Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Deep Learning Techniques For Paddy Plant Disease Detection In Agricultural Research

Jeya Bharathi R. and Dr. Arpana Bharani

Department of Computer Science, Dr. A. P. J. Abdul Kalam University.

Corresponding Author Email: arpanabharani@gmail.com

Abstract— Disease identification in plants is important to avoid the losses in the measure and production of agricultural products. The problems in the agricultural sector are lessoned by engaging various deep learning and image processing techniques. This paper mainly focus on paddy plant disease detection centred on image inputs of infected rice plant by using disparate deep learning and image processing techniques. Also, the notable deep learning and image processing concepts in detecting and classifying the paddy diseases are discussed. Support Vector Machine (SVM), k-Nearest Neighbor Classifier (KNN),Multiple Linear Regression, Randomly used Classifiers, Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) are the various classification techniques used in various applications in the agricultural research. Different input data yields varied quality of an outcome and so selecting a classification method is a critical task. Biological research, agriculture, etc. are the disparate fields where the paddy disease classifications are applied. The detailed study on the diseases of paddy, size of image dataset, pre-processing, segmentation techniques, classifiers are presented in this paper.

Keywords— Paddy, Convolutional Neural Network, Deep Learning, Segmentation.

I. INTRODUCTION

Agriculture is the prime income source in various countries in the world. Grounded on the significance of agriculture, farmers select their crops, paddies, and the related pesticide to ameliorate the growth of the plant in the limited time [1]. Paddy stands as the major food crop in several countries [2]. Nowadays the paddy plants are facing grave problems in the agricultural sector owing to diseases which affects the quality and quantity of the crops. Lack of enough expert availability at the farming field, lack of knowledge in fertilizer management, lack of awareness about diseases and pests are the disparate reasons for the lower production rate [3]. Diseases in plants contribute indirectly and directly to certain conservational damages. As these diseases spread worldwide, they cause damage to the general operational of the plant and also damage the financial complaint by significantly diminishing the number of crops grown [4]. Plants are affected by various bacterial and fungal diseases [5]. Sheath blight, NBSD Leaf blast and Brown spot are the disparate diseases that transpire on paddy plants [6]. The damages rooted from the Blast shows the degree of severity of the disease. Another notable paddy disease termed brown spot is transpired by the Bypolaris

Oryza (fungus) which is visible all through the growing season. Brown spots normally become severe when rice is grown on silicon-deficient soils [7]. In Asia 10 to 15% of production is destroyed because of paddy diseases [8]. Recently, an expert in the undeveloped field manually supports the analysis and monitoring of diseases which requires more effort and long processing time [9]. Sometimes the ranchers face difficulties in order to distinguish the diseases, which brings loss of the crop. One of the effective solution for farmers is processing the captured images of “seem to appear” infected leaves by an automatedsystem. The leaf that discloses the disease symptoms is used for detecting the diseases. The automation system, helps the farmers to be aware of diseases instantly. Certain plants failed to produce owing to the delay in disease detection. Therefore, early disease detection is supreme important.

Several image processing and Deep Learning algorithms are developed to diagnose the diseases of the paddy plant. In plant disease identification utilizing DL algorithms, the accuracy is contingent on three processes namely feature segmentation, feature extraction and classification algorithm. Deep learning methodologies have shown a propitious result in image classification. Recently, researchers have been utilized to examine the diseases of mango, apple, tomato, rice, and also wheat. The blast disease may affect very small portion of the leaf during the nursery stage and even spread to entire cultivation resulting in huge loss to the farmer. Agriculture forms the foundation of Indian economy.The images are captured as of the paddy fields. Those images are then pre-processed and the infected parts of the leaves in the images are segmented. Then, from the segmented images, the features are extorted and finally the ML techniques are used to perform classification of the diseases. The success of such a system relies on how accurately the system performs the image processing and ML operations.

The present paper is organized as follows sections 2 discuss about various techniques of rice yield prediction process, section 3 deal with proposed method and section 4 discuss about experimental results.

II. RELATED WORK

In paddy crops, the problems can occur by pests, diseases and conservational conditions. The pests can be controlled by spread over the proper chemical Pesticides in the field. The environmental parameters like water irrigation, soil moisture and animal invasion in the field are monitored and controlled by automated technology. Several diseases can be controlled by monitoring and implementing proper farming methods. Blast disease is the common factor affected by the rice crop wherever it is cultivated [10]. This disease is caused by bacterial or fungal and it could affect the crop either in nursery stage or main field. Inspite of applying pesticides, it is difficult to control the rapid spreading of this infectious disease. In our research work has been done to find out the efficient methods of an agriculture to predict paddy yield prediction of paddy production with accuracy rate.

