ISSN:0975-9646

Leaf Blast Rice Disease Prediction Model Based on Environmental Factors Using Binary Classifiers

***Shwetha M.K*[#1], *Dr. Nagarathna*[*2]**

[#] *Assistant Professor*
Dept. of Computer Science and Engineering
PESCE Mandya, Karnataka, India

1shwethamanju@gmail.com

[*] *Professor*
Dept. of Computer Science and Engineering
PESCE Mandya, Karnataka, India

2nagarathna@pesce.ac.in

***Abstract* - Rice is a very important and essential food for large part of the world's population. Rice crop yield is affected by the diseases caused by the pathogens, among which the rice blast disease stands out. The progression of the rice blast begins from the seedling stage and is followed by several stages namely leaf blast, neck blast, panicle and node blast. The rice productivity gets affected by the leaf blast. In order to address the productivity loss of rice due to leaf blast, a binary prediction model is developed for prediction of leaf blast disease. The leaf blast disease related environmental data's are given as input to the K- Nearest Neighbour (KNN) and Support Vector Machine (SVM) classification algorithms. This study demonstrates that KNN algorithm is better than SVM algorithm in predicting the leaf blast diseases.**

***Keywords* - Machine learning, K- Nearest Neighbour, Support Vector Machine, leaf blast disease.**

I. INTRODUCTION

Rice is one of the oldest agricultural plants in the world. It is a plant of tropical origins broadly cultivated in the world in a variety of areas [15]. India has the largest area under the rice accounting with 28.5% of the global rice and produce 22% of world production. In India, the major paddy growing states are Uttar Pradesh, West Bengal, Orissa, Chhattisgarh, AndraPradesh and Karnataka [10]. Rice production has been challenged by recent changes in crop production technologies such as heavy chemical fertilization, continuous flooding aggravates that impact disease occurrence [13]. Rice Diseases greatly reduce yield of rice. They are mainly caused by bacteria, viruses, or pathogens. The diseases caused by pathogens are Rice Blast, Sheath Blight, Brown Spot, Leaf Scald, Narrow Brown Spot, and Root knot, Stem Rot, Sheath Rot, Bakanae and False Smut.

Among rice diseases, the rice blast is one which is caused by the pathogen Magnaporthe grisea, causing significant losses in cultivated varieties yield when the environmental conditions are favourable [15]. The pathogen can infect and causes lesions on the shoot. The progression of the rice blast begins from the seedling stage and is followed by several stages namely leaf blast, neck blast, panicle and node blast. In Leaf Blast the smallest and greyest spots appear on the leaves. It is as shown in Fig.1. If the disease is severe, the plants die at the seedling stage. In Node Blast the pathogen infects the nodes of the plant nodes and nodes that turn black and get weakened, as a result all parts of the plant that are above the infected node will die. It is shown in Fig.2. In Neck and Panicle Blast, the neck shrinks and is covered with grey mycelium. If neck blast occurs before the milk phase, the whole panicle can die prematurely as shown in Fig.3. The pathogen also causes brown lesions on panicle branches of the plant. It is shown in Fig.4 [18].

Fig. 1 Leaf Blast

Fig. 2 Node Blast

Fig. 3 Neck Blast

Fig. 4 Panicle Blast

A. Production loss of rice yield due to the incidence of rice diseases

As far as rice diseases are concerned, according to farmers experience rice blast disease leads to more than 25% yield loss and some time loss up to 50%. The other diseases namely udbatta disease, sheath blight, brown leaf spot and ganaka disease are also responsible for significant yield loss [14].

Field survey was undertaken in the major rice growing regions of Karnataka, a southern state of India which revealed the presence of bacterial leaf blight incidence ranging from 12 to 75% [20]. The pathogen was isolated from the seeds and infected plant materials collected during field survey [17]. Maintenance of perennial rice crops, especially the diseases affecting production and post-harvest life, requires close monitoring [6].

One of the major activities under All India Coordinated Rice Improvement Program of Indian Institute of Rice Research is Production Oriented Survey (POS). The POS reports of severe incidence of Leaf blast disease of rice in India during the period 2001-2014 is shown in Table I [13].

TABLE I
INCIDENCE OF LEAF BLAST DISEASE OF RICE IN INDIA DURING THE PERIOD 2001-2014.

