The Pharma Innovation Journal 2022; SP-11(6): 2974-2984

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(6): 2974-2984
© 2022 TPI
www.thepharmajournal.com
Received: 19-04-2022
Accepted: 21-05-2022

Manjari Singh
Department of Agrometeorology, GB Pant University of Agriculture & Technology, Pantnagar, India

Parul Setiya
Department of Agrometeorology, GB Pant University of Agriculture & Technology, Pantnagar, India

Anand Kumar Tewari
Department of Plant Pathology, GB Pant University of Agriculture & Technology, Pantnagar, India

Ajeet Singh Nain
Department of Agrometeorology, GB Pant University of Agriculture & Technology, Pantnagar, India

Corresponding Author
Manjari Singh
Department of Agrometeorology, GB Pant University of Agriculture & Technology, Pantnagar, India

Comparison of machine learning and regression approaches to forecasting *Alternaria blight* epidemic of Indian mustard

Manjari Singh, Parul Setiya, Anand Kumar Tewari and Ajeet Singh Nain

Abstract

In the present investigation, weather-based prediction models have been developed for predicting epidemic characteristics of Alternaria blight of mustard. Models used in the study are ENET, LASSO, Ridge, and ANN which have been developed using 14 years (2006 to 2020) of epidemics data on (1) crop age at first appearance of *Alternaria blight*, (2) crop age at highest disease severity, and (3) highest disease severity in a growing season. Models were trained with 70% of data (2006- 2016) and remaining 30% data (2017-2020) were used for testing the model. Performance evaluation was done using R^2, root mean square error (RMSE), normalized root mean square (nRMSE), Mean Biased Error (MBE), and modeling efficiency (EF). Results indicate that models performed well at the calibration stage for all variables at all sowing dates. However, at validation stage, ANN derived models gave excellent results (R^2val = 1.00, nRMSEV ~ 0.00 to less than 5, and MBEV less than 1 in most cases), and LASSO derived models gave satisfactory results. Evaluation metrics (including R^2val, nRMSEV, and MBEV) suggested that ENET- and Ridge-derived models do not perform satisfactorily, whereas ANN-derived models yielded reliable results.

Keywords: Alternaria blight, disease forecasting, weather data, ANN, LASSO, ENET, Ridge, model evaluation metrics

1. Introduction

Rapeseed mustard, currently being grown in more than 60 countries globally and occupying second position in the world after soyabean, and third position in India after soyabean and groundnut, represents closely related species of cultivated oilseed crops. These include *Brassica juncea*, *B. rapa*, *B. napus*, *B. carinata* and *Eruca sativa*; all belonging to family *Brassiceae* (*Cruciferae)*. B. juncea also known as Indian Mustard, is generally grown in marginal and sub-marginal lands, and accounts for more than 90% of the area under rapeseed-mustard in India. It is grown either as a pure crop or as a mixed crop with other rabi cereals and pulses. In India, it is commonly grown as an intercrop or a mixed crop. Indian mustard is an annual herbaceous plant of 4-5 months duration, sown in the month of October-November and harvested in March.

Alternaria brassicae (Berk.) Sacc. causing Alternaria blight of mustard has been reported to cause significant yield losses across the world. It has been reported from many countries where Brassica is grown. To name a few, it has been reported from India (Chattopadhyay *et al*. 2005; Sangeetha and Siddaramaiah, 2007; Singh *et al*. 2014; Singh *et al*. 2022) [3, 28, 32-33], Bangladesh (Meah *et al*. 2002; Rahman *et al*. 2020) [21, 26], Nepal (Shrestha *et al*. 2005) [31], Pakistan (Ahmad and Ashraf, 2016) [1], Estonia (Runno-Paurson *et al*. 2021) [27], Canada (Conn *et al*. 1990) [5], Australia (You *et al*. 2005; Al-Lami *et al*. 2019) [38, 2], Russia (Serdyuk *et al*. 2021) [30]. Alternaria blight of mustard appear on leaves as well as on pods and cause shriveling of the developing pod (Yadav, 2011) [37] resulting in poor quality and quantity of oil (Meena *et al*. 2011) [22] lowered 1000-grain weight (Kolte *et al*. 1986) [15] and thus significant yield reduction. Depending on the severity, yield losses have been reported up to 47% in Indian mustard (Kolte, 1984) [14], even exceeding 70% in some *Brassica* species (Kumar *et al*. 2014), and more than 30% in canola (Conn *et al*. 1990) [5] in Canada.

In order to control or minimize the damage, a prior knowledge/information of the disease characteristics such as (1) expected time when the disease can appear, (2) important weather variables important for favoring pathogen multiplication and dispersal, (3) highest severity attainable under a given set of weather (4) and the yield as influenced by weather and disease, can be used in decision making and recommending appropriate necessary measures if the severity is expected to go above the threshold.

Weather has an important bearing on the growth and survival of all living forms present on earth. Most of the systems in nature are semi-open, where weather variables modify the rate of reaction of the processes and subprocesses, without being a part of that system. Process-based mechanistic simulation models have the potential to help us understand the underlying processes, why the system behaves in a particular way, and what will happen if the parameter's value is changed. However, they demand some very specific information to provide us with outputs that have some biological meaning. Alternatively, empirical statistical models provide us with an output by establishing relationship between the given dependent and independent variables. They have been widely used because of their simple structure and less data requirement.

In agriculture, statistical models have been widely used for forecasting yield of certain crops (Panwar *et al*. 2018; Srivastava *et al*. 2022) [34, 24], diseases (Kumar *et al*. 2013; Goswami *et al*. 2014) [8], and insect-pests (Narayanasamy *et al*. 2017; Jayakumar and Rajavel, 2019) [23, 12] of economic importance. However, complexity of the epidemiological processes, plant growth and their interaction with weather limits the forecasting skill of these statistical models. With the increasing awareness about the application of ICT technology in agriculture, focus of present-day research has shifted more towards the use of machine learning and AI techniques. Techniques such as Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), Long-term Short memory (LSTM) etc. have been utilized for the prediction of plant diseases: Septoria Leaf Blotch and Stripe Rust in Wheat (El Jarroudi *et al*. 2020) [7]; Rice blast (Malicdem and Fernandez, 2015; Kim *et al*. 2018) [20, 13]; grape diseases (Sannakki *et al*. 2013; Deshmukh *et al*. 2018; Chavan *et al*. 2019) [29, 6, 4] etc.

Therefore, in the present study, an attempt was made to investigate the potential of different regression approaches such as ENET, LASSO and Ridge, and machine learning approach of Artificial Neural Network (ANN). In order to test the usefulness of the model, the entire dataset was partitioned into training and testing datasets, in a ratio of 70:30, both independent of each other. Subsequently, statistical techniques were employed to evaluate the performance of the model for different sowing dates and arrive at the best model for a given set of conditions. The model once calibrated and validated can be used for forewarning diseases of interest. The present study aims to (1) develop empirical statistical models using ANN, LASSO, Ridge and ENET for predicting disease characteristics of *Alternaria blight* of mustard, (2) evaluate the performance of different statistical models, (3) establish relationship of disease characteristics with weather indices (using best performing model).

2. Materials and Methods

2.1 Data collection

Data for disease severity of *Alternaria blight* recorded on the mustard crop during 2006-2021 (excluding 2011 for which data is missing) were collected from the Dept. of Plant pathology, GBPUAT, Pantnagar, Uttarakhand India. Mustard plants were planted at the weekly interval, starting from Oct 1st to Nov. 19th each year. In total, we have eight sowing dates representing different climatic conditions available during the growing season. The date of observation and the number of observations taken during a particular year differed during different years. All plots were sampled on each observation date during a given year.

Data were analyzed to extract information on three parameters characterizing the disease incidence and its progression over time. These were crop age at the first appearance of disease (P1), crop age at highest severity (P2), and highest severity during the growing season (P3). The dataset was then statistically analyzed for the mean and standard deviation to see how it is varying over years.

