Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 24 (4) : 373-379 (December- 2022)

https://journal.agrimetassociation.org/index.php/jam

Research Paper

Multistage wheat yield prediction using hybrid machine learning techniques

SHREYA GUPTA[1], ANANTA VASHISTH[1*], P. KRISHNAN[1], ACHAL LAMA[2], SHIV PRASAD[1] and ARAVIND K.S.[1]

[1]*ICAR-Indian Agricultural Research Institute, Pusa, New Delhi - 110012*
[2]*ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi - 110012*
***Corresponding author e-mail**: ananta.iari@gmail.com*

ABSTRACT

Wheat being highly affected by the weather, adverse weather drastically reduces the wheat yield. Model was developed for multi stage wheat yield prediction by stepwise multi linear regression (SMLR), support vector regression (SVR), least absolute shrinkage and selection operator (LASSO) and hybrid machine learning LASSO-SVR and SMLR-SVR techniques. Wheat yield data and weather parameter for generating thermal and weather indices during different growth stage for more than 30 years were collected for study area. Analysis was carried out by fixing 70 % of the data for calibration and remaining 30 % dataset for validation in R software. Results showed that LASSO performed best having nRMSE value between 1.22 % at grain filling stage for IARI, New Delhi to 8.36 % for Hisar at flowering stage. The model performance of SVR is increased if a hybrid model in combination with LASSO and SMLR is applied. The hybrid model LASSO-SVR has shown more improvement than SVR model compared with SMLR-SVR.

Keywords: Weather variable, hybrid machine learning model, support vector regression, least absolute shrinkage and selection operator, stepwise multi linear regression, yield prediction

Wheat (*Triticum aestivum*) is one of the principal food crops of the country. Wheat production are significantly influenced and controlled by climatic factors such as rainfall, temperature, solar radiation, and relative humidity (Ji *et al*., 2007). Weather variability within the crop growing seasons is an important source of variability in yields. Weather variables affect the crop differently during different stages of development. Hence, predicting crop yield using weather variables is of foremost important. Crop yield forecast may be done by using biometrical characteristics, weather variables and agricultural inputs. These can be used individually or in combination (Agrawal *et al.,* 2001). Multiple linear regression has the biggest disadvantage of over-fitting when the number of samples is less than the number of variables. Also, another disadvantage is the multi-collinearity when independent predictors are correlated (Verma *et al*., 2016). To overcome these demerits, least absolute shrinkage and selection operator (LASSO), machine learning and hybrid machine learning technique can be used. Tibshirani (1996) proposed LASSO, which can be utilized in the crop yield forecasting technique. Aravind *et al.* (2022) reported that elastic Net and LASSO were found to be the best model followed by PCA-SMLR, SMLR, PCA-ANN and ANN respectively for wheat yield prediction of different locations of north-west India. Support vector machine is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fitting to the data. SVM model was used for crop yield forecast of barley, canola and spring wheat grown on the Canadian Prairies developed using vegetation indices derived from satellite data (Johnson *et. al.,* 2016). Alam *et. al.*, (2018) reported that the performance of the hybrid ARIMAX-SVM and ARIMAX-ANN model was superior than ARIMAX model for forecasting rice yield. In the present study multistage wheat yield prediction model was developed by stepwise multiple linear regression (SMLR), support vector machine (SVM), least absolute shrinkage and selection operator (LASSO). Model was also developed by two hybrid machine learning SMLR-SVR and LASSO-SVR techniques for improving the accuracy of multi stage wheat yield prediction for different location of north-west India.

