Tea Leaf Diseases Classification and Detection using a Convolutional Neural Network

[1]Vishesh Tanwar*,[2]Shweta Lamba
[1,2]Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab, India
[1]vishesh.tanwar@chitkara.edu.in*
[2]shweta.lamba@chitkara.edu.in

2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS) | 979-8-3503-3360-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCSS57650.2023.10169680

***Abstract*— It is worth noting the fact that tea is the most popular drink on the planet and India is one of the top producers and consumers of it. However, many diseases that affect crop quality and yield can interfere with tea production. Machine learning techniques such as deep learning are making it easier to identify and classify these diseases in tea leaves. The presence of disease symptoms in tea leaves can be used to classify and identify various diseases using deep learning techniques such as Convolutional Neural Networks (CNN). This strategy helps detect disease early and maintain good health. Both of these are essential to sustainable agricultural practices. As manual detection can be time-consuming and require specialized personnel, applying image processing models can greatly aid in identifying diseases in a large number of tea leaves. A proposed study using CNN layers classified submitted photos into one of his eight categories with a staggering 96% accuracy.**

Keywords— Disease detection, Deep learning, Convolutional Neural Networks (CNN), Tea-Leaf, Plant-Leaf Infection Classification.

I. INTRODUCTION

Conventional machine learning (ML) algorithms have found use in a variety of contexts. Deep learning is entering a new era of application in the agricultural industry, particularly in the field of plant and leaf disease identification, thanks to the development of several cutting-edge image-processing techniques[1].Controlling agricultural issues is very important and it is made possible by new technologies. This monitoring is essential for lowering food demand and achieving high productivity[1]. These days, the most common application for artificial intelligence methods such as machine learning and a variety of deep learning algorithms is found in a wide range of agricultural operations. Deep learning is one of these methods, and it is the one that has proven to be most effective in the field of agricultural research due to its capacity for automatic feature extraction[2]. Deep learning also comprises some different deep learning algorithms.

One of the most essential functions in agricultural production is the early diagnosis of crop plant diseases. When crops are afflicted by illnesses, it leads to a significant reduction in economic growth [2]. As a result, rapid and correct identification of the disease is necessary to prevent the reduction in both the amount and quality of the output. To diagnose an illness in the lab, a process known as disease culturing is necessary. Nevertheless, this process may not always provide accurate findings within the allotted amount of time. [3]. One of the most important crops, tea, also known as "chai," is only grown in certain geographical and climatic zones. More than other crops, it is cultivated in practically all hill stations around the nation. Tea illnesses have a significant impact on exporting and manufacturing of tea. Crops' leaves and stems are only two examples of where illnesses might be found[4]. The categorization of the illness is incorrect as a consequence of normal eye observation and detection. This results in the improper use of fertilizer and insecticides. Hence, identifying tea leaf diseases is crucial for producing high-quality tea with a high yield[4][5]. The categorization of tea leaf diseases has made use of several machine learning methods. For the image-based classification issue, deep learning architectures such as Alex Net, Visual Geometry Group, and ResNet performed well in comparison to machine learning models. In this study[5], CNN was used to categorize the seven different diseases that might affect tea leaves. Normal leaf identification was also suggested. Using a convolution neural network, Gray blight, White spot, Brown blight, Bud blight, Leaf blight, and Red scab are just a few of the diseases that are taken into consideration [6].

A Research Purposes

The key goal of the strategy that has been proposed here is to make use of machine learning technology to detect and categorize the prediction of tea leaf disease. This will be accomplished through the use of the technique that has been provided. to correctly diagnose the many illnesses that are associated with tea. improving performance by integrating the most productive aspects of proven deep-learning techniques.

II. LITERATURE REVIEW

Recognizing and classifying the illness associated with the tea leaf has been challenging for a long time. Quite a few investigations on this topic have shown encouraging results. These results are discussed in more detail in the next section. To identify and classify illnesses in tea leaves, the researchers used an approach that blended CNN, VGG16, and SVM. The findings of many research suggest in the direction of the potential for the very precise categorization of the different leaf diseases. In this paper, the author performs picture recognition of damaged tea leaves using CNN. picture pre-processing begins with picture segmentation and data augmentation. Second, regular adjustments are made to the learning rate, iteration count, and dropout rate during over-tuning in order to increase CNN recognition accuracy. Lastly, the experimental findings demonstrate that CNN's recognition accuracy is 93.75%.

