https://doi.org/10.48047/AFJBS.6.12.2024.3559-3576

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Deep Learning and Statistical Models for Predicting Occurrence of YSB in rice Based on Weather Parameters

Satish Kumar Yadav[1*], D. Pawar[1], Latika Yadav[2], Saurabh Tripathi[3] and Priyanka Mishra[3]

[1]Department of Statistics, Amity Institute of Applied Sciences, Amity University, Noida-201313
[2]Vijay Singh Pathik Government Post Graduate College Kairana, Shamli Uttar Pradesh, India
[3]Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar, Uttarakhand
*Communicating author email: satishkumaryadav.akash@gmail.com

Article History
Volume 6, Issue 12, 2024
Received: 02 Jun 2024
Accepted: 25 Jun 2024
doi: 10.48047/AFJBS.6.12.2024.3559-3576

Abstract
Incidence of Yellow stem borer (*Scirpophaga incertulas*)(YSB) on Rice (*Oryza sativa L.)*atChinsurah, West Bengal, India is modelled based on field data sets generated during six kharif seasons [2011-20]. The weather variables considered are maximum & minimum temperature (MaxT & MinT) (^{0}C), morning and evening humidity (RHM & RHE) (%), sunshine hours (SS) (hr/d), wind velocity (Wind) (km/hr), total rainfall (RF) (mm) and rainy days (RD).Long Short-Term Memory (LSTM)networks, which arecapable of learning long-term temporal dependencies, are used to overcome the limitations oftraditional machine learning techniques. The results indicate that LSTMand Gated Recurrent Unit (GRU) models, although morecomputationally expensive, provide a more accurate solution for pest prediction compared withother methods. Correlation analyses indicate significant positive influence of maximum and minimum temperature on YSB. An empirical comparison of the above models is carried out based on root mean square error (RMSE) and mean square error (MSE). It is observed that, for YSB, the MSE and RMSE values of LSTM and GRU are less as compared to other competing models. Diebold-Mariano (D-M) test was applied for comparison of forecasting performance among the applied models. It is observed that, in the studied pest, predictive accuracy ofLSTM is higher than that of other models. The analysis is carried out using R package.
Keywords: YSB, Rice, Accuracy, Deep Learning

Introduction

Natural and artificial agro-environments are vulnerable to the pernicious and occurring climate change phenomenon, and rice ecosystems are no exception. Rice (*Oryza sativa L.)* is farmed in various climatic zones and ecoregions throughout India, depending on temperature, soil type, water availability, rainfall, and other climatic factors, with a single crop to three crops being harvested each year. India grows rice in 43 Mha with production of 112 million tons (Mt) of milled rice and average productivity of 2.6 ton/ha (Pathak *et al.* 2022). Rice is one of the families of Poaceae's most important staple foods. India, the world's second-largest rice producer, produces 104.32 million tonnes and cultivates approximately 44.6 million hectares with an average yield of 2.34 tonnes per hectare (Anon., 2013; Rajasekar and Jeyakumar 2014). During Kharif, 84% of 42.7 million hectares are planted with rice, with sowings beginning in June and July (Annual 2019). Despite India's 28 percent pesticide use, insect pests cause an estimated 25% output loss in rice cultivation (Dhaliwal *et al.* 2010). West Bengal is India's second-largest rice producer (14,771,000 tonnes) and second-largest rice producer (5,386,000 acres) (Chatterjee *et al.* 2017). Around 78 percent of total rice acreage is classified as high or medium productivity, accounting for almost 84 percent of total rice production in the state. A comprehensive examination of the discrepancy between potential and actual rice yields across the country reveals various factors that operate as yield restrictions. Among these issues, insect pests significantly contribute to yield loss and productivity losses in rice production (Chatterjee *et al.* 2016). Around 100 bug species feed on rice in India, and 20 of these are regarded as severe pests, causing a 30% production loss. Stem borers cause damage to cereal crops worldwide (Lawani 1982; Heinrichs 1985; Kfir *et al.* 2002). Yellow stem borer (YSB) (Nos./weeks/trap), Scirpopagha in certulas Walker, and rice leaf folder, Cnaphalocrocis medinalis Guenee, are the most prevalent and damaging insect pests in the country, accounting for around 10% to 60% of overall production loss (Chatterjee and Mondal 2014). Yellow stem borer (YSB) (Scirpophaga in certul as (Walker) (Pyralidae; Lepidoptera) is a common and significant rice pest found throughout India's agro-climatic zones. During the vegetative stage, YSB larvae feed on the central shoots of rice tillers, causing 'dead heart' and 'white ear' symptoms if feeding occurs during the panicle initiation stage. YSB damage to rice is a severe problem across India's rice-growing regions, as it occurs at both vegetative and reproductive crop stages and results in output losses of 27–34 percent per year (Prasad *et al.* 2007). The abundance of insect pests and the severity of projected damage in rice ecosystems are represented by light trap catches, and YSB moth catches in light traps are frequently employed as a monitoring tool throughout the year. The weather has a significant impact on all agricultural operations, including the pest management practices used by farmers. Conducive weather circumstances favor an increase in the prevalence of insect pests, especially YSB, whose bionomics is inextricably linked to the prevailing weather. Establishing a relationship between YSB light trap catches and weather gives critical information about the population's timing and abundance. In India, short-term projections for YSB were tried using monthly seasonal indices of biotic variables based on light trap catches and weather characteristics (Ramakrishnan *et al.* 1994). As a result, it becomes vital to construct weather-based models that incorporate historical data sets to forecast the intensity of YSB for use in its forewarning. Climate change is a global phenomenon, and its impact on insect pests is unavoidable sooner or later. Predicting the future appears to be a skill everyone desires, mainly when it might result in advantages. Artificial neural networks (ANN) and autoregressive integrated moving average (ARIMA) models are utilized. Naturally, some researchers have claimed in recent years that a long short-term memory (LSTM) network has superior prediction accuracy. (Xue *et al*. 2020) built a high-precision short-term forecasting model for financial market time series in 2020 and compared it to the BP neural network, the standard RNN, and the upgraded LSTM deep neural network. -e findings demonstrated that the LSTM deep neural network has a high forecasting accuracy and can accurately anticipate stock market time series (Xue *et al*. 2020). ARIMA forecasts are generated using the values of the input variables and their associated error terms. (Tabachnick *et al*. 2001). ARIMA is a linear regression model, but it is not confined to it; ARIMA may exhibit certain variations when confronted with complex nonlinear practical problems. However, linear models typically outperform sophisticated structural models in short-term predictions (Meyler *et al*. 1998). A neural network (ANN) is a data-driven adaptive model that makes essentially no prior assumptions. (Khashei *et al*. 2010) It is frequently utilized in various sectors, including banking, business, and engineering, as a predictive model. The prediction of ANN is based on the findings acquired from the original data, which are used to create broad observations and then infer the possible component of the whole. Compared to the ARIMA model, this one is extremely good at solving non-linear issues. The stock market's movements are similarly non-linear. The causality of the price of the vegetable with the other contributing factors has not been addressed. only multi collinearity of the factors among themselves does not provides information regarding the causality co-efficient between the dependent and the independent variables. (Chen *et al.* 2019) suggests that using the wavelet analysis with LSTM technique could be used to predict the commodity pricing. The importance is given to the temporal information of the data and this forms the reason behind using the LSTM as this technique with RNN works good with long term data. The limitation of using RNN with LSTM with no external features to contribute to the independent variable.

