Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

Predictive Model for Rice Blast Disease on Climate Data Using Long Short-Term Memory and Multi-Layer Perceptron: An Empirical Study on Davangere District

Varsha M.[1*] , Dr. Poornima B.[2] , Dr. Vinutha H.P.[3] and Pavan Kumar M.P[4].
[1,2,3]Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology
[4]Department of Information Science and Engineering, Jawaharlal Nehru National College of Engineering
Email: [1]mvarsha16@gmail.com* (corresponding author),[2]poornimateju@gmail.com , [3]vinuprasad.hp@gmail.com , [4]pavankumarjnnce@gmail.com

ABSTRACT

Among various diseases of paddy affecting rice production and cultivation, blast majorly called as rice blast disease has the predominant impact. Thus, monitoring and early prediction of the occurrence of rice blast disease are very important and would be largely helpful for prevention of blast disease. Here, we have proposed LSTM and MLP based machine learning models for rice blast disease prediction and prevention. Historical seven metrological data are used to make prediction of blast disease, two days before its actual occurrence. According to the literature survey conducted in this study, we have made an observation that rice blast disease would outbreak when Minimum Temperature is between 20-26□C and Maximum Relative Humidity is ≥90%, hence region specific models are developed for four regions of Davangere district: Chanagiri, Davangere, Harihara, Honnalli. We have adopted curve shift method and two more user defined functions namely temporalize and scale in LSTM model. Performance of the proposed models are evaluated considering classification metrics such as accuracy, precision and recall. In the study conducted, dropout rate is varied from 0.1 to 0.9 for LSTM model and number of hidden layer are added from 1 to 4 in MLP model. For all the regions, both LSTM and MLP model predictions are accurate, and compared to LSTM model, performance of MLP model accuracy is high. These models will be very helpful for rice cultivator and researchers than using regular blast disease prediction model

Keywords:
Rice blast disease, Climate Data, Long Short Term Memory, Multi-Layer Perceptron, Dropout rate, FeedForward Neural Network, Recurrent Neural Network.

1.Introduction

Agriculture is the primary source of income for the major population of India. Agriculture generates 17% of the total GDP of India and India is the second-largest producer of rice and wheat. Rice (Oryza sativa) is a major food crop for many parts of India. India has the largest area under rice cultivation, hence rice is the important crop of the country. Rice is such a major cereal crop, which provides 20% of total energy and leads as the main food for more than 50% of the world's population[1].

Rice production has been challenged by recent changes in crop production technologies, that also has impact on disease occurrence. Thus crop management includes extensive use of fertilization, repeated flooding increases the disease problem, increased monoculture of rice helps in support of pathogens from one crop to another crop[2].The Rice crop in India has affected by many pathogens. Among 36 rice diseases, rice blast is the disease caused by Magnaporthe Oryae, is the major destructive disease of paddy crop. This disease having significant threat to the production of paddy crops all over the country. Rice blast continues to be a cryptic problem in several rice-growing regions (tropical and temporal) where the pathogen spreads exponentially and is difficult to manage by the farmers and thus reduces yield of paddy crop in the field.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

Rice blast can damage any aerial organ of a plant and plants get the highest disease at maximum tillering stage[4]. Rice blast reduces the photosynthetic area of the plant and leads to the death of the plant. Rice blast disease may lead to losses of up to 90% depending on the part of the plant affected. The loss in yield due to diseases varies depending on the season, weather conditions, and variety of cultivation.

In India, rice blast is a major concern due to favorable weather conditions during the crop season. Climate plays major role in the disease appearance, multiplication, and spread of the fungus. Along with climatic factors, the varieties of seeds also influence the occurrence of rice blasts, primarily the climate factors have a strong influence on the occurrence of blast disease even though a sufficient amount of nutrients are present in the plant. Thus, rice blast disease will occur and develop when certain weather conditions continue for the given period. Forecasting models that make predictions of possible blast disease occurrence may give important information to the producers of rice to manage the disease.

2. Literature Survey

So far in existing studies, diverse forecasting models have been implemented for rice blast disease. For example G. Miah, M. Y. Rafii et. al. have revied fundamentals of rice blast diseases and methods of controlling blast disease. Tomio Yamaguchi conducted test for about 10 days before the average date of leaf blast occurrence till the heading time in the region. And the test is repeated every 7 days and reported that the first occurrence of blast disease is noticed at the average temperature of 19-20℃ and the forecasting can be made by the variation of minimum temperature. Y. Padmanabhan observed incidence of rice blast disease each year on the genetic stocks, susceptibility trials, trials to control blast in years on manurial and fertilizer trials. Based on the observation, that each year is classified as very favorable or unfavorable, moderately favorable concerning the seedling infection(July-August), leaf infection (September- October), and neck infection(November). During these years metrological factors such as maximum and minimum temperature, rainfall, and humidity are obtained from the central rice research institute. Observed that minimum temperature has an association with the development of seed infection. Blast epidemic breaks out within 24-26 ℃ in the year 1957 out of 62 days, 50 days are characterized as blast epidemic in the seedbed. Year 1960 and 1961 was unfavorable to blast epidemic and for leaf infection, the maximum temperature did not show any relation with the blast in both the seasons and high infection having more days under 24℃. Neck infection had a clear association with the relative humidity, rainfall, and the number of rainy days.Chang Kyu Kim experimented and collected field data of rice blast to analyze the epidemiology and has developed a rice leaf blast simulation model called EPIBLAST. For the forecasting model, temperature, relative humidity, rainfall, dew period, and wind velocity are the metrological input factors. The developed model is field-tested during the year 1991. The model predicted rice leaf blast disease early in the month of July. S. B. Calvero et. al. have used the regression equation as a practical model to forecast rice blast disease at Icheon South Korea on two varieties IR50 and C22. The windowpane program identified weather factors that are highly correlated with rice blast disease. The program also identified consecutive days relative humidity ≥80%, number of rainy days ≥80% and precipitation are important variables for the prediction of rice blast disease at Icheon. Manibhushanrao K. Krishnan P. developed a computerized forecasting model called EPIBLA to simulate the incidence and growth of rice leaf blast disease in the field. To predict atmospheric spores and disease progress regression analysis was used on seed variety IR50 and IR20. Regression coefficients suggest that temperature and relative humidity influence

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

significantly spore generation and relative humidity (73-100%), temperature (14-25℃), and amount of dew are highly significant after disease incidence. Kowang-Hyung-Kim et. al has taken the first step to establish a seasonal disease by using a seasonal early forecasts dataset, and developed an EPIRICE model for rice blast disease. The model was used to predict the disease risk by using AUDPC model and implied that model output could be obtained only when the climate data for the whole season is available. Omkar Singh, Jagadeesh Bathula et. al. discussed climate factors that are known to influence sporulation and spore dissemination of rice blast disease. Temperature range between 19-29℃ particularly 23-26℃ and humidity above or equal to 90% are considered as highly favorable of rice blast disease development. Forecasting results could help to identify which year are conducive and fungicides or cost-effective or risky under those conditions. Several computerized forecasting systems such as EPIBLA, EPIBLAST, BLAST, BLASTM have been developed to simulate the incidence and progress of the rice blast disease in the field. They have discussed factors of the epidemic and their components should be known before the forecasting model was developed. Host factors include varieties of seeds, different stages of rice cultivation, density, and distribution of host in different regions. Laxman Singh Rajput et. al have studied the effects of temperature on different varieties. Temperature effects significantly in the growth of plant, maximum growth are observed at temperature 27℃ compared to 32 and 22℃. Maximum lesion size was observed at 27 ℃ on variety PRR 78. Dimitrios Katsantonis et. al. conducted review of literature of the rice blast disease forecasting models and revealed that 52 studies have predicted rice leaf blast disease. Input variables that are considered are air temperature, relative humidity, rainfall. Along with these, critical factors are also considered such as leaf wetness, nitrogen fertilization. This review revealed a low rate model due to inaccuracies and uncertainties in the predictions. The review is also useful for the modelers, users, and stakeholders to assist in developing and selecting the most suitable models for efficient rice blast forecasting.

Added to this Gianni Fenu et al. have proposed an artificial intelligence-based approach to predict potato late blight disease in the Sardinia region and a novel technique is imposed for potato disease risk. Predictions are made on historical data such as temperature, humidity, rainfall, speed of the wind, and solar radiation collected from various locations over four years (2016-2019). This aimed to identify the usefulness of the SVM classifier to determine the relationship between weather and disease for potato crops. The results obtained show that temperature, humidity, and speed of wind plays important role in forecasting.

