Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 25 (4) : 583-588 (December - 2023)
https://doi.org/10.54386/jam.v25i4.2272
https://journal.agrimetassociation.org/index.php/jam

Research Paper

Prediction of potato late blight disease severity based on weather variables using statistical and machine learning models: A case study from West Bengal

VAIDHEKI M.[1], DEB SANKAR GUPTA[1], PRADIP BASAK[1*], MANOJ KANTI DEBNATH[1], SATYAJIT HEMBRAM[2] and AJITH S.[1]

[1]Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, West Bengal
[2]Department of Plant Pathology, Uttar Banga Krishi Viswavidyalaya, West Bengal
**Corresponding author E-mail: pradip@ubkv.ac.in*

ABSTRACT

Late blight is one of the most devastating diseases on potato the world over, including West Bengal, India. The economic and yield losses from outbreaks of potato late blight can be huge. In this article, application of statistical models such as autoregressive integrated moving average (ARIMA), autoregressive integrated moving average with exogenous variables (ARIMAX) in combination with machine learning models such as, neural network auto regression (NNAR), support vector regression (SVR) and classification and regression tree (CART) have been explored to predict the percentage disease index (PDI) of potato late blight in the northern part of West Bengal. Models were developed to predict PDI at 3- and 7-days interval using the weather variables *viz.*, rainfall, maximum and minimum temperature, maximum and minimum relative humidity, and dew point temperature. Among the developed models, CART to predict PDI at 7 days interval was found to be the best fitted model on the basis of least RMSE, MAE and MAPE. The results of CART showed that dew point temperature had a significant effect on PDI at 7 days interval and the incidence of potato late blight was high when dew point temperature was greater than 12 ^{0}C in the preceding week.

Key words: Potato late blight, ARIMA, ARIMAX, NNAR, SVR, Classification and regression tree (CART).

Potato is an important commercial crop in West Bengal. As per 3rd advanced estimate of 2021-22, area under potato in West Bengal was 447.45 thousand hectares which contributes to 20% of India's total area coverage and its production was 12403.13 thousand tons which contributes to 23% of India's total production (Anonymous, 2023). The northern part of West Bengal comprises eight districts where potatoes are grown in large quantity. The area is characterised by three agro-climatic conditions viz., Terai zone, Hill zone and Old alluvial zone. During 2019-20, area under potato in this region was 110.06 thousand hectares which contributed to 25% of state's area coverage and its production was 3588.81 thousand Tons which contributed to 29% of state's total production (Anonymous, 2020). Late blight is one of the major diseases of potato which is brought on by the oomycete *Phytophthora infestans* (Fry, 2008) and causes severe yield loss in this region. Decision-making in late blight disease management is challenging due to the complex nature of the disease development and its interaction with weather parameters. Potato late blight disease severity is highly associated with the weather parameters. Empirical models such as Cook's System (1947) and Hyre's System (1954) predict the late blight disease development using weather parameters such as minimum temperature, relative humidity. Wallin (1962) predicted the initial appearance and subsequent spread of late blight based on temperature and relative humidity.

Various mathematical, statistical and simulation models are used for timely and accurate forecasting of disease incidence which will eventually benefit the farmers in minimizing the losses by following proper management for the diseases. Madden (1992) used several nonlinear statistical models such as exponential, monomolecular, logistic and Gompertz, to describe the progress of potato late blight. Singh *et. al.,* (2012) explained that by adoption of disease forecast models, the frequency of fungicides application could be reduced up to 50% as compared with conventional, calendar-based schedule. Dar *et. al.,* (2021) applied multiple linear regression model using weather parameters to predict the potato late blight disease severity in Northern Himalayas of India and it was

Article info - DOI: *https://doi.org/10.54386/jam.v25i4.2272*
Received: 6 July 2023; Accepted: 16 October 2023; Published online : 30 November, 2023

found that maximum temperature, relative humidity and rainfall are highly associated with the disease development. In the Indo-Gangetic plains of West Bengal, the buildup of the potato late blight was found to be positively associated with the maximum and minimum temperatures as well as the maximum and minimum relative humidity, and it was found that the disease spreads practically exponentially from the time of its initial appearance (Dey *et. al.*, 2022). Since duration and severity of winter, and other meteorological factors in Terai zone is different from Indo-Gangetic plains of West Bengal, where these relationships were established, there is need to fit models for Terai zone, our study area separately. Similarly, the predictive models developed for other hilly areas cannot work in Terai zone.