Materials and methods

Modern agriculture has to cope with several challenges, including the increasing call for food, as a consequence of the global explosion of earth's population, climate changes (Thayer *et al*. 2020) natural resources depletion (Nassani *et al.* 2019), alteration of dietary choices (Conrad *et al*. 2019), as well as safety and health concerns (Benos *et al.*2018). As a means of addressing the above issues, placing pressure on the agricultural sector, there exists an urgent necessity for optimizing the effectiveness of agricultural practices by, simultaneously, lessening the environmental burden. In particular, these two essentials have driven the transformation of agriculture into precision agriculture. This modernization of farming has a great potential to assure sustainability, maximal productivity, and a safe environment (Lampridi *et al.* 2019). In general, smart farming is based on four key pillars to deal with the increasing needs; (a) optimal natural resources' management, (b) conservation of the ecosystem, (c) development of adequate services, and (d) utilization of modern technologies (Zecca *et al.* 2019). An essential prerequisite of modern agriculture is, definitely, the adoption of Information and Communication Technology (ICT), which is promoted by policy-makers around the world. ICT can indicatively include farm management information systems, humidity and soil sensors, accelerometers, wireless sensor networks, cameras, drones, low-cost satellites, online services, and automated guided vehicles (Sorensen *et al.* 2019). The large volume of data, which is produced by digital technologies and usually referred to as "big data", needs large storage capabilities in addition to editing, analyzing, and interpreting. The latter has a considerable potential to add value for society, environment, and decision-makers (Sonka *et al.* 2016). Nevertheless, big data encompass challenges on account of their so-called "5-V" requirements; (a) Volume, (b) Variety, (c) Velocity, (d) Veracity, and (e) Value (Meng *et al.* 2016). The conventional data processing techniques are incapable of meeting the constantly growing demands in the new era of smart farming, which is an important obstacle for extracting valuable information from field data (Evstatiev *et al.* 2020). To that end, Machine Learning (ML) has emerged, which is a subset of artificial intelligence (Swiergosz *et al.* 2020), by taking advantage of the exponential computational power capacity growth.

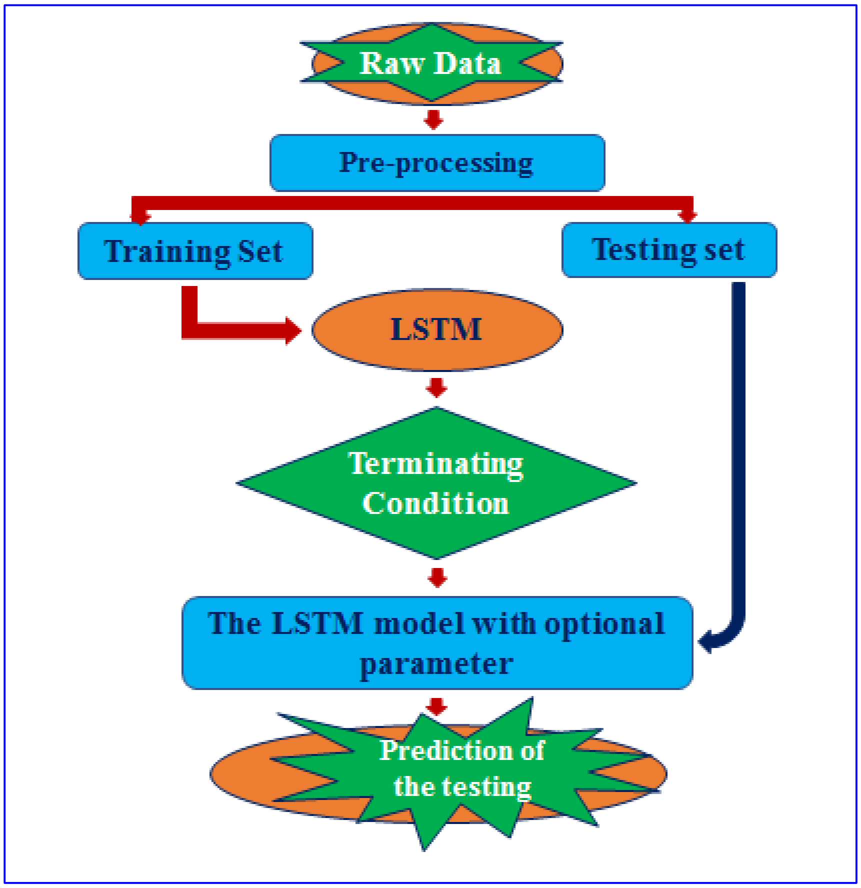

Figure 1: The above diagram showed the protocol followed in this study in implementation of deep learning.

Statistical analyses

Autoregressive Integrated Moving Average Model with exogenous variable (ARIMAX)

The ARIMAX model (Bierens 1987) is a generalization of the ARIMA model capable of incorporating an external input variable (X). Given a (k+1) time-series process $\{(y_t, x_t)\}$, where y_t and k-components of x_t are real valued random variables, the ARIMAX model assumes the form

$$\left(1 - \sum_{s=1}^{p} \alpha_s L^s\right) \Delta y_t = \mu + \sum_{s=1}^{q} \beta_s^{'} L^s x_t + \left(1 + \sum_{s=1}^{r} \gamma_s L^s\right) e_t, \quad (1)$$

Where L is usual lag operator, i.e. $L^s y_t = y_{t-s}, \Delta y_t = y_t - y_{t-1}$, $\mu \in$ R,$\alpha_s \in$ R,$\beta_s \in R^k$and$\gamma_s \in R$ are unknown parameters and e_t, s are the errors, and p, q and r are natural numbers specified in advance. The estimation of the parameters of ARIMAX model was carried out using maximum likelihood method in R software package.

Random forest

A random forest (RF) is a predictor consisting of a collection of randomized base regression trees $\{r_n(\mathbf{x},\Theta_m,D_n), m \geq 1\}$, where $\Theta_1,\Theta_2,...$ are i.i.d. outputs of a randomizing variable Θ. These random trees are combined to form the aggregated regression estimate

$$r_n(X,D_n) = E_\Theta\,[r_n(X,\Theta,D_n)], \quad (2)$$

where E_Θ denotes expectation with respect to the random parameter, conditionally on **X** and the data set D_n. In the following, to lighten notation a little, we will omit the dependency of the estimates in the sample, and write for example $r_n(\mathbf{X})$ instead of $r_n(\mathbf{X},D_n)$. Note that, in practice, the above expectation is evaluated by Monte Carlo, that is, by generating M (usually large) random trees, and taking the average of the individual outcomes. The randomizing variable Θ is used to determine how the successive cuts are performed when building the individual trees, such as selection of the coordinate to split and position of the split (Figure 2).

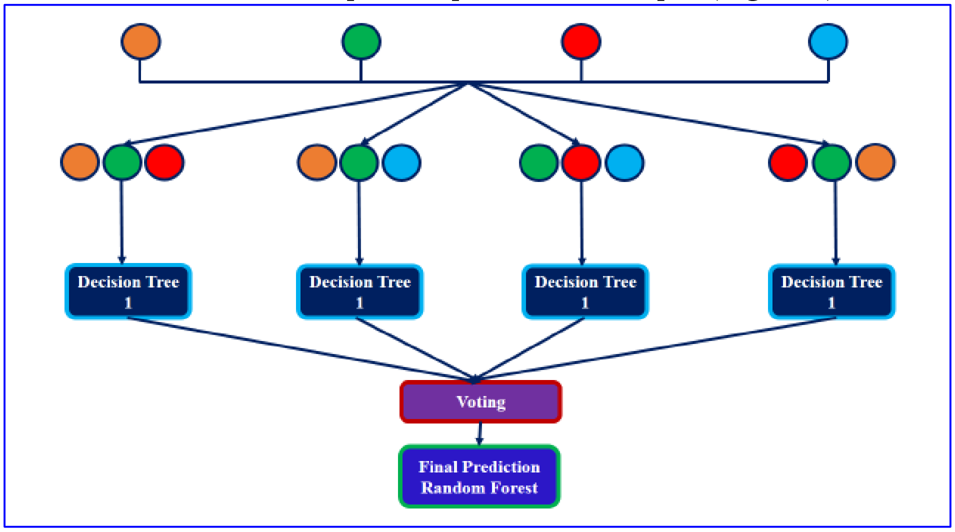

Figure 2: Workflow of random forest regression machine learning algorithm

Artificial neural network (ANN)

ANN, a powerful a self-adaptive approach for modeling nonlinear data was applied to datasets where the underlying data relationship was unknown. A general neural network consists of an input layer that accept external information, one or more hidden layers that provide non-linearity to the model and output layer that provides the target value. Each layer consists of one or more nodes. All the layers are connected through acyclic arc. Each input node in the input layer is associated with its corresponding weight. To compute the output, its activation function is applied to the weighted sum of the inputs. The activation function is either the identity function or sigmoidal function. Most commonly used ANN is multi-layer perceptron (MLP), a class of feed forward neural network. MLP consists of at least three layers of nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. An application of this approach can be found (Paul and Sinha 2016). A graphical presentation of MLP is given in Figure 3.