Further Sarinya Kirtphaiboon et. al developed a mathematical model for studying the dynamics and severity of rice blast disease. The severity of the disease is simulated under environmental conditions such as temperature, humidity, dryness, in different regions of Thailand. The developed model investigates the pathogen's life cycle together with disease development in the field of paddy. Temperature is the major climatic factor that affects rice plant. In addition to this, study also investigates severity is high when the temperature is low and humidity is high.

The main objective of this study is to develop LSTM and MLP models that can predict the rice blast disease in advance based on historical climate data. Environment and soil differ among regions, thus region-specific models were created. Agriculture and Horticulture Research Station (AHRS) Kathalgere, provided data on the occurrence of rice blast disease, according to the data provided by the research station, blast disease would outbreak seven days after satisfying each of the two conditions for seven consecutive days. First one is Relative Humidity ≥90% , second one is Minimum Temperature between 20-26℃. If these two conditions do not coincide it is impossible to forecast the occurrence of rice blast disease.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

3. Materials And Methods

For the LSTM and MLP models, the data for occurrence of rice leaf blast disease in the Davangere region in Karnataka state was obtained from the Agriculture and Horticulture Science Kathalgere Research Station of Davangere District. The data on the occurrence of blast disease for the years 2013-2015 was based on climate data and in literature study it is been observed that rice blast disease would outbreak when Minimum Temperature is between 21-26℃ and Maximum Relative Humidity ≥ 90%.
The climate data of Davangere District for the timestamp from 2015-2019 was collected from the Karnataka state natural disaster monitoring center. Thus, weather variables that are collected were Minimum Temperature, Maximum Temperature, Minimum Relative Humidity, Maximum Relative Humidity and Rainfall, Temperature, Difference and Relative Humidity Difference. Data provided at taluk level of Davangere District is analyzed and model developed for the four regions of Davangere District such as Harihara, Honnalli, Davangere, Channagiri. These regions are considered as active rice farming regions hence LSTM and MLP models are developed for these regions. Fifth region Jagaluru is not a rice farming region hence models are not developed for this region.

3.1 Correlation Coefficient

The Correlation Coefficient is useful in data analysis and formulating a model for better understanding of the relationship between variables. The statistical association between variables is referred to as correlation. If two variables move in the same direction then it is called a positive correlation. Negative correlation is defined if one variable increases and the other variable decreases. Correlation can also be zero when two variables are not related to each other.
The Pearson's Correlation Coefficient is also known as the Pearson Product-Moment Correlation Coefficient. It is a measure of linear association between two variables X and Y. In the proposed work to analyze strength of a linear coefficient of seven different climate input varaibles with target variable, we have used pearson's correlation coefficient. Correlation of seven different input varaibles with target variable is described in Table 3 in section Results and Discussion.

3.2.Development of Models for Prediction of Epidemics

We have developed models for forecasting the occurrence of blast disease in advance using two neural network models, namely Long Short Term Memory(LSTM) and Multi-Layer Perceptron(MLP). Both LSTM and MLP model predictions are accurate but, Multi-Layer Perceptron(MLP) had performed better than Long Short Term Mermory(LSTM) for the data that we have analyzed.

3.2.1 Multi-Layer Perceptron

A neural network is a combination of neurons. Each one performs using simple and standard functions. In the case of feedforward neural networks, neurons are represented into a graph without any cycles, this is referred to as sequential computation. The simplest kind of feed-forward neural network is a Multilayer Perceptron (MLP), the neurons are arranged into set of layers and each layer has some similar neurons. Every neuron in one layer is connected to every neuron of the next layer and this network is called fully connected network. The first layer is referred to as the input layer, each neuron in this layer will take input features. The last layer is

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

called the output layer and it has one neuron for each value (one neuron in case of binary classification or regression, k neurons in case of k class classification). And the layers between these two layers are called hidden layers. The neurons in these layers are called input neurons, hidden neurons, and output neurons. The total number of layers defines the depth and the number of neurons in each layer defines width. The perceptron is a machine learning algorithm for solving regression and classification problems. The structure of Multilayer Perceptron is depicted in fig. 1

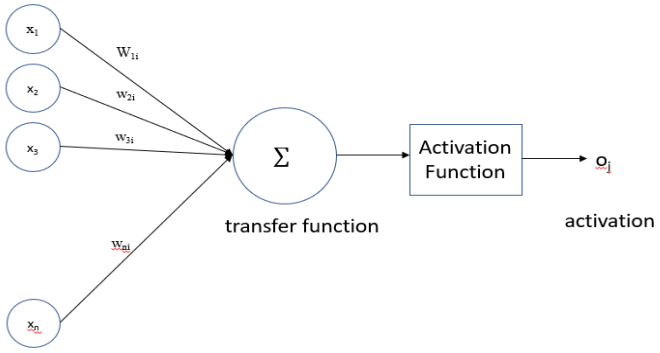

Fig. 1: Architecture of Multilayer Perceptron

In the proposed work, the Multi-Layer Perceptron training is done with backpropagation which is capable of demonstrating a variety of nonlinear decisions. It exercises gradient descent which attempt to minimize errors in the network, hence error correction is adapted in the Artificial Neural Network. The methodology used in the proposed work consists of updating weights by incorporating all the patterns of the input file and assemble all the weights. There is also need for stop criteria and the most widely used is cross-validation and this method is more efficient in stopping the training when the best generalization is achieved. It consists of separating tiny part of the training data and using it to evaluate the training module. Multi-Layer Perceptron model used in this study has adopted the activation functions in the hidden layers which is reLU function (rectified Linear Unit). There are many numbers of activation functions available which include tanh, sigmoid, softmax, hyperbolic tangent, reLU, etc since tanh and sigmoid suffers from vanishing gradient problem while reLU activation function overcome from above said drawback and provides quick convergence, hence based on the literature review reLU activation function is used.Another activation function called Sigmoid activation function is used in the output layer, since the objective of the work is to predict rice blast disease which is generally called a binary classifier.The optimizer adopted in the study is adam (adaptive moment estimation) and the optimizer which is a stochastic gradient method based on first hand and second hand moments. On the other hand, batch size used in the MLP neural network was equal to 20 and the number of epochs was 100.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

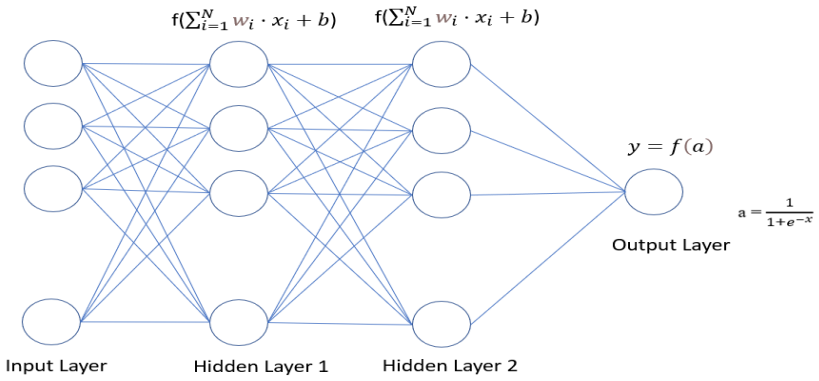

Fig. 2: Proposed Architecture of Multi-Layer P91erceptron

Input to the Hidden Layers is mathematically represented as:

$$z_{ik} = x_i * w_i + b_k \quad (1)$$

A neuron in the hidden layer can be called as summation hence, it is represented as $\sum$ which computes sum of inputs and weights of each input and a bias value followed by an activation function. In the proposed work, activation function in the hidden layer used was reLU and can be mathematically represented as:

$$z_i = b_i + \sum_{i=1}^{n} w_i + x_i \quad (2)$$

$$a=\sigma(z_i) \quad (3)$$

$$a = \max(0,z_i) \quad (4)$$

$$y(x) = \frac{1}{1+e^{-x}} \quad (5)$$

where x_i is the input variable, w_i is the weight assigned to each neuron in the hidden layer, and this layer computes the sum of inputs and weights, hence hidden layer is represented as $\sum$. Further, b_i is the bias value of each neuron. reLU is the activation function used in the h hidden layer, hence the sum of weights and input value followed by this activation function. The output layer is represented as y(x), sigmoid is the activation function that is used, output value greater than or equal to x returns 1 else 0.