Article info - DOI: *https://doi.org/10.54386/jam.v24i4.1835*
Received: 07 September 2022; Accepted: 06 October 2022; Published online : 1 December 2022

MATERIALS AND METHODS

Data collection

Daily Weather data during wheat crop growing period for last thirty to forty years were collected from the Regional met centre Chandigarh for Amritsar and Patiala, AMFU Ludhiana, AMFU Hisar and AMFU New Delhi. Wheat yield data was collected from the Directorate of Economics & Statistics (DES) and the state agricultural department. Reference evapotranspiration (ETo) was calculated by the FAO Penman-Monteith method. Different thermal indices (growing degree days, helio-thermal units, heat use efficiency and photo thermal index) were calculated from sowing up to harvest of the crop. Weather indices were developed using daily weather data (maximum temperature, minimum temperature, rainfall, morning relative humidity, evening relative humidity, bright sunshine hour, pan evaporation, reference evapotranspiration) during the crop growing period. For each weather variable, two types of weather indices were developed. The first one being the simple values of weather variables during the crop growing period [un-weighted index -Zi0] and the second one is weighted [weighted index Zi1] (Agrawal *et al*., 2001). Weights are taken as correlation coefficients between yield and weather variables in respective periods. In the same way, indices were also produced for interaction of weather variables by using weekly products of weather variables taking two at a time (Aravind *et al.,* 2022).

Development of models

For development of a multi-stage yield prediction model, weather indices were developed by weather parameters from 46 to 4th, from 46 to 8th and from 46 to 11th standard meteorological week for the yield prediction at tillering, at flowering and at grain filling stage respectively. Thermal and weather indices were used for developing a multi-stage wheat yield prediction model using SMLR, LASSO, SVR and hybrid machine learning techniques for five different locations, IARI, New Delhi, Hisar, Amritsar, Ludhiana and Patiala. R software (version 3.6.0) was used for developing the multistage wheat yield prediction model, package HDCI was used for LASSO and package e1071 was used for SVR, which are the inbuilt packages in R software (version 3.6.0). For hybrid machine learning combination of SMLR with SVR and LASSO with SVR was done. In the SMLR-SVR model, SMLR select variables from the data analysis and is used as an input variable for SVR. In the LASSO-SVR model, first variables are selected by LASSO techniques and these variables are used as an input variable for SVM. For data reduction, LASSO is a very efficient shrinkage technique. This technique reduces the number of regressors to be used in the SVM model and give precise estimation of wheat yield from the given set of observation. Models for predicting wheat yield at tillering, flowering and gain filling stage was developed using long-term crop yield data as well as weather and thermal indices developed from 46th to 4th, 46th to 8th and 46th to 11th standard meteorological week. Data used for model calibration, validation and prediction for different location are given in Table 1.

Model accuracy

The performance of statistical models was estimated by calculating, coefficient of determination (R^2), mean squared error (MSE), root mean square error (RMSE), normalized root mean square error (nRMSE) and percentage deviation using the following formula.

$$R2 = 1 - \frac{\sum_{i=1}^{N}(Pi-Oi)^2}{\sum_{i=1}^{N}(Pi-M)^2} \qquad MSE = \frac{1}{N}\sum_{i=1}^{N}(Pi-Oi)^2$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi-Oi)^2} \quad nRMSE = \frac{100}{M} * \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi-Oi)^2}$$

Percentage Deviation= (P_i-O_i) *100/ O_i,

where Pi, O_i, N and M are predicted value, observed value, number of observations and mean of observed value. nRMSE is considered excellent with the nRMSE value less than10 %, good if nRMSE value ranges between 10–20 %, fair if value ranged between 20–30 % and poor if value is more than 30 %.

RESULTS AND DISCUSSION

Multistage wheat yield prediction for IARI, New Delhi by different model

Performance of the models during calibration and validation for multistage wheat yield prediction for IARI are shown in Table 2. Model performance for wheat yield prediction done at tillering, at flowering and at grain filling stage was excellent for all the models having nRMSE < 10. At tillering stage nRMSE value during calibration was between 1.23 to 3.95 % and between 4.25 to 6.14 % during validation. During validation lowest value of nRMSE was for LASSO-SVR followed by LASSO, SVR, SMLR and SMLR-SVR. At flowering stage nRMSE value was less than 3 % and 7 % during calibration and validation respectively. During validation lowest value was found for LASSO followed by LASSO-SVR, SVR, SMLR and SMLR-SVR. At grain filling stage nRMSE value during calibration and validation was less than 5.92 and 6.63 % respectively. During validation lowest value was found for LASSO followed by LASSO-SVR, SVR, SMLR and SMLR-SVR. Based on model performance for wheat yield prediction done at tillering stage LASSO-SVR followed by LASSO was found to be best and at flowering and at grain filling stage LASSO followed by LASSO-SVR was found to be best for IARI, New Delhi.