A. Support Vector Machine (SVM)

SVM is considered as the technique of machine learning that is mainly utilized in image processing for the purpose of classification. In the initial years, the ability of SVM is very low which can separate the data in a linear fashion with only two classes. However, nowadays, it has the ability to classify the data into more than two classes. In order to enhance the accuracy in image processing, SVM is preferred. An imageprocessing based approach with the help of SVM [11] to classify and analyze the three various kinds of paddy crop diseases. In another research work, disease prediction using training was conducted. The SVM parameters like nu and gamma were optimized to enhance the overall efficiency. The offered approach successfully classifies and determines the type of the disease. Based on the results, the cure activities were performed on the affected rice crop. By using both recall and precision method, it was claimed that an accuracy of an image processing system can be enhanced upto 94.16% [12]Another review article uses, SVM to differentiate both normal and disease-affected plants in various environment, in which, 200 samples per image (100 for each class) considering the 70% of the Tomato, Maize and Carrot datasets. SVM algorithm had a high potential to work under the various working conditions and also it can differentiate the crop species[13].

To compare the performance of machine learning techniques such as support vector machine, decision tree and naive bayes based on image processing techniques,[14] had performed an experiment, in which, 800 images of the turmeric leaf were taken to perform the feature extraction by both segmentation and pre-processing. The extraction results of three various machine learning techniques exhibited that the accuracy of SVM was the highest as 93.75%. K-Means and Fuzzy Logic techniques were combined by [15] to perform the disease severity recognition in rice crop leaves. The critical condition of the diseases was found by the assistance of Support Vector Machine. Fuzzy Logic was used to separate the leaves based on the grades of the leaf disease. Both machine vision tool and fuzzy logic are involved in this work, and hence, it provides nearly 86.35 % accuracy.

B. K-Nearest Neighbors

KNN method is one among the training Machine algorithm(ML) which is taken into account as an easy method to be applied within the analysis of knowledge with many dimensions variable. This method are often used when it doesn't meet the classical assumptions. The performance of KNN and ensemble KNN. Although this method is straightforward but this method has advantages over other method. For instance, it can generalize from a comparatively small training set. In This method is extremely important to settle on the amount of k-nearest neighbors. Ensemble technique may be a method that has accuracy of prediction and efficiently utilized in the KNN method, so it's not necessary to look the optimal number k. The result shows that MAPE, MAE, and RMSE of prediction are going to be small if the amount of k-nearest neighbors large. Overall, KNN ensemble method features a better performance than KNN method. The K Nearest Neighbors (KNN) is applied for simulating daily precipitations and other weather variables [16], and different possible changes of the weather scenarios are analyzed using KNN. In KNN method is very important to choose the value of k –nearest neighbors number, because this can affect the predicted results. Small values of k-nearest neighbors number can produce a great variety on the prediction results, whereas a large value of k-nearest neighbor number can lead to a large biased models. KNN uses weighted average concept to parameters estimator of some single KNN models. This method is predicted to yield better prediction than single KNN method.

C. Randomly Used Classifiers

An experiment was performed [17] to monitor the disease in Soya bean. The pooled data was

analysed by various techniques such as linear correlation analysis, logistic discriminant analysis(LgDA) and step wise linear discriminant analysis (LDA) to identify the Wave bands with the most disease discrimination capability. A classification function was developed with the help of identified band combinations to determine the conditions of the plant disease. Both LgDA and LDA identified several bands in the blue, red, NIR and green regions as critical for disease discrimination. Under the cross-validation, the discriminant models deliver nearly 80% accuracy to detect the healthy plants among the overall images. But, in disease affected plant detection, it exhibited very low accuracy. A discriminant model with two-class, provides 58% accuracy in the detection of diseased plants and 97% accuracy in the detection of healthy plants.
Random forest is considered as an algorithmic method of bagging and the in stability is reduced in decision trees when performing the averaging. Begue et al. had used both machine learning and computer vision techniques to develop a fully automated method to differentiate the plants pecies. A smart phone was employed to photograph and collect the leaves from 24 various medicinal plants pecies. The features such as area, perimeter, length, number of vertices, width and colour of hull are extracted from the images of medicinal plant species. A 10-fold cross validation technique was utilized by random forest classifier to obtain the finest outcomes. The random forest classifier delivered 90.1% accuracy which is higher than other machine learning approaches like neural networks, naïve Bayes, SVMs and k-nearest neighbour. Computer vision techniques were employed in this work to extract several shape-based features present in the leaves of medical plants. The finest performance of random forest classifier extends its use to various applications such as classification of biological specimens and taxonomic in agriculture.

D. Multiple Linear Regression

A regression model that involves quite one variable is named multiple correlation Model. Multiple rectilinear regression (MLR) is that the method, wont to model the linear relationship between a variable and one or more independent variables. The dependent variable is sometimes termed as predictant and independent variables are called predictors. Multiple rectilinear regression (MLR) technique is predicated on method of least squares and doubtless the foremost widely used method in climatology for developing models to reconstruct climate variables from tree ring services. This crop yield prediction model is presented with the utilization of Multiple rectilinear regression (MLR) technique where the predictant is that the Production and there are seven predictors namely Year, Rainfall, Area of Sowing, Yield, Temperature, Season and Fertilizers (Nitrogen, Phosphorous and Potassium).