3.2.2 Long Short-Term Memory

In the feed-forward neural networks, all cases are independent and while fitting a model for the current day, observations of previous days are not considered. This kind of dependency on time is achieved by the network called Recurrent Neural Networks (RNN). Recurrent Neural Networks work well with short-term dependencies that is RNN remembers things for a small duration of time. When a large number of data is fed to the model, information might lose somewhere, we refer to this problem as the vanishing Gradient problem. This drawback we can overcome using an advanced version of RNN called Long Short-Term Memory (LSTM).

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

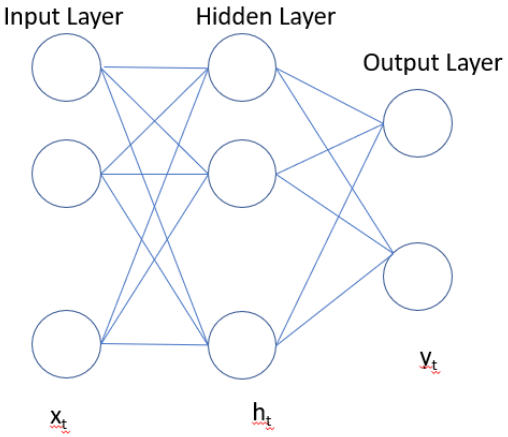

Fig. 3: Structure of Feed Forward Neural Network

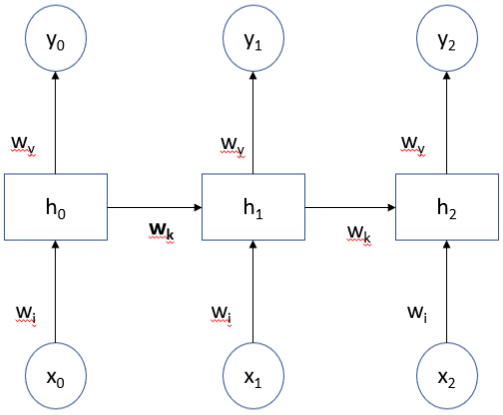

Fig. 4: Strucutre of RNN

Above fig. 3 depicts Feed Forward Neural Network where x_t is the input layer, h_t is the hidden layer and y_t is the output layer. This type of neural network is mathematically represented as follows:

$h_t = f1((w*x_t) + b_1$ (6)

$y_t = f2((w*x_t) + b_2$ (7)

Structure of Recurrent Neural Network is represented in the above figure 4, this neural network has three layers namely input layer, hidden layer and output layer. To solve hidden layer and output layer mathematically, we have equations (8) and (11)

$h^{(t)} = g_h(w_i x^{(t)} + w_r h^{(t-1)} + b_h)$ (8)

in solving hidden layer equation(3) at time t = 0

$h^{(0)} = g_h(w_0 x^{(0)} + w_r h^{(0-1)} + b_0)$ (9)

h(0-1) = h(-1) time cannot be -1 hence it is 0

$h^{(0)} = g_h(w_0 x^{(0)} + b_0)$ (10)

$y^{(t)} = g_y(w_y h^{(t)} + b_y)$ (11)

Recurrent Neural Networks are trained using backpropagation neural networks over time. This is the same mechanism as standard backpropagation. Except that it is backpropagated over time rather than through its layers. RNN is a deep neural network through time but it has two limitations vanishing or exploding gradients. And also RNN is a short-term memory, it has shortcomings over long-term dependencies. We can analyze the vanishing gradient problem in recurrent neural networks by the following equations (12), (13), and (14). RNN is trained using backpropagation and it is used to update new weights with the old weights. Error in the output is calculated as the square of actual output minus model output this is written as in equation (12). Change in weights is equal to change in error over the weight change and multiplied with the learning rate and weight change is added to the old weight that gives new weight and is represented by equations (13) and (14). If change in error is too small over the weight change that is almost less than 1, which is too small and this will be almost equal to old weights. Under this condition old weights are not replaced with new weights.

Error = (Actual output – Model Output)^2 (12)

$\Delta w = n \frac{de}{dw}$ (13)

$w = w + \Delta w$ (14)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

To overcome these limitations of RNN, LSTM has been developed to hold short-term memory for long-term dependencies. Long Short-Term Memory is comprised of memory blocks similar to neurons in RNN. This memory block is the combination of cells and gates and plays a vital role in training long-term dependencies. A common LSTM unit has cell state, input gate, output gate, and forget gate. The cell state remembers information over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell state. The memory blocks are held responsible for remembering information and doing manipulations and this is done through three gates.

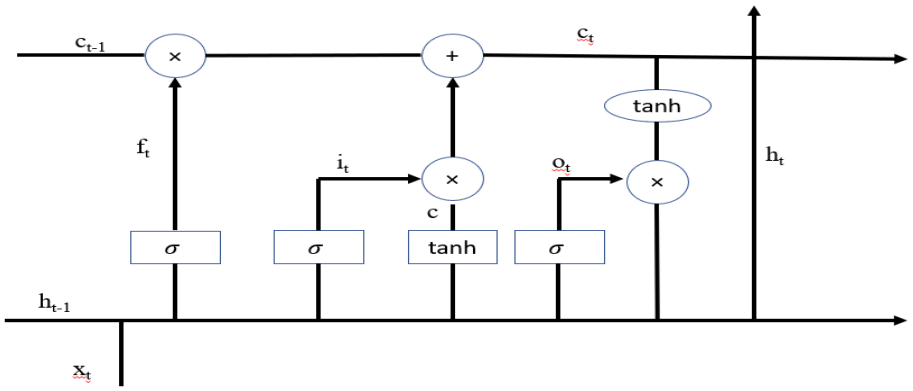

Fig. 5: Structure of LSTM Networks

A forget gate is responsible for removing information from the cell. The data which is no longer required for LSTM or data which is least important is removed through the filter. This gate takes two inputs h_{t-1} and x_t. x_t is input at a particular timestamp and h_{t-1} is the output of the previous cell state or information of the previous hidden state. These two inputs are multiplied with weight matrices and bias vectors are added. The sigmoid function is applied and the output of this function is 0 or 1. 0 means that forget gate removes particular information in the current cell state and 1 means forget gate holds the particular information in the current cell state. The forget gate in a cell can be represented as in equation (15)

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (15)$$

Input gate is responsible for adding information to the cell state. This regulates what values are needed to be stored in the cell state and this involves sigmoid function and which is similar to the forget gate layer. Later creates a vector containing all possible information that can be stored in the cell state and this is done by tanh function which brings out values from 1 to -1. Multiplying the values from the sigmoid layer to the vector created using tanh function then storing this information in the cell state by performing an addition operation.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (16)$$

$$c = \tanh(w_c[h_{t-1}, x_t] + b_c) \quad (17)$$

$$ct = f_t * c_{t-1} + i_t * c_t \quad (18)$$

Output gate creates Vector by applying tanh function to the cell state thereby scaling values between the range -1 to +1. Employing sigmoid function to filter using the values h_{t-1} and x_t. Multiplying the values from the sigmoid layer and the vector and sending this as output and also to the hidden state of the next cell.

$$O_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (19)$$

$$h_t = o_t * \tanh(c_t) \quad (20)$$

In the above equations 15-19, w_i, w_c,w_f,w_o are weight matrices and b_f,b_o,b_i,b_c indicates bias.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

Further σ sigmoid function and * symbol indicate element-wise multiplication. Equations (15),(16),(19), are the formula to calculate f_t, i_t, o_t at time t. These three gates take x_t and h_{t-1} as inputs which are then multiplied with the weight matrices. And results are added to the bias values and the sigmoid function is applied to the term to take the result. If the result is near zero then the gate is fully closed and the gate does not accept information if the result is one then the gate is fully entered. Hence these three gates are important components of the LSTM memory block. C_t is computed as shown in equation(18) where the cell state of previous block c_{t-1} is multiplied by f_t and new input information x_t by i_t. So f_t decides the amount of information on c_{t-1} and it decides how much newly added information is stored in cell state c_t. Finally, calculation of h_t is done as shown in equation (20) where o_t is multiplied by taking activation function tanh in c_t at time t. The calculated c_t and h_t are passed to the next time calculation. In the h_t calculation of the LSTM, both c_{t-1} and h_{t-1} are passed.