Multistage wheat yield prediction by different models for Hisar

Performance for different models for wheat yield prediction done at tillering stage during calibration was excellent for SVR, having nRMSE value 7.87 % (Table 3). During validation models was excellent for LASSO, good for SVR and LASSO-SVR, fair for SMLR-SVR and SMLR. At flowering stage during calibration model performed excellent for SVR having nRMSE value 7.39 %, good for LASSO-SVR, SMLR, SMLR-SVR and LASSO having nRMSE value 16.26, 17.52, 17.77 and 18.92 % respectively. During validation the model performance was excellent for LASSO, good for SVR, LASSO-SVR and SMLR-SVR, fair for SMLR having nRMSE value 8.36, 13.56, 17.44, 19.72 and 21.66 respectively. At grain filling stage performance during calibration was excellent for SMLR-SVR and SVR having nRMSE value 7.39 and 7.56 %, good for LASSO-SVR, LASSO and SMLR having nRMSE value 15.45, 15.88 and 16.74 % respectively. During

Table 1: Period of data used for model calibration, validation and prediction for different locations

Particulars	IARI, New Delhi	Hisar	Amritsar	Ludhiana	Patiala
Model calibration	1985-2008	1985-2008	1971-2003	1972-2003	1971-2003
Validation	2009-2017	2009-2017	2004-2016	2004-2016	2004-2016
Prediction	2018	2018	2017	2017	2017

Table 2: Performance of wheat yield prediction done at multistage by different models for IARI, New Delhi

Model	Modal accuracy during calibration				Modal accuracy during validation		
	R^2 (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)
At tillering stage							
SMLR	0.89	18572	136.3	3.95	64701	254.4	6.00
SVR	0.99	1807	42.5	1.23	43512	208.6	4.92
LASSO	0.92	18050	134.4	3.84	35687	188.9	4.45
LASSO-SVR	0.92	17799	133.4	3.85	32469	180.2	4.25
SMLR-SVR	0.96	8304	91.1	2.63	67872	260.5	6.14
At flowering stage							
SMLR	0.95	9837	99.2	2.84	57830	240.5	5.67
SVR	0.99	1938	44.0	1.27	24918	157.9	3.72
LASSO	0.99	2297	47.9	1.37	3038	55.1	1.30
LASSO-SVR	0.99	1584	39.8	1.15	17157	131.0	3.09
SMLR-SVR	0.99	1983	44.5	1.27	83914	289.7	6.83
At grain filling stage							
SMLR	0.93	25539	159.8	4.57	72478	269.2	6.35
SVR	0.80	42230	205.5	5.92	36133	190.1	4.41
LASSO	0.98	3674	60.6	1.73	2685	51.8	1.22
LASSO-SVR	0.98	4612	67.9	1.94	6913	83.1	1.88
SMLR-SVR	0.99	1596	40.0	1.15	79052	281.2	6.63

validation model performance was excellent for LASSO, good for SMLR-SVR, LASSO-SVR and SVR, fair for SMLR. Based on model performance wheat yield prediction done at tillering and at flowering, LASSO followed by SVR was found to be best and at grain filling stage LASSO followed by SMLR-SVR was found to be best for Hisar.

