International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Plant Disease Detection Using ImageProcessing and Machine Learning

SANDEEP R
Department of Artificial Intelligence and Data Science,
Bannari Amman Institute of Technology,
Sathyamangalam, India
sandeep.ad20@bitsathy.ac.in

SHANGAMESHWAR K
Department of Artificial Intelligence and Data Science,
Bannari Amman Institute of Technology,
Sathyamangalam, India
shangameshwar.ad20@bitsathy.ac.in

GOWTHAM S
Department of Artificial Intelligence and Data Science,
Bannari Amman Institute of Technology,
Sathyamangalam, India
gowtham.ad20@bitsathy.ac.in

***Abstract*—** Agriculture acknowledges a fundamental part by temperance of the fast improvement of everybody and expanded interest in food in India. Consequently, it is expected to increase gather yield. One serious reason for low gather yield is contamination achieved by microorganisms, disease, and creatures. Plant disease examination is one of the major and fundamental errands in the piece of developing. It will in general be thwarted by using plant illness identification strategies. To screen, notice or deal with plant infections physically is an exceptionally complicated task. It requires immense extents of work, and needs absurd arranging time; thus, picture handling is used to recognize sicknesses of plants. Plant infection grouping should be possible by utilizing AI calculations which incorporate advances like dataset creation, load pictures, pre-getting ready, division, highlight extraction, preparing classifier, and arrangement. The primary target of this exploration is to build one model, which groups the solid and infected reap leaves and predicts disease of plants. In this paper, the researchers have trained a model to recognize some unique harvests and 38 diseases from the public dataset which contains 70,306 images of the diseases and healthy plant leaves that are collected under controlled conditions. This paper worked on the ResNets algorithm.

Keywords—Plant leaf disease, Deep Learning, CNN algorithm

I. INTRODUCTION

Agriculture is the mother of all civilizations. The focus is on improving productivity without considering the environmental impact caused by environmental degradation. Plant diseases are of great concern in agricultural development as they can affect both crop quality and quantity. In general, plant diseases include fungi, bacteria, viruses and molds. A farmer or a professional usually detects and diagnoses plant diseases with the naked eye. However, this approach can be time-consuming, expensive, and inaccurate, so using deep learning techniques to detect and classify plant diseases is a fast and accurate method. Photographs of plant infections are used for plant disease detection, research, education, and analysis. The use of computer image processing and deep learning technology aims to enable fast and accurate detection. Studies have shown that deep learning techniques are an effective way to classify plant diseases. A major concern is to improve the reliability, accuracy, and precision of image analysis for plant disease detection and classification. An automated plant disease diagnosis system based on the presence and visible signs of plants is very useful not only for growing learners, but also for qualified professionals as a system for confirming infection diagnoses. Researchers used visualization techniques to extract plant disease representations from trained CNNs. Every year many researchers deal with the growing part of machine vision and computer vision. In this study, we propose a system to detect and classify plant diseases using machine learning techniques. A study based on our deep learning and CNN. The data set that was taken from a global data set which is (New Plant Diseases Dataset (Augmented)) includes a number of the plants.

II. WORK

This section describes the different systems for detecting illness in plant leaves using deep learning techniques. The main objective of this research is to discover a solution to the difficulty of detecting tomato leaves disease using the easiest method while utilizing minimal computational resources to produce outcomes that are comparable to state-of the-art methods. Neural network systems use an automated extraction of features to help differentiate the input image into several classes of illness. This system has obtained an overall accuracy of 94-95 %, showing even under unfavorable conditions the viability of the neural network approach. CNN is applied to classify tomato plant leaf images dependent on the obvious impacts of diseases. In addition to transfer learning as a compelling methodology, training a CNN from scratch utilizing the Deep Residual Learning strategy tests. To do that, design of CNN is proposed and applied to a subset of the New Plant Diseases Dataset, including plant leaf picture. The outcomes show that the proposed design outflanks VGG models, pre-trained on the Image Net dataset, in both exactness and the time required for re-training, and it tends to be utilized with a standard PC.