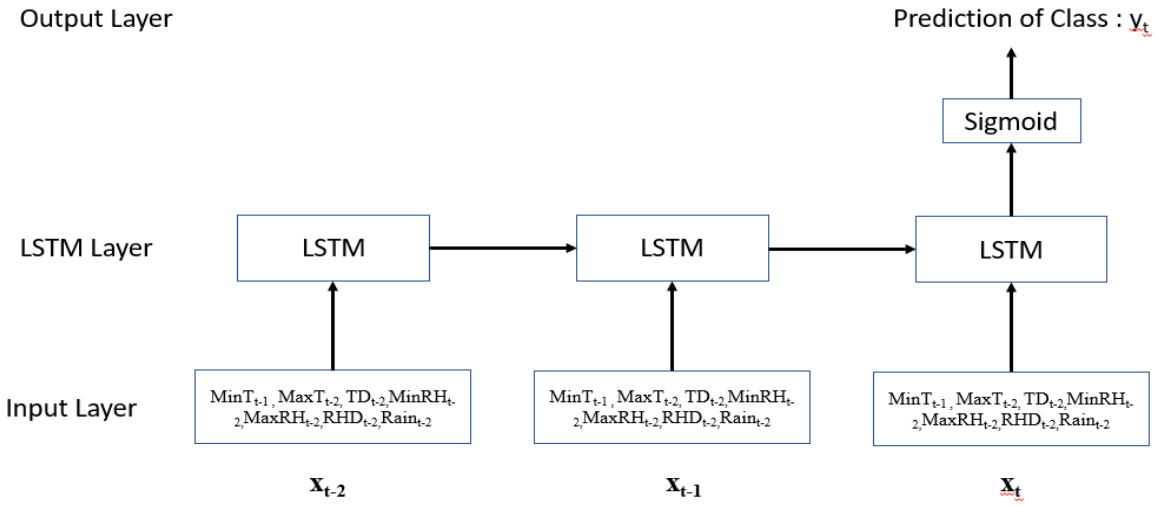

Fig. 6 : LSTM Model for Prediction of Blast Disease Occurrence

We have developed another rice blast prediction model using LSTM network structure with the 3 hidden layer and timesteps of 5. In the fig. 6 x_{t-2},x_{t-1},x_t are Minimum Temperature(MinTt), Maximum Temperature(MaxTt), Temperature Difference(TDt), Minimum Humidity(MinHt), Maximum Humidity(MaxHt), Humidity Difference(HDt), Rainfall(Rt) for each of the years from 2013-2019. Given input value passes through the LSTM layer according to the equations from (16) – (20). The value h_t at last layer called the output layer in the time t predicts the occurrence of rice blast disease prior to the 4 days. This prediction y_t goes through the layer called as sigmoid layer . The output classes are numbered as 0-1. y_t is calculated as follows

$\backslash z_t = w_z h_t + b_t$ (21)

$\backslash y_t = \text{sigmoid}(z_t)$ (22)

$\text{Sigmoid}(z_{t,i}) = \frac{1}{1+e^{-z}}$ (23)

W_z is the weight matrix, b_t is called the bias value, Z_t is the three-dimensional vector from which we have obtained the sigmoid equation for binary classification as in equation (23). In sigmoid equation, $Z_{t,i}$ is the i^{th} unit value called logit. y_t is the probability of class.

To study the performance of the LSTM model, Binary Cross Entropy consider as a loss function:

$BCE = - \sum_{i=1}^{c=2} t_i \log(s_i)$ (24)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

With the above loss function, we have considered two classes C_1 and C_2 . s_1, t_1[0,1] are the score and groundtruth for C_1 . s_2 = 1-s_1, t_2=1-t_1 are score and groundtruth of s_2. We trained the LSTM model using batch size of 20, epochs of 100, and varying dropout layer from 0.1 to 0.8 and Adam Optimizer were implemented using TensorFlow 1.3.0.

3.2.3. Methodology of The Proposed Work

To develop predictive models for rice blast disease, experiments are conducted to predict blast disease for the year 2019, based on the climate data of previous four years from 2015 to 2018. As described in section 2, minimum temperature ranging from 20-26℃ and Relative Humidity ≥ 90% are considered as class 1 which means favorable factors for blast disease to develop, otherwise class 0 which means ranges of different climate variables are not favorable for disease development. The analytical procedures were implemented as shown in fig 7:

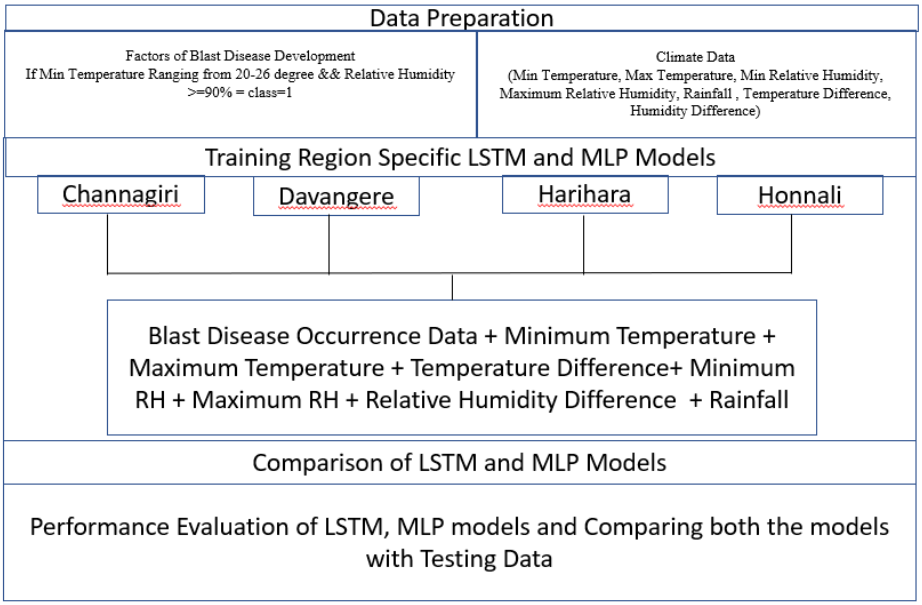

Fig. 7: Flowchart of Procedures followed in the Study

The experiment procedure is briefly described here. First, data were generated for the training of LSTM and MLP models. With this data, region-specific LSTM and MLP models for four regions, Channagiri, Honnali, Harihara, Davangere were developed and these regions are active rice farming regions and fifth region called Jagaluru region is not part of rice farming, hence models are not implemented for this region. Experiments were performed for seven input variables and considered input variables were Minimum Temperature, Maximum Temperature, Temperature Difference, Minimum Relative Humidity, Maximum Relative Humidity, Humidity Difference and Rainfall. These attributes were important for rice blast disease prediction. After the region-specific LSTM and MLP training phase, the model is tested by the region-specific test values hence this phase is referred to as testing phase. Here trained region-specific models were used to predict rice blast disease occurrence of next year. The results of both region-specific models are analyzed and results of LSTM and MLP models are compared to find the best model in making predictions of rice blast disease.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

3.2.4. Data Preparation for LSTM and MLP Models

LSTM and MLP models are data-driven approaches that learn attributes that are important in predicting rice blast disease, so the amount of data of input variables has significant effect on the performance of an algorithm. Daily raw climate data acquired from Karnataka state natural disaster monitoring center for five taluks of Davangere District in the period between 2015-2019 includes data of Channagiri, Davangere, Harihara, Honnalli, and Jagaluru taluks of Davangere District – regions that are targeted in the study. Daily data of five consecutive years are collected for each of the five regions. In this study initially, four years of climate data of five regions are used as input variables for the development of LSTM and MLP models, and fifth-year data are used for the testing phase.

Table 1: Sample Dataset used to conduct experiment

Date	District	Taluk Name	Minimum Temperature	Maximum Temperature	Minimum Humidity	Maximum Humidity	Humidity Difference	Rainfall
01-01-2015	Davangere	Channagiri	20.63	32.41	11.78	37.75	91.43	53.68
02-01-2015	Davangere	Channagiri	16.51	31.26	14.75	38.33	97.31	58.98
03-01-2015	Davangere	Channagiri	16.85	32.08	15.23	26.21	95.06	68.85
04-01-2015	Davangere	Channagiri	15.86	31.26	15.4	39.18	95.48	56.3
05-01-2015	Davangere	Channagiri	17.25	35.53	15.28	28.08	89.43	61.35

Using this approach, if any data is missing or found null for a particular region or any other 4 regions for the five years, that dataset is filled with mean values from the experimental dataset. Hence the dataset containing total of 1826 elements obtained for each region. The 80% of the dataset is used for training, 10% is used for validation and 10% of the dataset was used as test data.

The target values of the MLP and LSTM models were classified as classes 0-1 corresponding to climate factors influencing the rice blast disease development. Class is 0 when climate factors such as Minimum Temperature and Maximum Relative Humidity values are unfavorable for disease development and class is 1 when climate variables such as Minimum Temperature ranging from 20-26℃ and Maximum Relative Humidity ≥ 90%.

