INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print), 2222-5234 (Online)
http://www.innspub.net
Vol. 24, No. 5, p. 96-106, 2024

RESEARCH PAPER OPEN ACCESS

Predictive Analysis of Occurrence of Thrips in Tomato Subject to Weather Parameters Using Machine Learning Techniques

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

[1]*Department of Statistics, Amity Institute of Applied Sciences, Amity University, Noida-201313, India*

[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, India*

Key words: Accuracy, Machine Learning, Statistical Models, Thrips, Tomato, Weather

http://dx.doi.org/10.12692/ijb/24.5.96-106 Article published on May 04, 2024

Abstract

Thrips (*Thripidae*) on tomato (*Solanum lycopersicum L.*) at Rajendranagar, Andhra Pradesh, India is modelled based on field data sets generated during six kharif seasons [2011-18]. The weather variables considered are maximum & minimum temperature (MaxT & MinT) (^oC), morning and evening humidity (RHM & RHE) (%), sunshine hours (SS) (hr/d), wind velocity (Wind) (km/hr), total rainfall (RF) (mm) and rainy days (RD). Thrips incidence was higher during 2012 and lowest in 2014. Correlation analyses significant positive influence of maximum temperature and negative influence of wind of one lags, RHM both current and one lags, rainfall one lag of negative influence on thrips. Machine learning techniques namely. An empirical comparison of the above models [support vector regression (SVR), random forest (RF) and the other statistical models e.g., multiple linear regression (MLR), ridge regression (RR), least absolute shrinkage and selection operator (LASSO), and elastic net (EN)] is based on root mean square error (RMSE). It is observed that, for thrips, the RMSE values of RF and LASSO 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, predictive accuracy of RF and LASSO is higher than that of other models.

* **Corresponding Author:** Satish Kumar Yadav ✉ satishkumaryadav.akash@gmail.com

Introduction

Tomato (*Solanum lycopersicum* L.) is one of the most popular produced and extensively consumed vegetable crops in the world (Grandillo *et al.*, 1999). It is one of the most important vegetable crops in India that can be eaten raw in salads or as an ingredient in many dishes and in drinks (Alam *et al.,* 2007). Tomatoes and tomato-based foods provide a wide variety of nutrients and many health-related benefits to the body. In regions where it is being cultivated and consumed, it constitutes a very essential part of people's diet. Tomatoes production accounts for about 4.8 million hectares of harvested land area globally with an estimated production of 165 million tonnes (FAOSTAT, 2017). China leads world tomato production with about 50 million tonnes followed by India with 17.8 million tonnes. Tomato production can serve as a source of income for most rural and peri-urban producers in most developing countries. Despite all the numerous benefits from the crop, many challenges are making its production unprofitable in most developing countries, especially those in Africa. The challenges faced by producers are seen either in production, post-harvest, marketing or a combination of any of them. The purpose of this paper is to look at the post-harvest challenges that result in losses and recommend some low-cost intermediate technologies needed to remedy the situation. Accounting for about 8.23% of the total vegetable production in the country. Tremendous progress has been made in tomato production during the past four and half decades. At present, India is the fourth largest producer of tomato, accounting for 6.8% of the world production and the second largest in terms of acreage, accounting for 11.9% of area under tomato in the world. Tomato spotted wilt virus (TSWV) is widely distributed and has caused serious losses in the yield of this and many other crops in Australia, India, Nepal, China, Thailand, and USA. Early infections cause the most severe damage and can lead to total crop loss. Epidemics of insect-transmitted plant viruses in agricultural ecosystems require the interaction of 3 basic components: the host plant of the virus, the insect vector and the plant pathogenic virus. While this triad sounds quite straight forward, the relationships and interactions occurring between and among the basic triad components and the environment are complex and dynamic, frequently defying complete understanding by scientists and agricultural practitioners worldwide. Many plant viruses are transmitted by arthropod vectors (Nault, 1997). TSWV, the type species of the genus Tospovirus, family Bunyaviridae (Murphy *et al.,* 1995), is exclusively transmitted by several thrips species in a propagative manner (German *et al.,* 1992; Ammar, 1994; Goldbach and Peters, 1994). Tomatoes are susceptible to more than 200 diseases. Important achievements in chemical, biological, cultural and genetic control methods have greatly reduced economic losses and sometimes have eliminated them. Viral diseases are a special case since they cannot be controlled by chemical treatments. Crop protection must then rely on genetic resistance or on disease avoidance. TSWV was first reported in India in tomato in 1964 (Todd *et al.,* 1975). The occurrence of TSWV on a legume in India was first recorded in 1968 (Reddy *et al.,* 1968). Thrips (*Thysanoptera: Thripidae*) cause serious problems in the cultivation of a wide range of greenhouse and field crops. They create major damage on plants by causing reduction in plant growth, deformation of plant organs, and cosmetic damage in the form of silver scars on leaves and flowers. Thrips cause direct damage during feeding, plants should be released by breaking the leaf, fountain and fruits cells, leaving behind silvery patches and fruit sores to reduce plant yields and tomato market shortage (Riley and Pappu, 2004; Staford *et al.,* 2011). And these are dependent on weather conditions (Verhage *et al.,* 2017; Harvey *et al.,* 2018). Thus, there is a need for the development of predictive models for the incidence of pests and diseases that can improve the interpretation of the crop cycle according to the weather, incorporating weather-soil-plant factors (Malau *et al.,* 2018; Badnakhe *et al.,* 2018). 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). It is a field of computer science that works with the recognition of patterns using computational learning theory in artificial intelligence (Sahoo *et al.,* 2017). Machine learning algorithms are very promising for faster, more dependent variables and the meteorological elements are the independent variables of the models. Elastic Net (EN), a penalized variable selection approach that combines both ridge penalty and LASSO penalty. Different forecasting techniques e.g., Multiple Linear Regression (MLR); K Neighbors Regressor (KNN); Random Forest Regressor (RFT), and Artificial Neural Networks–Multilayer Perceptron (MLP), EN, LASSO are applied. The ridge method of MLR is utilized. This method avoids poor conditioning of the matrix of the repressor variables, controlling the inflation and the general instability found in least squares estimators. Ridge avoids the multicollinearity problem without having to exclude repressor variables, so it has no information loss.

