Minor Project (6th semester)

Paddy leaf Disease detection using CNN

Project Synopsis

Project synopsis submitted in partial fulfillment of requirements for the award of the degree of B.Tech. in Information Technology under Institute of Technology, Guru Ghasidas Vishwavidyalaya ,Bilaspur(C.G.)

Submitted By

Dharmadev Yadav	**Suryam Kumar**	**Shyam Kumar**
GGV/20/01417	GGV/20/01467	GGV/20/01461
20107017	20107066	20107060

Under The Guidance of

Dr. Rajesh Mahule

(Asst. Prof., Dept. of Information Technology)

DEPARTMENT OF INFORMATION TECHNOLOGY

SCHOOL OF STUDIES (ENGINEERING & TECHNOLOGY)

GURU GHASIDAS VISHWAVIDYALAYA

BILASPUR (C.G.), INDIA

CERTIFICATE

I hereby certify that the work which is being presented in B.Tech. minor Project report entitled "Paddy leaf Disease detection", in partial fulfillment of the requirements for the award of the Bachelor of Technology in Information Technology and submitted to the Department of Information Technology, School of Studies (Engineering & technology), Guru Ghasidas Vishwavidyalaya, Central University, Bilaspur (C.G.), India is an authentic record of my own work carried out during a period from Dec. 2022 to March 2023 (6^{th} semester) under the supervision of Dr. Rajesh Mahule, Assistant Professor, IT Department.

The matter presented in this project has not been submitted by me or any one else for the award of any other degree elsewhere.

Signature of Students

Dharmadev Yadav	Suryam Kumar	Shyam Kumar
(20107017)	(20107066)	(20107060)

This is to certify that the above statement made by the student is correct to the best of my knowledge.

Signature of Supervisor Date:

Dr. Rajesh Mahule

Assistant Professor

Dr. Rohit Raja

Head of Department (Information Technology)

ACKNOWLEDGEMENT

We would like to express our profound gratitude to Dr. Rohit Raja, HOD of Information Technology Department for providing this opportunity.

We would like to express our special thanks to our mentor Dr. Rajesh Mahule sir (assistant professor, Department of IT) for his time, efforts and guidance he provided throughout the semester. Your useful advice and suggestions were really helpful to us during the project's completion. In this aspect, we are eternally grateful to you. Our deep and sincere gratitude to Rajesh Mahule sir for his unflagging support and continuous encouragement throughout the project work, without his guidance and persistent help this report would not have been possible.

We would like to acknowledge that this project was completed entirely by us and not by someone else.

Dharmadev Yadav

E. No.: GGV/20/01417

Roll No.: 20104017

Suryam Kumar

E. No.: GGV/20/01467

Roll No.: 20107066

Shyam Kumar

E. No.: GGV/20/01460

Roll No.: 20107060

ABSTRACT

Plant diseases function as a significant threat to the food industry. The projected system helps in detection of crop diseases and provides remedies which might defend in contradiction of the crop infection. The information from the web is divided and also the totally different plant types are known and are relabeled so that we can create accurate information then get a sample database which consists of various crop diseases which will help in identifying the accuracy levels of the application . So, by a training dataset we will train our classifier so the production will be predicted with best truth fulness. We tend to practice the CNN that includes various layers that are used for prediction.

Keywords: Convolutional Neural Network, Tensorflow software, Image Processing technique, Android application, Paddy leaf disease detection, Features extraction.

Dharmadev Yadav	**Suryam Kumar**	**Shyam Kumar**
GGV/20/01417	GGV/20/01467	GGV/20/01460
Sem: 6th	Sem: 6th	Sem: 6th
Branch: IT	Branch: IT	Branch: IT

Table of Contents

1.INTRODUCTION

The primary occupation in India is farming. Here, agriculturalists cultivate a great diversity of crops. There are various factors such as climatic, soil and various diseases etc which affect the production of the crops reducing the ability of the crop to grow properly. Most importantly identifying the plant crop diseases and classification of it has become an easy process compared to the earlier days. However quick time, easiness and accuracy are the major factors to be kept in mind. Paddy crop disease identification and classification are dependent on the symptoms of that particular plant. Statistics show that in Asia 20% to 30% of rice crop per year is destroyed by the diseases caused by the plant. Fungi and Bacteria are the main reason for the cause of these diseases. So identification of diseases in paddy crops using digital images has been chosen as the best methodology.

Deep learning is a machine language that adapts self-learning techniques used on big amounts of data. The most recent developments in hardware and big data analytics have made this technique more useful and practical as it has a fast performing system. One of the algorithms in deep learning is CNN and it is used to analyze an image and pre-process it very easily along with the help of image processing technique. The CNN models provide association between layers and also spatial information of the image to make it easy for classifying the image.

