

Review Article

Detection of Rice Blast Disease (*Magnaporthe grisea*) Using Different Machine Learning Techniques

Abstract

Rice is one of the most important staple food crops in the world. Most Asian countries are dependent on rice and huge quantities of rice are grown every year. However, there are many categories of diseases (e.g., blast) which affect rice production and can ultimately lead to huge financial loss to rice growers. Yield loss due to rice blast disease about 10 to 30 percent annually and under favourable condition, this disease can destroy the rice plant within 15 to 20 days and cause yield loss up to 100%. Therefore to ensure better quality, quantity and better productivity early disease detection should be done so that the right amount of pesticides can be administered at right time to curb the infection. Nowadays Machine Learning has been integrated into the agriculture sector. The aim of this review paper is to identify which Machine Learning algorithms work best in rice blast disease detection. The algorithms reviewed here include Naive Bayes, LSTM RNN, Random Forest Classifiers, Support Vector Machines, K Means, Decision Tree and Convolutional Neural Networks. This review paper also covers the future scope of improvement of some Machine Learning algorithms like Naive Bayes and Recurrent Neural Networks.

Keywords: Rice, *Magnaporthe grisea*, Detection, Machine learning, Algorithms.

Introduction

Rice (*Oryza sativa* L.) is the second most important staple crop for more than half of the world's population, providing more than 3.5 billion people with more than 20% of their daily caloric needs [1]. In addition, rice has also been growing for many years in Europe, primarily in the Mediterranean nations of Italy, Spain, Greece, Portugal, and France [9]. Today,

agriculture is the primary industry everywhere in the globe and is essential to the international economy but agricultural production is affected by many diseases and pests. The cost of rice cultivation has increased due to cost of fertilizers and pesticides under current climate change scenario. Early disease detection in field is a key to sustain production and lower financial loss. One effective solution for farmers is processing the captured images of “seem to appear” infected leaves by an automated system. Machine learning is being used to identify various types of crop diseases and damages in rice [1]. Numerous funding and research projects are constantly being implemented using the newest technologies [2]. In Asia 10 to 15% of production is destroyed because of rice diseases. [2]. The usual practice in a farmer’s fields is an expert visually monitor plant diseases which requires more effort and long processing time. Farmers need more crop output, automated crop-related tasks which will reduce the need for human labour and lower the disease ratio by spotting early stages with already-existing matching patterns, and eventually harvesting with machinery in the shortest amount of time [3]. From planting to harvesting, machine learning is used in agriculture employing a variety of methods from several fields of technology, including big data, artificial intelligence, drones, and data mining [4 and 39]. Using these technologies, for mapping with existing data to identify and fix the solution of frequent problems. For the purpose of identifying crop diseases, parameter values of machine learning models can also be manually assigned [5]. It has been discovered that utilizing a probabilistic neural network for the categorization of illnesses affecting commercial crops improves the accuracy of the frequency domain image [6].

Sometimes the farmers face difficulties in order to recognize the diseases, which bring loss of the crop [8]. The primary factor restricting rice production globally is blast disease, which is brought on by *Magnaporthe grisea*. The rice blast fungus can infect plants at various phases of development [10]. The primary factor restricting rice production globally is blast disease,

which is brought on by *Magnaporthe grisea*. It first manifests as white/gray and brownish leaf lesions, then nodal rot and neck blast, which can cause necrosis and frequently shattering of the panicle (complex raceme or branched cluster of flowers). The fungus is currently widespread in over 85 nations around the globe, and it is the most significant rice disease in China, Japan, and the USA, where it can result in significant output losses. According to estimates, a mild infection in the field is sufficient to reduce yield by 50%. It is found that more than 10% of rice yield loss is due to various diseases [11, 12].

Blast disease management significantly investigated by many researchers in several nations. In the face of climate change, rice blast has never been fully eliminated from a region in which rice is grown, changing the practices by which resistant genes are outreach as a result of outbreak of disease even after many years of successful management [13]. That's why fungicide is the most effective for controlling disease rice blasts, but due to the residual effect on the environment and fungicide resistance within the pathogen population, this is a doubtful method. Among the methods to manage and control a disease which can reduce the significant losses is played by forecasting systems. Due to the lack of excellent agricultural software, a system was created utilizing a self-organizing model to detect blast and brown spot illnesses in rice leaves [14, 15]. When used in agriculture with the use of information technology, machine learning enables farmers to gather knowledge and data in order to make the best decisions for maximizing farm output [16]. Applications for machine learning algorithms in agriculture include crop recommendations based on plant pest detection, soil fertility, weed detection, crop yield cultivation, and plant disease detection based on disease identification at an early stage that will recover the plant and thereby increase the crop [17].