TABLE I DESCRIPTION OF THE LITERATURE

The year of publication	Classification of Infections	The technique of training and evaluation	Model Correctness
[6]2018	Tea Leaf Disease	CNN	93.7%
[7]2019	3 types of Tea Leaf infection	VGG16 and AlexNet	92.5%

[8]20219	Rice Blast diseases	CNN	90.16%
[9]2020	4 types of Tea Leaf infection	LeNet	90.23%
[10]2020	5 classes of tea leaf Child leaf, Defect, Insect, Perfect, Adult	R-CNN and VGG16	95.83%
[11]2021	Tea leaf blight	VGG16	84.5%
[12]2021	Multiple plant diseases like rice, tomato, and corn leaf infection	CNN based AlexNet	96.76 %
[13]2023	Sugarcane Disease	CNN+SVN	97%
[14]2023	Paddy Disease	SVM+CNN	94%

III. METHODOLOGY

The deep learning technique is comprised of the processes that were used to develop the classifier model. The whole procedure may be broken down into five distinct stages: the gathering of datasets, the development of datasets based on infection, the pre-processing of pictures, the extraction of features, and the categorization of images. Figure one is a representation of an experimental layout that includes a graphic that uses leaf pictures to determine whether or not the cane crop is infected with a disease. After then, the dataset used for testing and training is divided into these photographs in an 80:20 ratio. After the completion of an evaluation of the proposed technique using the training dataset, the method is next examined using the testing sample.

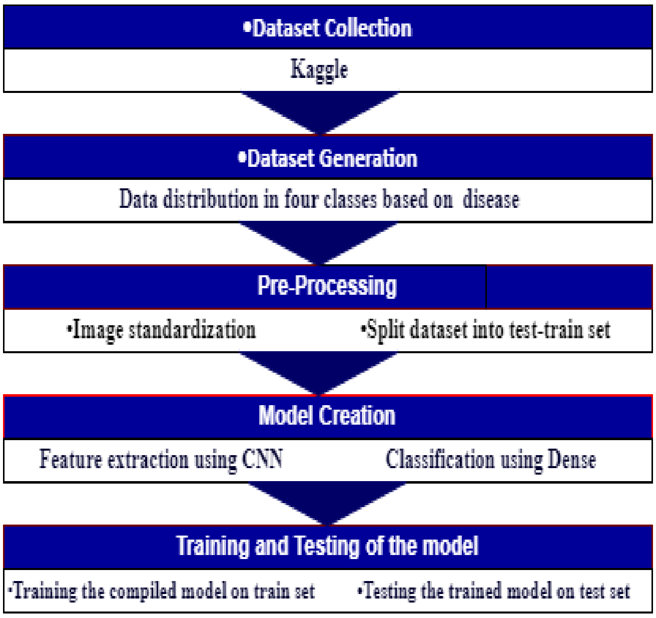

Fig 1. The Procedure of the Proposed Model

A. CONVOLUTION LAYER

The convolution layer has a large number of filters, all of which are utilized to extract the crucial details from the photographs given as input. The entire volume being fed into the image is less than the filters' height and width combined. To recognize the neuron-based feature map, each filter is combined with the input picture and adjusted.

