data required to create the dataset was collected by tea research centers and finally, a dataset of 4000 records was generated to train the machine learning model through a linear regression model. Figure 6 describes the main steps of the process.



4. RESULT AND DISCUSSION

The trained models were tested on test data to determine the classification accuracies of data collected from Tea research centers, tea factories, and tea farms using the phone camera. Table 4 displays the obtained accuracies.

Table 4: Accuracy table

Function Name	Accuracy
Disease Detection and recommending solutions	99%
Tea leaves classification	96%
Pest Diseases Identification and recommendation	90%
Yield Prediction	100%

The number of epochs committed has a direct impact on model accuracies. For Disease Detection, Figure 7 displays the accuracy achieved in training the images used in the model.

	6s 51ms/step	- loss: (0.0452 - acc:	0.9863 - val_loss:	0.0470 - val_acc: 0.9865	
Epoch 12/20	and a second second					
4657/4657 [=====] - 23 Epoch 13/20	6s Sims/step	- 1055; (0.0495 - acc:	0.9850 - val_loss:	0.0427 - val_acc: 0.9846	
4657/4657 [=====] - 23	5s 51ms/step	- loss: (0.0487 - acc:	0.9856 - val_loss:	0.0332 - val_acc: 0.9884	
Epoch 14/20		1				
4657/4657 [======] - 24 Epoch 15/20	0s 52ms/step	- 1055: (0.0345 - acc:	0.9880 - val_loss:	0.0537 - val_acc: 0.9884	
	9s 53ms/step	- loss: (0.0456 - acc;	0.9869 - val_loss:	0.0568 - val_acc: 0.9884	
Epoch 16/20						
4657/4657 [======] - 25	05 54ms/step	- 1055: 1	0.0218 - acc:	0.9921 - Val_1055:	0.0487 - val_acc: 0.9903	

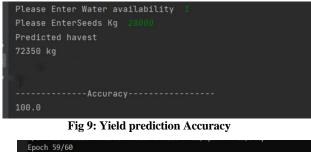
Fig 7: tea leaves Disease Detection Accuracy

The CNN model's training accuracy for Tea leaves classification was 96%, as shown in Table?? Figure 8 shows the accuracy obtained in training the images used in the model for the tea leaves classification

1/1 -	4s - loss:	0.1598	-	acc:	1.0000	-	4s/epoch	-	4s/step
Epoch	149/150								
1/1 -	3s - loss:	0.1769	-	acc:	0.9667	-	3s/epoch	-	3s/step
Epoch	150/150								
1/1 -	3s - loss:	0.0939	-	acc:	0.9667	-	3s/epoch	-	3s/step

Fig 8: Tea leaves classification Accuracy

For the Pest Diseases Identification, Figure 10 displays the accuracy acquired in training the previously collected data using the CNN model. Figure 9 displays the accuracy obtained in training the images used in the model for the Yield prediction.



1/1 - 1s - loss: 0.3260 - acc: 0.7500 - 781ms/epoch - 781ms/step

Epoch 60/60 1/1 - 2s - loss: 0.1864 - acc: 0.9000 - 2s/epoch - 2s/step

Fig 10: Pest Diseases Identification Accuracy

Each function of the system has the following accuracies: Yield prediction is 100% accurate, pest disease identification is 90% accurate, disease detection within tea leaves is 99% accurate, and tea leaf classification is 96% accurate.

Validation accuracy varies linearly with training accuracy, indicating highly accurate predictions of independent data. This shows that the model was successful in remembering the data.

5. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORK

The project's main limitation was the scarcity of tea leaf images in the data collection process. Since we would be able to collect tea leaves from Nuwara Eliya in high grown area, and Ruhuna, and Sabaragamuwa in the low country area, but the medium region tea leaves which are grown in Kandy could not be included in the model before the project was completed.

This study can be increasing the accuracy with more datasets from every region of Sri Lankan tea. Also, further, development could be done with a robotic arm to pluck the tea leaves by classifying them.

6. CONCLUSION

Farmers need timely information to make more informed decisions in their day-to-day farming operations. Lack of information visibility at all three key stages of crop selection, cultivation, and disease protection lead farmers to make suboptimal decisions leading to financial difficulties. Thus, it is extremely important to present the right information to the communities in need to help them make informed decisions precisely at the right time. A review of relevant research highlighted systems designed to deliver prevailing market prices have not been enthusiastically embraced by the farming community. An in-depth analysis has shown that after the farmers get a tea crop, there is nothing they can do about the market price except sell the crop at the prevailing price.

This emphasizes the importance of providing accurate information right from the crop selection stage. Currently, agricultural data is scattered in different places, which makes it difficult to find the right information at the right time.

