Comparison of Phenological Weather indices based Statistical and Machine Learning Models for Soybean Yield Forecasting in Pantnagar, Uttarakhand

YUNISH KHAN[1], VINOD KUMAR[2]

[1]Department of Mathematics, Statistics and Computer science,
G.B. Pant University of Agriculture and Technology, Pantnagar 263145, Uttarakhand, India
khanyunish263145@gmail.com

[2]Department of Mathematics, Statistics and Computer science,
G.B. Pant University of Agriculture and Technology, Pantnagar 263145, Uttarakhand, India
vinod_kumarbcb@yahoo.com

Abstract

Early information exchange regarding predicted crop production could play a role in lowering the danger of food insecurity. Predicting crop yields is one of the more difficult tasks in the farming industry. Several investigations have been conducted in the agricultural field to predict increased crop production using the machine learning algorithm Artificial Neural Network (ANN) and statistical model Stepwise Multiple Linear Regression (SMLR). In this study eight multivariate weather-based models including stepwise multiple linear regression (SMLR), principal component analysis (PCA), artificial neural network (ANN) and combinations of them using weather indices and direct weather variables were investigated by fixing 80% of the data for calibration and the remaining dataset for validation to predict soybean yield for Pantnagar, Uttarakhand. Based on the value of R^2 (0.95) and nRMSE (7.16%) during calibration stage, the PCA-ANN-W model performed excellent, becoming the best model for soybean prediction compared to other models in the study region. The overall ranking based on the performances of the models can be given as: PCA-ANN-W > ANN-WI > SMLR-W > SMLR-WI ≈ PCA-SMLR-WI > ANN-W > PCA-ANN-WI > PCA-SMLR-W. The study results indicated that PCA-ANN-W and ANN-WI model performed well for the study region as compared to other models.

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

I. Introduction

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 (FAO, 2015). 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 (Khan et al., 2023). Hence, making accurate crop yield forecasting is more important than ever (Mann et al., 2019). 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. (Khan et al., 2021; Banerjee et al., 2021; Ihsan et al., 2016; Ransing et al., 2014). Phenological weather indices are useful because they provide information on the physiological status of crops, which is a crucial determinant of yield (Seo et al., 2019). 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; Setiya 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, no study has covered the comparison related to the effect of direct weather variables and weather indices on yield prediction. Addition to this, no research work has assessed the internal relationships between PCA (principal component analysis)-SMLR (Stepwise multiple linear regression) and PCA-

ANN (Artificial neural network). Hence, in this study we developed hybrid models such as PCA-SMLR and PCA-ANN based on both direct weather variables and weather indices. This study aims to examine the accuracy and reliability of both statistical and machine learning models in forecasting soybean yield, with the goal of identifying the most effective approach for soybean yield forecasting. The findings of this study have significant implications for the agricultural industry, providing insights into the effectiveness of different approaches to crop yield forecasting.

II. Materials and Methods

2.1 Study Area

Yield prediction models were developed based on the kharif soybean yield data (kg/ha) and the weather data of GBPUAT, Pantnagar, Uttarakhand in this study (Figure1). The Pantnagar region lies in the Tarai belt of Uttarakhand state of India at 29°3' N latitude and 79°31' E longitude at an elevation of 243m above the mean sea level. The agricultural potential of this area is enhanced by its fertile soil and retains sufficient moisture to yield good crops.

