Asian Journal of Agricultural Extension, Economics & Sociology

Volume 41, Issue 11, Page 334-345, 2023; Article no.AJAEES.103172
ISSN: 2320-7027

Weather Based Prediction Models for Disease and Pest Using Machine Learning: A Review

Dayana David [a++*]

[a] *Department of Agricultural Statistics, College of Agriculture, Kerala Agricultural University, Vellannikkara, Kerala, India.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJAEES/2023/v41i112290

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/103172

Review Article

Received: 16/05/2023
Accepted: 21/07/2023
Published: 07/12/2023

ABSTRACT

Critical review of weather based prediction models of disease and pest attack on crops using machine learning (ML) algorithms are performed in the study. Since suitable weather conditions are the accelerators for the growth and spreading of disease or pest, the prediction models based on weather condition achieves high degree of accuracy. Due to the advancement of technology ML algorithms remarks successful application in prediction of diseases and pest on crops. The scope of the review work lies in the fact that the accurate forewarning system helps for the timely application of pest and disease management techniques which have greater significance in controlling and solving the damages due to diseases or pest infestation in plants. Stages in prediction models are analysed and the applied techniques are compared in detail in this review. Consequently, importance of weather parameters in perdition and, performance metrics used for evaluating the prediction models are compared and presented. The review presents the detailed discussion on machine learning algorithms used in the prediction models. The review reveals that new models with high degree of accuracy need to be developed for the prediction of diseases or pest outbreak of various crops.

[++] *Assistant Professor*
**Corresponding author: E-mail: dayana.d@kau.in;*

Keywords: Weather parameters; machine learning; prediction model; evaluation metric.

ABBREVIATIONS

Area Under the ROC Curve	*ROC-AUC*	*Mann-Whitney-Wilcoxon*	*MWW*
Area Under the ROC Curve	*ROC-AUC*	*Mann-Whitney-Wilcoxon*	*MWW*
Artificial Intelligence	*AI*	*Maximum Temperature*	*MaxTemp*
Artificial Neural Network	*ANN*	*Mean Biased Error*	*MBE*
Autoregressive Integrated Moving Average	*ARIMA*	*Minimum Temperature*	*MinTemp*
Balanced Accuracy	*BA*	*Modelling Efficiency*	*EF*
Balanced Accuracy	*BA*	*Moving-Average*	*MA*
Bayesian Learning Network	*BLN*	*Multilayer Perceptron*	*MLR*
Coefficient of Determination	*R2*	*Multiple linear regression*	*MLR*
Compact classification tree	*CCT*	*Neural Network*	*NN*
Efficient Neural Network	*ENET*	*Normalized Root Mean Square*	*nRMSE*
Ensemble Feature Ranking	*EFR*	*Pace Regression*	*PR*
Extreme Learning Machine	*ELM*	*Rainfall*	*RF*
Feature Elimination Algorithm	*FEA*	*Rainy Days*	*RD*
Geometric Mean	*GM*	*Random Forest*	*RFT*
Geometric Mean	*GM*	*Recurrent Neural Networks*	*RNN*
Gradient Boosting	*GB*	*Relative Humidity*	*RH*
Gradient Boosting	*GB*	*Rice Blast Disease*	*RBD*
Importance Sampling	*IMPS*	*Root Mean Square Error*	*RMSE*
Integer-valued Generalized Autoregressive	*INGARCH*	*SunShine Hours*	*SSH*
K Neighbours Regressor	*KNN*	*Support Vector Machine*	*SVM*
Least Absolute Shrinkage and Selection Operator	*LASSO*	*Support Vector Regression*	*SVR*
Linear Regression	*LR*	*Synthetic Minority Over-sampling Technique*	*SMOTE*
Long Short-Term Memory	*LSTM*	*Temperature*	*Temp*
Machine Learning	*ML*	*Water Accounting Rice Model*	*WARM*

1. INTRODUCTION

Sustainable growth in agriculture helps to achieve the global development goal of poverty eradication. Agriculture development raises the income of the poorest people as they depend more on agriculture than any other sector. However, over the past few decades agriculture sector faces various problems especially yield loss caused by the infestation of diseases and pest. Farmer's expectation of the profitable yield is challenged by the uncertainties created by the disease outbreak and pest attack of the crops. According to the study [1], every year 30% of the yield loses due to upsurge of diseases and pest in farmlands. Application of pesticides in large quantity leads to crop deportation and endangerment to humanity. The widely adopted practice of naked eye observation of crops by farmers and experts are unfeasible and impractical in large farms. With the advancement of technology automatic prediction models would have a significant role in timely application of control measures to decrease the usage of chemical treatment and reducing the damage caused by the outbreak of diseases and pests.

The development of diseases or pest are not cause serious threat until favourable weather conditions occur in the field. Since, weather condition had a major role in the outbreak of diseases and in the life cycle duration or spreading of pests, weather parameters are the most useful information for building a forewarning system for diseases or pest attack in crop field. Literature review shows that weather based predictive models, utilizes the influence of weather condition in occurrences of diseases or pest, results in high degree of predictive accuracy [2,3,4]. Recently, farmers adopt modern agricultural practices like AI and ML for prediction of development of disease and pest attacks. ML algorithms along with high performance computing are capable to analyse huge volumes of data and overcome difficulty in the representation of non-linear agricultural structures and ensure high predictive accuracy. In ML technique, algorithm uses historical dataset to train the system for acquiring knowledge and thus predict the outbreak of disease or pest. The architecture of machine learning prediction model is shown in Fig. 1.

