RICE PLANT DISEASE DETECTION USING MACHINE LEARNING

Shital Patil[1], Harshali Ragite[2], Aasha Sangole[3]

[1] Assistant Professor, Computer Science Department, Wainganga College of Engineering and Management, Nagpur, India
[2] Assistant Professor, Computer Science Department, Wainganga College of Engineering and Management, Nagpur, India
[3] Student, M. Tech, Computer Science Department, Wainganga College of Engineering and Management, Nagpur, India

ABSTRACT

Plant diseases have a very serious affect on the farming industry. As a result of this, there is a bad impact on the productivity of the crops. This leads to huge losses to farmers. To ensure better quality, quantity and productivity of the yield, it is very crucial for identifying the diseases at early stage for reducing the use of pesticides to reduce damage of the crops and environment. In this research our aim was to detect and classify the rice plant leaf diseases, divided into four categories of classes as healthy, hispa, brown spot and leaf blast. We have used convolutionl neural network for the feature extraction from the rice images. Along with i , some machine learning classifiers such as K-Nearest Neighbors and Random Forest were used for the classification of the diseases based on the categories. The CNN Model performed well for the feature extraction with the accuracy of 80 percent. Along with this second model was classification of diseases using some machine learning classifiers such as Random Forest and K-Nearest Neighbors, accomplished the accuracy of 96% and 72% respectively.

Keyword : - *Deep Learning, Machine Learning, CNN, Random Forest, KNN, IOT, Image Processing*

1. INTRODUCTION

Agricultural sector plays a very important role for economic development of any country. In terms of raw materials, the majority of countries are dependent on agricultural goods. Rice is a crop which is mostly cultivated 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 climate shifts. The agricultural sector is suffering as a result of these changes in 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. Different forms of illnesses are wreaking havoc on rice fields these days, having a negative impact 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. These diseases can be found on any part of the rice plant, including the leaf, neck and ear. Plant diseases have a negative impact on not just agriculture but also on the environment in terms of pollution. Plant diseases are responsible 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. 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. Advancement in IT field, help to increase productivity of the crops with the minimum use of fertilizers. In today's Artificial Intelligence (AI) environment, convolutional neural networks (CNN) and machine learning can play important role in classifying diseases.

The Image Processing domain may be able to provide a solution to the agricultural sector's challenges. CNN will be used for extracting features from the leaf images. Diseases can be classified using machine learning classifiers. CNN has a faster processing speed and is more accurate in categorization. This approach could be useful for disease classification and detection. In this study, a convolutional neural network is utilized to extract features from rice photographs. For classifying diseases, machine learning classifiers such as random forest and K-Nearest Neighbors are utilized.

2. PROPOSED METHODOLOGY

The objective of this research is to build a deep learning and machine model which will detect and classify the diseases in rice crop using the database. In this research Convolutional Neural Network (CNN) is used to extract the features from the rice leaf images. Further some machine learning classifiers such as Random Forest and K-Nearest Neighbors will classify the disease in which class it belongs. This is kind of hybrid model.

For this implementation, we have used KDD Methodology and the figure 1 shows the process flow of KDD approach.

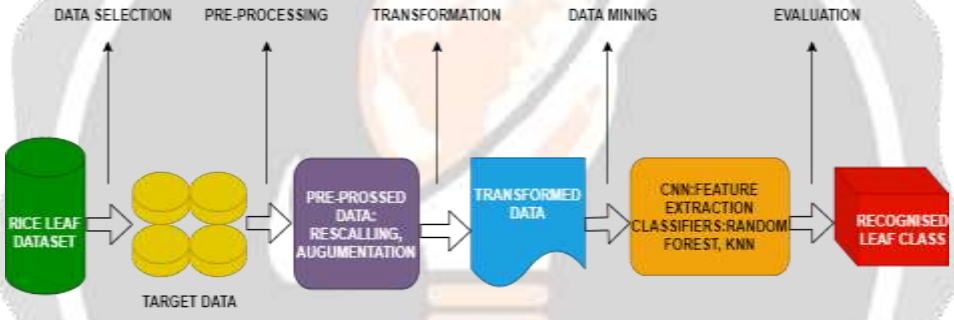

Figure -1: Proposed Methodology

2.1 Data Selection

The dataset for rice leaf diseases was obtained from the web and is available on kaggle.com. There are 3355 photos in the collection, and are in.jpg format. The dataset is 7 GB in size due to a high resolution photos. Further data set was divided into training and validation for each and every class. In total 2684 images are kept for training and 671 images were kept in validation phase. This dataset was generated for the aim of study and can be used to identify and classify rice leaf diseases. This data selection method completes the image acquisition process.

