International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Identification of Diseases in Paddy Crops Using CNN

[1]Gayatri Parasa, [2]M. Arulselvi, [3]Dr. Shaik Razia

Submitted: 10/02/2023 **Revised:** 08/04/2023 **Accepted:** 11/05/2023

Abstract: In ancient times, agriculture is one of the most predominant occupations of Indian civilizations and it has a great impact in contributing to our country's economy. Unfortunately, due to several reasons like pests and unpredictable climatic conditions, there has been poor productivity in certain crops, especially paddy. This has been drawn attention towards enhancing the productivity of the paddy crops. Through lots of research, it has been identified that paddy crops are infected by various diseases, and this is one of the reasons that directly affects the overall productivity of the crop. Hence, there emerged an immediate need to take preventive measures and improve the overall productivity rate of paddy crop. In this regard, an Intelligent deep learning algorithm called Convolution Neural Network (CNN) is proposed with an increased structure of 15 layers which predict various diseases that may affect the rice leaves. The developed model efficiency was evaluated in terms of Accuracy, Precision, F-measure, and Recall.

Keywords: *Paddy Crop, Leaf Smut, Bacterial Leaf Blight, Brown Spot, Convolution Neural Network, Deep Learning, Artificial Intelligence.*

1. Introduction

The world's rising population demands for food grain cultivation, and this causes a major challenge for the agriculture industry [1]. Unfortunately, due to several reasons the overall productivity is affected. Rice is the major crop in Southern parts of India as it is highly consumed in southern regions of India [3]. Due to several reasons, the overall productivity of paddy crop is getting affected and one of the main reasons for getting infectious diseases is by pathogens, bacteria, fungi, and viruses [13]. Therefore, a lot of techniques were deployed to get the plant infection prediction. Consequently, evaluating crop efficiency is a reasonable strategy of anticipating crop productivity so study on crop yield using different machine learning and deep learning techniques were improved [4]. Study on crop yield is fundamental to guarantee sufficient accessibility of food all through the decades regardless of environment and market elements [5]. Past few decades, the IT sector is playing a major role and rendering its services to meet the demand of the agriculture industry. This prevention saves the crop as further spreading of the disease is controlled by taking appropriate measures. [14]. Image processing techniques are incorporated in tracking out the data about the various diseases that affects the plants [6]. One of the key procedures in Deep Learning is to avoid potential risk and it is one of the methods for successful prevention of the disease [2]. A few samples of rice leaf images along with corresponding diseases are depicted in Fig.1.

1Research Scholar,
Department of Computer Science & Engineering
Annamalai University.
gayathriparasa20@gmail.com
2Associate Professor,
Department of Computer Science & Engineering
Annamalai University.
marulcse.au@gmail.com
3Associate Professor
Department of Computer Science & Engineering Koneru
Lakshmaiah Education Foundation.
Vaddeswaram Andhra Pradesh 522302, India.
razia28sk@gmail.com

Fig.1 sample images of rice leaf disease

There was plenty of research done in the past to predict rice leaf disease that could increase paddy production, such as Random Forest (RF) classifier [17], Support Vector Machine (SVM) [18], K-Nearest Neighbor (KNN) [19], etc. But nevertheless, there is a need of more appropriate and accurate solutions. Furthermore, the traditional strategies lacked classification accuracy and were prone to many errors [20]. Therefore, the rate at which the disease is predicted is of very low accuracy. The aim of the neural networks is not only to maximize the prediction rate but also to identify the disease more effectively [7]. The modified CNN structure in the proposed framework helps in predicting the leaves with diseases by identifying the affected location. Moreover, the affected location is shaped out and it also predicts the name of the disease efficiently.

In this work, the following paddy crop diseases are identified:

1.1 Brown Spot:

Brown Spot is one of the most common diseases that affect rice plants. These are round earthy brown colored spots (Fig.2) which can be seen up on the paddy leaf. The crop is mostly targeted by this disease between the seedling and milky stages [9]. Helminthosporiose or sesame leaf spot are other names for brown spot. Loses due to brown spot affect both quantity and quality.