There will be detection here only for diseases and not a classification of plant types. The unhealthy leaf image is caught. HSV features are derived because of the color segmentation. Artificial Neural Network, (ANN) is then trained to identify samples that are sick and healthy. The performance of the ANN classification is 80 % better than others approach in accuracy. As for the loophole of this work, it is a little accuracy, which is 80%. This accuracy, using the neural network, is few and contains some detection errors for plant leaf diseases. Kawasaki et al. suggested the CNNs method that is used to differentiate nutritious cucumbers from diseased cucumbers by utilizing photos of leaves. Three convolutional layers, pooling layers, and normalization layers. The activation method used in this system is the function Rectified Linear Unit (ReLU). The precision obtained in this analysis is 94.9 %.

III. The Proposed System

In this part, we present a computer vision system, for the systematic identification of leaf diseases. The proposed system consists of several steps. Figure 1 shows the main structure of the proposed system.

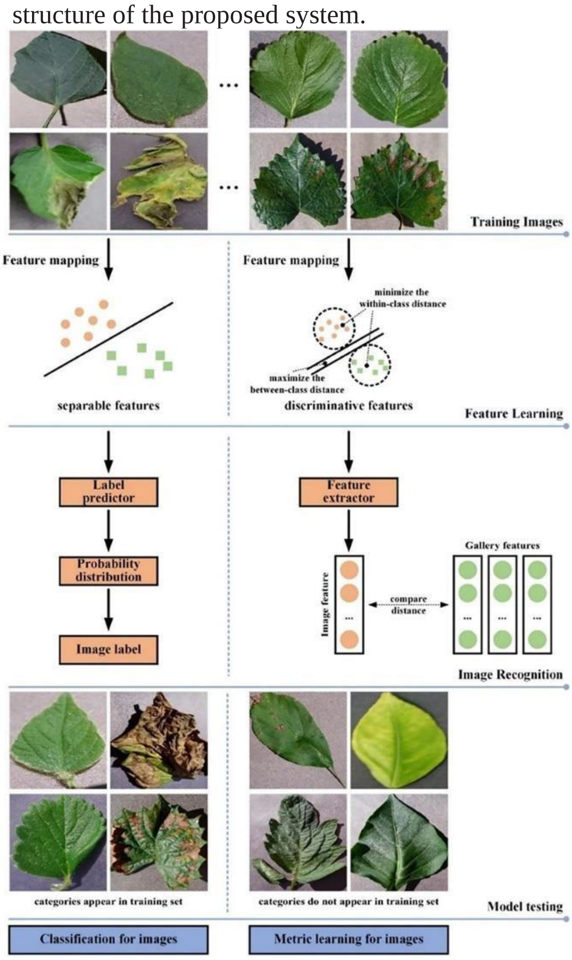

Fig. 1. General Diagram for Plant Leaf Diseases Detection and Classification System

As in Fig. 1, which shows the construction of our proposed system for detection and classification of plant leaf diseases, the first stage in it is the stage of image acquisition.

A. *Image acquisition (Dataset)*

The plant leaves disease pictures have been obtained from the (New plant diseases) treasury from website (kaggle). The collected dataset comprises of around 70436 images relating to 38 different classes as shown in Fig. 2 The dataset includes images of all the major foliar diseases that can affect three crops: tomato, pepper, and potato. They were chosen because they are one of the most famous plants in the world at large and in the world. Each uploaded image defaults to an RGB color space and is saved as an uncompressed JPG.