E. Support Vector Machine

SVM is considered as one of the major classification algorithms which can solve variety of classification tasks. Both nonlinear and linear data can be classified by using SVM. In the beginning, with the help of kernel functions, the data is mapped by SVM nonlinearly into a high-dimensional space. The linear optimal hyper plane divides the data with high margin, in which, the hyper plane is found by high-dimensional space.The advantages of SVM include the ability to handle high dimensional data, higher accuracy and robustness even in the presence of distortion in the training samples.However, the mapping of data to produce higher dimensional data and the selection of both kernel parameters and kernel function are very critical and it consumes much time.

F. Artificial Neural Network (ANN)

Artificial neural networks may be referred to as the combination of many classification techniques. ANN works on the principle of operation human biological nervous system, and its robustness is high at the interpolation of large amounts of ambiguous data. Artificial neural networks are popular among the commonly used techniques for the detection of plant diseases. It is used with various image pre-processing techniques for the extraction of useful information.

ANN has the advantage of being robust and providing better accuracy even with complex and noisy data. Nevertheless, the high accuracy could be achieved through repetitive training, which might take longer duration. It also suffers with scalability issues.

G. Convolutional Neural Network

In modern agriculture field, pest and disease identification is a major role of rice cultivation. Image classification by the use of deep convolutional neural networks of training and methodology used the facilitate a quick and easy system implementation. Pests and diseases are a threat to paddy production, especially in India, but identification remains to be a challenge in massive scale and automatically. Collecting images from Image Net dataset. The results [18] show that we can effectively detect and recognize the rice diseases and pests including healthy plant class using a deep convolutional neural network, with the best accuracy of 96.50%. The significantly high success rate makes the model a really useful advisory or early warning tool, and an approach that would be further expanded to support an integrated plant disease identification system to work in real cultivation conditions.

III. PROPOSED METHOD

The flow diagram shown in Fig.1 indicates the process steps used for Blast disease recognition using SVM, ANN and CNN classifier algorithms. This process consists of five various stages such as Image Acquisition, Image pre-processing, Image Segmentation, Feature Extraction and Classification.

In the image acquisition step, the digital camera is employed to capture the image512×512 pixels of rice leaf. The pixel of images is reduced to 256×256 pixels for increasing the accuracy. The gray scale image is created from there sized colour image for carrying out subsequent process. The HSV image is extracted from the RGB image during the phase of image pre-processing. The normal and the blast affected imageswhichareused in the processing steps are shown in Fig 2 (a) and (b).

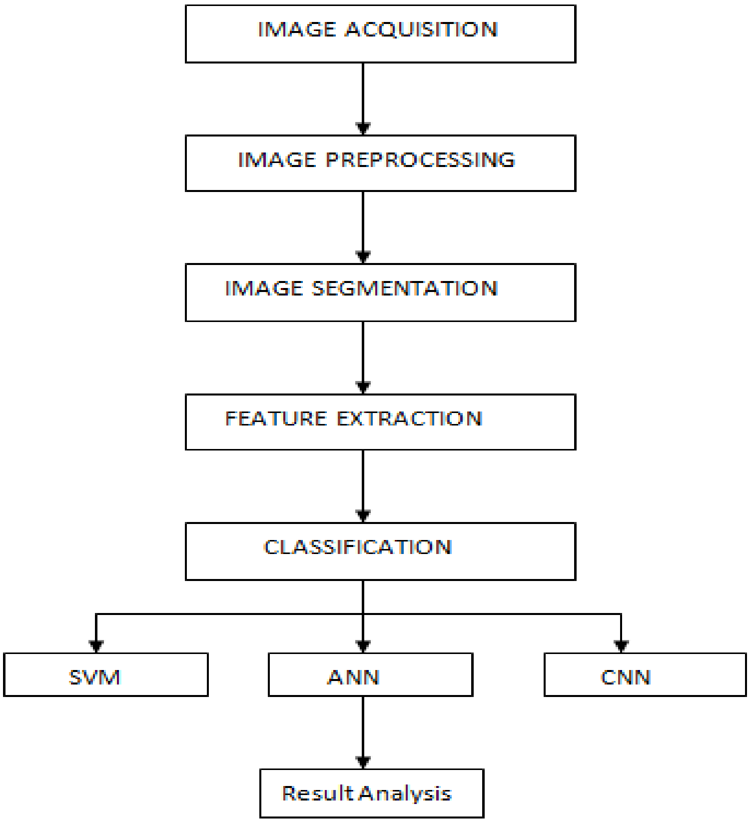


Figure 1: Flow diagram of proposed methodology

After the pre-processing of paddy leaf images, k-means clustering is employed for various k values for Image segmentation. Fig. 2exhibits the sample result of the k-Means clustering implemented images affected by blast disease.

