© May 2025 | IJIRT | Volume 11 Issue 12 | ISSN: 2349-6002

An Approach Early Disease Detection in Plants using CNN

Pranali Bawankule[1], Vruchika Kamble[2], Vishakha Satishendre[3], Rasna Meshram[4], Shubhangi Shende[5], Jayant Rajurkar[6]
[1,2,3,4,5]*Student, Manoharbhai Patel institute of Engineering and Technology Shahapur,Bhandara*
[6]*Assistant Professor, Manoharbhai Patel institute of Engineering and Technology Shahapur, Bhandara*

***Abstract*—The Indian economy benefits from early disease detection in plant leaves. According to reports, 10-30% of crops suffer harm from diseases that are not discovered during the curing process. Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have paved the way for transformative solutions in disease identification for agriculture. This study addresses the critical issue of early disease detection by harnessing the power of deep learning models, specifically CNNs. Different leaf disease detection technologies are used for different crops. The pre-trained deep learning model is used in this study to identify and categorize leaf diseases. A dataset of tomato, potato, and bell pepper leaf pictures from the plant village repository was employed for the current investigation. The developed model can detect 12 plant diseases in normal leaf tissue. The quantitative assessment of our CNN-based technique reveals an impressive accuracy rate of 86.21%. This notable accuracy underscores the efficacy of our approach in the challenging domain of plant disease detection. Our findings show potential for the larger agricultural environment beyond the immediate quantitative advantages. This technical advance not only paves the way for improved crop output and lower losses, but also brings in a new era of data-driven sustainability in agriculture. The bottom line is that our research "offers a tangible pathway for leveraging AI-powered solutions to address the long-standing challenges of plant disease detection, thereby significantly contributing to the well-being of farmers and the sustenance of the Indian agricultural sector.**

***Index Terms*—Deep Learning; classification; CNN; leaf disease detection.**

I. INTRODUCTION

India is a developing nation where agriculture supports around 70% of the people. Plant disease results in a significant decline in the quantity and quality of agricultural goods. Therefore, to ensure food security, both early disease detection and the development of better crop varieties is essential.

The traditional technique of disease identification has been manual examination by farmers or professionals, but this may be time-consuming and expensive, making it impracticable for millions of small and medium-sized farms worldwide. The development of automated systems for plant disease identification by checking for telltale signals on leaves has been sped up by the construction of computer vision models. These methods attempt to make the detecting process as straightforward and error-free for farmers as possible.

Studies have concentrated on employing image processing and feature extraction to build methods for disease detection. In order to recognise and categorise plant diseases, numerous machine learning mo dels have been used. Since deep learning (DL) models have been developed, this area of research currently seems to offer a lot of potential for increased accuracy. Numerous customised DL architectures are used with various visualisation approaches to detect and classify the symptoms of plant diseases. In addition, various performance metrics are used to evaluate these frameworks and procedures. Teaching a neural network to understand patterns and symptoms shared by various ailments is one step in the deep learning process for detecting plant diseases. The deep learning model gains the ability to distinguish between healthy and unhealthy plants by examining plant photos, making automated detection possible. By doing away with manual inspection, this method yields quicker and more precise results. Convolutional neural networks and other deep learning algorithms extract pertinent elements from the photos, allowing the model to categorise the state of the plant's health. By enabling early identification and intervention, this technology has the potential to revolutionise agricultural practices, resulting in higher crop yields and decreased economic losses.

II. LITERATURE SURVEY

A preliminary literature survey was conducted for plant disease detection in which existing research and studies related to the detection and diagnosis of plant diseases were reviewed. The survey aimed to gather information on the methods, techniques, and algorithms used by researchers in this field. Researchers have used various techniques, including spectral data for hyperspectral imaging, machine learning algorithms like support vector machines (SVM) and deep learning models like CNNs, and image processing techniques like segmentation and feature extraction. These techniques were used to precisely diagnose and categorize illnesses by analysing spectral data or plant image analysis.