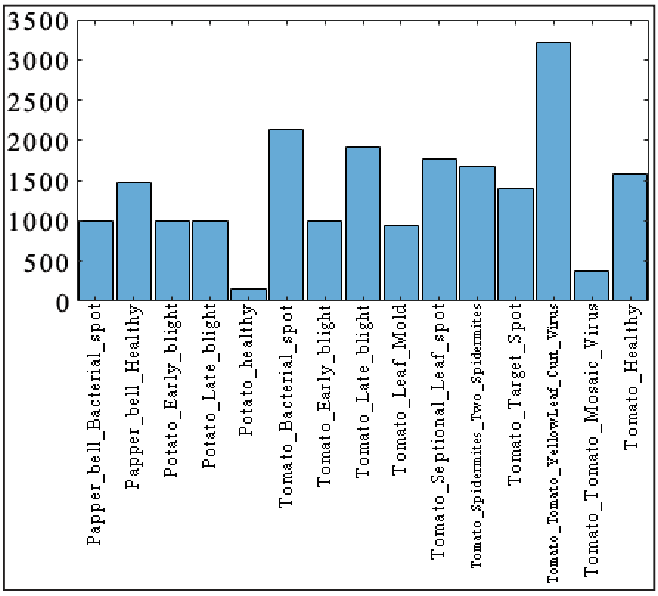

Fig. 2. Crops and their Disease that been used as a Dataset

B. *Image Pre-processing*

To speed up the training process and get the actual model-check computation, the dataset image is scaled to 128 x 128 resolution. Optimization of the input or target variables helps to increase training processing speed. It also preserves the integrity of the image data from loss.

C. *CNN Structure Design*

CNN is a very popular approach in deep learning in which multiple layers are robustly trained. It has been observed to be highly effective and is additionally the most widely used in various applications of computer vision. CNN can be applied to construct a computational form that operates on unorganized image inputs and transforms them into the correct output categories for classification. In our work, a structure has been built for this algorithm. This structure is made up of several layers as shown in Fig. 3, which illustrates the architecture that we used to construct the CNN.

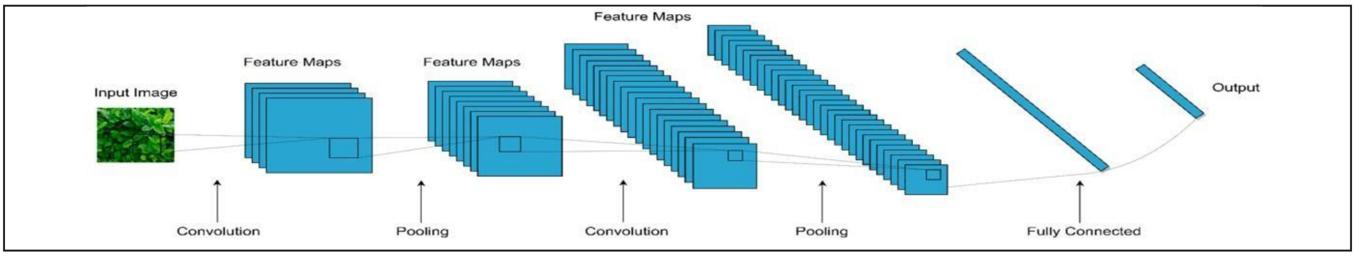

Fig. 3. CNN Model Architecture

1. Input Layer

The input layer contains the input images and their pixel values.

2. Convolution Layer

A CNN uses different kernels, in the convolution layers to convert the entire object as well as the optimal feature maps, creating different feature maps. In our research, we have used 4 convolution layers, in the first convolutional layer, we used 10 filters that have a height and width of 3 with padding character arrays 'same'. The padding applied to the input along the edges. The 'same' means Padding is set so that the output size is the same as the input size. In the second convolutional layer, we used 20 filters that have a height and width of 3 with padding character arrays 'same'. In the third convolutional layer, we used 64 filters that have a height and width of 3 with Padding character array 'same' and In the fourth convolutional layer, we used 30 filters that have a height and width of 3 with Padding character array 'same' as shown in Fig. 3.

3. Pooling Layer

A pooling layer often follows a convolutional layer and can be utilized to depreciate the dimensions of feature maps and parameters of the network. Pooling layers are also invariant in interpretation, alike to convolutional layers because their calculations consider neighboring pixels. The most widely used approaches are average pooling and max pooling. In our research, we used a max-pooling layer.