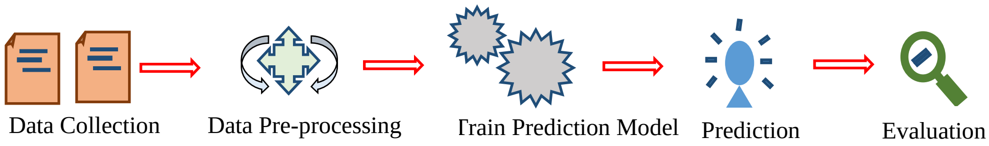

Fig. 1. Architecture of ML prediction model

In literature several informative reviews [5,6] for ML based diseases or pest attack prediction and crop yield prediction has been published. However, analysis of prediction accuracy based on weather parameters and influence of weather parameters in prediction models were not discussed in those reviews. As weather parameters are the best predictors for disease or pest attack, the readers who are interested in disease or pest attack prediction based on weather parameter may not be benefited from these articles. Therefore, there is a research gap for a comprehensive review article that give emphasize on weather based ML prediction algorithm for diseases or pest attack. This article provides a systematic review on the application of ML algorithm in disease or pest prediction with special emphasis on weather parameter based prediction techniques. The contribution of this review article is as follows: 1) Critical review of weather based ML prediction algorithms in literature. 2) Describe the significance of weather parameter in prediction 3) Discussion about the techniques for identifying best weather parameter predictor. 4) Detailed discussion of ML techniques used for prediction of disease or pest attack. 5) Comparative analysis of pre-processing techniques used for data cleaning and performance metrics used for evaluation of prediction techniques. To focus on recent techniques, the articles published in the last eight years in the peer-reviewed journals from scientific databases are considered for the review. Among 40 identified papers from the area, duplicates articles, irrelevant conference papers and studies not based on weather parameters are removed and the twenty articles which considered most relevant are reviewed. Out of twenty papers two were published in 2022 and 2017, seven in 2020, five in 2019 and, one paper from each 2021,2018, 2016 and 2015, are reviewed. The remainder of the paper organized as follows: Initially the selected articles in the literature are comprehensively reviewed. The different sources of weather data and comparison of pre-processing techniques are explained in section 2.1 and section 2.2 respectively. In the section 2.3, the influence of weather parameters and widely used weather parameters in prediction are discussed. The detailed explanation of ML algorithms used in prediction are described in section 2.4. In last section performance measures used for evaluating the accuracy of prediction models are compared.

2. ANALYSIS AND DISCUSSION

The overview of the research works selected for review are discussed in this section. A technique for predicting occurrence and non-occurrences of RBD using ANN and SVM classifiers was developed [2]. The models trained and tested with eight years of data from 2006 to 2013. Classifiers also compared with regression models and performance measures showed the supremacy of SVM classifiers than other models. A prediction model for potato late blight using SVR method, based on 13 weather parameters was designed [7]. The study compared the performance of SVR with NICS moving-average method, pace regression and LR method. The models analysed and compared with data from 1976 to 1985 and from 2009 to 2012. The evaluation results showed that SVR performed superior than other methods.

The scientists claimed that it was the first time in research LSTM used for prediction of disease or pest outbreak of crops [8]. The model trained and tested with 14 years of data for the period of 2003 to 2016 of four different locations and model tested with 17 different cultivars. The performance metric showed that model predict with high accuracy. Researchers suggested a prediction technique for Ganoderma Basal stem rot of oil palm using ANN based on specific spectral signatures and weather data [3]. The method shows 100% accuracy in prediction and identified that high relative humidity increases the possibility of disease. Scientists compared different mathematics-based techniques and ML based techniques for prediction of diseases on cherry fruit using eight years of data from the period of 2009 to 2016 [4]. The author uses the prediction models like Linear discriminant analysis, Quadratic discriminant analysis, Pseudo linear discriminant analysis and CCT to forecast the most important diseases like monilinia laxa and coccomyces hiemalis of cherry fruit. According to the study CCT outperforms the other techniques.

A prediction model for RDB, which utilizes Auto-Sklearn and neural network algorithms as the classification model using five years of data ranging from 2014 to 2018 was designed [9]. They also developed a FEA to identify the best weather predictor variable. Scientists also suggested that using weather parameters obtained from sensors in farmland along with weather data obtained from weather stations increases the prediction accuracy. The evaluation results showed that FEA along with Auto-Sklearn classifiers performed well in prediction. Scientists compared different prediction algorithm like MLR, KNN, RFT, and ANN for predicting incidence of coffee rust, cercospora, coffee miner, and coffee borer [10]. The experimental results revealed the RFT model gave more accurate results than other methods.

Researchers compared two rule based models and two machine learning models for the prediction of RBD [11]. According to the authors this was the first time performance of process models like Yoshino and WARM and ML models like M5Rules and RNN were compared for plant disease management. The data set consists of data from two different regions for a year period to build unique train and test dataset. The performance comparisons showed that all four approaches perform well in prediction of RBD and also indicates that if high quality dataset was available the ML models outperform all other models. Scientists developed virtual sensors for apple scab disease using RFT algorithm [12]. The behaviour of the sensors modelled using weather parameters and performance measures showed that the prediction model was a good replacement for the real hardware.