Materials and methods

The techniques used in the present investigation as depicted in fig. 1 are described below briefly.

Multiple Linear Regression (MLR)

Generalized MLR for a data set of N observations on a dependent or response variable Y and p predictor or explanatory variables, $X_1, X_2, \ldots, X_p$ is

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

where regression coefficients $\beta_0, \beta_1, \ldots \beta_p$ are constants and ε is a random disturbance or error which is assumed to follow the normal distribution with mean '0' and a constant variance. It is assumed that Y is approximately a linear function of the X's, and ε measures the discrepancy in that approximation (Chatterje and Hadi, 2012). In the present investigation, stepwise selection procedure for selecting the significant variable in the model is adopted.

Support vector regression (SVR)

SVR model is given for a data set as $D = \{(x_i, y_i)\}_{i=1}^{N}$, 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 is

$$f(x) = w^T \varphi(x) + b \quad (1)$$

where $\varphi(.): R^n \to R^{n_h}$ 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

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

This regularized risk function minimizes both empirical error and regularized term simultaneously and implements structural risk minimization (SRM) principle to avoid under and over fitting of training data (Vapnik, 2000). In Equation (2), 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 functionA schematic representation of the loss function under nonlinear SVR model is given in Fig.2. Applied SVR technique for prediction of SMD of Pigeonpea in Banaskantha Region of Gujarat (Paul *et al.*, 2018). The SVR model was applied using R software package (e1071).

Random Forest (RF)

Random forest (Breiman, 2001) 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 (Liaw and Wiener, 2002), because of its simplicity and diversity (it can be used for both classification and regression tasks). Random forest is a supervised learning algorithm.

The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result. The schematic representation is given in fig. 3. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

Ridge Regression (RR)

The ordinary least square (OLS) estimates depend upon $(X^TX)^{-1}$, where X is the matrix of regressors. If X is ill-conditioned then small changes to the elements of X lead to large changes in $(X^TX)^{-1}$. Thus, the OLS estimates become computationally unstable, and the individual component estimates may either have the wrong sign or too large magnitude. Ridge regression shrinks the regression coefficients by imposing a penalty on their size. The ridge coefficients minimize the penalized residual sum of squares (RSS) and are given by,

$$\hat{\beta}^{ridge} = \arg\min\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda_2\sum_{j=1}^{p}\beta_j^2\}_\beta$$

Here $\lambda \geq 0$ is a complexity parameter that controls the amount of shrinkage, the larger the value of λ, the greater the amount of shrinkage. Hence, the ridge regression solution is,

$\hat{\beta}^{ridge} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$, where I is the p × p identity matrix.

Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO is a constrained OLS minimization problem and the LASSO estimate is defined by

$$\hat{\beta}^{LASSO} = \arg\min\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda_1\sum_{j=1}^{p}|\beta_j|\}$$

Here the L_2 ridge penalty $\sum_{j=1}^{p}\beta_j^2$ is replaced by the L_1 LASSO penalty y $\sum_{j=1}^{p}|\beta_j|$ with tuning parameter $\lambda_1 \geq 0$. Finally, the regression coefficients are estimated as

$$\hat{\beta}^{LASSO} = (X^TX)^{-1}(X^Ty - \frac{\lambda_1}{2}w),$$

where the elements w_j of w are either +1 or −1, depending on the sign of the corresponding regression coefficient β_j.

Tibshirani (1996) used a quadratic programming algorithm to estimate the LASSO coefficients. As variable selection becomes increasingly important in modern data analysis, the lasso has become more appealing due to its sparse representation.

Elastic Net (EN)

Zou and Hastie (2005) introduced the elastic-net penalty, which is a compromise between RR and LASSO. The EN simultaneously does automatic variable selection and continuous shrinkage similar to LASSO, and can select groups of correlated variables The penalized loss function is given by,

$$\hat{\beta}^{EN} = \arg\min\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$$

The above procedure can be viewed as a penalized least-squares method. By considering, $\alpha = \lambda_1/(\lambda_1+\lambda_2)$ the equivalent optimization problem becomes

$$\hat{\beta}^{EN} = \arg\min|y - X\beta|^2 \quad \text{subject to}$$

$$(1-\alpha)|\beta|_1 + \alpha|\beta|^2 \leq t \quad \text{for some t}$$

when $\alpha = 1$, the EN becomes RR and for $\alpha = 0$ it becomes a LASSO situation. Thus, EN has characteristics of both RR and LASSO. In matrix notation,

$RSS(\lambda 1, \lambda 2) = (y - X\beta)^T (y - X\beta) + \lambda\, w^T\beta + \lambda 1\, \beta^T\beta$

For the EN the regression coefficients are estimated as $\hat{\beta}^{XN} = (X^TX + \lambda I)^{-1}(X^Ty - \frac{\lambda_1}{2}w)$

Validation of forecasts

The dataset of pest population and weather was divided into two parts before analysis. For each location 90% of the observations were used for estimation (model development) and remaining 10% observations were used for validation. Comparative assessement of prediction performance of different models namely MLR, RF, SVR, RR, LASSO, EN models was carried out in terms of root mean square error (RMSE) based on the following formluae:

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

where h dentoes 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.

Results and discussion

Study locations

Study was a part of information and communication technology (ICT) based pest surveillance on tomato implemented at experimental research stations of Rajendranagar, AP, India. Two fields grown with tomato 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 thrips (*Thysanoptera*) on whole plant basis from early vegetative stage till crop harvest. While the surveillance plans were fixed during season the sampling plan for thrips 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.

Table 1. Comparative analysis of thrips occurrence across the years.

2011	2012	2013	2014	2015	2016	2017	2018
0.47^{abc}	0.16^{c}	1.04^{a}	-	0.82^{ab}	0.23^{bc}	0.40^{abc}	0.44^{abc}

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

Table 2. Pearson Correlation Coefficients of thrips occurrence with Climatic variables.

Lag	MaxT	MinT	RHM	RHE	Rainf	SunS	Wind	RainyD
Current Lag	0.99***	-0.17	-0.32*	-0.06	-0.14	-0.11	0.16	-0.07
One Lag	-0.03	-0.05	-0.37**	-0.17	-0.29*	-0.10	0.28*	-0.09

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

Data accrual and reporting system

Proforma of pest surveillance for tomato (ref *http://www.ncipm.res.in/Nicra2015/NICRAAdminPanelNew/rvLogin.aspx*) which also had variables of thrips was used for recording spot wise observations during each week. A client software developed for offline entry and online 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. Provisions were also kept for data entry of daily weather data for varibles *viz.,* MaxT & MinT (°C), RHM & RHE (%), SS (hr/d), Wind (km/hr), RF (mm) and RD from the experimental station in the client software along with insect pest and predatory data and SMW based weather data calculations in reporting system.

Seasonal dynamics and status of Thrips

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. Studies on thrips in tomato carried out for seven consecutive kharif seasons (2011-13 and 2015-2018) at Rajendranagar, AP location showed the commencement of infestation from second week of August with peak incidence between third week of October and November. Thrips were higher during 2012 and lowest in 2014. The seasonal variation in occurrence of thrips in tomato studied location is graphically represented in Fig. 4.

Table 3. MSE and RMSE values in relation to MLR, SVR, RF, RR, LASSO and EN models predicting of Thrips.