Multistage wheat yield prediction by different models for Amritsar

Performance of different models for wheat yield prediction done at tillering stage during calibration was excellent for all the models having nRMSE value <10% (Table 4). During validation the value of nRMSE was between 4.91 to 11.26 %. The lowest value of nRMSE during validation was for LASSO (4.91 %) followed by, SMLR (8.13%), SVR (8.47%), LASSO-SVR (9.80%) and SMLR-SVR (11.26%). At flowering stage model performance during calibration was excellent for all models having nRMSE value between 3.77 and 7.98 % respectively. During validation performance of the model was excellent for LASSO, LASSO-SVR and SVR having nRMSE value 4.90, 9.19 and 9.29 %, good for SMLR-SVR and SMLR having nRMSE value 11.09 and 13.78 % respectively. At grain filling stage model performance during calibration was excellent for all the models having nRMSE value less than 6.63 %. During validation model performance was excellent for LASSO, LASSO-SVR, SMLR-SVR, good for SVR and SMLR. Based on model performance LASSO followed by SVR was found to be best for wheat yield prediction done at tillering, and LASSO followed by LASSO-SVR was found to be best for wheat yield prediction done at flowering and at grain filling stage for Amritsar.

Multistage wheat yield prediction by different models for Ludhiana

Performance of the model for wheat yield prediction done at tillering stage during calibration was excellent, having nRMSE value < 5% for all the model (Table 5). During validation model performed excellent for LASSO, LASSO-SVR and SMLR having nRMSE value 2.89, 9.73 and 9.94 % respectively, good for SVR and SMLR-SVR having nRMSE value 11.32 and 11.71 % respectively. At flowering performance of the model during calibration was excellent for all the models having nRMSE value between 1.95 and 4.93 %. During validation model performance was excellent for LASSO and LASSO-SVR having nRMSE value 4.10 and 8.88 %, good for SVR, SMLR and SMLR-SVR model having nRMSE value 10.37, 10.47 and 11.26 % respectively. At grain filling stage during calibration model performance was excellent for all the models having nRMSE value less than 4.66 %. During validation model performance was excellent for LASSO and LASSO-SVR, having nRMSE value 4.14 and 9.51 %, good for SVR, SMLR-SVR and SMLR model having nRMSE value 10.51, 11.35 and 11.54 % respectively. Based on model performance multistage wheat yield prediction was found to be best for LASSO followed by LASSO-SVR for Ludhiana.

Multistage wheat yield prediction by different models for Patiala

Performance of the model for multistage wheat crop yield prediction

Table 3: Performance of wheat yield prediction done at multistage by different models for Hisar

Model	Modal accuracy during calibration				Modal accuracy during validation		
	R^2 (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)	MSE (kg ha^{-1})	RMSE (kgha^{-1})	nRMSE (%)
At tillering stage							
SMLR	0.92	217192	466.0	18.78	977116	988.5	22.72
SVR	0.99	38198	195.4	7.87	532053	729.4	16.77
LASSO	0.93	226758	476.2	19.18	114795	338.8	7.79
LASSO-SVR	0.94	175617	419.1	16.88	669586	818.3	18.81
SMLR-SVR	0.92	221375	470.5	18.96	869873	932.7	21.44
At flowering stage							
SMLR	0.93	189122	434.9	17.52	888152	942.4	21.66
SVR	0.99	33677	183.5	7.39	347748	589.7	13.56
LASSO	0.94	220446	469.5	18.92	132326	363.8	8.36
LASSO-SVR	0.94	162953	403.7	16.26	575785	758.8	17.44
SMLR-SVR	0.93	194562	441.1	17.77	735610	857.7	19.72
At grain filling stage							
SMLR	0.93	172707	415.6	16.74	846468	920.0	21.15
SVR	0.99	35242	187.7	7.56	757158	870.2	20.00
LASSO	0.95	155359	394.2	15.88	66106	257.1	5.91
LASSO-SVR	0.94	147063	383.5	15.45	401087	633.3	14.56
SMLR-SVR	0.99	33677	183.5	7.39	347748	589.7	13.56

Table 4: Performance of wheat yield prediction done at multistage by different models for Amritsar