There are a variety of methods for detecting plant diseases in their early stages. The traditional method of plant disease detection is naked eye monitoring, which is ineffective and inaccurate for large crops[2].The major goal of this paper is to research and diagnose rice leaf illnesses in advance, as well as to identify the disease's name so that appropriate precautions can be followed. Rice is a standout amongst the most vital food plantations [4] in our country, as well as one of the crops having a variety of purposes and high nutritional worth, with a production volume of 104.80 million tons coming from various Indian states. Because our country is the largest producer of rice at the second position in the world, the country's rice-growing region is constantly expanding. It contains a high amount of carbohydrates and protein, as well as a significant amount of dietary fiber and minerals. Plant illnesses are caused by pathogens, fungus, bacteria, viruses, and other microbes in the majority of cases. Rice leaves are sensitive to diseases that are caused by fungus, viruses, and the varied field environment makes it simple for pathogens to infect the leaves[5]. Figure 1 shown below is an example of different diseases of rice leaf.

Climate changes [7] will create an ideal environment for those pathogens to thrive. The growth of crops is hampered in their initial stages due to fungi-caused illnesses. If illness strikes while the crop is still growing, it might reduce the crop's yield.

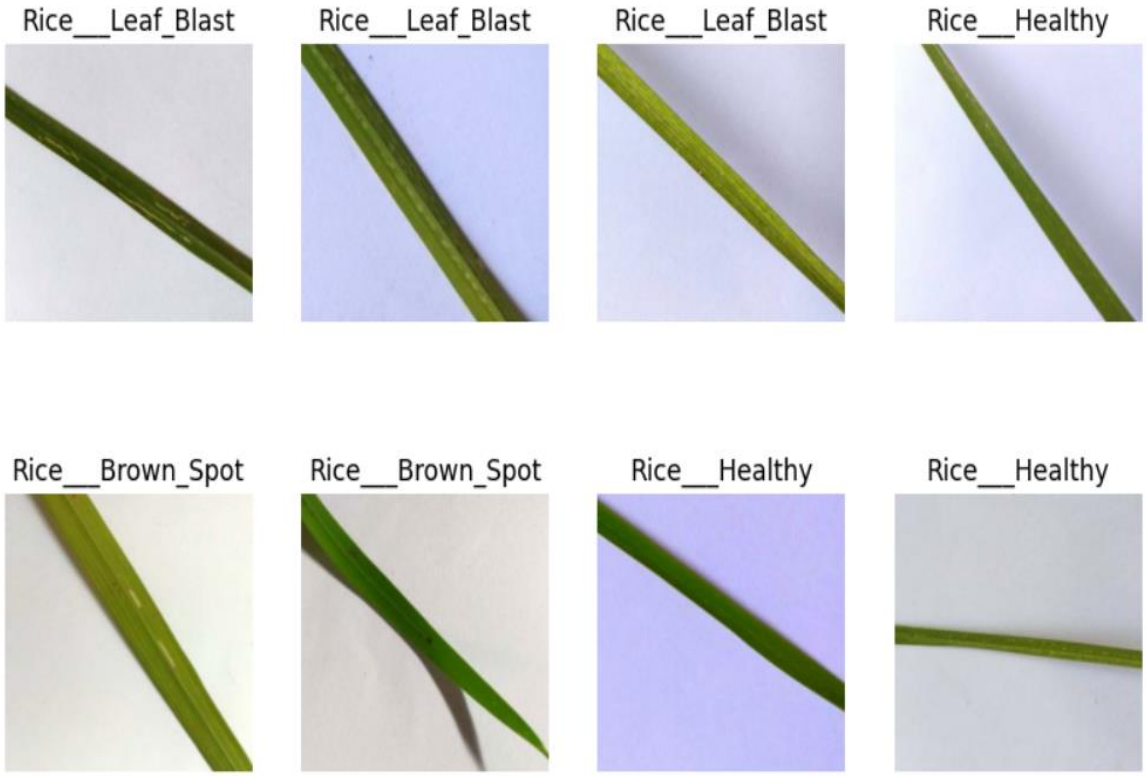

Fig.1.1. Different type of Paddy leaf Diseases.

Manually determining the presence of illnesses in large agriculture regions is quite challenging. Diseases, particularly in rice plants, have become a problem since farmers are unable to identify leaf disease with the naked eye, and they must consult the expert in order to discover that specific disease, which takes more time and requires much expense. The most frequently occurring diseases in rice leaf are Brown spot, LeafBlast, Healthy.

Thus, disease detection in leaves is an important topic that provides many benefits in monitoring large fields of crops. Rice leaf disease can affect yield and quality by damaging the green layer from the leaves. The way to control these rice diseases is to rapidly and precisely detect the disease type and then implement appropriate corrective actions in a timely manner[8]. Using digital image processing techniques and deep learning networks, the detection of disease is efficient, consumes less time, and is accurate. Advances in Computer vision offer an opportunity to extend and increase plant protection[9].