4. Non-Linear Layer

A non-linear transformation is applied to the input by the CNN, the object of which is to classify the features within per hidden layer. In CNN structure we use Rectified linear units (ReLU). Rectified linear units are commonly used as non-linear transformation. This kind of layer executes a simple operation with a threshold where any input value smaller than zero is set to zero.

5. Fully Connected Layer

The data arrives at the last layer of the CNN, which is the fully connected node, later much iteration of the prior layers. In the two neighbouring layers, the neurons are connected directly to the neurons within the fully connected network as shown in Fig. 3.

6. Normalize Layer

In our proposed system we use a batch normalize layer. Batch normalization layer form normalizes any channel through a mini-batch. This can help to decrease sensitivity to data variations.

7. Softmax Layer

The network's performance can be difficult to interpret. It is normal to finish the CNN with a softmax function in classification issues. After extracting values of 15 classes of plant diseases in the fully connected step, a Softmax will be made for them, so that the class will be selected in each process and according to the features that were extracted through the previous layers that the images of plant diseases went through it. In this layer, the correct class of disease is determined by applying the Softmax function.

D. Training

Training a network is a procedure of obtaining kernels in convolution layers and weights in fully connected layers that reduce differences on a training dataset between output predictions and specified ground truth labels. In our work, we used 70% of the data for training, through this stage so that the network that has been built learns by extracting features from plant leaf disease images in order to learn from these features for each image to be distinguished on its basis

E. Testing

The testing is a dataset utilized to provide an impartial final design fit evaluation on the training set of data. In this stage, we use the groups that were trained in the previous step that was trained in CNN, and the features were extracted by learning the network when the data set passes from plant leaf diseases on this network, we used 70% of the data for testing.

F. Detection for Plant Leaf Diseases

After the previous operations, plant species diseases are detected and classified according to three types of plants, namely tomatoes, peppers, and potatoes. All results, detection and classification will be presented in the next section.

III. Experiment Results

The experiments are performed on collab, RAM 8 GB and processor core(TM) i5_12500H CPU @ 2.50 GHz. We train the network and save the trained network so that the training process is not repeated and so the time taken for training is provided, after which the network is also tested by testing data and showing the accuracy. In the last stage, through which the diseases of plant leaves are detected and classified, a random selection is made for any image, here we have chosen an example of the potato plant paper after loading it to the system and choosing the network that was trained and press the Detection button, the disease and the type of the affected plant will be revealed as in the Fig. 4 that illustrates the detection and classification process using the CNN algorithm for Late blight disease, to the potato plant leaf.

