Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 25 (3) : 425-431 (September - 2023)
https://doi.org/10.54386/jam.v25i3.2232
https://journal.agrimetassociation.org/index.php/jam

Research paper

Comparison of phenological weather indices based statistical, machine learning and hybrid models for soybean yield forecasting in Uttarakhand

YUNISH KHAN[1], VINOD KUMAR[1], PARUL SETIYA[2] and ANURAG SATPATHI[2*]

[1]*Department of Mathematics, Statistics and Computer science, G.B. Pant University of Agriculture and Technology, Pantnagar 263145, Uttarakhand, India*

[2]*Department of Agrometeorology, College of Agriculture, G.B. Pant University of Agriculture and Technology, Pantnagar 263145, Uttarakhand, India*

**Corresponding Author: anuragsatpathi50@gmail.com*

ABSTRACT

Early information exchange regarding predicted crop production could play a role in lowering the danger of food insecurity. In this study total six multivariate models were developed using past time series yield data and weather indices viz. SMLR, PCA-SMLR, ANN, PCA-ANN, SMLR-ANN and PCA-SMLR-ANN for three major soybean producing districts of Uttarakhand viz. Almora, Udham Singh Nagar and Uttarkashi. Further analysis was done by fixing 80% of the data for calibration and the remaining dataset for validation to predict soybean yield. Phenology wise average values were computed using the daily weather data. These average values are subsequently employed in the computation of both weighted and unweighted weather indices. The PCA-SMLR-ANN, SMLR-ANN and PCA-ANN models were found to be the best soybean yield predictor model for Almora, Udham Singh Nagar and Uttarkashi districts, respectively. The overall ranking based on the performances of the models for all locations can be given as: SMLR-ANN > PCA-ANN > PCA-SMLR-ANN ≈ ANN > PCA-SMLR > SMLR. The study results indicated that hybrid models outperformed the individual models well for all the study regions.

Keyword: Crop yield prediction, Stepwise Multiple Linear Regression (SMLR), Principal component analysis (PCA), Artificial Neural Network (ANN).

Agriculture plays a vital role in the global as well Indian economy. The world's growing population and climate change has put increasing danger on agricultural production. There is no way to completely reduce these occurrences, it would be much better if information about the future was known early so that farmers could make appropriate plans and take actions accordingly (FAO, 2017). Early information exchange regarding crop production forecasting could play a key role in lowering the danger of food insecurity. Hence, making accurate crop yield forecasting is more important than ever. Accurate crop yield production is a vital aspect of agriculture, providing valuable insights into the expected yield, not only allowing timely decision making for farmers but also for other stakeholders (Cao *et al*., 2021). Accurate crop yield forecasting can help farmers to optimize their resources, reduce the amount of waste produced and improve overall efficiency (Setiya *et al*., 2022). In addition to this, it can help policymakers to make timely informed decisions relating to food security, grain storage, transportation, marketing and price stabilization (Satpathi *et al*., 2023).

Worldwide, soybean (Glycine max (L.) Merrill) is one of the most significant oilseed crop produced around the world. The Brazil (38%) is highest producer of soybean followed by United States of America (31%). India ranks fifth in leading soybean producing countries (Soystats, 2022). In the Northwestern Himalayan hill region, soybean is grown as a major Kharif crop. In the Northwestern Himalayan region, the state of Uttarakhand contributes maximum approximately 90-95% of total soybean acreage and production (Bhartiya *et al*., 2017).