2.LITERATURE SURVEY

Image analysis techniques for measuring rust sickness found on soybean leaves. The methodology used is division of contaminated parts from multi-phantom pictures of soybean plant leaves which are finished utilizing quick manual edge setting strategy in light [1-2, 22].

A Convenient application for paddy illness with recognizable proof structure demonstrating Fluffy entropy and Probabilistic neural framework classifier that continues running on Android which is a flexible working operating system. It incorporates forms of ailments specifically darker spots, leaf impact, tungro and microorganism leaf curse [3,13,21].

A conventional surveys on different methods of image processing applications in the agriculture field. The main focus of the area of their survey was in two streams, one is weed detection and the other one is fruit grading systems. Weed are the dangerous crops grown along with the main crops so they found them out by applying image processing techniques and collecting the images of the particular crops [4,17,26].

Entire summary on disease classification in plant crops using image recognition techniques. These strategies are expected to be valuable for scientists giving a far detailed reaching diagram of vegetable pathology and also programmed discovery of plant ailments utilizing design acknowledgment systems [5,12,16].

The main essential paddy infections. His research was more precise in finding out the leaf impact formed because of spots, bacterial infections, leaf strength and many other factors. Shape patterns are used to dismember the piece of the injuries [6]. Discussed various data mining techniques for paddy crop disease prediction and classification [7,15].

Image Edge detection and Segmentation techniques. Initially the taken pictures are processed for enrichment at the initial stage of the process. R, G, B color, color, shape, boundaries, texture etc are the features which are extracted out from the target regions (disease spots) and later on other steps of the process continue to follow. This research includes the pest recommendations and also remedies for the diseases identified. This analysis work had a lot of great qualities but if failed to show the remedies and also could not reach the needed accuracy level [8,10,14].

A process that has only eyesight and needs precise observation and additional scientific methods and technologies. Here the images of unhealthy leaves of that particular plant are

captured and the features (Hue, Saturation, and Value) are extracted once the segmentation phase is completed. ANN is trained to distinguish the healthy and unhealthy plant samples and gives 80% accuracy [19,23,24].

A model that explains about the images which are captured and stored in the database by mobile phones or any digital application. After this stage these images are analysed and judged by the senior experienced persons and a final conclusion is passed out. Computer vision techniques help in recognising the diseased places in the images and then undergo classification process. Color difference process which is a primary approach is used for the segmentation of the not well plant areas [27,20]. The system allows the conclusion of the senior persons and also sends the feedback to the farmers through any means of digital communication. The main goal of this proposal is to develop an image recognition system that can identify the crop diseases and help out the agriculture field and also the future generations from food crisis and economic downfall. Image processing is a picture identification system which helps in the digitization of the color image. Arithmetic science is the primary logic behind this segmentation process. Then after a little more research it is found that this process depends on mankind (drawback of many other proposals) and also the time taken to complete this process is much more than many other existing and upcoming methods [11,15,18,25].

In summary, various image processing techniques have been proposed for plant disease detection. Yet none of them proved to be efficient for identification of all kinds of paddy crop diseases.

3.PROPOSED MODEL

Unlike any other existing paddy crop disease detection methods, we have developed a model which runs in a very less time and also gives more accuracy along with the remedies for the disease detected in the paddy crop leaves. In this method we are using Mat lab software as the platform to perform the process.

3.1 ALGORITHM

The algorithm for this proposed model is explained in the following steps:

1. Collecting the database and segregating it according to the features of the plant.

2. Taking an image as the test image/ input image.

3. Conversion of the image in to arrays and storing the data.

4. Training the images from datasets for the CNN classification.

5. Comparison of these both test and train images.

6. After comparison, all the features are extracted.

7. Diseases in the plant are found if there are any, remedies and the accuracy is given

3.2 ARCHITECTURE

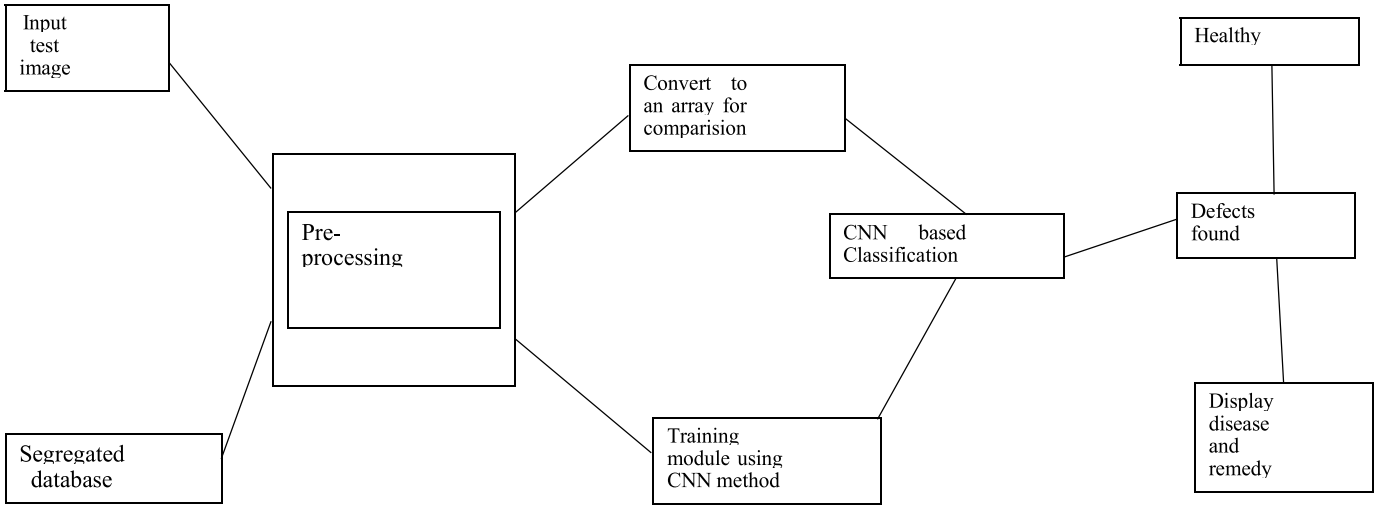

Fig. 3.2.2. Architecture of the System

4.MODULES

There are 4 modules in this entire process. They are:

a. Database
b. Data pre-processing
c. CNN classification
d. Comparison of test and train images

a. Database

An image database means storing high quantities of digital images in a particular location. It also means organizing photos so that they can be shared, accessed quickly

and easily. In this project we have taken large number of paddy leaf images into datasets.

