Proceedings of the Second International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2023)
IEEE Xplore Part Number: CFP23BC3-ART; ISBN: 978-1-6654-5630-2

Proposed CNN Model for Tea Leaf Disease Classification

[1]Rahul Singh
[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[1]rahul.2414@chitkara.edu.in

[2]Neha Sharma
[2]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[2]sharma.neha@chitkara.edu.in

[3]Rupesh Gupta
[3]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[3]rupesh.gupta@chitkara.edu.in

2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 978-1-6654-5630-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAAIC56838.2023.10140680

Abstract— **Tea leaf diseases have had a substantial impact both on the quantity and quality of the tea produced. The high-precision automatic detection and identification of illnesses that can be found in tea leaves is beneficial to the accurate prevention and control of those diseases. Manual procedures, which require a lot of time and effort, are still the primary tool for diagnosing tea illnesses and determining the severity of their effects and also effect the agriculture. This situation has persisted for some time. It is helpful to the tea leaf disease prevention and control efforts to have accurate and speedy disease detection. This research presents a technique for tea leaf disease classification that is based on an improved version of a deep convolutional neural network. This project aims to develop a deep convolutional neural network model capable of identifying diseases affecting tea plants based on image sets of their leaves. The results of the studies indicate that the proposed method has an average identification accuracy of 73%, which is higher than the accuracy of more conventional manual approaches. In later applications, the CNN model was utilized to improve the diagnostic measurement of tea leaves as well as the measurement of leaves from other plants.**

Keywords— Tea Leaf Disease, Deep learning, Image Classification, Convolution Neural Network, Early Detection.

I. INTRODUCTION

Tea is one of the world's most popular beverages, and tea leaves are one of its most important components. These are the origins of tea's distinct aroma, flavor, and color, all of which contribute to tea's global popularity among billions of people [1]. The plant used to make tea, Camellia sinensis, is native to China and other Southeast Asian countries [2]. The leaves of the tea plant are harvested. Tea leaves have been used for centuries for a variety of medicinal purposes in addition to brewing tea [1]. They have antioxidant and anti-inflammatory compounds such as catechins, flavonoids, and polyphenols. One of these characteristics is the ability to reduce inflammation. Several compounds have been studied for potential health benefits such as lowering the risk of cardiovascular disease, lowering cholesterol levels, and improving brain function. Furthermore, tea leaves contain caffeine, a natural stimulant that can help increase alertness and concentration. Caffeine can be found in tea leaves [3]. Theanine is an amino acid that calms the brain and can help relieve anxiety and tension. Tea leaves contain theanine, so drinking tea is a great way to get your daily dose. Tea leaves are an essential component of the tea-making process and may provide a variety of health benefits. They have been used for a long time throughout history, and millions of people around the world continue to benefit from them. Tea leaf disease is a fungal infection that damages the leaves of the Camellia sinensis plant, also known as the tea plant [4]. Tea blight is also known as tea leaf disease. The fungus Exobasidium vexans causes the disease, which can infect tea plants at any stage of development. The fungus first appears on the underside of the leaves as white or light pink dots. These patches will eventually turn dark, causing the leaf to fall off [3,4]. The fungus produces spores, which are then dispersed by the wind or rain, resulting in the disease's eventual development. The disease's spread is difficult to control because the spores can survive in the soil for months without being killed. Because the disease is more common in areas with high humidity and precipitation, tea-growing regions in tropical and subtropical countries are more vulnerable to it [1-3]. Tea farmers may suffer significant financial losses as a result of tea leaf diseases. When the disease strikes, it has the potential to result in significant yield losses, resulting in lost income and higher production costs. When the disease is severe enough, it can destroy entire tea plantations, forcing growers to abandon their crops. The global impact of tea leaf disease is difficult to quantify due to the fact that different outbreaks of the illness vary greatly in terms of severity and geographic spread [5]. However, in areas where it is common, the illness has had a significant impact on the tea industry. For example, the first recorded epidemic of tea leaf disease in Sri Lanka (formerly Ceylon) in the late 1800s caused significant damage to the island's tea plantations, forcing producers to seek out new, disease-resistant tea varieties. Despite the fact that tea leaf disease poses a number of challenges to tea cultivation, tea farmers have devised a number of methods for controlling and managing it. Among these preventative measures are the use of fungicides, the breeding of disease-resistant tea plants, and the implementation of cultural practices that reduce the likelihood of infection [6]. As a result, tea leaf disease is no longer as common as it once was; however, tea remains one of the most valuable commodities on the global market. Tea leaf disease is a fungal infection that damages the leaves of the Camellia sinensis plant, also known as the tea plant. Tea blight is also known as tea leaf disease. The fungus Exobasidium vexans causes the disease, which can infect tea plants at any stage of development. The fungus first appears on the underside of the leaves as white or light pink dots [5]. These patches will eventually turn dark, causing the leaf to fall off. The fungus produces spores, which are then dispersed by the wind or rain, resulting in the disease's eventual development. The disease's spread is difficult to control because the spores can survive in the soil for months without being killed. Because the disease is more common in areas with high humidity and precipitation, tea-growing regions in tropical and subtropical countries are more vulnerable to it [4-6]. Tea farmers may suffer significant financial losses as a result of tea leaf diseases. When the disease strikes, it has the potential to result in significant yield losses, resulting in lost income and higher production costs. When the disease is severe enough, it can destroy entire tea plantations, forcing growers to abandon their crops. The global impact of tea leaf disease is difficult to quantify due to the fact that each outbreak varies greatly in terms of severity and geographical distribution [7]. In contrast, where the disease is prevalent, it has had a significant negative impact on the tea industry. The first reported outbreak of tea leaf disease occurred in Sri Lanka (formerly known as Ceylon) in the late nineteenth century. This epidemic severely harmed the island's tea plants, forcing producers to seek disease-resistant tea cultivars. Despite the fact that tea leaf disease poses a number of challenges to the tea growing industry, tea farmers have developed a number of prevention and management strategies. Among these preventative measures are the use of fungicides, the breeding of disease-resistant tea plants, and the implementation of cultural practices that reduce the likelihood of infection. As a result, tea leaf disease is no longer as common as it once was; however, tea remains one of the most valuable commodities on the global market. Early detection of Tea Leaf Disease (TLD) is critical for several reasons. Second, it can prevent large output losses in tea harvests by allowing farmers to act quickly and limit disease spread. This not only saves farmers money, but it also reduces the environmental impact of chemical treatments required to cure a disease in its advanced stages. Second, early TLD detection can help farmers identify and address the disease's underlying causes, leading to more sustainable and successful crop management strategies. This can lead to better soil health, water management, and insect control, resulting in environmental and economic benefits in the long run [8]. Finally, preserving the quality of tea leaves is critical for the reputation of the tea brand, and early identification of TLD can aid in preserving this quality [9]. As a result, early detection of TLD is a critical component of preventative crop management, which can lead to healthier and more sustainable tea production. Using a CNN (Convolutional Neural Network) model to detect Tea Leaf Disease (TLD) early is an efficient way for the agriculture industry to harness technology [10]. The CNN model is trained on a set of healthy and sick tea leaf photos, allowing it to recognize the visual patterns associated with TLD. The program can quickly detect the illness and notify farmers to take preventative measures by evaluating new photos. Farmers can save time and money by administering the necessary treatments or modifying crop management procedures. Furthermore, as the illness progresses, the CNN model can be retrained on new data, making it a versatile and adaptive detection tool for TLDs. Overall, the use of CNN models for early TLD detection can help tea farmers optimize crop management and increase yields, resulting in a more sustainable and profitable tea sector.