Phenological weather indices can be widely used in agricultural research to predict crop yield (Ji *et al*., 2021). These indices measure the timing of specific developmental stages in crops viz. flowering and maturity, in response to environmental factors like temperature, rainfall and sunlight etc. (Banerjee *et al*.,

Article info - DOI: https://doi.org/10.54386/jam.v25i3.2232
Received: 25 May 2023; Accepted: 20 July 2023; Published online : 31 August 2023

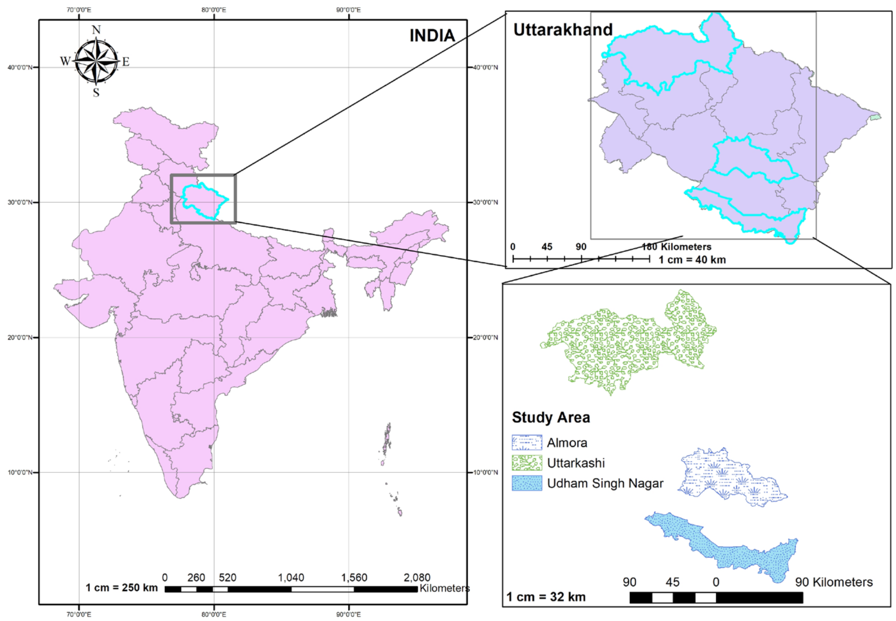

Fig. 1: Map of study locations

2021). Formerly, researchers estimate the crop yield using crop cutting experiment (Ahmad *et al*., 2021) and by employing only statistical approaches such as Multivariate linear regression (MLR) technique (Basso *et al*., 2013) but due to lower prediction accuracy now a days machine learning models are very popular among the researchers. Statistical models use mathematical equations to identify relationships between weather variables and crop yields. Machine learning models, on the other hand, utilize algorithms to comprehend patterns and connections within data, enabling them to generate predictions based on those patterns. There has been number of studies, machine learning methods used to predict crop yield in number of crops and plants viz. rice (Satpathi *et al*., 2023), wheat (Aravind *et al*., 2022), pigeon pea (Sridhara *et al.,* 2023), cashew (Das *et al*., 2022), sorghum (Sridhara *et al*., 2020) and coconut (Das *et al*., 2020).

In the available past studies, several studies have covered the comparison related to the effect of direct weather variables and weather indices on yield prediction but only few research work has assessed the variety of hybrid models or internal relationships between individual models. Hence, in this study we developed hybrid models such as PCA-SMLR, PCA-ANN etc. to examine the accuracy and reliability of hybrid models compared to individual statistical and machine learning models. The findings of this study have significant implications for the agricultural industry, providing insights into the effectiveness of different approaches to crop yield forecasting.

MATERIALS AND METHODS

Yield prediction models were developed based on the *kharif* soybean yield data and historical weather datasets. Three major soybean producing districts of Uttarakhand were selected for this study, viz. Almora, Udham Singh Nagar and Uttarkashi (Fig. 1). The soybean production potential of this area is enhanced by its fertile soil and retains sufficient moisture to yield good crops.

Time series data of soybean yield of 20 years (2001–2020) were obtained from the Soybean Breeding Laboratory, GBPUAT Pantnagar for Udham Singh Nagar and from Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare for Almora and Uttarkashi districts. The data on weather variables namely maximum and minimum temperature, daily average relative humidity, daily rainfall, daily sunshine hour and wind speed were collected from the Department of Agrometeorology, GBPUAT for Udham Singh Nagar district and from NASA POWER web portal (https://power.larc.nasa.gov/data-access-viewer/). The weather variables data have been transformed into phenological stage wise averages to process further.

