# ABSTRACT

Rice is one of the major cultivated crops in India which is affected by various diseases at various stages of its cultivation. It is very difficult for the farmers to manually identify these diseases accurately with their limited knowledge. Recent developments in Deep Learning show that Automatic Image Recognition systems using Convolutional Neural Network (CNN) models can be very beneficial in such problems. We have used publicly available dataset and we have used Machine Learning to develop our deep learning model. The proposed CNN architecture is based on Machine Learning models and is trained and tested on the dataset collected from rice fields and the internet.

**KEYWORDS:** Rice leaf diseases . Image recognition . Convolutional Neural Network.

## Chapter 1 : INTRODUCTION

### DESCRIPTION:

Agricultural sector plays a crucial role for economic development of any country.In terms  
of raw materials, the majority of countries are dependent on agricultural goods. Rice is  
mostly cultivated crop around the globe. Rice is cultivated in over 100 countries around  
the world. A total of 158 million hectares are harvested each year, yielding more than  
700 million tons of rice. In comparison to other continents, Asia produces the majority  
of rice . Because of increasing population it is affecting environment in terms of global  
warming, rapid climatic shifts Yadav et al. (2021). The agricultural sector is suffering as  
a result of these changes in the environment.

Crops are becoming infected with a variety of illnesses as a result of environmental  
changes. This has a significant impact on crop quality, quantity, and productivity. Dif-  
ferent forms of illnesses are wreaking havoc on rice fields these days, having a negative  
effect on crop production around the world. Many illnesses have been seen in recent  
years, including rice leaf blast, brown spot, Hispa, rice curl disease, and many others  
(Jiang et al.; 2020). These diseases can be found on any part of the rice plant, including  
the leaf, neck, and ear.

Plant leaf disease has a negative impact on not just agricultural  
productivity but also on the environment in terms of pollution. Plant diseases are re-  
sponsible for 10 to 15% of total productivity losses. In the worst-case scenario, farmers  
could lose up to 50% of their crop, which is a significant loss for farmers and the country’s  
economy (Tian et al.; 2021). As a result, it is critical to detect a plant disease early on  
in order to ensure sustainable and accurate agriculture and to avoid waste of financial  
and other resources. As a result, early detection of pests on crops is critical for avoiding  
large use of fertilizers and pesticides for obtaining higher productivity.For large-scale crops, naked-eye participant observation are not practical nor sufficient.

In today’s Artificial Intelligence (AI) environment, convolutional neural networks (CNN)  
and machine learning can play important role in classifying diseases.Research in this field is imperative, whether for the rice grains consumed by three billion people or other agricultural products that are as significant or even more popular. The main aim of this work is to develop a system utilizing novel optimized ML and deep learning (DL) techniques that will accurately detect, classify, and diagnose rice disease automatically without human intervention. The ultimate aim is also to propose novel methods that can achieve higher diagnostic accuracy than other techniques in the extant literature that use similar datasets .

This study is focused on the identification of rice leaf diseases. This involves image aquisition,segmentation,feature extraction,image analysis and classification techniques are used to detect plant diseases. Photos leaves from rice plants are taken with a digital camera or similar unit, and the images are used to classify the affected region in the leaves. To detect disease in rice leaves, we use a Convolution neural network and a Deep neural network in the proposed framework. This paper proposes a framework that employs low-cost,open-source software to achieve the task of reliably detecting plant disease.

### PROBLEM STATEMENT:

To detect the disease present in Rice leaf using Image Processing and Machine Learning techniques.

### OBJECTIVES:

* To completely & thoroughly understand the biological jargon behind rice leaf to devise a better system.
* To study the nature of Rice leaf using image processing techniques that involves:

1.Image Acquisition .

2.Image Pre-processing.

3.Image Enhancement.

4.Image Segmentation.

5.Image Analysis.

* To detect and classify the diseases present in Rice leaf using Algorithms in Machine Learning

and choose the right parameters for the algorithm.

* To compare the following algorithms and get the best results.

## Chapter 2 : REVIEW OF LITERATURE

In the literature survey we provide a brief summary of the different methods that have been proposed for clustering over the period of 2002 to 2021. We have been though 7 papers each of which has a unique approach towards segmentation in some parameter or the other. The summaries of each of the papers are provided below.