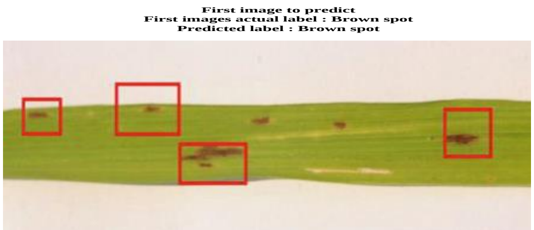

Fig.2 Sample leaf for Brown spot disease

1.2 Bacterial Leaf Blight:

In irrigated rice, Bacterial Leaf Blight (BLB) is a significant vascular disease that can reduce crop yield up to 50% and is the most challenging task to control [15]. This infection normally happens in the leaf of the rice plant and without doing much investigation it can be found out by checking out at the yellow and white strips on the leaves (Fig.3). This infection can be perceived by taking the youngest leaf which will be light yellow in shading assuming the plant is experiencing bacterial leaf curse [10]. The preferred scientific name of the bacterial leaf blight is Xanthomonasoryzae.

Fig.3 Sample image for Bacterial Leaf Blight disease

1.3 Leaf Smut:

Leaf Smut (Fig.4) is a small infection that occurs in paddy fields but is widely distributed. By looking at the leaf, which will be completely encircled by growth, it may very well be distinguished [8]. Also, this disease is infected by Entyloma Oryza. This infection can be constrained by performing the clean development and rising safe assortments. The preferred scientific name of the bacterial leaf smut is Ustilaginoidea virus.

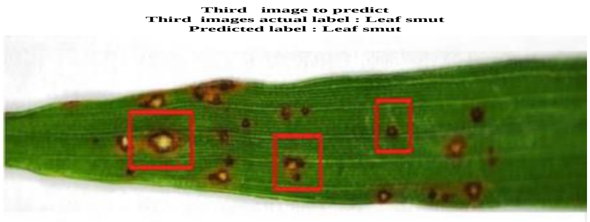

Fig.4 Sample image of Leaf smut disease

2. Related works

Agriculture plays a vital role and contributes a major part to the Indian economy. Nevertheless, plant disease is the most significant challenge which needs immediate attention to meet the demands of the rising population [11]. It is observed that diseases in plants lead to affect the agriculture yield. Therefore, prevention and detection are the key parameters to avoid such diseases. Moreover, the IT industry extends its service with its powerful strategies such as Deep learning and Image Processing to detect plant disease in an efficient way. But Paddy crop disease is the most demanding challenge, for the South Asian region. Therefore, Sharma *et al.* [21] have developed the CNN module with a decision-making process to predict the various stages of Paddy crop disease in an initial stage. Here, the proposed module can efficiently enhance classification accuracy.

Picon*et al.* [22] has introduced the Deep Residual Neural Network (DRNN) to perform the detection and prediction process for multiple disease classifications with real-time conditions. Moreover, it is an intelligent image processing replica incorporated with ML strategies. Various plant diseases effectively decrease the productivity of rice. If accurate detection is not performed the disease can spread and affect the overall agriculture productivity. Therefore, Chen et al. [23] have proposed the Deep Learning-based attention mechanism that is utilized to improve the learning ability of the entire system performance. Moreover, transfer learning is used twice for the training model. Here, loss functions are reduced using the multi-classification module and enhance the accuracy. However, the process is time consuming. Jundeet al. [24] has proposed the Deep Convolutional Neural Network (DCNN) based transfer learning module that helps in recognition of plant disease. Here, in the classification and feature extraction process the layers are extended towards the high dimensional structures. Moreover, typical massive datasets are used to keep the pre-trained module. Also, the performances are compared with state-of-the-art techniques.

3. Proposed Methodology

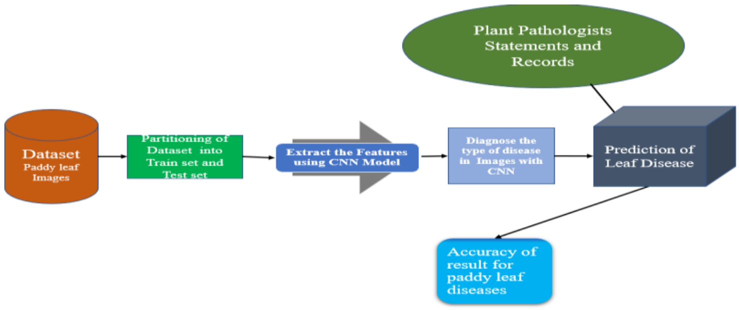

Fig 5: Proposed Framework.