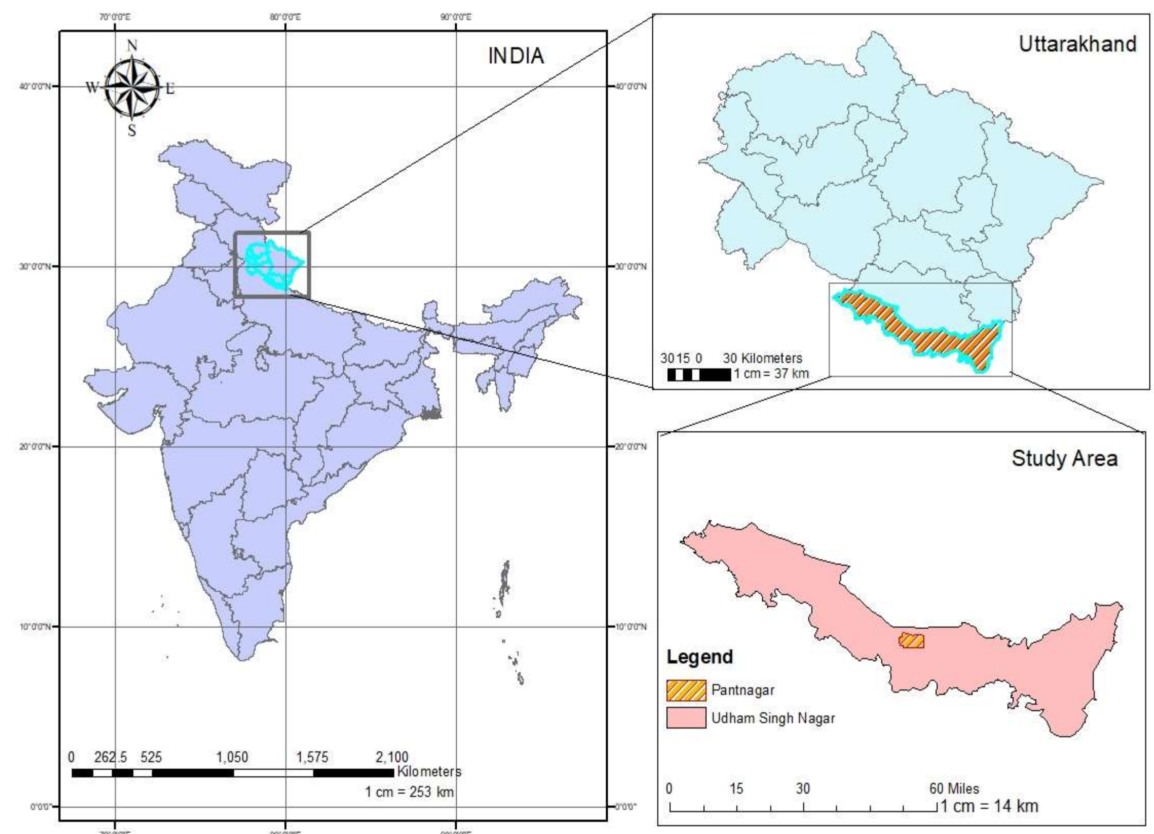

Fig. 1: Map of Pantnagar, Udham Singh Nagar, Uttarakhand

2.2 Data Collection

Time series data of soybean yield of 20 years (2001–2020) were obtained from the Soybean Breeding Laboratory, Department of Genetic & Plant Breeding, GBPUAT Pantnagar, Uttarakhand. The data on weather variables were collected from the Meteorological Observatory, Department of Agrometeorology, GBPUAT, Pantnagar, Uttarakhand.

2.3 Steps involved in the Model Development

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 (Figure 2), 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 Setiya et al. (2022) and Satpathi et al. (2023). The steps involved in the model development are illustrated in Figure 3.

