ISSN No:-2456-2165

Plant Disease Detection Using Machine Learning Algorithm

Pradip Chougala, MTech Student, Dept ISE, SDMCET Dharwad,
Dr. Rajashekharappa, Assistant Professor, SDMCET Dharwad.

Abstract:- Pest infestations have an impact on the nation's agricultural output when they harm plants and crops. Farmers or experts typically keep a close check on the plants to spot any signs of disease. However, this procedure is frequently time-consuming, expensive, and unreliable. Results from automatic detection employing image processing methods are quick and precise. This study uses deep convolutional networks to establish a new method for developing illness recognition models that is supported by leaf image categorization. The field of precision agriculture has a chance to grow and improve the practise of precise plant protection as well as the market for computer vision applications. A quick and simple system implementation in practise is made possible by a wholly original training methodology. The entire process of putting this disease recognition model into practise, from gathering photos to create a database to having it reviewed by agricultural specialists and using a deep learning framework to carry out the deep CNN training, is comprehensively documented throughout the research. With the help of a deep convolutional neural network that has been trained and fine-tuned to accurately match the database of plant leaves that was compiled independently for various plant illnesses, the technique paper presented here may represent a novel way for identifying plant diseases. The innovation and advancement of the developed model reside in its simplicity; healthy leaves and backdrop images are consistent with other classes, allowing the model to use CNN to differentiate between ill and healthy leaves or from the environment. On earth, food is produced by plants. As a result, plant infections and diseases pose a serious threat, and the most common method of diagnosis is by looking for visible symptoms on the plant's body. Diverse research projects intend to identify workable methods for safeguarding plants as an alternative to the customarily time-consuming practise. The development of technology in recent years has led to the emergence of several alternatives to laborious old procedures. Deep learning methods are particularly effective at solving picture classification issues.

I. INTRODUCTION

Sustainable agriculture issues are directly related to the challenge of effective disease prevention. Inexperienced pesticide use can result in microorganisms developing long-term resistance, greatly weakening their ability to defend themselves. One of the foundational elements of precision agriculture is the prompt and precise identification of plant diseases. In this changing climate, it is essential to stop needless waste of money and other resources in order to achieve healthy production. Appropriate and timely illness identification, including early prevention, has never been more significant. There are numerous methods for identifying plant disorders. In cases where there are no obvious symptoms of a disease or when it is too late to take action, a sophisticated study is required. However, since most diseases produce subtle visible symptoms, the primary method used in practise for disease identification is an eye exam performed by a qualified specialist.

The most important sector of our Economy is Agriculture. Various types of disease damages the plant leaves and effects the production of crop there for Leaf disease detection is important. Regular maintenance of plant leaves is the profit in agricultural products. Farmers do not expertise in leaf disease so they producing lack of production. Leaf disease detection is important because profit and loss depend on production. So that here use deep learning techniques to detect apple, grape, corn, potato, and tomato plant leaves diseases. That contains twenty-four types of leaf diseases and twenty-four thousand leaves images are used.

There are a total of 24 different sorts of labels for the leaves of apple, grape, corn, potato, and tomato plants, including "apple scab," "apple rust," and "healthy." Black rot, Esca, healthy, and Leaf blight are the specific grape labels. The following are listed on the corn label: Corn Northern Blight, Corn Rust, Corn Healthy, and Corn Cercospora spot. Early blight, healthy, and late blight are the three potato labels. The following pests and diseases are listed on the tomato label: bacterial spot, early blight, healthy, late blight, leaf mould, septoria leaf spot, spider mite, target sport, and mosaic virus.

The dataset consist of 31,119 images of apple, grape, potato and tomato, all Images are resized into 256 x 256,that images divided into two parts training and testing dataset.

ISSN No:-2456-2165

1. Apple scab 2. Grape Esca 3.Corn leaf spot 4.potato Early 5.Tomato Bacterial
Fig 1:- Blight Spot

➢ *Leaves with Disease part*

In Leaves with Disease part we can see vegetable and fruit leaves like potato, tomato, corn, apple, grape with diseased part this disease can be easily detected using deep learning techniques.

This disease detected using convolutional neural network (CNN), and also this model is compared with VGG16. Images are resized into 224 x 224.