A different approach proposed for the prediction of Citrus Gummosis disease based on weather, soil and plant related information [13]. The approach developed different prediction models based on these factors separately and also a model incorporating all three factors together using SVM and MLR technique. The study utilized two years of data from the period of 2014 to 2016. The experimental results showed in both types SVR outperforms MLR. An experiment conducted to predict the Septoria tritici blotch on wheat using ANN based on weather parameters [14]. The study reveals that ANN models performs well in prediction of disease.

Scientists successfully developed a prediction model using RFT algorithm to forecast thrips disease of Mango Fruit Crop using past 20 years of data [15]. Researchers developed a different approach for prediction of Powdery mildew of Tomato on highly imbalanced dataset with 244 samples of different weather parameters [16]. The approach used non-iterative ELM algorithm and performance measures showed that the algorithm gave high accuracy in prediction. An approach developed for forecasting Late Blight on Potato using SVM classifier [17]. The approach also analysed the relationship between disease severity and weather parameters. The performance measures depicts the better performance of the model.

Researcher investigated a method for accurate prediction of Downy mildew diseases on Grape fruit using LASSO, GB, and RFT algorithm [18]. The dataset included 9 years of disease severity information along with weather parameters. The performance of algorithms were compared and identified that LASSO, RFT and GB algorithms showed better performance than generalized linear models algorithms. A different model to predict the Powdery mildew of grape fruits for three susceptible cultivars using probabilistic BLN based on weather, pathogen and host factors was developed [19]. The 12 years data for the period ranging from 2000- 2011 were used for the study. The author claimed that model can predict the disease by 16 days ahead with high accuracy. Scientists proposed a method for the prediction of Sterility mosaic disease on Pigeon pea using SVR, ANN and their combination with ARIMA [20]. The study carried out by data of kharif seasons for the period of 2011 to 2016. Performance measures showed that hybrid model outperform other models. A method suggested to build a prediction model for forecasting the attack of gall midge on rice crop using 6 years of data for period of 2013 to 2018 from four different locations [21]. The authors compared the performance of INGARCH, ANN, and SVR models for predicting the gall midge population and found that ANN outperforms all other methods. The performance of ENET, LASSO, Ridge, and ANN for predicting epidemic characteristics of Alternaria blight of mustard compared using data of 14 years from 2006 to 2022 [22]. According the evaluation metrics ANN performs better than other models. Scientist suggested an EFR method to identify the best weather predictor variable and also compared performance of five different classification algorithm like Multilayer perceptron, SVM, Naive Bayes, Decision Tree, and KNN in prediction of outbreak of RBD [23]. The study used the data of 96 instance of rice blast during the 2013-2019 period. The performance of classification algorithms was satisfactory and EFR method capable enough to identify the best weather parameter for prediction. The summary of review articles techniques used and weather parameters used are given in Table 1.

2.1 Data Sources

Disease outbreak or pest incidence severity data and weather parameters were required for building the prediction model. The scientists collects data for the research works from various sources like weather stations, field sensors, third party organizations and web service. The collection of reliable weather data and disease severity or pest incidence data is really challenging. Literature review shows that the data used for the studies range from years to days in different period. As the prediction models are based on ML algorithms the amount of data used for training and testing have a significant influence on performance of the algorithm. The large amount of data for training and testing gave higher accuracy in prediction [24]. Analysis showed that 40% of research works depend on weather stations for weather data and 35% collected data from sensors installed in their own fields and remaining depends on other sources. The experimental results showed that data collected from field sensors shows more accurate prediction than using the data from weather stations [5].

The disease severity data were collected from heterogeneous sources like the third party organizations [22,23], conducting surveys [15], and directly from fields with agreement of the farmers [18]. The traditional method of installation of light trap was used for the collection of pest incidence data. The research studies used the daily or weekly or monthly cumulative count of pest caught form the filed [21].

2.2 Pre-Processing

The collected data from heterogeneous sources are in different format which includes data in different numerical ranges, incomplete and missing values and, outliers. As the prediction models were sensitive to input parameters, the raw data obtained from the sources has to be transformed by applying various statistics techniques like normalization, scaling, dimensionality reduction.

To understand the relationship between weather parameter and manifestation of diseases or pest the reviewed experiments either utilizes existing or derived feature selection algorithms [2,9,10, 13,17] or wrapper method i.e performance evaluation of ML technique used for the prediction [7,18,22]. Literature review showed that Pearson's product moment correlation analysis was the most widely used technique for identifying the inter relationship between weather parameters and disease severity or first date of pest incidence [20,21,22]. An EFR algorithm was introduced for identifying the weather parameters which would have more influence on the outbreak of disease and pest [23].

2.3 Impact of Weather Parameters on Prediction

The major challenge in agriculture sector was the reduction in crop productivity and thus income of farmers due to the disease or pest outbreak. A prediction system mitigates these problem by giving an early forewarning to farmers, which help them to implement effective pest or disease management strategy to reduce the risk. As the nonlinear nature of sequences of ecological processes, accurate prediction of disease outbreak or pest attack is very much challenging task. The weather conditions, soil peculiarities and plant characteristics influences the outbreak of diseases or pest. The presence of pathogens at canopy or field doesn't make a sever outbreak until the weather condition favours its development and spreading. The weather conditions capable to modify the rate of reactions in a process without being part of a system [22]. Therefore, weather condition exerts significant influence on the occurrence of diseases or pest than other factors [25].