II. LITERATURE REVIEW

Karmokar et al. [11] developed a project called the Tea Leaf Diseases Recognizer (TLDR) to detect and classify different tea leaf illnesses. In the TLDR method of image processing, the tea leaf picture is transformed by adjusting its dimensions, parameters, and parameters' limits. They have also employed a feature extraction method. The Ensemble Neural Network was utilized for the purpose of pattern recognition. The ANN is trained using the retrieved characteristics and the illness type. There was a 91% success rate after the testing phase. Hossain et al. [12] created an automated system for identifying three types of tea leaf disease using the Support Vector Machine (SVM), a machine learning (ML) technique with less feature counts. The proposed method may classify a leaf 300 milliseconds quicker than previous studies that used SVM as a classifier. Ramesh et al. [13] created an algorithm to detect irregularities reported on crops in their controlled or outdoor greenhouse habitat. Using the Random Forest classifier and a total of 160 photographs of papaya tree leaves, the suggested model was trained. The Histogram of Oriented Gradients (HoG) method of feature abstracting was utilized to build the training dataset's feature vector. Around 70% of the time, the model's classifications were accurate. In [14] a deep CNN was created with the goal of being able to recognize different forms of tea leaf illnesses from photographs of the leaf. The CNN model LeafNet utilizes feature extractor filters of varying sizes to automatically extract the features of the tea leaf disease from a collection of pictures. which employs SVM and MLP classifiers, is used to identify diseases. It is constructed using characteristics obtained from the Dense Scale Invariant Feature Transform. The LeafNet method recognized tea leaf disease with the highest average accuracy, or 90.16 percent, as compared to the SVM method's accuracy of 60.62 percent and the MLP algorithm's accuracy of 70.77 percent. Mukhopadhyay et al. [15] used cutting-edge computational methods including NSGA-II, principal component analysis (PCA), and multi-class support vector machine (SVM) to build a model for illness detection in tea leaves. The Silhouette index is used to verify the accuracy of the proposed NSGA-II-based photo clustering algorithm. After that, a multi-class support vector machine (SVM) is used to diagnose illness based on a selected set of features obtained using principal component analysis (PCA). The suggested model can read tea leaves and detect five different diseases. Overall, this model has an accuracy rate of 83%.