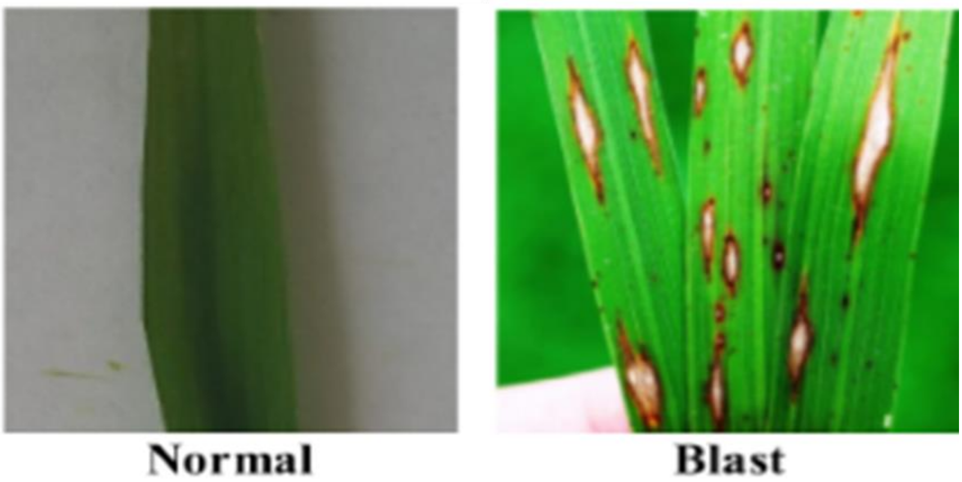
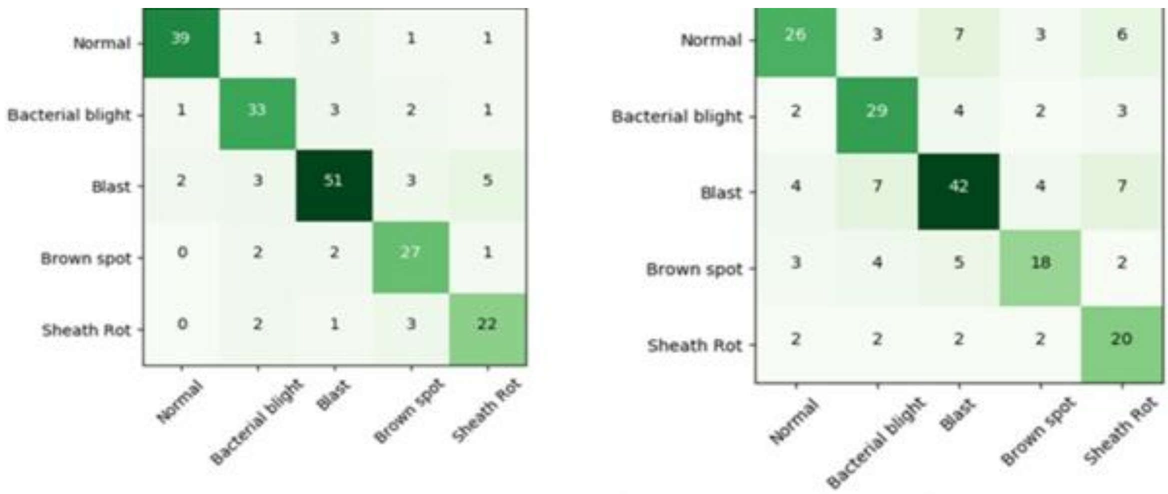


Figure 2: (a) Normal image (b) Blast affected image

IV EXPERIMENTAL RESULTS

Once the classification using CNN is completed, the confusion matrix is used to estimate its performance. In addition to this a comparative analysis is performed for this proposed method with existing the SVM classifier algorithm. The confusion matrix for the CNN based detection is shown in Fig. 3 (a) and (b) that of SVM is presented in Fig. 3(b). True positive, false positive, true negative and false negative values are evaluated from the confusion matrix.

CNN **SVM**



(a) CNN Confusion matrix (b) SVM Confusion matrix

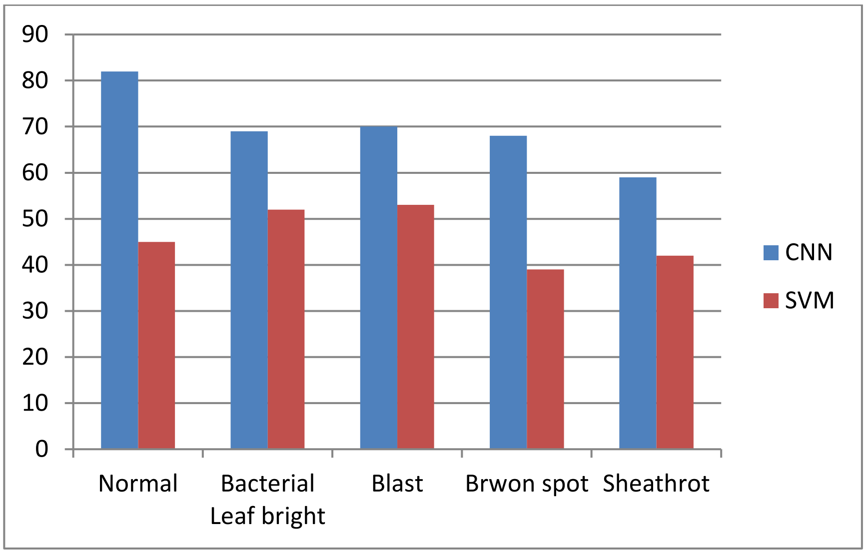
Figure 4: Confusion Matrix

The performance measures of the two classification methods tabulated and given the Table 1.