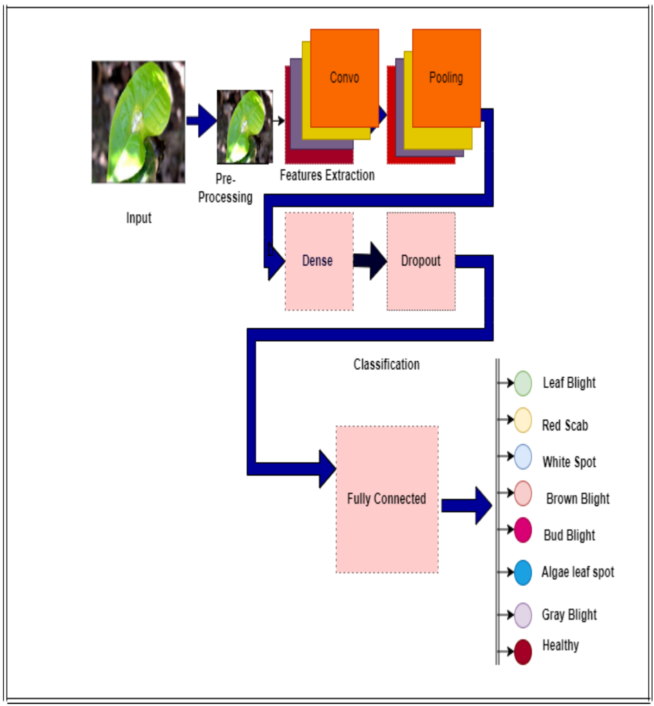

Figure 2. Proposed Model Aarchitecture

B. POOLING LAYER

The practice of merging, also known as pooling, may assist in minimizing the amount of data. The process that we call max pooling happens when we take the submatrix's highest value and add it all together. If we take the average of the values, then what we have here is an example of average pooling. Figure 3 illustrates this point quite well.

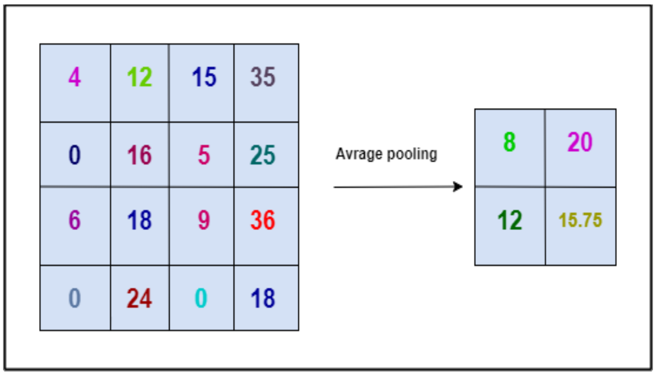

Figure 3 Pooling

C. FLATTENING LAYER

A completely integrated feed-forward neural network receives its input in the form of a one-dimensional vector, which is generated by the flattened layer after the data from a two-dimensional array has been converted into it. One single lengthy feature vector is produced when the outcome of the convolutional layers is flattened. With the use of this information, the NN(Neural-Network) is educated to recognize the many illnesses that might affect tea leaves.

D. RECTIFIED LINEAR UNIT(ReLU)

ReLU is a very useful activation function that is used extensively across the CNN network. It is not like other activation mechanisms in that it does not enable all neurons at once. This suggests that the ReLU will convert a -ve input to a value of zero, hence avoiding the activation of the neuron when it receives the input.

E. FLOW DIAGRAM

The process flow chart for the recommended system for identifying diseases in tea leaves is shown in Figure 4. It consists of some processes, including pre-processing, the extraction of features, prediction, and mapping of the pathogen area to identify illnesses. The architecture of CNN is based on stacked layers, each of which is processed in turn.

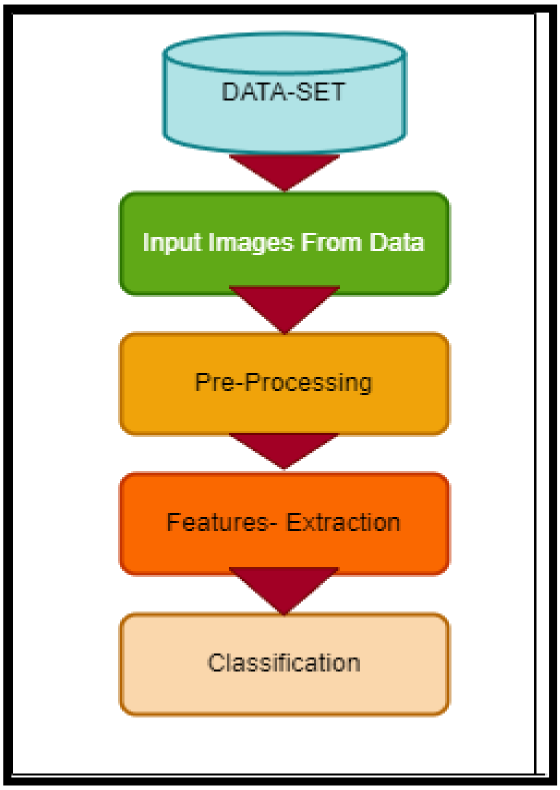

Figure 4 Diagrammatic representation of the categorization of diseases Pre-Processing