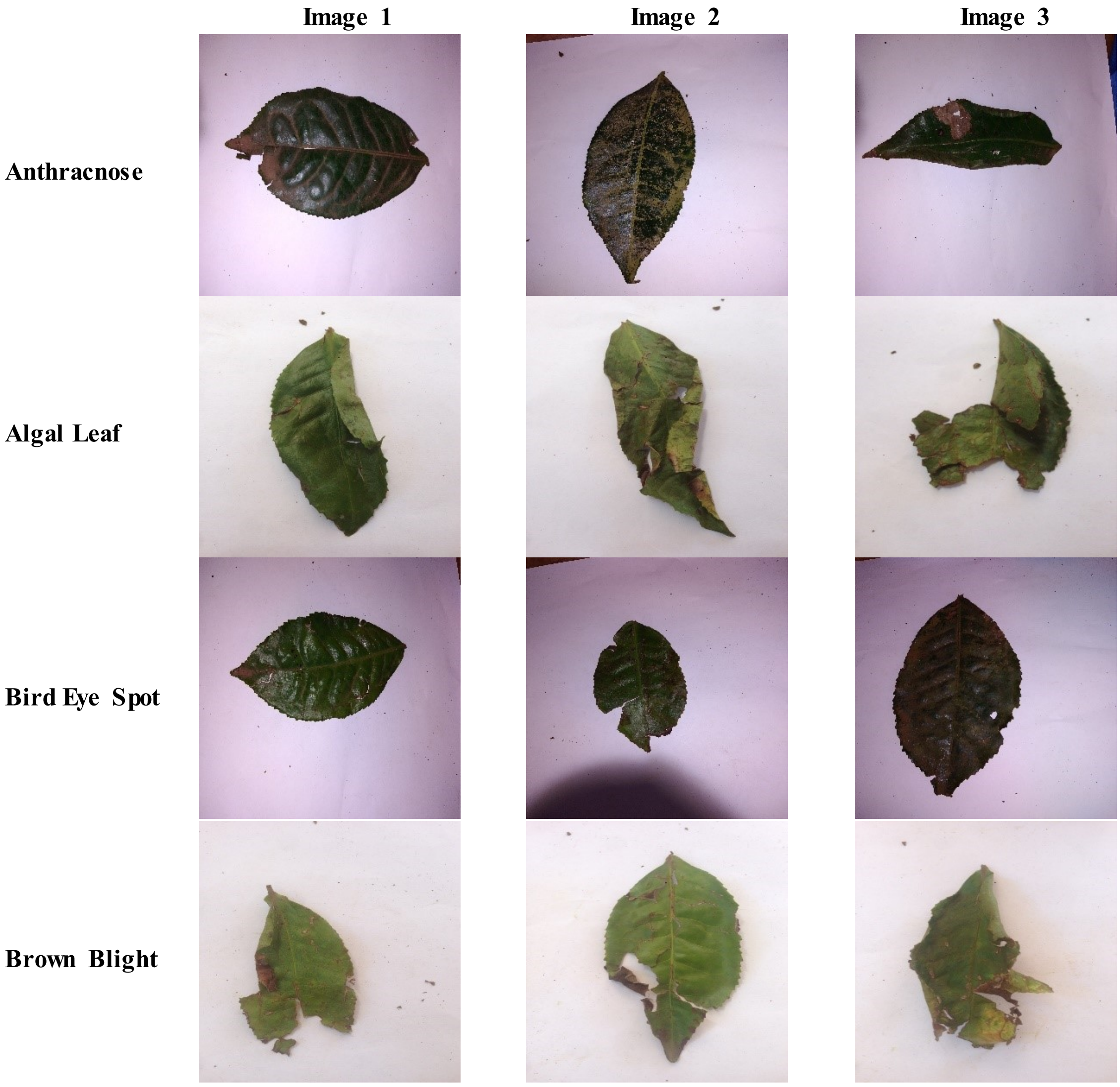
III. DATASET DESCRIPTION

The table 1 provides information on the number of photos in each class of tea leaf disease. The dataset contains 8 classes of tea leaf diseases, including Anthracnose, Algal Leaf, Bird Eye Spot, Brown Blight, Gray Light, Healthy, Red Leaf Spot, and White Spot, with a total of 885 images. Among them, the largest number of images belong to the Red Leaf Spot and White Spot classes with 143 and 142 images, respectively, while the smallest number of images are in the Healthy class with only 74 images. The images in the dataset is same size i.e., 224x224. To examine the implementation of the machine learning model, the dataset is divided into two parts, where 708 photos are employed for training and 177 photos are reserved for validation. The availability of this dataset will enable researchers to develop accurate and reliable algorithms for the automatic recognition and diagnosis of tea leaf diseases, which can significantly improve the performance and quality of tea production some images of input dataset are shown in figure 1.