In fact, disease projections can help farmers and other end users plan ahead for fungicide applications, define fertilization methods by minimizing luxury consumption (which makes

plants more susceptible), and even estimate yields. [18]. The "disease triangle," which has the three sides "favourable conditions," "virulent pathogen," and "vulnerable host", is a popular metaphor used to describe how a pathogen interacts with its host and its environment in biological terms. The accessibility of the development of trustworthy early-warning systems would enable the prevention of the disease's explosive nature through the effective implementation of preventative measures [19, 20]. This would result in a decrease in crop losses and fungicide usage, lowering the environmental impact of rice farming. So that's why through machines we can identify the disease with more precision and accuracy so that we can manage accordingly.

Problem Statement

India is the second largest producer of staple crops and agriculture contributes the most to the Indian economy. When the disease spreads to epidemic levels in India, yield losses from blast could reach 50%. Disease incidence during natural outbreaks of blast disease during the wet season varied from 14 to 27% (above the economic threshold), causing a yield loss of approximately 27–35%. So close monitoring is required so that diseases do not cause fatal damage to production. To combat this modernized agriculture has been introduced. To detect the diseases many Machine learning algorithms and Convolutional Neural Networks have been developed.

About the Paper

From planting to harvesting, machine learning is used in agriculture employing a variety of methods from several fields of technology, including big data, artificial intelligence, drones, and data mining. In this paper, we have investigated that there are so many algorithms, but selecting the correct algorithm plays an important role. We have analyzed various algorithms like Random Forest Classifiers, SVM (Support Vector Machines) classifiers, K Means, Naive

Bayes, Recurrent Neural Networks, Decision Trees, Convolutional Neural Networks and Linear Regression.

Methodologies

Data collection and Pre-processing

The most crucial step in Machine learning is Data Collection. Researchers/ML practitioners usually collect data from farms. Some conditions have to be kept in mind while collecting the data. The requirements are temperature, weather, pressure, soil moisture content, humidity, pH, water content and other environmental factors. To measure these factors researchers use sensors and cameras for capturing images. Also, many devices are used to measure soil nutrients. For instance, an optical transducer was designed to measure soil nutrients such as Nitrogen, Phosphorus and Potassium. Some data is available on various websites such as Kaggle, GitHub etc. The government also publishes agricultural reports every year. After collecting, data pre-processing is done. Sometimes the researchers also conduct interviews with the personnel of the Department of Agriculture. The pre-processing works on raw data so as to enhance its quality. It involves cleaning data and data augmentation. Suppose some data is missing. For that, we should eliminate the null values. Sometimes mean values are also inserted in case of null values. We also use forward and backward values to fill up the null ones. For the image part, we flip the images, rotate them, and crop them to remove unnecessary pixels. For Deep learning, the images are also converted to tensors for the computation in neural networks.

Algorithms

After data pre-processing comes the choice of algorithms. There are many ML algorithms, and selecting the correct algorithm plays a crucial role. For example, if we take decision trees, it might give fruitful results but it has a tendency to over fit. So, while choosing one algorithm one must take note of the advantages and disadvantages of any algorithm. For

disease prediction researchers have mostly used Random Forest Classifiers, SVM classifiers, K Means, Naive Bayes, Recurrent Neural Networks, Decision Trees, Convolutional Neural Networks and Linear Regression. Let us discuss each of them in detail.

Random Forest Classifiers

Random Forest Classifiers are a branch of the Supervised Machine Learning Technique. It is used both for classification as well as regression. It is basically a combination of many classifiers and they are nothing but decision trees. Each decision tree provides a result and based on the majority of the votes we get the result. The individual predictors also depend upon the random variables. Random Forest is also known as Ensemble Classifier. It divides each value by using the best one among a subset of values which are randomly selected at that value. It is a multi-step procedure[21]:First, we find the Average of individual tree predictions in regression. Then we find the majority vote among individual prediction trees in classification. Then using random feature selection a tree is built. Research shows that Random forests perform better than Convolutional Neural Networks [22]. For the detection of rice plant diseases, the accuracy was 100% [22]

Support Vector Machines (SVM):