As the dataset originates from a separate source, the photos are not standard. So, it is necessary to standardize and feed the model with all of the photos. A straightforward pre-processing procedure is utilized to normalize the photos. There are several pre-processing methods, including resizing, scaling, normalizing, and dimensionality reduction of the photos. Since the bigger picture will take up more space, the temporal complexity will rise, creating a massive neural network. Hence, resizing photos is done to make them uniform. All of the photographs in the suggested system are scaled down to 100 X 100 X 3. The convolution layer will then receive the pre-processed pictures as input to the begin extraction process.

F. Feature Extraction

To extract the features of the photos, convolutional neural network architecture could be employed. The input picture is often presented as a two-dimensional array of neurons, with each neuron standing in for a different pixel group in the image. When the images have been preprocessed, CNN architecture is applied to them to recognize and categorize the images by extracting attributes from them. The CNN architecture is made up of convolutional layers, pooling layers, and FC layers, all of which are employed for feature extraction. The convolution layer is one of them, and it comprises numerous filters that search for a certain feature at each place on the input image. The majority of convolution layers consist of 64 or 128 filters, which leads to exponential growth. The input is divided into two halves by the filters at a default stride of 1X1. Each layer of the convolutional representation has a ReLU function and a pooling layer.

The output of the last layer of convolution is the data that is sent into the flattening layer as its input. By the use of the flattening layer, the two-dimensional matrix is converted into a vector format. It must be finished before access may be granted to the layer that is connected. Using the ReLU activation technique, all 2048 nodes that make up the first layer have been activated. This layer is called the root layer.

The second layer is referred to as the output layer. The category of the picture that was provided as input is identified according to which label has the highest likelihood. The output layer is distinguished into 8 categories, such as healthy leaves and leaves with illness.

G. Dataset

The dataset was obtained via an online source similar to Kaggle [15]. It incorporates every single leaf that was afflicted by the plant disease from which photographs of tea leaf disease were obtained. The use of a camera in the wild allows for the creation of certain sorts of photographs. The overall amount of pictures in the collection is 860. We have decided to use 80% of those images for training, while the remaining 20% will be used for testing. The following is a list of the total number of photographs included in each category: Tea leaves with a variety of ailments are seen in Table II.

Table II. Dataset Details

Tea Leaf's Disease Data		
Class	**Disease**	**No of images**
1	Leaf Blight	120
2	Red Scab	98
3	White Spot	92
4	Brown Blight	122
5	Bud Blight	112
6	Algae leaf spot	111
7	Gray Blight	91
8	Healthy leaf	114

Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS 2023)
IEEE Xplore Part Number: CFP23DJ3-ART; ISBN: 979-8-3503-3360-2

Fig 5 Sample Images of Leaf Diseases

IV. Result Analysis

While evaluating the suggested job, accuracy is regarded as one of the most crucial criteria to take into consideration. The accuracy is measured as the ratio of successfully detected picture classes to the total number of image classes concerning the entire amount of photos that were taken into consideration.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Calculating accuracy in multiclass classification is as easy as dividing the number of accurate categories by the total number of classifications.

$$\text{Accuracy} = \frac{\text{correct classifications}}{\text{all classifications}}$$

- True-Positive (TP)- Tea leaves that have been contaminated with infection are appropriately labelled as infected leaves.
- TN-The non-infected leaves that have been appropriately recognized as not being harmed by the contaminant are called true negatives.
- False-Positive (FP) is a term that describes the situation in which undamaged tea leaves are incorrectly classified as being impacted.
- FN-The leaf that is infected with illnesses is incorrectly categorized as the typical normal leaf, which is a false negative.

The effectiveness of the suggested approach is shown in Figures six and seven, which are evaluated in terms of accuracy and loss versus epoch. One full iteration of the training process on all of the data constitutes an epoch. It is pretty obvious that increasing the number of epochs would result in an improvement in the correctness of the job that is being presented. It is possible that increasing the number of epochs would improve the accuracy of the model, but this is not something that can be said with absolute certainty. An example of overfitting is when a model is trained for an excessively long period of time (an excessive number of epochs), which causes the model to remember the training data rather than learn how to generalize to new data.