Now-a-days due to advanced computational power, the machine learning models such as ANN, SVR etc. can be designed for an accurate and precise forecasting (Durgabai and Bhargavi, 2018). Sarkar *et. al.,* (2023) used ARIMA, ARIMAX and SVR models to predict the peak pest population incidences in jute crop based on weather variables in West Bengal. In the present investigation an attempt has been made to predict potato late blight disease severity in northern part of West Bengal using statistical and machine learning models with regional weather variables in order to achieve recommendations and efficiently resolve the issue.

MATERIALS AND METHODS

Description of Data

A field experiment was carried for six years from 2016-17 to 2021-22 *rabi* seasons at Instructional Farm, the Regional Research Station, Terai Zone, Uttar Banga Krishi Viswavidyalaya (UBKV), Cooch Behar, West Bengal. The variety viz. Kufri Jyoti were used for the experiment. The Kufri Jyoti variety of potato was planted on five different sowing dates at fortnightly interval starting from November to December in randomized block design with 4 replications and at spacing of 60 cm X 30 cm in a plot size of 3m X 3m (Hembram *et. al.*, 2018). Late blight disease incidence was measured in terms of PDI (Wheeler, 1969) and estimated as

$$PDI = \frac{Total\ sum\ of\ numerical\ ratings}{number\ of\ leaves\ observed \times maximum\ disease\ ratings} \times 100.$$

These data were taken from control fields, without spray operation from the tubers planted during first week of December. Initially, PDI data were recorded at unequal interval of days starting from the initial appearance of the disease. From these unequal interval data, PDI at 3- and 7-days interval were considered such that we could fit time series models which are based on the assumptions of equal interval data. In case of PDI at 3 days interval, total number of observations was 138 from 2016-17 to 2021-22, whereas for PDI at 7 days interval, it was 63. Therefore, PDI at 3- and 7-days interval have been predicted using weather variables. Daily data on weather variables *viz*., rainfall (RF in mm), maximum and minimum temperature (T_{max} & T_{min} in ^{0}C), maximum and minimum relative humidity (RH_{max} & RH_{min} in %), and dew point temperature (T_{dew} in ^{0}C) were collected from the records of Agromet observatory, Agrometeorological Field Unit (AMFU), Pundibari, UBKV for the year 2016-17 to 2021-22. The daily data of weather variables were converted into 3 days and 7 days interval weather data by taking the average value of preceding 3 days and 7 days weather variables from the respective dates of PDI data collection. In case of rainfall, total is taken instead of average.

Different statistical and machine learning models have been used to predict the incidence of disease using weather variables. For PDI data at 3 days interval, initial data for the years 2016-17 to 2020-21 which consists of 124 observations were used for model building and the remaining data for the year 2021-22 which consists of 14 observations were used for validation purpose maintaining the training to testing ratio of 90:10. Similarly, in case of PDI data at 7 days interval, initial data for the years 2016-17 to 2020-21 which consists of 57 observations were used for model building and the remaining data for the year 2021-22 which consists of 6 observations were used for validation purpose maintaining the training to testing ratio of 90:10. The analysis has been carried out using R software.

ARIMA model

ARIMA model (Box and Jenkins, 1976) is suitable for forecasting up to 3 to 5 advance time points. It uses the past values of the study variable to build the forecasting model which can be expressed as;

$$\phi_p(B)\nabla^d y_t = \theta_q(B)\varepsilon_t$$

where, B is the backshift operator such that,

$(B^p)\nabla^d y_t = \nabla^d y_{t-p}$, $\nabla^d y_{t-p} = (1-B)^d y_{t-p}$, $(B^q)\varepsilon_t = \varepsilon_{t-q}$ and $\varepsilon_t \sim N(0, \sigma^2)$, known as white noise errors. $\phi_p(B)$ and $\theta_q(B)$ are the AR and MA components of order p and q respectively, where d denotes the order of differencing.