Various input variables considered in this model are Minimum Temperature, Maximum Temperature, Temperature Difference, Minimum Relative Humidity, Maximum Relative Humidity, Humidity Difference, Rainfall in a day. Hence data standardization is needed, climate input variables incorporated in the model are floating values and these values are rescaled or input variables are standardized using the standard scaler method and expressed as

$$y=(x-\text{mean})/\text{standard\_deviation} \quad (25)$$

$$\text{mean}= \text{sum}(x)/\text{count}(x) \quad (26)$$

$$\text{standard\_deviation}=\text{sqrt}(\text{sum}(x-\text{mean})^2/\text{count}(x)) \quad (27)$$

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

where x is the value of the climate input variable or original input value, mean is the mean of the input variable, count(x) is the total number of values in the input variable x, y is rescaled or standardized input variable x.

Table 2: Parameters and their values initialized in the Developed Models

Sl.No.	Parameter	MLP(value)	LSTM(value)
1	Input Variables	7 (two dimensional)	7 (three Dimensional)
2	Batch Size	20	20
3	Epochs	100	100
4	Lr	0.0001	0.0001
5	Hidden Layer Activation Function	ReLU	ReLU
6	Number of Hidden Layers	1,2,3,4	2
7	Dropout	0.5	0.1,0.2,0.3,0.4,0.5 0.6,0.7,0.8,0.9
8	Output layer Activation function	Sigmoid	Sigmoid
9	Optimizer	Adam	Adam
10	Loss	Binary Cross Entropy	Binary Cross Entropy

4 Results And Discussion

In order to conduct experiment with climate variables as described in section 2, the daily Minimum Temperature, Maximum Temperature, Temperature Difference, Maximum Humidity, Minimum Humidity, Humidity Difference, Rainfall of the months January to December in the period of 2015 to 2019 were obtained for the target regions(Channagiri, Davangere, Honnalli, Harihara) and this was the same period in which references for the occurrence of blast disease were obtained. There are 1826 datasets of the daily Minimum Temperature, Maximum Temperature, Temperature Difference, Maximum Relative Humidity, Minimum Relative Humidity, Relative Humidity Difference, Rainfall were obtained for each region of all the 5 years. Above said input variables are graphically shown in fig. 8(a) to 8(e).

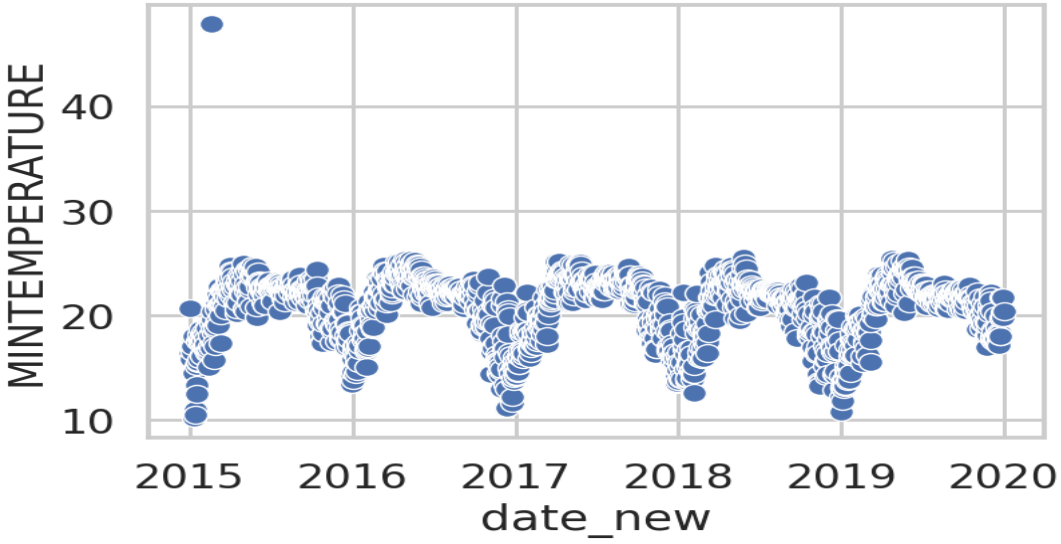

Fig. 8(a): Daily Minimum Temperature of Channagiri Region from the year 2015 to 2020

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

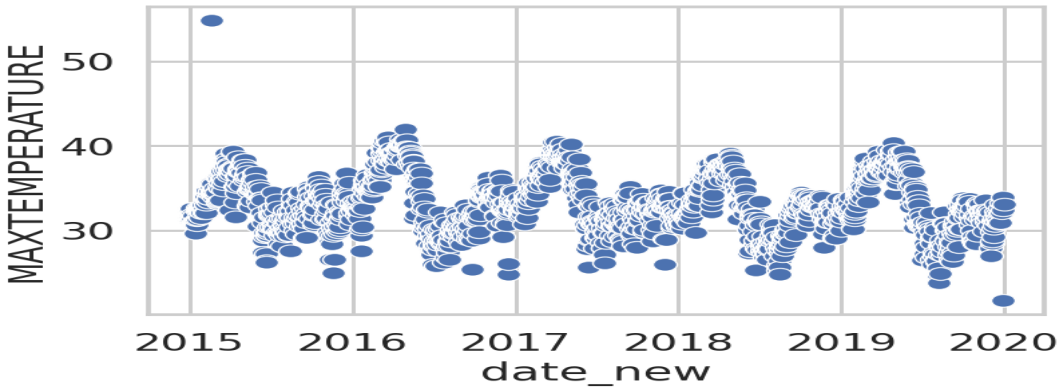

Fig.8(b): Daily Maximum Temperature of Channagiri Region from the year 2015 to 2020

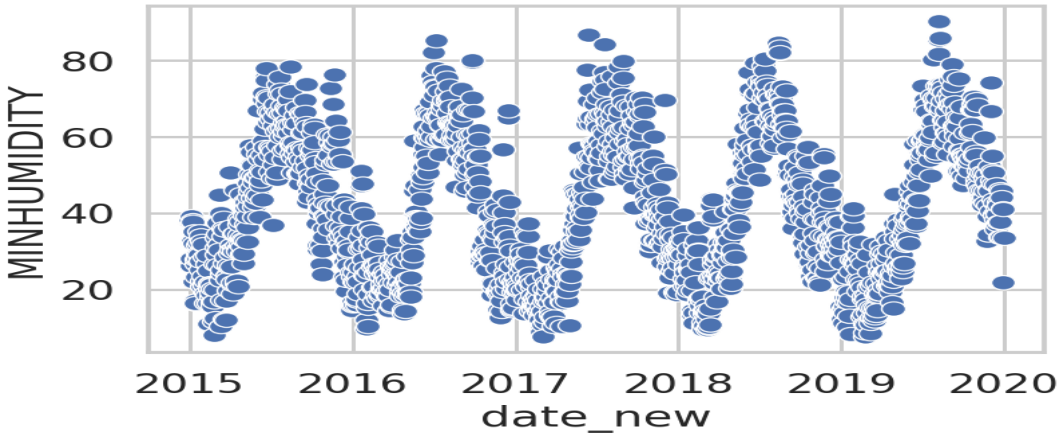

Fig. 8(c): Daily Minimum Relative Humidity of Channagiri Region from the year 2015 to 2020

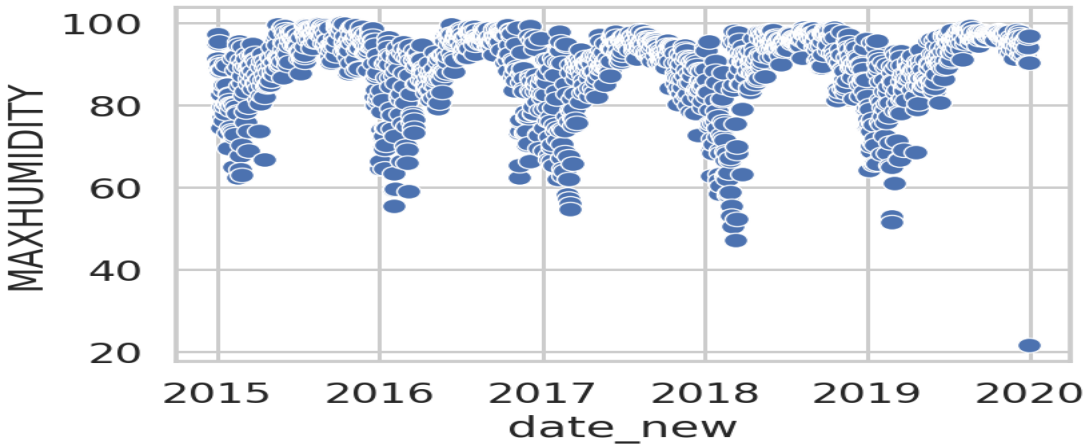

Fig. 8(d): Daily Maximum Relative Humidity of Channagiri Region from the year 2015 to 2020