Saleem et al. (2019) [2] proposed plant disease detection and classification using deep learning methods. It was a pioneering effort that used hyperspectral imaging to identify plant disease automatically. The experimental study aimed to detect Tomato Spotted Wilt Virus (TSWV) infection in capsicum plants. After inoculation, a routine automated system was developed to collect hyperspectral hypercubes of plant leaves in the VNIR and SWIRspectrum regions. Additionally, the technology was used to collect pertinent meta-data about the plants. After processing the data, it was possible to see how effectively plant leaves could be automatically differentiated using machine learning and image processing techniques. The results showed an accuracy of over 90%. They took advantage of the VNIR (400– 1000 nm) and SWIR (900–2500 nm) portions of the electromagnetic spectrum.

Durmuş et al. (2017) [3] proposed disease detection on the leaves of tomato plants by using deep learning. The dataset of tomato leaf images was downloaded from the Plant Village Dataset, and the proposed research was based on the six different types of tomato leaf diseases. The error-correcting output code was utilized to train a deep convolutional neural network model, which was employed for detection and classification. The classification results were computed using parameters including accuracy, specificity, sensitivity, F-Score, and true negative rate.

The trained model was compared to the basis paper and had a 98.8% accuracy rate. Sharma et al. (20 proposed a performance analysis of deep-learning CNN models for plant disease detection using image segmentation. This study has demonstrated the viability of training a CNN model instead of whole images using segmented and annotated images. Model performance on independent data increases from 42.3% to 98.6% when the same CNN model was trained using segmented pictures (SCNN) instead of full images (F-CNN). With 82% of the test dataset displaying an increase in confidence, quantitative examination of self-classification confidence demonstrated a considerable improvement. Before CNN model training, picture preprocessing has proven invaluable in achieving excellent real-world performance.20) [4] khirade et al. (2015) [5] proposed an approach to Plant disease detection using image processing. Various novel approaches to quickly identify and categorise any illness are essential for the successful production of crops and for ensuring their sustainability. Image Processing is one of the modern technologies that paved the path for this. This study investigated the segmentation of the sick plant components utilising several techniques.

This study also examined the methods for feature extraction and classification and the classification of plant diseases. Plant diseases can be effectively classified using SVMs, backpropagation methods, self-organizing feature maps, and other ANN techniques. These techniques enable accurate diagnosis and classification of numerous plant diseases using image processing tools. Sood et al. (2020) [6] describe a hybrid system for detecting and classifying plant disease using qualitative texture features analysis. In this work, the diseased affected capsicum piece is first subdivided, and then the features of the infected region are extracted using feature extraction. The classification of the capsicum illnesses is completed using an SVM classifier. In order to distinguish between healthy and diseased capsicum and its leaves, the proposed solution is tested for the five illnesses: anthracnose, bacteria processing, training and testing data, and assess ent through metrics. The Grey Level Co-occurrence Matrix was utilised for feature extraction, the K-means approach was employed for image segmentation, and a multi-class support vector machine algorithm with a linear kernel function was used to classify potato leaves. The proposed framework was assessed using specific assessment criteria, such as precision, recall, F1-score, and accuracy. The proposed model's F1-score is approximately 96.16%, its precision is roughly 96.12%, its recall is approximately 96.25%, and its total accuracy is approximately 95.99%.

III. PROPOSED METHODOLOGY

Figure 1 shows the entire process of the proposed model. The three main steps of the technique are the data preprocessing stage, model training using the training set, and prediction using the test set. The following subsections cover the specifics of each phase.

3.1. Data Preprocessing Step

The data preprocessing step involved the normalization of images.

Image normalisation is a preprocessing procedure used in the dataset for deep-learning plant disease detection to guarantee that all images have uniform sizes and ranges of pixel values. This stage is crucial for boosting the model's convergence during training and for enhancing its generalisation abilities.

Image normalisation frequently involves the following steps:

Rescaling: Images are scaled or resized to 224x224 pixels. Usually, this is done to ensure that all of the images are the same size so that feeding them to the neural network is simpler.