Among the complete dataset spanning 20 years, 16 years of data were employed for training of the models, while the remaining 4 years data were utilized for testing of the models (Li *et al.,* 2017, Rajaee *et al.,* 2018). In terms of phenology, the average values were computed using the daily weather data. These average values are subsequently employed in the computation of both weighted and unweighted weather indices. The details about calculation of weighted and unweighted weather indices can be found in previous paper of Satpathi *et al.,* (2023). In total six multivariate models were developed to train and test the models *viz.* SMLR, PCA-SMLR, ANN, PCA-ANN, SMLR-ANN and PCA-SMLR-ANN. The following are the provided specifics regarding the multivariate analysis techniques employed in this study:

Principal component analysis (PCA)

The objective of principal component analysis (PCA) is to decrease the dimensionality of a data set while retaining most of the information. PCA is executed to reduce the risk of overfitting due to the high dimensionality and interdependencies among the independent variables. It is also known as a variable reduction method or data reduction method or data dimension reduction method. All the input variables were standardized on dividing the values by the standard deviation after the mean has been subtracted. The principal components (PCs) with eigen-values more than 1 were only considered (Brejda *et al.,* 2000).

Stepwise multiple linear regression (SMLR)

The SMLR technique, based on the dataset of yield and weather parameters, is the simplest approach for developing the yield forecast model. It involves a systematic process of constructing the model by introducing or eliminating predictor variables. This method allows for the selection of the most effective predictors from a large pool of predictors (Das *et al*., 2018). Stepwise regression necessitates two significant levels: one for adding variables and another for removing variables. In the current study, the p-values of 0.50 and 0.10 were taken for addition and removal of the variables respectively.

Artificial neural network (ANN)

ANN is a type of computational model inspired by the central nervous system and designed for machine learning purposes. These models are typically represented as interconnected systems of "neurons" that can process input information and compute values by propagating data through the network (Dahikar *et al.,* 2014). The main challenge in implementing ANN is determining the optimal number of hidden neurons or nodes. In this study, the number of hidden nodes was selected using the "train" function of the "caret" package in R software, employing the "nnet" method with 10-fold cross-validation (Kuhn, 2008). All weather indices were used as inputs, while the yield served as the dependent variable.

PCA-SMLR, PCA-ANN, SMLR-ANN and PCA-SMLR-ANN

In these techniques, PCA scores were employed as input for the analysis (Aravind *et al.,* 2022). To address the issue of multicollinearity among weather variables, PC (Principal Component) scores were utilized as regressors for SMLR (Stepwise Multiple Linear Regression) and ANN (Artificial Neural Network) in order to construct crop yield models (Verma *et al*., 2016). PCA (Principal Component Analysis) is employed to decompose the original data matrix X into two matrices, P and T, denoted as $X = TP^t$. The matrix P is commonly referred to as the loading's matrix, while the matrix T represents an orthogonal score matrix. The superscript t denotes the transpose of a matrix.

Uttrakashi

All the developed models were compared based on the R^2 and nRMSE values provided in Table 7. For Almora during the calibration stage ANN based models viz. ANN, PCA-ANN, SMLR-ANN and PCA-SMLR-ANN performed excellent. Similar findings by Mishra *et al.,* (2017) were observed that the ANN can be more accurate and practical for yield prediction than the SMLR technique. Overall PCA-SMLR-ANN model was found to be the best soybean yield predictor model for Almora district. Among the used models two ANN based models viz. SMLR-ANN and PCA-SMLR-ANN were found to be excellent for Udham Singh Nagar region. Addition to this ANN based models again outperformed SMLR based models for the Uttarkashi district. The PCA-ANN model was found to be the excellent model followed by SMLR-ANN for Uttarkashi. The overall ranking based on the performances of the models for all locations can be given as: SMLR-ANN > PCA-ANN > PCA-SMLR-ANN ≈ ANN > PCA-SMLR > SMLR. The study results indicated that SMLR-ANN model performed well for all the study regions. These findings were also in line with the study done by Aravind *et al.,* (2022), Kumar (2019) and Setiya *et al.,* (2022) which concluded that the performance of ANN based models were better as compared to SMLR based models. Apart from that, this study also shows the superior performance of hybrid models over the individual models.