State	Intensity/ Incidence
AndhraPradesh	up to 40%
Telangana	up to 40%
ArunachalPradesh	up to 40%
Assam	up to 50%
Bihar	up to 40%
Chhattisgarh	up to 40%
Gujarat	5-30%
Haryana	up to 40%
Jammu & Kashmir	5-30%
Himachal Pradesh	up to 40%
Jharkhand	up to 50%
Karnataka	up to 35%
Madhya Pradesh	55-65%
Maharashtra	2-30%
Odisha	5- 30%
Punjab	up to 40%
Tamil Nadu	up to 40%
Tripura	up to 40%
Uttar Pradesh	5- 40%
Uttarakhand	15- 50%
West Bengal	up to 40%

Leaf blast is influenced by rice variety, environmental factors and pathogens. The environment controls whether the disease develops or not. The environmental conditions also affect the ability of the pathogen to cause diseases [2].

Expert's naked eye observation is one of the main methods implemented to detect and identify plant diseases [11]. However, large farms require constant and continuous monitoring by experts, which can be costly. Furthermore, in some developing countries, farmers may have to go a long way to contact experts, which can be very costly and time consuming to consultants. Thus prediction of rice diseases is essential.

II. RELATED WORK

In this paper [9] the rice blast disease is predicted by applying multiple regression, artificial neural network and support vector machine (SVM) machine learning algorithms on the environmental datasets. The relationship between the environment and disease (rice blast) development is comprehended using empirical approaches. This case study results weather-based prediction model of plant diseases developed using SVM machine learning algorithm is better than other approaches.

Alvin R Malicdem et al. [7] developed models for predicting the rice blast disease using ANN, SVM binary classifiers and regression models. Regression models are used to estimate the severity of the disease. The aim of this study is helping to prevent or at least mitigate the spread of such disease. This study results that SVM model developed for predicting the disease gives accurate results compared to ANN model.

S. Nithya et al. [4] has proposed an efficient recommendation system to predict the symptoms based paddy diseases with high similarity. The data related to paddy disease related are collected from heterogeneous websites such as Agropedia and blogs. The Hadoop, Hive and HiveQL tools are used to process and analyze the data. The collected documents are represented in the form vector using Vector Space model and the TF-IDF ranking is used to calculate the weight of the vector. The similarity between the document vector and query vector is calculated using the cosine similarity measure. The disease based similarity measure is used in the proposed method.

Wen-Liang Chen et al. [1] has developed the RiceTalk project to detect the rice blast using IoT devices. Non-image data that can be automatically generated using agriculture sensors. AI techniques are used to train and analyzed the data in real time. The RiceTalk project treats the AI model as an IoT device.

In this paper [5] various rice diseases like brown spot disease, leaf blast disease and bacterial blight etc. are identifies and classified by image processing system. In the proposed sysytem, first rice disease is identified by applying Haar-like features and AdaBoost classifier. Second, the disease is recognized or classified using Scale Invariant Feature Transform (SIFT) feature by applying binary classifiers namely k-Nearest Neighbour and Support Vector Machine.

Farhana Tazmim Pinkian et al. [3] has proposed an automated system for diagnosis three common paddy leaf diseases (Brown spot, Leaf blast, and Bacterial blight) and pesticides and/or fertilizers are advised according to the severity of the diseases. K-means clustering is used to separate the affected part from the paddy leaf image. Visual themes (color, texture and shape) are used as features for the classification of these diseases. Types of paddy leaf diseases are identified by the support vector machine (SVM) classification.

III. PROPOSED SYSTEM

The leaf blast prediction model is shown in Fig 5. A prediction model is developed based on the relationship between the environmental conditions and the leaf blast disease occurrence. This system helps to reduce the cost of production of ice and also helps to reducing chemical use promotes environmental safety for the operator and consumers.