Weekly data on weather variables starting from the week of sowing until the harvest was collected for the period 2006-2021 from the Dept. of Agrometeorology, GBPUAT, Uttarakhand, India. It included maximum temperature (TMAX), minimum temperature (TMIN), Relative humidity (I), Relative humidity (II), Bright Sunshine hours and Rainfall, and Potential Evapotranspiration (PET). Each weather variable does not have an equal weightage in all cases. So, in our study, weather indices have been computed and were used as predictors for model development (Panwar *et al*., 2016) [25]. All variables included in our analysis have been summarized in Table 1. These were used to establish a statistical relationship between disease severity and weather, and also for issuing forewarning of the variables of interest i.e. P1, P2, and P3.

Table 1: Unweighted and weighted weather indices for the development of multivariate models

Parameter	Unweighted Weather indices	Weighted Weather indices
Tmax	Z_{10}	Z_{11}
Tmin	Z_{20}	Z_{21}
Rainfall	Z_{30}	Z_{31}
Solar Radiation	Z_{40}	Z_{41}
Relative Humidity I	Z_{50}	Z_{51}
Relative Humidity II	Z_{60}	Z_{61}
PET	Z_{70}	Z_{71}
Tmax*Tmin	Z_{120}	Z_{121}
Tmax*Rainfall	Z_{130}	Z_{131}
Tmax*Solar Radiation	Z_{140}	Z_{141}
Tmax* Relative Humidity I	Z_{150}	Z_{151}
Tmax*Relative Humidity II	Z_{160}	Z_{161}
Tmax*PET	Z_{170}	Z_{171}
Tmin*Rainfall	Z_{230}	Z_{231}
Tmin*Solar Radiation	Z_{240}	Z_{241}

Tmin* Relative Humidity I	Z_{250}	Z_{251}
Tmin*Relative Humidity II	Z_{260}	Z_{261}
Tmin*PET	Z_{270}	Z_{271}
Rainfall*Solar Radiation	Z_{340}	Z_{341}
Rainfall* Relative Humidity I	Z_{350}	Z_{351}
Rainfall*Relative Humidity II	Z_{360}	Z_{361}
Rainfall*PET	Z_{370}	Z_{371}
Solar Radiation * Relative Humidity I	Z_{450}	Z_{451}
Solar Radiation *Relative Humidity II	Z_{460}	Z_{461}
Solar Radiation*PET	Z_{470}	Z_{471}
Relative Humidity I *Relative Humidity II	Z_{560}	Z_{561}
Relative Humidity I*PET	Z_{570}	Z_{571}
Relative Humidity II*PET	Z_{670}	Z_{671}

2.2 Data analysis

2.2.1 Computation of weather indices (Weighted and Unweighted) and detrending time-series data

For each weather variable two indices have been developed i.e., (1) simple total of values of weather variables in different weeks, and (2) weighted total, where weights are assigned based on the correlation coefficients between variable to forecast and weather variable in respective weeks. These indices are computed as follows:

$$Z_{ij} = \sum_{w=1}^{n} X_{iw} \; , \; Z_{ii'j} = \sum_{w=1}^{n} X_{iw} X_{i'w} \qquad \text{(i),}$$

$$Z_{ij} = \sum_{w=1}^{n} r_{iw}^{j} X_{iw} \; , \; Z_{ii'j} = \sum_{w=1}^{n} r_{ii'w}^{j} X_{iw} X_{i'w} \qquad \text{(ii)}$$

Here, n is the week of forecast, $X_{iw} / X_{i'w}$ is the value of i^{th}/ i'^{th} weather variable and $r_{iw}^{j} / r_{ii'w}^{j}$ is the value of correlation coefficient of detrended variable of forecast with i^{th} weather variable/ product of i^{th} and i'^{th} weather variables in w^{th} week. By following the above procedure, 56 weather indices have been generated and are presented in Table 1.

In the present study, simple linear regression has been applied to detrend variables of interest using least square method.

$Y_t = \beta_0 + \beta_1 t$

Where, Y_t is the variable of interest (P1, P2, P3); time t is the predictor, and β_0 and β_1 are the coefficients. If the result comes out to be insignificant, the residuals of the model were used for indices calculation. This shows that time has an effect. Otherwise, we use the same time series data for indices calculation when $p > 0.05$, i.e., the result comes out to be significant.

2.2.2 Multivariate techniques

In the present study, multivariate analysis methods such as LASSO, ENET, Ridge, and ANN have been used to give forewarning. The model has been trained with 70% dataset and the remaining 30% of the dataset from subsequent years were used for validation purpose. Out of the 14-year dataset available with us, 2006-2016 was used for training the model, and an independent set of 2017-2020 was used to test performance of the model. Details of these methods are given as follows:

Ridge regression is a method of estimating the coefficients of multiple regression models where linearly independent variables are highly correlated (Hilt and Seegrist 1977) [9]. In this case, the coefficients derived from a least square regression are brought closer to zero by multiplying by a constant (shrinkage factor). Also, it retains all predictors in the final model by imposing different penalties. On the contrary, in the case of LASSO regression, coefficients are brought closer by adding or subtracting a constant. It performs both variable selection and regularization in order to enhance the prediction accuracy (Tibshirani 1996) [35].

ENET regression models combine both the LASSO and Ridge regression by learning from the models' shortcomings to improve the performance of the final model (Zou and Hastie 2005) [39]. The two parameters, namely lambda and alpha, are needed to be optimized. The alpha is fixed at 0 for ridge regression and 1 for LASSO regression. In ENET the alpha can take any value between 0 and 1. In the present study, the 'glmnet' package was used for implementing LASSO and ENET in R software version 3.6.1

Artificial Neural Network models have been designed to simulate the method of information processing by a human brain, involving pattern recognition, establishing relationships, and learning through experience, and not from programming (Kustrin and Beresford 2000) [19]. The principal issue in the usage of ANN is to find the ideal number of hidden neurons or nodes. In the present study, the number of hidden nodes is selected by the "train" function of the "caret" package using the method "nnet" with 10-fold cross-validation in R software version 3.6.1 (Kuhn 2008) [16].

2.2.3 Model Performance evaluation

The performance of statistical models was evaluated using R^2, root mean square error (RMSE), normalized root mean square (nRMSE), Mean Biased Error (MBE), and modeling efficiency (EF). R^2 values nearer to 1 and RMSE values close to 0 indicate better model performance. The developed model is considered as excellent, good, fair and poor if values of nRMSE lies in the range of <10%, 10–20%, 20–30%, and >30% respectively (Jamieson *et al*. 1991). In the present study, all evaluation metrics were computed in R software using the 'applyStats' function in "tdr" package.

3. Results and Discussion

Mustard is a rabi season crop that is grown mostly as a rainfed crop. Weather is a well-established parameter that affects its production potential in different years and therefore is widely used in yield forecasting models. However, yield is often moderated as a result of the prevalence of insect-pests and diseases during the growing season. It is something that is most often ignored in statistical models, the reason may be the unavailability of these datasets of good quality, or increased complexity of the model. In the present study, we have attempted to test the ability of different models in predicting disease characteristics.

3.1 Cross-comparison of models in their predictions of disease severity characteristics

3.1.1 Crop age at first disease appearance (P1)

3.1.1.1 Alternaria blight infection on leaves

Wide variation in model performances was found with respect to the prediction of first disease appearance (P1) of *Alternaria blight* infection on leaves. ANN and LASSO models yielded satisfactory results for late sowing dates whereas, the performance of ENET and Ridge were poor. R^2 values at calibration were in nearly all cases very high for all sowing dates (Table 2): ENET: 0.85 to 1.00; Ridge: 0.85 to 0.93; LASSO: 0.56 to 0.98; and ANN: 1.00. There was wide variation in the R^2 values at validation: 0.00 to 0.85; Ridge: 0.00 to 0.99; LASSO: 0.06 to 0.99; ANN: 1.00. nRMSE values at validation were fair to good for ENET, and Ridge, fair to excellent for LASSO, and excellent for ANN at all sowing dates. Ranges of additional criteria are listed in Table 2.