**Ms.Kiran.R,Gavhale,Prof.Ujwalla Gawande**

The present paper reviews and summarizes image processing techniques for several plant species that have been used for recognizing plant diseases. The major techniques for detection of plant diseases are: BPNN, SVM, K-means clustering, and SGDM. These techniques are used to analyses the healthy and diseased plants leaves. Some of the challenges in these techniques viz. effect of background data in the resulting image, optimization of the technique for a specific plant leaf diseases, and automation of the technique for continuous automated monitoring of plant leaf diseases under real world field conditions.

**Bulent Tugrul , Elhoucine Elfatimi and Recep Eryigit \***

CNN methods are widely used in the detection of plant diseases. It has solved the problems of traditional object detection and classification methods. In this study, we presented a detailed review of CNN-based research on plant leaf disease detection in crops over the last five years.

The following abbreviations are used in this manuscript:

DL Deep Learning

CNN Convolutional Neural Networks

DCNN Deep Convolutional Neural Networks

FOA The Food and Agriculture Organization

**Jashraj Karnik and Dr. Anil Suthar**

In this paper [1] They have performed classification using convolutional neural network algorithm to classify the tomato leaf disease by doing image processing techniques like DTW and GLCM and segmentation process for extracting the feature by categorizing leaf image into smaller parts of texture and determines the boundary of the tomato leaf and label on it. The Classification and detection using convolutional neural network to get the results.The author has used the pre-trained model using transfer learning i.e., VGG16, Res-Net, Inception such as combination of deep learning, colour, and Position augmentation dataset gives the highest classification performance than all other datasets. The data manipulation technique by PCA algorithms and neural style transfer. The result outcome of transfer learning through data manipulation is 89%.In this paper [3] in this they have performed techniques of machine learning algorithm using KNN (K-nearest neighbors) classifier. The texture of the leaf disease image is use for the feature extraction for classification. They have performed algorithm classifier to classify the diseases they are Alternaria, anthracnose, bacterial blight, leaf spot, and canker of various plant species is using KNN classification approach.

**Sunil S. Harakannanavar , Jayashri M. Rudagi , Veena I Puranikmath , Ayesha Siddiqua , R Pramodhini**

. In this paper the authors pre-processed by image resizing,contrast enhancement and color-space conversion.K-means clustering is used for segmentation and feature extraction using GLCM is performed. For segmentation, based on a set of marks generated by analysis of the color and luminosity components of different regions of image is L∗A∗B∗ color spaces. The GLCM is used for feature extraction. . Arya et al., [21,22] takes input RGB image and creates color transformation then conversion of the input samples to HIS format. Finally, segment the components using Otsu’s method. The model is developed based on the IP and ML approaches for detection of leaf disease in presented in this section. The proposed model (DWT+PCA+GLCM+CNN) using computer vision and machine learning approaches for leaf disease detection. The tomato samples having six disorders are considered to evaluate its accuracy and to recognize the leaf disease as Healthy or Unhealthy. As a part of image processing, the samples of tomato are resized to 256 × 256 pixels to maintain equal in their size throughout the experiment. The HE and K-means clustering are employed to maximize the quality and segment the leaf samples. The DWT, PCA and GLCM are used to extract the informative regions/features of the samples. In the next stage as a part of machine learning approaches the SVM, KNN and CNN are used to classify the features and the performance of the model is recorded.

**G. Geetha , S.Samundeswari , G.Saranya ,K.Meenakshi and M. Nithya**

Machine learning in detecting and classifying diseases of a plant leaf. Machine learning in detecting and classifying diseases of a plant leaf. The paper gives an elaborate view about the techniques which can be implemented for detecting and classifying the various plant leaf diseases caused by bacteria, viruses and fungi. Based on their morphology i.e., their particular form, shape, or structure the diseases detected through classification are categorized. In this paper detection of diseases of the stem plant is done, using the HSV algorithm, GLCM algorithm and SVM to perform and initiate the segmentation process, followed with feature extraction and classification respectively. It discusses the removal of noise, conversion from RGB to HSV and vice versa. This paper talks about the detection and recognition of abnormalities of plants for training and study papaya leaves were taken. The digital image processing provides a vast area for identification of diseases through the various algorithms it supports. The most common bacterial, fungal and viral diseases are studied which affects the plant leaves and roots on wide scale and reduces the productivity of the plants can be easily studied and identified through RGB to grey scale conversions.Authors used HOG(Histogram of Oriented Gradients )to extract the spatial dependency of the texture.In this they also used other genetic algorithm and Naïve Bayes algorithm for classification.