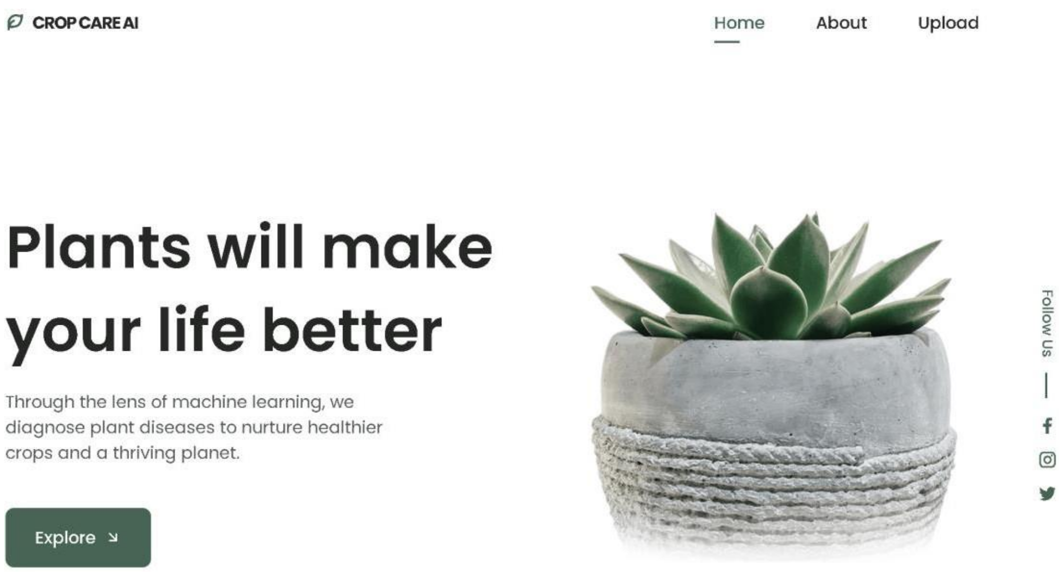

Fig. 4. Completed product

The process of calculating the accuracy was performed to be displayed later in the confusion matrix according to TABLE I and similar to “(1)” and “(2)” through which the accuracy is calculated for each class and also the total accuracy is calculated for all classes.

TABLE I. CONFUSION MATRIX TABLE

		Predicated	
		Positive	Negative
Actual	Positive	a	b
	Negative	c	d

$$\text{Accurcy} = \frac{a + d}{a + b + c + d} \qquad (1)$$

$$\text{Accurcy} = \frac{\text{number of correctly claSSified imageS}}{\text{Total number of imageS}} \times 100\% \qquad (2)$$

The reason for the difference in accuracy between the classes, despite the close proximity of this accuracy, is due to the data set, the difference in these images, the difference in the capture process, as well as the lighting, as well as the difference in the number of pictures for each class in the data set. This greatly affects the accuracy of each class.

In the network training process as in Fig. 5, which illustrates the training process and also shows the loss function that occurred during the training of the network and also explains the training options used, such as epochs and iterations, as in the following figure.

In addition, we divided the training data set we used in our work 15 epochs and for each epoch, 112 iterations and the number of iterations was 1680 divided by 15 epochs. As shown in Fig. 5.

In the classification with the CNN algorithm, we obtained a high accuracy of the performance was shown in the confusion matrix for each class that was classified in training data. At the top of the matrix, rate the overall performance accuracy of the training data which (98.29) appears as in Fig. 6. Also the rate the overall performance accuracy of the test data that we obtained was calculated (98,028) and the accuracy of each class in the testing process as in Fig. 7 that shows the confusion matrix for the accuracy of the performance of the test data.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