Obs.	MLR	SVR	RF	RR	LASSO	EN
0.30	0.48	0.33	0.45	0.45	0.48	0.50
0.54	0.47	0.50	0.58	0.63	0.50	0.50
1.13	0.53	0.46	0.99	0.44	0.50	0.50
0.24	0.52	0.42	0.31	0.44	0.51	0.50
0.47	0.54	0.43	0.47	0.45	0.52	0.50
0.41	0.56	0.37	0.45	0.44	0.50	0.50
0.10	0.52	0.43	0.29	0.47	0.51	0.50
0.40	0.54	0.44	0.47	0.48	0.53	0.50
0.44	0.61	0.40	0.52	0.53	0.59	0.50
0.48	0.74	0.50	0.54	0.56	0.55	0.50
MSE	0.08	0.52	0.01	0.07	0.01	0.04
RMSE	0.28	0.72	0.10	0.26	0.09	0.19

Table 4. Testing of Prediction Accuracy of Thrips.

Combinations	Alternative Hypothesis	D-M Statistic	*p-value*
RF and MLR	Predictive accuracy of MLR is less than that of RF	-5.24	0.04
RF and SVR	Predictive accuracy of SVR is less than that of RF	-4.32	0.02
RF and Ridge	Predictive accuracy of Ridge is less than that of RF	-2.41	<0.0001
RF and Lasso	Predictive accuracy of Lasso is equal to RF	-1.81	0.03
RF and Elastic net	Predictive accuracy of Elastic net is less than that of RF	-6.12	<0.0001

***: significant at p< 0.001; **: significant at p< 0.01; *: significant at p< 0.05.

Comparative analysis of Thrips occurrence across the years

Comparisons of thrips mean incidence for levels across seasons carried out using Duncan's Multiple Range Test (DMRT) are presented in Table 1. Thrips incidence was significantly lower in 2016 as compared to 2012 and 2013 with on par incidence during other seasons.

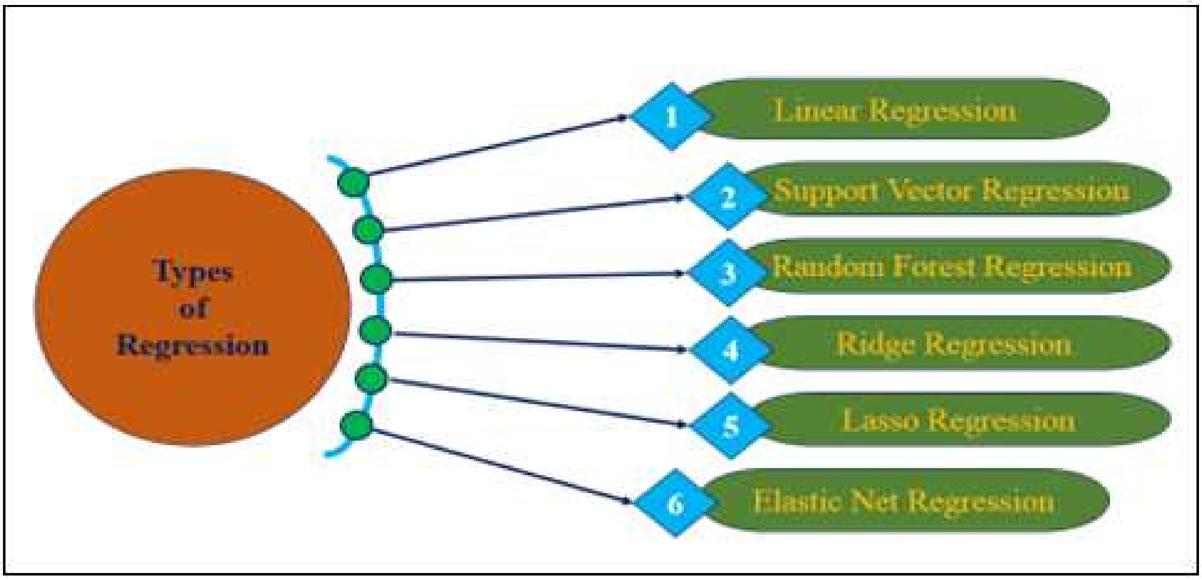

Fig. 1. Schematic representation of types of regression.

Correlation coefficients between thrips with weather factors

Pearson's correlation analysis was carried out to find the significant weather variables influencing the occurrence of thrips in tomato and the same is reported in Table 2. Correlation analyses of significant positive influence of maximum temperature and negative influence of Wind of one lags, RHM both current and one lags, Rain fall one lags of negative influence in thrips.