**Jayshree Hajgude, Jayesh Kriplani, Dhiraj Chhabria, Anish Verliani**

In 2018 in International Conference on Design Innovations for 3Cs Compute Communicate Control Shima Ramesh, Mr. Ramachandra Hebbar, Niveditha M, Pooja R, Prasad Bhat N, Shashank N,Mr. P V Vinod proposed According to past studies, 42% of agricultural production is in Loss, and that too solely as a result of plant leaf diseases. With this technique, plant leaf diseases can be detected from input images to overcome this major issue. This process involved steps like image preprocessing, image segmentation, feature extraction. On the basis of these three steps, K Nearest Neighbor (KNN) classification is applied. As a result of the proposed implementation, A 98.56% accuracy rate in predicting plant leaf diseases. In 2021 in BMC(Biomedcentral) Jun Liu and Xuewei Wang proposed plant disease and pest detection using SVM algorithm. SVM is used for recognizing image samples. The accuracy of implementation is about 92% with a sufficient and stable dataset used for the project.

**Kowshik B ,Savitha V ,Nimosh madhav M ,Karpagam G , Sangeetha K**

The proposed system tracks the cultivated field on a regular basis. The CNN and DNN algorithms are used to identify crop diseases at an early stage. Machine learning methods are used to train the model, which aids in making appropriate disease decisions. To contain infected diseases, the farmer is advised to use pesticides as a cure. In the future, the proposed scheme could be expanded to provide additional facilities such as nearby government markets, pesticide price lists, and a nearby open market, among others. d classification of plant leaf diseases in the future. Some of the organisms on which the proposed algorithm is evaluated include banana, beans, jackfruit, lemon, mango, potato, tomato, and sapota. As a result, similar diseases for these plants were investigated. The best results were obtained with very little computational effort, demonstrating the efficacy of the proposed algorithm in recognising and classifying crop diseases. Another benefit of this approach is that plant diseases can be detected at an early stage, or even at the beginning. Convolution neural network and Deep neural network algorithms may be used to increase recognition rates in the classification process.

**Chapter 3 : SYSTEM REQUIREMENT SPECIFICATIONS**

**3.1HARDWARE CONFIGURATION**

* + - Processor: Intel core i3 or above.
    - 64-bit, quad-core, 1.5 GHz minimum per core
    - Ram: 4 GB or more
    - Hard disk: 10 GB of available space or more.
    - Display: Dual XGA (1024 x 768) or higher resolution monitors
    - Operating system: Windows 8 and above

**3.2 SOFTWARE REQUIREMENTS**

Windows: Python 3.6.2 or above, PIP and NumPy 1.13.1

**Python:**

Python is an interpreted, high-level, general purpose programming language created by Guido Van Rossum and first released in 1991, Python's design philosophy emphasizes code Readability with its notable use of significant Whitespace. Its language constructs and object- oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming.

**PIP:**

It is the package management system used to install and manage software packages written in Python.

**NumPy:**

NumPy is a general-purpose array-processing package. It provides a highperformance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

* A powerful N-dimensional array object
* Sophisticated (broadcasting) functions
* Tools for integrating C/C++ and Fortran code
* Useful linear algebra, Fourier transform, and random number capabilities

**Pandas :**

Pandasis the most popular python library that is used for data analysis. It provides highly optimized performance with back-end source code is purely written in *C* or *Python*. We can analyze data in pandas with

1. Series
2. Data frames

**Anaconda:**

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution includes data-science packages suitable for Windows, Linux, and macOS. Anaconda distribution comes with 1,500 packages selected from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator, as a graphical alternative to the command-line interface (CLI).

**Jupyter Notebook:**

Anaconda distribution comes with 1,500 packages selected from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator, as a graphical alternative to the command line interface (CLI). A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text mathematics, plots and rich media, usually ending with the “. ipynb" extension.

**Tensor Flow:**

Tensor flow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

**Keras:**

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or Plaid ML.

Designed to enable fast experimentation with deep neural networks, it focuses on being user- friendly, modular, and extensible. Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code.

