TOMATO LEAF DISEASE DETECTION USING CNN

Thejaswini k Lamani[1]

Department of Master of Computer Applications, Dayananda Sagar Academy of Technology and Management Bangalore 560082, India
tejulamani720@gmail.com

Shreedhara N H[2]

Department of Master of Computer Applications, Dayananda Sagar Academy of Technology and Management Bangalore 560082, India
Shreedhara-mca@dsatm.edu.in

Abstract

The detection of tomato leaf is the critical task in agriculture, as it can help to prevent crop loss and increase yields. It is the most crucial and widely utilised crops in the world, Considering the quantity and quality of the yield can be greatly impacted by leaf diseases. These diseases cannot only affect the agricultural industry, but also human health and financials. One of the major challenges in managing leaf diseases is detecting infected plants in the early stages, as diseases can spread rapidly and infest entire farms if left untreated. These systems typically involve capturing images of tomato leaves, pre-processing the images to enhance visual features, and using machine learning techniques to develop illness detection models. CNN is one of the most well-known and effective image classification techniques, and it has been effectively used to address a number of tomato disease- related issues including object identification, image classification, and semantic segmentation. The suggested technique makes use of image processing methods and a machine learning classifier, specifically a CNN, to diagnose the disease. The results show good accuracy in disease identification using the suggested approach, which uses a dataset of tomato leaf pictures to test performance. The proposed method can be a useful early detection tool and management of tomato leaf diseases, potentially leading to increased crop yields and reduced use of pesticides. The study also aims to implement a CNN specifically for the application of plant disease detection in tomato leaves.

Keywords—leaf, healthy, unhealthy, remedies, detection

I. INTRODUCTION

The device uses machine learning-based approach for disease detection of tomato plant. This detector's objective is to enhance the e the efficiencies and accuracy of disease detectionin tomato plants, which can help farmers to take appropriate action to prevent growth of diseases and protect their crops. To perform this objective, it focus on collecting data of diseases onplants and training a model for disease detection in the early stage of tomato. The device will employ image processing methods to pre- process the images and extract the features that used by a ml algorithm for disease detection. The ml algorithm used will likely be CNN, which shown effective in image recognition, classification, and detection of colour and size of leaves. It also classify the leaves of plant into unhealthy or healthy, if the result is unhealthy, the system will show name ofpredicted disease. and also suggest remedies for particular diseases that help farmers. This strategy can be utilised to identify various leaf diseases. It is significant to remember that while this system is likely to be accurate, it is important to consult with a local expert or extension service for further diagnosis and treatment recommendations to ensure the best results. The datasets and pretrained models are also available for detecting leaf disease which is usable as starting point for the system's expansion. Designed a automated machine-learning technique for detecting tomato leaf disease or algorithm that can accurately and efficiently identify and diagnose tomato leaves might reveal many diseases that damage tomato plants. The system needs to be capable of discriminate between infected and healthy leaves also identify the specific illness kind that is present. To increase the effectiveness and precision of disease detection in tomato plants, which can help farmers and gardeners to take the necessary preventing the spread of diseases and safeguard their crops, ultimately resulting in increased crop yields and reduceduse of pesticides.

II. LITERATURE SURVEY

The many deep learning techniques used by researchers toidentify plant diseases are presented. According to

Prajwala T. M. used a ML method known as Convolutional Neural Network (CNN) in their research. Their study's purpose was to evaluate this's efficience algorithm using fourdistinct instruction sets: lenet, vgg16, resnet50, and xception. And emphasis of investigation was on analyzing images of tomato leaves. The hypothesis was that by utilizing multiple instruction sets and varying layers within the CNN, the software would be able to recognise more subtle characteristics and patterns compared to previous versions.[1]