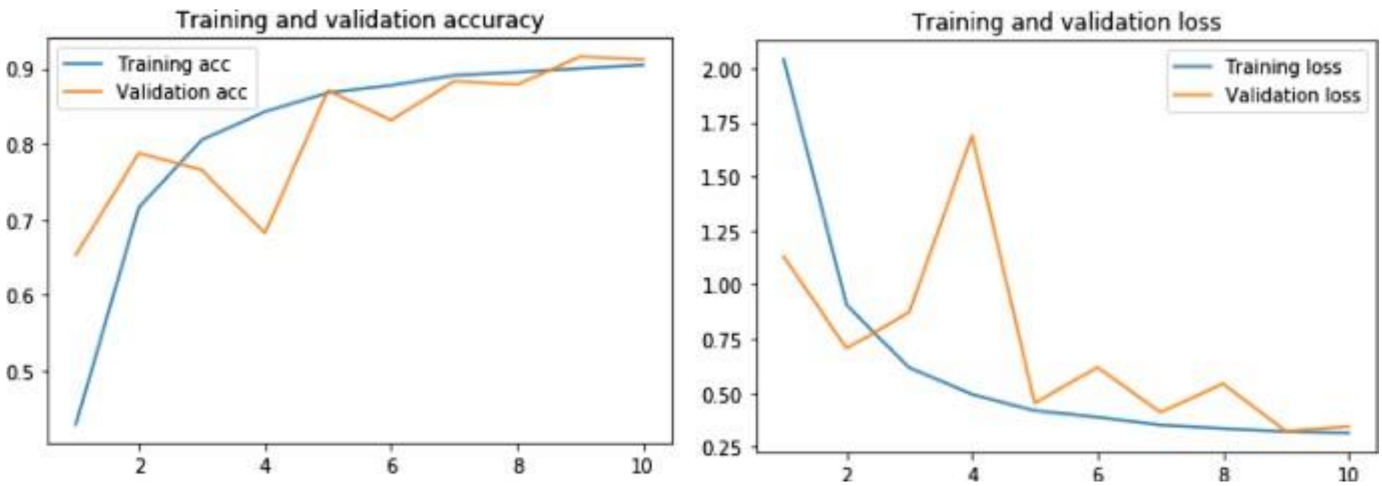

Fig.5. Training Progress (Accuracy) and Loss Function

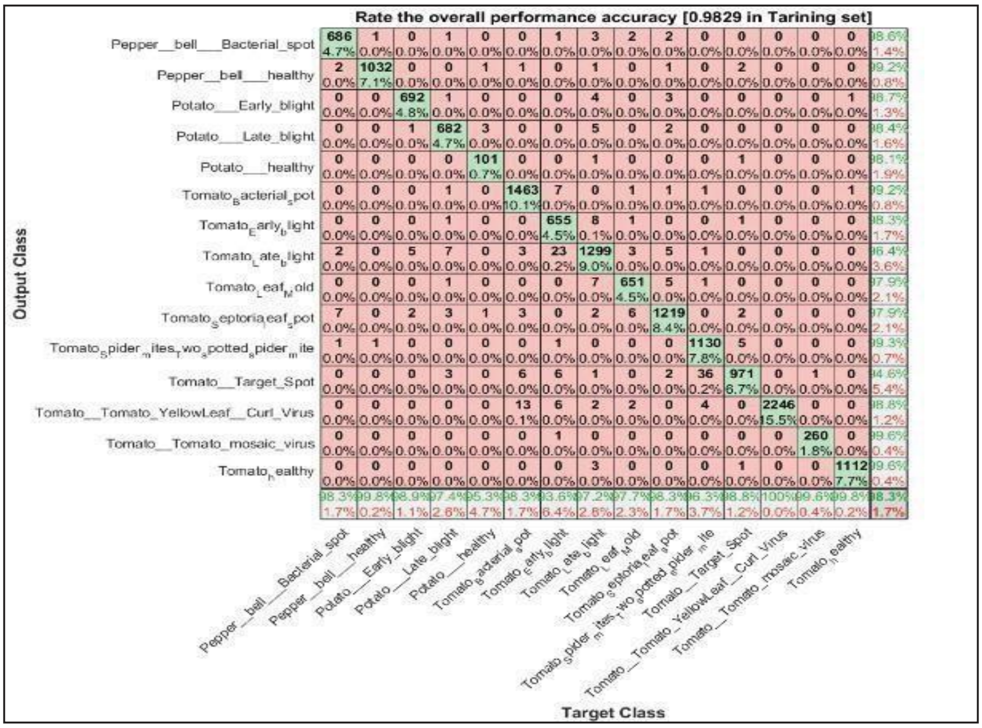

Fig. 6 performance accuracy of the training data

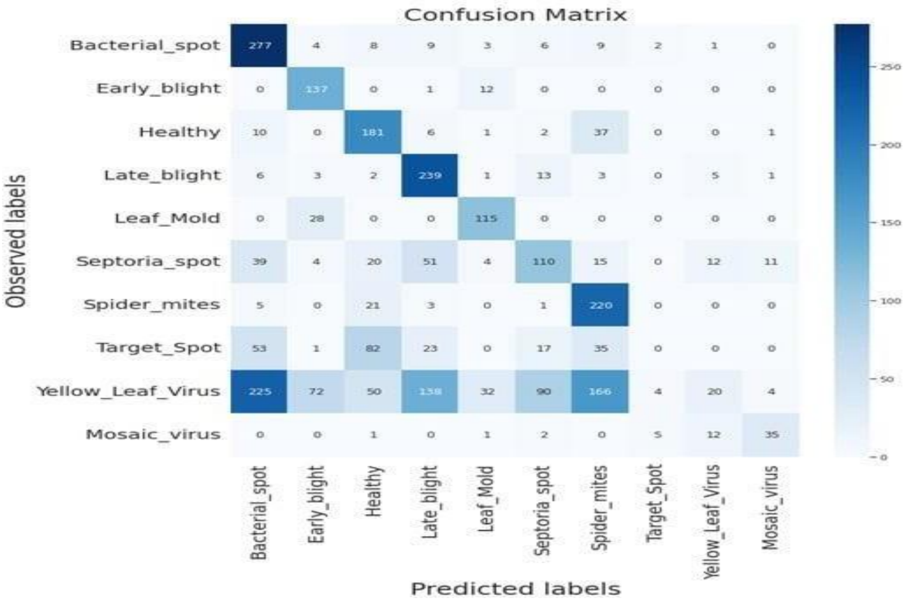

Fig.7. Confusion Matrix for CNN Performance Accuracy for Testing Data

TABLE II. ACCURACY OF EACH CLASS WITH THE OVERALL ACCURACY OF THE TRAINING AND TESTING STAGES

No	Class Name (Plant Diseases)	Accuracy in CCN Training	Accuracy in CCN Testing
1	Pepper bell Bacterial spot	98.60	97.60
2	Pepper bell healthy	99.20	99.30
3	Potato Early blight	98.70	99.30
4	Potato healthy	98.10	93.30
5	Potato Late blight	98.40	99.00
6	Tomato Target Spot	94.60	93.90
7	Tomato_mosaic_virus	99.60	99,10
8	Tomato Yellow Leaf_ Curl Virus	98.80	98.40
9	Tomato Bacterial spot	99.20	99.40
10	Tomato Early blight	98.30	96.90
11	Tomato healthy	99.60	99.60
12	Tomato Late blight	96.40	96.80
13	Tomato Leaf Mold	97.90	97.20
14	Tomato_Septoria leaf spot	97.90	97.00
15	Tomato Spider mites	99.30	99.60
16	Rate overall accuracy of the training for CNN	98.29	
17	Rate overall accuracy of the testing for CNN	98.029	

It should be noted here that the confusion matrix is a global and approved view and the reason for its lack of clarity here is just the large numbers of classes used, to clarify the results more we will display them in the form of a table. As in TABLE II, this shows the accuracy obtained in the test and examination stages for each class, as well as the total accuracy of all classes in the two mentioned stages.

In TABLE III we showed a comparison of the proposed system with some work related to our work. It clarifies the things that have been compared, including the algorithms used for each of the related work and the algorithms used as we have shown in the table are the CNN algorithm and also the ANN algorithm. Also among the things that within the comparison