https://doi.org/ 10.33472/AFJBS.6.9.2024.598-605

African Journal of Biological Sciences

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

Research Paper Open Access

Modeling incidence of leaf miner in tomato in Rajendranagar (AP), India using machine learning techniques

Satish Kumar Yadav[1], D. Pawar[1], Latika Yadav[2] and Saurabh Tripathi[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 9, 2024

Received:22 Mar 2024

Accepted : 22 Apr 2024

doi: 10.33472/AFJBS.6.9.2024.598-605

Abstract

Studies on population of *leaf miner (Liriomyza trifolii)* in tomato (*Solanuml ycopersicum* Linnaeus) compared with weather data was carried out for eight consecutive years (2011-18) during *kharif* season. 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). The study concluded that the comparatively average population of *leaf miner* in experimental protected field was found higher than the other (1.3 Nos/5 larvae/plant) during 31 SMW in 2012 followed by farmer's field, and the lowest pest population (0.1 Nos/5 larvae/plant) was recorded in experimental unprotected field during in the 2016. Correlation analyses indicate while both current and one lag wind and RainyD had negative and positive influence respectively MinT and RHE had negative influence on leaf miner incidence. Among all variables, MaxT (current) and Rainy D (current and one lag) had highly significant positive effect on leaf miners. Machine learning techniques namely support vector regression (SVR), random forest (RF) and the other statistical models e.g., multiple linear regression (MLR), General regression neural network (GRNN), and Feed forward neural network (FFNN) are used. An empirical comparison of the above models is carried out based on root mean square error (RMSE). It is observed that, for leaf miner, the RMSE values of RF less as compared to other competing models. To this end, 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 of RF is higher than that of other models. The analysis is carried out using R package.

Keywords: Accuracy, Machine Learning, Statistical Models, Leaf minar, Tomato, Weather

Introduction

Tomato (*Solanuml ycopersicum* Linnaeus) native to south America is a major source of dietary antioxidant grown mainly in China, India, USA, Turkey, Egypt, Iran, Italy and Spain in 4.78 million hectares with production and productivity of 177.0 million tones and 37.0 t/ha respectively (Anon, 2018). In India, it is grown throughout the year under diverse agro-climatic regions largely in Madhya Pradesh, Karnataka, Andhra Pradesh, Telangana, Odisha, Gujarat and West Bengal in 0.79 million hectares with production of 19.76 million tons (Anon 2018). In Telengana, tomato is very popular and is cultivated in 0.41 lakh ha with production of 1.17 million tones and productivity of 28.2/t/ha. Among several tomato production constraints, poor agronomic practices and crop loss by major pests are important. (Kumari et al., 2015a) reported changed disease scenario of tomato in Telengana due to altered climatic conditions. Recently (Kumari et al., 2018) reported upto 90 % infestation by tomato pinworm, *Tuta absoluta* in various regions of Telangana. Tomato pinworm is a new invasive pest established in Malnad and Hyderabad-Karnataka region of Karnataka (Sridhar et a*l.*, 2014) which entered Telangana during November 2014 with crop loss upto 60% (Kumari et al., 2015b). While the first two instars mine leaves by feeding on mesophyll leaving epidermis intact, create tunnels known as “mines”, third and fourth instar larvae mines and bore into stalks, apical buds and fruits. Fruits infested by *T. absoluta* are easily identified by the presence of characteristic pin holes.

Occurrence and progress of insect pests depend on climatic factors of temperature, relative humidity, precipitation etc. (Aheer et al., 1994). Using climatic factors, predictive models have been developed for occurrence of crop pests warranting timely application of pest management practices (Vennila et al., 2018). Multiple linear regression model (MLR) is commonly used to investigate the relationship between dependent variables with two or more independent variables. However, due to the nature of linear relationships, linear regression models may not provide accurate predictions in some complex situations (Paswan and Begum, 2013). Recently, with the advent of data mining, machine learning technology in the field of agriculture is getting much attention. Machine learning is a method that works with data analysis and seeks to automate the construction of analytical models (Shekoofa et al., 2014; Li et al., 2016). Precise and timely forecasting of pest incidence helps farmers in planning effective management strategies. In the present study, some of the machine learning techniques such as Support vector regression (SVR), Generalized regression neural network (GRNN), Random Forest (RF) and Feed forward neural network (FRNN) and linear multiple regression (MLR) analyses were applied and compared to forecast tomato pinworm using nine years weather data of 2011-2018 under Nation initiative on climate resilient agriculture (NICRA) in the region of Telangana.