Testing the performances of the models for the Uttarkashi

Prediction accuracy of models were evaluated based on R^2 (Coefficient of determination), RMSE (Root Mean Square Error), nRMSE (Normalized Root Mean Square Error), MAE (Mean Absolute Error), MBE (Mean Biased Error) and modeling efficiency (EF). Formulas of these can be found in previous paper of Setiya *et al.,* (2022). The developed models were compared based on the value of R^2, as $R^2 > 0.90$, excellent, $R^2 = 0.90$-0.75, good, $R^2 = 0.75$-0.50, fair and $R^2 < 0.50$, poor, similarly value of nRMSE, as nRMSE < 10%, excellent, nRMSE = 10-20%, good, nRMSE = 20-30%, fair and nRMSE > 30%, poor.

RESULTS AND DISCUSSION

Performances by the individual as well hybrid models were categorized based on value of R^2, RMSE, nRMSE, MBE and EF and presented here under according to different locations.

Almora

The values of prediction accuracy statistics of all models for Udham Singh Nagar can be found in Table 1 and Table 2. Initially the performance of models during calibration was compared based on the R^2 value, which is ranging from 0.71 for PCA-SMLR to 0.97 for ANN. The value of RMSE, nRMSE and EF during the calibration stage suggests that the best model performance was observed for ANN followed by SMLR-ANN, PCA-SMLR-ANN, PCA-ANN, SMLR and PCA-SMLR. The MBE values during the calibration stage were positive for SMLR suggesting over estimation, zero for PCA-SMLR and positive for other models suggesting under estimation of predicted soybean yield. During the validation stage based on R^2 value model performances were poor except for PCA-SMLR-ANN (0.99), SMLR-ANN (0.87) and ANN (0.77). The RMSE values during validation range from 9.83 t ha^{-1} for PCA-SMLR-ANN to 124.21 t ha^{-1} for SMLR. The values of nRMSE and EF also suggest that the best model during validation is PCA-SMLR-ANN followed by ANN, SMLR-ANN. The lowest error percentage was achieved by the PCA-SMLR-ANN (0.04% to 1.38%), while highest was SMLR (-28.46% to 25.77%) during soybean yield forecasting. Hybrid models performed better compared to the individual models for soybean yield prediction in Almora region.

Table 1: Quantitative measures obtained during calibration and validation for Almora

Model	R^2	MBE (t ha^{-1})	RMSE (t ha^{-1})	nRMSE (%)	EF	R^2	MBE (t ha^{-1})	RMSE (t ha^{-1})	nRMSE (%)	EF
	Calibration					Validation				
SMLR	0.85	0.65	124.21	11.28	0.83	0.33	85.73	124.21	21.73	-1.02
PCA-SMLR	0.71	0	173.16	15.72	0.6	0.09	2.66	108.21	10.5	-1.94
ANN	0.97	-13.82	56.45	5.12	0.97	0.77	28.2	52.97	5.14	0.44
PCA-ANN	0.9	-21.05	103.16	9.37	0.9	0.29	50.8	85.49	8.29	-0.45
SMLR-ANN	0.95	-12.21	84.6	7.68	0.92	0.87	12.6	77.83	6.97	0.76
PCA-SMLR-ANN	0.9	-16.54	87.34	7.93	0.9	0.99	-8.34	9.83	0.95	0.98