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

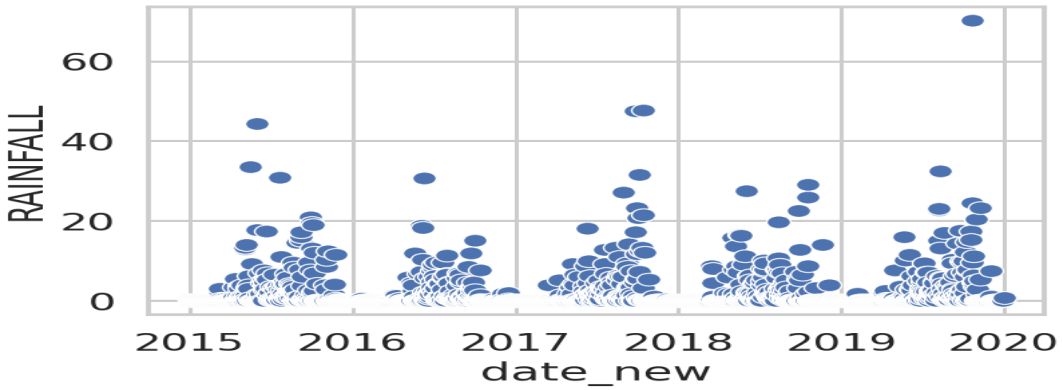

Fig. 8(e): Daily Rainfall of Channagiri Region from the year 2015 to 2020

As shown in fig. 8, Minimum Temperature from January to December for the year 2015-2019 in Chanagiri Taluk which is located in southeren transistion zone (7) of Karnataka state is ranging 10-27℃. It has been confirmed and validated that the Minimum Temperature for the occurrence of Rice blast disease are within range 20-26℃. Rice blast disease is also caused by the high humidity occurring post monsoon season in Karnataka, that is end of the June or mid of the July. Average Maximum Humidity in Chanagiri region is ranging from 45-99 % , it is confirmed that for the occurrence of rice blast disease, Maximum Humidity are within range of 90 and above in line with Minimum Temperature ranging from 20-26℃. Correlation between seven input variables of each region is analyzed by applying pearson's correlation equation as described in section 2.1. Table 2 shows the correlation between seven input variables and target variable, it is observed that in all regions Minimum Temperature, Minimum Humidity and Maximum Humidity are positively correlated with the target variable, Temperature Difference and Humidity Difference are negatively correlated with the target variable, hence these five input variables are considered as observable variables.

Table 3: Correlation coefficient between input variables and target variable

Input Variables	Correlation with Target Variable
Minimum Temperature	0.52
Maximum Temperature	-0.38
Temperature Difference	-0.69
Minimum Humidity	0.71
Maximum Humidity	0.65
Humidity Difference	-0.51
Rainfall	0.30

The main importance of this study is to construct machine learning based models that can make prediction of the occurrence of blast disease on historical climate data. Climate and soil values differes from region to region, thus region specific models were created. As discussed in section 2, Chanagiri, Davangere, Harihara, Honalli are active rice growing regions hence machine learning models are developed for these active rice farming regions.

The meteorogical data are used as input variables for LSTM and MLP models and prediction of blast disease occurrence through these models are the target variable. In this study, blast disease occurrence is predicted 2 days prior to the incidence based on climate data, this is adopted through curve shift method, we moved positive label of row n to n-1 and n-2 timestmaps and

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

dropped row n. Also there is a difference of one day between disease row and the next row. This is because when disease occurs it remains in the same status until action is taken by the farmer. In the study, consecutive disease rows are deleted to prevent the classifier from learning to predict disease after it has already occurred.

LSTM is important and more demanding model in machine learning for the analysis of time series data. Much time and attention have been given to the data that has fitted to the LSTM model. In the study LSTM model is fitted with 3 dimensional array, the shape of the array is represented as Array = samples * lookback * features, here samples are the number of observation in the data, lookback is at time t, LSTM model will look (t-lookback) to make current prediction and features are the number of input varaibles in the data. In our study we have considered Array = 1826 * 20 * 7 which is three dimensional array for the LSTM model. Initially data is two dimensional array of sample * features and transform 2D to 3D array we have developed temporalize function which converts 2D dimensional of sample * features to 3D samples * lookback * features. Data is normalized, an X matrices are 3 dimensional and standardization happen to the 2 dimensional. To perform this, we have user defined function called flatten, this will recreate 2 dimensional array. This function is inverse of temporalize and another user defined function scale is used to scale 3 dimensional array as inputs to the LSTM.

To ensure the better performances of the LSTM and MLP models, major hyper-parameter in LSTM and MLP models were tested such as the dropout rate for LSTM and number of Hidden Layers for MLP, this dropout rate is a technique which ignores certain rate that avoids overfitting of model. Fig. 9(a) is a LSTM model validation loss for Honnalli region,fig. 9(b) is a LSTM model validation loss for Harihara region ,fig. 9(c) is a LSTM model validation loss for channagiri region,fig 9(d) is a LSTM model validation loss for Davangere region, change in validation loss for the varying dropout rate hyper-parameter are analyzed for all the 4 regions.

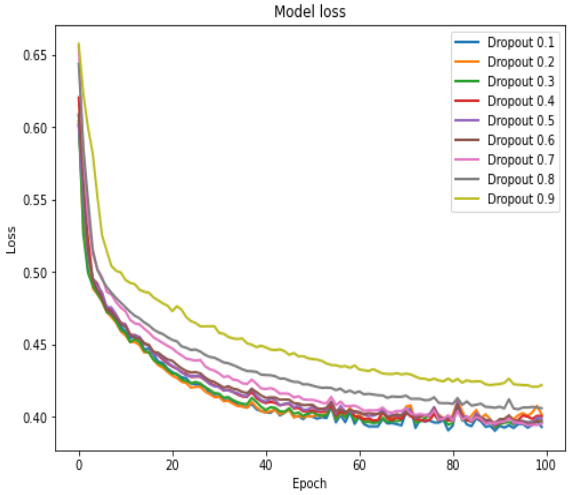

Fig. 9(a)

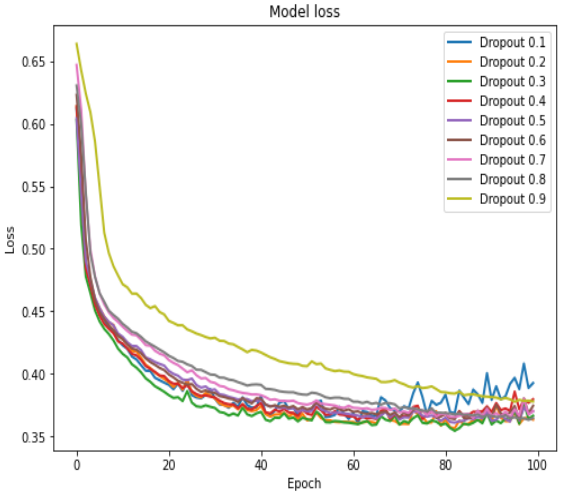

Fig. 9(b)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

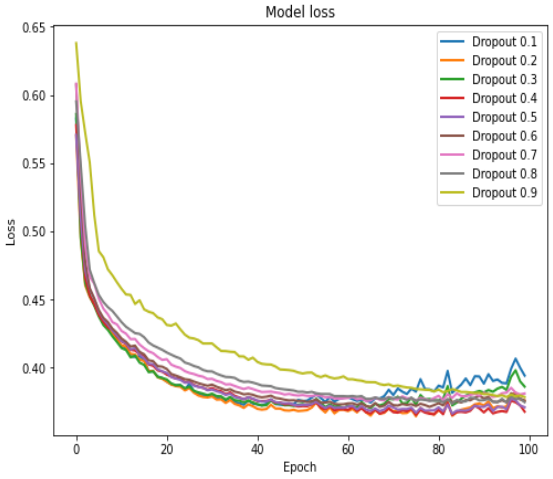

Fig. 9(c)

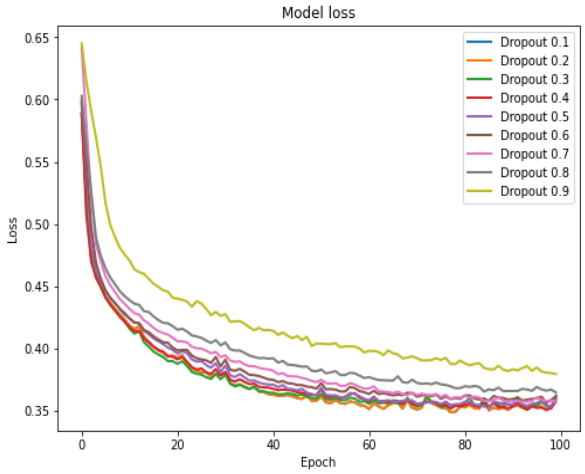

Fig. 9(d)

Table 4: Best Dropout rate for four region specific model.