Akshay Kumar, et al used a machine learning method calledSvm utilising to identify illnesses in tomato plant leaves pictures of the leaves of tomatoes. That time This method isnew and has not been use before in this field, and results were good showing a high rate of correct detection. However, Theresearch's usage of a dataset with a maximum size of 16500 may have had an influence on findings. To improve the model's output, it is advisable to use a bigger, more varied dataset. It's crucial to evaluate model's effectiveness in actualscenarios as well.[2]

Rangarajan et al. conducted a study in which they trained twodistinct neural network topologies,, namely AlexNet and VGG16net. They experimented with various hyper□parameters, such as minimum batch size, weight learning rate, and bias learning rate. Surprisingly, they determined that the accuracy of of VGG16net was negativelycorrelated with the minimum batch size, meaning that smaller batch sizes led to lower accuracy for this particular network. In addition, the researchers explored the concept of dimension reduction by combining convolution and pooling layers into a single module. They applied this technique to the Google Net architecture, specifically to the Inception V4model, aiming to reduce the data's dimensions being processed.[3]

VGG16 model with a CNN foundation in their study. The model has received widespread acclaim for its ability to effectively address image categorization and localizationissues. The proposed model achieved an impressive classification accuracy of 89%. A significant benefit of the proposed model is its high effectiveness in addressing the problem at hand. However It's significant to remember that that one study limitation is the utilization of relatively high□quality test images exclusively, which might impact the generalizability of the outcomes to actual-world scenarios with diverse image qualities.[4]

Robert G. de Luna et al. introduced a trained model based onCNN (Convolutional Neural Network) and F-RCNN (Feature-based Region Convolutional Neural Network). Their approach achieved an impressive performance of 91.67%. The model demonstrated several advantages, including the utilization of transfer learning, which led to accurate recognition. The method's disadvantage is that the CNN must be retrained in order to work at its best.[5]

Melike Sardogan et al, suggested a CNN model that the learning vector quantization (LVQ) algorithm to identify andcategorise plant leaf ailments. The model had an average classification accuracy of 86%. This means that it was able to correctly identify 86% of the plant leaf diseases and that itwas tested on. The model could Early detect the plant leaf diseases enables farmers to take proactive measures to stop the disease's progress. Future study increase the model's categorization rate, which is still a concern.[6]

Dheeb Al Bashish et al, proposed a technique to find diseasesof the leaves and stems. The research focused on utilizing a dataset that is collected from the Al-Ghor area in Jordan. Five specific Using image processing methods, the followingplant diseases were examined: Cottony Mould, Tiny Whiteness, Ashen Mould, and Late Scorch. The suggested method has numerous steps. The leaf photos were first obtained by image acquisition. The damaged and healthy parts of the image were then separated using the K-Means clustering technique. The texture of the diseased leaf and stem was examined using the Colour Co-occurrence Method(CCM) during obtaining the feature stage. Finally, a Backpropagation algorithm a neural network was used for disease classification. With a high precision rate of almost 93%, the outcomes of this image processing technologyshowed precise identification and categorization of plant diseases.[7]
Chowdhury, Muhammad EH, et al, emphasize the significance of automatic recognition for various mobile phone users, especially within the setting of identifying and extracting leaf regions from complex

backgrounds. Extracting the leaf region from intricate backgrounds poses several challenges, like the fact that changing background patterns, clutter, diverse leaf shapes and sizes, and fluctuating illumination due to unpredictable environmental conditions. The authors suggest that as a solution to this that gathers a substantial dataset comprising 18,161 images of tomato leaves, both in their original form and after being segmented.These images serve the basis for building a deep learning architecture based on an efficient net model that seeks to address the complexity associated with leaf region identification when there are complicated backgrounds, and extraction.[8]

H. Sabrol et al. proposed a classification tree model using supervised learning methods. The model is trained on a dataset of labeled data, which are image of tomato plants with and without diseases. The model then uses a classification tree to classify tomato plants into different disease categories. The model's accuracy was 97.3%, meaning that 97.3% of the test dataset's images were properly classified. The model's advantages include that it generates reliable findings. There are many more modern categorization techniques, In some situations, approaches to deep learning have demonstrated to be more accurate than classification tree methods.[9]