MATERIALS AND METHODS

Multiple Linear Regression Analysis

The general form of MLR for a data set of N observations on a response variable Y and p predictor variables, $X_1, X_2, \ldots, X_p$ is

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon.$$

where β_0 is intercept, $\beta_1, \ldots \beta_p$ are the regression coefficients and ε is error term which is assumed to follow the normal distribution with mean zero and a constant variance. In the present investigation, the stepwise selection procedure for selecting the significant variable in the model was adopted.

Support vector regression (SVR)

For a given data set $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in R^n$ input vector is, $y_i \in R$ is scalar output and N corresponds to size of data set, general form of Nonlinear SVR estimating function (Fig. 1) is:

$$f(x) = w^T \varphi(x) + b$$

Where $\varphi(.) : R^n \rightarrow R^{n_h}$ is a nonlinear mapping function from original input space into a higher dimensional feature space, which can be infinitely dimensional, $w \in R^{n_h}$ is weight vector, b is bias term and superscript T indicates transpose. The coefficients w and b are estimated from data by minimizing the following regularized risk function:

$$R(\theta) = \frac{1}{2}\left|\left|w\right|\right|^2 + C\left[\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)\right].$$

In above equation, first term $\frac{1}{2}||w||^2$ is called 'regularised term', which measures flatness of the function. Second term $\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)$ called 'empirical error' is estimated by Vapnik ε-insensitive loss function, C referred to as regularized constant. The SVR model was applied using R software package (e1071). Data on mean and maximum severity of early blight for available seasons under each location along with the weather variables (lagged by one and two weeks) considered under MLR were used for SVR.

Artificial neural network (ANN)

ANNs are powerful tools for modeling nonlinear data and are a self-adaptive approach especially when the underlying data relationship is unknown. A general neural network consists of an input layer that accepts external information, one or more hidden layers and the output layer that provides the target value. Each layer consists of one or more nodes. Most used ANN is multi-layer perceptron (MLP), a class of feed forward neural network. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. An application of this approach can be found in Paul and Sinha (2016).

Random forest (RF)

Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity. Random forest is a supervised learning algorithm. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. One big advantage of random forest is that it can be used for both classification and regression problems, which form most current machine learning systems (Fig. 2).

Generalized regression neural network (GRNN)

Generalized regression neural network is related to the radial basis neural networks, which are found on kernel regression. It can be treated as a normalized radial basis neural network in which there is a hidden neuron centered at every training case. These radial basis function units are generally probability density function such as the Gaussian (Celikoglu 2006). GRNN approximates any arbitrary function between input and target vectors; fast training and convergence to the optimal regression surface as the training data becomes very large (Specht 1991). This makes GRNN a very advantageous tool to perform predictions. Figure 4 is a representation of the GRNN architecture with four layers: an input layer, a hidden layer, a summation layer, and an output layer.

Feed forward neural networks (FFNN)

Deep Feed forward networks or also known multilayer perceptrons are the foundation of most deep learning models. These networks are mostly used for supervised machine learning tasks where we already know the target function i.e. the result we want our network to achieve and are extremely important for practicing machine learning.

Validation of forecasts

The dataset of pest population and weather was divided in two parts before analysis for each location with 90% of the observations for estimation (model development) and remaining 10% for validation. Comparative assessment of prediction performance of different models namely MLR, RF, GRNN, FNN, and SVR models was carried out in terms of root mean square error (RMSE) based on the following formulae (Fig. 3).

$$\text{RMSE}=\sqrt{1/h\sum_{i=1}^{h}\left\{y_{t+i}-\hat{y}_{t+i}\right\}^{2}}$$

where h denotes the number of observations for validation, yi 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.