Best Dropout rate for region specific LSTM model				
Regions	**Channagiri**	**Davangere**	**Harihara**	**Honnalli**
Droprate	0.4	0.4	0.2	0.1

Multi-Layer Perceptron is also demanding model in machine learning for the analysis of time series data. In the current study MLP model is fitted with 2 dimensional array and the shape of the array is represented as Array = samples * features. For the better optimization of the MLP model one of the hyper-parameter in MLP were tested, which is number of hidden layers between input and output layer. For each of the region specific model, in the proposed study result is demonstrated for each region with single hidden layer model and has achieved better accuracy compared to two hidden layers in LSTM Model.

4.1.Performannce Evaluation and Comparision of LSTM and MLP Model

Performance evaluation is needed to quantify performace of model. The choice of evaluation metrics depends on the type of the problem we study. This section presents the results obtained when applying LSTM and MLP models. Dropout rate hyper-parameter for LSTM and Number of hidden layers as hyper-parameter is analyzed and demonstrated in section 3.

In order to evaluate the performance of LSTM and MLP models with imbalanced distribution of class labels we have provided with following performance evaluation metrics namely accuracy, precision and recall. Base definition of accuracy is the number of correct predictions made out of all predictions that model has made. Precision is the percentage of positive instances out of total positive predicted instances. Recall is the percentage of positive instance out of the total actual positive instances.considering these definitions

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (29)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (30)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (31)$$

The terms TP and TN, FP and FN are called True Positive, True Negative, False Positive and False Negative. Table 3, Fig. 10(a) and Fig 10(b) compares accuracy, precision and recall of two models of all the four regions, according to results obtained for occurrence or non occurrence of

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

rice blast disease based on climate data, both LSTM and MLP models obtains good accuracy, highest percentage of precision and recall for the dataset used in the study, however if we compare LSTM and MLP models directly, MLP obtains highest accuracy, precision and recall.

Table 3: Performance Evaluation of LSTM and MLP models

Performance Evaluation of LSTM and MLP models						
	LSTM			**MLP**		
Regions	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Channagiri	79.7	78.8	75.14	76.50	71.35	80.00
Davangere	63.5	57.89	66.6	87.00	84.75	85.80
Harihara	70.52	71.35	72.4	83.60	84.26	82.40
Honnalli	66.6	62.2	67.87	83.87	82.35	79.70

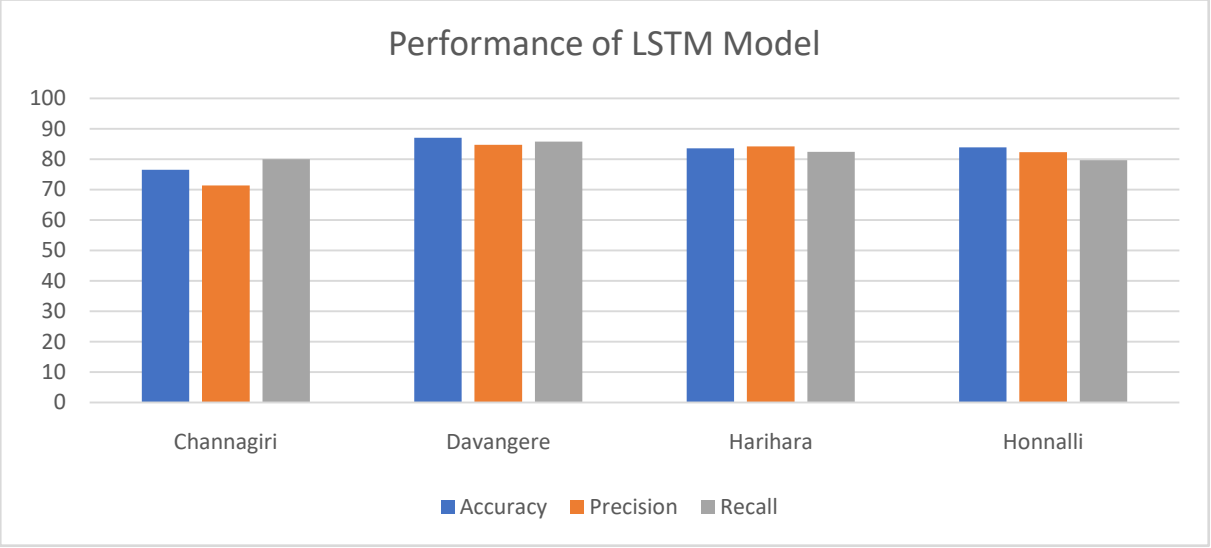

Fig. 10(a): Performance Evaluation metrics of LSTM Model

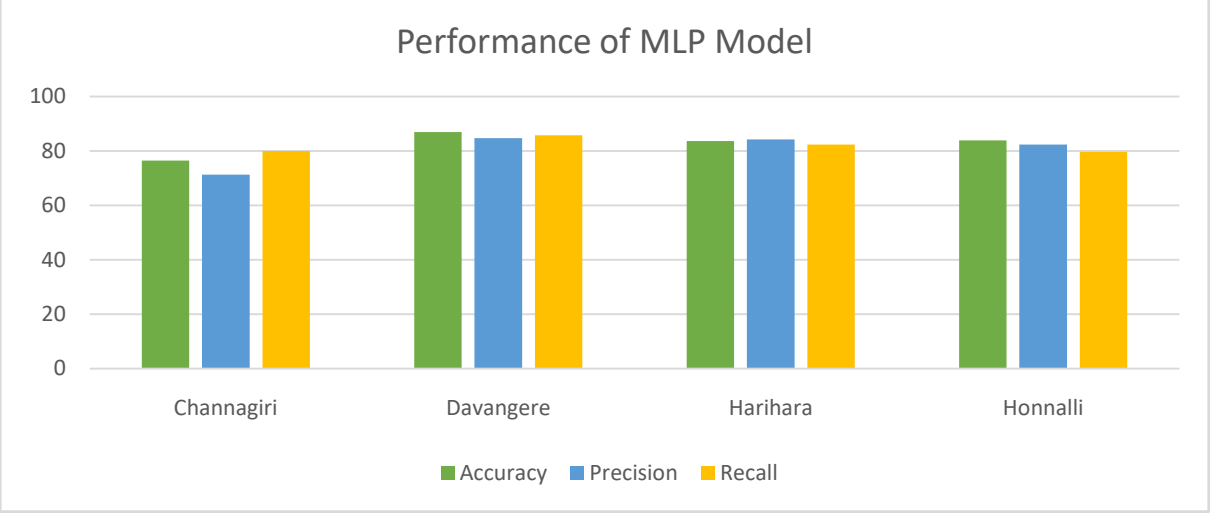

Fig. 10(b): Performance Evaluation Metrics of MLP Model

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 4703 - 4722
Received 25 April 2021; Accepted 08 May 2021.

5.Conclusion:

In this study we developed the LSTM and MLP rice blast disease occurrence prediction models using seven different kinds of climate data as input varaibles. i.e. the Mini Temperature, Max. Temperature, Temperature Difference, Minimum Humidity, Maximum Humidity, Humidity Difference and Rainfall. Blast disease occurrence and non occurrence as output variable for four regions of Davangere District. Hence LSTM and MLP models are referred as region specific models, here blast disease occurrence is predicted 2 days early before disease actually occurs, this is made possible through curve shift method. While developing LSTM model three user defined functions are written namely temporalize, flatten and scale, MLP model data is represented 2 dimensionally. Regularization is adopted in LSTM model, while validation loss of model is analyzed by varying from 0.1 to 0.9 for all the four regions respectively. Similarly number of hidden layers is considered as hyper-parameter in MLP model, number of hidden layers added from 1 to 4, model achieved good accuracy when it has one hidden layer between input layer and output layer. For all the four regions MLP model achieved highest accuracy, precision and recall compared LSTM model. The utility of the proposed LSTM and MLP is expected to be high because models used in the study is capable of transfer learning. Same methodology can be applied to the climate data of other regions, this study was based on the data obtained from Karnataka State Natural Disaster And Monitoring Centre, Karnataka and these developed models can be used for other states of India in which rice is considered and grown as primary crop.