Disease	**SVM**				**CNN**			
	TN	TP	FN	FP	TN	TP	FN	FP
Sheath rot	163	20	8	18	173	22	6	8
Brownspot	166	18	14	11	168	27	5	9
Blast	127	42	22	18	136	51	13	9
Bacterialblight	153	29	11	16	161	33	7	8
Normal	156	6	16	11	161	39	6	3

Table 1: Confusion Matrix

(a) Precision Comparison



(b) F1 Score

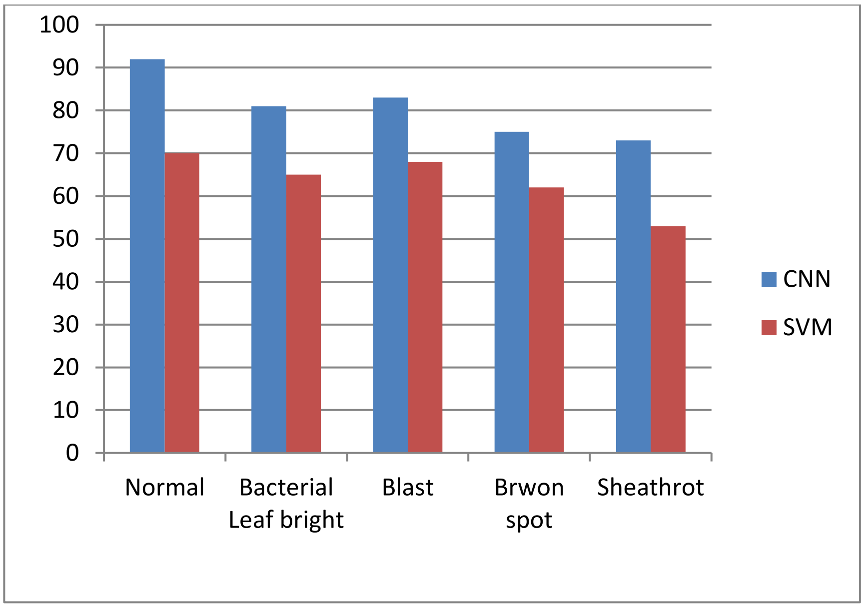
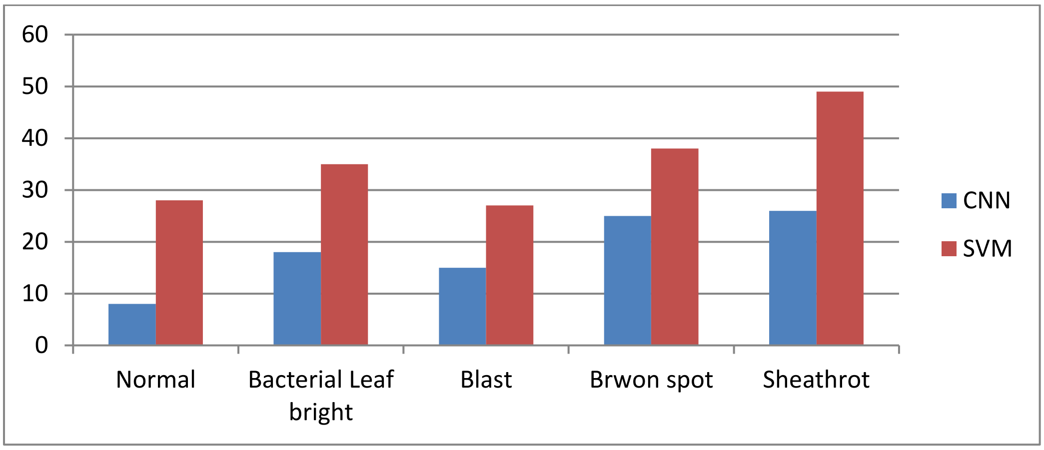


Figure 5: CNN and SVM Comparison

Precision Performance Measure

Fig.5(a) and (b) represents the F1 score and precision comparison plots for five different classes considered in this work. When employing the proposed CNN, it could be inferred from Fig.5 that, the achieved F1 score and precision values are 81% and 92% of normal, 69% and 79% of bacterial leaf blight, 71% and 85% of blast, 66% and 75% of brown spot and 61% and 74% of sheath rot. When using the existing KNN, the achieved F1 score and precision values are 45% and 69% of normal, 52% and 63% of bacterial leaf blight, 51% and 68% of blast, 43% and 62% of brown spot and 42% and 52% for sheath rot.