The proposed framework of the entire process has been depicted in Fig. 5 and the steps are defined as follows: Initially data is collected from Kaggle, and few images were referenced from Google. Then the data will be divided into two partitions that is training and testing. After features are extracted using proposed CNN model. Finally, the CNN model predicts the type of disease that the leaf was infected and performance metrics are calculated.

3.1 Dataset:

Various images of rice leaves were collected from Kaggle, and a few images were referenced from Google. The dataset contains the different categories of healthy rice leaves and leaves with different diseases like Leaf Smut, Bacterial Leaf Blight and Brown Spot. Altogether 5000 images were considered for this work. The collected data goes through various data augmentation techniques that in turn make it suitable to train deep learning models.

Image resizing and rescaling has been performed because every image in the collected dataset has RGB coefficients in the range 0-255 and varying (height x width). The original dimension of every image in the collected dataset is (3081 x 897) and all the images were resized to (256 x 256). Different data augmentation techniques were used to improve the original dataset. Image random rotation and flipping has been applied for real-time data and this operation was employed by ImageDataGenerator class provided by the Keras deep learning library. Dataset with 5000 images has been expanded to 8000 images after applying different data augmentation techniques. In these 8000 images 2250 were found to be healthy, 2100 images of leaves were affected by Bacterial Leaf Blight, 1385 images of leaves were affected by Leaf Smut and 2265 images of leaves were affected by Brown Spot.

3.2 Proposed CNN:

CNN frameworks give better performance for feature detection in images processing and this CNN

The CNN structure includes different layers like input layer, convolution layer, pooling layer, dense layer and finally output layer. But to detect the leaf disease efficiently the proposed CNN framework will include an increased number of layers. Therefore, the proposed CNN model is developed with a depth of 15 layers, and they are input layer, five convolution layers (Conv1, Conv2, Conv3, Conv4 and Conv5), five max pooling layers (Pooling1, Pooling2, Pooling3, Pooling4 and Pooling5), four dense layers (Dense1, Dense2, Dense3, and Dense4), 1 flatten layer and output layer. Consequently, the SoftMax activation function is utilized in the output layer.

Initially, an image is passed through the convolution layer 1 of size 256x256 consisting of 32 filters as each filter identifies certain features of the input image. The output of the above layer acts as input to the second convolution layer consisting of 64 filters with an image size of 128x128 by identifying different features for prediction of diseases in the leaves. In the convolution layer 3, comprising of 128 filters is subject to further reduction of image size with a dimension of 128x128. In convolution layer 4 there are 256 filters, and the image is reduced to dimension of 64x64. Finally, in the fifth convolution layer the image is further reduced with a dimension of 32x32, consisting of 256 filters. Layer1 is comprised of 891 parameters similarly layers 2,3,4 and 5 are comprised of 8256, 18496, 36928, 84571 parameters respectively. The total trainable parameters are 3346115.

replica takes the collected images in the form of matrices and entire functions are processed. After the collection of entire data, preprocessing and augmentation techniques are effectively performed. At first input layer of the developed replica is fed by an RGB image with size is represented as $w' \times h'$ where, w' is denoted as width and h' is height of the collected images respectively. Then convolution layer which is the significant layer of CNN module extracts specific features from the collected input images with the help of kernel function. Here, multiple features of the input image are extracted after coordinating the convolution kernel function more times. In a kernel function, forward and backward propagation is expressed in following eqn. (1) and (2),

$$o_{i'j}(k) = \sum_{i'=0}^{p-1} \sum_{j'=1}^{q-1} w_{mn}(k) y_{(i'+m)(j'+n)} + a(k) \quad \text{----} \quad (1)$$

$$\frac{\partial C}{\partial w_{mn}(k)} = \sum_{i'=0}^{P-p} \sum_{j'=0}^{Q-q} \frac{\partial C}{\partial a_{i'j'}(k)} y_{(i'+m)(j'+n)} \quad \text{----} \quad (2)$$