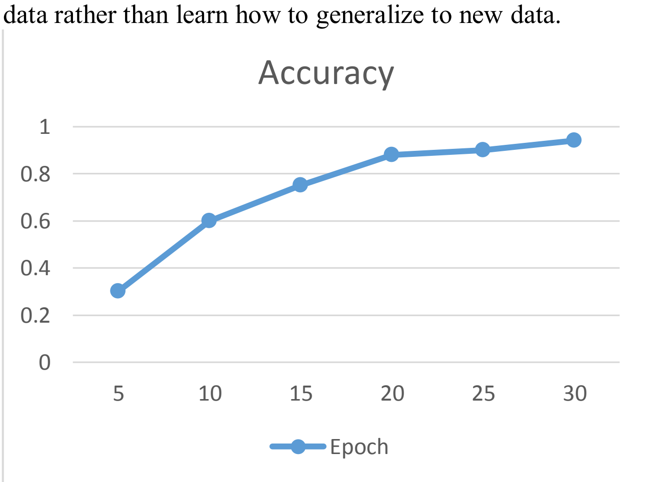

Fig. 6 Accuracy curve of the presented model.

In addition, the dataset and model architecture that is being used are the two factors that decide the best number of epochs to employ. Throughout the training process, it is essential to keep a close eye on the validation loss and to halt the procedure if the loss of validation starts to increase, as this is an indication that the model is being overfitted.

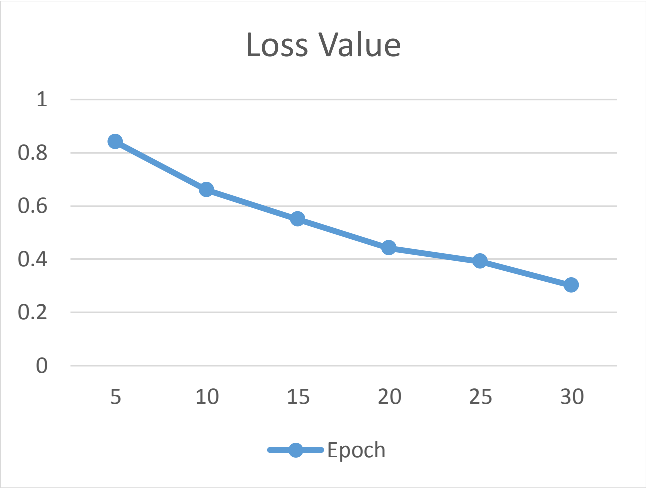

Fig. 7 Loss valued curve of the presented model.

In most cases, the accuracy of the training keeps getting better since the model that is supposed to discover the best fit for the training data has a tendency to overfit. The choice of the epoch value ought to be made in such a manner as to prevent the model from being overfitting.

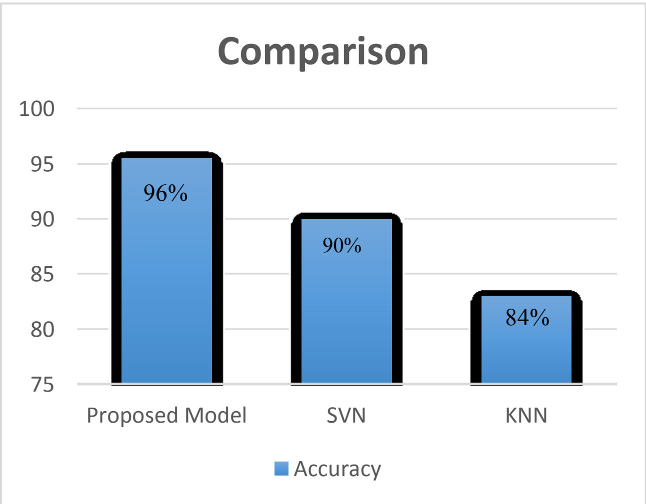

Fig 8 Assessment of suggested approach with other approaches.