Santosh Adhikari et al. introduced a Ml method based on CNN (Convolutional Neural Network) architecture. Their approach incorporated augmentation strategies for data to mitigate overfitting during the training process The result is the model's overall accuracy was 89One of the prominent benefits of of their method is the utilization of data augmentation, which enhanced the model's performance. a drawback of the approach is the requirement of transferlearning to accurately classify all diseases.[10]

Srdjan Sladojevic et al. proposed an approach to for classifying images. 10,000 photos were as a part of the dataset used to train the CNN, and achieved an accuracy of 96.3%,with precision of CNN, according to the authors' evaluation of its performance on a held-out test set, was 95.7%. The proposed CNN technique is a promising method for classifying images, according to the authors' analysis. On a sizable image collection, the suggested technique can produce results with a high degree of accuracy. It is also relatively simple to implement For a particular dataset, accuracy may need to be improved through tweaking and augmentation. Overall, the proposed CNN technique is a promising approach for image classification.[11]

Belal A.M.Ashqar introduced a deep CNN that includes both models in grayscale and in full colour. The system demonstrated remarkable performance, achieving an accuracy of 99.84% when utilizing the full-colour model. The grayscale model focused on learning The leaf's form and pattern, while the colour model's primary objective was to distinguish damaged leaves. This strategy has several benefits, one of which is is its high feasibility, making it practical for real-world applications. However, a slight The system's drawback is that the recognition aspect may have some minor shortcomings.[12]

Keke Zhang et al, proposed the optimization methods through a deep neural network The study focused on employing the Stochastic Gradient Descent (SGD) algorithm with the ResNet technique to achieve optimal results. Remarkably, this combination yielded the maximum level of accuracy of 97.28%.The ability to adjust the network parameters is one of this strategy's benefits. And helps to conserve computational resources and reduces the required time for training. Consequently, the proposed method showcases good overall performance. However, it's important to remember that the optimization process can be time□consuming, which in some cases could be viewed as a negative.[13]

Anand.H.Kulkarni et al. suggested an artificial neural network (ANN)-based system for categorising and identifying damaged leaves. The input image is first filtered and segmented using a Gabor filter. The ANN is then trained using the filtered image. With a 91% accuracy rate, the ANN can classify and identify sick leaves. The strategy is promising for identifying and classifying damaged leaves. It is simple to implement and achieves high accuracy. However, it may not be as accurate to the same level as other methods uses more complex classifiers. This research has two key advantages: it can achieve high accuracy, and it is reasonably easy to use. The main demerit of the method is maybe incapable of achieving the same level of as other methods use more complex classifiers.[14]

III. METHODOLOGY

Figure 1 depicts the proposed system's intricate system architecture. The application of a ML system has been developed for the purpose of finding disease on tomato plants concentrating on identifying Mosaic Virus, Bacterial Spot, etc Healthy plants, Leaf Mold, and Septoria Leaf Spot. Modern machine learning techniques are used by this system. to accurately classify and detect these specific diseases in tomato plants. By automating the process, it enables efficient and reliable disease detection, aiding in the timely and targeted management of tomato plant health. The purpose of the project is to discover six different disease categories that damage tomato crops using an automated deep learning-based tomato disease detection system. The system we've suggested is summarised in the figure 1. Additionally, a thorough explanation of each component of system for detecting tomatodisease is provided below.