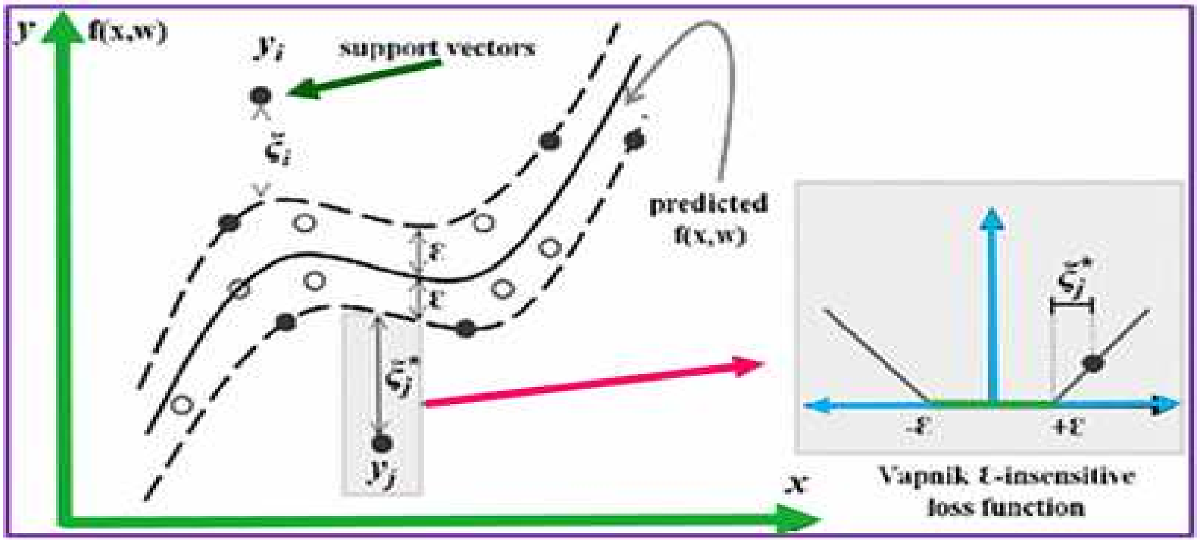

Fig. 2. A schematic representation of Vapnik ε-insensitive loss function and accuracy tube under nonlinear SVR model.

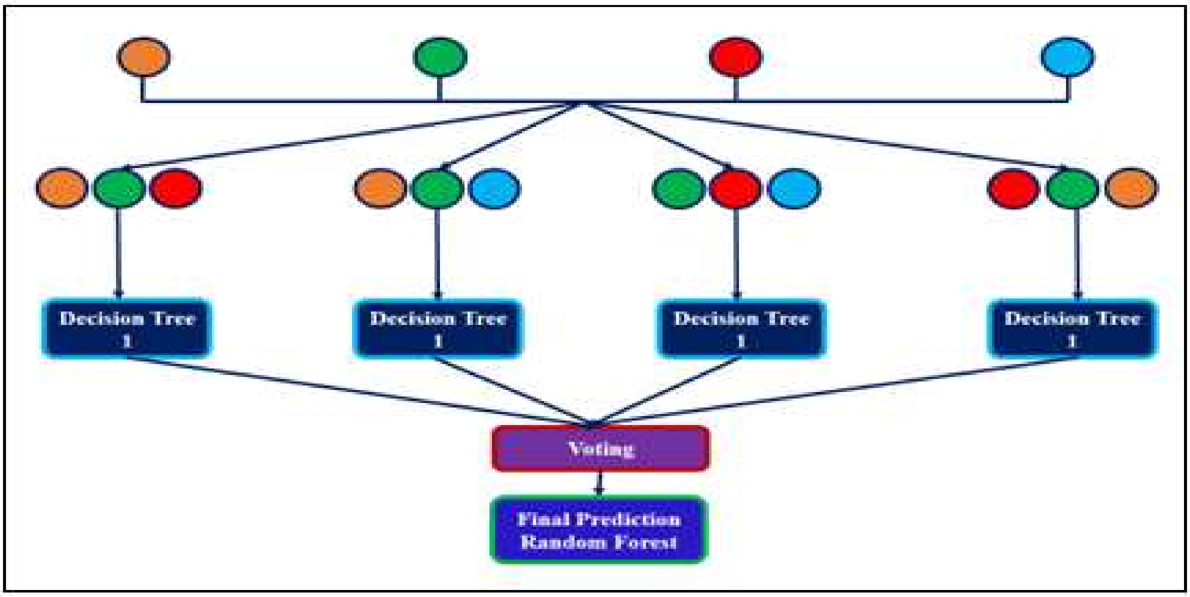

Fig. 3. Schematic representation of RF methodology.