RESULTS AND DISCUSSION

Seasonal dynamics and status of Leaf miner

In recent years, pest epidemics have been due to climate change and there is a need to understand the impact of climate change crop pests to outline appropriate management strategies (Chowdappa, 2010). Leaf miner infestation dynamics in tomato was studied for eight consecutive years of kharif 2011-18 in Rajendranagar (AP), Telangana (Fig. 4). Leaf miner infestation appeared as early as during 27 SMW in 2014, 2015 and 2016 and latest by 31 SMW during 2018 reflecting varied incidence initiation period during eight years of study. Peak population of leaf miner appearance also varied across the seasons with higher (1.3) during 2012 closely followed by 2014 (1.2) with lowest peak (0.3) recorded during 2015. During the study, the influence of weather factors on peak occurrence of leaf miner was evident. Maximum (1.1 to 1.3/5 leaves/plant) occurred between 29 to 34 SMW during 2012 to 2014. During 2015, leaf miner population remained almost static at 0.1 to 0.3 /5 leaves/plant.

Comparative analysis of Leaf mineroccurrence across the years

Seasonal average reflects seriousness of a pest in a given locality. Comparisons of leaf miner mean incidence for levels across seasons carried out using Duncan's Multiple Range Test (DMRT) is presented in Table 1. While seasonal average varied from a maximum and significantly higher population of 0.62 during 2012 to a lowest of 0.14 during 2015 and 2016 which were at par.

Correlation coefficients between leaf miner with weather factors

Pearson's correlation analysis was carried out to find out the influence of current and one lag weather variables on the occurrence of leaf miner in tomato (Table 2). Insect pests and diseases are well known to be governed and influenced by climatic and weather conditions. While both current and one lag wind and RainyD had negative and positive influence respectively MinT and RHE had negative influence on leaf miner incidence. In a similar type of study, Choudary and Rosaiah (2000) also reported that minimum temperature and evening relative humidity were negatively correlated with *L. trifolii* incidence in tomato. Among all variables, MaxT (current) and Rainy D (current and one lag) had highly significant positive effect on leaf miner population contradicting the report of Reddy and Kumar (2005) with respect to RainyD. Negative non-significant correlation obtained between morning and evening relative humidity reported by Reddy and Kumar (2005) is in agreement with our findings.

Validation

Once forecast values were obtained for the leaf miner population through five models viz., MLR, RF, GRNN, FFNN and SVR the performance of prediction was tested using root mean square error (RMSE) (Table3). The RMSE values of RF model was less as compared to other models. In order to check the adequacy of fitted models, residuals diagnostics were carried out and it revealed that there are no autocorrelations among the residuals. Leaf miner population trend predicted by RF followed almost a similar trend of actual observed values (Fig.5).

Test results

Diebold-Mariano test (Diebold and Mariano, 1995) was applied for comparing the forecasting performance of RF with MLR, SVR, GRNN and FNN models based on the null hypothesis that the predictive accuracy of two competing models is equal. Different combinations of comparisons with specific alternative hypothesis along with test statistics and their significance are reported in table 5. Results revealed that predictive accuracy of MLR is found to be less than that of RF model. Similarly for other comparisons i.e., RF *vs* MLR, RF *vs* SVR, RF *vs* GRNN and RF *vs* FNN the test is found to be significant implying that there is statistically significant difference in predictive accuracy in the pair of comparisons. As can be seen by the specification of alternative hypothesis given in table 5, it may be concluded that RF performs better than all other models for the present data set. Recently, (Balaban et al. 2019) showed RF models had accuracy over 99% to predict the nymphal stage of the sunn pest in the Middle Eastern region.

CONCLUSION

Climate change has serious impact on seasonal dynamics of the pest and diseases. Leaf miner infestation dynamics in tomato was studied for eight consecutive years of kharif 2011-18 in Rajendranagar (AP), Telangana. Present study revealed that Leaf miner infestation appeared as early as during 27 SMW in 2014, 2015 and 2016 and latest by 31 SMW during 2018. The peak population of leaf miner appearance also varied across the seasons. Statistical models *i.e.*, MLR including machine learning techniques i.e., RF, GRNN, FNN, SVR were used to analyze the Leaf miner infestation. Empirically, RF was found out to be the best in terms of predicting the infestation for the present data set. The same is supported by the DM test. The approaches discussed here may be replicated to forewarn the incidence of other important pest and diseases in important agricultural crop which in turn will help the farmers to take necessary preventive measures to minimize the loss due to pest and diseases attack.