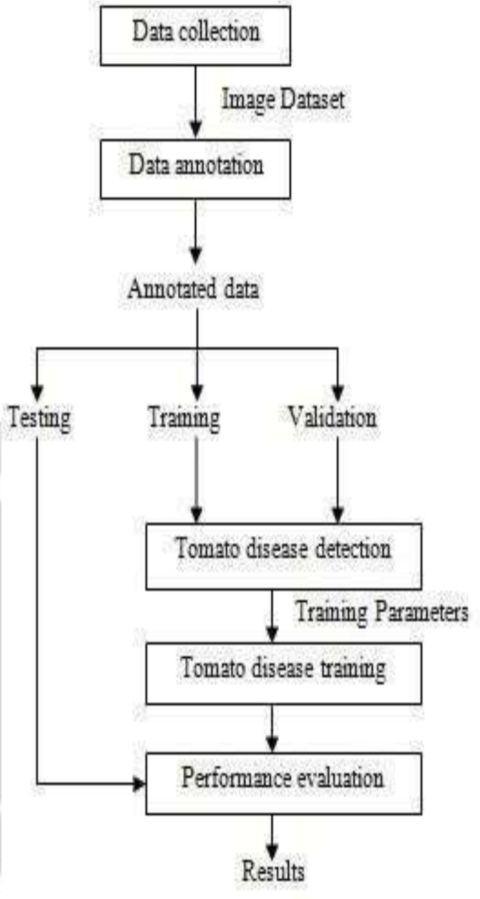

Fig. 1. Brief summary of machine learning based disease detectionsystem for tomato.

A . Data Gathering

To acquire a lot of data is the aim of data collection. and diverse dataset of wholesome and unwholesome leaf of tomato. A large number images of healthy leaves and unhealthy tomato leaf images are collected from the internet or from real-world sources like kaggle.com, data set search, Research, google.Com etc.

Fig. 2. Image that is collected

B. Data Pre-processing

The objective of Pre-processing raises the standard of the data, and minimize noise, and prepare it for training, ensuring its compatibility with the CNN model. The pre- processing steps are carried out in order to improve the data and make suitable for input into the CNN, thereby improving the overall model's efficiency and accuracy. These steps involve various techniques such as data cleaning, normalization ,resizing, and feature extraction, among others, to guarantee the data is in a suitable format and condition for effective utilization by the CNN.

C. Image resizing

The tomato leaf images could come in various sizes, so theyneed to be resized equivalent in size. This makes the data more consistent and suitable for CNN. Image normalization: The images need to be normalized to have zero mean and unit variance. This helps the CNN to converge faster and improves its performance.

D. Image augmentation

The dataset was enhanced using data and making feasible to create larger and that can be used to create new images by transforming the existing images. This enhances the model's generalisation capabilities and prevents overfitting.

E. Data balancing

With enough photographs, the model may begin to favour one class over another if the dataset is imbalanced, meaningthat there are more images of one class than the other. To overcome this, the data can be balanced by oversampling the class with fewer images or under sampling the class with more images.

F. Data Splitting

This step is performed to separate the given data into training,validation, and test sets.

Training data: The CNN model is trained using this data. A large portion of the available data (e.g. 70-80)percent is typically used for training

Validation data: The model's performance is assessed using this data during training. The model is trained using the training data, while the. Utilising the validation data, the performance of data is evaluated. This information avoid overfitting, which happens when a model becomes extremely complex and starts to memorise training data instead ofidentifying larger patterns.

Test data: This information is utilised to the model's final performance. The data is solely used to obtain a final evaluation of the model's performance and should not be used during training. This phase is crucial to make sure the model can generalise adequately to brand-new, untested data.

G. Model Training

CNNs is for recognition of the diseased leaf. The idea is to train the CNN to learn the patterns and features in leaf of tomato, and then use this learned information to classify newimages as either healthy or unhealthy. An image of a tomatoleaf is the input to CNN, and the prediction of whether the leaf is healthy or unhealthy is the output. The CNN can be train on the large dataset of healthy and unhealthy leaf,following it to learn the differences between the two. Prior to generating the final prediction, multiple convolutional and pooling layers are typically employed in a deep learning model, with the intention of extracting and learning relevantfeatures from the input data. And it can be incorporated intoCNN's layout.