Validation

After estimation of the models, forecasts were obtained for the validation set. The performance of prediction of thrips occurrence in tomato through various models *viz.* MLR, RF, Ridge, Lasso and Elastic net were tested using root mean square error (RMSE) and reported in Table 3. For thrips, the RMSE values of RF model are less in comparison to other competing models. To check the adequacy of fitted model, residuals diagnostics were carried out and it revealed that there are no autocorrelations among the residuals.

Fig. 4. Infected of the leaf and fruit.

Test results

Diebold-Mariano test (Diebold *et al.*,1995) applied for comparison of forecasting performance among Random Forest (RF), Multiple Linear Regression (MLR), Ridge, Lasso and Elastic net models was based the null hypothesis that the predictive accuracy of any two competing models is equal revealed.

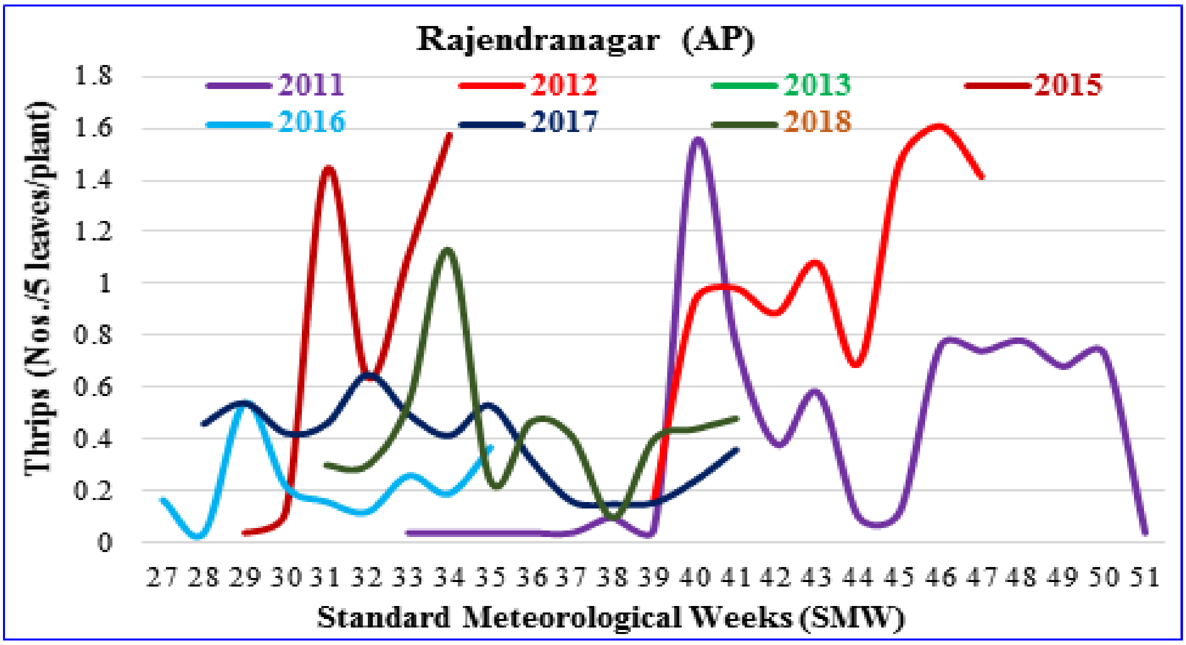

Fig. 5. The seasonal variation of occurrence of thrips in tomato.

Different combinations of comparisons, the specific alternative hypothesis along with test statistics and their significance are reported in Table 4.

In thrips the predictive accuracy of MLR is found to be less than that of RF model whereas in other two comparisons i.e., RF *vs* MLR, RF *vs* SVR, RF *vs* RR, RF *vs* LR and RF *vs* ER the test is found to be significant implying that there is no statistically significant difference in predictive accuracy in the pair of comparisons.