Table 2: Details of the models employed and error percentage for Almora

Model	Equation	Error Percentage
SMLR	Y = - 6537.7 + 0.8 * Z351 + 0.7 * Z250	-28.46 % to 25.77 %
PCA-SMLR	Y = 1087.8 + 43.8*PC1 + 72.7*PC4; No. of PC's: 6	-13.78 % to 15.45 %
ANN	No of hidden neurons: 10	-10.49 % to 2.11 %
PCA-ANN	3; No of PC's: 6	-15.83 % to 2.33 %
SMLR-ANN	o of hidden neurons: 11	-12.06 % to 6.63 %
PCA-SMLR-ANN	5; No of PC's: 6	0.04 % to 1.38 %

Table 3: Quantitative measures obtained during calibration and validation for Udham Singh Nagar

Model	R^2	MBE (t ha^{-1})	RMSE (t ha^{-1})	nRMSE (%)	EF	R^2	MBE (t ha^{-1})	RMSE (t ha^{-1})	nRMSE (%)	EF
	Calibration					Validation				
SMLR	0.87	-0.02	215.74	10.96	0.84	0.01	379.35	581.41	36.49	-2.08
PCA-SMLR	0.81	0	259.3	13.17	0.76	0.35	118.86	423.18	26.56	-11.02
ANN	0.86	-110.82	246.95	12.54	0.83	0.99	1.5	36.59	2.3	0.99
PCA-ANN	0.74	-150.73	360.69	18.32	0.63	0.88	-109.75	172.83	10.85	0.71
SMLR-ANN	0.99	-8.17	25.52	2.02	0.99	0.99	-3.26	6.48	0.02	1
PCA-SMLR-ANN	0.97	-16.13	59.7	4.74	0.96	0.99	-11.05	24.44	1.81	0.99

Table 4: Details of the models employed and error percentage for Udham Singh Nagar

Model	Equation	Error Percentage
SMLR	Y = - 15138.3 + 102.5 * Z31 + 1.5 * Z381	-39.19 % to 14.22 %
PCA-SMLR	Y = 1917.2 + 103.2 * PC1; No of PC's: 11	0.17 % to 17.02 %
ANN	No of hidden neurons: 3	-8.69 % to 2.1 %
PCA-ANN	3; No of PC's: 11	-10.28 % to 3.64 %
SMLR-ANN	No of hidden neurons: 3	-0.62 % to 0.63 %
PCA-SMLR-ANN	3; No of PC's: 7	-0.92 % to 0.97 %

Udham Singh Nagar

The values of prediction accuracy statistics of all models for Udham Singh Nagar are shown in Table 3 and Table 4. During the calibration stage the R^2 value for all the models were in excellent to good range. The best R^2 among all the models is for SMLR-ANN (0.99) followed by PCA-SMLR-ANN (0.97), SMLR (0.87), ANN (0.86), PCA-SMLR (0.81) and PCA-ANN (0.74). The RMSE and nRMSE values during calibration stage were also in line with the findings based on R^2 value. The values of MBE during calibration suggest under estimation of predicted values except for the PCA-SMLR (0). During the validation stage best R^2 values were obtained

Table 5: Quantitative measures obtained during calibration and validation for Uttarkashi

Model	R^2	MBE (t ha^{-1})	RMSE (t ha^{-1})	nRMSE (%)	EF	R^2	MBE (t ha^{-1})	RMSE(t ha^{-1})	nRMSE (%)	EF
	Calibration					Validation				
SMLR	0.99	-1.85	2.07	0.22	1	0.32	-161.5	301.2	27.68	0.03
PCA-SMLR	0.78	0	126.09	13.51	0.73	0.88	-55.98	79.66	7.32	0.76
ANN	0.92	2.5	84.17	9.02	0.9	0.6	-4.32	88.79	8.16	0.6
PCA-ANN	0.99	-3.93	22.4	2.4	0.99	0.96	9.21	33.08	3.03	0.94
SMLR-ANN	0.96	-18.93	66.96	7.19	0.94	0.91	52.05	108.28	11.69	0.87
PCA-SMLR-ANN	0.62	11.25	148.18	15.87	0.62	0.98	-52.95	69.88	6.77	0.82