➢ *Applications*

- Biological research
- Plant leaf disease detection also useful in agricultural institute
- Some plant leaf disease detection automatic technique are beneficial for large work of monitoring in farm of crops disease detection

➢ *Objectives*

Farmers or experts keep a close eye on the plants to spot and recognise sickness.

However, this procedure is frequently time-consuming, expensive, and unreliable. Results from automatic detection employing image processing methods are quick and precise. This study uses deep convolutional networks as a substitute method for developing a disease recognition model, supported by leaf image classification.

- *The purpose of this deep convolutional network research*

The goal of this study is to use CNN to focus on the identification of potato, corn, grape, and apple leaf diseases.

The leaves of both healthy and sick plants are examined using CNN.

➢ *Motivation*

Identifying and recognition of leaves disease is the solution for saving the reduction of large farm in crop disease detection and profit in productivity, it is beneficial in agricultural institute, Biological research.

II. EXISTING WORK AND IMPLEMENTATION WORK

➢ *Overview of Existing Work*

Existing work related to leaf disease detection using CNN show to detect and classify leaf disease using image processing techniques that follow steps like

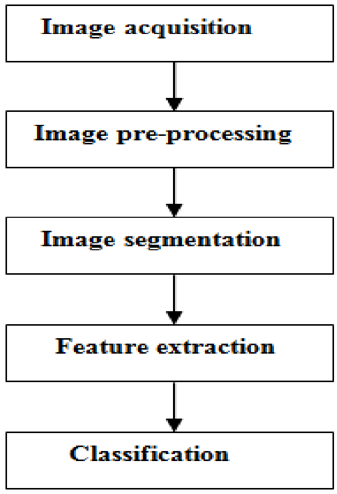

Fig 2:- General Block Diagram of Feature Based Approach

Image Acquisition: Image Acquisition is the process of loading an image into a digital camera and then storing it on a digital medium for use in subsequent MATLAB processes. B. Image pre-processing: The main goal of image pre-processing is to strengthen certain image features or enhance the image information that contains undesired distortions in preparation for any processing. Pre-processing techniques include dynamic image size and shape, noise filtering, image conversion, image enhancement, and morphological processes, among others. C. Image Segmentation: To divide images into clusters for image segmentation, the K-means cluster algorithm is used. At least one component of each cluster must have an image with the majority of the unhealthy area. Applying the k means cluster algorithmic rule, the objects are divided into K different groups for each collection of characteristics. D. Texture feature extraction: Using GLCM, texture features are extracted once clusters have formed (Gray-Level Co-occurrence Matrix). E. Classification: To test for leaf disease, classification is used. For classification, the Random Forest classifier is employed.

➢ *Implementation work*

Machine learning Model: There are a total of 24 different sorts of labels for apple plant leaves, including apple scab, black rot, apple rust, and healthy. Corn label, specifically: Corn Blight, Corn Rust, Corn Healthy, and Corn Cercospora Gray spot. Black rot, Esca, healthy, and Leaf blight are the specific grape labels. Early blight, healthy, and late blight are the three potato labels. Specifically, the following tomato diseases are included on the label: bacterial spot, early blight, healthy, late blight, leaf mould, septoria leaf spot, spider mite, target sport, mosaic virus, and yellow leaf curl virus.

The dataset includes 31,119 photos of tomatoes, apples, maize, grapes, potatoes, and other produce; 24000 of those photographs were used. All photos have been scaled to 256 by 256, and the training and testing datasets have been separated into two portions using an 80-20 split (80 percent of the whole dataset used for the training and 20 percent for the testing). The CNN model after that.