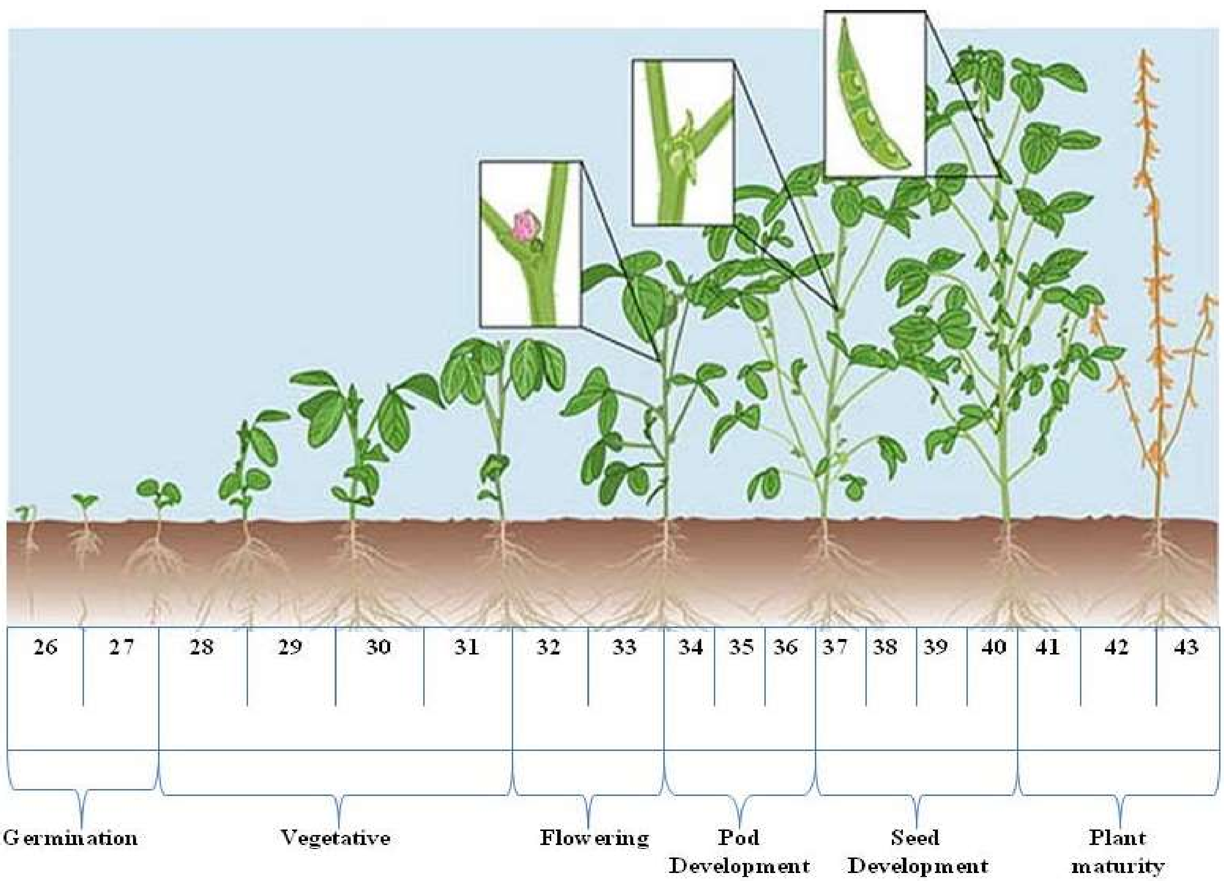

Fig. 2: Crop growth stages of soybean in terms of standard meteorological weeks

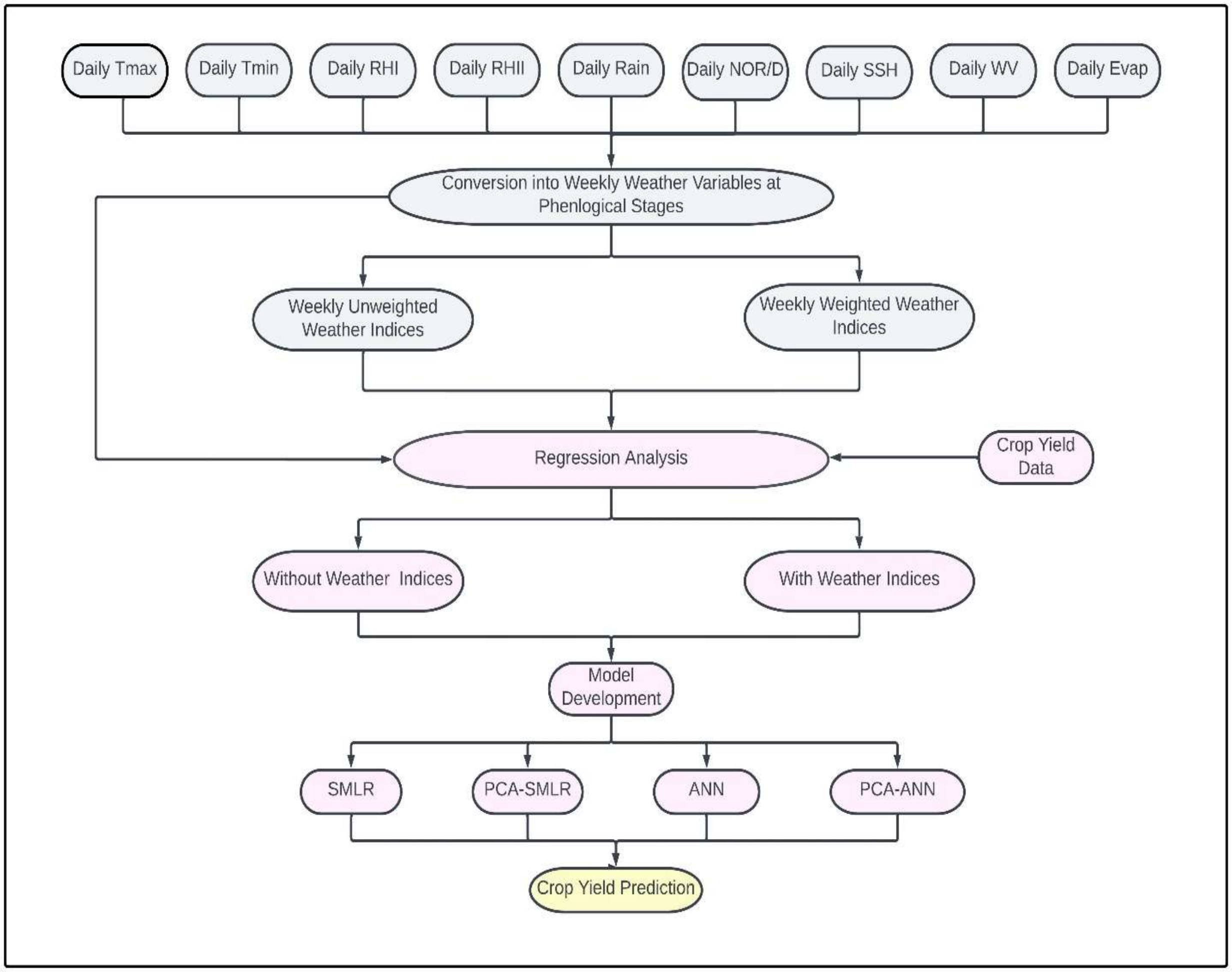

Fig. 3: Flowchart representing different stages in development of yield prediction model

2.4 Multivariate Analysis Techniques

In total eight multivariate models were developed by using historical data on soybean yield and phenological weather indices to train and test the models *viz.* SMLR-WI, SMLR-W, PCA-SMLR-WI, PCA-SMLR-W, ANN-WI, ANN-W, PCA-ANN-WI and PCA-ANN-W. W represents use of weather variables directly as model input and WI represents use of weather indices as model input. The following are the provided specifics regarding the multivariate analysis techniques employed in this study for the development of the crop yield prediction model:

2.4.1 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).

2.4.2 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 (Singh et al., 2014; Das et al., 2018). Stepwise regression necessitates two significant levels: one for adding variables and another for removing variables. To avoid an infinite loop, the cutoff probability for adding variables should be lower than the cutoff probability for removing variables (Singh et al., 2018). In the current study, the p-values of 0.50 and 0.10 were taken for addition and removal of the variables respectively.

2.4.3 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). They consist of three layers: the input layer, the hidden layer, and the output layer. In this technique, data flows from the input layer through the hidden layer to the output layer (Kaul et al., 2005). The number of nodes in the input layer depends on the number of independent predictors. Each layer is composed of interconnected neurons or nodes. The number of neurons in the input and output layers is determined by the dataset used. 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.

2.4.4 PCA-SMLR and PCA-ANN

In PCA-SMLR and PCA-ANN 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.

2.5 Testing the Performance of the Models

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 papers of Setiya et al. (2022) and Satpathi et al. (2023). 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.