The weather parameters influence the life cycle of pathogens and accelerate to the development of pest at the suitable climatic condition. The influence of weather on pathogen or pest and interaction between these two leads the rapid growth of diseases [1]. Hence, weather parameters are considered the best predictors used for the accurate perdition of disease or pest attack outbreak. In literature, the experimental results showed that the weather parameters collected from fields were more accurate to predict the outbreak or attack than the data collected form weather stations [9,17].

The Fig.2 shows the most widely used weather parameters in disease or pest attack prediction and their distribution in research works. The literature review shows that a wide range of weather parameters have been utilized to predict the outbreak of disease or pest attack. The parameters like maximum temperature, average temperature, minimum temperature, temperature difference, temperature range, relative humidity, maximum relative humidity, minimum relative humidity, humidity difference, precipitation, amount of evaporation, solar radiation, sunshine hours, visibility sunny days, wind velocity, average wind speed, amount of cloud cover, rainfall, rainy days, and amount of vapour collected daily, weekly or monthly have been used by the prediction models. According to the literature review, the most significant weather parameter that contribute to prediction was temperature difference [17,18,20-23]. The parameter maximum temperature also equivalently important in prediction model of various crops [1,7,20,22]. Relative humidity [2,9,2,23] and difference in relative humidity [2,9,20,22] also plays key role in prediction. Other than these parameters sunshine hours [22,23], wind speed [17,23], and precipitation [18] were also identified as parameters that have significant influence on prediction of diseases or pests of various crops.

2.4 Machine Learning Models

The goal of using ML algorithms in disease or pest prediction is to give a forewarning to framers and thus avoid yield loss and limit the excess usage of chemical fertilizers. The most widely used algorithm is ANN, and second most is SVR. The brain inspired ANN model consists of many layers with neurons and it can model complex relationship with hidden layers. ANN models worked based on Gradient Descent algorithm which minimize the error function by iterative method. The widely used SVM model capable to build regression or time series model in a transformed high dimensional feature space [26]. The SVM techniques used in many prediction algorithms matches the nonlinear problems with linear problems of high dimensional feature space [27]. The performance evaluation of prediction models of RBD showed that SVM outperforms ANN and regression models [2].