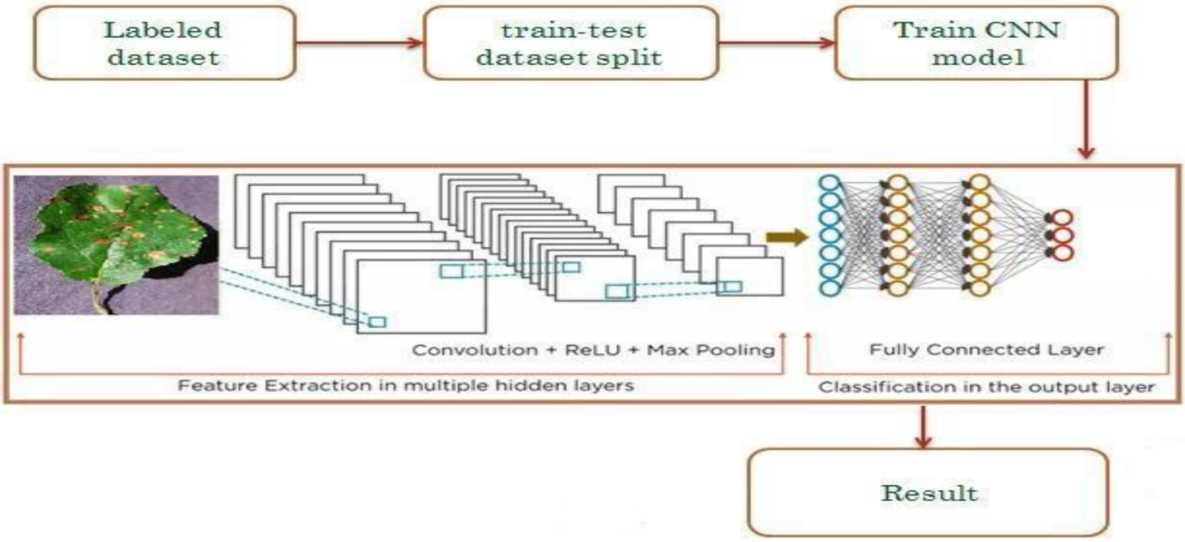

Fig 3:- Proposed workflow

Convolutional neural networks can be used to create a computer model that takes unstructured visual inputs and transforms them into output labels of matched categorisation (CNN). It is a type of multi-layer neural network that may be instructed to learn the features needed for classification. Less pre-processing is required compared to conventional methods, and automatic feature extraction is done for better performance. For the goal of identifying leaf sickness, a LeNet architecture version produced the best results. LeNet is a simplistic CNN model that has four layers: fully connected, convolutional, activation, and max-pooling. This architecture is used in the LeNet model to classify leaf diseases. It features an additional block of convolution, activation, and pooling layers in comparison to the original LeNet architecture. The model utilised in this investigation is shown in Fig. 2. A convolution layer, an activation layer, and a max pooling layer are present in each block. Three such blocks, completely connected layers, and soft-max activation are utilised in this architecture. Convolution and pooling layers are utilised for feature extraction, and fully linked layers are employed for classification. Through the use of activation layers, the network is given nonlinearity. maintain the size of the image. The max pooling layer is used to minimise the size of the feature maps, expedite training, and make the model more resistant to tiny changes in input. The largest kernel size used in maximum pooling is 22. Re-LU activation layers are used in each of the blocks to introduce non-linearity. To avoid over-fitting the train set, the Dropout regularisation technique has also been used with a 0.5 maintain probability. Dropout regularisation randomly removes neurons from the network during training iterations, reducing the variance of the model and simplifying the network. This method lessens the complexity of the network, hence preventing overfitting. The classification block is composed of two neural network layers of 500 and 10 neurons each, each of which is fully connected. After the second dense layer, a soft max activation function is used to determine the probability scores forthe ten classes.