**OpenCV:**

OpenCV (Open source computer vision) is a library of programming functions mainly aimed at real-time computer vision. Originally developed by Intel, it was later supported by willow garage then Itseez (which was later acquired by Intel). The library is cross platform and free for use under the open source BSD license. OpenCV supports some models from deep learning frameworks like TensorFlow, Torch, PyTorch (after converting to an ONNX model) and Caffe according to a defined list of supported layers. It promotes Open Vision Capsules. which is a portable format, compatible with all other formats.

**Chapter 4 : DATASET USED**



Fig. Brown spot disease

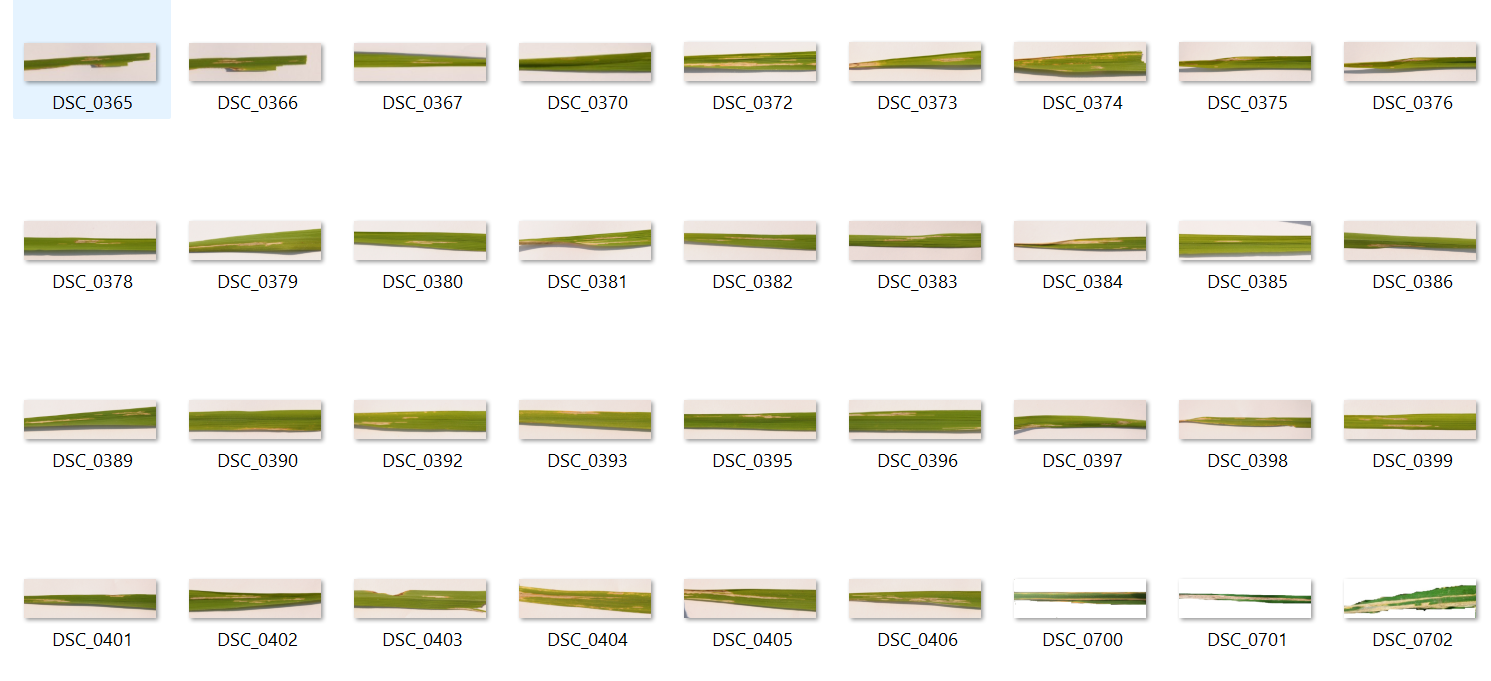


Fig. Leaf Smut disease

## Chapter-5: ALGORITHM

## Step 1 : Collect a dataset of images of healthy and diseased rice plant

## leaf .

## Step 2 : Take an input image known as image acquisition.

## Step 3 : Pre-process the images to remove noise and resize them to

## a consistence size.

## Step 4 : Extract the features of the leaves using Image Processing Techniques.

## Step 5 : Enhance the Images Using Image Enhancement Techniques.

## Step 6 : Segment the images Acquired to distinguish among the regions of

## pixels that are present in the image.

## Step 7 : Split the dataset into training and testing sets.

## Step 8 : Define the CNN architecture ,such as the number of layers

## and the size of convolutional filters.

## Step 9 : Train the CNN on the training set using supervised learning,

## where the labels are the corresponding diseases for each

## image.

## Step 10 : Evaluate the performance of CNN on the testing set.

## Step 11 : Use the trained CNN to predict the diseases present in the

## rice leaves.

## Description of the algorithm:

**Image Acquisition**

The first step is to gather data from a publicly accessible repository. The picture is used as the  
input for further processing. We've chosen the most common image domains so that we can accept