Where, $'y'$ is the collected input image, $o(k)$ is output image which means convoluted image, $'k'$ is represented as kernel coefficient, $'w'$ is the weighting function, $'a'$ is denoted as bias function and finally $'c'$ is the cost function. Here, five convolution layers are utilized to attain the feature maps from the collected rice leaf images. Once the image features in the convolution layer are filtered then the image has passed through the pooling layer for dimensionality reduction. Moreover, the main of this pooling function is to decrease the spatial size of the rice leaf images. The process of pooling layer is depth dimension of each image and remains are intact. In our work the model has been implemented in max pooling operation.

Here, the invariant features from the convolution layer are extracted with the help of max-pooling function by extracting the special features that are formed in different partitions. Thus, remaining functions are processed, and summarized features make the system robust for finding the disease location from the rice leaf images.

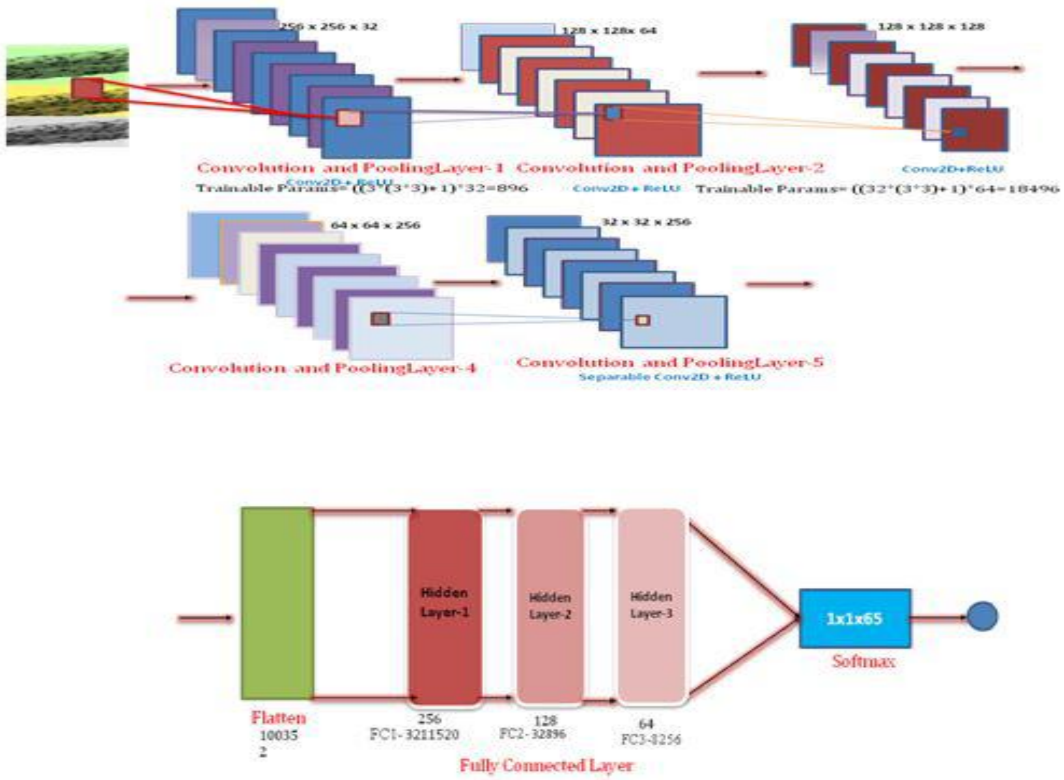

Fig 6: Proposed CNN Architecture

Therefore, the max-pooling stage is implemented in following eqn. (3),

$$o_{i'j}(k) = \max\_pool(0, y_{(i'+m)(j'+n)}) \quad \text{----} \quad (3)$$

After the pooling operation, nonlinear coordination of the pooling output is flattened and fed into the fully connected dense layer. This layer can provide the one-dimensional vector and fully connected layer gets activated. Here, the rice leaf images are renovated into column vector utilizing the flatten operation. After that, linear operations are functioned in the hidden layer of the sense function. ReLU and pooling layers are used in feature extraction and SoftMax is used in the last layer to predict which class that the leaf image belongs to. Primarily, in neural networks, the main responsibility of activation function is to transform the summed weighted input to output. The input to any activation function is the sum of weight multiplied by the input and the bias will be summed up to this input (y) in each layer and to every neuron. The Rectified linear activation function or simply ReLU activation function always gives the maximum value in 0 and the value of y as the output for the given input.