(a) False discovery rate



(b) False Positive rate

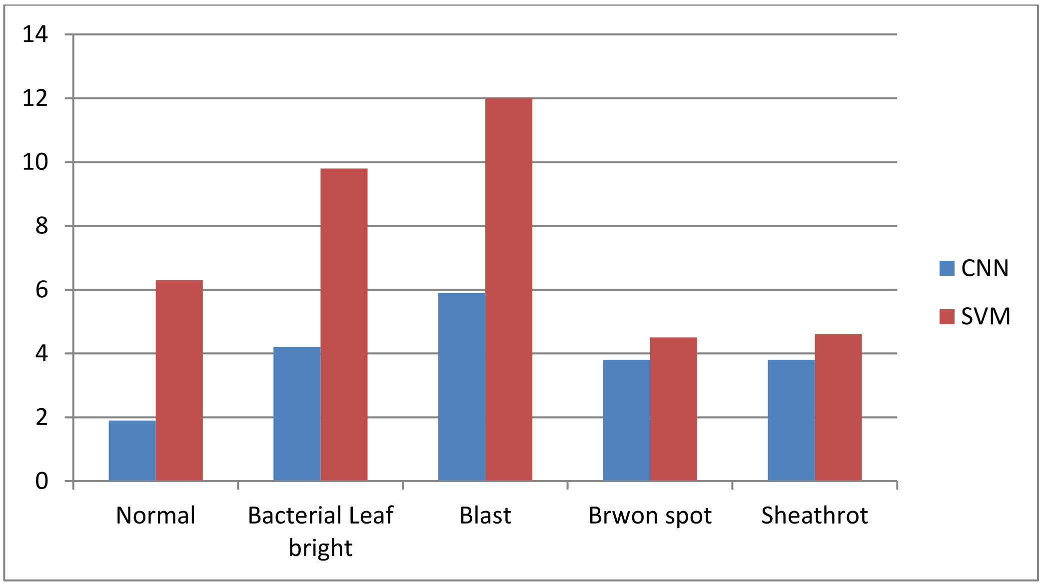
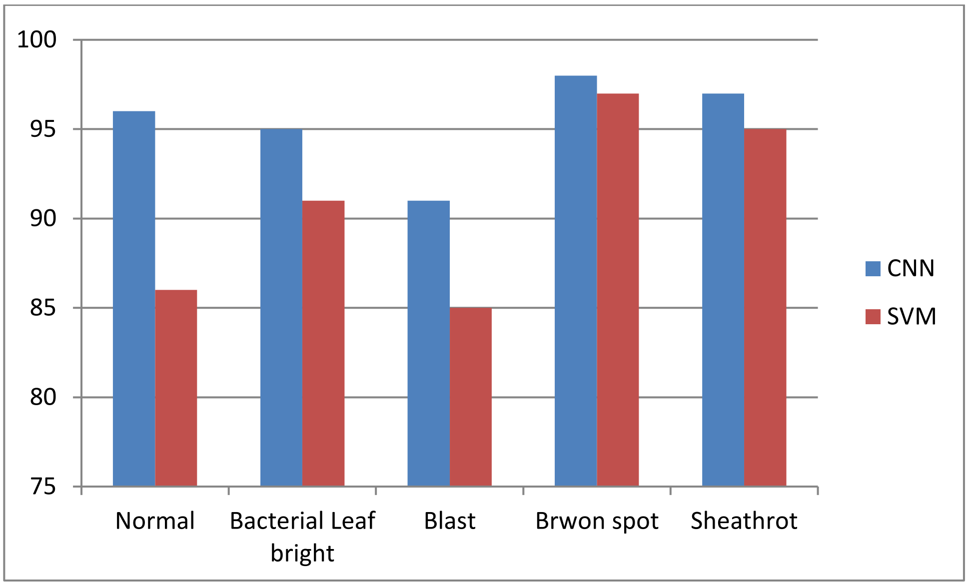


Figure 6: CNN and SVM Comparison

Fig. 6(a) and (b) depicts the FDR and FPR comparison results for five classes. While using the proposed CNN method, the achieved FPR and FDR values are 7% and 1.9% of normal, 19% and 4.4% of bacterial leaf blight, 19% and 6.1% of blast, 13% and 4.5% of brown spot and 27% and 4.1% of sheath rot. When existing SVM is used, the achieved FPR and FDR values are 29% and 6.3% of normal, 34% and 9.8% of bacterial leaf blight, 30% and 6.1% of blast, 25% and 5.9% of brown spot and 48% and 9.8% of sheath rot.