Model	Modal accuracy during calibration				Modal accuracy during validation		
	R^2 (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)	MSE (kg ha^{-1})	RMSE (kgha^{-1})	nRMSE (%)
At tillering stage							
SMLR	0.92	58738	242.4	7.21	129135	359.4	8.13
SVR	0.97	22369	149.6	4.45	140200	374.4	8.47
LASSO	0.92	76397	276.4	8.22	47102	217.0	4.91
LASSO-SVR	0.92	38033	195.0	5.80	187518	433.0	9.80
SMLR-SVR	0.92	41451	203.6	6.05	247680	497.7	11.26
At flowering stage							
SMLR	0.92	56365	237.4	7.06	370644	608.8	13.78
SVR	0.98	16038	126.6	3.77	168673	410.7	9.29
LASSO	0.92	72106	268.5	7.98	46892	216.5	4.90
LASSO-SVR	0.92	41333	203.3	6.04	164814	406	9.19
SMLR-SVR	0.91	44417	210.8	6.27	223156	472.4	11.09
At grain filling stage							
SMLR	0.95	34760	186.4	5.54	363639	603.0	13.64
SVR	0.99	6870	82.9	2.46	277126	526.4	11.91
LASSO	0.95	49690	222.9	6.63	33897	184.1	4.17
LASSO-SVR	0.96	15297	123.7	3.68	178062	422.0	9.55
SMLR-SVR	0.94	14856	121.9	3.62	184775	429.9	9.73

for Patiala during calibration and validation is shown in table 6. Model performance for wheat yield prediction done at tillering stage, during calibration was excellent, having nRMSE value < 10 % for all the models. During validation model performed excellent for LASSO, SVR and LASSO-SVR having nRMSE value 5.06, 8.18 and 9.69 % respectively and performed good for SMLR and SMLR-SVR having nRMSE value 11.37 and 12.17 % respectively. At flowering model performance during calibration was excellent for all the models having nRMSE value between 4.41 and 6.70 %. During validation performance was excellent for all the models having nRMSE value 4.94, 7.12, 7.51 and 8.83 % for LASSO, SVR, LASSO-SVR and SMLR-SVR respectively and good for SMLR having nRMSE value 12.42 % respectively. At grain filling stage model performance during calibration was excellent for all the models having nRMSE value less than 6.16 %. During validation model performance was excellent for all the models, except good for the SMLR model having lowest value of nRMSE for LASSO followed by SVR, LASSO-SVR, SMLR-SVR and SMLR. Based on model performance multistage wheat yield prediction was found to be best for LASSO followed by SVR for Patiala.

Table 5: Performance of wheat yield prediction done at multistage by different models for Ludhiana

Model	Modal accuracy during calibration				Modal accuracy during validation		
	R^2 (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)
At tillering stage							
SMLR	0.91	39847	199.6	5.14	224666	474.0	9.94
SVR	0.97	11696	108.2	2.78	291249	539. 7	11.32
LASSO	0.93	32634	180.7	4.65	18935	137.6	2.89
LASSO-SVR	0.94	25414	159.4	4.10	215293	464.0	9.73
SMLR-SVR	0.91	15920	126.2	3.25	311716	558.3	11.71
At flowering stage							
SMLR	0.95	22459	149.9	3.86	248867	498.9	10.47
SVR	0.99	5707	75.5	1.95	244568	494.5	10.37
LASSO	0.93	36693	191.6	4.93	38178	195.4	4.10
LASSO-SVR	0.95	17540	132.4	3.41	179340	423.5	8.88
SMLR-SVR	0.95	22873	151.4	3.89	288293	536.9	11.26
At grain filling stage							
SMLR	0.96	19966	141.3	3.64	302381	549.9	11.54
SVR	0.98	7083	84.2	2.17	250835	500.8	10.51
LASSO	0.94	3270	180.9	4.66	39022	197.5	4.14
LASSO-SVR	0.93	20865	144.5	3.72	205274	453.1	9.51
SMLR-SVR	0.95	13001	114.0	2.94	292471	540.8	11.35

Table 6: Performance of wheat yield prediction done at the multistage by different models for Patiala