any format as input to our method, including.bmp,.jpg,and.gif. The camera feeds the real-time images  
directly. Since most leaves colour varies from red to green for exact segmentation, a white  
background is provided for further study, proper visibility, and easy image analysis. Cotton images  
are captured using an image capturing system in this process. The picture is taken in such a way that  
any distortion is avoided. The photo was not taken in direct sunlight because it would distort the  
picture.

**Image pre-processing**  
The use of computer algorithms to perform image processing on digital images is known as image  
pre-processing. We can detect the plant by analysing the image with a specific algorithm. We  
use a similar approach for image processing and detection with a specific algorithm. The image  
quality is critical in this process; we can't use the algorithm if the image isn't clear.

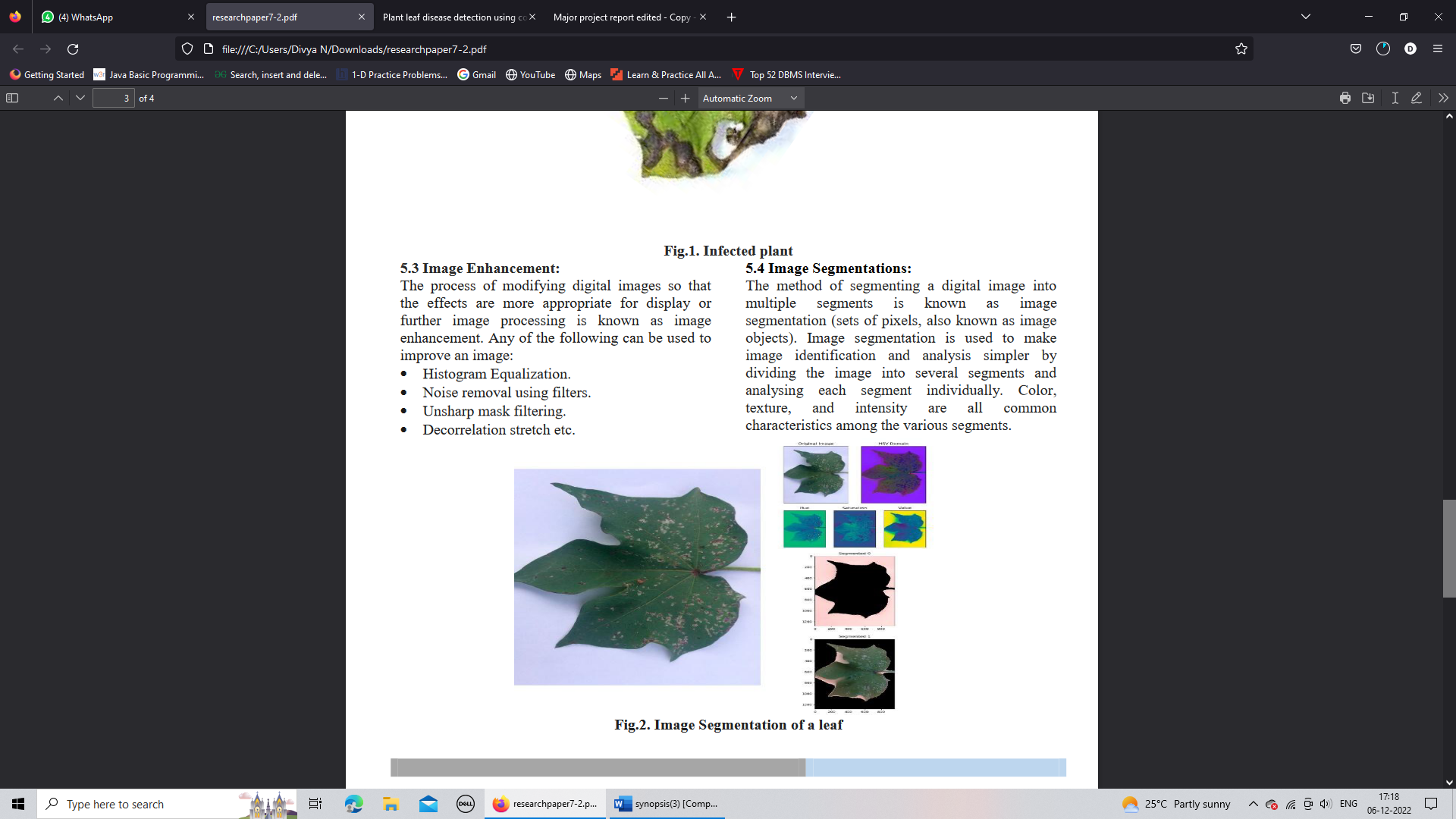
**Feature extraction**After segmentation the area of interest i.e. diseased part extracted. In the next step, significant features  
are extracted and those features can be used to determine the meaning of a given sample. Actually, image  
features usually includes color, shape and texture features. Currently most of the researchers targeting plant leaf texture as the most important feature in classifying plants. With the help of texture features, plant diseases are classified into different types. There are various methods for feature extraction as discussed below.