Pixel Value Normalisation: The image's pixel values are altered to fall between a range of values, frequently between 0 and 1 or -1 and 1. This normalisation aids in minimising the effects of

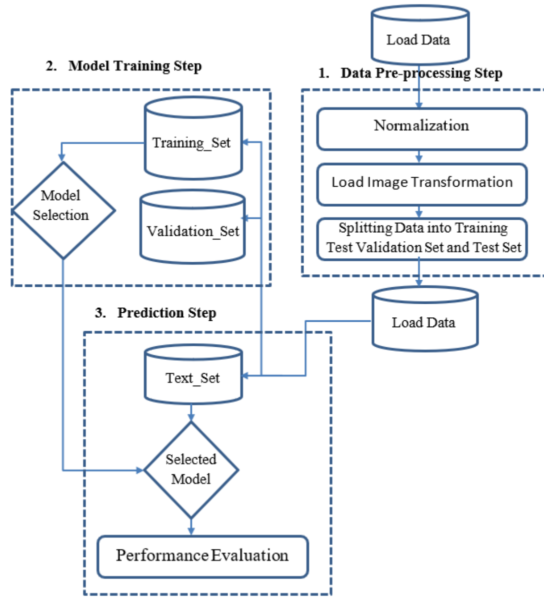

Fig.3.1.1 Overall Proposed Methodology

variations in pixel intensity and enhances the neural network's convergence during training.

Mean Subtraction: Each pixel has its mean pixel value from which the entire dataset is subtracted. This process prevents neural network activation functions from becoming saturated and centres the data distribution around zero.

Standardisation: The pixel values are further scaled by dividing the pixel values by the dataset's standard deviation. The unit variance of the data is ensured using this procedure, which also speeds up neural network learning.

The dataset is divided into three sets: the training set, the validation set, and the test set once it has been normalised.

Training Set: Most of the dataset, known as the training set, is used to train the deep learning model. Through forward and backward passes throughout the training phase. It is used to update the model's parameters.

Validation Set: Using the validation set, the validation performance of the model is assessed, and its hyper parameters are modified. As it enables tracking of the model's performance on data it hasn't seen during training, it is crucial for preventing overfitting.

Test Set: The test set is a distinct dataset that is only utilized following the conclusion of model training. It measures the trained model's ultimate performance and ability to generalize to new data. Data augmentation generates new images that are altered versions of our input data by applying image modification, contrast adjustments, and image blurring without affecting the labels on the original photos. The training set images were augmented for the model's robustness. The data augmentation steps are depicted in Figure 2.

Even the issue of over-adjustment is also decreased by the expansion of the entire “train”. To boost data, the “Keras Image data Generator” class was used, and the following options produced the best results:

(a) Rotation range: 20
(b) Width and height shift ranges: 0.2
(c) Horizontal flip: True
(d) Vertical mode: True

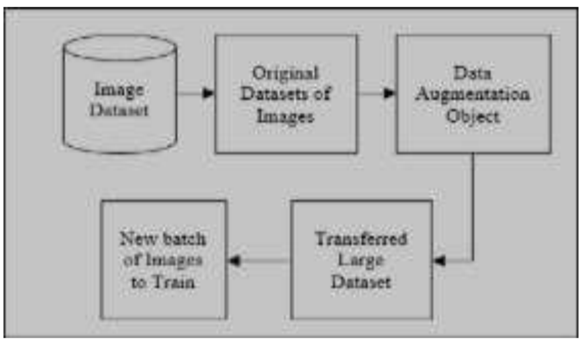

Fig.3.1.2 Data Augmentation Steps.