Support Vector Machines are used both for classification and Regression. It is a popular non-linear machine learning tool. It divides two classes by drawing a hyperplane. By introducing hyperplanes, the vectors can be isolated into different classes [23]. Many models have been developed using SVM. The models are easy to train but the accuracy is not up to the mark. But other SVM frameworks have also been developed in which some diseased images are used to train the SVM model. Later in the testing phase, the highlights of the image are separated using a procedure similar to the training phase. At that point, a component vector is constructed and forwarded to the classifier (for perceiving the paddy leaf ailments). The precision of the model was 92% [23]. Another Framework was developed using SVM. The

model is able to classify the shapes and text features of the diseases. The accuracy of the model was 97.2%[24]. A quadratic SVM classifier was developed. The classifier is able to classify the area, roundness, and lesion ratio. The results are good and the accuracy was estimated to be 81.6%[25].

K Means:

K means is a popular unsupervised Machine Learning Technique. It is an iterative technique in which the data points are grouped into clusters of different groups. In each of the clusters lies a centroid. The distance between each data point and the centroid of each cluster is calculated and the data point is assigned to that cluster that has a minimum distance. First, an image is fed to the K Means. Here the pixels are grouped into clusters and the classification of colours is also done. Finally, the cluster containing the diseased part is segregated and the results are shown.[26] The result of the K Means algorithm are plotted on histograms so as to make a comparative analysis between the faulty and the healthy ones[28].For 3 clusters the accuracy was 85% on training data[29].

Naive Bayes:

Naive Bayes is a popular supervised machine learning algorithm. It is a simple method based on probability. It produces the decisions based on the training data. The major advantage of this algorithm is that only independent variables are assumed. There are some steps in the algorithm which are as follows: We determine the prior value of each class. The prior value defines the probability of the occurrence of the disease in training data based on the symptoms. The likelihood for each class is calculated. Finally, the posterior values are calculated and the results of the existing ones and the new ones are compared. The highest posterior value is selected as the result of the classification.The accuracy of the model was calculated to be 73.91%. The NB modelling does a better classification process. Another NB

classifier was developed but the accuracy was 59% and therefore was not able to classify the three diseases of the rice accurately [29].

Recurrent Neural Networks:

RNNs are a type of neural network in which the output from the previous step is fed as input to the next step. The recurrent relations allow for the establishment of a link between the current state and the previous states. It can be compared to the Markov Chains.[30]The most commonly used RNNs are LSTMs. Long Short-Term Memory is a sequential network that predicts early blast disease. LSTMs are capable of avoiding vanishing gradient problems and is capable of remembering only the patterns that are required for long periods of time.[31]It is evaluated by varying the input variables like humidity, temperatures etc.[32] The accuracy of the LSTMs fluctuates from one region to another as the conditions vary. However, the accuracy was between 49%-65%.

Decision Trees

Decision Trees is a supervised machine learning technique. It can be compared to a flowchart and uses decisions at each step.[33] It can be used for both classification and regression techniques. Basically, it partitions the dataset so as to differentiate the features at each step. The algorithm breaks the datasets into unmixed partitions. The measure of heterogeneity is calculated using Entropy. Positive entropy indicates that the instances are heterogeneous. The accuracy of the model is estimated to be 97%.

Convolutional Neural Networks

Convolutional Neural Network is a Deep Learning technique which takes the image as input, processes the input and gives the output label. The pre-processing in this technique is not much required. The Convolutional Nets are able to capture the features of the images. It is also to be noted that with the proper tuning of the hyper parameters these models perform very well. A standard CNN model comprises eight layers: the input layer, Convolutional

layer, Pooling layer, Max Pooling, Fully Connected layer, Dropout Layer, Softmax Layer and Output Layer. A CNN model has been developed using 4199 diseased rice images. The training accuracy of the model was 99.78% and validation was 97.35% [34]. Another CNN model was developed which accepts 2-dimensional data. It comprises different weights and linear operations with layers. The model has the capacity to classify four categories of rice diseases. The estimated accuracy was 91%[35]. A CNN model was developed in which 500 images were fed. The accuracy of the model was 95.48%[36,37]. Many other famous models like VGG16, VGG19, Xception model, ResNet, and 5-layer convolution were used. Let us look at the accuracy of the model.

Table 1. Model accuracy

| Name of the model | Accuracy | Reference |
|---------------------|----------|-----------|
| VGG16 | 58.4% | [37] |
| VGG16 | 72.2% | [37] |
| VGG19 | 72.4% | [37] |
| Xception model | 72.2% | [37] |
| ResNet | 72.2% | [37] |
| 5-layer convolution | 78.2% | [37] |

It can be concluded VGG16 performs poorly so tuning of parameters plays an important role for a good accurate model.