H. Model evaluation

After the model has been trained, its accuracy will be assessed R- Squared metrics can be used for this. This step is important as it helps in understanding the model’s performance and identifying any area of improvement.

IV. RESULT AND DISCUSSION

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 64, 64, 64)	640
conv2d_1 (Conv2D)	(None, 32, 32, 64)	36928
batch_normalization (BatchNormalization)	(None, 32, 32, 64)	256
max_pooling2d (MaxPooling2D)	(None, 11, 11, 64)	0
conv2d_2 (Conv2D)	(None, 6, 6, 128)	73856
conv2d_3 (Conv2D)	(None, 3, 3, 128)	147584
max_pooling2d_1 (MaxPooling2D)	(None, 1, 1, 128)	0
conv2d_4 (Conv2D)	(None, 1, 1, 256)	295168
conv2d_5 (Conv2D)	(None, 1, 1, 256)	590080
max_pooling2d_2 (MaxPooling2D)	(None, 1, 1, 256)	0
flatten (Flatten)	(None, 256)	0
dropout (Dropout)	(None, 256)	0
dense (Dense)	(None, 512)	131584
dense_1 (Dense)	(None, 6)	3078

Total params: 1,279,174
Trainable params: 1,279,046
Non-trainable params: 128

Fig. 3. Model summary

The fig 3, gives the summary of all the parameters and the model parameters. Where total params is 1,279,174, trainable param is 1,279,046 and the non-trainable params is128.

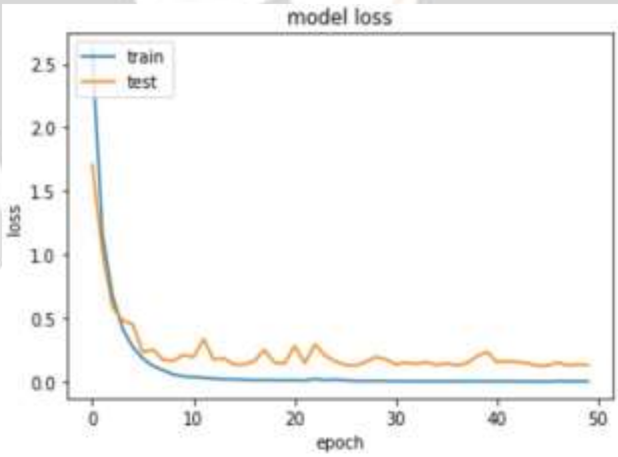

Fig. 4. Graph for Model loss

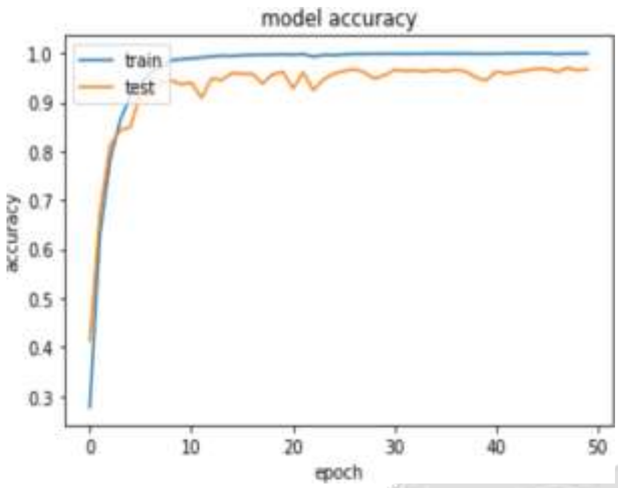

Fig. 5. Graph for Model accuracy

The graphs in Figures 4 and 5 show the accuracy whereby the model is internally evaluated the photos over a period of50 epochs. The data does not suffer from overfitting issues as the model's performance is excellent, accurately predicting the desired outputs. Both the training and testing stages 10 consistently achieved a low model loss of approximately 0.2. Moreover mode’s accuracy is significantly improved, reaching a value close to 0.9. These positive results indicate that the model is effectively based on the data and generalizing well to new instances, resultingin highly accurate predictions.