Table 1: Statistical measures derived for different models in predicting crop age at first appearance of *Alternaria blight* (P1) on mustard leaves and pods

	Model	Range of values (Min - Max)*	R^2cal	nRMSEC	MBEC	RMSEC	R^2val	nRMSEV	MBEV	RMSEV	Predicted value
AB severity on leaves	ENET	Max	0.85	0.84	0.00	0.66	0.00	14.07	5.12	7.14	56.42
		Min	1.00	16.04	0.00	9.35	0.85	57.19	23.46	32.03	86.48
	Ridge	Max	0.85	10.17	0.00	6.50	0.00	14.07	5.12	7.14	56.42
		Min	0.93	16.04	0.00	9.35	0.99	52.02	24.85	29.13	80.54
	LASSO	Max	0.56	3.66	0.00	1.93	0.06	2.16	0.70	1.11	51.97
		Min	0.98	24.56	0.00	18.12	0.99	23.56	11.05	14.78	73.79
	ANN	Max	1.00	0.02	-0.30	0.01	1.00	0.04	-0.03	0.02	50.62
		Min	1.00	2.31	0.20	1.66	1.00	1.32	0.40	0.83	76.39
AB severity on pods	ENET	Max	1.00	0.55	0.00	0.53	0.10	13.95	-24.15	13.01	85.9
		Min	1.00	1.06	0.00	0.98	0.95	23.17	-12.73	26.24	109.9
	Ridge	Max	0.89	7.28	0.00	6.52	0.04	12.00	-22.04	10.05	87.3
		Min	0.92	8.37	0.00	8.77	0.98	20.22	-9.59	22.90	107.6
	LASSO	Max	0.72	0.79	0.00	0.87	0.01	0.47	-4.35	0.57	80.0
		Min	1.00	11.99	0.00	11.01	0.98	4.72	-0.47	4.50	123.9
	ANN	Max	0.97	0.03	-0.86	0.03	0.94	0.01	-0.50	0.01	50.5

3.1.1.2 Alternaria blight infection on pods

Wide variation in model performances was found with respect to the prediction of variable "P1" of *Alternaria blight* infection on pods. This was especially true for late sowing dates (DOS4 and beyond). In particular, most of the models yielded satisfactory results for late sowing dates, except Ridge. R^2 values at calibration were in nearly all cases very high for all sowing dates (Table 2): ENET: 1.00; Ridge: 0.89 to 0.92; LASSO: 0.72 to 1.00; and ANN: 0.97 to 1.00. There was wide variation in the R^2 values at validation: 0.10 to 0.95; Ridge: 0.04 to 0.98; LASSO: 0.01 to 0.98; ANN: 0.94 to 1.00. nRMSE values at validation were fair to good for ENET, and Ridge, while they were excellent at all sowing dates for LASSO and ANN. Cross comparison revealed that ANN is the best model for predicting P1. Ranges of additional criteria are listed in Table 2.

3.1.2 Crop age at maximum disease severity (P2)

3.1.2.1 Alternaria blight infection on leaves

Wide variation in model performances was found with respect to prediction of crop age at which maximum disease severity is attained (P2). This was especially true again for late sowing dates (DOS6 and beyond). In particular, LASSO and ANN models yielded satisfactory results for late sowing dates. R^2 values at calibration were in nearly all cases very high for all sowing dates (Table 3: ENET: 0.90 to 1.00; Ridge: 0.82 to 0.92; LASSO: 0.85 to 0.97; and ANN: 0.97 to 1.00. R^2 values at validation were acceptable, except for few sowing dates. These were: ENET: 0.00 to 0.98; Ridge: 0.17 to 0.91; LASSO: 0.74 to 0.99; ANN: 1.00. nRMSE values were fair to good for ENET, Ridge, and LASSO models, while they were excellent at all sowing dates for ANN. Ranges of additional criteria are listed in Table 3.

Table 2: Statistical measures derived for different models in predicting crop age at highest disease severity (P2) of Alternaria blight

	Model	Range of values (Min - Max)*	R^2cal	nRMSEC	MBEC	RMSEC	R^2val	nRMSEV	MBEV	RMSEV	Predicted value
AB severity on leaves	ENET	Max	0.90	0.35	0.00	0.39	0.00	8.39	-7.62	10.13	48.56
		Min	1.00	4.71	0.00	4.97	0.98	41.14	-17.29	21.08	117.28
	Ridge	Max	0.82	4.04	0.00	4.23	0.17	7.98	-23.56	9.63	114.40
		Min	0.92	6.69	0.00	7.21	0.91	20.35	-8.72	25.49	119.77
	LASSO	Max	0.85	2.36	0.00	2.61	0.74	1.73	-6.57	2.13	102.17
		Min	0.97	6.02	0.00	6.83	0.99	8.45	-1.52	10.73	135.44
	ANN	Max	1.00	0.07	-0.21	0.07	1.00	0.31	-4.12	0.40	88.52
		Min	1.00	0.69	0.11	0.78	1.00	6.63	-0.20	8.16	129.92
AB severity on pods	ENET	Max	0.89	0.90	0.00	0.79	0.00	12.83	-32.96	17.23	117.7
		Min	1.00	4.81	0.00	6.24	0.69	27.31	-8.92	36.46	150.0
	Ridge	Max	0.89	4.09	0.00	5.24	0.01	12.82	-10.42	17.23	115.2
		Min	0.93	5.46	0.00	6.24	0.98	27.80	-34.06	37.11	140.1

	LASSO	Max	0.57	6.19	0.00	8.05	0.53	7.53	-11.32	10.11	123.3
		Min	0.77	11.00	0.00	13.69	0.73	12.29	-3.04	16.71	131.0
	ANN	Max	0.97	0.07	-0.56	0.08	0.80	0.16	-5.32	0.21	101.7
		Min	1.00	3.19	1.75	3.68	1.00	7.02	-0.12	9.48	151.5

3.1.2.2 Alternaria blight infection on pods
Wide variation in model performances was found with respect to prediction of crop age at which maximum disease severity is attained (P2). This was especially true again for late sowing dates (DOS6 and beyond). In particular, LASSO and ANN models yielded satisfactory results for late sowing dates. R^2 values at calibration were in nearly all cases very high for all sowing dates (Table 3: ENET: 0.89 to 1.00; Ridge: 0.89 to 0.93; LASSO: 0.57 to 0.77; and ANN: 0.97 to 1.00. R^2 values at validation were acceptable, except for few sowing dates. These were: ENET: 0.00 to 0.69; Ridge: 0.01 to 0.98; LASSO: 0.53 to 0.73; ANN: 0.8 to 1.00. Model performance in terms of nRMSE values were poor, good, and excellent for Ridge, ENET, and LASSO respectively with exceptions for late sowing dates DOS 7 and DOS8. ANN performed excellently for all sowing dates. Cross comparison revealed that ANN is the best model for predicting P2. Ranges of additional criteria are listed in Table 3.

3.1.3 Prediction of the highest level of disease severity (P3)
3.1.3.1 Alternaria blight infection on leaves
Wide variation in model performances was found as well with respect to prediction of highest level of disease severity (P3). In particular, LASSO and ANN models gave satisfactory results for late sowing dates. R^2 values at calibration were in nearly all cases very high for all sowing dates (Table 4: ENET: 1.00; Ridge: 0.84 to 0.96; LASSO: 0.82 to 1.00; and ANN: 1.00. R^2 values at validation acceptable, except for few sowing dates. These were ENET : 0.00 to 0.90; Ridge: 0.00 to 0.92; LASSO: 0.29 to 1.00; ANN: 1.00. nRMSE values were poor early and late swoing dates in case of ridge model, fair to good for ENET and LASSO, while they were excellent for all sowing dates for ANN. Cross comparison revealed that ANN is the best performing model. Ranges of additional criteria are listed in Table 4.