Table:1 Input Dataset

Tea Leaf Disease	Images in each classis
Anthracnose	100
Algal Leaf	113
Bird Eye Spot	100
Brown Blight	113
Gray Light	100
Healthy	74
Red Leaf Spot	143
White Spot	142

Total images in 8 class: 885
Images used for training: 708
Images used for validation: 177



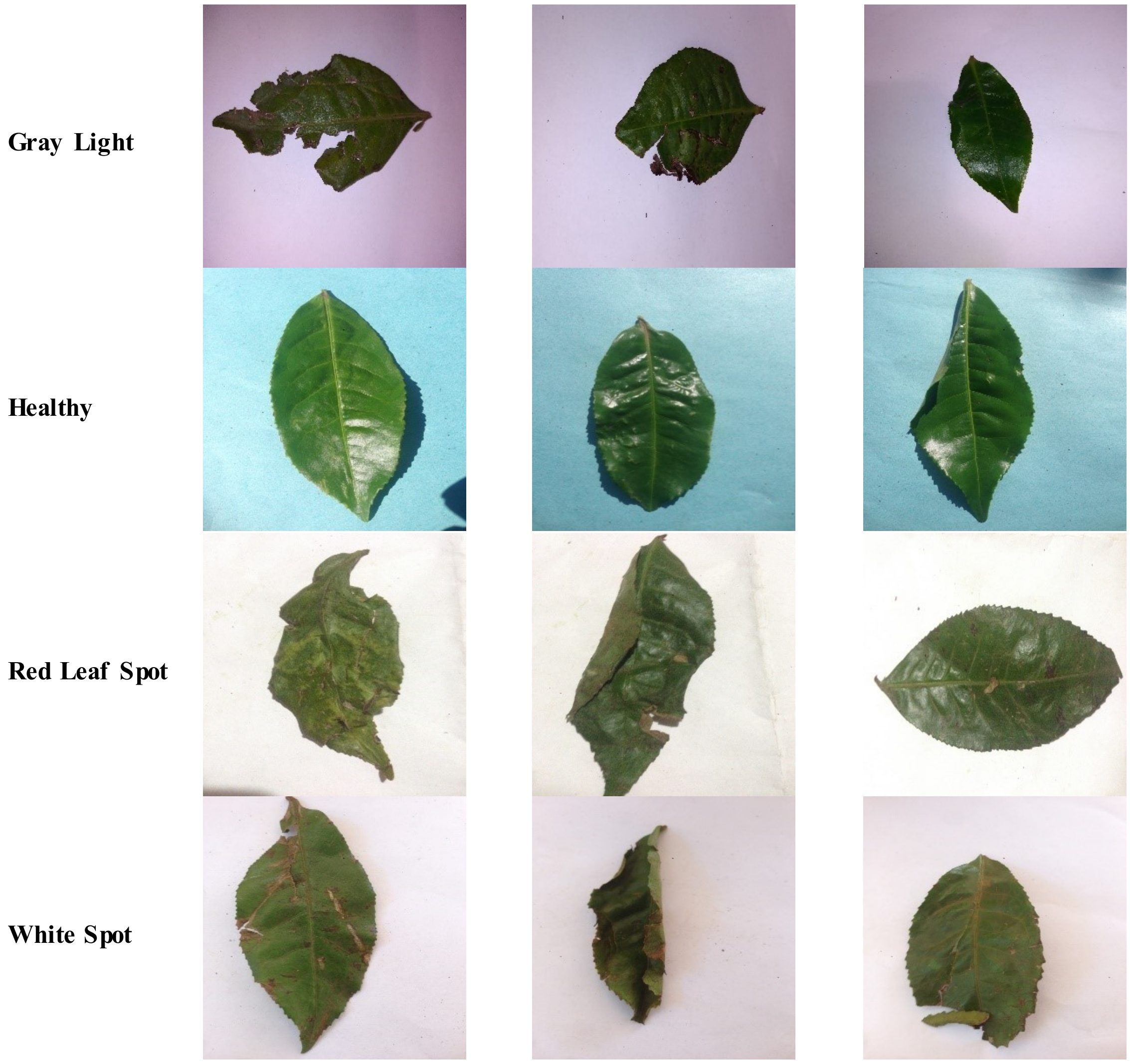


Figure: 1 Input Dataset

A. Data Augmentation

In the field of research, data augmentation is a frequently used technique in machine learning and computer vision for developing new variations of an existing dataset. It is especially beneficial when the original dataset is tiny or when the model needs to learn different versions of the same object. Data augmentation can help increase the robustness and generalization of the model by applying various alterations to the original images, resulting in better results. Some of the most prevalent types of data augmentation are: Horizontal Flip: This transformation flips the image horizontally, generating a mirror image of the original as shown in figure 2. It can be useful when the orientation of the object has no effect on its classification, and it can help to balance the dataset. Vertical Flip: Like horizontal flip, this transformation flips the image vertically, generating a mirror image of the original. It can also be useful when the orientation of the object has no effect on its categorization. Random Rotation: This transformation rotates the image by a random angle, imitating fluctuations in the object's orientation. The model can learn to detect objects at varied angles by introducing a modest bit of rotation, leading to better generalization. Random Zoom: This transformation zooms in or out of the image by a random factor, imitating fluctuations in the object's scale. By using a little amount of zoom, the model can learn to recognize objects of various sizes, resulting in increased resilience.