Logistic Regression

Logistic Regression is used to find a relationship between a dependent and an independent variable. In this case, it is applicable if the target is categorical[38]. Since targets have different categories we use multiclass regression. For each class, the model has been trained and therefore the model makes predictions based on the maximum value of probability. Later the probabilities are used to make an S-shaped sigmoid curve. For the logistic regression algorithm, the training accuracy is 75.463% and 70.8333% on the test set. The model is capable of classifying 3 diseases [39].

Results and Analysis

Let us compare the results of each of the methods.

For the prediction of diseases, many methods have been used. Out of these Random Forest Classifiers, Support Vector Machines, Decision trees and Convolutional Neural Networks have accuracy greater than 90%. The Naive Bayes and LSTMs perform poorly. Naive Bayes is a simple method and LSTMs depend a lot on environmental factors. We must develop a model that can make accurate predictions. A slightly incorrect prediction can lead to substantial financial losses. While developing the model, there are some factors: hardware, time, number of images, and pre-processing. If we look carefully, pre-processing plays a major role in the cases of SVM, Decision Trees, and Random Forest Classifiers. But these models can be run on CPUs. Convolutional Neural Networks do not require much data pre-processing but need lots of time to train and GPUs are required for faster training.

Table 2. Comparative analysis among machine approach

| Methodologies | Accuracy | Reference |
|---------------------------|----------|-----------|
| Random Forest Classifiers | 100 | [22] |
| Support Vector Machines | 92 | [23] |

| | | |
|-------------------------|-------|---------|
| Support Vector Machines | 97.2 | [24] |
| Support Vector Machines | 81.6 | [25] |
| K Means | 85 | [29] |
| Naive Bayes | 73.91 | [28] |
| Naive Bayes | 59 | [29] |
| LSTM | 49-65 | [32] |
| Decision Trees | 97 | [29] |
| CNN | 97.35 | [34] |
| CNN | 91 | [35] |
| CNN | 95.48 | [36,37] |

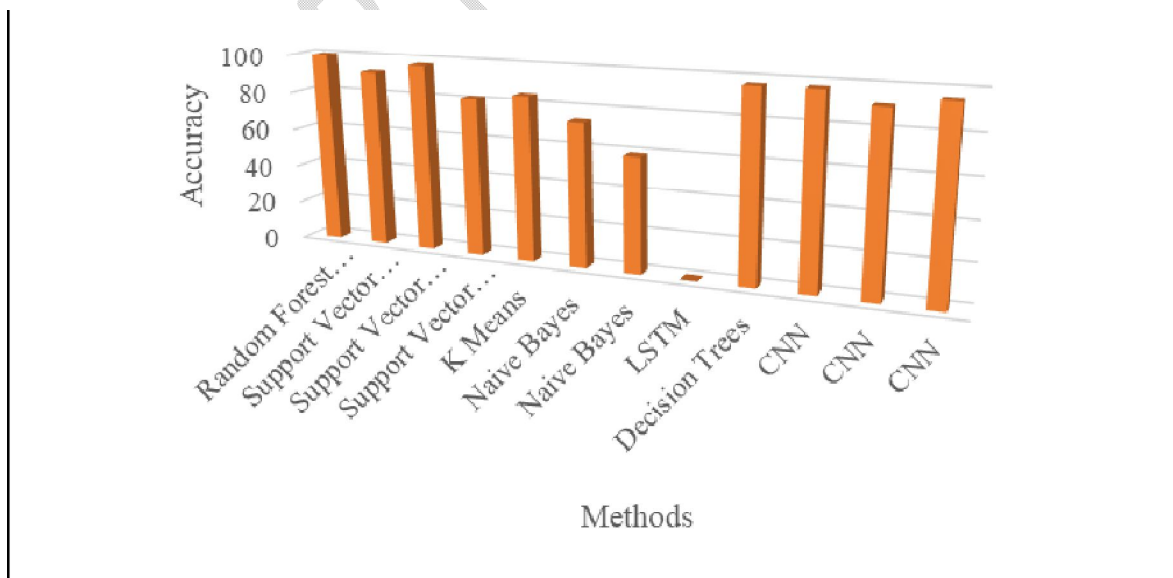


Fig 1. Accuracy of different methods of machine learning

Future Works

We have analysed different Machine Learning algorithms. Some algorithms have very good accuracy and some do not perform well. Convolutional Neural Networks and Random Forest Classifiers have excellent accuracies while LSTMs and Naive Bayes do not perform up to the mark. So, in the future, we would like to work on LSTMs and Naive Bayes because these two algorithms have very less accuracy and we would like to improve the accuracy by developing those using new parameters. We will also develop a CNN model that will take an input image and will predict it accordingly.