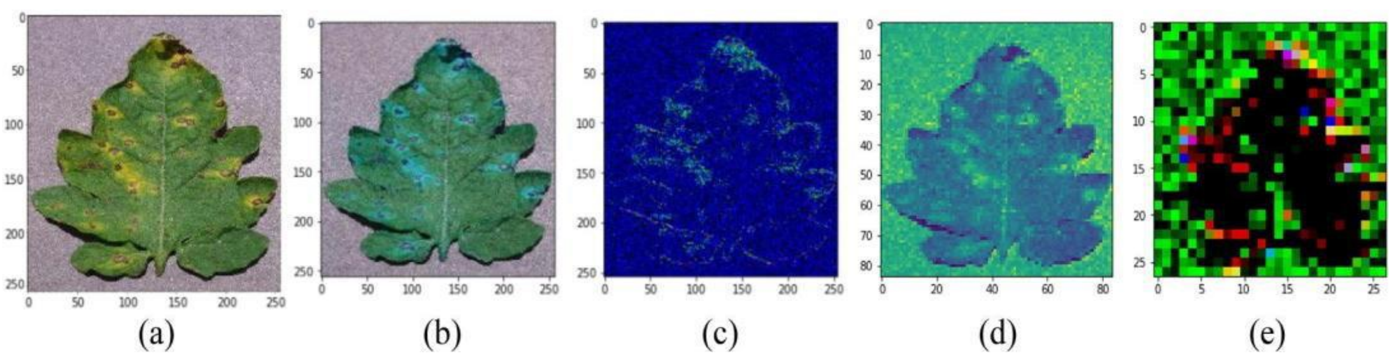

Fig 4:- Experimental result (a) input image (b) convolution layer-1(c) convolution layer-2 (d) convolution layer-3 (e) flatting layer.

In addition, each experiment will compute the overall accuracy over the course of training and testing (for each epoch). The overall accuracy score will be used to evaluate performance. Transfer learning is a method of knowledge exchange that uses 224*224 fixed-size images and requires the least amount of training data. Transfer learning is useful for transferring knowledge from one

ISSN No:-2456-2165

model to another. Sentiment analysis, activity recognition, software defect prediction, and plant classification are just a few of the tasks to which transfer learning has been applied. The effectiveness of the suggested Deep CNN model is compared to that of the well-known transfer learning technique VGG16 in this study.

Fig 5:- VGG16 layered architecture

It employs the VGG16 convolutional neural network. The convolution layer's input image must be 224 x 224 RGB fixed in size. The image is then passed on to convolutional layers, where the filters are applied with the lowest possible receptive field—33—to capture the ideas of left, right, up, and down in addition to centre. In some configurations, it makes use of 11 convolution filters, which might cause the input channels to undergo linear modification before becoming nonlinear. One pixel is the fixed convolution stride. The spatial padding of the input to the convolution layer is such that the spatial resolution is preserved after convolution when 33 convolution layers are used. There were five max-pooling layers that carried out spatial pooling after some of the convolution layers (not all the conv. layers are followed by max-pooling). There is a 2 * 2 pixel Max-pooling.

A stack of convolutional layers is followed by three layers. The first two devices each have 4096 channels, while the third device uses a 1000-way ILSVRC classification and has 1000 channels. The soft-max layer is the final one. The fully connected layer arrangement helps to pinpoint the leaf illness.

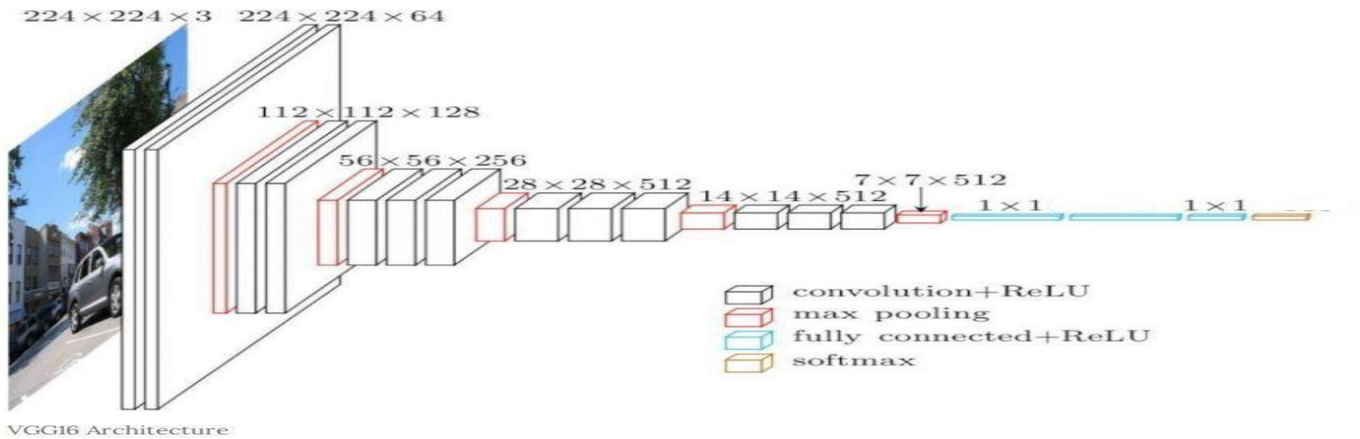

Fig 6:- VGG16 architecture

All concealed layers have correction capabilities. Re-Lu It should also be highlighted that none of the networks feature Local Response Normalization (LRN), which does not enhance the performance on the dataset. Repaired linear units contain non-linearity on networks.