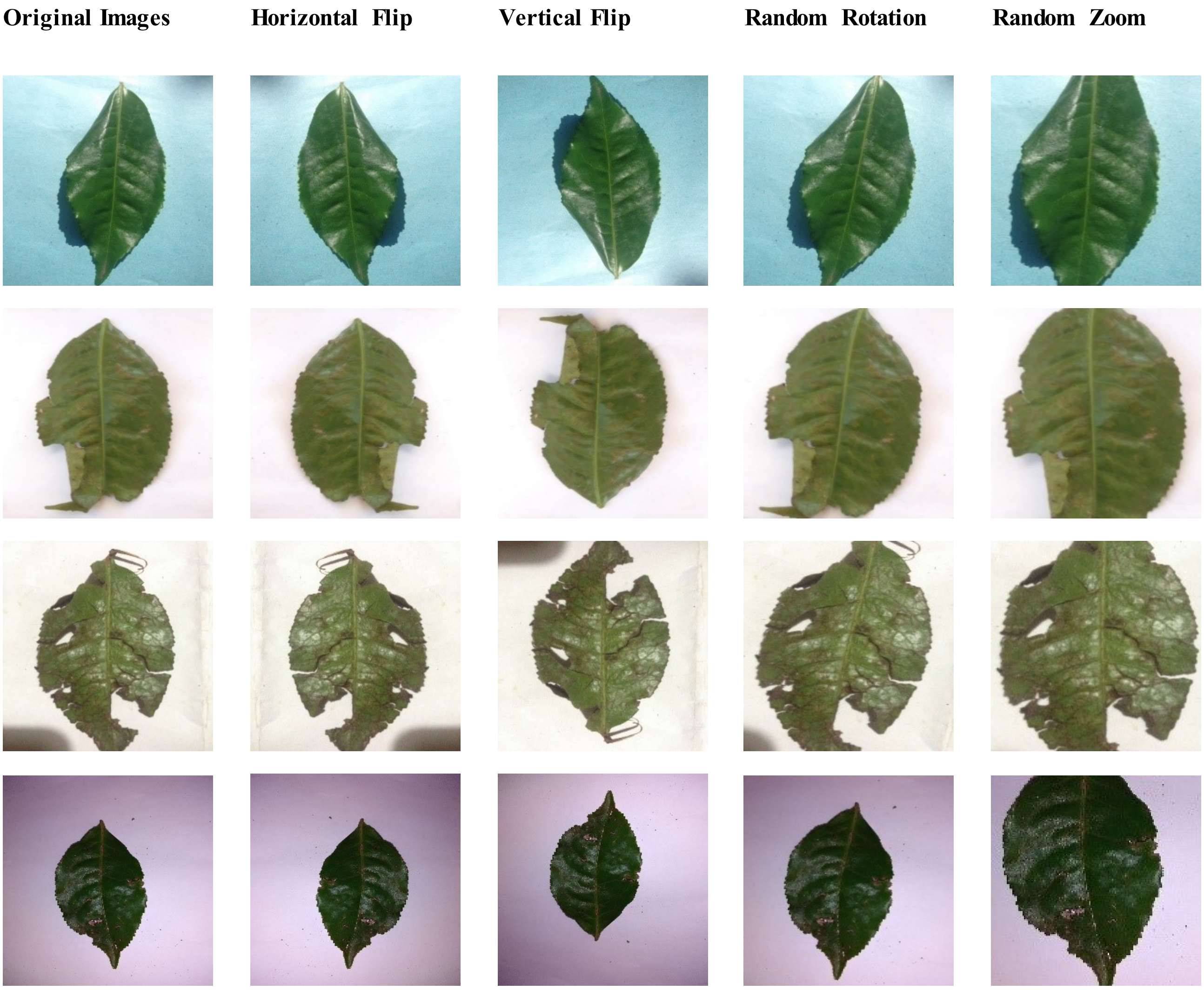


Figure: 2 Data Augmentation

B. Proposed CNN Model

The proposed CNN model for image classification is outlined in Figure 3. The CNN model proposed in this work has been designed from scratch. A lot of experimentation has been performed to finalize the number and sequence of layers in the CNN model. The final CNN model has two three convolution and maxpooling layers followed by flatten and three dense layers. The model begins with an input layer, which is followed by rescaling to ensure that all images are of the same dimensions. This is crucial for the convolution neural network to operate effectively. Next, a set of convolution and max pooling layers are applied in bundles of three. This is followed by a Flatten layer and a Dense layer with dropout to prevent overfitting. A batch normalization layer is then applied, followed by another Dense layer with dropout. The classification is determined by repeating these layers without dropout. The model is designed to identify eight categories: Anthracnose, Algal Leaf, Bird Eye Spot, Brown Blight, Gray Light, Healthy, Red Leaf Spot, and White Spot.