Model	Modal accuracy during calibration				Modal accuracy during validation		
	R^2 (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)	MSE (kg ha^{-1})	RMSE (kg ha^{-1})	nRMSE (%)
At tillering stage							
SMLR	0.96	43114	207.6	6.01	277548	526.8	11.37
SVR	0.96	37512	193.7	5.61	143474	378.8	8.18
LASSO	0.95	53792	231.9	6.72	54452	233.4	5.06
LASSO-SVR	0.95	42266	205.6	5.95	201501	448.9	9.69
SMLR-SVR	0.95	44891	211.9	6.14	317740	563.7	12.17
At flowering stage							
SMLR	0.97	26392	162.5	4.70	330896	575.2	12.42
SVR	0.98	23176	152.2	4.41	108912	330.0	7.12
LASSO	0.96	53568	231.5	6.70	51890	227.8	4.94
LASSO-SVR	0.97	30760	175.4	5.08	121152	348.1	7.51
SMLR-SVR	0.96	33001	181.7	5.26	167502	409.3	8.83
At grain filling stage							
SMLR	0.96	39604	199.0	5.76	355923	596.6	12.88
SVR	0.98	22906	151.4	4.38	169982	412.3	8.90
LASSO	0.96	45228	212.7	6.16	49211	221.8	4.81
LASSO-SVR	0.96	35926	189.5	5.49	172670	415.5	8.97
SMLR-SVR	0.95	40293	200.7	5.81	176781	420.5	9.07

Percentage deviation of predicted yield done at different stage by observed yield for different location using different models

Percentage deviation of yield prediction done at the tillering stage for IARI, New Delhi for year 2018 by observed yield was lowest for SMLR-SVR followed by LASSO-SVR, LASSO, SMLR and SVR respectively (Table 7). For Hisar the percentage deviation was lowest for SMLR followed by SVR, LASSO, LASSO-SVR and SMLR-SVR respectively. For Ludhiana percentage deviation was lowest for LASSO followed by SMLR, LASSO-SVR, SVR and SMLR-SVR respectively. The percentage deviation for Patiala was lowest for SMLR followed by SMLR-SVR, SVR, LASSO and LASSO-SVR respectively. At tillering stage, LASSO had a percentage deviation less than 5% for four locations Ludhiana (0.62 %), Amritsar (2.98 %), IARI, New Delhi (-5.33 %), Patiala (-5.36 %) and 12.07% for Hisar.

Percentage deviation of yield prediction done at flowering stage for IARI, New Delhi by observed yield was lowest for LASSO-SVR followed by LASSO, SVR, SMLR-SVR and SMLR respectively. For Hisar percentage deviation was lowest for SMLR, followed by SVR, LASSO-SVR , LASSO and SMLR-SVR

Table 7: Percentage deviation of predicted yield done at the different stages by observed yield using different models for different location

	SMLR	SVR	LASSO	LASSO-SVR	SMLR-SVR
At the tillering stage					
IARI, New Delhi	-6.62	-7.41	-5.33	-4.71	-3.12
Hisar	7.58	10.15	12.07	14.97	20.62
Amritsar	11.21	-9.85	2.98	-5.09	-6.99
Ludhiana	5.23	9.59	0.62	9.09	10.33
Patiala	1.39	3.94	-5.36	8.02	2.31
At the flowering stage					
IARI, New Delhi	-7.56	-2.05	-0.91	-0.85	-3.60
Hisar	6.38	9.94	14.44	12.87	17.54
Amritsar	9.22	-10.18	2.93	-5.36	-9.02
Ludhiana	7.31	0.07	4.43	6.99	9.87
Patiala	7.29	7.32	0.28	5.87	7.67
At the grain filling stage					
IARI, New Delhi	-7.60	-0.67	-0.02	-1.03	-2.24
Hisar	5.36	-4.48	6.09	12.81	9.94
Amritsar	12.65	-12.11	-0.29	-5.90	-2.82
Ludhiana	6.23	-1.70	4.48	8.23	10.12
Patiala	1.74	6.71	-2.03	7.75	2.06

respectively. For Amritsar percentage deviation was lowest for LASSO followed by LASSO-SVR, SMLR-SVR, SMLR and SVR respectively. For Ludhiana the percentage deviation was lowest for SVR followed by LASSO, LASSO-SVR, SMLR, and SMLR-SVR respectively. For Patiala percentage deviation was lowest for LASSO followed by LASSO-SVR, SMLR, SVR, and SMLR-SVR respectively. At flowering stage LASSO had percentage deviation less than 5% for four stations, Patiala (0.28 %), IARI, New Delhi (-0.91 %), Amritsar (2.93 %), Ludhiana (4.43 %) and more than 10 % for Hisar (14.44%).