Conclusion

Different machine learning algorithms have been analyzed. Each of them has its own advantages and disadvantages. Out of different machine learning methods the Naive Bayes and LSTMs perform poorly. Naive Bayes is a simple method and LSTMs depend a lot on environmental factors. We must develop a model that can make accurate predictions. Out of these Random Forest Classifiers, Support Vector Machines, Decision trees and Convolutional Neural Networks have accuracy greater than 90%. We must develop a model that can make accurate predictions. A slightly incorrect prediction can lead to substantial financial losses. While developing the model, there are some factors: hardware, time, number of images, and pre-processing. If we look carefully, pre-processing plays a major role in the cases of SVM, decision Trees, and Random Forest Classifiers.

Ethical Approval:

Research involving Human Participants and/or Animals:

This article does not contain any studies with human participants or animals performed by any of the authors.

References:

1. Rice Knowledge Bank. www.knowledgebank.irri.org/ericeproduction/ Importance of Rice.htm.
2. Food and Agriculture Organization Of The United Nations Statistics. 2016. Available:<http://faostat.fao.org/>.
3. Wang JC, Correll JC, Jia Y. Characterization of rice blast resistance genes in rice germplasm with monogenic lines and pathogenicity assays. *Crop Prot.* 2015; 72:132–8.
4. Katsantonis D, Kadoglidou K, Dramalis C, Puigdollers P. Rice blast forecasting models and their practical value: a review. *PhytopatholMediterr.* 2017; 56(2):187–216.
5. Wang GL, Valent B. *Advances in genetics, genomics and control of rice blast disease.* Springer Science and Business Media, New York USA, 2009.
6. Groth DE. Azoxystrobin rate and timing effects on rice head blast incidence and rice grain and milling yields. *Plant Dis.* 2006; 90:1055–8.
7. Zeng, J., Feng, S., Cai, J., Wang, L., Lin, F., Pan, Q. Distribution of mating type and sexual status in Chinese rice blast populations. *Plant Dis.* 2009; 93:238–42.
8. Devi S, Sharma G. Blast disease of Rice caused by *Magnaporthe grisea*: a review. *Assam University Journal of Science & Technology: Biological and Environmental Sciences.* 2010;6:144–54.
9. Koizumi S, Kato H. Effect of mixed plantings of susceptible and resistant rice cultivars on leaf blast development. *Research Bulletin of the Aichi-ken.*

10. Meng X, Yu J, Yu M, Yin X, Liu Y. Dry flowable formulations of antagonistic *Bacillus subtilis* strain T429 by spray drying to control rice blast disease. *Biol Control*.2015;85:46–51.
11. TeBeest DO, Guerber C, Ditmore M. Rice blast. *The Plant Health Instructor*.2007. <http://www.apsnet.org/edcenter/intropp/lessons/fungi/ascomycetes/Pages/RiceBlast.aspx>.
12. Todorova S, Kozhuharova L. Characteristics and antimicrobial activity of *Bacillus subtilis* strains isolated from soil. *World J MicrobiolBiotechnol*. 2010; 26:1207–16.
13. Pagani V, Guarneri T, Busetto L, Ranghetti L, Boschetti M, Movedi E, Campos-Taberner M, Garcia-Haro FJ, Katsantonis D, Stavrakoudis D, Ricciardelli E, Romano F, Holecz F, Collivignarelli F, Granell C, Casteleyn S, Confalonieri R. A high-resolution, integrated system for rice yield forecasting at district level. *Agricultural Systems*. 2018. <https://doi.org/10.1016/j.agsy>.
14. Agrios GN. *Plant pathology*, 5th ed., Elsevier Academic Press, Inc; 2005.
15. Kaundal R, Kapoor AS, Raghava PS. Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC Bioinformatics*. 2006; 7:485.
16. Yoshino R. Ecological studies on the penetration rice blast fungus, *Pyriculariaoryzae* into leaf epidermal cells [in Japanese, English summary]. *Bulletin of the Hokuriku National Agricultural Experiment Station*. 1979; 22: 163 –221.
17. Confalonieri R, Rosenmund AS, Baruth B. An improved model to simulate rice yield. *Agron Sustain Dev*. 2009;29:463 –74.
18. Bregaglio S, Titone P, Cappellia G, Tamborini L, Mongiano G, Confalonieri R. Coupling a generic disease model to the warm rice simulator to assess leaf and panicle blast impacts in a temperate climate. *Eur J Agron*. 2016;76:107 –17.