2.2 Data Preperation

This section provides steps involved in pre-processing and transformation of data in this research work. Data Pre-processing involves working on database before the implementation phase. For any research work it is crucial to pre-process the data before using for any modeling purpose. In this research project normalization is done on images to use the images in proper format which will help to do modeling faster.

In data transformation process image augmentation is carried out. Image augmentation is one of the most popular method of data augmentation, can be used to increase the size of training by using resized version of images in dataset. Image augmentation process comprises of transforming the pictures which are present in the training data. In this research work image augmentation for CNN is done using image data generator.

2.3 Data Modeling

Data Modeling gives explanation about the algorithms and models that were used to classify the pictures. This research work shows the usage of both classic machine learning and deep learning methods. Convolutional Neural Network

The convolutional neural network model works well with the image processing and gives better accuracy and builds an efficient model.

The features retrieved from the leaf photos are represented by an embedding, which is a vector. The vectors created for other leaf pictures can then be compared to this. Another vector that is close to the first could be the same leaf class, while another vector that is remote could be a different leaf class. The classification model will accept an embedding as inputs and predict the leaf's identification.

Random Forest classification algorithm is implemented for classification of rice leaf im- ages. This classification algorithm is used because it give better accuracy with the CNN model

K-Nearest Neighbors classification algorithm is implemented for classification of rice leaf images. This algorithm is used because it performs better in classification task.

3. DESIGN AND IMPLEMENTATION

The figure 3 shows the process flow on the implementation.

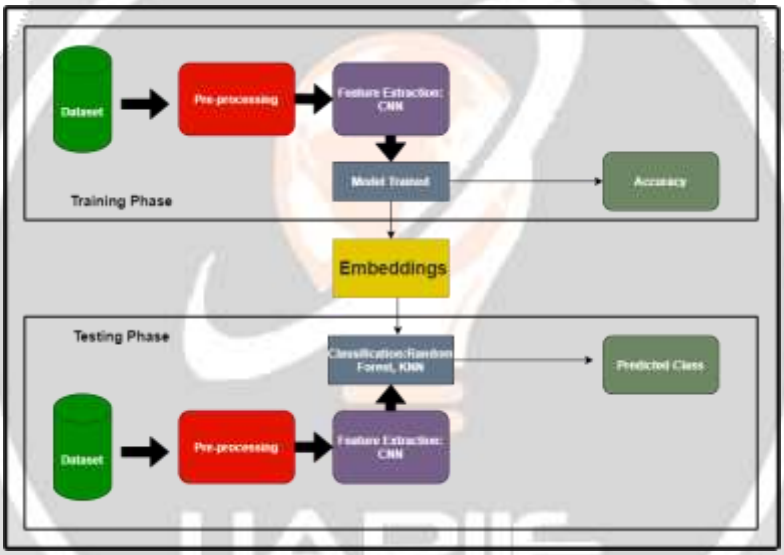

Figure - 2: Architecture Design

Initially data was imported and further pre-processing and transformation is done so it can used used for model training. Pre-processing and normalization is very necessary before applying any model on data. Further the results of this is transformed data. Transformed data was split in two for training and testing purpose. Same process has been done on both train and test data. This data then further passed to the CNN model and feature from the images were extracted. Next embbedings process was done to obtained feature vector using array. The feature map from embbedings was compared with the features from the CNN model and further it store in the labels i.e four categories available in the dataset. Further this feature map is given as an input to the classification model such as Random Forest and KNN. This two classifiers will classify the images according to the classes present and will show output with the evaluation measures.

3.1 Implementation

In this study, Python version 3.8 is used for the implementation of the project work along with the spyder as IDE (Integrated Development Environment). CNN model takes long time to train the model. So, system configuration and GPU should be good and there should be proper installation of all the libraries before training the model and to execute the model smoothly. Various librabries were installed such as TensorFlow, Keras, Scikit- learn, Matplotlib, Pillow and Opencv. TensorFlow is one of the most popular library which is used in image processing to build an efficient

model. Keras is one of the most powerful library for developing deep learning models. It Also acts as an interface for TensorFlow library. Scikit-learn is used for machine learning algorithms for classification and regression tasks. Matplotlib is use for plotting in python programming language. Pillow and Opencv is used for working on image data. All these libraries were installed to build an environment for the actual implementation part.