The proposed prediction model uses data samples for the prediction of rice leaf blast disease where each data sample describing a leaf blast report, Observation Year, Standard Week and the corresponding weather information. Feature Selection technique is used to select most significant features such as Maximum Temperature (Tmax) , Minimum Temperature (Tmin), Relative humidity (RH1 and RH2), and Rain fall (RF) in mm, Wind Speed (WS) in km/h, Sunshine hour (SSH) and Evaporation (EVP) in mm from the dataset. Feature scaling was then applied to ensure that disease and weather features were in similar scale. The general rule is to take a feature value X_{mn} (at row m, column n) and replace it with a new value using the formula.

$$fs(\mathrm{x_m}, \mathrm{n}) = \frac{\mathrm{x_m} - \mu_\mathrm{n}}{\mathrm{s_n}} \quad (1)$$

Where μ_n and s_n are the average and standard deviation of all X_{mn} values within column n of the data set [9].

Binary classification models can be used to predict the occurrence of rice leaf blast disease. In this paper K-Nearest Neighbours (KNN) and Support Vector Machine (SVM) machine learning techniques are used in predicting rice leaf blast disease.

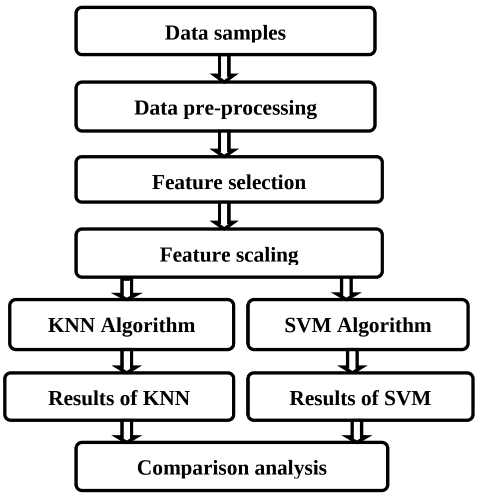

Fig. 5 The work flow of leaf blast disease prediction model.

A. K-Nearest Neighbours (KNN)

In training phase KNN algorithm simply stores all the data samples and it pompons the classification until the new data point is encountered. KNN classifies a new data point based on the similarity. Each data sample is a set of features represented as X, where

$$X = \{ Tmax_x, Tmin_x, RH1_x, RH2_x, RF_x, WS_x, SSH_x, EVP_x\} \quad (2)$$

and new data sample is represented as Y, where

$$Y = \{ Tmax_y, Tmin_y, RH1_y, RH2_y, RF_y, WS_y, SSH_y, EVP_y\} \quad (3)$$

Then Euclidean distance metric is used to calculate the distance between the each data sample (X) and new data sample (Y). This Euclidean distance (D) metric is defined as

$$D= \sqrt{Z_{Tmax} + Z_{Tmin} + Z_{RH1} + Z_{RH2} + Z_{RF} + Z_{WS} + Z_{SSH} + Z_{EVP}} \quad (4)$$

Where $Z_{Tmax} = (Tmax_y - Tmax_x)^2$ (5)

$Z_{Tmin} = (Tmin_y - Tmin_x)^2$ (6)

$Z_{RH1} = (RH1_y – RH1_x)^2$ (7)

$Z_{RH2} = (RH2_y – RH2_x)^2$ (8)

$Z_{RF} = (RF_y – RF_x)^2$ (9)

$Z_{WS} = (WS_y – WS_x)^2$ (10)

$Z_{SSH} = (SSH_y – SSH_x)^2$ (11)

$Z_{EVP} = (EVP_y – EVP_x)^2$ (12)

Then select k, where k indicates the number of nearest neighbours. Usually k value is an odd number if the number of classes is 2. To determine the exact value for k, the model is to be tested for each and every expected value of k. In this paper KNN is implemented with k value as 1, 3 and 5. And finds the nearest neighbours based on k value.

B. Support Vector Machine (SVM)

It is a binary linear classification whose decision boundary is explicitly constructed to minimize generalization error [7]. The SVM classifies data points by assigning them to one of two disjoint half spaces either in the pattern space or in a higher-dimensional feature space. This classification is done using Kernel Methods. In this paper, linear kernel and RBF kernel methods are used [16].