ARIMAX model

ARIMAX model (Bierens, 1987), being an extension of ARIMA model, increases the explanatory nature of the model by incorporation of exogenous independent variables which have possible influence over the predicted values. A time series process $\{(y_t, x_t)\}$ having (k + 1) terms, where y_t and k values of x_t are real valued random variables, can be formulated as,

$$\nabla^d y_t = \phi_1 \nabla^d y_{t-1} + \ldots + \phi_p \nabla^d y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} + \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt}$$

where, ε_t 's are the errors, but here interpreting β is difficult. So, it is expressed as,

$$\nabla^d y_t = \beta_0 + \beta_1 \nabla^d x_{1t} + \ldots + \beta_k \nabla^d x_{kt} + \nabla^d \eta_t$$

$$\nabla^d \eta_t = \phi_1 \nabla^d \eta_{t-1} + \ldots + \phi_p \nabla^d \eta_{t-p} + z_t - \theta_1 z_{t-1} - \ldots - \theta_q z_{t-q}$$

where, η_t = eta at t and z_t = error.

Neural network auto regression (NNAR)

In NNAR model, the inputs are typically the past observations series and the output is the future value. The NNAR model performs the following nonlinear function mapping between the input and output

$$y_t = f(y_{t-1}, y_{t-2}, ..., y_{t-p}, w) + \varepsilon_t$$

where, w is a vector of all parameters and f is a function of network structure and connection weights. Since it is a machine learning model, there are no assumptions on the error distribution which is one of the major advantages over the traditional statistical models. Therefore, the NNAR model resembles a nonlinear autoregressive model. NNAR model is characterized by a network of three layers of simple processing units. The first layer is input layer, the middle layer is the hidden layer and the last layer is output layer. The relationship between the output (y_t) and the inputs ($y_{t-1}, y_{t-2}, \ldots, y_{t-p}$) can be represented as follows

$$y_t = f\left(\sum_{j=0}^{q} w_j g\left(\sum_{i=0}^{p} w_{ij} y_{t-i}\right)\right)$$

where, w_j ($j = 0,1,2,...,q$) and w_{ij} ($i = 0,1,2,..., p; j = 0,1,2,...,q$) are the model parameters often called the connection weights at the hidden node and input node respectively, p and q are the number of input nodes and hidden nodes respectively, g and f are the activation functions at the hidden and output layer respectively.

Support vector regression (SVR)

The SVR method is a nonlinear modelling procedure which utilizes the principle of structural risk minimization, in which the upper bound of the generalization error is minimized (Vapnik, 2000). Consider a vector of data set $Z = \{x_i\, y_i\}_{i=1}^{N}$ where $x_i \in R^n$ which contains both, the vector of input and $y_i \in R^n$ is the scalar output and N is the size of data set. The general expression of non-linear SVR estimation function is given by

$$f(x) = W^T \phi\ (x) + b$$

where $\phi(.)$: $Rn \rightarrow R^{nh}$ is a nonlinear mapping function from original input space into a higher dimensional feature space, which can be infinite dimensional, $w \in R^{nh}$ is weight vector, b is bias term and superscript T indicates transpose.

Classification and regression tree (CART)

CART *(Breiman, 1984)* is an algorithm or procedure that a decision tree uses to decide how/where to split a node into two or more sub-nodes. In the first step, a maximal tree or overgrown tree is to be built in such a way that it describes the training set as accurately as possible. At each level, the parent node that is split into two exclusive child nodes. In the next step, every child node becomes a parent node. Every split is described by one value of a split variable which is to be chosen in such a way that all objects in a child node have similar characteristic features of the response variable under study. In the next step, the overgrown tree is to be pruned to avoid the problem of over-fitting. In the final step, optimal tree is selected using a cross-validation (CV) procedure corresponding to a specific complexity parameter (Cp) value for which the cross-validation error is minimum.

RESULT AND DISCUSSION

Fitting of ARIMA model

Augmented- Dickey-Fuller (ADF) test has been applied for checking the stationarity and the results are presented in Table 1. The p-value of ADF test is 0.01 (<0.05), thus, both 3 days and 7 days interval PDI data are stationary. ARIMA (1, 0, 0) and ARIMA (0,0,1) are found to be the best fitted model using "auto.arima" function in R software for 3 days and 7 days interval PDI data respectively. ARIMA (1, 0, 0) indicates first order auto-correlation, i.e., 3 days interval PDI is correlated only with its past lag. ARIMA (0, 0, 1) indicates first order moving average model, i.e., averages of 7 days interval PDI with its past lag, which is used to find the trend in the data. The estimate of parameters, its standard error (S.E.) and respective p-values are presented in Table 2. Here, AR1 indicates auto regression of first order which indicates PDI is associated with its first order lagged values and MA1 indicates moving average of first order which indicates PDI is associated with first order lagged error of the model.