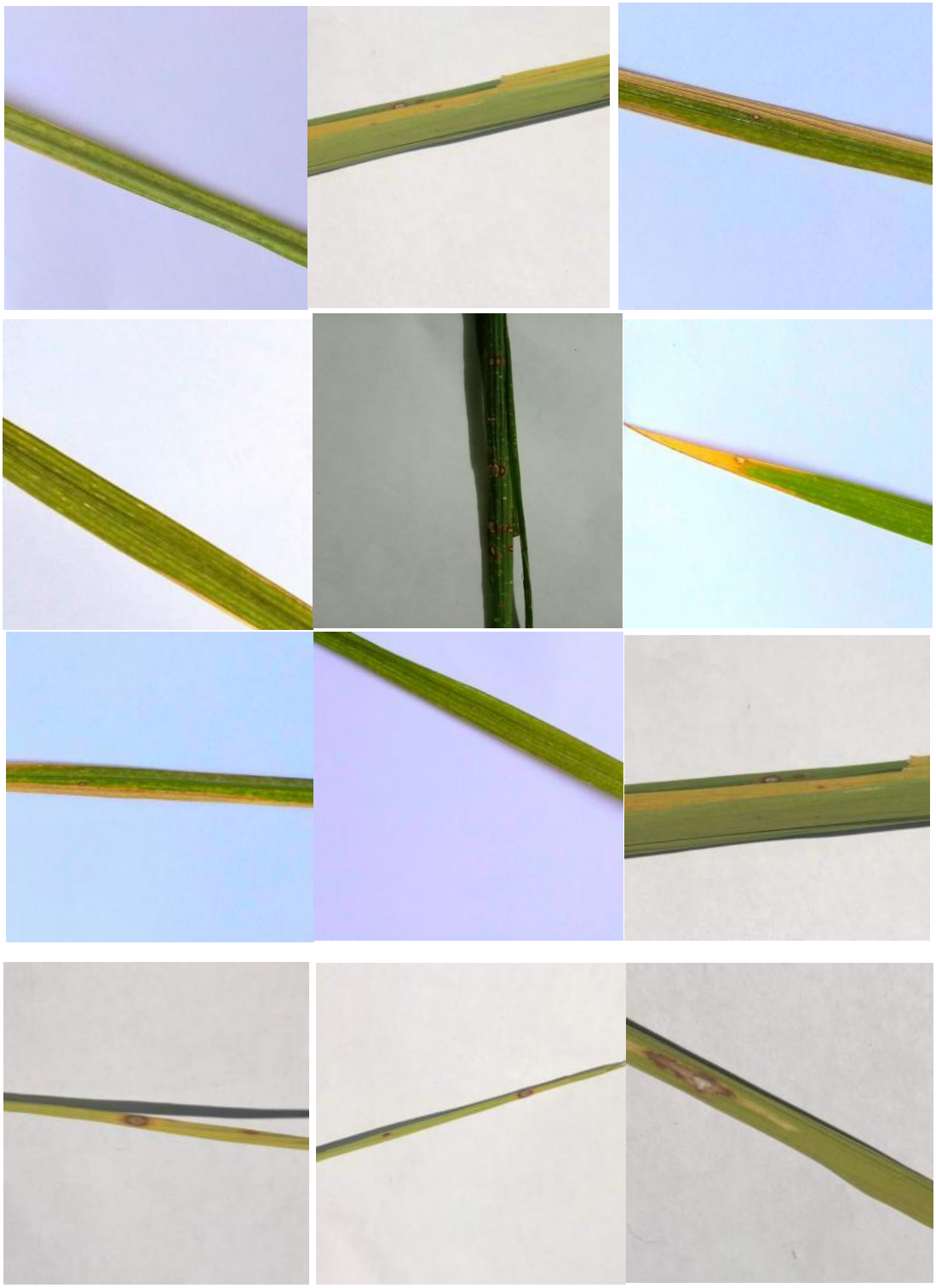

Fig. 4.a.3. Different leaves images

b. Data pre-processing

In our present world scenario[11], input data involves a lot of noise, has missing values, outliers, and is inconsistent. Data preprocessing involves removal of the noise, missing data, and organizing data in a proper format so that accuracy is increased. It enhances the quality of the data.

This step involves data cleaning, data transformation, and data reduction (data compression) .

Data cleaning involves cleaning the data. It removes the noise present in the data. Data transformation is transforming high-level data into low-level data for easier calculations.Data reduction involves reducing the data dimensions so that the data is not high dimensional but the quality of data remains the same. Data cube aggregation which is summarizing the data.

c. CNN classification

A CNN is a higher level language of deep learning. CNN models are trained using large collections of images taken across different sources. From these large collections of databases CNN models can learn rich feature extraction and presentation for a wide range of images. CNN is considered as one of the best techniques used for features extraction.

CNNs are typically composed of three main facts namely convolution layer, pooling layer, and dense layer. A CNN consists of multiple convolution and pooling layers that are followed by dense layers. The convolution and pooling layers can be arranged in several different ways; and their arrangements are conventionally based on the complexity of the problem. The typical final dense layer's dimension is equal to the number of output classes

Our architecture mainly contains the following layers

1. Convolution layer
2. Pooling layer
3. Fully connected layer

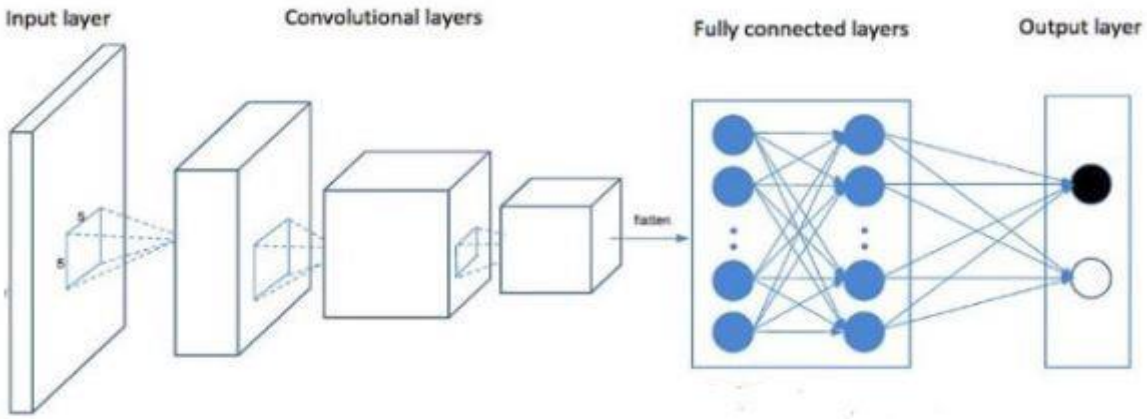

Fig.4.c.4. Layers of CNN

The above figure represents the working of CNN. The input in the form of the image after preprocessing the data and extracting the required features when passed through CNN passes through 3 layers of CNN and it is precisely represented. The final output is then displayed.

● **Input Layer:** The input layer of CNN consists of the dataset. The input data will be represented as a 3X3 matrix.

● **Convolution Layer:** A layer that uses filters to learn from smaller sections of input data to obtain features from an image.

● **Pooling Layer:** This layer is used to shrink the image's dimensionality, lowering the processing power required for subsequent layers. There are two variations of pooling. They are:

● **Max pooling:** The pixel with the maximum value as input is selected and transferred to the output while parsing input. It is the most used approach compared to average pooling.