(a) Negative Predictive value



(b) Specificity

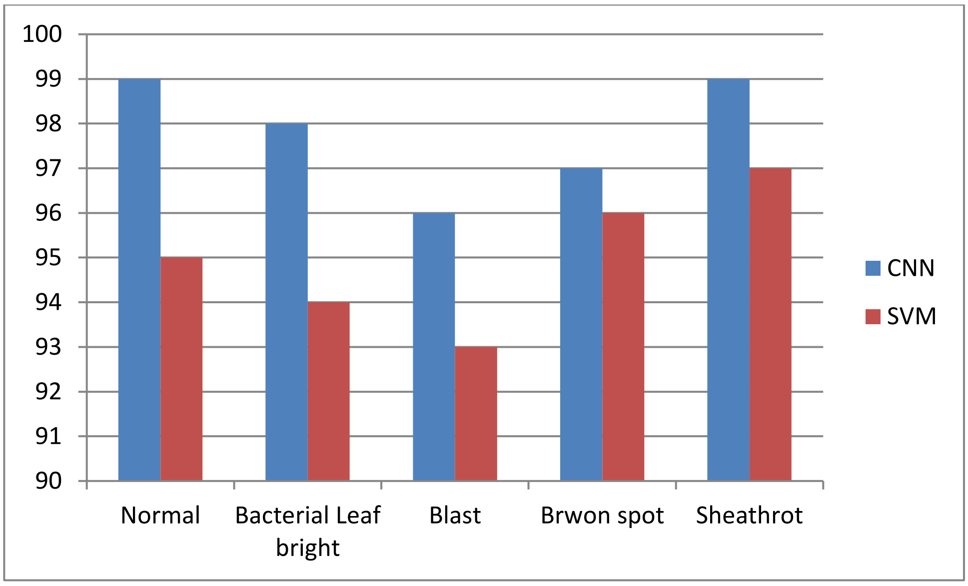
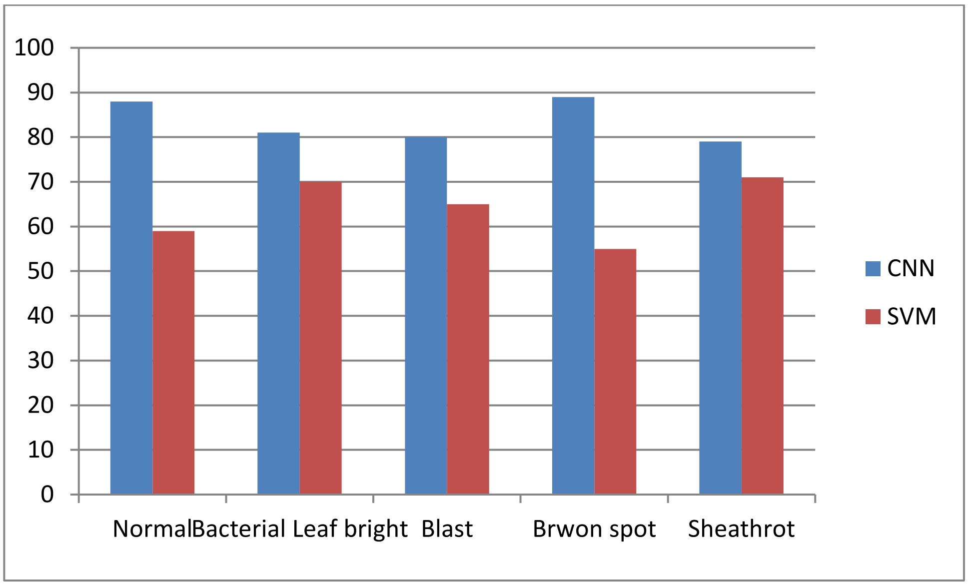


Figure 7: CNN and SVM Comparison

Fig.7 (a) and (b) shows comparative analysis of the NPV and specificity for five classes. Based on the results obtained for the proposed DNN, the achieved NPV and specificity values are 96% and 98% of normal, 95% and 96% of bacterial leaf blight, 93% and 95% of blast, 97% and 96% of brown spot and the same 97% and 98% of sheath rot. When using the existing KNN, the achieved NPV and specificity values are 88% and95%of normal, 92% and 93% of bacterial leaf blight, 85% and 88% of blast, 91% and 94% of brown spot and 93% and 89% of sheath rot.

(a) **True Predictive Rate**



(b) Accuracy

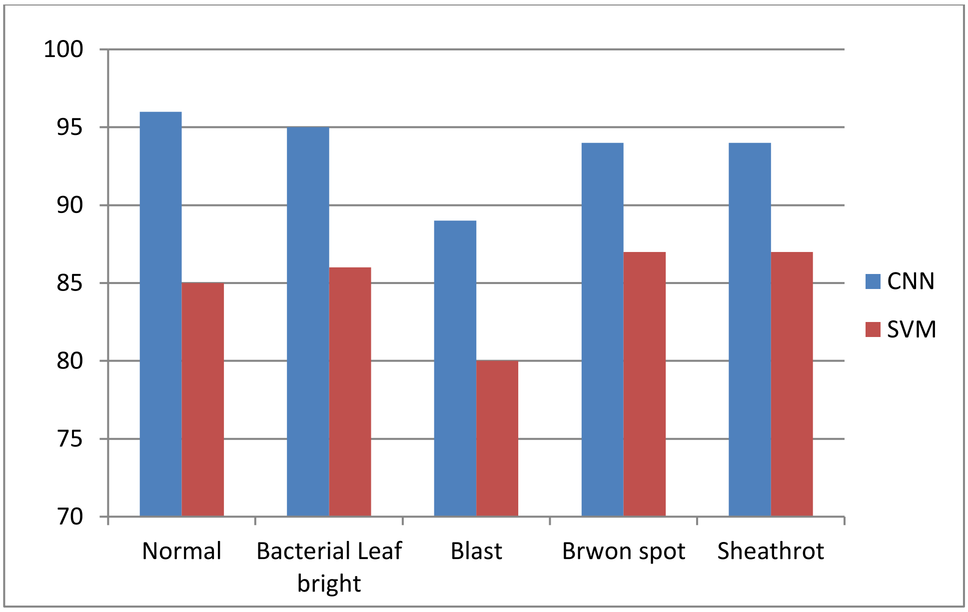


Figure 8: CNN and SVM Comparison

Fig.8 (a) and (b) represents the TPR and accuracy comparison graphs for five classes. When using our proposed CNN, the achieved NPV and specificity values are 86% and 96% of normal, 84% and 93% of bacterial leaf blight, 78% and 89% of blast,86% and 93% of brown spot and the 76% and 92% of sheath rot. When using the existing SVM, the achieved NPV and specificity values are 58% and 86% of normal,72% and 88% of bacterial leaf blight.