3.2. Proposed Model Architecture

The architecture of CNN, which enables the effective processing of image data, is depicted in Figure 3.2.1 A deep CNN architecture is made up of numerous layers of various kinds. It is the most often used technique for sifting through large datasets and extracting useful information

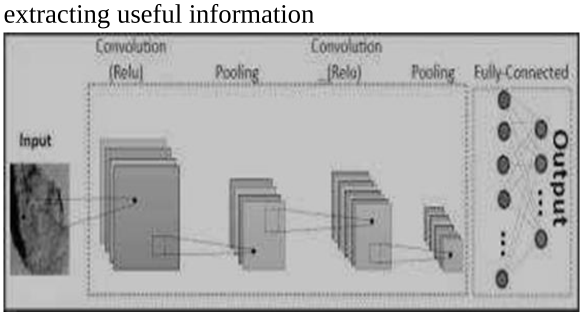

Fig.3.2.1 Convolution Neural Network (CNN) Architecture

Convolutional layers:Convolutional layers make up the first few layers of the CNN. Convolutional layers perform feature extraction by applying filters to the input image. These filters learn spatial patterns and local features present in the image. From one block to another, the number of filters gradually increases, but the size of the filter remains fixed at 3 x 3. There are 32 filters in the first convolutional block, 64 in the second, and 128 in the third. The size of the feature maps was lowered due to the use of pooling layers in each block, necessitating an increase in the number of filters.

Convolutional layers are particularly effective in capturing hierarchical representations of images, allowing the model to learn discriminative features at different scales. Each convolutional layer applies a set of filters (kernels) to the input image.

Pooling Layer:A max pooling layer with a pool size of 2x2 is added after each convolutional layer. The feature maps created by the convolutional layers are downsampled using pooling layers. Doing this makes it possible to keep key features while reducing the feature maps' spatial dimensions. By condensing local data, pooling layers simplify computation and increase the model's spatial invariance.

Flattened Layer: The flattened layer creates a 1-dimensional vector from the output of the preceding levels. This flattening phase is essential before transferring the data into a fully connected layer.This process does not affect the underlying data values but restructures their layout to suit the input requirements of fully connected layers. The usage of the Flatten layer is essential in scenarios where high-level, abstracted features extracted from images or signals need to be aggregated into a format suitable for decision-making layers within a neural network.

Fully Connected Layers: The CNN architecture's fully connected layers, which are found near the bottom, are in charge of categorising the learned features into particular illness categories. Fully connected layers capture global relationships between features and enable the model to make predictions based on the extracted representations. The first dense layer consists of 64 neurons with the ReLU activation function. It captures global relationships between features.

IV. DATASET AND IMPLEMENTATION DETAILS

4.1 Dataset Creation

The 'Plant Village Dataset' is the source of all Potato, Tomato, and Pepper Bell pictures. The 20,639 leaf images in the Plant Village dataset are categorised into 15 groups based on species and illness, as shown in Table 1. The leaf pictures dataset includes both healthy and ill leaves that have been impacted by various biotic causes. Images of every significant type of leaf disease are included in this dataset. These vegetables are part of the staple diet in many Indian households. Thus, plant disease detection in such plants was important. We build the model using a subset of 70% and test it using a subset of 30%. Examples of the "Train" folder images from the adopted distribution are included in Table 2.

Table 2. Images from Dataset

Pepper Bell Healthy Leaf	
Pepper Bell Bacterial Spot	
Tomato Healthy L	
TomatoLeaf Mold	
TomatoEarly Blight	
Potato Healthy Leaf	
Potato Early Blight	
Potato Late Blight	

V. RESULTS AND DISCUSSION

The suggested convolution neural network achieved an overall testing accuracy of 86.21% using data augmentation and transfer learning. The improvement in accuracy and loss over time from epoch 13 to 20 is seen in Table III. The accuracy improves, and the loss is significantly reduced with each epoch. The results have been reported for 20 epochs due to system limitations. Although the loss value continues to drop significantly after iteration 10, there has been no appreciable gain in our model's accuracy.