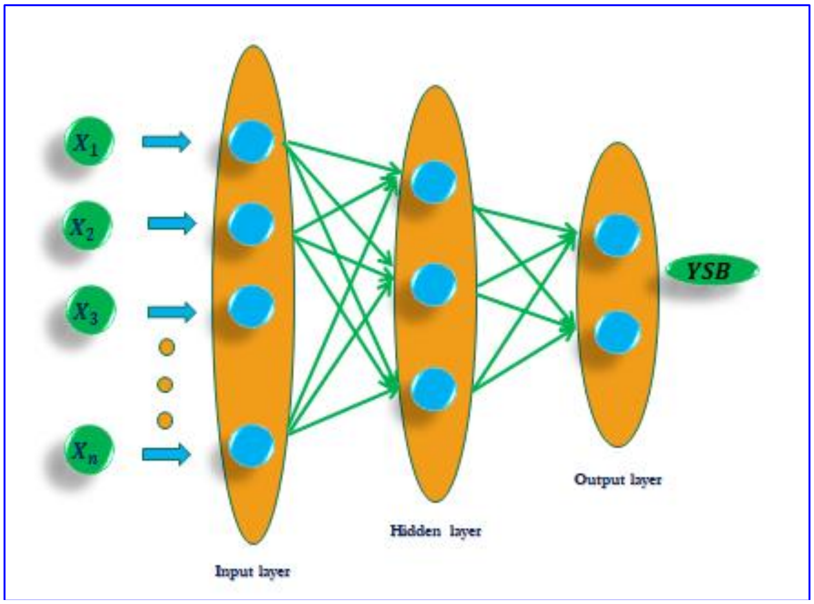

Figure 3: A multilayer perceptron (MLP) architecture with one hidden layer

Long Short-Term Memory (LSTM) Model

Hochreiter and Schmidhuber introduced the LSTM neural network in 1997 to include the benefits of addressing long-term data dependencies (Hochreiter & Schmidhuber, 1997). Because of these characteristics, the LSTM model was very practical for financial high-frequency time series data. LSTM model is developed to solve the problems of recurrent neural network (RNN) models, gradient expansion, and gradient disappearance (Ta *et al.*, 2020). LSTM model has three memory modules: input gate, output gate and forget gate (figure 5). The main functions of these three gates are retaining important information and discarding irrelevant information from the units. A variety of LSTM models are available in the literature, we use Hochreiter and Schmidhuber's LSTM model in our study.

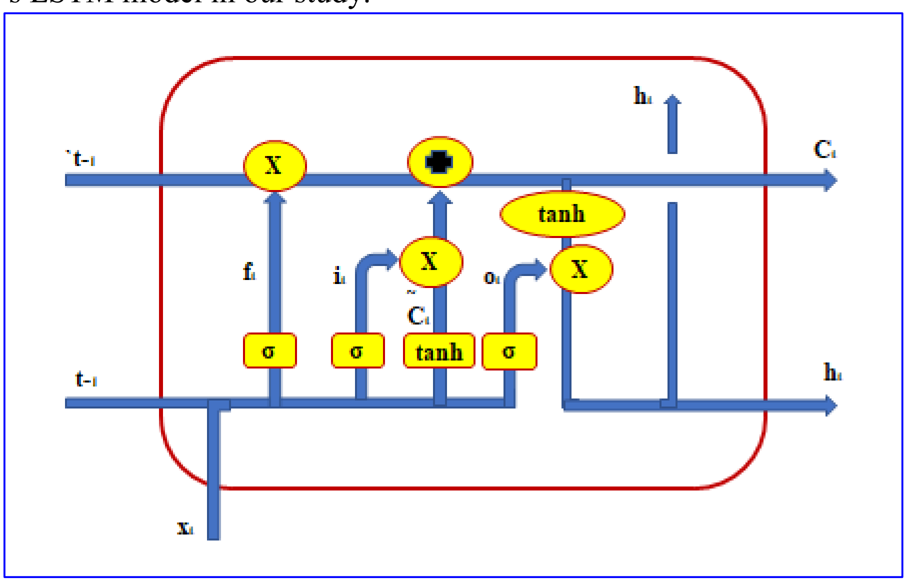

Figure 4. Schematic representation of LSTM

The operational premise of LSTM is to analyze the information at time t, first, it discards the unnecessary information through forget gate, then filters the useful information with a given probability by the input gate, and ultimately extract useful information through the output gate which participates in the next LSTM unit. The selection of activation function is an important step in the LSTM process. Here we have used the standard sigmoid function and the tanh function as activation function. The LSTM process can be summarized in five steps.

Step1: The previous unit's output value and the current unit's input value are integrated into the forget gate. The forget gate's output value is calculated by the following formula:

$$f_t = \sigma\{W_f * (h_{t-1} * x_t)\} + b_f \tag{3}$$

where W_f is the forget gate's weight, and b_f is the bias, x_t is the input value, and h_{t-1} is the output value of the prior unit.

Step 2: The output value of the prior unit and the input value of the current time are incorporated into the input gate. The output value and candidate cell state values are computed by the following formulas

$$i_t = \sigma\{W_i * (h_{t-1} * x_t)\} + b_i \tag{4}$$

$$\widetilde{C}_t = tanh\{W_c * (h_{t-1} * x_t)\} + b_c \tag{5}$$

where W_i is the weight of this gate, and b_i is the bias W_c and b_c are the weight and bias of the candidate input respectively.

Step 3: Updation of the current cell is done using the formula given below:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{6}$$

Step 4: the output gate takes h_{t-1} and x_t as input values, and the output of the output gate is calculated by the following formula

$$o_t = \sigma\{W_o * (h_{t-1} * x_t)\} + b_o \tag{7}$$

where, W_i and $b_i$are the weight and bias of this gate respectively.

Step 5: The final output of the LSTM unit is generated by computing the output gate output and the cell state, as indicated in the following formula

$$h_t = o_t * tanh(C_t) \tag{8}$$

Validation of forecasts

The dataset of YSB population and weather was divided in two parts before analysis. For each locationwith 90% of the observations were used for estimation (model development) and remaining 10% 0bservations were used for validation. Comparative assessment of prediction performance of different models namely RNN, GRU, LSTM, Bidirectional LSTM, Deep LSTM, ARIMA-X, ANN, SVR and RF models was carried out in terms of root mean square error (RMSE) based on the following formulae:

$$\text{RMSE}=\sqrt{\sum_{i=1}^{n}\frac{\left(\text{Predicted}(y_i)-\text{observed}(y_i)\right)^2}{n}}\ \text{RMAPE}=\frac{1}{n}\frac{\sum_{i=1}^{n}\left|\left(\text{observed}(y_i)-\text{predicted}(y_i)\right)\right|}{\text{observed}(y_i)} \times 100$$

$$\text{Model Accuracy}=100-\text{RMAPE}$$

where h denotes the number of observations for validation, y_i is the observed value and $\hat{y}_i$ is the predicted one. Diebold Mariano test (Diebold and Mariano 1995) was also conducted for different pairs of models to test for differences in predictive accuracy between any two competing models. Other common evaluation metrics were the coefficient of correlation (R), coefficient of determination (R^2; basically, the square of the correlation coefficient), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE), which can be given via the following relationships (De Myttenaere *et al.* 2016; Lehmann *et al.* 1998). Where X(t) and Z(t) correspond to the predicted and real value, respectively, t stands for the iteration at each point, while T for the testing records number. Accordingly, low values of MAE, MAPE, and MSE values denote a small error and, hence, better performance. In contrast, which demonstrates better model performance and also that the regression curve efficiently fits the data.