ReLU activation function is expressed in following eqn. (4),

$$\mathrm{Re}LU\,(y) = \max\_\ pool(0, y) \text{ ------ (4)}$$

Then SoftMax function is represented for the classification performance, that is expressed in following eqn. (5),

$$\alpha(x)_{i'} = \frac{e^{x_{i'}}}{\sum_{j'=1}^{k} e^{x_{i'}}} \text{ ------- (5)}$$

The proposed model is trained on 80 percentage of data and the features are extracted on series of convolution and pooling layers and then rice leaf disease is classified and predicts the specific disease that the paddy crop is suffering with. Furthermore, the developed workflow is demonstrated in the proposed methodology.

3.3 Performance Metrics

The proposed CNN model is trained and tested on the collected dataset. The efficiency of the model has been evaluated using the following performance metrics like Accuracy, Precision, Recall and F-score.

3.3.1 Confusion Matrix

The expected outcomes for a classification task are summarized in a confusion matrix and shown in Table 1. With count values, the number of accurate and inaccurate predictions is totaled up and separated by each class. This is the key contribution of confusion matrix. The confusion matrix demonstrates the various ways in which the classification model generates incorrect predictions. In the below table, H represents Healthy, BLB denotes Bacterial Leaf Blight, BS stands for Brown Spot and LS terms Leaf Smut.

	H	BLB	BS	LS
H	TH	FBLB	FBS	FLS
BLB	FH	TBLB	FBS	FLS
BS	FH	FBLB	TBS	FLS
LS	FH	FBLB	FBS	TLS

Table 1: Confusion matrix

3.3.2 Accuracy

The accurate score of rice leaf disease prediction is validated to compute the accuracy; the proposed strategy has achieved finest value of accuracy score and is found too more efficient. In other words, accuracy is the ratio of the number of proper predictions to the total number of input samples. The accuracy score of the developed replica was evaluated using the following equation.

$$Accuracy = \frac{TH + TBLB + TBS + TLS}{TH + TBLB + TBS + TLS + FH + FBLB + FBS + FLS}$$

Where, TH stands for True Healthy, TBLB is True Bacterial Leaf Blight, TBS is True Brown Spot, TLS is True Leaf Smut, FH is False Healthy, FBLB is False Bacterial Leaf Blight, FBS is False Brown Spot, and FLS is False Leaf Smut.

3.3.3. Precision

The parameter precision is defined to enumerate the detection function on the fundamentals of true positive and predicted true positive scores to estimate this prediction outcome. Additionally,

$$Recall = \frac{TH}{TH + FH}$$

3.3.5 F-measure

In a prediction system, f-measure is the significant parameter to prove the efficiency of the developed system. Here, F-measure is validating mainly based on two parameters such as precision (P) and recall

precision is evaluated based on the accurate positive scores to that of total number of sample dataset and is given in equation.

$$Precision = \frac{TH}{TH + FBLB + FBS + FLB}$$

3.3.4 Recall

Recall is the important key parameter in DL and ML paradigms, which is defined as correct predictions of healthy divided by the total number of true healthy and false healthy scores. It helps in effectively classifying the diseases in rice leaves from the collection of sample images. Consequently, recall measure is computed by following equation.

(R). Therefore, the formula for calculating these parameters is demonstrated as the following equation,

$$F\ measure\ = \frac{2 * P * R}{P + R}$$

4. Results and Discussion

The data augmentation has been applied to the 5000-image dataset and it was expanded to 8000 images. Moreover, 80% of the data (6400 images) was used for training and the remaining 20% of data (1600 images) has been used for testing. The proposed model efficiency was evaluated by comparing the key parameters in terms of Accuracy, Precision, F-measure, and Recall.