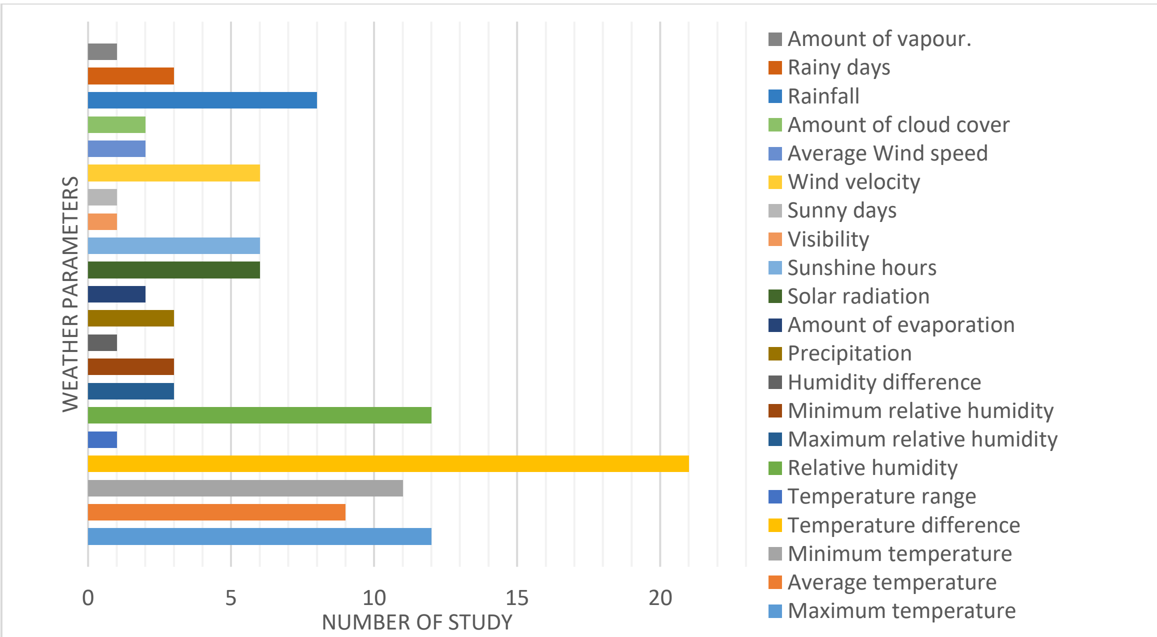

Fig. 2. Most widely used weather parameters

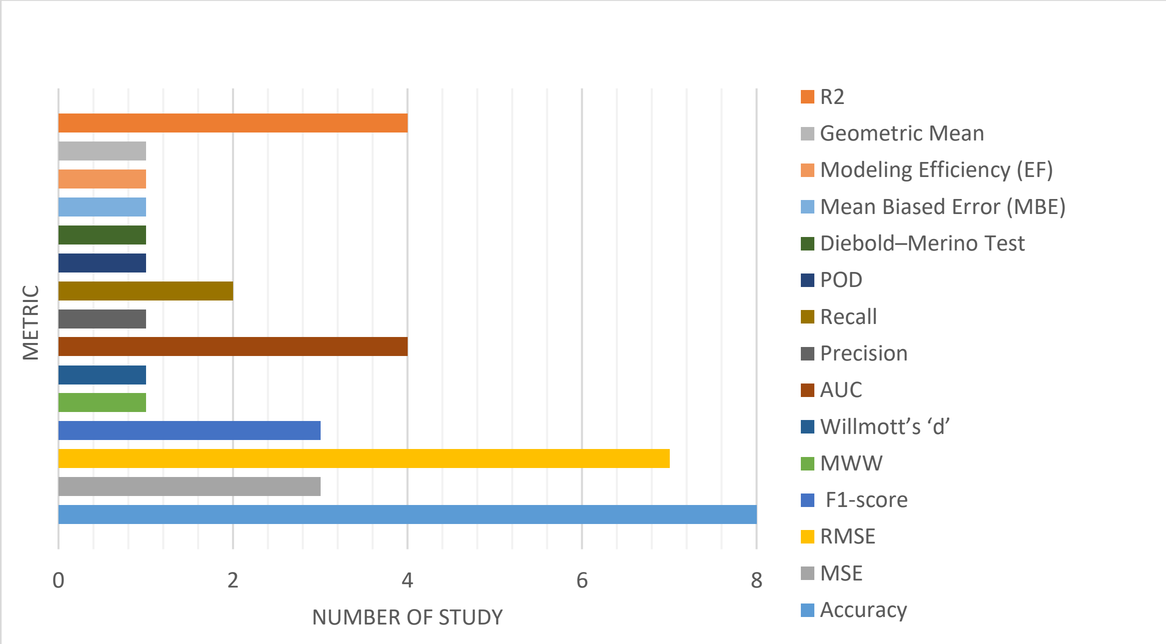

Fig. 3. Most widely used metrics for evaluation

Table 1. Comparison of the prediction techniques

Ref	Year	Crop	Diseases/pest	Framework	Weather Parameters	Identified Predictors	Performance Measures	Results	Inference
2	2015	Rice	RBD	SVM and ANN, Regression Model	Evaporation, MaxTemp, MinTemp, Average air speed, Average Precipitation Count, Average Radiant Energy emitted by the sun, Average brightness, Average amount of vapour.	RF, MinTemp, MaxTemp,RH	Accuracy, R2 and MSE	SVM - Accuracy : 96.44% SVM - R2 : 0.7758 SVM - MSE : 0.2374 ANN - Accuracy : 92.32%	Performance metric shows the supremacy of the ANN
7	2016	potato	Late blight	MA,SVR,PR,LR	Average Temp, Peak Temp, MinTemp, Temp Range, Min grass Temp, RH, Lowest RH, Amount of Evaporation, Solar Radiation, Sunshine, Average wind velocity, Ground-surface Temp, Amount of cloud cover	Temp, Lowest Grass Temp, Peak Temp, Lowest Temp	Accuracy	MA - Accuracy : 42.9% SVR - Accuracy : 64.3% PR - Accuracy : 42.9% LR - Accuracy : 35.7%	SLR outperforms other prediction techniques
8	2017	Rice	RBD	LSTM and RNN	Air Temp, RH, SSH		Accuracy and F1-	LSTM-Accuracy - 79.4%	LSTM performs better