Table 6: Details of the models employed and error percentage for Uttarkashi

Model	Equation	Error Percentage
SMLR	$Y = -7084.1 - 22.3 * Time + 22.6 * Z_{20} + 64.3 * Z_{21} + 0.6 * Z_{50} + 1.3 * Z_{151} - 0.4 * Z_{160} + 0.1 * Z_{250} + 19.3 * Z_{261} + 0.7 * Z_{351} + 0.1 * Z_{360} - 5.5 * Z_{361} - 0.1 * Z_{450} + 0.6 * Z_{451} - 1.2 * Z_{461} - 29.3 * Z_{561}$	-45.5 % to 18.92 %
PCA-SMLR	$Y = 973.0 - 1.9 * Time + 32.0 * PC1 + 45.7 * PC2 + 18.1 * PC3 - 64.7 * PC4 + 42.6 * PC5 - 21.7 * PC6$; No. of PC's: 6	-16.60 % to 1.15 %
ANN	No of hidden neurons: 17	-11.70 % to 11.87 %
PCA-ANN	4; No of PC's: 6	-3.92 % to 4.46 %
SMLR-ANN	No of hidden neurons: 19	-31.29 % to 1.40 %
PCA-SMLR-ANN	1; No of PC's: 6	0.28 % to 13.48 %

for ANN (0.99), SMLR-ANN (0.99) and PCA-SMLR-ANN (0.99) while poor R^2 values were also obtained for SMLR (0.01) and PCA-SMLR (0.35) suggesting poor model performance. The value of RMSE and nRMSE were also lowest for SMLR-ANN (6.48 t ha-1 and 0.02%), PCA-SMLR-ANN (24.44 t ha^{-1} and 1.81%) and ANN (36.59 t ha^{-1} and 2.3%) suggesting again excellent model performance. Similarly, the values of EF were best during validation stage for PCA-SMLR-ANN (1), ANN (0.99) and PCA-SMLR-ANN (0.99). The lowest model error percentage was for SMLR-ANN (-0.62% to 0.63%) followed by PCA-SMLR-ANN (-0.92% to 0.97%), which shows the superior performance of hybrid models again over individual models.

The performance of the different models for Uttarkashi is shown in Table 5 and Table 6. During the calibration stage the R^2 for all models ranging from 0.62 for PCA-SMLR-ANN to 0.99 for SMLR and PCA-ANN. The RMSE and nRMSE values suggest, excellent performance for SMLR (0.22%), PCA-ANN (2,4%), SMLR-ANN (7.19%) and ANN (9.02%) to good performance for PCA-SMLR-ANN (15.87%) and PCA-SMLR (13.51%). The values of EF during calibration were also excellent for SMLR, PCA-ANN, SMLR-ANN and ANN. The MBE during the calibration stage suggest under estimation of yield for SMLR (-1.85 t ha^{-1}), PCA-ANN (-3.93 t ha^{-1}) and SMLR-ANN (-18.93 t ha^{-1}), while over estimation for PCA-SMLR-ANN (11.25 t ha^{-1}) and ANN (2.5 t ha^{-1}). During the validation stage based on R^2 value PCA-SMLR-ANN (0.98) model was found to be best followed by the PCA-ANN (0.96), SMLR-ANN (0.94) and PCA-SMLR (0.73). The RMSE values during validation stage were ranged from 33.08 t ha-1 for PCA-ANN to 301.2 t ha-1 for SMLR. Based on nRMSE values during the validation stage model performance was excellent for PCA-ANN (3.03%), PCA-SMLR-ANN (6.77%), PCA-SMLR (7.32%) and ANN (8.16%), good for SMLR-ANN (11.69%) and fair for SMLR (27.58%). The MBE values were negative, suggesting under estimation of predicted yield except for PCA-ANN and SMLR-ANN. Based on EF values also PCA-ANN (0.94) was the best model followed by SMLR-ANN (0.87) and PCA-SMLR-ANN (0.82). Among the all models the minimum error percentage ranging from -3.92% to 4.46% was observed for PCA-ANN, while maximum for SMLR ranging from -45.5% to 18.92%. The lower values of error percentage of hybrid models over individual models signifies the better performance of hybrid models.