Table 3: Statistical measures derived for different models in predicting highest disease severity (P3) of *Alternaria blight*

	Model	Range of values (min - Max)*	R^2cal	nRMSEC	MBEC	RMSEC	R^2val	nRMSEV	MBEV	RMSEV	Predicted value
AB severity on leaves	ENET	Max	1.00	-70.11	0.00	0.55	0.00	16.37	-11.38	9.00	9.07
		Min	1.00	1.59	0.00	1.00	0.90	261.52	-5.90	20.30	62.33
	Ridge	Max	0.84	-432.65	0.00	6.17	0.00	20.84	-11.74	9.27	-1.07
		Min	0.96	16.44	0.00	9.03	0.92	292.10	-5.24	18.65	64.35
	LASSO	Max	0.82	4.25	0.00	0.45	0.29	4.72	-3.17	0.46	0.84
		Min	1.00	19.02	0.00	10.15	1.00	28.38	0.00	14.54	52.07
	ANN	Max	1.00	-145.40	-0.40	1.25	1.00	0.16	-0.09	0.09	2.20
		Min	1.00	3.01	0.18	2.07	1.00	3.80	-0.21	0.41	75.80
AB severity on pods	ENET	Max	1.00	1.61	0.00	0.60	0.03	60.08	-8.29	18.87	15.1
		Min	1.00	3.87	0.00	1.17	0.51	78.31	18.36	25.86	45.2
	Ridge	Max	0.86	13.25	0.00	5.36	0.00	51.23	-6.59	17.29	35.2
		Min	0.94	20.44	0.00	7.45	0.21	82.14	16.68	24.48	47.4
	LASSO	Max	0.49	12.73	0.00	5.37	0.01	20.04	-4.92	5.26	29.9
		Min	0.93	46.16	0.00	15.53	0.88	57.46	8.42	21.55	50.8
	ANN	Max	1.00	0.25	-0.33	0.09	0.94	0.27	-0.90	0.06	21.3
		Min	1.00	4.31	0.50	1.63	1.00	16.59	1.15	6.22	53.6

3.1.3.2 Alternaria blight infection on pods
Wide variation in model performances was found as well with respect to the prediction of highest level of disease severity (P3). In particular, none of the models yielded satisfactory results, except ANN. R^2 values at calibration were in nearly all cases very high for all sowing dates (Table 4: ENET: 1.00; Ridge: 0.86 to 0.94; LASSO: 0.49 to 0.93; and ANN: 1.00. R^2 values at validation were much lower: 0.03 to 0.51; Ridge: 0.00 to 0.21; LASSO: 0.01 to 0.88; ANN: 0.94 to 1.00. nRMSE values were poor for ENET, Ridge, and LASSO, while they were excellent for all sowing dates for ANN (except DOS1 and DOS3). Cross comparison revealed that ANN is the best model for predicting P3. Ranges of additional criteria are listed in Table 4.

3.2 Contribution of input variables
Figure 1 to 6 presents the relative contribution of predictors (weather variables and disease severity) to the prediction of variables P1-P3 for AB infection on leaves and pods. The vertical axis represents top ten input variables, plotting the highest contributing variable at the top and others below it in decreasing order of their contribution.

3.2.1 Crop age at first disease appearance (P1)
Figure 1 presents the weather variables which are having significant contribution to the prediction of onset of disease of *Alternaria blight* (P1) on leaves during different sowing dates. There is a wide variation in the percent contribution of top ten weather variables for different dates. It is more than 50% for a majority of the sowing dates (except DOS7 and DOS8). In particular, joint effect of maximum temperature (Tx) and morning relative humidity (RHm) is the highest contributing variable for all sowing dates (except DOS2, where joint effect of Tn x RHa has the highest contribution). Other weather variables contributing to the prediction of P1 are minimum temperature (Tn), afternoon relative humidity (RHa), solar radiation (RAD), and PET.

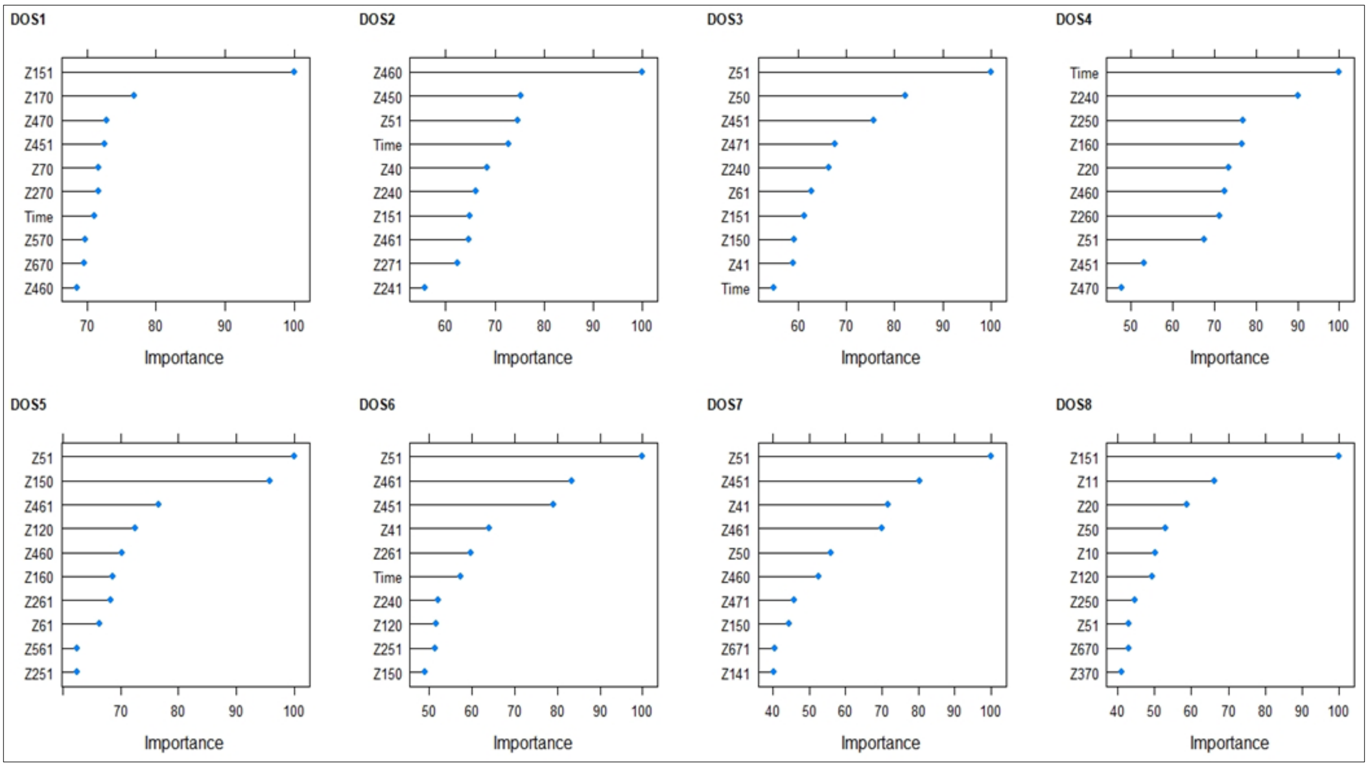

Fig 1: Important predictors for P1 (crop age at first appearance of AB disease) on mustard leaves

In case of AB infection on pods (Figure 2), there is wide variation in the highest contributing variable. Unweighted interaction effect of RAD and RHa (for DOS1 and DOS2), R and Tn (for DOS 3 and DOS5), Tx and RHm for DOS 4 and remaining late sowing dates, have the highest contribution as predictor of P1. Rainfall has come out to be an important predictor for DOS 3 and DOS5. The joint effect of rainfall with minimum temperature (R x Tn), R x RHa, R x Tx have more than 90% contribution as a predictor of P1 for these sowing dates. This indicates that appearance of AB infection on pods is primarily governed by the sunshine hours and resulting afternoon RH for early sowing dates. However, rainfall associated with western disturbance in the region is affecting disease severity on pods for crops sown in late October. Thereafter, for late sown crops i.e., in November and afterwards, pod formation occurs late in January, when temperature starts rising. So, maximum temperature and morning RH decides the level of severity on crops sown during this window. Details of other predictors are provided in figure 2.