From the the first model i.e. CNN extraction of features is done from the images. All the necessary libraries for this model building were imported from TensorFlow library using keras. The libraries were installed using pip install command. Second step is embbedings which will create an array and try to extract features from the images and further those features will be compared with the feature obtained from the CNN model and it will store the feature vector according to the particular categories. Second model is a classification of images based on the categories. There are two classification models which are implemented such as Random Forest and KNN. These classification models will take the output i.e. feature vector as an input and then it will classify the images.

4. RESULT ANALYSIS

The models and architectures developed in this study attempt to alleviate the difficulties experienced by farmers in diagnosing and classifying rice leaf diseases. As a result, CNN is used to demonstrate the performance and robustness of the models developed in comparison to other machine learning algorithms such as Random Forest and K-Nearest Neighbors. The accuracy, recall, precision, and F-1 score were used as evaluation measures in this study to analyze the outcomes. Then, using the negative and positive classifications of each class, a confusion matrix is generated to determine the right amount of predictions made by model for identifying rice leaf diseases. The performance of the models developed is evaluated using this matrix and the classification report. The table bellow summarizes the accuracy findings for all of the models used in this investigation.

Table 1: Results Summary

Model	Accuracy
CNN	80%
Random Forest	96%
K-Nearest Neighbors	72%

The above table represents the result summary for all the models which are implemented in this research work. In this research project first model CNN performed quite well with the accuracy of 80% for feature extraction from images. Whereas, second classification model contains two classifiers such as Random Forest and K-Nearest Neighbors. Accuracy of random forest is 96 percent, which is very good. And the accuracy of the KNN classifier is 72 percent. Among both the classifications model Random Forest performed well with the CNN model as an feature extractor. Whereas, KNN have not given the best performance with the CNN model.

5. CONCLUSIONS

To ensure better quality, quantity and productivity of the yield, it is very crucial for identifying the diseases at early stage for reducing the use of pesticides to reduce damage of the crops and environment. The main aim of this research was to detect and classify the diseases in the rice leaf, having four categories of classes as healthy, hispa, brown spot and leaf blast. In this research study convolutionl neural network was used for the feature extraction from the rice images. Whereas, Some machine learning classifiers such as Random Forest and K-Nearest Neighbors were used for the classification of the diseases based on the categories. The implementation was done on total 3,355 images of rice leaf. The first model CNN performed well for the feature extraction with the accuracy of 80 percent. Along with this second model was classification of diseases using some machine learning classifiers such as

Random Forest and K-Nearest Neighbors, accomplished the accuracy of 96% and 72% respectively. In this research project Random forest achieved better accuracy than KNN with the CNN model.

In future, accuracy can be improved for CNN model to get more good results and also for KNN model. This can be achieved with more good quality of image database. Various image enhancement techniques and also increase or decrease in number of epochs can be done for checking the improvements in the model performance.

6. ACKNOWLEDGEMENT

My heartfelt gratitude to our college, Wainganga College of Engineering And Management, and, department of Computer Science And Engineering for allowing me to successfully complete this research project. Gratitude to my guide Prof. Shital Patil ma'am and co-guide Harshali Ragite Ma'am for their insightful supervision.

7. REFERENCES

[1]. Atila, Ü, Uçar, M., Akyol, K. and Uçar, E. (2021). Plant leaf disease classification using efficientnet deep learning model, *Ecological Informatics* 61: 101182.

[2]. Azadbakht, M., Ashourloo, D., Aghighi, H., Radiom, S. and Alimohammadi, A. (2019). Wheat leaf rust detection at canopy scale under different lai levels using machine learning techniques, *Computers and Electronics in Agriculture* 156: 119–128.

[3]. Bao, W., Zhao, J., Hu, G., Zhang, D., Huang, L. and Liang, D. (2021). Identification of wheat leaf diseases and their severity based on elliptical-maximum margin criterion metric learning, Sustainable Computing: Informatics and Systems 30: 100526.

[4]. Chouhan, S. S., Singh, U. P., Sharma, U. and Jain, S. (2021). Leaf disease segmentation and classification of jatropha curcas l. and pongamia pinnata l. biofuel plants using computer vision based approaches, Measurement 171: 108796.

[5]. Hu, G., Wang, H., Zhang, Y. and Wan, M. (2021). Detection and severity analysis of tea leaf blight based on deep learning, *Computers & Electrical Engineering* 90: 107023.