Thus, this study identifies the need for a farmer-centric information flow model that engages all stakeholders to support the decision-making process. Information needs to depend on the current stage the farmer is in the farming life cycle as shown in the study; Crop selection stage, growing stage, or disease awareness stage. The proposed information flow mode classified the sources of information from which the farmer can get accurate information for the identified key stages of the farmer's life cycle.

Her findings indicate that a mobile phone-based information system would be an effective method of intervention as most farmers have access to a mobile phone. Then planning the next few phases of the action research life cycle. Based on the above findings, the next step is to design interventions to address the problems faced by tea farmers. Future research will be conducted to identify how information from identified sources can be gathered using emerging technologies. Next, the intervention is evaluated using a set of farmers. Reflecting on the findings will improve the mode to ensure sustainability in this domain. In addition, how to create a meta-level model that can be generalized to other domains to solve such problems is another challenge that will be addressed soon. The future. This work contributes to ensuring the sustainability of farmers by providing timely information on their livelihood activities.

7. REFERENCES

- S. L. T. E. Association, "Statistics," Tea Export Association, Sri Lanka, [Online]. Available: https://teasrilanka.org/statistics.
- [2] S. L. T. R. Institute, "Health benefits of Tea," Tea Research Institute, Sri Lanka, [Online]. Available: https://teasrilanka.org/health-benefits.
- [3] S. L. Tea Research Institute, "Priority Research Areas," Tea Research Institute, Sri Lanka, [Online]. Available: https://www.tri.lk/major-research-areas/.
- [4] H, Karunasena G & Priyankara, "Tea Bud Leaf Identification by Using Machine Learning and Image Processing Techniques," Journal of Scientific and Engineering Research, August 2020.
- [5] Betty Sitienei, Juma Shem G.Everline Opere, "On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County, Kenya," Climate, July 2017.
- [6] R. D. Baruah, S. Roy, R. M. Bhagat, and L. N. Sethi, "Use of Data Mining Technique for Prediction of Tea Yield in the

Face of Climate Change of Assam, India," 2016 International Conference on Information Technology (ICIT), pp. 265-269, 2016.

- [7] K. A. Reddy, N. V. M. C. Reddy, and S. Sujatha., "Precision Method for Pest Detection in Plants using the Clustering Algorithm in Image Processing," 2020 International Conference on Communication and Signal Processing (ICCSP),pp. 894-897, 2020.
- [8] S. Sameer, B. D. Niharika, S. Vasavi, M. Rohith, and V. Abhishek, "Pest and Disease Detection from Plant Leaves using Enhanced AlexNet Model," 2021 IEEE International Conference on Electronics, Computing, and Communication Technologies (CONNECT), vol. 10.1109, pp. 01-05.2021.
- [9] Karan Chaudhary, Prof. Farhana Kausar, "PREDICTION OF CROP YIELD USING MACHINE," International Journal of Engineering Applied Sciences and Technology, 2020, vol. 4, pp. 153-156, 2020.
- [10] S. L. T. R. Institute, "Annual Report 2017," 2017., [Online]. Available: https://www.parliament.lk/uploads/documents/paperspresented/annual-report-tea-research-institute- 2017. pdf.
- [11] Karunamoorthy B. & Somasundereswari D, "A Defect Tea Leaf Identification Using Image Processing," Przegląd Elektrotechniczny, vol. 89,pp. 318-320, 2013.
- [12] S. L. T. e. associaion, ""The importance of origin,"

[Online]. Available: https://teasrilanka.org/growing-regions. [Accessed 30 10 2022].

- [13] Qian Y., Li G., Lin X., Zhang J., Yan J., Xie B. & Qin J.Fresh, "Tea Leaves Classification Using Inception-V3," 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), no. 2019, pp. p. 415-419.
- [14] M. F. & Y. S. Yanyan Z, "Study on sorting technology of fresh tea leaves based on the convolutional neural network model," " 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), pp. 614-617, 2019.
- [15] M. S. Hossain, R. M. Mou, M. M. Hasan and S. &.R. M. A. Chakraborty, "Recognition and detection of tea leaf's diseases using support vector machine," 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA), pp. 150-154, 2018.
- [16] Wenxia Bao, Tao Fan, Gensheng Hu*, Dong Liang & Haidong Li, "Detection and identification of tea leaf diseases based on AX-RetinaNet," Scientific Reports, 2022.
- [17] Mohammad Samawat Ullah,Bikash Chandra Karmokar,Kazi Md. Rokibul Alam, "Tea Leaf Diseases Recognition using Neural Network Ensemble," International Journal of Computer Applications, March 2015.
- [18] H. Nagar and R. S. Sharma, "Pest Detection on Leaf using Image Processing," 2021 International Conference on Computer Communication and Informatics (ICCCI), vol. 10.1109, no.ICCCI50826.2021.9402606, pp. 1-5, 2021.