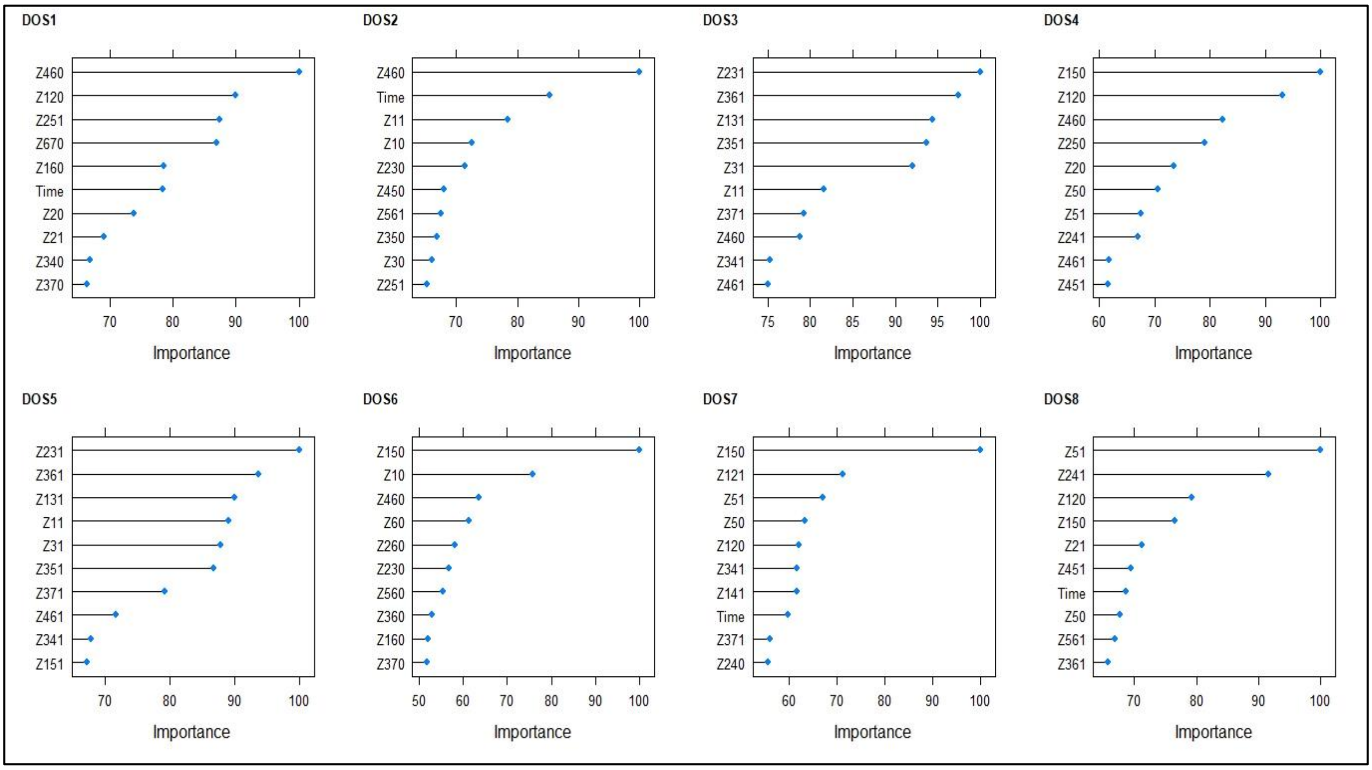

Fig 2: Important predictors for P1 (crop age at first appearance of AB disease) on mustard pods

3.2.2 Crop age at maximum disease severity (P2)

The relative contribution of different weather variables to the prediction of P2 i.e., crop age at maximum disease severity has been presented in Figure 3. The prediction of the time of maximum disease severity (P2) involved weighted and unweighted RHm for early sown crops. Results indicate that morning relative humidity (RHm) is a deciding factor for a majority of the sowing dates, either alone or jointly with maximum temperature and Sunshine hours. Sunshine hours is also an important predictor jointly with rainfall and humidity (both RHa and RHm). Rainfall is important in case of DOS 5 only. The percent contribution of different weather indices varies for all sowing dates.

In case of AB infection on pods (Figure 4), again, there is wide variation in the highest contributing variable. The weighted and unweighted joint effect of Tx and RHm (for DOS1 to DOS4), Tn with RAD and RHm (for DOS7 and DOS8) and weighted maximum temperature for DOS5 AND DOS6 have the highest contribution as the predictor of P2. Rainfall has come out to be an important predictor for DOS1, DOS2, and DOS5, DOS6. This indicates that crop age at highest pod infection is primarily governed by the morning RH for early sowing dates. Details of other predictors are provided in figure 4.

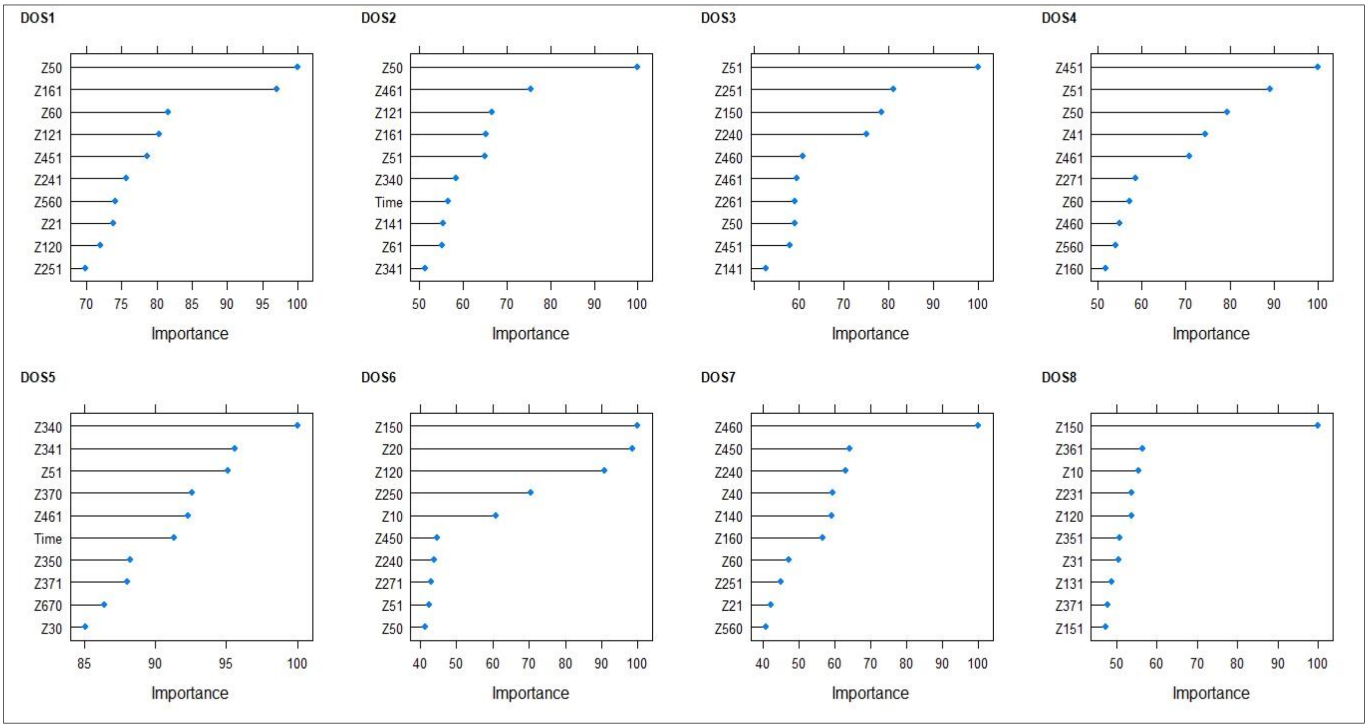

Fig 3: Important predictors for P2 (crop age at highest severity of AB disease) on mustard leaves