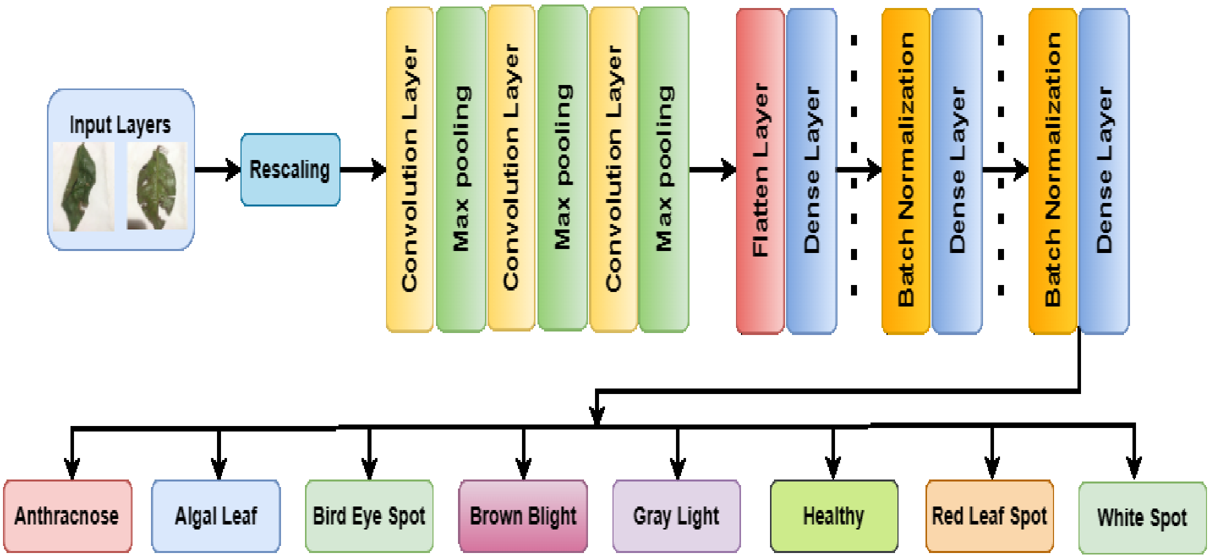


Figure: 3 Proposed CNN Model

Table 2 depicts the architecture of a deep neural network, which includes several layers such as Rescaling, Conv2D, Maxpooling2D, Flatten, Dense, Dropout, and Batch Normalization. The model's input shape is 224x224x3, and the output shape is not explicitly defined for most of the layers. The model begins with three filters in the first Conv2D layer and gradually increases to 32 in the final Conv2D layer. The model has a total of 811,832 parameters, of which 811,704 are trainable and the remaining 128 are non-trainable. The model architecture appears to be intended for image classification tasks, as it includes convolutional and pooling layers, succeeded by fully connected classification layers. By reducing overfitting, the addition of Batch Normalization and Dropout layers can improve the model's generalization ability.

Table:2 Architecture of Proposed Model

Layers	Input shape	Output shape	Number of filters	Parameters
Sequential	224x224x3	224x224x3	3	0
Rescaling	224x224x3	224x224x3	3	0
Conv2d	224x224x3	224x224x16	16	448
Maxpooling2d	224x224x16	112x112x16	16	0
Conv2d_1	112x112x16	112x112x16	16	2320
Maxpooling2d_1	112x112x16	112x112x16	16	0
Conv2d_2	112x112x16	56x56x32	32	4640
Maxpooling2d_2	56x56x32	28x28x32	32	0
Flatten	-	-	-	0
Dense	-	-	-	802848
Dropout	-	-	-	0
Batch Normalization	-	-	-	128
Dense_1	-	-	-	1056
Dropout _1	-	-	-	0
Batch Normalization_1	-	-	-	128
Dense_2	-	-	-	264

Total Parameters: 811,832
Trainable Parameters: 811,704
Non-Trainable Parameters: 128

IV. RESULTS

The research presents a CNN model with a proposed configuration for categorizing tea leaf disease into eight unique kinds. The model was trained for 320 epochs with a batch size of 128 and its accuracy and loss were examined alongside its classification parameters. The results show that the suggested CNN model accurately classified the various forms of tea leaf disease, with excellent precision and recall scores. As a result, the study implies that the proposed CNN model could be a reasonable option for accurately diagnosing tea leaf illness. The following sections of this study paper present and thoroughly examine these findings.