The CNN classification techniques were used to construct the model that is suggested for tea leaf disease predictions. This model has a success rate of 96% and was built using these methods. Conventional SVM and KNN models, when applied to the same dataset, yield results with an accuracy that is lower than that of the results obtained by the CNN model. Figure 8 provides a visual representation of this point.

V Conclusion

The study that is being suggested has as its ultimate objective the categorization of illnesses and an improvement in accuracy when it comes to properly forecasting their occurrence. By accounting for some different illnesses that might affect tea leaves, the model that was suggested has an accuracy rate of 96% when it comes to properly identifying the condition. This is realizable by using the CNN architecture and utilizing deep learning. By making changes to the number of layers and other parameters like the optimizer that is now being used, the performance of the algorithm may be increased even more. Making use of the concept of transfer learning, one is able to do a comparison between the output of the model and the CNN architecture that is presently being used.

References

[1] D. Krisnandi et al., “Diseases classification for tea plant using concatenated convolution neural network,” CommIT (Commun. Inf. Technol.) J., vol. 13, no. 2, 2019.

[2] X. Sun, S. Mu, Y. Xu, Z. Cao, and T. Su, “Image recognition of tea leaf diseases based on convolutional neural network,” in 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2018.

[3] X. Sun, S. Mu, Y. Xu, Z. Cao, and T. Su, “Image recognition of tea leaf diseases based on convolutional neural network,” 2019.

[4] Kumar, Atul, and Ishu Sharma. "Enhancing Cybersecurity Policies with Blockchain Technology: A Survey." In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), pp. 1050-1054. IEEE, 2022.

[5] S. Bhowmik, A. K. Talukdar, and K. Kumar Sarma, “Detection of disease in tea leaves using convolution neural network,” in 2020 Advanced Communication Technologies and Signal Processing (ACTS), 2020.

[6] X. Sun, S. Mu, Y. Xu, Z. Cao, and T. Su, “Image recognition of tea leaf diseases based on convolutional neural network,” in 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2018.

[7] G. Hu, X. Yang, Y. Zhang, and M. Wan, “Identification of tea leaf diseases by using an improved deep convolutional neural network,” Sustain. Comput. Inform. Syst., vol. 24, no. 100353, p. 100353, 2019.

[8] W. jie Liang, H. Zhang, G. feng Zhang, and H. xin Cao, “Rice Blast Disease Recognition Using a Deep Convolutional Neural Network,” Sci Rep, vol. 9, no. 1, pp. 1-- 10, Dec. 2019, doi: 10.1038/s41598-019-38966-0.

[9] M. H. Kamrul, M. Rahman, M. R. I. Robin, M. S. Hossain, M. H. Hasan, and P. Paul, “A deep learning based approach on categorization of tea leaf,” in Proceedings of the International Conference on Computing Advancements, 2020.

[10]M. H. Kamrul, M. Rahman, M. R. I. Robin, M. S. Hossain, M. H. Hasan, and P. Paul, “A deep learning based approach on categorization of tea leaf,” in Proceedings of the International Conference on Computing Advancements, 2020.

[11]G. Hu, H. Wang, Y. Zhang, and M. Wan, “Detection and severity analysis of tea leaf blight based on deep learning,” Comput. Electr. Eng., vol. 90, no. 107023, p. 107023, 2021.,

[12]P. B R, A. Ashok and S. H. A V, "Plant Disease Detection and Classification Using Deep Learning Model," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 1285-1291, doi: 10.1109/ICIRCA51532.2021.9544729.

[13]V. Tanwar, S. Lamba, and B. Sharma, “Deep learning-based hybrid model for severity prediction of leaf smut sugarcane infection,” in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2023.

[14]S. Lamba, V. Kukreja, A. Baliyan, S. Rani, and S. H. Ahmed, “A novel hybrid severity prediction model for blast paddy disease using machine learning,” Sustainability, vol. 15, no. 2, p. 1502, 2023.Chen, J., Liu, Q., & Gao, L. (2019).

[15]“tea-leaf-disease-classify,” Kaggle.com, 2023. [Online]. Available: https://www.kaggle.com/code/visheshtanwar26/tea-leaf-disease-classify-monai-pytorch/edit. [Accessed: 11-Apr-2023].