are the comparison with the quantities of the data set used In related work and also in relation to our work, we have proven that our work has used more types and numbers of images than other related work, and the most important thing has been proven that the overall accuracy in our work is the highest accuracy compared to other related work as shown in the following table.

TABLE III. A COMPARISON BETWEEN THE PROPOSED SYSTEM VS. THE RELATED WORK

Author(s), Year	**Ref. No.**	**Algorithm (classification)**	**Dataset size (Images Number)**	**Accuracy**
Raghad Mula Alyas, 2022	[16]	SVM	18160	93.65%
. Habibollah Agh Atabay , 2017	[17]	CNN	19742	95.50-96.48%
Malvika Ranjan ,2020	[18]	ANN	Not Mentioned	80%
Kawasaki and et al,2015	[19]	CNN	800	94.9%
Our proposed system	-	CNN	20636	98.029%

IV. CONCLUSIONS

Given the importance of agriculture and plants in the whole world and in our country, Iraq, and because of many plant diseases that exist today, this research proposed a robust methodology to detect and classify these diseases with accurate and fast results based on computer facilities and Deep Learning Techniques. We conducted this work to obtain the results by using CNN algorithm. We obtained high results over 98%, and this led to very accurate and fast detection of the type of disease and also the type of plant that carries this disease through the leaf of that plant. Fifteen different classes are classified as plant diseases include different plants, internationally famous, and essential in our country, Iraq, are tomatoes, potatoes, and peppers. Various learning rates and optimizers could too be applied as part of future work to experiment with the proposed system. Also, we aspire to increase a large number of different types of plants with other types. Use multiple techniques and create an expert system that detects and classifies plant leaf diseases.