For the Large-Scale Image, CNN was employed. The best way to identify plant diseases is to complete two tasks. The first step is object localization, which is the detection of objects in an image that come from various classes. The second is picture classification, which involves labelling each image with one of various categories. There are seven distinct layers in the CNN model. Certain information is handled in each layer. Here are those seven layers: Convolutional layers with fully connected, Soft-max, input, output, and pooling layers.

Input layer: Data in the form of images are contained in the input layer. The parameters include the image's dimensions (height, width, depth, and colour information) (RGB). The size of the input is a fixed 224 x 224 RGB picture.

Convolutional layer: Another name for this layer is feature extraction layer. With the use of dot products of the picture dimensions, this layer extracts the salient features from the provided collection of photographs.

Pooling Layer: By lowering (or reducing) the dimensions of the featured matrix produced by utilising dot products, the pooling layer aids in lowering the processing power required to process the data.

Fully connected layer: It is made up of biases, neurons, and loads. It links neurons in one convolutional layer to neurons in another layer.

Softmax Layer/ Logistic Layer: Multi-classification is carried out using Softmax. The binary classification is carried out by the logistic layer. It establishes the likelihood that a specific object will appear in the image. The likelihood is "1" if the object is visible in the image and "0" if it is not.

Activation Function- ReLU: The node is activated after the node transforms the total weighted input and passes it into the operation. An activation function used in neural networks for convolutional operations is the Rectified Linear Unit (ReLU).

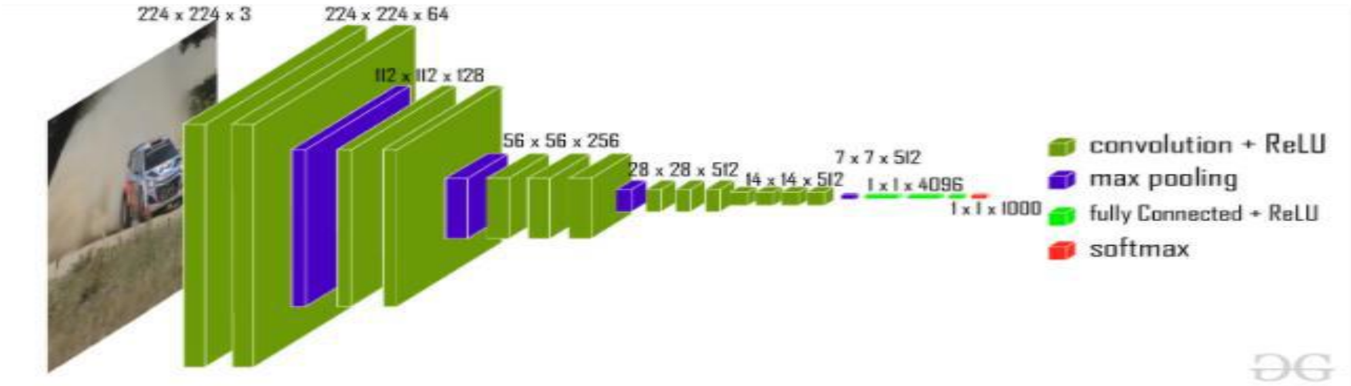

Fig 7:- System Architecture

Therefore, it is suggested to construct an image processing system to automate the detection and classification of leaf batches into particular illnesses in order to determine the cause of the symptom using an automated tool. The system is comprised of three basic pieces, as depicted in the above diagram: Image Analyser, Feature Database, and Classifier, respectively [9]. The two steps of the processing that these blocks attempt to propose are as follows: offline Phase: A picture analyser processes a large number of defective photos to extract aberrant features.