Table 1: ADF test for PDI at 3 days and 7 days interval

Variable	Test Statistic	p-value
3 days interval data	-4.1769	0.01
7 days interval data	-4.3028	0.01

Table 2: Parameter estimates of ARIMA model for PDI at 3 days and 7 days interval

Model	Parameter	Estimate	S.E.	p-value
ARIMA	AR1	0.585	0.106	<0.000
(1,0,0)	Intercept	32.744	2.021	<0.000
ARIMA	MA1	0.244	0.212	0.2511
(0,0,1)	Intercept	33.570	2.210	<0.000

Fitting of ARIMAX model

In order to fit the ARIMAX model, only statistically significant independent variables with PDI at 3- and 7-days interval data are included in the model. The correlations among PDI and weather variables are presented in Table 3. To predict PDI at 3 days interval, maximum, minimum and dew point temperature are the exogenous variables included in the model whereas for PDI at 7 days interval, minimum temperature, minimum RH and dew point temperature are the exogenous variables. ARIMAX (0,0,1) is found to be the best-fit model for both PDI at 3 and 7 days interval data using "auto.arima" function in R software. The estimated parameters of the models along with S.E. and p-values are presented in Table 4 and 5 respectively, and these are used to obtain the fitted model equation. MA1 is Moving Average of first order which is one of the parameters of ARIMAX (0,0,1) model.

Table 3: Correlations among PDI and weather variables

Parameters	RF	T_{max}	T_{min}	RH_{max}	RH_{min}	T_{dew}
3 days interval data	0.081	0.374**	0.455**	-0.180	-0.153	0.348**
7 days interval data	0.255	0.303	0.499**	-0.235	-0.347*	0.552**

:** Significant at 5% level of significance; *:** Significant at 1% level of significance

Table 4: Parameter estimates of ARIMAX (0,0,1) model for PDI at 3 days interval

Parameters	Estimate	S.E	p-value
MA1	0.564	0.095	0.000
Intercept	22.967	8.177	0.005
T_{max}	0.640	0.368	0.082
T_{min}	-0.223	0.066	0.001
T_{dew}	2.062	0.443	0.000

Table 5: Parameter estimates of ARIMAX (0,0,1) model for PDI at 7 days interval

Parameters	Estimate	S.E	p-value
MA1	0.058	0.292	0.843
Intercept	25.636	19.015	0.178
T_{max}	0.794	0.710	0.264
T_{min}	-0.276	0.178	0.121
T_{dew}	2.045	0.590	0.001

Table 6: Parameters of NNAR (p, k) model for PDI at 3 days and 7 days interval

Dataset	Number of lags (p)	Number of hidden nodes (k)
3 days interval PDI data	3	5
7 days interval PDI data	2	4

Fitting of NNAR model

The data are normalized before fitting the NNAR model in the scale 0 to 1 such that the variables which are measured in different scale become scale invariant. Normalization was done as, $X_n=(X_o-X_{min})/(X_{min}-X_{max})$, where X_o is the original data, X_{min} and X_{max} are the minimum and maximum value of X_o respectively. Twenty NNAR models were fitted each with different random starting weights. These are then averaged for producing the final weights. For PDI at 3 days interval, maximum, minimum and dew point temperature are the exogenous variables whereas for PDI at 7 days interval, minimum temperature, minimum RH and dew point temperature are the exogenous variables included in the model along with lagged values of PDI to fit the NNAR model. In NNAR (p,k) model, p indicates number of lags for PDI data and k indicates number of nodes in the hidden layer respectively. As per the thumb rule, half or one-third number of input nodes are used as the number of hidden nodes. The parameters of NNAR model for 3 days and 7 days interval data are given in Table 6.