III. Results

3.1 Stepwise Multiple Linear Regression Models (SMLR)

The values of prediction accuracy statistics of all SMLR based models can be found in Table 1. Initially the performance of Stepwise Multiple Linear Regression Model based on weather indices (SMLR-WI) was evaluated. Coefficient of determination (R^2) value was 0.87 which indicated that approximately 87% of the variation in soybean yield was explained by the predictors which were found to be significant (Z_{31} and Z_{381}). RMSE during calibration was found to be 215.74 kg/ha but on the other hand, RMSE during calibration was found to be 581.41kg/ha. nRMSE value during calibration was 10.60% and that of validation was 36.49%. MBE at calibration stage and validation stage was found to be -0.02 kg/ha and 379.35 kg/ha respectively. Decrease in R^2 value and increase in errors (RMSE, nRMSE and MBE) during validation were observed. SMLR-WI model performed consistently during calibration and validation. The error percentage ranged from -12.88% to 39.38%. Graphical analysis of predicted and observed values of yield for SMLR-WI model is shown in figure 4.

During the development of Stepwise Multiple Linear Regression Model based on direct weather variables (SMLR-W) the coefficient of determination (R^2) value was 0.99 which indicated that approximately 99.51% of the variation in soybean yield was explained by weather parameters at different growth stages of soybean. RMSE during calibration was found to be 41.20 kg/ha but on the other hand, RMSE during calibration was found to be 583.07 kg/ha. nRMSE value during calibration was 2.09% and that of validation was 36.60%. MBE at calibration stage and validation stage was found to be 0.02 kg/ha and 443.32 kg/ha respectively.

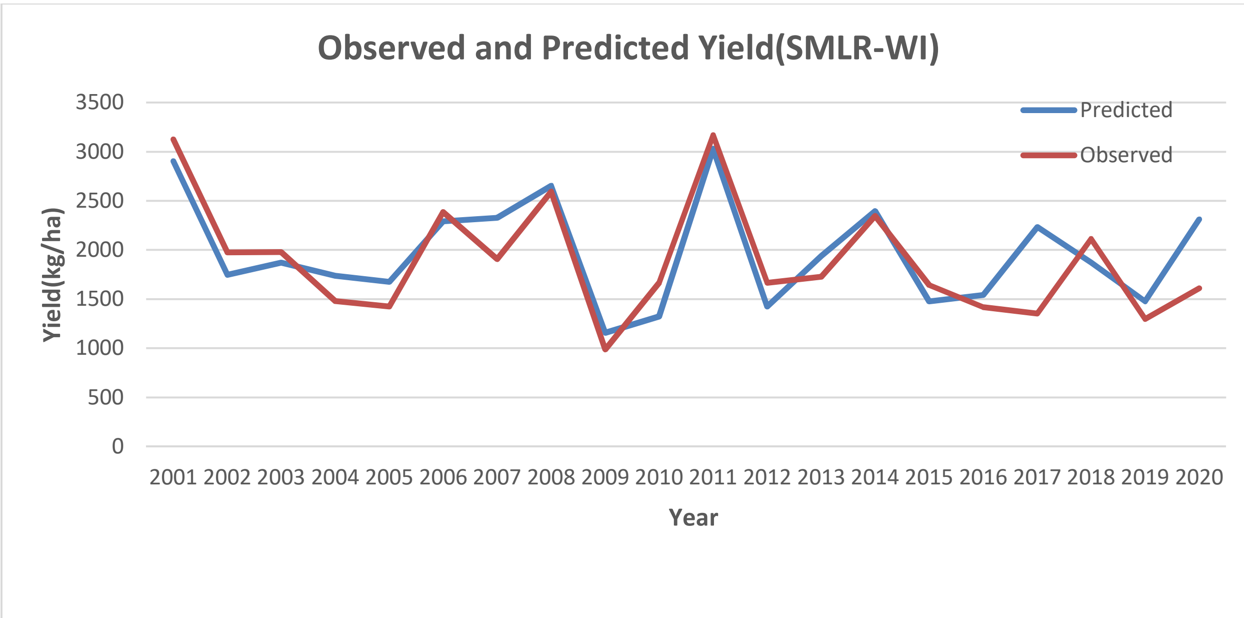

Fig. 4: Observed and Predicted Yield (SMLR-WI)