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.

REFERENCE

Aheer, G.M., Ahmad, K.J. and Ali, A. (1994). Role of weather in fluctuating aphid density in wheat crop. Ayub Agricultural Research Inst., Faisalabad (Pakistan).

Anon, Horticultural statistics at a glance (2018). Department of Agriculture, Cooperation & Farmers' Welfare Horticulture Statistics Division, p. 458.

Balaban, I., Acun, F., Arpalı, O.Y. Murat, F. Babaroglu, N.E., Akci, E., Çulcu, M., Ozkan, M. and Temizer, S. (2019). Development of a Forecasting and Warning System on the Ecological Life-Cycle of Sunn Pest.1905.01640.

Biau, G. (2012). Analysis of a Random Forests Model. *J. Mach. Learn. Res. 13*, 1063–1095.

Choudary, D.P. and Rosaiah, R.B. (2000). Seasonal occurrence of Liriomyza trifolii (Burgess) (Agromyzidae:Diptera) on tomato crop and its relation with weather parameters. *Pest Management Eco Zoology*. 8:91-95.

Chowdappa, P. (2010). Impact of climate change on fungal diseases of Horticultural crops: *In*: Challenges of climate change-Indian Horticulture (Singh, H. P, Singh, J.P, Lal, S. S. (eds.) Westville publishing house, New Delhi. p.144-151.

Fawagreh, F., Mohamed Medhat Gaber∗ and Eyad Elyan (2014). Systems Science & Control Engineering: An Open Access Journal, 2014 http://dx.doi.org/10.1080/21642583.2014.956265 Random forests: from early developments to recent advancements .

Haiguang Wang (2011). Prediction of Wheat Stripe Rust Based on Support Vector Machine, 2011 Seventh International Conference on Natural Computation, pp.378–382, 2011.

Kumari, D.A., Anitha, G., Anitha, V., Lakshmi, B.K.M., Vennila, S and Rao, N.H.P. (2015b). New record of leaf miner *Tuta absoluta* (Meyrich) in Tomato. *Insect Environment*. 20:136-138,.

Kumari, D.A., Anitha, G., Vennila, S. and Hanuman Nayak, M. (2018). Incidence of tomato pinworm, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in Telangana (India). *Journal of Entomology and Zoology Studies*; 6: 1085-1091.

Kumari, D.A., Lakshmi, B. ., Sireesha, K., Vennila, S. and Hariprasad, N., Rao. (2015a). Changing scenario of diseases in tomato in Telangana. 1st To 3rd 66 December, 2015 - Sri Konda Laxman Telangana State Horticultural University, Rajendranagar, Hyderabad, Telangana State.

Li, Y.H., Xu, J.Y., Tao, L., Li, X, F, Li, S. and Zeng, X. (2016). SVM-Prot 2016: a web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PloS one*;11: e0155290.

Paswan, R.P. and Begum, S.A. (2013). Regression and Neural Networks Models for Prediction of Crop Production. *Int. J. Sci. Eng. Res*. p 4- 11.

Reddy, N.A., and Kumar, C.T.A. (2005). Influence of weather factors on abundance and management of Serpentine leaf miner, Liriomyza trifolii (Bugress) on tomato. *Annals of Plant Protection Sciences*. 13:315-318.

Shekoofa, A., Emam, Y., Shekoufa, N., Ebrahimi, M. and Ebrahimie, E. (2014). Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: a new avenue in intelligent agriculture. *PloS one;* 9: e97288.

Sridhar, V., Chakravarthy, A.K., Asokan, R., Vinesh, L.S., Rebijith, K.B. and Vennila, S. (2014). New record of *Tuta absoluta* (Meyrick) (Lepidoptera*:* Gelechiidae) in India. *Pest Management in Horticultural Ecosystems*; 20:148-154.

Vennila, S., Satish Kumar Yadav, Priyanka Wahi, S., Kranthi, A., Amutha & Dharajothi, B. (2018). Seasonal Dynamics, Influence of Weather Factors and Forecasting of Cotton Sap Feeders in North India. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*; p467–476.

Younes C., (199). A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease", Chemometrics and Intelligent Laboratory Systems, vol. 48, pp.47–58.