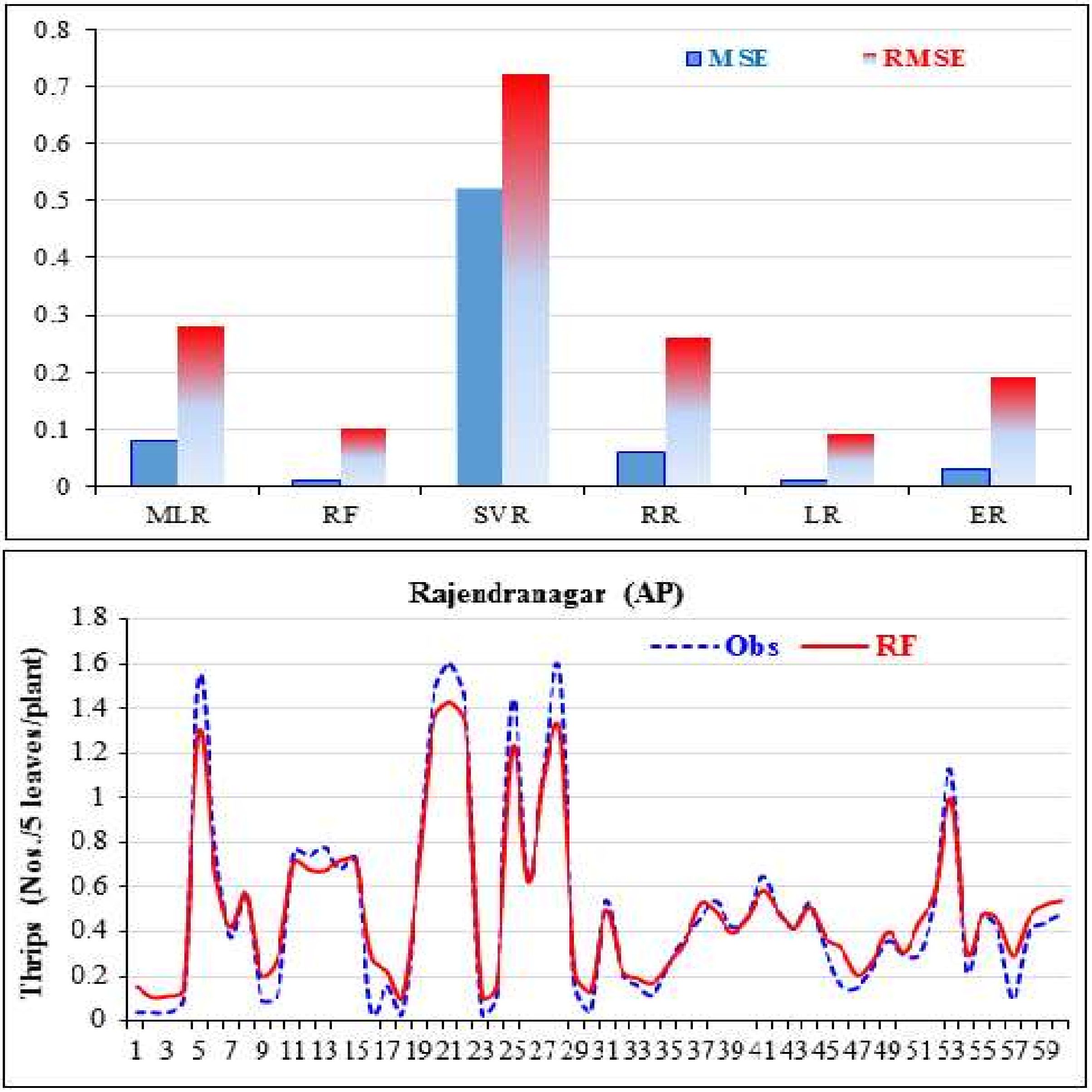

Fig. 6. Obs.: Thrips, MLR: Multiple Linear Regression, SVR: Support vector regression, RF: Random forest, RR: Ridge regression, LR: Least Absolute Shrinkage and Selection Operator regression and ER: Elastic net Regression.

Conclusion

Climate change has an adverse impact on tomato 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 thrips of tomato at Rajendranagar, A.P. is on the decline with 2013 having the lowest incidence over 2015-2016. Approaches including machine learning techniques used for modeling incidence of thrips indicated varying performances. Empirically, RF and LR model outperformed MLR, SVR, RR and ER models. The same conclusion may be drawn from the result of D-M test. Utilizing disease-weather interactions has resulted improved models with higher prediction accuracy. The current models could form a part of prediction for future seasons and for estimating scenario of thrips 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

Alam T, Tanweer G, Goyal GK. 2007. Stewart Postharvest Review, Packaging and storage of tomato puree and paste. Research article. **3(5),** 1-8.

Ammar ED. 1994. Propagative transmission of plant and animal viruses by insects: factors affecting vector specificity and competence. *Advances in* Disease Vector Research **10,** 289-331.

Badnakhe MR, Durbha SS, Jagarlapudi A, Gade RM. 2018. Evaluation of Citrus Gummosis disease dynamics and predictions with weather and inversion-based leaf optical model. Computers and Electronics in Agriculture **155,** 130-141.