● **Fully Connected Layer (Dense):** This is one of CNN's last layers, and it can recognize features that are significantly linked with the output class. The result is a one-dimensional vector created by flattening the pooling layer results.

● **Dropout Layer:** Used to reduce model overfitting problem by removing a random set of neurons in that layer. It is connected with the FC layer.

● **SoftMax Layer:** This is the network's last layer that assists in classifying individual input images of the dataset into several classes depending on the learned properties from the network.

● **Output Layer:** The output layer holds the final classification result.

d. Comparison of test and train images:

The final stage of the process is finding whether the plant is healthy or not healthy. If there are any defects found in the image then the output will be given as yes along with the remedies to be followed for eradicating the disease. This happens when the both test and train images are brought together and are compared. The accuracy of the problem found is more when this is executed.

5.EXECUTION AND RESULTS

To get the exact and accurate results we are using the Tenserflow software library for the execution. Android is a operating system for mobile devices,such as small phone ,Tablets by google, which allows devise manufectures to costomize the operating system their needs is used for technical computing. Tenserflow software library is designed to perform various operations like computation, visualization of image, associated programming in all kinds of surrounding environments. It also involves high-level commands for 2-dimensional and 3-dimensional image visualization, image processing, presentation graphics and animation process which is one of the biggest assets of this software.

The below screenshot presented the training process. Whereas training and validation accuracy were shown in Fig 5.5, where the y-axis shows the accuracy obtained after each iteration represented in the x-axis. Similarly, Fig 5.6 contains the training and validation Accuracy and loss, where the y-axis shows the loss (in percentage) when training started and thereafter increase or decrease in loss after each iteration represented in the x-axis.

```
scores=model.evaluate(test_ds)

52/52 [==============================] - 1s 22ms/step - loss: 0.1360 - accuracy: 0.9459

[ ] scores

[0.1360405683517456, 0.9459134340286255]
```

Fig.5.5. Test Accuracy of Paddy leaf diseases.

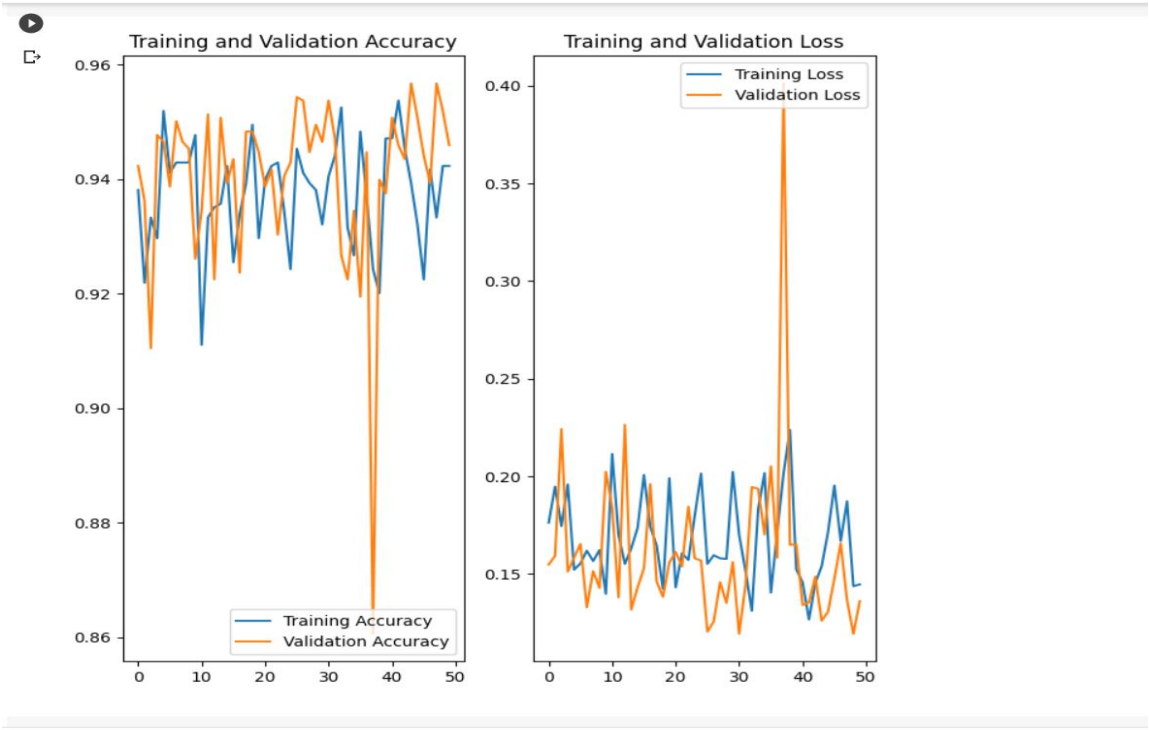

Fig.5.6. Graph of Training and Validation Accuracy and Loss.