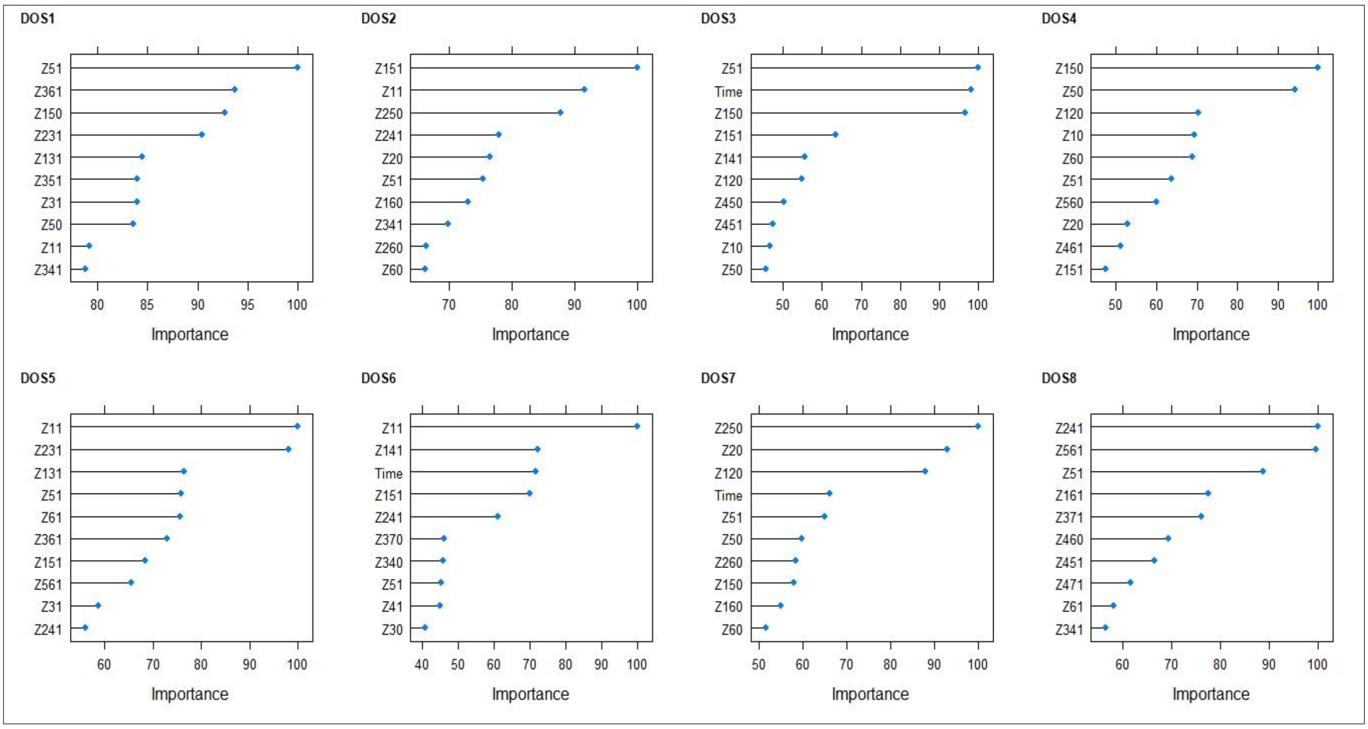

Fig 1: Important predictors for P2 (crop age at highest severity of AB disease) on mustard pods

3.2.3 Prediction of the highest level of disease severity (P3)
The relative contribution of different weather variables to the prediction of P3 i.e., highest disease severity has been presented in Figure 5. The prediction of the maximum disease severity in the growing season (P3) involved weighted and unweighted RHm for different sowing dates. Results indicate that joint effect of minimum temperature and PET is the highest contributing weather index for DOS 1 and DOS2. Joint effect of sunshine hours and humidity (RHm and RHa) is an important predictor for DOS 4,5 and 7. Results of important weather indices for predicting AB severity on pods (Figure 6) indicate that sunshine hours jointly with maximum temperature, humidity (RHa and RHm), and PET have very high contributions. Details of other predictors are provided in figure 5&6.

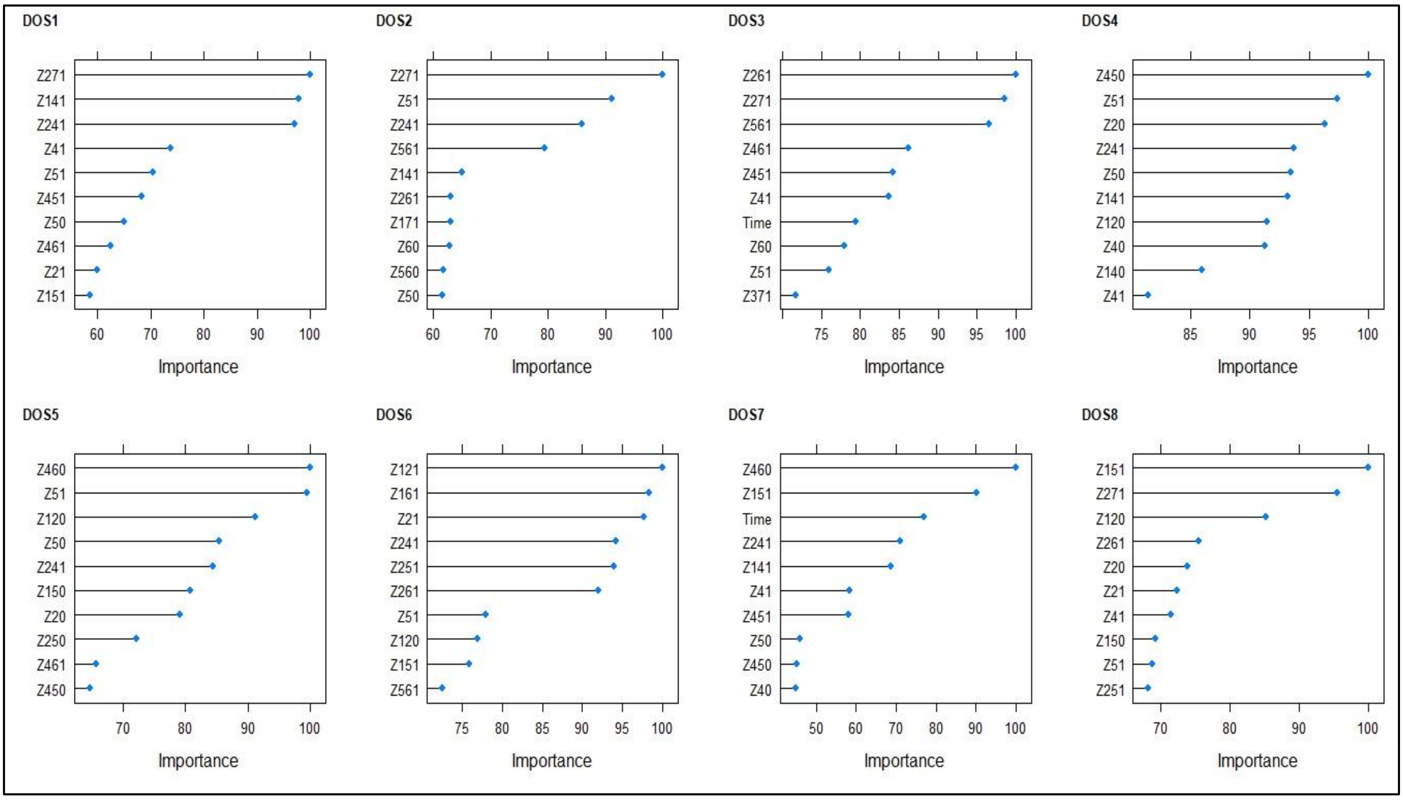

Fig 2: Important predictors for P3 (highest severity of AB disease) on mustard leaves

4. Discussion

Association between weather variables and plant disease severity has been reported previously for Alternaria blight (Shrestha *et al*. 2005; Fagodiya *et al*. 2022) [31]. The present study addresses the epidemics characteristics across and between growing seasons as influenced by the weather.