REFERENCES

[1] P. Narvekar, and S. N. Patil, "Novel algorithm for grape leaf diseases detection" International Journal of Engineering Research and General Science Volume 3, Issue 1, pp.no 1240-1244, 2015.

[2] R. Kaur and M.. M. Kaur" An Adaptive Model to Classify Plant Diseases Detection using KNN"International Journal of Trend in Scientific Research and Development (IJTSRD) ISSN: 2456-6470.

[3] P. V,Rahul Das,and V. Kanchana , "Identification of Plant Leaf Diseases Using Image Processing Techniques" 978-1-5090-4437-5/17/$31.00 ©2017 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development.

[4] M. J. Vipinadas. [1], A.Thamizharasi. [2] "A Survey on Plant Disease Identification" International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 6, Nov-Dec 2015

[5] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, no. September 2017, pp. 311–318.

[6] J. M. AL-Tuwaijari and S. I. Mohammed, "Face Image Recognition Based on Linear Discernment Analysis and Cuckoo Search Optimization with SVM", International Journal of Computer Science and Information Security. ISSN 1947 5500, IJCSIS November 2017 Volume 15 No. 11

[7] S. P Mohanty, D. P Hughes, and Marcel Salath´e."Using deep learning for image-based plant disease detection".In: Frontiers in plant science 7 (2016), p. 1419.

[8] Gittaly Dhingra1 & Vinay Kumar1 & Hem Dutt Joshi 1" Study of digital image processing techniques for leaf disease detection and classification" © Springer Science+Business Media, LLC, part of Springer Nature 2017 https://doi.org/10.1007/s11042-017-5445-8\ .

[9] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," Computational Intelligence and Neuroscience, vol. 2016, Article ID 3289801, 11 pages, 2016.

[10] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," Computers and Electronics in Agriculture, pp. 418–424.

[11] Jamal. M. AL-Tuwaijari, S. I. Mohammed, "Performance Evaluation of Face Image Recognition Based Voila–Joins with SVM", Iraqi Journal of Information Technology. V.9 N.3. 201

[12] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in Proceedings of the 2018 IEEEWinter Conference on Applications of Computer Vision (WACV), pp. 839–847, 2018.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[14] Deng J et al. (2009) Imagenet: A large-scale hierarchical image database in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. (IEEE), pp. 248–255.

[15] D.P. Hughes , M. Salathé (2015) An open access repository of images on plant health to enable the development of mobile disease diagnostics. CoRR abs/1511.08060.

[16] T.M. Prajwala , A. Pranathi, Kandiraju Sai Ashritha, Nagaratna B. Chittaragi*, Shashidhar G. Koolagudi" Tomato Leaf Disease Detection using Convolutional Neural Networks"Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018, Noida, India 978-1-5386-6835-1/18/$31.00 ©2018 IEEE.

[17] H. A. Atabay ''Deep Residual Learning for Tomato Plant Leaf Disease Identification"' 1Department of Computer, Gonbad Kavous University, Journal of Theoretical and Applied Information Technology 31st December 2017. Vol.95. No 24 © 2005 – ongoing JATIT & LLS.

[18] M. Ranjan1, M. R. Weginwar,NehaJoshi,Prof.A.B. Ingole, detection and classification of leaf disease using artificial neural network,International Journal of Technical Research and Applications e-ISSN: 2320-8163, Volume 3, Issue 3 (May-June 2015), PP. 331-333

[19] Y. Kawasaki, H. Uga, S. Kagiwada, S., Iyatomi, H.: Basic study of automated diagnosis of viral plant diseases using convolutional neural networks. In: Advances in Visual Computing: 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, 14–16 December 2015,Proceedings, Part II, pp. 638–645 (2015).