Materials and Methods

Study locations

Study was a part of information and communication technology (ICT) based pest surveillance on rice implemented at experimental research stations of Chinsurah (WB) India. Two fields grown with rice each of one acre from 10 villages located within 30 km radius of meteorological observatory of experimental station constituted surveillance plan during study seasons. Five spots per field and two plants per spot selected randomly were accounted for weekly observations on YSB whole plant basis from early vegetative stage till crop harvest. While the surveillance plans were fixed during season the sampling plan for YSB followed random pattern for selection of spots and plants. For all surveillance fields, general information relating to field area, cultivar grown, dates of sowing and other production practices were also collected.

Data accrual and reporting system

Proforma of pest surveillance for rice (ref *http://www.ncipm.res.in/Nicra2015/NICRAAdminPanelNew/rvLogin.aspx)* which also had variables of YSB was used for recording spot wise observations during each week. Client software developed for offline entry and on-line upload served to accrue data collected in respect of each field during each week. Reporting system developed worked online for extraction of data of each field across spots for each week of observation along standard meteorological weeks (SMW) in respect of seasons. Field observations on yellow stem borer (YSB) (Nos./weeks/trap) (figure 5) was carried out on weekly basis in 10 villages located within 30 km radius of meteorological observatory of Chinsurah district in West Bengal (23:02:06 lat. and 88:21:13 log.) Through Real Time Pest Dynamics (RTPD) surveillance under National Initiative on Climate Resilient Agriculture (NICRA). In each field, five spots were randomly selected and observations were made on ten plants randomly selected per spot from each field at weekly intervals right from vegetative stage and continued till the harvest crop. Major rice commonly grown by the farmers during kharif with row and plant spacing of 90 x 30 cm were considered for YSB (Nos./weeks/trap) surveillance during 2011-20 seasons.

Figure 5. Infected of the leaf and fruitby YSB

Meteorologicalobservations

The weather data viz. maximum and minimum temperature (^{0}C), morning and evening humidity (%), total rainfall (mm), and rainy days on standard meteorological week (SMW) basis were recorded from Chinsurah meteorological observatory of West Bengal. The influence of weather parameters on YSB (Nos./weeks/trap) (Chinsurah) was assessed.

Seasonal dynamics and status of YSB

Graphical representation of data related to the reviewed studies,Epidemics of incidence have increased in recent years due to climate change and there is a need to understand the impact of climate change on host pathogen interaction to outline appropriate management strategies (Shepard 1995). Studies on YSB in rice carried out for ten consecutive kharif seasons (2011-20) in Chinsurah (WB) location showed the commencement of infestation from second week of August with peak incidence between third week of October and November. YSB was higher during 2019. Lowest of YSB was in 2014 (figure 6). Considering the type and the accessibility of data criteria, (Baldini *et al.* 2017) distinguish the data in experimental and model data.The seasonal variation in occurrence of YSB in rice studied location is graphically represented in Figure 2. Among them, the yellow stem borer (YSB), Scirpophaga incertulas (Walker) (Lepidoptera), is the most destructive widely occurring pest that attacks rice throughout the growing season (Rubia 1989). YSB on rice gains top priority across rice-growing regions of India as its damage is at both vegetative and reproductive crop stages and causes yield losses to an extent of 27–34% every year (Prasad 2007). Light trap catches represent the abundance of insect pests and severity of anticipated damage in rice ecosystems, and YSB moth catches in light traps are often used as monitoring tool throughout the year. Documented evidences of changing weather over onset, peak and severity of insect pests across crops generated through laboratory experiments and field data in India are available (Annual 2016).

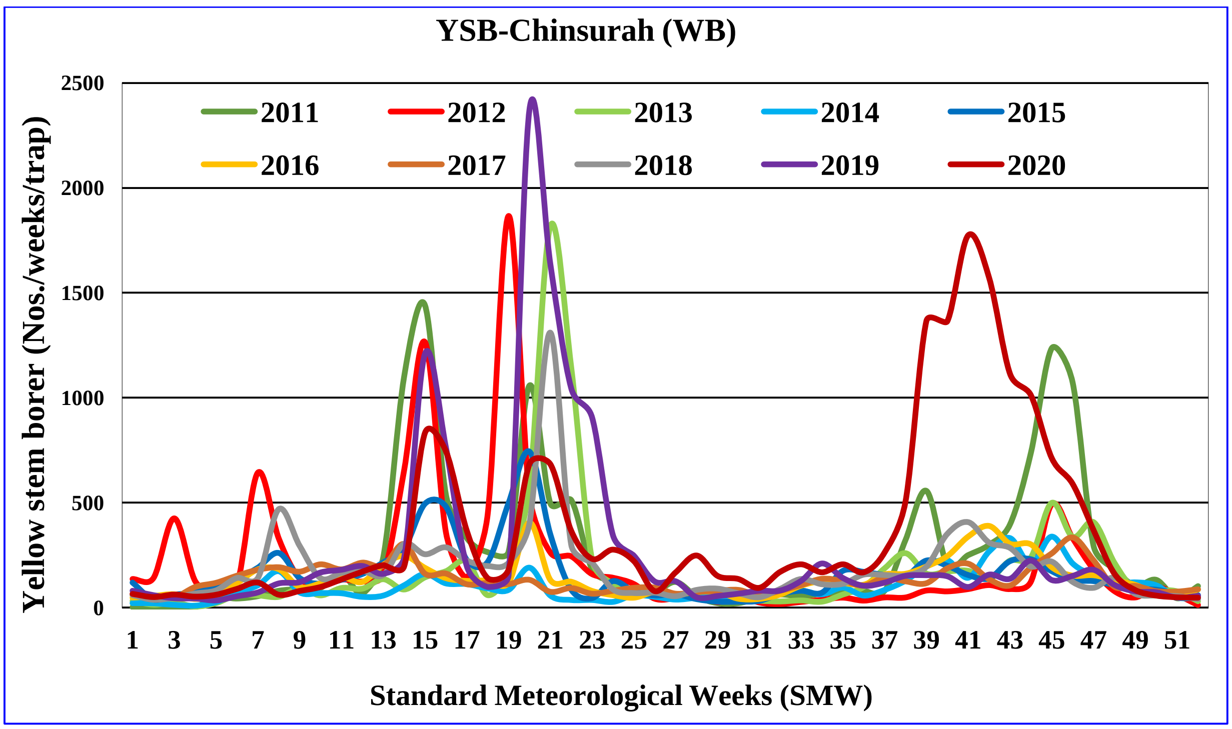

Figure 6. The seasonal variation of YSB occurrence in rice

Analysis of variance (ANOVA)

Comparative analysis of YSB occurrence across the years

Comparisons of YSB mean population for levels across seasons carried out was analysed using one-way analysis of variance (ANOVA) after arcsine transformation with mean comparisons made through using Duncan's Multiple Range Test (DMRT) are presented in table 1 (Vargas *et al.* 2010).YSB was significantly lower in 2014 as compared to others and 2020 with on par incidence during other seasons (Shelly *et al.* 2014).