CONCLUSION

In the present study, six multivariate models were examined for soybean yield prediction based on different weather variables. The results revealed that the PCA-SMLR-ANN, SMLR-ANN and PCA-ANN models were found to be the best soybean yield predictor model for Almora, Udham Singh Nagar and Uttarkashi districts, respectively. The performance of SMLR-ANN model was found to be best compared to other multivariate models considered in this study. The next best model was PCA-ANN. Thus, it can be concluded that hybrid models viz. SMLR-ANN and PCA-

Table 7: Cross comparison of the model performances based on R^2 values

Location	Model	R^2_{cal}	$nRMSE_{cal}$	R^2_{val}	$nRMSE_{val}$
Almora	SMLR	Good	Good	Poor	Fair
	PCA-SMLR	Fair	Good	Poor	Good
	ANN	Excellent	Excellent	Good	Excellent
	PCA-ANN	Excellent	Excellent	Poor	Excellent
	SMLR-ANN	Excellent	Excellent	Good	Excellent
	PCA-SMLR-ANN	Excellent	Excellent	Excellent	Excellent
Udham Singh Nagar	SMLR	Good	Good	Poor	Poor
	PCA-SMLR	Good	Good	Poor	Fair
	ANN	Good	Good	Excellent	Excellent
	PCA-ANN	Fair	Good	Good	Good
	SMLR-ANN	Excellent	Excellent	Excellent	Excellent
	PCA-SMLR-ANN	Excellent	Excellent	Excellent	Excellent
Uttarkashi	SMLR	Excellent	Excellent	Poor	Fair
	PCA-SMLR	Good	Good	Good	Excellent
	ANN	Excellent	Excellent	Fair	Excellent
	PCA-ANN	Excellent	Excellent	Excellent	Excellent
	SMLR-ANN	Excellent	Excellent	Excellent	Good
	PCA-SMLR-ANN	Fair	Good	Excellent	Excellent

ANN outperforms the individual models viz. SMLR and ANN. Additionally, this study also concluded that PCA-SMLR-ANN, SMLR-ANN and PCA-ANN can be used for Almora, Udham Singh Nagar and Uttarkashi districts, respectively for reliable soybean yield prediction.

ACKNOWLEDGEMENT

Authors are thankful to the Govind Ballabh Pant University of Agriculture and Technology, Pantnagar for providing facility to conduct the research. Authors would like to express heartfelt gratitude to the editor and reviewers' invaluable contribution and dedication in reviewing and editing the manuscript.

Funding: No external funding is involved in this study.

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

Data availability: To be provided on request.

Authors contribution: **Y. Khan**: Investigation, Formal analysis, Writing-original draft; **V. Kumar**: Supervision, Writing-review; **P. Setiya**: Conceptualization, Methodology, Writing; **A. Satpathi**: Writing-review, Editing.

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

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

REFERENCES

Ahmad, T., Sahoo, P. M., Singh, M. and Biswas, A. (2021). Crop Cutting Experiment techniques for determination of yield rates of field crops. DOI:10.13140/RG.2.2.16327.62883.

Aravind, K., Vashisth, A., 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. DOI:https://doi.org/10.54386/jam.v24i1.1002

Banerjee, P., Mukherjee, B., Venugopalan, V. K., Nath, R., Sarath Chandran, M. A., Dessoky, E. S., Ismail, I. A., El-Hallous, E. I. and Hossain, A. (2021). Thermal Response of Spring–Summer-Grown Black Gram *(Vigna mungo L. Hepper)* in Indian Subtropics. *MDPI.* DOI:https://doi.org/10.3390/atmos12111489.

Basso, B., Cammarano, D. and Carfagna, E. (2013). Review of crop yield forecasting methods and early warning systems. In Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy (Vol. 18, p. 19).