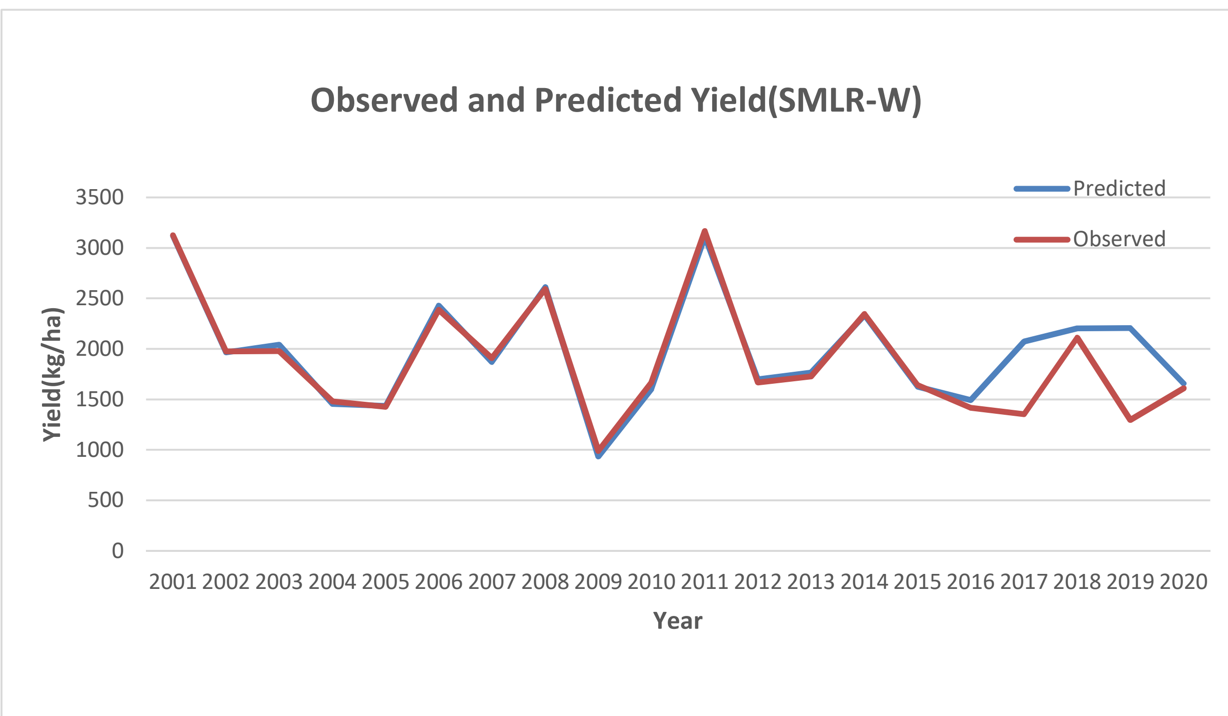

Fig. 5: Observed and Predicted Yield (SMLR-W)

Decrease in R^2 value and increase in errors (RMSE, nRMSE and MBE) during validation were observed. The SMLR-W model performed consistently during calibration and validation. The error percentage ranged from 2.91% to 41.26%. Graphical analysis of predicted and observed values of yield for SMLR-W model is shown in figure 5.

3.2 Principal Component Analysis-Stepwise Multiple Linear Regression Model (PCA-SMLR)

During the performance evaluation of Principal Component Analysis-Stepwise Multiple Linear Regression Model based on weather indices (PCA-SMLR-WI), the coefficient of determination (R^2) value was 0.81 which indicated that approximately 81% of the variation in soybean yield was explained by the predictors which were found to be significant (PC1). RMSE during calibration was found to be 259.30 kg/ha but on the other hand, RMSE during calibration was found to be 423.18 kg/ha. nRMSE value during calibration was 13.17% and that of validation was 26.56%. MBE at calibration stage and validation stage was found near to 0 and 118.86 kg/ha respectively. Decrease in R^2 value and increase in errors (RMSE, nRMSE and MBE) during validation were observed. SMLR-WI model performed consistently during calibration and validation. The error percentage ranged from -35.91% to 28.76%. Graphical analysis of predicted and observed values of yield for PCA_SMLR-WI model is shown in figure 6.