Table 1. Comparative analysis of YSB occurrence across the years

2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
268.44^{b}	215.77^{b}	180.56^{bc}	99.71^{c}	163.23^{b}	137.38^{b}	132.1^{b}	186.48^{b}	263.65^{b}	365.31^{a}

*** Means followed by the superscript of same at p<0.05 based on DMRT**

Descriptive statistics of YSB occurrence

The descriptive statistics of YSB (Nos./weeks/trap) occurrence has been reported in table 2. A perusal of table 2 indicates that. Variability in YSB population measured in terms of Standard Deviation was high and ranged from 281.89. Maximum 2374 (Nos./weeks/trap) YSB occurrence. YSB showed positively skewed and leptokurtic distribution. The variability in Rainfall measured in terms of coefficient of variation (CV) was high and ranged with minimum 2.00 (Nos./weeks/trap) recorded. The understanding shape of data is a crucial action. It helps to understand where the most information is lying and analyze the outliers in a given data. In the one-dimensional case, the interesting parameters are the population mean and variance. The effect of non-normal data on statistical inference for these two parameters can be totally characterized by skewness and kurtosis. The concepts of skewness and kurtosis in the one-dimensional case are well known to graduate students in social sciences (Tabachnick *et al.* 2001). YSB occurrence is positively skewed and leptokurtic. On the other hand, among the regressor variables, except MaxT and MinT all others are positively skewed (Blanca *et al.* 2013). As far as kurtosis is concerned, MaxT, MinT and RHE follow platykurtic distribution whereas other regressors follow leptokurtic distribution (An L *et al.* 2008). Analyses were carried out using R software (R Core Team 2013).

Table 2: Descriptive statistics of response variable with repressor variables

Statistic	**YSB**	**MaxT**	**MinT**	**RHM**	**RHE**	**Rainfall**	**RainyD**
Mean	201.26	31.23	20.84	93.14	62.21	30.40	1.51
Median	120.50	32.13	23.93	93.86	62.14	6.95	1.00
Minimum	2.00	16.32	5.63	47.00	29.86	0	0
Maximum	2374	39.4	29.36	221.29	115.29	418.2	7.00
Mode	59.00	32.40	25.93	96.29	56.14	0.00	0.00
Range	2372.00	23.08	23.73	174.29	85.43	418.20	7.00
SD	281.89	3.96	6.34	7.30	14.99	48.60	1.79
Skewness	3.86	-0.72	-0.69	9.70	0.03	2.76	1.03
Kurtosis	17.72	0.09	-0.96	185.70	-0.57	11.54	0.18
CV	140.06	12.69	30.41	7.84	24.10	**159.87**	118.45

SD: standard deviation; CV: coefficient variation

Goodness of fit tests

Before further analysis, normality check was carried out by means of Kolmogo-rov-Smirnov test and Anderson-Darling tests and it was observed that the YSB population in the Chinsurah (WB) location significantly deviated from normality in table 3 (Ramesh *et al.* 2019). Non-normality of the data triggered nonparametric method for modeling YSB (Nos./weeks/trap) based on climatic variables.

Table 3: Goodness-of-fit tests for normal distribution

Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p-Value**	
Kolmogorov-Smirnov	D	0.26	Pr > D	<0.010
Cramer-von Mises	W-Sq	13.82	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	72.58	Pr > A-Sq	<0.005

*****: significant at $p< 0.01$; *: significant at $p< 0.05$**

Pearson Correlation Coefficients

Correlation analyses between standard meteorological weeks (SMW) based YSB(Nos./weeks/trap) and weather parameters prior were made by simple correlation coefficient based on ten years' performance data obtain crop season 2011-20. Correlation analysis indicate that of YSB found to be significantly positively correlated with maximum and minimum temperature. Influence of different weather parameters on fluctuations of YSB was carried in Purulia, West Bengal (Animesh 2018). The correlation workout between meteorological parameters and population dynamics of YSB are presented in table 4. Weather is an important determinant in the population dynamics of the pests (Agrawal and Mehta 2007; Laxmi and Kumar 2011a, 2011b). Most of the earlier workers have utilized regression models (both linear and nonlinear) for insect pest disease forewarning (Desai *et al*. 2004; Chattopadhyay *et al.* 2005a, 2005b; Dhar *et al*. 2007; Kumar *et al.* 2012; Kumar *et al.* 2013).

Table 4: Correlation Coefficients of YSB with Climate variables

Pearson Correlation Coefficients, Prob > |r| under H0: Rho=0						
Trap	MaxT	MinT	RHM	RHE	Rainfall	RainyD
YSB	0.23***	0.17***	-0.07	0.01	-0.07	-0.06

*****: significant at $p< 0.01$; *: significant at $p< 0.05$**

Data has been divided in structural change in three groups. First, second and third groups is pink, green and blue respectively. Here also correlation analysis indicates that of YSB found to be significantly positively correlated with maximum and minimum temperature significantly positively correlated. Here maximum temperature found all three groups to be significantly positively correlated and minimum temperature in third group, RHM in first group to be significantly positively correlated respectively in figure 7. The two major factors that are responsible for considerable yield loss in rice are regular pest outbreaks and adverse weather conditions (Pathak 1994).