Ref	Year	Crop	Diseases/pest	Framework	Weather Parameters	Identified Predictors	Performance Measures	Results	Inference
							score	LSTM-F1-score - 17.1%	in prediction
3	2017	Oil Palms	Ganoderma Basal stem rot of oil palm	ANN	Temp, RH, RF, Speed Wind, Solar Radiation	RH	Accuracy	Accuracy : 100%	ANN gives the remarkable prediction
4	2018	Cherry	Monilinia laxa and coccomyces hiemalis	CCT	MnTemp,MaxTemp, Avreage Temp, RH, RF, Wind speed		Accuracy	Accuracy - 95.8%	CCT achieves high accuracy in prediction
9	2019	Rice	RBD	Auto-Sklearn and NN algorithms	Atmospheric Climate Raw Data - MaxTemp, MinTemp, Temp Difference, Maximum Humidity, Minimum Humidity, Humidity Difference Micro Climate Raw Data - MaxTemp, MinTemp, Temp Difference, Max. Hum, Min. Hum, Hum. Diff, RH.	RH, Temp	Accuracy and Recall	NN : Accuracy - 66.6% , Recall - 83.8 Auto-Sklearn : Accuracy - 67.4% , Recall - 74.5%,	NN reveals the supremacy
10	2019	Coffee	Coffee rust, cercospora, coffee miner, and coffee borer	RLM, KNN, RFT and MLP	MiniTemp, MaxiTemp, Rainfall, Average RH,	MaxTemp,RH	Willmott's 'd', RMSE , and R2	KNN - RMSE : 19.59, Willmott's 'd - 0.823, R2 - 0.667 MLP - RMSE - 24.15, Willmott's 'd - 0.847, R2 - 0.507 RFT - RMSE - 8.51, Willmott's 'd - 0.939, R2 - 0.736 RLM - RMSE - 15.48, Willmott's 'd - Not Sepcified, R2 -0.655	Superior performance of RFT has shown
11	2019	Rice	RBD	WARM, M5Rules and RNN	MinTemp, MaxTemp, RH, Leaf Wetness		R, R2, %MAE, AUC	M5Rules - r : 0.59, r2 : 0.39 , %MAE : 0.63, AUC : 0.80 RNN - r : 0.70, r2 : 0.50 , %MAE : 0.75, AUC : 0.76 YOSHINO - r : 0.55, r2 : 0.35, %MAE : 0.49, AUC : 0.74 WARM - r : 0.59, r2 : 0.40 , %MAE : 0.82, AUC : 0.77	M5Rules and RNN gives the better accuracy
12	2019	Apple	Apple scab	RFT	Temp, RH and Wind Speed		MWW test		Virtual Sensors performs well
13	2019	Citrus	Citrus Gummosis	SVR and MLR	Temp, Humidity, Rainfall	Temp, RH	RMSE, R2	SVR - RMSE :0.225,MLR - RMSE : 0.272, SVR - R2 :0.906,MLR - R2 : 0.851	SVR outperforms other models
14	2020	Wheat	Septoria Leaf Blotch and Stripe Rust	ANN	Leaf wetness, Temp, RH, Precipitation		POD	POD - 0.9 to 1	ANN give better prediction
15	2020	Mango	Thrips	RFT	MinTemp, MaxTemp, Humidity		MAE,MSE and RMSE.	MAE : 0.046, MSE : 0.006 RMSE : 0.081	RFT gives the better result
16	2020	Tomato	Powdery mildew	ELM	Global radiations, Wind Speed, Humidity, Temp, Leaf Wetness		AUC and CA	Accuracy : 89.19% , AUC : 88.57%	ELM gives high accuracy in prediction
17	2020	Potato	Potato Late Blight	SVM	Temp, Humidity, RF, Speed Wind, Solar Radiation	RH, Temp, Wind Speed	Accuracy, Precision, Recall, and F1-score	Accuracy : 82% , Precision : 72%,Recall : 82% F1-score : 77%	SVM shows the outstanding performance
18	2020	Grape	Downy mildew	LASSO and GB and RFTs	Temp, Precipitation	Precipitation	AUC	GB - AUC : 0.86 LASSO - AUC : 0.80 RRFTs - AUC : 0.78 Generalized Linear Models - AUC : 0.76	GB gives high accuracy
19	2020	Grape	Powdery mildew	BLN	Temp, Wind Velocity, RH, Rain Intensity,		MAE and RMSE.	MAE - 0.55, RMSE - 0.86	BLN shows the better

Ref	Year	Crop	Diseases/pest	Framework	Weather Parameters	Identified Predictors	Performance Measures	Results	Inference
					Precipitation				performance
20	2020	Pigeonpea	Sterility mosaic disease	SVR, ANN,SVR-ARIMA,ANN-ARIMA	MinTemp, MaxTemp,Morning RH, Evening RH, SSH, Wind Velocity, RF	MinTemp, MaxTemp, Morning RH, Evening RH,	RMSE	SVR - RMSE : 0.12 ANN - RMSE : 0.22 ANN-ARIMA - RMSE : 0.26 SVR-ARIMA - RMSE : 0.34	SVR-ARIMA models gives the high accuracy
21	2021	Rice	Gall midge	INGARCH, ANN and SVR	MaxTemp, MinTemp, RF, Morning RH, Evening RH, SSH	Morning RH, Evening RH, RF and SSH	MSE and RMSE	INGARCH - MSE: 1.9 , RMSE : 1.38 ANN - MSE: 0.42, RMSE : 0.64 SVR - MSE: 12.4 , RMSE : 3.52	ANN shows the supremacy
22	2022	Mustard	Alternaria blight	ENET, LASSO, Ridge, and ANN	MaxTemp,MinTemp, RH, SSH, RF, and Potential Evapotranspiration	Temp, RH, SSH	R2,RMSE,nRMSE, MBE and EF	ENET - R2 : 0.90 ,RMSE : 1,nRMSE : 1.59, MBE : 0 , EF : 62.33 Ridge - R2 : 0.92 ,RMSE : 9.03,nRMSE : 16.44, MBE : 0 , EF : 64.35 LASSO - R2 : 1.00, RMSE : 10.15 ,nRMSE : 19.02 , MBE : 0 , EF : 52.07 ANN - R2 : 1.00, RMSE : 2.07,nRMSE : 3.01, MBE : 0.18, EF : 75.80	ANN model gives the excellent result
23	2022	Rice	RBD	MLP,SVM,Naïve bayes, KNN	MaxTemp, MinTemp, AvgTemp,RD, Average Wind, Pressure, Cloud, Humidity, UV, SSH, SunDays, Visibility	Average Visibility, RF, SSH, Wind Speed, RD	F1 Score, BA, GM and ROC-AUC	MLP - F1 Score : 0.65+/-0.09, BA : 0.85+/-0.05 , GM : 0.85+/-0.05, ROC-AUC : 0.90+/-0.02 SVM - F1 Score : 0.54+/-0.04, BA : 0.82+/-0.03 , GM : 0.82+/-0.03, ROC-AUC : 0.88+/-0.03 NB - F1 Score : 0.36+/-0.03 , BA : 0.67+/-0.03 , GM : 0.66+/-0.03 , ROC-AUC : 0.76+/-0.05 DT - F1 Score : 0.61+/-0.05 , BA : 0.81+/-0.02, GM : 0.81+/-0.02, ROC-AUC : 0.86+/-0.05 KNN - F1 Score : 0.61+/-0.05, BA : 0.83+/-0.03, GM : 0.83+/-0.03 , ROC-AUC : 0.87+/-0.04	MLP perform very well