Linear kernel is denoted as LK(X,Y) where X and Y are the data samples shown in equations (2) and (3), which is defined as

$$LK(X,Y) = (\mathrm{Tmax}_x * \mathrm{Tmax}_y + \mathrm{Tmin}_x * \mathrm{Tmin}_y + \mathrm{RH1}_x * \mathrm{RH1}_y + \mathrm{RH2}_x * \mathrm{RH2}_y + \mathrm{RF}_x * \mathrm{RF}_y + \mathrm{WS}_x * \mathrm{WS}_y + \mathrm{SSH}_x * \mathrm{SSH}_y + \mathrm{EVP}_x * \mathrm{EVP}_y) \quad (13)$$

RBF kernel is denoted as RK(X,Y) where X and Y are the data samples shown in equations (2) and (3), which is defined as

$$RK(X,Y) = \exp(-\gamma * \mathrm{S}) \quad (14)$$

The $\boldsymbol{\gamma}$ parameter defines how far the impact of a single training example reaches**.** It ranges from 0 to 1, in this kernel 0.1 is the value of it. Then S is defined as

$$S = (\mathrm{Tmax}_x - \mathrm{Tmax}_y{}^2 + \mathrm{Tmin}_x - \mathrm{Tmin}_y{}^2 + \mathrm{RH1}_x - \mathrm{RH1}_y{}^2 + \mathrm{RH2}_x - \mathrm{RH2}_y{}^2 + \mathrm{RF}_x - \mathrm{RF}_y{}^2 + \mathrm{WS}_x - \mathrm{WS}_y{}^2 + \mathrm{SSH}_x - \mathrm{SSH}_y{}^2 + \mathrm{EVP}_x - \mathrm{EVP}_y{}^2) \quad (15)$$

IV. EXPERIMENTATION AND RESULT DISCUSSION

A. Dataset

The important environmental features that affect the occurrence of rice leaf blast are Rainfall, Minimum Temperature, Maximum Temperature, Humidity, Solar Radiation, Evaporation, Sun shine hour, Brightness and Wind Speed [7]. Temperature, sunshine hour, relative humidity, rain fall, evaporation time and the length of dew period can influence multiplication and infection of the causal fungus to host plants [9].

The crop-pest-weather database for rice documents weekly pest records (34,472) for 11 insect pests and diseases in rice along with corresponding weather across 12 important locations spread across India [19]. The database is useful for the researches to predict rice diseases and develop forecasting models. Example for Rice leaf blast disease data sample is shown in Fig 6.

Tmax(°C)	Tmin(°C)	RH1(%)	RH2(%)	RF(mm)	WS(kmph)	SSH(hrs)	EVP(mm)	Pest Value
23.7	12.8	76	60.9	73.4	5.6	6.8	2.9	0
24.4	11.7	62.3	48.3	0	5.3	10	3.6	0
23.6	11.1	65.4	47	0	5	8.8	2.7	0
21.9	8.8	65.6	49.1	0	5.4	9.1	2.5	0
22.2	9.8	64.6	46.3	0	5.1	9.3	2.7	0

Fig 6. Rice leaf blast disease data example from 1984.

B. *Result discussion*

In this work, the data set is divided into two sets namely training set and the testing set, 70% of the data set was used as a training set and 30% of the data set was used as a test set.

In KNN classifier the data samples are randomly assigned to training and testing set and the classification accuracy is reported [21]. Performance of KNN classifier is better for the small value of k. If there are n data samples in a dataset, all these data samples are used by KNN classifier while classifying new data. Majority voting scheme is used by KNN classifier to classify the new data [12]. In view of this 1, 3 and 5 values are used for k. The performance outcome of the KNN classifiers for these k values is reported in Table II and their comparison is shown in Fig 7 respectively. This comparison shows KNN classifier results better accuracy when K=5.

TABLE II
PERFORMANCE OUTCOME OF THE KNN CLASSIFIER

Data samples	KNN classifier with K=1	KNN classifier with K=3	KNN classifier with K=5
52	0.75	0.625	0.8125
104	0.84375	0.875	0.9375
156	0.829787	0.851064	0.893617
208	0.85714	0.84127	0.92063
260	0.9359	0.85897	0.88462
312	0.92553	0.88298	0.94681
364	0.88182	0.88182	0.90909
416	0.928	0.936	0.944
468	0.92199	0.9078	0.9078
520	0.90385	0.87821	0.91026

In SVM classifier the data samples are randomly assigned to training and testing set and the classification accuracy is reported. This paper presents the linear kernel and RBF kernel functions for SVM classifier. The performance outcome of the SVM with linear kernel and SVM with RBF kernel classifiers is reported in Table III and their comparison is shown in Fig.8 respectively. This comparison results the accuracy of linear kernel is high compared to RBF kernel.