The proposed CNN with 15-layer architecture has attained finest outcomes to classify and detect various diseases in the rice plant. Here, three types of rice leaf diseases are detected, and the collected datasets are tunneled through several layers for resizing of image, training, testing, and classification using CNN. Consequently, the implementation results are depicted in fig.7. As a part of the results, all the images in the testing dataset were examined, and in the below figures, the results are shown with the actual label and the predicted label. The first figure shows an image of leaf smut, and the designed CNN successfully predicted the correct title, and the same is shown for the remaining diseases too.

Fig.7 Predicted results

Furthermore, the confusion matrix is utilized to validate the proposed network outcome (predicted) and actual output. The confusion matrix is shown in Table 2. In the below table2, H represents healthy, BLB denotes Bacterial Leaf Blight, BS stands for Brown Spot and LS terms Leaf Smut.

	H	BLB	BS	LB
H	1144	0	11	0
BLB	0	189	0	2
BS	5	0	180	0
LS	0	4	0	65

Table 2: Confusion matrix

The training and testing accuracy and loss of the developed CNN with 15 layered frameworks is illustrated in fig.8.

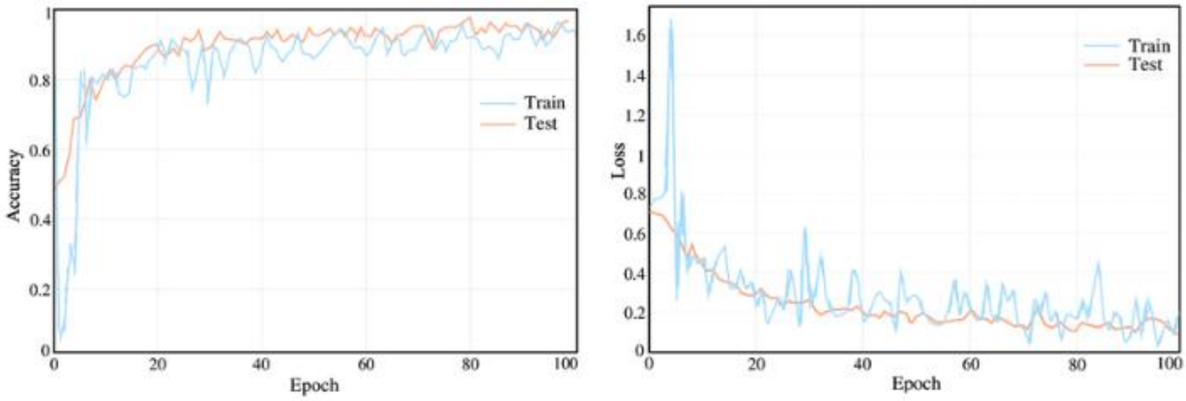

Fig 8. Training and Testing loss and accuracy

The performance of the proposed model is measured in terms of accuracy, f-measure, recall based on disease prediction is shown in Table 3. Moreover, 80% of data has been considered for the training process and the remaining 20% of data has been considered as testing. The overall accuracy of the proposed model is 95%.

Disease name	Precision	Recall	f-measure	Accuracy
Bacterial Leaf Blight	0.89	1.00	0.94	0.93
Brown Spot	1.00	0.83	0.91	0.92
Leaf Smut	1.00	1.00	1.00	0.99
Healthy	0.96	1.00	1.00	1.00
Average	0.96	0.95	0.99	0.96

Table 3 Predicted results

5. Conclusion

The classification and prediction of rice leaf disease through the computer aided strategy has been very helpful in analyzing the diseases effectively. In the past several techniques were incorporated but unfortunately, they could not attain the best performance and there was an emerging need in effective and in-time disease identification. Hence, this work introduces a modified CNN for predicting the diseases in paddy crop. Moreover, the projected replica has achieved finest outcomes and best prediction rate. The achieved accuracy of this proposed model is 95%.

References:

[1] Das, Sujay, et al. "A model for probabilistic prediction of paddy crop disease using convolutional neural network." Intelligent and Cloud Computing. Springer, Singapore, 2021. 125-134.

[2] Hossain, Khandkar Asif, et al. "Paddy Disease Prediction Using Convolutional Neural Network." International Conference on Intelligent Computing & Optimization. Springer, Cham, 2021.