**Image Enhancement**

The process of modifying digital images so that the effects are more appropriate for display or  
further image processing is known as imag enhancement. Any of the following can be used to  
improve an image:  
• Histogram Equalization.  
• Noise removal using filters.  
• Unsharp mask filtering.  
• Decorrelation stretch etc

**Image Segmentation**

The method of segmenting a digital image into multiple segments is known as image  
segmentation (sets of pixels, also known as image objects). Image segmentation is used to make image identification and analysis simpler by dividing the image into several segments and analyzing each segment individually. Color ,texture, and intensity are all common  
characteristics among the various segments.

****

**Image Analysis**

In this step, image segmentation is used to locate the region of interest. The technique used in segmentation is region-based segmentation, which uses the colour of the leaf to distinguish between healthy and diseased regions of the plant leaf.

**Disease Classifications:**It is the method of using our qualified deep learning model to recognise plant disease. A  
digital camera or equivalent system should be used to take an image of the contaminated plant's leaf.Opencv was used to scan the image. Then it determines what kind of plant it is. It determines what kind of disease the plant has after finding it.

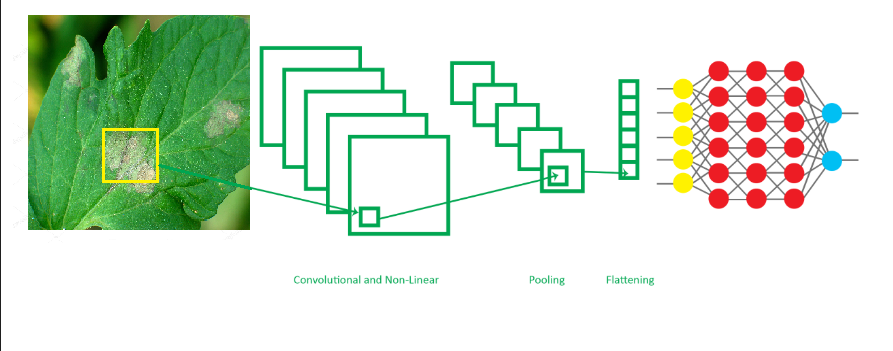


Fig.1 General Architecture of CNN

**CONVOLUTION:**

* + - * To add the convolution layer, we call the *add* function with the classifier object and pass in Convolution2D with parameters. The first argument *feature detectors* which is the number of feature detectors that we want to create. The second and third parameters are dimensions of the feature detector matrix.
      * We used 256 feature detectors for CNNs. The next parameter is *input shape* which is the shape of the input image. The images will be converted into this shape during pre-processing. If the image is black and white it will be converted into a 2D array and if the image is coloured it will be converted into a 3D array.
      * In this case, we’ll assume that we are working with coloured images. *Input shape* is passed in a tuple with the number of channels, which is 3 for a coloured image, and the dimensions of the 2D array in each channel. If you are not using a GPU it’s advisable to use lower dimensions to reduce the computation time. The final parameter is the activation function. Classifying images is a nonlinear problem. So, we use the rectifier
      * function to ensure that we don’t have negative pixel values during computation. That’s how we achieve non-linearity.
      * ***classifier.add (Convolution2D (256, 3, 3, input\_shape = (256, 256, 3), activation=’relu’))***