The ELAs like RFT and GB were also used as prediction models. The RFT algorithm builds decision trees with high variance and prediction was done by averaging the trees [28]. In GB algorithm successive trees were built where each tree predicts the residuals of previous one. RNN [8,11] also follows the architecture of ANN which capable to capture the temporal relationships. BLN, where structural learning and inference learning, were possible, represents the complex relationship between weather parameters and diseases outbreak [19]. Inference learning uses multiple learning algorithms to link random variables within a network into a directed acyclic graph. KNN is the classification technique based on distance measure and gave more weights to close neighbours. Minkowski, Euclidean, and Manhattan, were the various methods used in KNN to calculate the distance. CCT used the rules inferred from the training set for prediction of outbreak.

In literature the performance of ML based prediction model was compared with each other and with regression technique and, identified that ML based prediction models outperforms other models. SVM [2], M5Rules and RNN [11] models showed the better predictive power in predicting RBD. RFT technique was more accurate in prediction of Coffea arabica pests [10]. In prediction of Grape downy mildew disease of grapevines, GB algorithm showed the highest predictive accuracy [18]. ANN shows the high accuracy in prediction of gall midge population of rice [21] and Alternaria blight disease of Mustard [22].

2.5 Performance Evaluation Metric Comparison

The efficiency of the prediction models was evaluated by the performance metric. There is various performance metric exist in literature for evaluating the models. In literature MAE, MSE, RMSE, R-squared, F1 Score, GM, ROC-AUC, Accuracy, Precision, Recall, MWW test, Willmott's 'd' etc. were employed. As shown on Fig. 3 accuracy is the most widely used metric followed by RMSE and AUC. The evaluation metric Accuracy measures out of total data elements how many of them classified correctly [8]. RMSE measures the spread of data elements from the regression line [20]. MAE is the average of the difference between the predicted data elements and actual data elements [29]. AUC measures the area under famous ROC curve which represents the degree of separability [16]. Determination coefficient (R-squared) is the square of the correlation coefficient measure. F1 score assess the predictive skill of the model by combining score of precision and recall [8].

3. CONCLUSION

In this paper, we discussed twenty recent research works of weather based prediction models for various crops. Majority of the review works belongs to the development of prediction models for RBD in rice. According to the study, research works used data from different sources and geographical locations. This review discussed the pre-processing stage, impact of weather parameters in prediction models, ML models used in prediction, and performance metric of prediction models in detail. Analysis showed that Pearson Correlation Coefficient was the most prominently used pre-processing technique for identifying the influence of weather parameters in prediction. The majority of the research works for prediction of pest or disease incidence used ANN as a prediction model and experiment results shows its supremacy in prediction. The prediction power of the models investigated by performance metric and this review showed that accuracy is the most widely used metric for determining the performance of a model. As research works used data from different sources and using different feature sets on variety of crops there is no specific result can be made from the analysis. The review woks will enhance the scope of research the works in the prediction of pest or disease outbreak.

4. FUTURE SCOPE

In present study ML algorithms used for pest or disease prediction based on weather parameters were discussed. However, influence of weather parameters in prediction are investigated and summarized, other important factors are not addressed in the study. The influence of soil characteristics and plant related features in outbreak of diseases and pest can be analysed in the future work.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Dhaliwal GS, Jindal V, Mohindru B. Crop losses due to insect pests: Global

and Indian scenario. Indian Journal Entomology. 2015;77(2):165–168.