Bhartiya, A., Chandra, N., Pal, R. S., Aditya, J. P. and Bajeli, J. (2017). Comparative yield gaps, economic impact and constraint analysis in front line demonstrations of soybean and black soybean in Uttarakhand hills. *Indian J. Agric. Res.*, 51(5): 483-487.

Brejda, J. J., Moorman, T. B., Karlen, D. L. and Dao, T. H. (2000). Identification of Regional Soil Quality Factors and

Indicators, I. Central and Southern High Plain. *Soil Sci. Soc. Am. J*., 64:2115–2124.

Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z. and Tao, F. (2021). Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *Eur. J. Agron.*, *123*: 126204.

Dahikar, S.S. and Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. *Int. J. Inov. Res. Electr. Electron. Instrum. Control Eng.,* 2: 683–686.

Das, B., Nair, B., Reddy, V. K. and Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *Int. J. Biometeorol*., 62(10): 1809-1822.

Das, B., Nair, B., Arunachalam, V., Reddy, K. V., Venkatesh, P., Chakraborty, D. and Desai, S. (2020). Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the west coast of India. *Int. J. Biometeorol*., 64: 1111-1123.

Das, B., Murgaonkar, D., Navyashree, S. and Kumar, P. (2022). Novel combination artificial neural network models could not outperform individual models for weather-based cashew yield prediction. *Int. J. Biometeorol*., 66(8): 1627-1638.

FAO (2017). The future of food and agriculture – Trends and challenges. Rome. Available online: https://www.fao.org/3/i6583e/i6583e.pdf (Accessed on 20 April 2023).

Ji, Z., Yaozhong, P., Xiufang, Z., Jinyun, W. and Qiannan L. (2021). Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. *Sensors.,* 21(4): 1406. DOI:https://doi.org/10.3390/s21041406.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *J. Stat. Softw.*, 28: 1–26.

Kumar, S. (2019): Comparison of linear regression and artificial neural network technique for prediction of a soybean biodiesel yield. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*., DOI:https://doi.org/10.1080/15567036.2019.1604858

Li, X., Sha, J. and Wang, Z. L. (2017). Chlorophyll-A Prediction of lakes with different water quality patterns in China based on hybrid neural networks. *Water*, 9: 524.

Mishra, M., Bhavyashree, S. and Sharma, H. (2017). Comparative study of the performance of artificial neural network and multiple linear regression techniques for predicting the soybean yield using its attributing characters. *Int. J. Life Sci.*, 12(1): 177-181.

Rajaee, T. and Jafari, H. (2018). Utilization of WGEP and WDT models by wavelet denoising to predict water quality parameters in rivers. *J. Hydrol. Eng.,* 23.

Satpathi, A., Setiya, P., Das, B., Nain, A. S., Jha, P. K., Singh, S. and Singh, S. (2023). Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India. *Sustainability*, 15(3): 2786.

Setiya, P., Satpathi, A., Nain, A. S. and Das, B. (2022). Comparison of weather-based wheat yield forecasting models for different districts of Uttarakhand using statistical and machine learning Techniques. *J. Agrometeorol.*, 24(3): 255–261. DOI:https://doi.org/10.54386/jam.v24i3.1571

Soystats (2022). A Reference Guide to Important Soybean Facts & Figures, http://soystats.com/wp-content/uploads/Soy-Stats-2022-Web.pdf (Accessed on 20 April 2023).

Sridhara, S., Ramesh, N., Gopakkali, P., Das, B., Venkatappa, S. D., Sanjivaiah, S. H. and Elansary, H. O. (2020). Weather-based neural network, stepwise linear and sparse regression approach for rabi sorghum yield forecasting of Karnataka, India. *Agron.*, 10(11): 1645.

Sridhara, S., Manoj, K. N., Gopakkali, P., Kashyap, G. R., Das, B., Singh, K. K. and Srivastava, A. K. (2023). Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. *Int. J. Biometeorol*, 67(1): 165-180.

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