Table 7: Parameters of SVR model for PDI at 3 days and 7 days interval

Dataset	Kernel	Cost	Gamma	Epsilon	Number of support vectors
3 days interval data	RBF	1	0.142	0.1	55
7 days interval data	RBF	1	0.142	0.1	26

Table 8: Decision tree Cp parameters estimates for PDI at 3 days interval

Size of the tree	Cp	Number of splits	Cross validation error
1	0.363	0	1.008
2	0.189	1	0.932
3	0.034	2	0.698
4	0.017	3	0.686
5	0.012	4	0.750

Fitting of SVR model

All weather parameters have been used as the input variables. First eighty percent of dataset were used for training and remaining datasets were utilized for testing. The SVR models have been fitted using the Radial Basis Function (RBF). The three hyperoperators of RBF function viz., Cost (C), Epsilon (ε) and Gamma (*YY*) were fixed at 1, 0.1 and 0.142 respectively following the result from the analysis. Based on SVR model fitting, number of support vectors were 55 and 26 for PDI data at 3- and 7-days interval respectively.

Fitting of CART model for 3 days interval data

The CART model explaining the initiation of potato Late blight disease in 3 days interval is given in Fig. 1. The overgrown tree has 4 splitting that consists of 5 root nodes. Cross validation has been done in order to prune the overgrown tree. The results of cross validation procedure are given in Table 8. The cross-validation error was lowest at the Cp value 0.017 and the corresponding number of splits was 3. Hence the optimal tree has obtained by retaining 4 terminal nodes and provided in Fig. 1.

In the optimum tree, first splitting was done using dew point temperature and the splitting point was 8.9^0C. The PDI was 36 when dew point temperature was greater than 8.9 ^{0}C for the preceding 3 days. The next splitting was done using minimum relative humidity and it was observed that when minimum relative humidity was greater than 67% for the preceding 3 days, PDI was high (43). However, the analysis does not indicate the severe status of the disease.

Fitting of CART model for 7 days interval data

The CART model explaining the initiation of potato Late blight disease in 7 days interval led to a regression tree consisting of 3 terminal nodes and 2 splitting nodes in the training data (Fig.2).