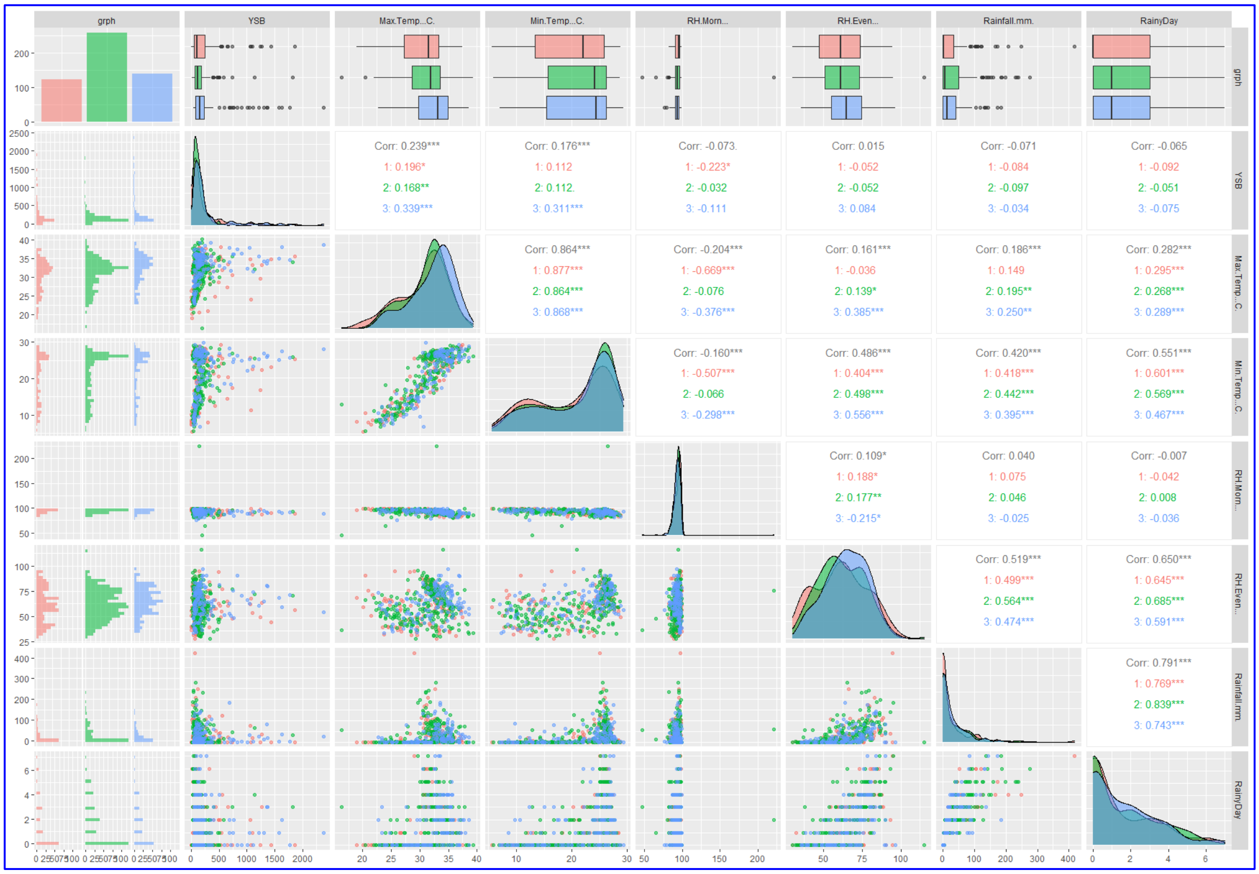

Figure 7: Correlation Coefficients of YSB occurrencewith Climate variables

Validation of forecasts

Short-term forecasts for YSB using light trap catches and weather parameters attempted in India using SMW seasonal indices of abiotic factors (Ramakrishnan *et al.* 1994), day degree and regression models (Krishnaiah *et al.* 1997) during last decade are available with their field use limited at present. Therefore, it becomes necessary to develop weather-based models including the data sets of the recent past to predict YSB occurrence for use in its forewarning. While variability in data sets over years of YSB occurrence was considered for development of categories of pest occurrence levels, congenial conditions of weather in respect of YSB occurrence (Krishnaiah 2004)

Table 5. Values in relation to RNN, GRU, LSTM, Bidirectional LSTM, Deep LSTM, ARIMA-X and ANN predicting of YSB.

Model	RMSE	MAE
RNN	336.44	162.81
GRU	**305.73**	150.82
LSTM	315.72	**141.35**
Bidirectional LSTM	316.10	155.06
Deep LSTM	346.76	165.19
ARIMA-X	461.52	231.43
ANN	452.43	264.04
SVR	412.86	217.67

RF	334.11	177.33

The success of the model does not lie in predicting the expected output based on the ten years' data but to efficiently predict with the same level of accuracy for all the incoming and future data. A good model should be robust enough to handle the changes in the data yet produce the same results. When comparing the evaluation metrics such as RMSE and MSE in table 5, the LSTM and GRU model outperforms the other models by some margin in figure 8. This some margin of difference could be explained by the following reasons figure 9. Forecasting of the development of population occurrence with as much accuracy as possible and describing the population dynamics correctly enable better management of this pest thus minimizing the yield loss caused by them. Efficient, economical and environmentally friendly management of the YSB can be done through knowledge of its timing of attack in relation crop phenology and the prevalent weather factors modelled to enable prediction of its occurrence that allows growers to take timely action in an efficient manner for crop management (Amrender *et al.* 2013; Kumar *et al.* 2012).

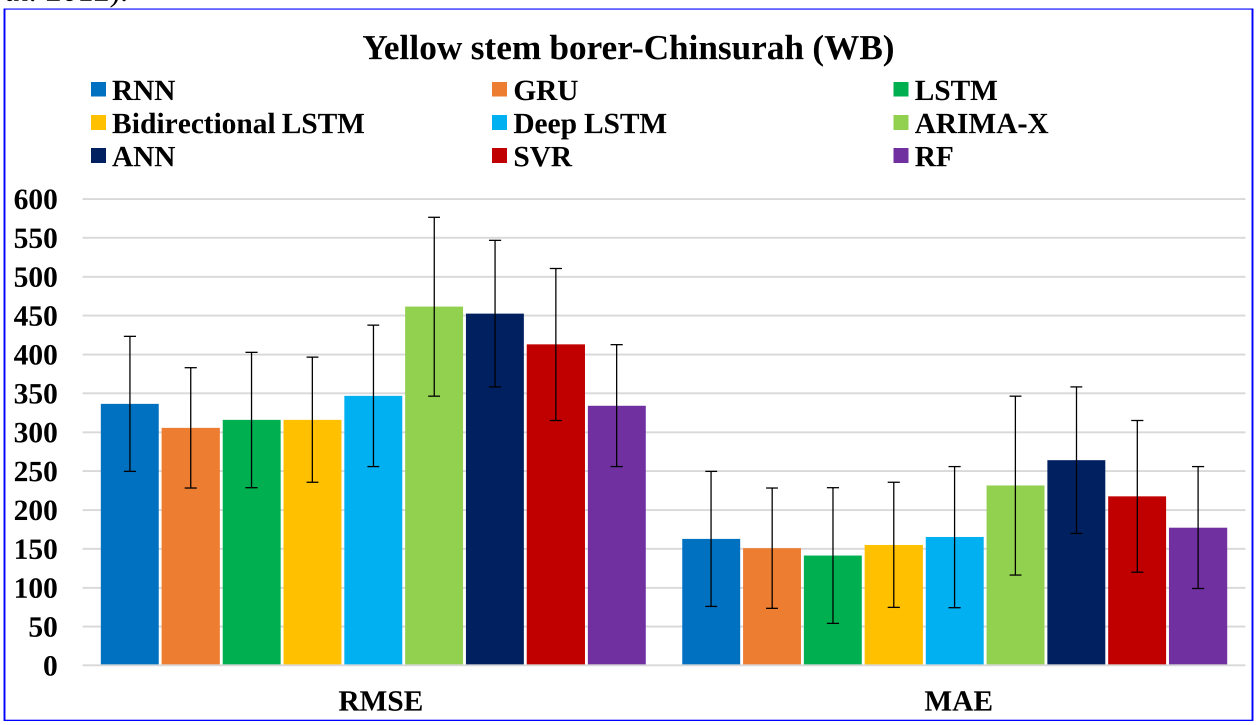

Figure 8. RMSE and MSE of different models for predicting YSB

Test results for yellow stem borer (WB)