TABLE III
PERFORMANCE OUTCOME OF SVM CLASSIFIERS WITH LINEAR KERNEL AND RBF KERNEL

Data samples	SVM with linear kernel	SVM with RBF kernel
52	0.619048	0.714286
104	0.833333	0.785714
156	0.809524	0.809524
208	0.809524	0.809524
260	0.884615	0.865385
312	0.888	0.856
364	0.890411	0.883562
416	0.892216	0.874251
468	0.87766	0.867021
520	0.889423	0.889423

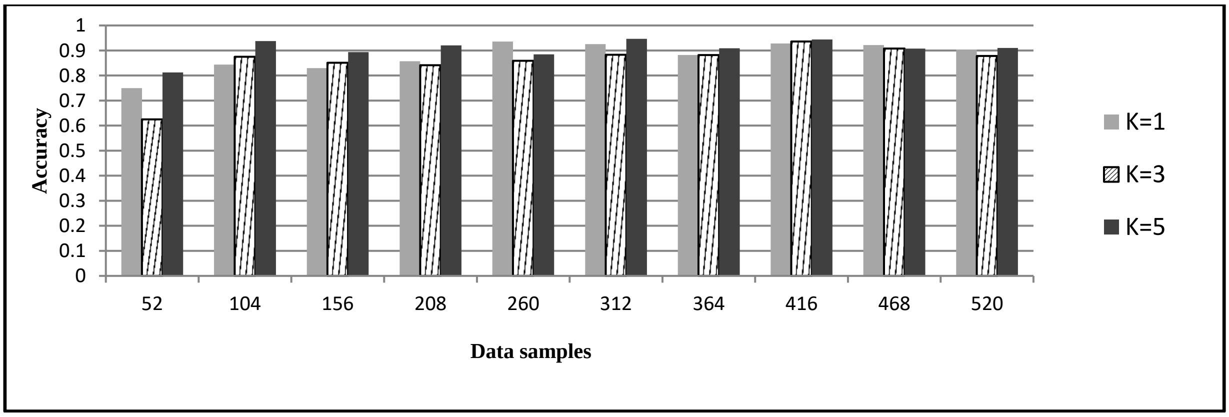

Fig. 7 Comparison of outcome of the KNN classifier.

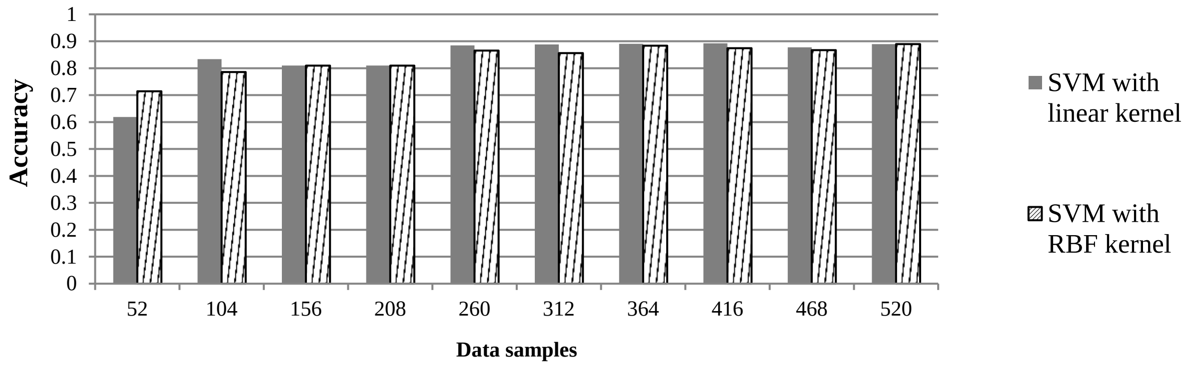

Fig 8. Comparison of outcome of SVM classifiers with linear kernel and RBF kernel.