A. Accuracy and Loss Plots

During training, machine learning models are monitored using accuracy and loss charts. An accuracy plot shows how accurately the model predicts the output class label for each input sample. It displays model accuracy as a proportion of the number of training epochs. The accuracy plot typically climbs as the model learns from data during training. A loss plot displays the model's ability to minimize the difference between its predicted and actual output for each input sample. It plots model loss against training epochs. The loss plot normally lowers with training. A lower loss values indicates that the model is making more accurate predictions, which improves its performance. To evaluate the model's prediction abilities, it was trained with 128 and 320 epochs. The training accuracy and model loss charts are shown in Figures 4(a) and 4(b). These graphs depict training accuracy and loss with time. The blue and orange lines represent training and validation accuracy and loss. The plots depict the model's learning behavior during training, whereas the validation data assesses the model's generalization.

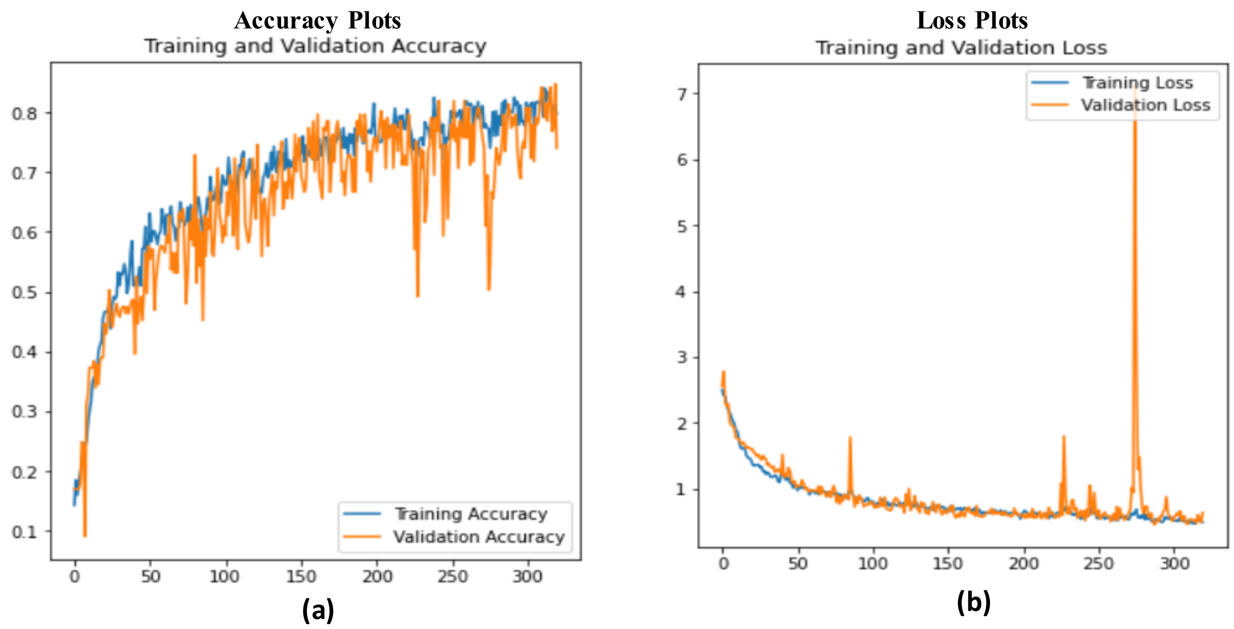


Figure: 4 Accuracy and Loss Plots

B. Classification Parameters

The table 3 provided the performance metrics of a classification model for identifying different diseases affecting plant leaves. The model was evaluated on various measures such as precision, recall, F1-score, and accuracy, for different classes of diseases. The results indicate that the model performed well for most classes, with high precision and recall scores, indicating good predictive power. However, there were some classes where the model performed relatively poorly, such as bird eye spot and red leaf spot, which had lower F1-scores. These classes may require further investigation and refinement of the model to improve its accuracy. Overall, the model achieved an accuracy of 0.73, indicating that it correctly classified 73% of the samples in the test dataset. The findings suggest that machine learning algorithms can be a useful tool for disease diagnosis and management in agriculture, helping to improve crop yield and productivity.

Table:3 Performance parameter

Name of the disease	Class	Precision	Recall	F1- Score	Accuracy
Anthracnose	0	0.75	0.50	0.60	0.73
Algal Leaf	1	0.86	0.92	0.89	
Bird Eye Spot	2	0.48	0.91	0.62	
Brown Blight	3	0.86	0.83	0.84	
Gray Light	4	0.94	0.75	0.83	
Healthy	5	1.00	1.00	1.00	
Red Leaf Spot	6	1.00	0.33	0.50	
White Spot	7	0.52	0.76	0.62	

V. CONCLUSION

Tea is the most popular and widely consumed beverage in India and globally. However, tea production is heavily impacted by various diseases. The identification and evaluation of the severity of tea illnesses continue to rely on arduous and time-consuming manual approaches. Thus, precise and prompt detection of tea leaf diseases is required for their prevention and management. This study proposes an enhanced deep convolutional neural network classification approach for tea leaf diseases. The aim is to build a deep CNN model that can detect tea plant diseases from leaf image sets. Experimental results indicate an average identification accuracy of 73%, which is superior to traditional manual methods. In order to enhance the diagnostic evaluation of tea leaves and other plant leaves, the CNN model is implemented in further applications. These findings have significant implications for the tea industry, as they can improve disease prevention and control, ultimately increasing the quality and quantity of tea production.