19. Zhou L. N., Yu HY, Zhang L., Ren S., Sui YY., Yu LJ. Rice blast prediction model based on analysis of chlorophyll fluorescence spectrum. *Guang Pu Xue Yu Guang Pu Fen Xi*. 2014; 34(4):1003.
20. Kim, Y, Roh JH, Kim HY. Early forecasting of Rice blast disease using long short-term memory recurrent neural networks. *Sustainability*. 2018; 10:34
21. Patil, S., Ragite, H., Sangole, A. Rice Plant Disease Detection using Machine Learning. *IJARIIIE*. 2022; 8 (3): 1489-1493.
22. Singh, AK., Sreenivasu, SVN., Mahalaxmi, USBK., Sharma, H., Patil, DD., and Asenso, E. Hybrid Feature-Based Disease Detection in Plant Leaf Using Convolutional Neural Network, Bayesian Optimized SVM, and Random Forest Classifier. *Journal of Food Quality*. 2022.
23. Aggarwal, S., Suchithra, M., Chandramouli, N., Sarada, M., Verma, A., Vetrithangam, D. and Ambachew and Adugna, B. Rice Disease Detection Using Artificial Intelligence and Machine Learning Techniques to Improve Agro-Business. *Scientific Programming*. 2022.
24. Yao, Q., Guan, Z., Zhou, Y., Tang, J., Hu, Y., and Yang, B. Application of support vector machine for detecting rice diseases using shape and color texture features. In *international conference on engineering computation*. 2009; (pp. 79-83).
25. Bhartiya, VP., Janghel, RR. and Rathore, YK. Rice Leaf Disease Prediction Using Machine Learning. In *Second International Conference on Power, Control and Computing Technologies. (ICPC2T)*; 2022; (pp. 1-5). IEEE.

26. Sethy, P. K., Negi, B. and Bhoi, N. Detection of healthy and defected diseased leaf of rice crop using K-means clustering technique. *International Journal of Computer Applications*; 2017; 157(1), 24-27.
27. Sarkar, S., Biswas, S., Tapadar, A., and Saha, P. AI based fault detection on leaf and disease prediction using K-means clustering. *Int. Res. Eng. Technol.* 2018.
28. Maneesha, A., Suresh, C., and Kiranmayee, B. V. Prediction of rice plant diseases based on soil and weather conditions. In *Proceedings of International Conference on Advances in Computer Engineering and Communication Systems* (pp. 155-165). Springer, Singapore; (2021).
30. Lu, Y., Yi, S., Zeng, N., Liu, Y. and Zhang, Y. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing.* (2017); 267, 378-384.
31. <https://arxiv.org/ftp/arxiv/papers/2004/2004.01602.pdf>
32. Kim, Y., Roh, JH. and Kim, HY. Early forecasting of rice blast disease using long short-term memory recurrent neural networks. *Sustainability.* (2017); 10(1), 34.
33. Song, YY., and Ying, LU. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry.* (2015); 27(2), 130.

34. Hossain, SM., Tanjil, M., Morhsed, M., Ali, MAB., Islam, MZ., Islam, M., and Islam, S. M. Rice leaf diseases recognition using convolutional neural networks. International Conference on Advanced Data Mining and Applications Springer, Cham. (2020, November); pp. 299-314.
35. Purbasari, IY., Rahmat, B., and PN, C. P. Detection of Rice Plant Diseases using Convolutional Neural Network. In IOP Conference Series: Materials Science and Engineering. IOP Publishing. (2021, May); (Vol. 1125, No. 1, p. 012021).
36. Lu, Y., Yi, S., Zeng, N., Liu, Y., and Zhang, Y. Identification of rice diseases using deep convolutional neural networks. Neurocomputing. (2017); 267, 378-384.
37. <https://iopscience.iop.org/article/10.1088/1755-1315/1032/1/012017/pdf>
38. Peng, CYJ., Lee, KL., and Ingersoll, G. M. An introduction to logistic regression analysis and reporting. The journal of educational research. (2002); 96(1), 3-14.
39. Aravind k. Survey and assessment of rice blast disease incidence from major rice growing locations of Telangana state, India. Scientist, USA. 2022 - Vol 1 - Issue 3.