References

[1] Laxman Singh Rajput, Taru Sharma, Puchakayala Madhusudhan and Parimal Sinha. “Effect of Temperature on Growth and Sporulation of Rice Leaf Blast Pathogen Magnaporthe oryzae” Int.J.Curr.Microbiol.App.Sci (2017) 6(3): 394-401.

[2] Chethana B.S., Deepak, C.A., Rajanna, M.P., Ramachandra, C. and Shivakumar, N. “Current Scenario Of Rice Diseases In Karnataka”. I.J.S.N., VOL.7 (2) 2016: 405-412.

[3] Yoshihiro Taguchi , Mohsen Mohamed Elsharkawy , Naglaa Hassan, Mitsuro Hyakumachi. “A novel method for controlling rice blast disease using fan-forced wind on paddy fields.” Crop Protection 63(2014) 68-75.

[4] Rajendra Prasad, Anupam Sharma and Sweta Sehgal. “Influence of weather parameters on occurrence of rice blast in mid hills of Himachal Pradesh.” Himachal Journal of Agricultural Research 41(2): 132-136 (2015).

[5] Tomio Yamaguchi. “Forecasting Techniques of Rice Blast.” JARQ Vol. 6, No. 4, 1970.

[6] G. Miah, M. Y. Rafii , M. R. Ismail, M. Sahebi, F. S. G. Hashemi,O. Yusuff and M. G. Usman. “Blast Disease Intimidation Towards Rice Cultivation: A Review Of Pathogen And Strategies To Control. “The Journal of Animal & Plant Sciences, 27(4): 2017, Page: 1058-1066 ISSN: 1018-7081.

[7] S. Y. Padmanabhan. “Studies On Forecasting Outbreaks Of Blast Disease Of Rice. Influence of Meteorological Factors on Blast Incidence at Cuttack.” Received January 16, 1965.

[8] Chang Kyu Kim And Choong Hoe Kim. “The rice leaf blast simulation model

EPIBLAST." F. W. T. Penning de Vries et al. reds.), Systems Approaches for Agricultural Development, 309-321.

[9] [S. B. Calvero, S. M. Coakley, P. S. Teng. "Development of Empirical Forecasting models for rice blast on weather factors." Plant Pathology(1996)45,667-668.

[10] Kwang-Hyung Kim and Imgook Jung . "Development of a Daily Epidemiological Model of Rice Blast Tailored for Seasonal Disease Early Warning in South Korea." Plant Pathol. J. 36(5) : 406-417 (2020),p ISSN 1598-2254 eISSN 2093-9280.

[11] Omkar Singh, Jagadeesh Bathula and DK Singh. "Rice blast modeling and forecasting." International Journal of Chemical Studies 2019; 7(6): 2788-2799.

[12] Dimitrios Katsantonis, Kalliopi Kadoglidou , Christos Dramalis And Pau Puigdollers. "Rice blast forecasting models and their practical value: a review." Phytopathologia Mediterranea (2017), 56, 2, 187−216 www.fupress.com/pm ISSN (print): 0031-9465 Firenze University Press ISSN (online): 1593-2095 DOI: 10.14601/Phytopathol_Mediterr-18706.

[13] Sarinya Kirtphaiboon , Usa Humphries , Amir Khan , Abdullahi Yusuf." Model of rice blast disease under tropical climate conditions." Chaos, Solitons and Fractals 143(2021)110530.

[14] Kwang-Hyung Kim, Jaepil Cho, Yong Hwan Lee , Woo-Seop Lee. "Predicting potential epidemics of rice leaf blast and sheath blight in South .Korea under the RCP 4.5 and RCP 8.5 climate change scenarios using a rice disease epidemiology model, EPIRICE." Agricultural and Forest Meteorology 203(2015)191-207.

[15] Kwang-Hyung Kim & Jaepil Cho. "Predicting potential epidemics of rice diseases in Korea using multi-model ensembles for assessment of climate change impacts with uncertainty information." Climatic Change (2016) 134:327–339.

[16] Shafaullah, Muhammad Aslam Khan, Nasir Ahmed Khan And Yasir Mahmood. "Effect Of Epidemiological Factors On The Incidence Of Paddy Blast (Pyricularia Oryzae) Disease." Pak J. Phytopayhol., Vol 23(2):108-111, 2011.

[17] Rupankar Bhagawati, Kaushik Bhagawati, D. Jini, R. A. Alone, R. Singh, A. Chandra, B. Makdoh, Amit Sen and Kshitiz K. Shukla. "Review on Climate Change and its Impact on Agriculture of Arunachal Pradesh in the Northeastern Himalayan Region of India." Nature Environment and Pollution Technology An International Quarterly Scientific Journal Vol. 16 No. 2 pp. 535-539 2017.

[18] Y.H. Gu , S.J. Yoo , C.J. Park , Y.H. Kim , S.K. Park , J.S. Kim , J.H. Lim "BLITE-SVR: New forecasting model for late blight on potato using support-vector regression." Computer and Electronics in Agriculture 130(2016) 169-176.

[19] Varsha M., Dr. Poornima B., "A Comparative Study of Ensemble Classifiers for Paddy Blast Disease Prediction Model." International Journal on Recent and Innovation Trends in Computing and Communication. ISSN: 2321-8169 Volume: 8 Issue: 9.

[20] David F. Nettleton , Dimitrios Katsantonis, Argyris Kalaitzidis, Natasa Sarafijanovic-Djukic, Pau Puigdollers and Roberto Confalonieri4 . "Predicting rice blast disease: machine learning versus process-based models." Nettleton et al. BMC Bioinformatics

(2019) 20:514.

[21] Spyridon D. Koutroubas, Dimitrios Katsantonis, Dimitrios A. Ntanos, Elisabetta Lupotto. “Blast Disease Influence On Agronomic And Quality Traits Of Rice Varieties Under Mediterranean Conditions.” Turk J Agric For 33 (2009) 487-494.

[22] Aziiba Emmanuel Asibi, Qiang Chai and Jeffrey A. Coulter. “Rice Blast: A Disease with Implications for Global Food Security.” Agronomy 2019, 9, 451; doi:10.3390/agronomy9080451.

[23] Serge Savary , Andrew Nelson , Laetitia Willocquet , Ireneo Pangga , Jorrel Aunario. “Modeling and mapping potential epidemics of rice diseases globally.” Crop Protection 34(2012)6-17.

[24] unyong Bae, Jeeyea Ahn & Seung Jun Lee. “Comparison of Multilayer Perceptron and Long Short-Term Memory for Plant Parameter Trend Prediction.” Nuclear Technology ISSN: 0029-5450 (Print) 1943-7471 (Online).

[25] 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, 7:485 doi:10.1186/1471-2105-7-485.

[26] Jia-You Hsieh, Wei Huang, Hsin-Tieh Yang, Chia-Chieh Lin, Yo-Chung Fan, Huan Chen. “Building the Rice Blast Disease Prediction Model based on Machine Learning and Neural Networks.” Easy Chair Preprint.

[27] Alvin R. Malicdem, Proceso L. Fernandez, “Rice Blast Disease Forecasting For Northern Philippines.” Wseas Transactions On Information Science And Applications. Volume 12, 2015 E-Issn: 2224-3402.

[28] Yangseon Kim , Jae-Hwan Roh and Ha Young Kim. “Early Forecasting of Rice Blast Disease Using Long Short-Term Memory Recurrent Neural Networks.” Sustainability 2018, 10, 34; doi:10.3390/su10010034.

[29] Bitzel Cortez, Berny Carrera, Young-Jin Kim, Jae-Yoon Jung. “An architecture for emergency event prediction using LSTM recurrent neural networks.” Experts System with Application 97(2018)315-324.

[30] Miguel Martínez Comesaña, Lara Febrero-Garrido, Francisco Troncoso-Pastoriza and Javier Mar’tínez-Torres. “Prediction of Building’s Thermal Performance Using LSTM and MLP Neural Networks. Appl. Sci. 2020, 10, 7439; doi:10.3390/app10217439.