The steps involved in the execution of our model are given below respectively

Step1 - Take an image from the dataset and give it as an input image.

Step2 - Now enter the number of the particular cluster which you want to test.

Step 3 - After this the images from the datasets are setup in the training phase which later gives us segmented image, b/w image and query leaf image.

Step 4 - In this last step, we finally get the output of the process with the disease name and the remedies to be taken to eradicate the disease. We will also get the accuracy of the disease detection.

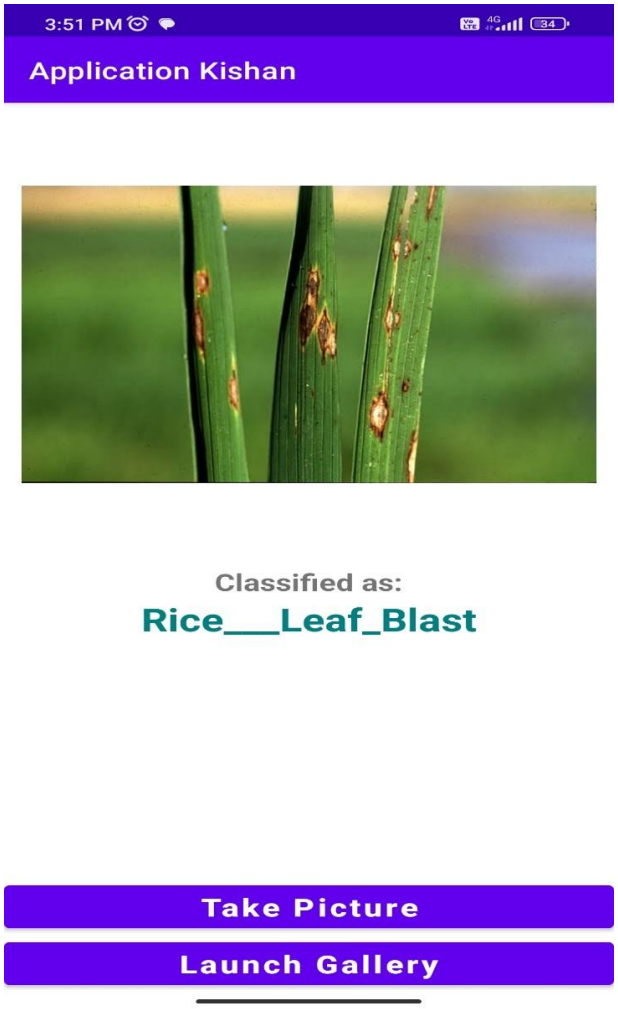

Fig.5.7. Output display in Application Kisan

6. CONCLUSION

We proposed this system to help the farmers in gaining productivity and also improvising agriculture systems. This methodology can detect the diseases in paddy crops and along with this it also specifies the remedies that can be used to control that particular disease. By this we can also start using new and smart technologies in the agriculture field which may be

bring a huge change in the Indian economy. This system is based on python language libraries and is giving an accuracy around 94% to 96%.

The Paddy leaf disease images are used for simulation purposes are collected from Kaggle. For detection of Leaf disease in paddy leaf using dataset images based on deep learning features.

The better we can detect infections, the simpler it will be for farmers to protect their crops. In the future, we will broaden the scope to include more diseases and algorithms, making disease detection vast, easier and faster.

7. REFERENCES

1. Cui D., et.al.,"Image Processing Methods for Quantitatively Detecting Soybean Rust from Multi-spectral Images", Biosystems engineering, Vol.107, Issue.3, 2010, pp.186193.

2. Vibhute A. and Bhode S K, "Applications of Image Processing in Agriculture: A Survey", International Journal of Computer Applications, Vol. 52, Issue. 2, August 2012, pp.34-40.

3. Barbedo J C A, "Digital Image Processing Techniques for Detecting, Quantifying and Classifying Plant Diseases", SpringerPlus, Vol.2, Issue.660, 2013, pp.1-12.

4. AuziAsfarian, et.al.,, "Paddy Diseases Identification With Texture Analysis Using Fractal Descriptors Based On Fourier Spectrum", International Conference On Computer, Control, Informatics And Its Applications,2013.

5. S.A. Ramesh Kumar, K.Ramesh Kumar, "a Study on Paddy Crops Disease Prediction Using DataMining Techniques", International Journal of Data Engineering (IJDE), Singaporean Journal of Scientific Research (SJSR), Vol.7, No.1, Pp. 336- 347, 2015.