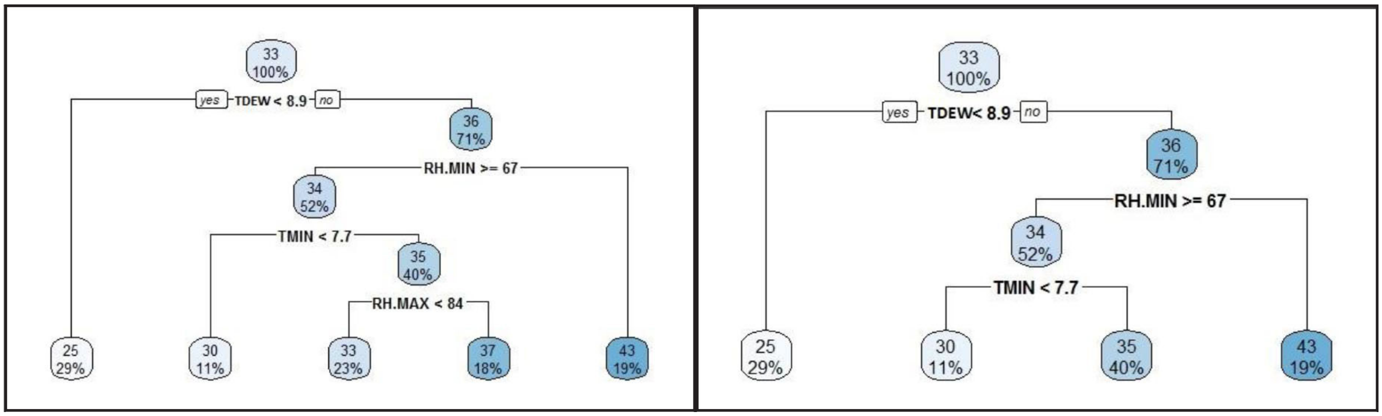

Fig. 1: Overall (left) and optimum (right) tree for PDI at 3 days interval

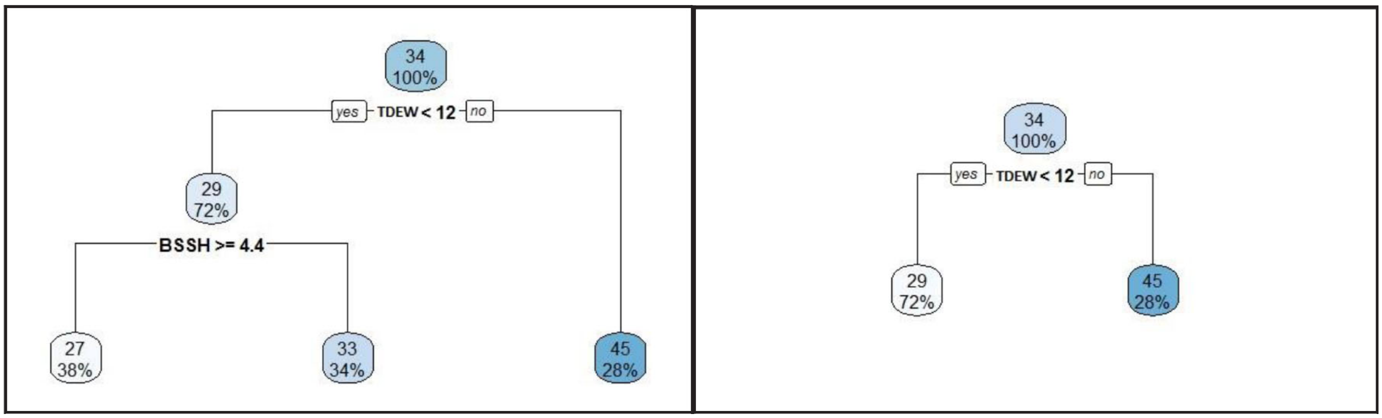

Fig. 2: Overall (left) and optimum (right) tree for PDI at 7 days interval

Decision tree cross validation procedure was used to obtain the optimal tree which contains two terminal nodes by a single split. The optimal tree is also provided in Fig. 2. The Dew point temperature was the important factor that influences the occurrence of Potato Late Blight severity. The best splitting point was at the dew point temperature 12 ^{0}C. When the dew point temperature was greater than 12 ^{0}C, the PDI value was high (45).

Model accuracy evaluation

In order to compare the performance of different statistical and machine learning models, the model accuracy criteria such as RMSE, MAE and MAPE have been used for both training and testing data set. The results are presented in Table 9 and 10.

Table 9: Model accuracy evaluation for PDI at 3 days interval

Model	Training			Testing		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
ARIMA	6.75	4.34	16.24	10.87	9.23	28.54
ARIMAX	5.11	3.92	14.08	14.98	13.54	34.54
NNAR	6.88	5.48	17.55	17.27	15.73	38.65
SVR	4.53	3.08	11.55	10.35	9.40	32.66
CART	6.55	5.49	19.53	10.32	9.13	31.32

From the Table 8, it is observed that SVR produces least RMSE, MAE and MAE for training dataset of 3 days interval PDI. However, during validation stage, the SVR had comparatively higher RMSE, MAE and MAPE values than CART. Hence, there are chances of overfitting in the SVR model. Therefore, CART is the best fit model as it has least RMSE, MAE and MAPE values for the testing dataset.

Table 10: Model accuracy evaluation for PDI at 7 days interval

Model	Training			Testing		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
ARIMA	9.82	8.11	28.82	13.01	10. 163	36.55
ARIMAX	6.39	5.10	18.20	13.34	10.86	39.33
NNAR	16.72	10.03	30.01	14.94	13.18	40.81
SVR	4.57	3.17	12.43	10.80	9.53	30.21
CART	5.29	3.85	13.66	7.88	6.98	25.15

From the Table 9, it can be seen that SVR produces the lowest RMSE, MAE, and MAPE for training dataset of 7 days intervals PDI. For testing dataset, CART produces lowest RMSE, MAE and MAPE followed by SVR. Thus, there are chances of overfitting in the SVR model. Therefore, CART is the best fit model to predict PDI at 7 days interval as it has least RMSE, MAE and MAPE values for the testing dataset. However, CART model to predict PDI at 7 days interval found to be superior in comparison to PDI at 3 days interval. Forecast for the season 2022-23 can be generated by the CART model for PDI at 7 days interval using the PDI and dew point temperature data from 2016-17 to 2021-22.