LASSO models are suitable for predicting P1 (crop age at first appearance of *Alternaria blight* on leaves), P2 (crop age at highest severity), and P3 (Highest disease severity) with high accuracy for most sowing dates. Our results further suggest that the weighted weather indices involving joint effect of weather variables are key predictors in these models. The performance of Ridge model is very variable. For P1, R^2 values at the validation stage is high for early sowing dates but its performance in terms of nRMSE is very poor. For variable P2, R^2 values at the validation stage are acceptable for most of the sowing dates (except DOS7 & DOS8), and model performance in terms of nRMSE value also falls in the "good" category. Therefore, Ridge model can be used for predicting crop age at highest severity of *Alternaria blight*. In case of variable P3, model performance in terms of R^2 and nRMSE values at the validation stage indicate that Ridge can be used with acceptable results only for a sowing window extending from SMW 43 (October, 22) to SMW 47 (November 19).

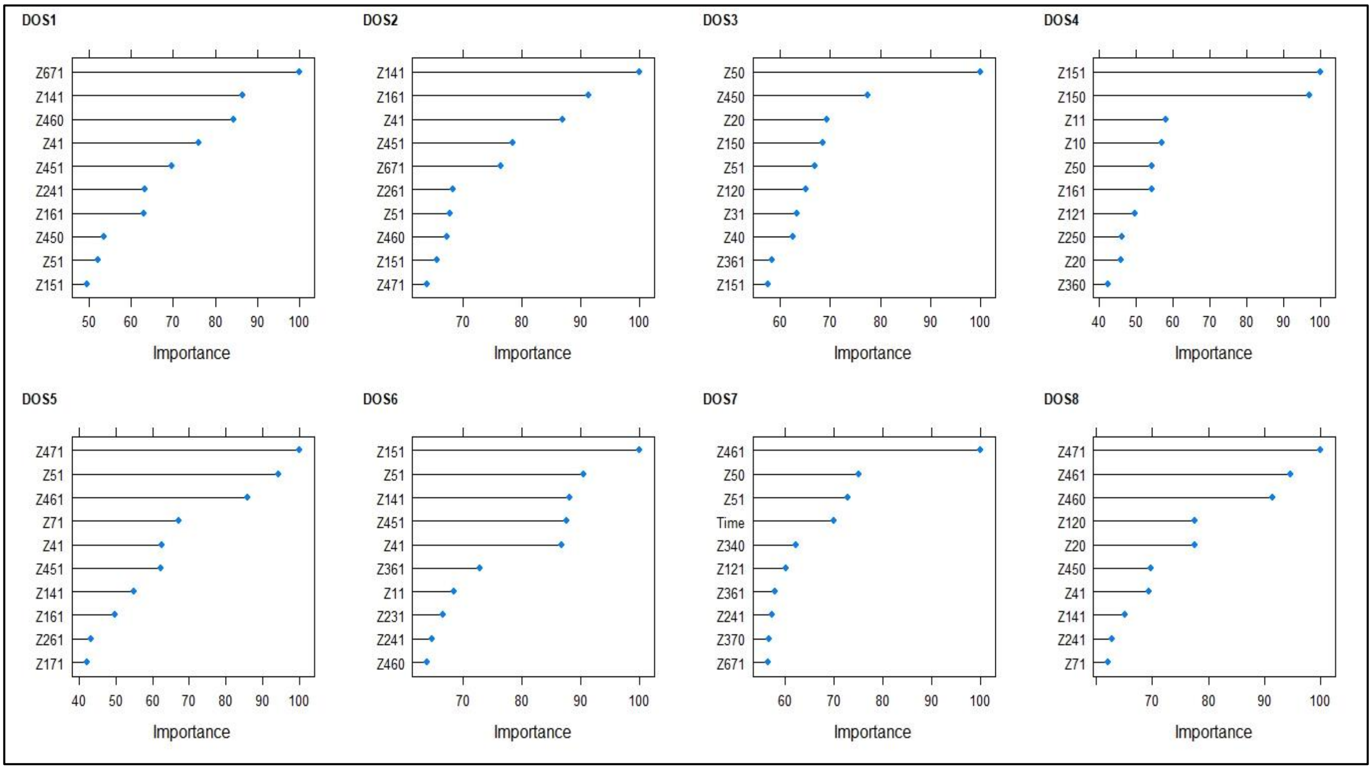

Fig 3: Important predictors for P3 (highest severity of AB disease) on mustard pods

The statistical measures derived from the ENET procedure indicate that models had been very well calibrated for all variables at all sowing dates. At validation stage, however, performances of models were poor for most of the sowing dates for all variables. Our results do not agree with the findings of Das *et al*. (2018) who used LASSO, ENET, ANN, and ANN-PCA for predicting rice yield, and found that LASSO followed by ENET to be generating best performing models while ANN-PCA was the worst model development procedure.

Cross comparison of models suggest that ANN-derived models perform very well for predicting all variables at all sowing dates. Therefore, the top ten predictors used by ANN models have been plotted according to their increasing contributions to the predicted variables (Gevrey *et al*. 2003) using the neural network approach. These are shown in Figure 1-6 for variables P1 to P3 for AB infection on leaf as well as on pod.

Prediction of the onset of epidemic of Alternaria using ANN reveals that the joint effect of temperature, humidity and sunshine hours were the most important predictor variable. This coincides with the findings of Chattopadhyay *et al*., 2005 [3], who reported that relative humidity, temperature and sunshine hours are important for the occurrence of Alternaria. Verma and Saharan, 1994 suggested that the germination of Alternaria conidium starts in free water and penetrates host tissues on a day with 25ºC temperature with a minimum of leaf surface wetness period of 6-16 hours for initiation of infection. Sporulation requires more than 90% of relative humidity (Humpherson-Jones and Phelps, 1989) [10]. The weighted and unweighted weather indices of temperature and humidity were other predictor variables of lower importance. Temperature and humidity have overwhelming significance in the prediction process for all sowing dates.

5. Conclusion

Four different multivariate methods were used to derive models for the prediction of epidemic characteristics of Alternaria blight on leaves as well as on pods. The variables to be predicted are (1) crop age at the first appearance of disease, (2) crop age at highest severity, and (3) highest disease severity in a growing season. Our results indicate that LASSO and ANN gave satisfactory results for variable P1 and P2 with respect to R^2 values and nRMSE, whereas Ridge and ANN performance was poor. However, the prediction of disease severity (P3) was found to be excellent for ANN only. Evaluation metrics for other models indicates poor performance for most of the sowing dates. In the present study, performance of ANN is excellent for all variables (representing epidemic characteristics) for most of the sowing dates followed by LASSO and ENET, whereas Ridge is not suggested for such studies. Most important predictor in terms of weather indices have also been computed and have been presented for the best performing model. Results indicate that weather indices computed using the temperature, humidity and sunshine hours are important predictors.