Fig. 6. Bacterial disease with remedies

Fig. 6 illustrates the user interface of or application. On this screen, users are able to view an image of a tomato leaf image, and receive information regarding the specific disease name that affecting the leaf along with recommended remedies.

.

	precision	recall	f1-score	support
Bacterial_spot	0.89	0.93	0.91	215
healthy	0.98	0.99	0.99	222
Leaf_Mold	0.87	0.83	0.85	176
mosaic_virus	0.94	0.92	0.93	204
Septoria_leaf_spot	0.85	0.87	0.86	189
Yellow_Leaf_Curl_Virus	0.94	0.93	0.94	194
accuracy			0.92	1200
macro avg	0.91	0.91	0.91	1200
weighted avg	0.92	0.92	0.91	1200

Fig. 7. Evaluation metrices

The evaluation metrics for detection of tomato leaf disease are presented in Figure 7. To evaluate the performance and precision of the illness model across different disease classes such as Mosaic Virus, Bacterial Spot, YellowLeaf Curl Virus, Healthy, Leaf Mold, and Septoria Leaf Spot, various measurements including precision, recall, and f1- score are utilized. These measurements give us useful information about different aspects of the model's performance

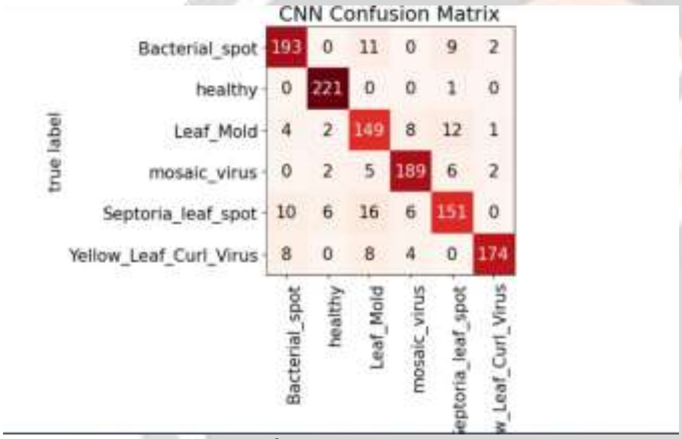

Fig. 8. Confusion matrix

The Fig. 8 matrix represents the classification performance of the CNN model for different disease types, including Bacterial Spot, Healthy, Leaf Mold, Mosaic Virus, Septoria Leaf Spot, and yellow leaf curl virus. In the Confusion Matrix, each disease type is represented by both rows and columns. The values within the matrix indicate the number of instances or observations that fall into specific combinations of predicted and actual disease classes. For instance, the value of 193 in the Bacterial Spot row and column denotes that 193 instances were correctly classified as Bacterial Spot. Similarly, the values of 221, 149, 189, 151, and 174 in the corresponding rows and columns indicate the correct classifications for Healthy, Leaf Mold, Mosaic Virus, Septoria Leaf Spot, and Yellow Leaf curl virus, respectively.

V. CONCLUSION

The use of machine learning techniques in tomato leaf disease detection has the potential to be a powerful tool in managing and controlling these plant pathogens. The abilityof ml algorithms to analyse a collection of data, to identify patterns, and make predictions can be leveraged to improve the efficiency and accuracy of tomato leaf disease detection.

tomato leaf diseases detection is a critical component of tomato crop management and plays major role in protecting the health and productivity of tomato plants. In tomato leaf disease detector, we could see that CNN works best for all types of images with 92.3% accuracy rate. According to the classification report precision of bacterial spot class is 89%,healthy class is 98%, leaf mold is 87, mosaic virus is 94, Septoria class is 85%, and yellow leaf curl virus is 94%. Early detection allows for prompt action to be taken to prevent or limit the spread of disease, minimizing the potential for yield and quality losses. A variety of methods and techniques, including visual inspection, laboratory testing, and technology such as digital imaging and machinelearning, can be employed to detect tomato leaf diseases.