Percentage deviation of yield prediction done at the grain filling stage by observed yield for IARI, New Delhi was lowest for LASSO followed by SVR, LASSO-SVR, SMLR-SVR, and SMLR respectively. For Hisar percentage deviation was lowest for SVR, followed by SMLR, LASSO, SMLR-SVR and LASSO-SVR respectively. For Amritsar percentage deviation was lowest for LASSO followed by SMLR-SVR, LASSO-SVR, SVR and SMLR respectively. For Ludhiana percentage deviation was lowest for SVR followed by LASSO, SMLR, LASSO-SVR and SMLR-SVR respectively. For Patiala percentage deviation was lowest for SMLR followed by LASSO, SMLR-SVR, SVR and LASSO-SVR respectively. At grain filling stage, LASSO had percentage deviation less than 5% for four station IARI, New Delhi (-0.02 %), Amritsar (-0.29 %), Patiala (-2.03 %), Ludhiana (4.48 %) and 6.09 % for Hisar.

In our study, wheat yield prediction done at different crop stages by model developed by SMLR, SVR, LASSO and hybrid machine learning techniques had a percentage deviation between 0.02 to 20.62 % for different location of north-west India. This showed that these models are capable to predict the pre harvest wheat yield. Agrawal *et al.,* (2001) reported that reliable forecasting of wheat yield could be obtained when the crops were at twelve weeks. *Singh et al., (2014)* reported that statistical models based on weather indices can successfully simulate multi-stage yield forecast of wheat at mid-season and at pre-harvest for Amritsar, Bhatinda and Ludhiana districts. Vashisth *et al.,* (2018) reported that the percentage deviation of estimated yield by actual yield of maize crop done at flowering and at grain filling stage was 10.3 and 7.1 % by a weather based statistical model. In our study the percentage deviation of estimated yield by observed yield done at tillering, flowering and grain filling stage by the SVR model was between 1.03 to 17.54 %. Parviz (2018) observed that the minimum correlation coefficient between the observed and simulated yield for barley was found in the Gilan province using a support vector machine. In our study LASSO modal performed well for yield prediction done at tillering, flowering and grain filling stages. Vashisth and Aravind (2020) reported that on the basis of percentage deviation and model accuracy elastic net and LASSO model was found best for multistage mustard yield estimation done at vegetative, flowering and grain filling stage during *Rabi* 2018-19 and 2019- 20. Kumar *et al.,* (2019) used stepwise and LASSO techniques for developing forecast model forty-five days before harvest. He found that the SMLR forecast model over fit, whereas the LASSO performs better fit model. Also, the percent error by the LASSO model was less than SMLR. The models developed by machine learning techniques using weather parameters had lower value of percentage deviation, nRMSE and RMSE for the yield prediction done at the grain filling stage as compared to yield prediction done at flowering and tillering stage. This indicates better performance of the model at the grain filling stage. Vashisth *et al*. (2014) used weather based statistical model for multistage crop yield forecasting of the wheat crop at 45 and 25 days before harvesting with 10.7% and 7% percentage deviation, respectively. Palanivel and Surianarayanan (2019) reviewed several types of machine learning big data techniques such as linear regression, artificial neural network and support vector machine and found that SVM based prediction models are found to be more suitable for crop yield prediction. Goyal and Vashisth (2021) reported that model developed by variable extraction by PCA and SVM performed best for mustard yield prediction for IARI, New Delhi.