Breiman L. 2001. Random forests. Machine Learning **45(1),** 5–32.

Chatterje S, Hadi AS. 2012. Regression Analysis by Example, John Wiley & Sons, Inc, New York.

Chowdappa P. 2010. Impact of climate change on fungal diseases of Horticultural crops: In: Challenges of climate change-Indian Horticulture (Eds.: H.P. Singh, J.P. Singh and S.S. Lal). Westville publishing house, New Delhi. 144-15.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics **13,** 253-263.

Efron B, Hastie T, Johnstone I, Tibshirani R. 2004. Least angle regression. The Annals of Statistics. **32,** 407–499.

FAOSTAT 2017. Global tomato production. Rome, FAO.

German TL, Ullman DE, Moyer JW. 1992. Tospoviruses: diagnosis, molecular biology, phylogeny, and vector relationships. Annual Review of Phytopathology **30,** 315–348.

German TL, Ullman DE, Moyerm JW. 1992. Tospoviruses: diagnosis, molecular biology, phylogeny, and vector relationships. Annual Review of Phytopathology **30,** 315–348.

Goldbach R, Peters D. 1994. Possible cause of the emergence of tospovirus diseases. Seminars in Virology **5,** 113–120.

Grandillo S, Zamir D, Tanksley SD. 1999. Genetic improvement of processing tomatoes: A 20 years perspective. *Euphytica*. **110,** 85–97.

Harvey CA, Saborio-Rodríguez M, Martinez-Rodríguez MR, Viguera B, Chain-Guadarrama A. 2018. Climate change impacts and adaptation among smallholder farmers in Central America. Agriculture and Food Security **7(1),** 1–20.

Li YH, Xu JY, Tao L, Li XF, Li S, Zeng X, Prot SVM.-2016. A web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PloS one*. **11(8),** e0155290.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News. **2,** 18-22.

Malau S, Lumbanraja P, Pandiangan S, Tarigan JR, Tindaon F. 2018. Performance of Coffea arabica L In Changing Climate of North Sumatra of Indonesia. Scientia Agriculturae Bohemica **49(4),** 340–349. https://doi.org/10.2478/sab-2018-0041.

Murphy FA, Fauquet CM, Bishop PHL. Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD. 1995. Virus taxonomy. Sixth report of the international committee on taxonomy of viruses. Archives of Virology **(10),** 313–314.

Nault LR. 1997. Arthropod transmission of plant viruses: a new synthesis. Annals of Entomological Society of America **90,** 521–541.

Paul RK, Vennila S, Singh N, Chandra P, Yadav SK, Sharma OP, Sharma V, K Nisar S, Bhat MN, Rao MS, Prabhakar M. 2018. Seasonal Dynamics of Sterility Mosaic of Pigeonpea and its Prediction using Statistical Models for Banaskantha Region of Gujarat, India. Journal of the Indian Society of Agricultural Statistics **72,** 213-223.

Reddy M, Reddy DVR, Appa Rao A. 1968. A new record of virus disease on peanut. Plant Disease Reporter **52,** 494-5.

Riley D, Pappu H. 2004. Tactics for management of thrips (Tysanoptera: Tripidae) and Tomato spotted wilt virus in tomato. Journal of Economic Entomology **97,** 1648–1658.

Sahoo S, Ta R, Elliott J, Foster I. 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. *Water* Resources Research **53(5),** 3878–3895.

Sakimura K. 1961. Field observations on the thrips vector species of the tomato spotted wilt virus in the San Pablo area, California. Plant Disease Reporter **45,** 772-776.

Shekoofa A, Emam Y, Shekoufa N, Ebrahimi M, 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(5),** e97288.

Staford CA, Walker GP, Ullman DE. 2011. Infection with a plant virus modifes vector feeding behavior. Proceedings of the National Academy of Sciences of the United States of America **108,** 9350–9355, https://doi.org/10.1073/pnas.1100773108.

Tibshirani R. 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society **58(B),** 267–288.

Todd JM, Ponniah S, Subramanyam CP. 1975. First record of tomato spotted wilt virus from Nilgiris in India. Madras Agricultural Journal **2,** 162-3.

Vapnik VN. 2000. The Nature of Statistical Learning Theory. Springer- Verlag, New York.

Verhage FYF, Anten NPR, Sentelhas PC. 2017. Carbon dioxide fertilization off sets negative impacts of climate change on Arabica coffee yield in Brazil. Clim Chang **144(4),** 671–685. https://doi.org/10.Journalpone.0211508.

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. B67 **(2),** 301–320.