CONCLUSION

The study concludes that CART model was the best fitted model in terms of lowest RMSE, MAE and MAPE in the test data to predict PDI at 7 days interval among the developed models to predict PDI at 3 and7 days interval in the northern part of West Bengal. The results of CART showed that dew point temperature had a significant effect on PDI at 7 days interval. The disease incidence of potato late blight was high when average dew point temperature was greater than 12 ^{0}C in the preceding week.

ACKNOWLEDGEMENT

The authors are profoundly thankful to the editors and the anonymous referee for their valuable comments and suggestions which led to improvements in the article.

***Funding*:** No external fund received for this study.

***Data availability statement*:** Data shall be available from the corresponding author on request.

***Conflict of Interests*:** The authors declare that there is no conflict of interest related to this article.

***Authors contributions*:** **Vaidheki M**.: Conceptualization, Data Analysis and original draft preparation; **D.S. Gupta**: Conceptualization, Supervision; **P. Basak**: Critical Review and Shaping of final manuscript; **M.K. Debnath**: Supervision and Critical Review; **S. Hembram**: Disease data supply; **Ajith S**.: Critical Review.

***Disclaimer*:** The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Publisher's Note: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Anonymous (2020). Area and Production Statistics 2019-20, Directorate of Economics and Statistics, Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India. https://aps.dac.gov.in/APY/Public_Report1.aspx

Anonymous (2023). Agricultural Statistics at a Glance 2022, Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India.

Bierens, H. J. (1987). ARMAX model specification testing, with an application to unemployment in the Netherlands. *J. Econom.*, 35(1): 161-190.

Box, G. E. P., and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Revised Edition, San Francisco: Holden Day.

Breiman, L. (1984). Classification and Regression Trees (1st ed.). Routledge. https://doi.org/10.1201/9781315139470

Cook, H. T. (1947). Forecasting tomato late blight. *Food Packer*, 69-70.

Dar, W. A., Parry, F. A. and Bhat, B. A. (2021). Potato late blight disease prediction using meteorological parameters in Northern Himalayas of India. *J. Agrometeorol.*, 23(3): 310-315. https://doi.org/10.54386/jam.v23i3.35

Dey, S., Chakraborty, A. and Adhikary, N. K. (2022). Epidemiology of late blight of potato, its progress and apparent rate of infection. *Int. J. Environ. Clim. Chang.*, 12(7): 34-41.

Durgabai, R. P. L. and Bhargavi, P. (2018). Pest management using machine learning algorithms: a review. *Int. J. Comp. Sci. Eng. Info. Tech. Res.*, 8(1): 13-22.

Fry, W. (2008). Phytophthora infestans: the plant (and R gene) destroyer. *Mol. Plant Pathol.*, 9(3): 385-402.

Hembram, S., Patsa, R., Baskey, S. and Hansda, M. (2018). Effect of weather indices on late blight of potato and its management in northern district of West Bengal, India. *Int. J. Curr. Microbiol. Appl. Sci.*, *7*(9): 1737-1743.

Hyre, R. A. (1954). Progress in forecasting late blight of potato and tomato. *Plant dis. rep.*, 38(4): 245-253.

Madden, L. V. (1992). Rainfall and the dispersal of fungal spores. *Adv. Plant Pathol.*, 8: 39-79.

Sarkar, P., Basak, P., Panda, C. S., Gupta, D. S., Ray, M. and Mitra, S. (2023). Prediction of peak pest population incidences in jute crop based on weather variables using statistical and machine learning models: A case study from West Bengal. *J. Agrometeorol.*, 25(2):305-311. https://doi.org/10.54386/jam.v25i2.1915

Singh, V.K., Shailbala and Pundhir, V.S. (2012). Forecasting models: an effective tool for potato late blight management. *In:* Eco-friendly Innovative Approaches in Plant Disease Management. (Eds. Singh, V.K., Singh, Y. and Singh, A.). pp. 101-112, International Book Distributors and Publisher, New Delhi.

Vapnik, V. (2000). The nature of statistical learning theory. Springer Verlag New York Inc, New York.

Wallin, J. R. (1962). Summary of recent progress in predicting late blight epidemics in United States and Canada. *Am. Potato J.,* 39(8): 306-12.

Wheeler, B. E. J. (1969). An introduction to plant diseases. Wiley London, New York.