V. CONCLUSION

Images were collected from the UCI repository, once the classifiers are trained with both blast affected and normal rice leaves they are tested for the performance accuracy. The convolutional neural network is used as the classification method instead of ANN classifier. In addition to this, four different diseases of paddy such as, brown spot, sheath rot, blast and bacterial blight are considered in this chapter. The background of the image is eliminated in pre-processing stage, texture and colour extraction of images are done in the feature extraction stage. Finally, convolutional neural network is used for the purpose of classification of diseased and normal leaves using the data set of images. The existing SVM method is also involved for the performance comparison of entire dataset.

References

1. Gayathri Devi T and Neelamegam P, “Paddyleaf disease detection using SVM with RBFN classifier”, International Journal of Pure and Applied Mathematics, vol. 117, no. 15, pp. 699-710, 2017.
2. Vanitha V, “Rice disease detection using deep learning”, International Journal of Recent Technology and Engineering, vol. 7, pp. 2277-3878, no. 5S3, 2019.
3. Hemanta Kalita, Shikhar Kr Sarma and Ridip Dev Choudhury, “Expert system for diagnosis of diseases of rice plants: prototype design and implementation”, International Conference on Automatic Control and Dynamic Optimization Techniques, pp. 723-730, 2016.
4. Sujeet Varshney and Tarun Dalal, “A novel approach for the detection of plant diseases”, IJCSMC, vol. 5, no. 7, pp. 44-54, 2016.
5. Mohan K J, Balasubramanian M and Palanivel S, “Detection and recognition of diseases from paddy plant leaf images”, International Journal of Computer Applications, vol. 144, no. 12, 2016.
6. Shaik Anzala Noor, Nimmala Anitha, Sreenivasulu P, “Identification of fungal diseases in paddy fields using image processing techniques”, International Journal of Advance Research, Ideas and Innovations in Technology, vol. 4, no. 3, 2018.
7. Santanu Phadikar, Jaya Sil and Asit Kumar Das, “Classification of rice leaf diseases based on morphological changes”, International Journal of Information and Electronics Engineering, vol. 2, no. 3, pp. 460-463, 2012.
8. Leonard Gianessi P, “Importance of pesticides for growing rice in south and south east asia”, International Pesticide Benefit Case Study, vol. 108, 2014.
9. Sachin Khirade D and Patil A B, “Plant disease detection using image processing”, International conference on computing communication control and automation, pp. 768-771, 2015.
10. Kahar M A, Sofianita Mutalib and Shuzlina Abdul-Rahman, “Early detection and classification of paddy diseases with neural networks and fuzzy logic”, in 17th International Conference on Mathematical and Computational Methods in Science and Engineering, pp. 248-257, 2015.
11. Agricultural crops: Cereals: Paddyhttp://agritech.tnau.ac.in/crop_protection/crop_prot_crop%20diseases_cereals_paddy.html (Last accessed December 2018).
12. Gupta, S.K.; Rai, A.K.; Kanwar, S.S.; Chand, D.; Singh, N.K.; Sharma, T.R. The single functional blast resistance gene Pi54 activates a complex defence mechanism in rice. J. Exp. Bot. 2012, 63, 757–772.

13. Gowda, M., Shirke, M. D., Mahesh, H., Chandarana, P., Rajamani, A. and Chattoo, B. (2015), 'Genome analysis of rice-blast fungus magnaporthe oryzae field isolates from southern India', Genomics data **5**,284–291.
14. Martin-Urdiroz, M., Oses-Ruiz, M., Ryder, L. S. and Talbot, N. J. (2016),'Investigating the biology of plant infection by the rice blast fungus magnaporthe oryzae', Fungal Genetics and Biology **90**,61–68.
15. Maithili Desai, Anmolika Goyal, Shreya Ghose, Sheona Das, "Plant disease Detection: A Comprehensive Survey 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), IEEE, published in 2020.
16. Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K. J., et al. (2011). Discovery of Rare Mutations in Populations: TILLING by Sequencing. Plant Physiol. 156, 1257–1268. doi: 10.1104/pp.110.169748.
17. Yan, X. and Talbot, N. J. (2016), 'Investigating the cellbiology of plant infection by thericeblast fungus magnaportheoryzae', Current opinion in microbiology **34**, 147–153.
18. Rao, R. (2016),'Jaya: Asimple and new optimization algorithm for solving constrained and unconstrained optimization problems', International Journal of Industrial Engineering Computations **7** (1), 19–34.