**POOLING:**

* + - * The Pooling layer is responsible for reducing the spatial size of the convolved feature. This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model.
      * There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel. Generally, we use max pooling.
      * In this step we reduce the size of the feature map. Generally, we create a pool size of 2x2 for max pooling. This enables us to reduce the size of the feature map while not losing important image information.
      * ***classifier.add (MaxPooling2D (pool\_size= (2,2)))***

**FLATTENING:**

* + - * In this step, all the pooled feature maps are taken and put into a single vector for inputting it to the next layer.
      * The *Flatten* function flattens all the feature maps into a single long column.
      * ***classifier.add (Flatten ())***

**FULLY CONNECTION:**

* + - * The next step is to use the vector we obtained above as the input for the neural network by using the *Dense* function in Keras. The first parameter is *output* which is the number of nodes in the hidden layer. You can determine the most appropriate number through experimentation. The higher the number of dimensions the more computing resources you will need to fit the model. A common practice is to pick the number of nodes in powers of two.
      * ***classifier.add (Dense (output = 64))***
      * The next layer we have to add is the output layer. In this case, we’ll use the *sigmoid* activation function since we expect a binary outcome. If we expected more than two outcomes, we would use the *SoftMax* function.
      * The *output* here is 1 since we just expect the predicted probabilities of the classes.

## *classifier.add (Dense(output=1, activation=’sigmoid*

## IMPLEMENTATION CODE:

## def convert\_img\_to\_tensor2(fpath):

## img = cv2.imread(fpath)

## img = cv2.resize(img, (180, 180))

## res = img\_to\_array(img)

## res = np.array(res, dtype=np.float16) / 255.0

## res = res.reshape(-1, 180, 180, 3)

## res = res.reshape(1, 180, 180, 3)

## return res

## Model Generation:

## Code:

**Chapter 6 : EXPECTED RESULTS**

Sample Input 1: Sample Input 2:

Expected Output:Bacterial leaf blight Expected Output:Brown Spot

Sample Input 3:



Expected Output:Leaf Smut

**Chapter 7 : CONCLUSION**

The proposed methodology in the following rice plant leaf disease detection system focus on generating an

advance and efficient system which makes the process of creating high yield of rice much easier for the

farmers. The project aims to detect the most common diseases occurring on a rice leaf, namely early blight,

bacterial spot , leaf smut, brown spot using image processing technique under upbringing technology i.e.,

machine learning.In easier terms, the farmer will be able to accurately detect the type of disease a particular

plant is having using the image of the plant. The proposed system is based on four important modules:

• Pre-processing.

• Segmentation.

• Feature extraction.

• Classification using CNN.

In this study, we classify the diseases present in the rice leaves.

**Chapter 8: BIBLIOGRAPHY**

[1] Ms. Kiran R. Gavhale1, Prof. Ujwalla Gawande , “Research on Plant Leaves Disease detection using Image Processing Techniques”.Jan 2004

[2] G. Geetha, S.Samundeswari, G.Saranya ,K.Meenakshi and M. Nithya, plant Leaf Disease Classification and Detection System Using Machine Learning”,2012

[3] Kowshik B,Savitha V,Nimosh madhav M,Karpagam G, Sangeetha K,” Plant Disease Detection Using Deep Learning”,Volume 3,March 2021

[4] Jashraj Karnik and Dr. Anil Suthar,” Agricultural Plant Leaf Disease Detection Using Deep Learning Techniques”,2021

[5] Bulent Tugrul , Elhoucine Elfatimi Recep Eryigit,”Convolutional Neural

Networks in Detection of Plant Leaf Diseases: A Review. Agriculture” 2022, https://doi.org/10.3390/agriculture12081192

[6] Sunil S. Harakannanavar , Jayashri M. Rudagi , Veena I Puranikmath , Ayesha Siddiqua ,R Pramodhini,” Plant leaf disease detection using computer vision and machine learning algorithms”, https://doi.org/10.1016/j.gltp.2022.03.016

[7] Jayshree Hajgude, Jayesh Kriplani, Dhiraj Chhabria, Anish Verliani,”plant leaf disease detection using machine learning”,Volume 9,