REFERENCES

[1] Hu, G., Wang, H., Zhang, Y. and Wan, M., 2021. Detection and severity analysis of tea leaf blight based on deep learning. *Computers & Electrical Engineering*, *90*, p.107023.
[2] Hu, G., Yang, X., Zhang, Y. and Wan, M., 2019. Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustainable Computing: Informatics and Systems*, *24*, p.100353.
[3] Latha, R.S., Sreekanth, G.R., Suganthe, R.C., Rajadevi, R., Karthikeyan, S., Kanivel, S. and Inbaraj, B., 2021, January. Automatic detection of tea leaf diseases using deep convolution neural network. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE.
[4] Bao, W., Fan, T., Hu, G., Liang, D. and Li, H., 2022. Detection and identification of tea leaf diseases based on AX-RetinaNet. *Scientific Reports*, *12*(1), p.2183.
[5] Chen, J., Liu, Q. and Gao, L., 2019. Visual tea leaf disease recognition using a convolutional neural network model. *Symmetry*, *11*(3), p.343.
[6] Lanjewar, M.G. and Panchbhai, K.G., 2023. Convolutional neural network based tea leaf disease prediction system on smart phone using paas cloud. *Neural Computing and Applications*, *35*(3), pp.2755-2771.
[7] Gayathri, S., Wise, D.J.W., Shamini, P.B. and Muthukumaran, N., 2020, July. Image analysis and detection of tea leaf disease using deep learning. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 398-403). IEEE.
[8] Sun, X., Mu, S., Xu, Y., Cao, Z. and Su, T., 2019. Image recognition of tea leaf diseases based on convolutional neural network. *arXiv preprint arXiv:1901.02694*.
[9] Xiaoxiao, S.U.N., Shaomin, M.U., Yongyu, X.U., Zhihao, C.A.O. and Tingting, S.U., 2018, December. Image recognition of tea leaf diseases based on convolutional neural network. In *2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)* (pp. 304-309). IEEE.
[10] He, M., Huang, F. and Liu, Z., 2017. Occurrence and Control of Major Leaf Diseases in Tea [J]. *Modern Agricultural Science and Technology*, *14*, pp.110-111.
[11] Karmokar, B.C., Ullah, M.S., Siddiquee, M.K. and Alam, K.M.R., 2015. Tea leaf diseases recognition using neural network ensemble. *International Journal of Computer Applications*, *114*(17).
[12] Hossain, S., Mou, R.M., Hasan, M.M., Chakraborty, S. and Razzak, M.A., 2018, March. Recognition and detection of tea leaf's diseases using support vector machine. In *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)* (pp. 150-154). IEEE.
[13] Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Shashank, N. and Vinod, P.V., 2018, April. Plant disease detection using machine learning. In *2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C)* (pp. 41-45). IEEE.
[14] Chen, J., Liu, Q. and Gao, L., 2019. Visual tea leaf disease recognition using a convolutional neural network model. *Symmetry*, *11*(3), p.343.
[15] Mukhopadhyay, S., Paul, M., Pal, R. and De, D., 2021. Tea leaf disease detection using multi-objective image segmentation. *Multimedia Tools and Applications*, *80*, pp.753-771.
[16] Singh, R., Sharma, A., Anand, V. and Gupta, R., 2022, December. Impact of EfficientNetB3 for stratification of Tomato Leaves Disease. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology* (pp. 1373-1378). IEEE.
[17] Anand, V., Gupta, S., Koundal, D. and Singh, K., 2023. Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images. *Expert Systems with Applications*, *213*, p.119230.
[18] Sharma, N., Gupta, S., Mehta, P., Cheng, X., Shankar, A., Singh, P. and Nayak, S.R., 2022. Offline signature verification using deep neural network with application to computer vision. *Journal of Electronic Imaging*, *31*(4), pp.041210-041210.
[19] Anand, V., Gupta, S., Koundal, D., Nayak, S.R., Barsocchi, P. and Bhoi, A.K., 2022. Modified U-net architecture for segmentation of skin lesion. *Sensors*, *22*(3), p.867.
[20] Sharma, N., Sharma, A. and Gupta, S., 2022, December. A Comprehensive Review for Classification and Segmentation of Gastro Intestine Tract. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology* (pp. 1493-1499). IEEE.