- *CNN Model Steps:*

Conv2D: It is a 2D Convolution Layer, this layer creates a convolution kernel that's wind with layers input which helps produce a tensor of outputs.
Keras.layers.Conv2D(filters, kernel_size, strides=(1, 1), padding='valid', data_format=None, dilation_rate=(1, 1), activation=None, use_bias=True, kernel_initializer='glorot_uniform', bias_initializer='zeros', kernel_regularizer=None, bias_regularizer=None, activity_regularizer=None, kernel_constraint=None, bias_constraint=None

Max-pooling: A pooling method called max pooling may select the best element from the feature map area that the filter has covered. Therefore, the output following the maximum pooling level would be a feature map that included the most crucial elements of the prior feature map.

Flatten: There is a "Flatten" layer sandwiched between the convolutional layer and, consequently, the fully connected layer. A fully connected neural network classifier is supplied with a vector created by flattening a two-dimensional feature matrix.

Image Data Generator: Image Data Generator rapidly learned about Python generators that will automatically convert batches of unprocessed tensors from image files on disc.

Training Process: Before a trainer conducts a private training session, effective training starts, and it continues after the session is over. Assessment, motivation, planning, delivery, and evaluation are the five connected processes or activities that make up training

Epochs: A word used in machine learning is called an epoch, which describes how many rounds the machine learning algorithm has made across the entire training dataset. Typically, datasets are organised into batches (especially when the quantity of knowledge is extremely large).

Validation Process: The process of evaluating a trained model against a testing data set is called validation. The training set's corresponding data set may contain a separate section that serves as the testing data set. The main goal of using the testing data set is to evaluate a trained model's capacity for generalisation.

- *Training and Testing Model:*

The dataset is pre-processed, including image scaling, reshaping, and array form conversion. On top of that, similar processing is applied to the test image. Any image from a dataset of approximately 38 distinct plant leaf diseases is frequently used as a test image for the software.

ISSN No:-2456-2165

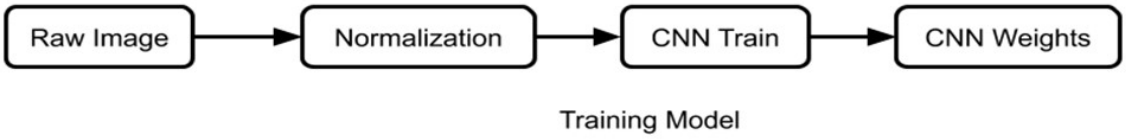

Fig 8:- Training Model

The train dataset is used to train the model (CNN), enabling it to recognise the test image and, consequently, the disease it represents. Dense, Dropout, Activate, Flatten, Convolution 2D, and Maxpooling 2D are some of the layers that CNN has. If the plant species is included in the dataset and the model has been successfully trained, the programme can detect the illness. The test image and trained model are compared after effective training and pre-processing in order to identify the disease.

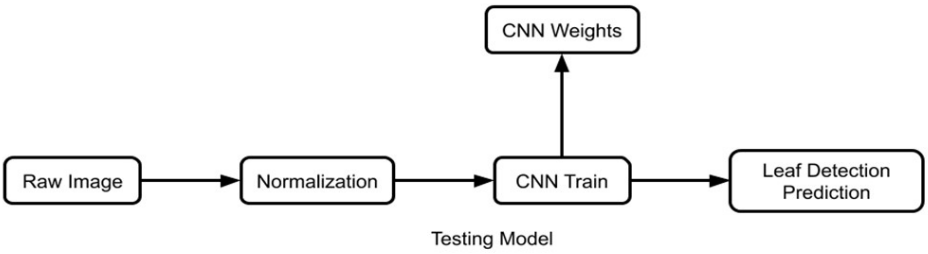

Fig 9:- Testing Model

III. DATASET, IMPLEMENTATION AND RESULT

Dataset: The dataset was obtained from the online Kaggle of Plant Village dataset, and the code was added to the Kaggle online kernel for efficient computation and the analysis of training loss and validation.