6. References

1. Ahmad A, Ashraf J. *In vitro* and *in vivo* Management of Alternaria Leaf Spot of Brassica campestris L. Plant Pathol Microbiol. 2016;7(7):365. DOI: 10.4172/2157-7471.1000365
2. Al-Lami HF, You MP, Barbetti MJ. Incidence, pathogenicity and diversity of *Alternaria* spp. associated with Alternaria leaf spot of canola (*Brassica napus*) in Australia. Plant Pathol. 2019;68(3):492-503
3. Chattopadhyay C, Agrawal R, Kumar A, Bhar LM, Meena PD, *et al*. Epidemiology and forecasting of Alternaria blight of oilseed Brassica in India - A case study. J Plant Dis Prot. 2005;112(4):351-365
4. Chavan R, Deoghare A, Dugar R, Karad P. IoT Based Solution for Grape Disease Prediction Using Convolutional Neural Network and Farm Monitoring. Int J Sci Res Eng Dev. 2019;2(3):494-500.

5. Conn KL, Tewari JP, Awasthiz RP. A disease assessment key for Alternaria blackspot in rapeseed and mustard. Can. Plant Dis. Surv. 1990;70:1.
6. Deshmukh N, Gaikwad A, Ghutukade P, Moogi D, Phadke M. Grape Disease Prediction Band Identification. Int. J res. appl. sci. eng. technol. 2018;6(V):2448-2451.
7. El Jarroudi M, Lahlali R, El Jarroudi H, Tychon B, Belleflamme A, Junk J. Employing Weather-Based Disease and Machine Learning Techniques for Optimal Control of Septoria Leaf Blotch and Stripe Rust in Wheat. Adv. Intell. Syst. Comput. 2020;1103:1-9.
8. Goswami P, Goyal R, Rao EVSP, Ramesh KV, Sudarshan MR, Ajay D. A weather-based forecast model for capsule rot of small cardamom. Curr. Sci, 2014, 107(6)
9. Hilt DE, Seegrist DW. Ridge, a computer program for calculating ridge regression estimates, 1977. Doi:10.5962/bhl.title.68934.
10. Humpherson-Jones FM, Phelps K. Climatic factors influencing spore production in Alternaria brassicae and *Alternaria brassicicola*. Annals of Applied Biology. 1989;114:449-458.
11. Jamieson PD, Porter, JR, Wilson DR. A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. Field Crop. Res. 1991;27:337-350.
12. Jayakumar M, Rajavel M. Weather based pest forewarning models for mealybug infestation in Robusta coffee (*Coffea canephora*). J Agromet. 2019;21(4):488-493
13. Kim Y, Roh JH, Kim HY. Early forecasting of Rice blast disease using long short-term memory recurrent neural networks. Sustainability. 2018;10:34.
14. Kolte S. Diseases of Annual Edible Oilseed Crops. Peanut Diseases. Boca Raton, FL, USA: CRC Press. 1984, I.
15. Kolte SJ, Awasthi RP, Vishwanath. Effect of planting dates and associated weather factors on staghead phase of white rust and downy mildew of rapeseed and mustard. Indian J Mycol Plant Pathol. 1986;16(2):94-102.
16. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat. Softw. 2014;28:1-26.
17. Kumar D, Maurya N, Bharati YK *et al*., Alternaria blight of oilseed brassicas: a comprehensive review. Afr. J Microbiol. Res. 2008;8:2816-29.
18. Kumar A, Agrawal R, Chattopadhyay C. Weather based forecast model for diseases in mustard crop. Mausam. 2013;64:663-670.
19. Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modelling and its application in pharmaceutical research. J Pharm. Biomed. Anal. 2000;22(5):717-27.
20. Malicdem AR, Fernandez PL. Rice blast disease forecasting for northern Philippines. WSEAS Trans. Inf. Sci. Appl. 2015;12:120-129.
21. Meah MB, Hau B, Siddiqua MK. Relationships between disease parameters of Alternaria blight (*Alternaria brassicae*) and yield of mustard. J Plant Dis Prot. 2002;109:(3):243-251.
22. Meena PD, Chattopadhyay C, Kumar A, Awasthi RP, Singh R, Kaur S. Comparative study on the effect of chemicals on Alternaria blight in Indian mustard- A multilocation study in India. J Environ Biol. 2011;32:375-379.
23. Narayanasamy M, Kennedy JS, Geethalakshmi V. Weather Based Pest Forewarning Model for Major Insect Pests of Rice – An Effective Way for Insect Pest Prediction. Annu. Res. Rev. Biol. 2017;21(4):1-13.
24. Panwar S, Kumar A, Singh KN, Paul RK, Gurung B, *et al.*, Forecasting of crop yield using weather parameters–two step nonlinear regression model approach. Indian J. Agric. Sci. 2018;88(10):1597-9.
25. Panwar S, Singh KN, Kumar A, Paul RK, *et al*., Performance evaluation of yield crop forecasting models using weather index regression analysis. Indian J. Agric. Sci. 2016;87(2):270–2.
26. Rahman MA, Fatema-Tuz-Zohura, Hamim I, Meah MB, Hossain MA. Sensitivity of Alternaria blight pathogen (*Alternaria brassicicola*) to fungicides and its effects on yield contributing parameters of mustard. Fundamental and Applied Agriculture. 2020;5(3):435–442. https://doi.org/10.5455/faa.120733
27. Runno-Paurson E, Lääniste P, Nassar H, *et al*. Alternaria Black Spot (*Alternaria brassicae*) Infection Severity on Cruciferous Oilseed Crops. Appl. Sci. 2021;11(18):8507. https://doi.org/10.3390/app11188507
28. Sangeetha CG, Siddaramaiah AL. Epidemiological studies of white rust, downy mildew and Alternaria blight of Indian mustard (*Brassica juncea* (Linn.) Czern. and Coss.). African J Agric Res. 2007;2(7):305-308
29. Sannakki S, Rajpurohit V, Sumira F, Venkatesh H. A neural network approach for disease forecasting in grapes using weather parameters. In Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 4–6 July 2013, 1–5.
30. Serdyuk O, Trubina V, Gorlova L. The effect of fungicides on reducing the harmfulness of Alternaria blight of brown mustard. IOP Conf. Series: Earth and Environmental Science. 2021;699:012018 doi:10.1088/1755-1315/699/1/012018
31. Shrestha SK, Munk L, Mathu SB. Role of weather on Alternaria Leaf Blight Disease and its effect on Yield and Yield Components of Mustard. Nepal Agric. Res. J. 2005;6:62-72.
32. Singh J, Kumar A, Chaubey AK. Evaluation and validation of disease management module for Alternaria Blight in Mustard. Environ Conserv. 2022. https://doi.org/10.36953/ECJ.0211096-2277
33. Singh D, Prasad R, Yadav SP, Kumar S. Characterization of Alternaria brassicae causing Alternaria blight of rapeseed- mustard (*Brassica juncea* L.) and it's management The Bioscan. 2014;9(3):1207-1212.
34. Srivastava AK, Safaei N, Khaki S, Lopez G, *et al*., Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. Sci. Rep. 2022;12:3215 https://doi.org/10.1038/s41598-022-06249-w
35. Tibshirani R. Regression Shrinkage and Selection via the lasso. Journal of the Royal Statistical Society. Series B (methodological). Wiley. 2011;58(1):267–88.
36. Verma PR, Saharan GS. Monograph on Alternaria Diseases of Crucifers. Research Branch. Agriculture and Agri Food Canada, 1994, 162.
37. Yadav MS, Gupta PP. Magnitude of losses in yield and management of Alternaria blight of rapeseed and mustard. Haryana Agric. Univ. J Res. 2011;17:14-18.
38. You MP, Simoneau P, Dongo A, *et al*. First Report of an

Alternaria Leaf Spot Caused by Alternaria brassicae on *Crambe abyssinica* in Australia. Plant Dis. 2005;89(4):430-430.
39. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R. Stat. Soc. Ser. B. 2005;67:301-320.