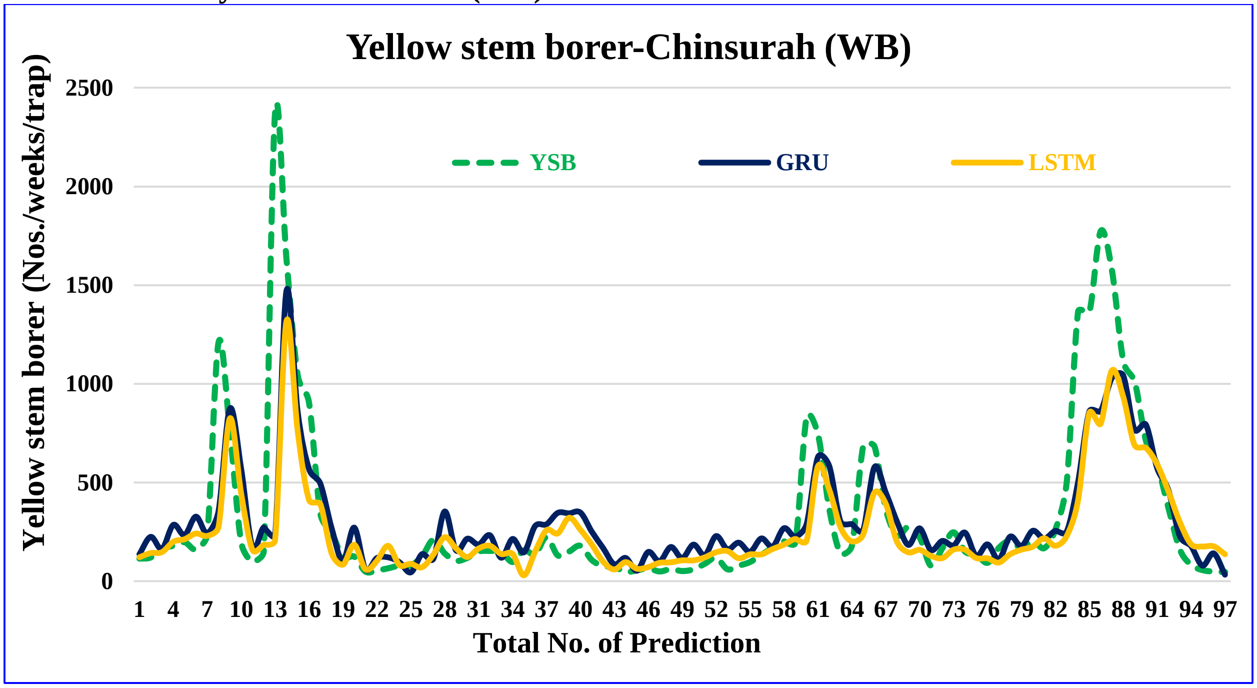

Figure 9. Different models for predicting YSB occurrence

Concussion

Climate change has an adverse impact on rice growing areas due to wide fluctuation in temperature and erratic rainfall patterns and assessment of seasonal dynamics of any pest in relation to weather variations is of significance. Present study revealed that YSB of rice at Chinsurah, West Bengal is on the decline with 23-34 SMW having the maximumoccurrence over 2019-2020. Approaches including deep learning techniques used for modelling incidence of YSB indicated varying performances. Empirically, LSTM and GRU model outperformed RNN, Bidirectional LSTM, Deep LSTM, ARIMA-X and ANN predicting of YSB models. The same conclusion may be drawn from the result of D-M test. Utilizing disease-weather interactions has resulted in improved models with higher prediction accuracy. The current models could form a part of prediction for future seasons and for estimating scenario of YSB for projected period of climate change. As the techniques used in the present investigation are mainly data driven, it is difficult to generalize the conclusion for all the disease and pests, but the same techniques can be replicated to other pests and diseases for gaining prediction accuracy.

ACKNOWLEDGEMENTS

Authors are grateful for funding by Indian Council of Agricultural Research, through National Innovations in Climate Resilient Agriculture implemented by Central Research Institute for Dryland Agriculture, Hyderabad.

REFERENCES:

1. Anonymous (2013) Directorate of Economics and Statistics. New Delhi, Govt. of India.
2. Agrawal R; Mehta SC (2007) Weather Based Forecasting of Crop Yields, Pests and Disease forecast system for oilseed Brassicas, The Indian Journal of Agricultural Sciences 82:608-614.

3. An L; Ahmed E (2008) Improving the performance of kurtosis estimator. Computational Statistic & Data Analysis, 52, 2669-2681.
4. Animesh M; Rajendra PM (2018) Impact of weather parameterson yellow stem borer RJLBPCS 4(6):731.
5. Annual report of Department of Agriculture Cooperation and Farmers Welfare (2014) http://agricoop.nic.in/Annualreport 2013-14/artp13-14ENG.pdf. Accessed 21 Aug 2017.
6. Annual Report of ICAR - National Research Centre for Integrated Pest Management (2016). http://www.ncipm.org.in/NCI PMPDFs/Annual%20Report%202015-16.pdf.
7. Anonymous (2013) Directorate of Economics and Statistics. New Delhi, Govt. of India.
8. Baldini C; Gardoni D; Guarino M (2017) A critical review of the recent evolution of Life Cycle Assessment applied to milk production. J. Clean. Prod 140:421–435.
9. Benos L; Bechar A; Bochtis D (2020) Safety and ergonomics in human-robot interactive agricultural operations. Biosyst. Eng 200:55–72.
10. Benos L; Bechar A; Bochtis D (2020) Safety and ergonomics in human-robot interactive agricultural operations. Biosyst. Eng 200:55–72.
11. Blanca MJ; Arnau J; López-Montiel D; Bono R; Bendayan R (2013) Skewness and kurtosis in real data samples. Methodology, 9:78-84. doi:10.1027/1614-2241/a000057
12. Chatterjee S; Dana I; Gangopadhyay C; Mondal P (2017) Monitoring of yellow stem borer, Scirpophaga incertulas (Walker) using pheromone trap and light trap along with determination of field incidence in kharif rice. J. Crop and Weed 13:15659.
13. Chatterjee S; Ghose M; Gangopadhyay C (2016) Field screening of different rice entries against different insect-pests of rice during kharif season. International Journal of Agriculture, Environment and Biotechnology 9(4): 667-671.
14. Chatterjee S; Mondal P (2014) Management of rice yellow stem borer, Scirpophaga incertulas Walker using some biorational insecticides. Journal of Biopesticides 7:143-147.
15. Chattopadhyay C; Agrawal R; Kumar A; Bhar LM; Meena PD; Meena RL; Khan SA; Chattopadhyay AK; Awasthi RP; Singh SN; Chakravarthy NV; Kumar KA; Singh RB; Bhunia CK (2005b) Epidemiology and forecasting of Alternaria blight of oilseed Brassica in India - a case study, Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz 112:351-365.
16. Chattopadhyay C; Agrawal R; Kumar A; Singh YP; Roy SK; Khan SA; Bhar LM; Chakravarthy NVK; Srivastava A; Patel BS; Srivastava B; Singh CP; Mehta SC (2005a) Forecasting of Lipaphis erysimi on oilseed Brassicas in India - A case study, Crop Protection 24:1042-1053.
17. Chen Q (2019) Price prediction of agricultural products based on wavelet analysis-lstm, in Proceedings - IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019. Institute of Electrical and Electronics Engineers Inc., pp. 984–990. doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00142.
18. Conrad Z; Niles MT; Neher DA; Roy ED; Tichenor NE; Jahns L (2018) Relationship between food waste, diet quality, and environmental sustainability. PLoS ONE 13.
19. Conrad Z; Niles MT; Neher DA; Roy ED; Tichenor NE; Jahns L (2018) Relationship between food waste, diet quality, and environmental sustainability. PLoS ONE 13.
20. De Myttenaere A; Golden B; Le-Grand B; Rossi F (2016) Mean Absolute Percentage Error for regression models. Neuro computing 192:38–48.