VI. FUTURE WORK

There are several areas of potential future work that could bedone to further improve tomato leaf disease detection: Improved image analysis algorithms: As mentioned earlier, current image analysis algorithms may not be able to accurately identify all types of diseases or may have a high rate of false positives. Developing more advanced image analysis algorithms, such as deep learning-based methods, could improve the accuracy of disease detection. Multi- modal data integration: Incorporating other types of data such as temperature, humidity, and nutrient levels from sensors, can provide additional information that could be used to identify diseases. Combining this information with images using multi-modal data fusion techniques could potentially improve the accuracy of disease detection. Realtime monitoring: Developing a system that can continuously monitor plants in real-time could allow for more timely detection and treatment of diseases. This could be achieved through the use of cameras or other sensors that are placed in the field, and using computer vision algorithms to analyse the data in real-time. Automated treatment Developing systems that can automatically apply treatments such as pesticides or nutrient solutions to infected plants based on the disease detection output could help to reduce the time and labour required to manage diseases.

REFERENCES

[1] Prajwala T M, et al. Tomato Leaf Disease Detection usingConvolutional Neural Networks, Proceedings of 2018Eleventh International Conference on ContemporaryComputing, 2-4 August, 2020, Noida India.

[2] Akshay Kumar, et al. Image Based Tomato Leaf DiseaseDetection, 10th ICCCNT, 68 July, 2019, IIT-Kanpur, Kanpur, India.

[3] Rangarajan, A.K., Purushothaman, R., Ramesh, A., 2018.Tomato crop disease classification using pre-trained deep learning algorithm, Procedia computer science 133, 1040– 1047.

[4] Shijie, Jia, Jia Peiyi, and Hu Siping. "Automatic detection of tomato diseases and pests based on leaf images." 2017 Chinese Automation Congress (CAC). IEEE, 2017.

[5]de Luna, R.G., Dadios, E.P. and Bandala, A.A., 2018, October. Automated image capturing system for deep learning-based tomato plantleaf disease detection and recognition. In TENCON 2018-2019 IEEE Region 10Conference (pp. 1414-1419). IEEE.

[6] Sardogan, Melike, Adem Tuncer, and Yunus Ozen. "Plantleaf disease detection and classification based on CNN with LVQ algorithm." 2019 3rd International Conference on Computer Science and Engineering(UBMK). IEEE, 2018.

[7] Dheeb Al Bashish, Malik Braik and Sulieman Bani- Ahmad "A Framework for Detection and Classification of Plant Leaf and Stem Diseases" 2019 IEEE International Conference on Signal and deep learning, pp. 978-1-4244- 8594-9/10

[8] Chowdhury, Muhammad EH, et al. "Automatic and reliable leaf disease detection using deep learning techniques." Agri Engineering 3.2 (2021): 294-312.

[9] Sabrol, H., and K. Satish. "Tomato plant disease classification in digital images using classification tree."2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2016.

[10] Adhikari, Santosh, et al. "Tomato plant diseases detection system using image processing." 1st KEC Conference on Engineering and Technology, Lalitpur. Vol. 1.2021.

[11] Sladojevic, Srdjan, et al. "Deep neural networks based recognition ofplant diseases by leaf image classification." Computational intelligence and neuroscience 2016 (2016).

[12] Ashqar, Belal AM, and Samy S. Abu-Naser. "Image□Based Tomato Leaves Diseases Detection Using Deep Learning." (2018).

[13] Zhang, Keke, et al. "Can deep learning identify tomato leaf disease?. “Advances in Multimedia 2018 (2018).

[14] Kulkarni, Anand H., and Ashwin Patil. "Applying image processing technique to detect plant diseases." International Journal of Modern Engineering Research 2.5 (2012): 3661- 3664.