CONCLUSION

In the current research five models were developed for multi stage wheat yield prediction using long term weather data. Results showed that LASSO performed best followed by SVR. The model performance of SVR is increased by the hybrid machine learning techniques. Accuracy of multistage wheat yield prediction by LASSO-SVR was better as compared with SMLR-SVR. Since LASSO performed best for wheat yield prediction done at all three stages for study area as compared to other model. Hence this could be used for district level multi stage wheat yield prediction for different location of north-west India.

ACKNOWLEDGEMENT

The first author acknowledges PG school, Indian Agricultural Research Institute, New Delhi for providing fellowship for conducting research work. The Authors acknowledge the research facilities extended by Director, ICAR-Indian Agricultural Research Institute, New Delhi. The authors are highly grateful, and thankful to the President, Managing Editor and reviewers of Journal of Agrometeorology for their fruitful, constructive comments and suggestions, which improved the content of the paper.

Conflict of Interest Statement: The author(s) declare(s) that there is no conflict of interest.

Disclaimer: The content and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

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

REFERENCES

Agrawal, R., Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis. *Indian J. Agri. Sci.,* 71(7):487-490.

Alam, W., Ray, M., Kumar, R.R., Sinha, K., Rathod, S. and Singh K. N. (2018). Improved ARIMAX modal based on ANN and SVM approaches for forecasting rice yield using weather variables. *Indian J. Agri. Sci.,* 88 (12): 1909–13.

Aravind K S, Vashisth Ananta, Krishanan P and Das B. (2022). Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models *J. Agrometeorol.,* 24(1):18-25. https://doi.org/10.54386/jam.v24i1.1002

Goyal, A. and Vashisth, Ananta (2021). Mustard Yield Prediction using Machine Learning Approach. *J. Agri. Phys.,* 21(2): 445-456.

Ji, B., Sun, Y., Yang, S. and Wan, J. (2007). Artificial neural networks for rice yield prediction in mountainous regions. *J. Agri. Sci.*, 145:249–261.

Johnson, M.D., Hsieh,W., Cannon, A., Davidson, A. and Bedard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agri. Fore. Meteorol.,* pp.74–84.

Kumar, S., Attri, S. D. and Singh, K. K. (2019). Comparison of Lasso and stepwise regression technique for wheat yield predication. *J. Agrometeorol.*, 21(2): 188-192. https://doi.org/10.54386/jam.v21i2.231

Palanivel, K. and Surianarayanan, C. (2019). An approach for prediction of crop yield using machine learning and big data techniques. *Int. J. Comp. Engi. Tech.,* 10(3): 110-118.

Parviz L. (2018). Assessing accuracy of barley yield forecasting with integration of climate variables and support vector regression. *Anna Biologia*, 73(1):19-30.

Singh, A.K., Vashisth, Ananta, Sehgal,V.K, Goyal, A., Pathak, H. and Parihar, S.S. (2014). Development of Multi Stage District Level Wheat Yield Forecast Models. *J. Agri. Phys.,* 14(2): 189-193

Tibshirani, R. 1996. Regression shrinkage and selection via lasso. *J. Roy. Stat. Soc. B.*, 58: 267–288.

Vashisth, Ananta, Singh R. and Choudary, Manu. (2014). Crop yield forecast at different growth stage of wheat crop using statistical model under semi arid region. *J. Agroecol. Nat. Res. Manag*., :1-3.

Vashisth, Ananta, Goyal, A. and Roy, Debasish. (2018). Pre harvest maize crop yield forecast at different growth stage using different model under semi arid region of India. *Int. J. Trop. Agr.,* 36(4): 915-920.

Vashisth, Ananta and Aravind, K S. (2020). Multistage Mustard Yield Estimation Based on Weather Variables using Multiple Linear, LASSO and Elastic Net Models for Semi Arid Region of India. *J. Agri. Phys.,* 20(2): 213-223.

Verma, U., Piepho, H. P. and Goyal, A. (2016). Role of climatic variables and crop condition term for mustard yield prediction in Haryana. *Int. J. Agric. Stat. Sci.,* 12, 45–51.