The computational complexity of the KNN classifier with k=5 is compared with that of SVM classifier with linear model. The performance outcome of the KNN classifier with k=5 and SVM with linear kernel classifiers is reported in Table IV and their comparison is shown in Fig 9 respectively. This comparative study results that the accuracy of the KNN classifier is significantly higher than the SVM classifier in detecting leaf blast disease.

TABLE IV
PERFORMANCE OUTCOME OF THE KNN CLASSIFIER WITH K=5 AND SVM WITH LINEAR KERNEL CLASSIFIER

Data samples	KNN classifier with k=5	SVM with linear kernel
52	0.8125	0.619048
104	0.9375	0.833333
156	0.893617	0.809524
208	0.92063	0.809524
260	0.88462	0.884615
312	0.94681	0.888
364	0.90909	0.890411
416	0.944	0.892216
468	0.9078	0.87766
520	0.91026	0.889423

The number of data samples used in this study is very low, this is the one of the limitation of the proposed system. The analysis of data should be carried out in future with a larger database. This study can be further extended by applying other machine learning algorithms for the prediction of rice diseases [8].

V. CONCLUSION

Disease severity of rice leaf blast on different varieties with significant relationship with environmental variables such as maximum temperature, minimum temperature, precipitation and relative humidity. The proposed system compares the performance of the KNN and SVM classifiers for the prediction of leaf blast disease. The maximum classification accuracies for the KNN and SVM classifiers were found to be 94.68% and 89.22%, respectively. The maximum classification accuracy of KNN classifier was obtained for a k value of 5. The maximum classification accuracy of the SVM classifier was obtained with the linear kernel. This comparison results that the accuracy of the KNN classifier is significantly higher than the SVM classifier in predicting the leaf blast disease.

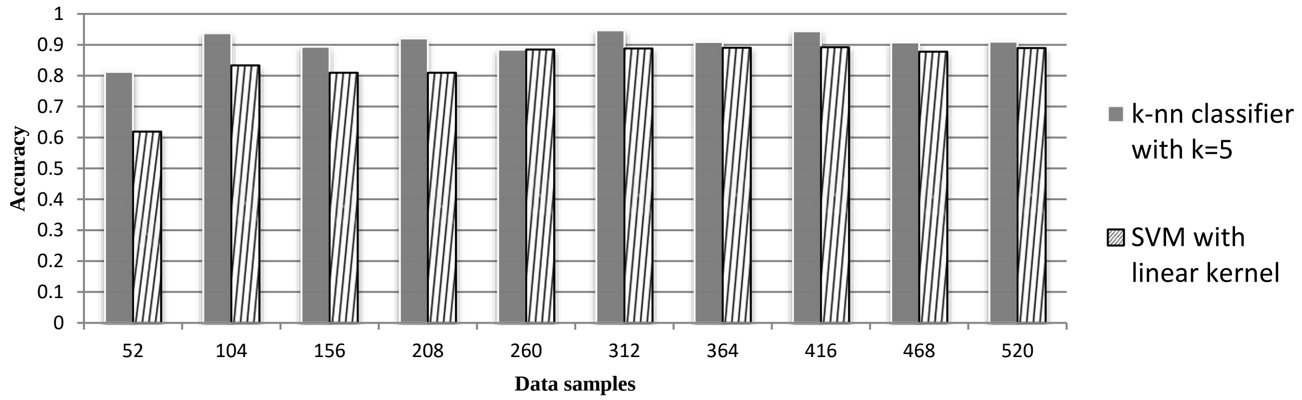

Fig. 9 Comparison of outcome of the KNN classifier with k=5 and SVM with linear kernel classifier.

REFERENCES

[1] Wen-Liang Chen, Yi-Bing Lin, Fung-Ling Ng, Chun-You Liu, Yun-Wei Lin, " RiceTalk: Rice Blast Detection using Internet of Things and Artificial Intelligence Technologies", *IEEE Internet of Things Journal,* Volume: 7, Issue: 2, Feb. 2020, DOI: 10.1109/JIOT.2019.2947624.

[2] Shital Shinde and Atul Srivastava, "Plant Protection And Pest Control Using Low Cost Sensor Based Embedded System", *International Journal Of Advanced Research And Engineering*, Volume- 07, Issue -07, July 2017, ISSN: 2319-8346.