6. P.Revathi, M.Hemalatha, "Advance Computing Enrichment Evaluation of Cotton Leaf Spot Disease Detection U sing Image Edge detection", ICCCNT'12.

7. Malvika Ranjan, et.al., "detection and classification of leaf disease using artificial neural network", International Journal of Technical Research and Applications, 2015

8. Heeb Al Bashish, et.al., "A Framework for Detection and Classification of Plant Leaf and Stem Diseases", IEEE 2010.

9. R. Mrunalini, “developer of an application of K-means clustering and artificial intelligence in pattern recognition for crop diseases”, 2011.

10. Prajapati, H. B., (2017).“Detection and classification of rice plant diseases”, (2017).

11. S. Raj Kumar, S. Sowrirajan, “Automatic Leaf Disease Detection and Classification using Hybrid Features and Supervised Classifier”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 5, Issue 6, 2016.

12. Surbhi Jain, Joydip Dhar, “Image based search engine using deep learning”, 2017 10th International conference on contemporary computing(IC 3), August 2017.

13. G.Balakrishna and MoparthiNageshwara Rao “ Study report on Indian agriculture with IoT “International Journal of Electrical and Computer Engineering.

14. Jin Wang, Yu Gao , Kai Wang , Arun Kumar Sangaiah , Se-Jung Lim, “An Affinity Propagation based Self-Adaptive Clustering Method for Wireless Sensor Networks”, Sensors, 2019, 19(11), 2579. /10.3390/s19112579.

15. Mekala, M.S., Viswanathan, P. “(t,n): Sensor Stipulation with THAM Index for Smart Agriculture Decision-Making IoT System”. Wireless Pers Commun 111, 1909–1940 (2020).

16. G.Balakrishna and MoparthiNageshwara Rao,” ESBL: Design and Implement A Cloud Integrated Framework for IoT Load Balancing” International Journal Of Computers Communications & Control ISSN 1841-9836, e-ISSN 1841-9844, 14(4), 459-474, August 2019.

17. Yuyu Yin, Lu Chen, Yueshen Xu, Jian “Wan:Location-Aware Service Recommendation With Enhanced Probabilistic Matrix Factorization”. IEEE Access 6: 62815-62825 (2018).

18. Jin Wang, Yu Gao, Wei Liu, Arun Kumar Sangaiah, Hye-Jin Kim, “An Improved Routing Schema with Special Clustering using PSO Algorithm for Heterogeneous Wireless Sensor Network”, Sensors, vol.19, no.3, Feb. 2019.

19. Jin Wang, Yu Gao, Wei Liu, Wenbing Wu, Se-Jung Lim, “An Asynchronous Clustering and Mobile Data Gathering Schema based on Timer Mechanism in Wireless Sensor Networks”, Computers, Materials & Continua, Vol.58, No.3, pp.711-725, 2019.

20. Anita Gehlot,e.al.,"IoT and Zigbee based Street Light Monitoring System with LabVIEW", International Journal of Sensor and Its Applications for Control Systems, SERSC Australia, vol.4, no.2, November (2016), pp. 1-8.

21. Navulur, S., Sastry, A. S. C. S., &Giri Prasad, M. N. (2017). “Agricultural management through wireless sensors and internet of things”. International Journal of Electrical and Computer Engineering, 7(6), 3492-3499.

22. Qinli An, "Localization of Sound Sensor Wireless Network with a Mobile Beacon", International Journal of Sensor and Its Applications for Control Systems, SERSC Australia, ISSN: 2287-8467 (Print); 2207-6344 (Online), vol.6, no.1, May (2018), pp. 1-14.

23. Jin Wang, Chunwei Ju, Yu Gao, Arun Kumar Sangaiah, Gwang-jun Kim,” A PSO based Energy Efficient Coverage Control Algorithm for Wireless Sensor Networks”, Computers Materials & Continua, vol.56, no.3, pp.433-446, 2018.

24. Chunyong Yin, , “Location Privacy Protection based on Differential Privacy Strategy for Big Data in Industrial Internet of Things”, IEEE Transactions on Industrial Informatics, vol.14, no.8, pp.3628-3636, Aug. 2018.

25. Jianming Zhang, You Wu, XiaokangJin, Feng Li, Jin Wang, “A Fast Object Tracker Based on Integrated Multiple Features and Dynamic Learning Rate, Mathematical Problems in Engineering”, vol.2018, pp.1-14.

26. Jiang,e.at., “VRAA: virtualized resource auction and allocation based on incentive and penalty. Cluster Computing”The Journal Of Networks Software Tools And Applications, 2013, 16(4): 639-650.

27. Jiang, e.at.,“Resource Allocation in Contending Virtualized Environments through Stochastic Virtual Machine Performance Modeling and Feedback”. Journal Of Information Science And Engineering, 2013, 29(2):299-327.