21. De Myttenaere A; Golden B; Le-Grand B; Rossi F (2016) Mean Absolute Percentage Error for regression models. Neuro computing 192:38–48.
22. Desai AG; Chattopadhyay C; Agrawal R; Kumar A; Meena RL; Meena PD; Sharma KC; Rao M; Srinivasa PYG; Ramakrishna YS (2004) Brassica juncea powdery mildew epidemiology and weather-based forecasting models for India - A case study", Journal of Plant Diseases and Protection 111(5):429-438.
23. Dhaliwal G; Jindal V; Dhawan A (2010) Insect pest problems and crop losses: changing trends. Indian J Ecol 37:1–7.
24. Dhar; Vishwa S; Kumar SKM; Agrawal R; Kumar A (2007) Prediction of pod borer (Helicoverpa armigera) infestation in short duration pigeonpea (Cajanus cajan) in central Uttar Pradesh, Indian Journal of Agricultural Sciences 77(10):701-704.
25. Evstatiev BI; Gabrovska-Evstatieva KG (2020) A review on the methods for big data analysis in agriculture. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Borovets, Bulgaria, 26–29 November 2020; IOP Publishing Ltd.: Bristol, UK 1032:012053.
26. Evstatiev BI; Gabrovska-Evstatieva KG (2020) A review on the methods for big data analysis in agriculture. In Proceedings of the IOP. Conference Series: Materials Science and Engineering, Borovets, Bulgaria, 26–29 November 2020; IOP Publishing Ltd.: Bristol 1032: 012053.
27. Heinrichs E (1985) Genetic Evaluation for Insect Resistance in Rice. International Rice.
28. Helm JM; Swiergosz AM; Haeberle HS; Karnuta JM; Schaffer JL; Krebs VE; Spitzer AI; Ramkumar PN (2020) Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. Curr. Rev. Musculoskelet. Med 13:69–76.
29. Kfir R; Overholt WA; Khan ZR; Polaszek A (2002) Biology and management of economically important lepidopteran cereal stem borers in Africa. Annu. Rev. Entomol. 47: 701–731.
30. Khashei M; Bijari M (2010) An artificial neural network (p, d, q) model for time series forecasting. Expert Systems with Applications 37(1):479–489.
31. Krishnaiah NV; Pasalu IC; Padmavathi L; Krisnhaiah K; Ram Prasad ASR (1997) Day degree requirement of rice yellow stem borer, Scirpophaga incertulas (Walker). Oryza 34:185–186.
32. Krishnaiah NV; Rama Prasad AS; Reddy CS; Pasalu IC; Mishra B; Ramakrishna YS; Prasad YG; Prabhakar M (2004) Forewarning and management of rice yellow stem borer, Scirpophaga incertulas (Walker). Directorate of Rice Research, Rajendranagar pp 48.
33. Kumar A; Agrawal R; Chattopadhyay C (2013) Weather based forecast models for diseases in mustard crop, Mausam 64(4):663-670.
34. Lampridi M; Sqrensen C; Bochtis D (2019) Agricultural Sustainability: A Review of Concepts and Methods. Sustainability 11:5120.
35. Lampridi M; Sørensen C; Bochtis D (2019) Agricultural Sustainability: A Review of Concepts and Methods. Sustainability 11:5120.
36. Lawani S (1982) A review of the effects of various agronomic practices on cereal stem
37. Laxmi RR; Kumar A (2011a) Forecasting of powdery mildew in mustard (Brassica juncea) crop using artificial neural networks approach, Indian J. Agric. Sci. 81:855-860.
38. Laxmi RR; Kumar A (2011b) Weather based forecasting model for crops yield using neural network approaches, Statistics and Applications, 9 (1&2 New Series), 55-69.

39. Lehmann EL; Casella N (1998) Theory of Point Estimation, 2nd ed.; Springer: New York, NY, USA, ISBN 978-0-387-98502-2.
40. Lehmann EL; Casella N (1998) Theory of Point Estimation, 2nd ed.; Springer: New York, NY, USA, ISBN 978-0-387-98502-2.
41. Meng T; Jing X; Yan Z; Pedrycz W (2020) A survey on machine learning for data fusion. Inf. Fusion 57:115–129.
42. Meng T; Jing X; Yan Z; Pedrycz W (2020) A survey on machine learning for data fusion. Inf. Fusion 57:115–129.
43. Meyler A; Kenny G; Quinn T (1998) Forecasting Irish Inflation Using ARIMA Models. Technical Paper 3/RT/1998, Central Bank of Ireland Research Department.
44. Nassani AA; Awan U; Zaman K; Hyder S; Aldakhil AM; Abro MMQ (2019) Management of natural resources and material pricing: Global evidence. Resour. Policy 64:101500.
45. Nassani AA; Awan U; Zaman K; Hyder S; Aldakhil AM; Abro MMQ (2019) Management of natural resources and material pricing: Global evidence. Resour. Policy 64:101500.
46. Pathak H; Tripathi R; Jambhulkar NN; Bisen JP; Panda BB (2020) Eco-regional Rice Farming for Enhancing Productivity, Profitability and Sustainability. NRRI Research Bulletin No. 22, ICAR National Rice Research Institute, Cuttack 753006, Odisha, India. pp 28.
47. Pathak MD; Khan ZR (1994) Insect Pests of Rice; International Rice Research Institute: Los Banos, Philippines.
48. Prasad SS; Gupta PK; Kanaujia BL (2007) Simulation study on yield loss due to Scirpophaga incertulas on semi deep-water rice. Ann Plant Prot Sci 15:491–492.
49. Prasad SS; Gupta PK; Kanaujia BL (2007) Simulation study on yield loss due to Scirpophaga incertulas on semi deep-water rice. Ann Plant Prot Sci 15:491–492.
50. Rajasekar N; Jeyakumar P (2014) Differential response of trifloxystrobin in combination with tebuconazole on growth, nutrient uptake and yield of rice (*Oryza sativa L.*). International Journal of Agriculture, Environment and Biotechnology 6(1): 87-93.
51. Ramakrishnan A; Sundaram A; Uthamasamy S (1994) Forecasting model for seasonal indices of stem borer population in rice. J Insect Sci 7(1):58–60 5.
52. Ramakrishnan A; Sundaram A; Uthamasamy S (1994). Forecasting model for seasonal indices of stem borer population in rice. J Insect Sci 7(1):58–60.
53. Ramesh SV Teegavarapu (2019) Methods for Analysis of Trends and Changes in Hydroclimatological Time-Series, Elsevier, pp 1-89, ISBN 9780128109854, https://doi.org/10.1016/B978-0-12-810985-4.00001-3.
    Research Institute, Manila, Philippines.
54. Shelly T; Epsky N; Jang EB; Reyes-Flores J; Vargas R (2014) Trapping and the Detection, Control, and Regulation of Tephritid Fruit Flies (New York (US): Springer Dordrecht Heidelberg.
55. Sonka S (2016) Big Data: Fueling the Next Evolution of Agricultural Innovation. J. Innov. Manag 4:114–136.
56. Sonka S (2016) Big Data: Fueling the Next Evolution of Agricultural Innovation. J. Innov. Manag 4:114–136.
57. Sørensen CAG; Kateris D; Bochtis D (2016) ICT Innovations and Smart Farming. In Communications in Computer and Information Science; Springer: Berlin/Heidelberg, Germany, 953:1–19.8.

58. Sorensen CAG; Kateris D; Bochtis D (2019) ICT Innovations and Smart Farming. In Communications in Computer and Information Science; Springer: Berlin/Heidelberg, Germany 953:1–19.
59. Swiergosz M; Haeberle AM; Karnuta HS; Schaffer MJ; Krebs JL; Spitzer EV; Helm JAI; Ramkumar PN (2020) Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. Curr. Rev. Musculoskelet. Med 13:69–76.
60. Tabachnick G; Fidell LS (2001) Using Multivariate Statistics, Pearson Education, Upper Saddle River, NJ, USA, 4th edition.
61. Tabachnick; Barbara G; Linda S; Fidell (2001) Using Multivariate Statistics 4th ed.New York: HarperCollins.
62. Thayer A; Vargas A; Castellanos A; Lafon C; McCarl B; Roelke D; Winemiller K; Lacher T (2020) Integrating Agriculture and Ecosystems to Find Suitable Adaptations to Climate Change. Climate 8:10.
63. Thayer A; Vargas A; Castellanos A; Lafon C; McCarl B; Roelke D; Winemiller K; Lacher T (2020) Integrating Agriculture and Ecosystems to Find Suitable Adaptations to Climate Change. Climate 8:10.
64. Vargas RI; Shelly TE; Leblanc L, Pinero JC (2010) Recent Advabces in Methyl Eugenol and Cue-Lure Technologies for Fruit Fly Detection (Monitoring and Control in Hawaii: Elsevier) 83:575-595.
65. Xue Y; Wang C; Miao C (2020) Research on financial assets transaction prediction model based on LSTM neural network, Neural Computing and Applications, vol. 1
66. Zecca F (2019) The Use of Internet of Things for the Sustainability of the Agricultural Sector: The Case of Climate Smart Agriculture. Int. J. Civ. Eng. Technol 10:494–501.
67. Zecca F (2019) The Use of Internet of Things for the Sustainability of the Agricultural Sector: The Case of Climate Smart Agriculture. Int. J. Civ. Eng. Technol 10:494–501.