Image Pre-processing and Labelling: Pre-processing often involves removing low-recurrence foundation disturbance, adjusting the power of the individual particle images, removing reflections, and obscuring portions of images. Pre-processing of images is a technique for enhancing information. Additionally, the pre-processing method for images involved physically manipulating the seeming variety of images, creating a square around the leaves to highlight the region of fascination (plant leaves). Photographs having a less ambitious purpose and measurements that weren't exactly 500 pixels were not regarded as significant pictures for the dataset during the period of collecting the images. In addition, the dataset was limited to only those images where the location of intrigue was closer to the objective. In this way, it was ensured that images contained all the information needed for highlight learning. The Internet makes it possible to find a lot of resources, yet their value is frequently disputed. Horticultural experts examined leaf images and labelled all the images with appropriate infection abbreviations, taking into account a real concern for confirming the accuracy of classes in the dataset that was first collected by a catchphrases search. As is common knowledge, using correctly defined images is important for the preparation and approval dataset. Only in this way is it possible to develop an accurate and reliable identifying model. At this stage, duplicate images that remained after the primary focus of grouping and classifying images was removed from the dataset.

IV. RESULTS

Convolutional networks are known to be capable of learning features when trained on larger datasets, hence the outcomes of training with only original photos won't be examined. An overall accuracy of 88% was attained once the network's settings were adjusted. The trained model was also put to the test for each class separately. Every image from the validation collection was put to the test. The results that were achieved should be compared to some other results, as recommended by good practise standards. Additionally, aside from those dealing with plant species identification using photographs of the leaf, there are presently no commercial solutions available. In this research, a technique for automatically classifying and identifying plant diseases from leaf photos was investigated. It was explained in detail every step of the way, from gathering the images used for training and validation through image pre-processing and augmentation to guiding the deep CNN and fine-tuning. To evaluate the performance of the newly developed model, various tests were run. There was no comparison with findings obtained using a similar method because, as far as we all know, the proposed method has not been used in the field of disease recognition. The test image we've provided in this case is a leaf spot.

➢ *Result and Conclusion:*

Convolutional networks are known to be capable of learning features when trained on larger datasets, hence the outcomes of training with only original photos won't be examined. An overall accuracy of 88% was attained once the network's settings were adjusted. The trained model was also put to the test for each class separately. Every image from the validation collection was put to the test. The results that were achieved should be compared to some other results, as recommended by good practise standards. Additionally, aside from those dealing with plant species identification using photographs of the leaf, there are presently no commercial solutions available. In this research, a technique for automatically classifying and identifying plant diseases from leaf photos was investigated. It was explained in detail every step of the way, from gathering the images used for training and validation through image pre-processing and augmentation to guiding the deep CNN and fine-tuning. To evaluate the performance of the newly developed model, various tests were run. There was no comparison with findings obtained using a similar method because, as far as we all know, the proposed method has not been used in the field of disease recognition.

REFERENCES

[1]. Parismita Bharali, Chandrika Bhuyan, Abhijit Boruah, "Plant Disease Detection by Leaf Image Classification Using Convolutional Neural Network" ICICCT 2019, pp.194-205.
[2]. Rahman Tashakkori, Timothy Jassmann, R. Mitchell Parry, "Leaf Classification Convolutional Neural Network App" , IEEE Southeast 2015.
[3]. M. B. Riley, M. R. Williamson, and O. Maloy, "Plant disease diagnosis. The Plant Health Instructor," 2002.
[4]. Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk and Darko Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification", 2016.
[5]. H. Cartwright, Ed., Artificial Neural Networks, Humana Press, 2015.
[6]. https://www.analyticsvidhya.com/blog/2017/06/architecture-of-convolutional-neural-networks-simplifieddemystified/
[7]. https://en.wikipedia.org/wiki/Convolutional_neural_network
[8]. S. Ajantha, G. Mahapriya, J. Mispa, S. Nithya, Measurement of Multimodal Brain Tumor Segmentation using Neural Network in IJERT vol. 5, issue 13.
[9]. Mohammad El-Helly, Ahmed Rafea, Salwa El-Gammal, "An Integrated Image Expert Processing System for Leaf Disease Detection and Diagnosis", Central Lab. For Agricultural System(CLAES), Agricultural Research Center(ACR), Giza,Egypt
[10]. https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/.