[3] Farhana Tazmim Pinki, Nipa Khatun and S.M. Mohidul Islam, "Content based Paddy Leaf Disease Recognition and Remedy Prediction using Support Vector Machine", *20th International Conference of Computer and Information Technology* (ICCIT), 2017 IEEE.

[4] S. Nithya, S. Savithri, G. Thenmozhi and Dr. Kalaimani Shanmugham, "Symptoms Based Paddy Crop Disease Prediction And Recommendation System Using Big Data Analytics", *International Journal of Computer Trends and Technology (IJCTT)* – Special Issue April – 2017, ISSN: 2231 - 2803.

[5] K. Jagan Mohan, M. Balasubramanian and S. Palanivel, "Detection and Recognition of Diseases from Paddy Plant Leaf Images", *International Journal of Computer Applications*, Volume 144 – No.12, June 2016.

[6] S.A. Ramesh Kumar and K. Ramesh Kumar, "A Novel and High Speed Technique for Paddy Crops Disease Prediction in Wireless Tele-Agriculture Using Data Mining Techniques", *Middle-East Journal of Scientific Research*, 22(9), 2014.pp. ISSN: 1990-9233.

[7] Alvin R. Malicdem and Proceso L. Fernandez, "Rice Blast Disease Forecasting for Northern Philippines", *Wseas Transactions On Information Science And Applications*, Volume 12, 2015, E-ISSN: 2224-3402.

[8] Rajkumar Palaniappan, Kenneth Sundaraj and Sebastian Sundaraj, "A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals", *BMC Bioinformatics* 2014, http://www.biomedcentral.com/1471-2105/15/22.

[9] Rakesh Kaundal, Amar S Kapoor and Gajendra PS Raghava, "Machine learning techniques in disease forecasting: a case study on rice blast prediction", *BMC Bioinformatics* 2006, http://www.biomedcentral.com/1471-2105/7/485.

[10] Mali Patil Vijay Kumar and Nahar Singh, "An Economic Analysis of Paddy Production in Raichur District, Karnataka, India", International Journal of Current Microbiology and Applied Sciences,Special Issue-9 pp. 183-193, 2019, ISSN: 2319-7706.

[11] "Effects of organic fertilizers on insect pest and diseases of rice", h*ttps://www.researchgate.net/publication /267260205* (Web link).

[12] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan and Uri shaft, "When Is Nearest Neighbor meaningful?" *https://www.researchgate.net/publication/2845566 (*Web link).

[13] Chethana B.S., Deepak, C.A., Rajanna, M.P., Ramachandra, C. and Shivakumar, N. , "Current Scenario of Rice Diseases in Karnataka", *International Journal of science and nature*, Volume- 7, 2016, ISSN: 2229-6441 (Journal).

[14] Elumalai Kannan Parmod Kumar Kedar Vishnu Hanna Abraham, "Assessment of Pre And Post Harvest Losses Of Rice And Red Gram In Karnataka", *Research Report*: Ix/Adrtc/153.

[15] Leonardo Cirilo da Silva Soares, Juan Piero Antonio Raphael, Rafael Pivotto Bortolotto, Douglas Dalla Nora and Eduardo Müller Gruhn, "Blast disease in rice culture", *Brazilian Journal of Applied Technology for Agricultural Science*, Guarapuava-PR, v.7, n.2, p.109-119, 2014.

[16] Palanian, Rajkumar , Kenneth Sundaraj and Sebastian Sundaraj ," A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals", *BMC Bioinformatics*, 2014, DOI: 10.1186/1471-2105-15-223.

[17] Shivalingaiah And S. Umesha, "Characterization Of Xanthomonas Oryzae Pv. Oryzae From Major Rice Growing Regions Of Karnataka", *The Bioscan, An International Quality Journal Of Life Sciences*, 6(1),5-10,2011.

[18] G.S. Laha, Ram Singh, D. Ladhalakshmi, S. Sunder, M. Srinivas Prasad, C.S. Dagar, and V. Ravindra Babu, "Importance and Management of Rice Diseases: A Global Perspective", *Springer International Publishing* AG 2017, DOI 10.1007/978-3-319-47516-5_13.

[19] https://www.crida.in

[20] https://www.mysciencework.com

[21] https:// www.biomedcentral.com