[3] Kodaty, SreeCharitha, and BalajiHalavath. "A New Approach for Paddy Leaf Blast Disease Prediction Using Logistic Regression." Advances in Information Communication Technology and Computing. Springer, Singapore, 2021. 533-542.

[4] Anandhan, K., and Ajay Shanker Singh. "Detection of paddy crops diseases and early diagnosis using faster regional convolutional neural networks." 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2021.

[5] Leelavathy, B., and Ram Mohan Rao Kovvur. "Prediction of biotic stress in paddy crop using deep convolutional neural networks." Proceedings of International Conference on Computational Intelligence and Data Engineering. Springer, Singapore, 2021.

[6] Anitha, A. "Disease prediction and knowledge extraction in banana crop cultivation using decision tree classifiers." International Journal of Business Intelligence and Data Mining 20.1 (2022): 107-120.

[7] Julie, J., J. JoshanAthanesious, and S. Adharsh. "Novel Disease detection for paddy crop using CNN with Transfer Learning." 2021 4th International Conference on Computing and Communications Technologies (ICCCT). IEEE, 2021.

[8] Rajmohan, R., et al. "Integrated Mathematical Modelling and Analysis of Paddy Crop Pest Detection Framework Using Convolutional Classifiers." Simulation and Analysis of Mathematical Methods in Real-Time Engineering Applications (2021): 293-315.

[9] Moparthi, Nageswara Rao, et al. "Paddy Leaf Disease Detection using CNN." NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO (2021): 800-806.

[10] Anandhan, K., and Ajay Shanker Singh. "Detection of paddy crops diseases and early diagnosis using faster regional convolutional neural networks." 2021 International

Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2021.

[11] Lijo, Jithy. "Analysis of Effectiveness of Augmentation in Plant Disease Prediction using Deep Learning." 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021.

[12] Shafi, Uferah, et al. "Wheat rust disease detection techniques: a technical perspective." Journal of Plant Diseases and Protection (2022): 1-16.

[13] Verma, Amit. "SVM, CNN and VGG16 Classifiers of Artificial Intelligence Used for the Detection of Diseases of Rice Crop: A Review." Sentimental Analysis and Deep Learning (2022): 917-931.

[14] Temniranrat, Pitchayagan, et al. "A system for automatic rice disease detection from rice paddy images serviced via a Chatbot." Computers and Electronics in Agriculture 185 (2021): 106156.

[15] Chen, Junde, et al. "Identifying crop diseases using attention embedded MobileNet-V2 model." Applied Soft Computing 113 (2021): 107901.

[16] Mekha, Panuwat, and NutnichaTeeyasuksaet. "Image Classification of Rice Leaf Diseases Using Random Forest Algorithm." 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering. IEEE, 2021.

[17] Ruan, Chao, et al. "Prediction of Wheat Stripe Rust Occurrence with Time Series Sentinel-2 Images." Agriculture 11.11 (2021): 1079.

[18] Jena, Kalyan Kumar, et al. "Rice Disease Classification Using Supervised Machine Learning Approach." 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2021.

[19] Picon, Artzai, et al. "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild." Computers and Electronics in Agriculture 161 (2019): 280-290.

[20] Chen, J., Chen, J., Zhang, D., Sun, Y., &Nanehkaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture, 173, 105393.

[21] Das, Sunanda, Debanik Roy, and Payel Das. "Disease feature extraction and disease detection from paddy crops using image processing and deep learning technique." Computational Intelligence in Pattern Recognition. Springer, Singapore, 2020. 443-449.

[22] Rallapalli, SethuMadhavi, and SaleemDurai. "A contemporary approach for disease identification in rice leaf." International Journal of System Assurance Engineering and Management (2021): 1-11.

[23] Maneesha, Akula, Chalumuru Suresh, and B. V. Kiranmayee. "Prediction of rice plant diseases based on soil and weather conditions." Proceedings of International Conference on Advances in Computer Engineering and Communication Systems. Springer, Singapore, 2021.

[24] Tian, Long, et al. "Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection." Remote Sensing ofEnvironment 257 (2021): 112350.