2. Alvin R. Malicdem, Proceso fernandez, rice blast disease forecasting for northern Philippine, Wseas Transactions On Information Science and Applications. 2015;12:120-129.
3. Parisa Ahmadi, Farrah Melissa Muharam, Khairulmazmi Ahmad, Shattri Mansor, Idris Abu Seman. Early detection of ganoderma basal stem rot of oil palms using artificial neural network spectral analysis. Plant Disease. 2017;101(6): 1009-1016.
4. Milos Ilic, Sinisa Ilic, Srdjan Jovic, Stefan Panic. Early cherry fruit pathogen disease detection based on data mining Predictio. Computers and Electronics in Agriculture. 2018;150:418–425.
5. Gianni Fenu, Francesca Maridina Malloci. Forecasting plant and crop disease: An explorative study on current algorithms. Big data and cognitive computing. MDPI. 2021;5(1):2.
6. Thomas van Klompenburga, Ayalew Kassahuna, Cagatay Catalb. Crop yield prediction using machine learning: A systematic literature review, Computers and Electronics in Agriculture. 2020; 177(10):105709.
7. Gu YH, Yoo SJ, Park CJ, Kim YH, Park SK, Kim JS, Lim JH. BLITE-SVR: New forecasting model for late blight on potato using support-vector regression, Computers and Electronics in Agriculture. 2016;130:169–176.
8. Yangseon Kim, Jae-Hwan Roh, Ha Young Kim. Early forecasting of rice blast disease using long short-term memory recurrent neural Networks, 7 Sustainability. MDPI. 2020;10(1):1-20.
9. Jia-You Hsieh, Wei Huang, Hsin-Tieh Yang, Chia-Chieh Lin, Yo-Chung Fan, Huan Che. Building the rice blast disease prediction model based on machine learning and neural Networks, Technical Report. EasyChair: Manchester, UK; 2019.
10. Lucas Eduardo de Oliveira Aparecido, Glauco de Souza Rolim, Jose Reinaldo da Silva Cabral De Moraes, Cicero Teixeira Silva Costa, Paulo Sergio de Souza. Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. International Journal of Biometeorology. 2020;64(4):671-688.
11. Nettleton DF, Katsantonis D, Kalaitzidis A, et al. Predicting rice blast disease: Machine learning versus process-based models. BMC Bioinformatics. 2019;20(1): 514.
12. Mariusz Wrzesień, Waldemar Treder, Krzysztof Klamkowski, Witold R. Rudnicki. Prediction of the apple scab using machine learning and simple weather stations. Computers and Electronics in Agriculture. 2019;161:252-259.
13. Mrunalini R. Badnakhe , Surya S. Durbha, Adinarayana Jagarlapudi, Rajendra M. Gade. Evaluation of citrus gummosis disease dynamics and predictions with weather and inversion based leaf optical mode. Computers and Electronics in Agriculture. 2018;155;130-141.
14. Moussa El Jarroudi, Rachid Lahlali, Haifa El Jarroudi, Bernard Tychon, Alexandre Belleflamme, Jürgen Junk,Antoine Denis, Mustapha El Jarroudi, Louis Kouadio. Employing weather-based disease and machine learning techniques for optimal control of septoria leaf blotch and stripe rust in wheat. Advances in Intelligent Systems and Computing, Springer. 2020; 1103:157–165.
15. Jawade PB, Dattatray Chaugule, Devashri Patil, Hemendra Shinde. Disease prediction of mango crop using machine learning and iot. advances in decision sciences, image processing. Security and Computer Vision. 2020;254-260.
16. Anshul Bhatia, Anuradha Chug, Amit Prakash Singh. Application of extreme learning machine in plant disease prediction for highly imbalanced dataset. 2020;23(6);1059-1068.
17. Gianni Fenu, Francesca Maridina Malloci, Artificial intelligence technique in crop disease forecasting: A case study on potato late blight prediction, proceedings of 12th KES International Conference on Intelligent Decision Technologies. 2020; 79–89.
18. Chen M, Brun F, Raynal M, Makowski D. Forecasting severe grape downy mildew attacks using machine learning. PLoS One. 2020;15(3).
19. Lu W, Newlands NK, Carisse O, Atkinson DE, AJ Cannon. Disease risk forecasting with Bayesian learning networks: Application to grape powdery mildew (*Erysiphe necator*) in vineyards. Agronomy, MDPI. 2020;10(5).

20. Ranjit Kumar Paul, Vennila S, Satish Kumar Yadav, Bhat MN, Kumar M, Chandra P, Paul AK. Prabhakar M. Weather based forecasting of sterility mosaic disease in pigeonpea (*Cajanu cajan*) using machine learning techniques and hybrid models. Indian Journal of Agricultural Sciences. 2020;9(10):1952–8.
21. Santosha Rathod, Sridhar Yerram, Prawin Arya, Gururaj Katti, Jhansi Rani,Ayyagari Phani Padmakumari, Nethi Somasekhar, Chintalapati Padmavathi, Gabrijel Ondrasek, Srinivasan Amudan, Seetalam Malathi, Nalla Mallikarjuna Rao, Kolandhaivelu Karthikeyan, Nemichand Mandawi, Pitchiahpillai Muthuraman, Raman Meenakshi Sundaram. Climate-based modeling and prediction of rice gall midge populations using count time series and machine learning approaches. Agronomy. MDPI. 2022;12(1).
22. Manjari Singh, Parul Setiya, Anand Kumar Tewari. Ajeet Singh Nain. Comparison of machine learning and regression approaches to forecasting Alternaria blight epidemic of Indian mustard. The Pharma Innovation Journal. 2022;11(6):2974-2984.
23. Kittakorn Sriwanna, Weather-based rice blast disease forecasting, Computers and Electronics in Agriculture. 2022;193: 106685.
24. Dauphin YN, Bengio Y. Big neural networks waste capacity. 2013; arXiv:1301.3583.
25. Eva Kocmanko, Miroslav Trnka, Jan Juroch, Martin DubroVsky. Impact of climate change on the occurrence and activity of harmful organisms, Plant Protection Science. 2009; 45:S48-S52.
26. Vapnik VN. The Nature of statistical learning theory. Springer: New York, NY, USA; 1995. Available:online:https://link.springer.com/book/10.1007%2F978-1-4757-2440-0 Accessed on 11 October 2021.
27. Hearst MA, Schölkopf B, Dumais S, Osuna E, Platt J. Trends and controversies-support vector machines. IEEE Intell. Syst. 1998;13:18–28.
28. Breiman L. Random Forests. Mach Learn. 2001;45(1):5–32.
29. Mumtaz Ali, Ravinesh C Deo, Nathan J Downs, Tek Maraseni. Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting. Computers and Electronics in Agriculture. 2018;152:149–165.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/103172