Furthermore, table 3 demonstrates that as more iterations are performed, the value of loss and accuracy gets closer to being constant with different epochs.over 20 training epochs, the maximum validation accuracy of 74.94% was recorded, while the best training accuracy of 86.21% was also noted. A validation accuracy of 86% on average has been found. This is a helpful measure of how accurately the Deep Learning algorithm has classified various objects. A method of visualising and indicating the rate of model convergence is provided by the plots of train and test accuracy and loss against the epochs in Fig. 4. It is clear that the model has stabilised after about 20 epochs, and the metrics do not indicate that the model has much improved in the last ten epochs. In Figure 5, The images are given as input to the trained model, and the predicted labels match with the actual label. The outcomes demonstrate the model's effectiveness on the dataset and its applicability for classifying leaf diseases.

The proposed CNN model architecture leverages the hierarchical and translation-invariant nature of CNNs to extract meaningful features from plant disease images. The model can effectively detect and classify plant diseases by combining convolutional, pooling, and fully connected layers. However, further experimentation and optimization may be necessary to finetune the architecture for specific datasets and disease detection tasks.

Fig.5.1 Prediction of diseases using CNN model.the label category denotes the predicted disease of the respective plant

VI. CONCLUSION

The majority of the Indian people continue to rely heavily on the agricultural industry, which is still one of the most significant sectors. Therefore, spotting diseases in these crops is essential to expanding the economy. Therefore, this research aims to discover and characterize distinct illnesses in the tomato, potato and pepper bell crops. Using a Convolutional Neural Network model, the suggested method classifies tomato leaf diseases, potato leaf diseases, and pepper bell leaf diseases from the Plant Village dataset. To categorize tomato, potato and pepper bell leaf illnesses into different classes, a straightforward Convolutional Neural Network with a minimal number of layers was utilized as the architecture. Future Scope: Future work may also involve experimenting with the suggested model using various learning rates and optimizers. It can also involve experiments with more modern architectures to improve the model's functionality on the train set. As a result, farmers can utilize the aforementioned model as a decision-making tool to assist and support them in recognizing diseases that can affect tomato, potato and pepper bell plants.

REFERENCES

[1] Department of Agriculture & Farmers Welfare, 2023. Agricultural Statistics at a Glance. [Online] Available at: http://agricoop.nic.in/en/Agricultural_Statistics_at_a_Glance [Accessed 7 August 2023].

[2] Saleem, M.H., Potgieter, J. and Arif, K.M., 2019. Plant disease detection and classification by deep learning. Plants, 8(11), p.468.

[3] Durmuş, H.W., Güneş, E.O. and Kırcı, M., 2017. Disease detection on the leaves of tomato plants using deep learning. In: Proceedings of the 6th International Conference on Agro-Geoinformatics, IEEE, pp. 1-5, August.

[4] Berwal, Y.P.S. and Ghai, W., 2020. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. Information Processing in Agriculture, 7(4), pp. 566-574.

[5] Khirade, S.D. and Patil, A.B., 2015. Plant disease detection using image processing. In: Proceedings of the 2015 International Conference on Computing Communication Control and Automation, IEEE, pp. 768-771, February.

[6] M., Singh, P.K., 2020. Hybrid system for detection and classification of plant diseases using qualitative texture features analysis. Procedia Computer Science, 167, pp. 1056-1065.

[7] Nandhini, S.A., Hemalatha, R. and Radha, S., 2019. IoT-enabled efficient detection and classification of plant diseases for agricultural applications. In: Proceedings of the 2019 International Conference on Wireless Communications, Signal Processing, and Networking (WiSPNET), IEEE, pp. 447-451, March.

[8] Singh, A. and Kaur, H., 2021. Potato plant leaf disease detection and classification using machine learning methodologies. IOP Conference Series: Materials Science and Engineering, 1022(1), p. 012121.

[9] Singh, A. and Kaur, H., 2021. Potato plant leaf disease detection and classification using machine learning methodologies. IOP Conference Series: Materials Science and Engineering, 1022(1), p. 012121.

[10] Parikh, A, Raval, M.S., Parmar, C. and Chaudhary, S., 2016. Disease detection and severity estimation in cotton plants from unconstrained images. In: Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE, pp. 594- 601, October.