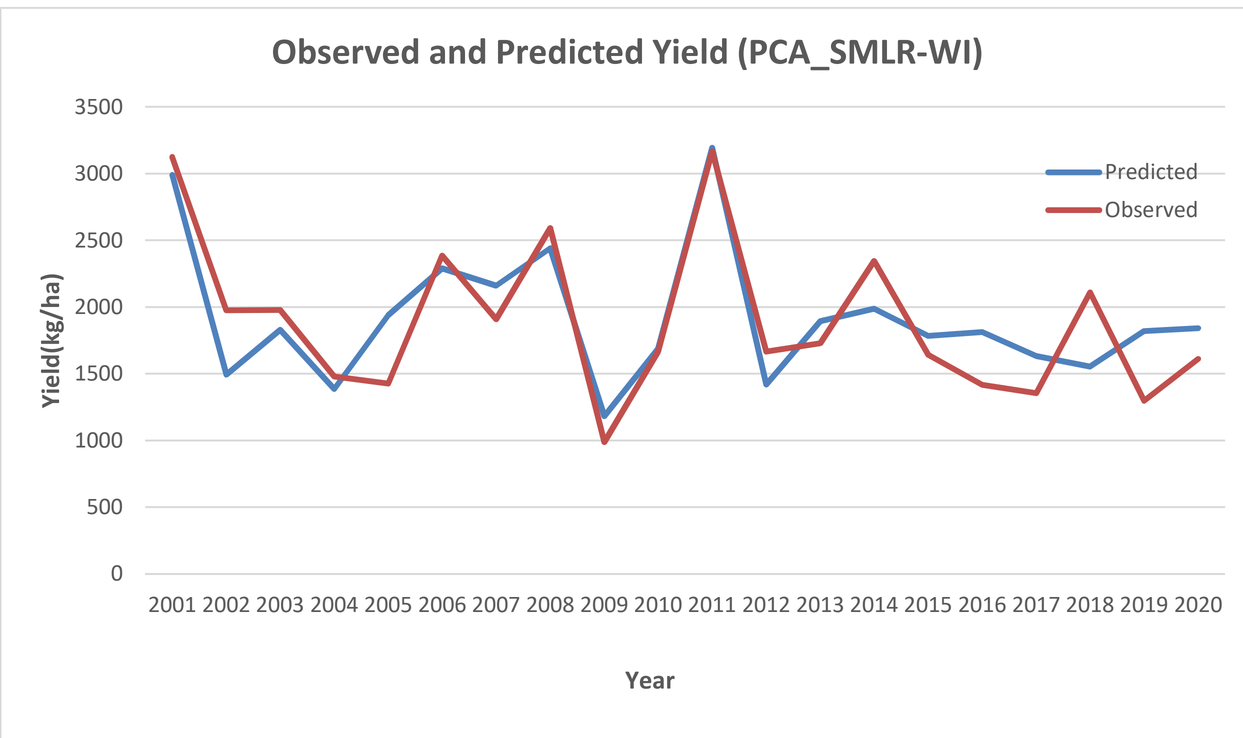

Fig. 6: Observed and Predicted Yield (PCA_SMLR-WI)

Coefficient of determination (R^2) value of Principal Component Analysis-Stepwise Multiple Linear Regression Model based on direct weather variables (PCA-SMLR-W) was 0.36 which indicated that approximately 36% of the variation in soybean yield was explained by the predictors which were found to be significant (PC15). RMSE during calibration was found to be 472.43 kg/ha but on the other hand, RMSE during calibration was found to be 851.55 kg/ha. nRMSE value during calibration was 24.00% and that of validation was 53.45%. MBE at calibration stage and validation stage was found to be near to 0 and 652.98 kg/ha respectively. Decrease in R^2 value and increase in errors (RMSE, nRMSE and MBE) during validation were observed. SMLR-W model performed consistently during calibration and validation. The error percentage ranged from -6.37% to 48.82%. Graphical analysis of predicted and observed values of yield for PCA_SMLR-W model is shown in figure 7.

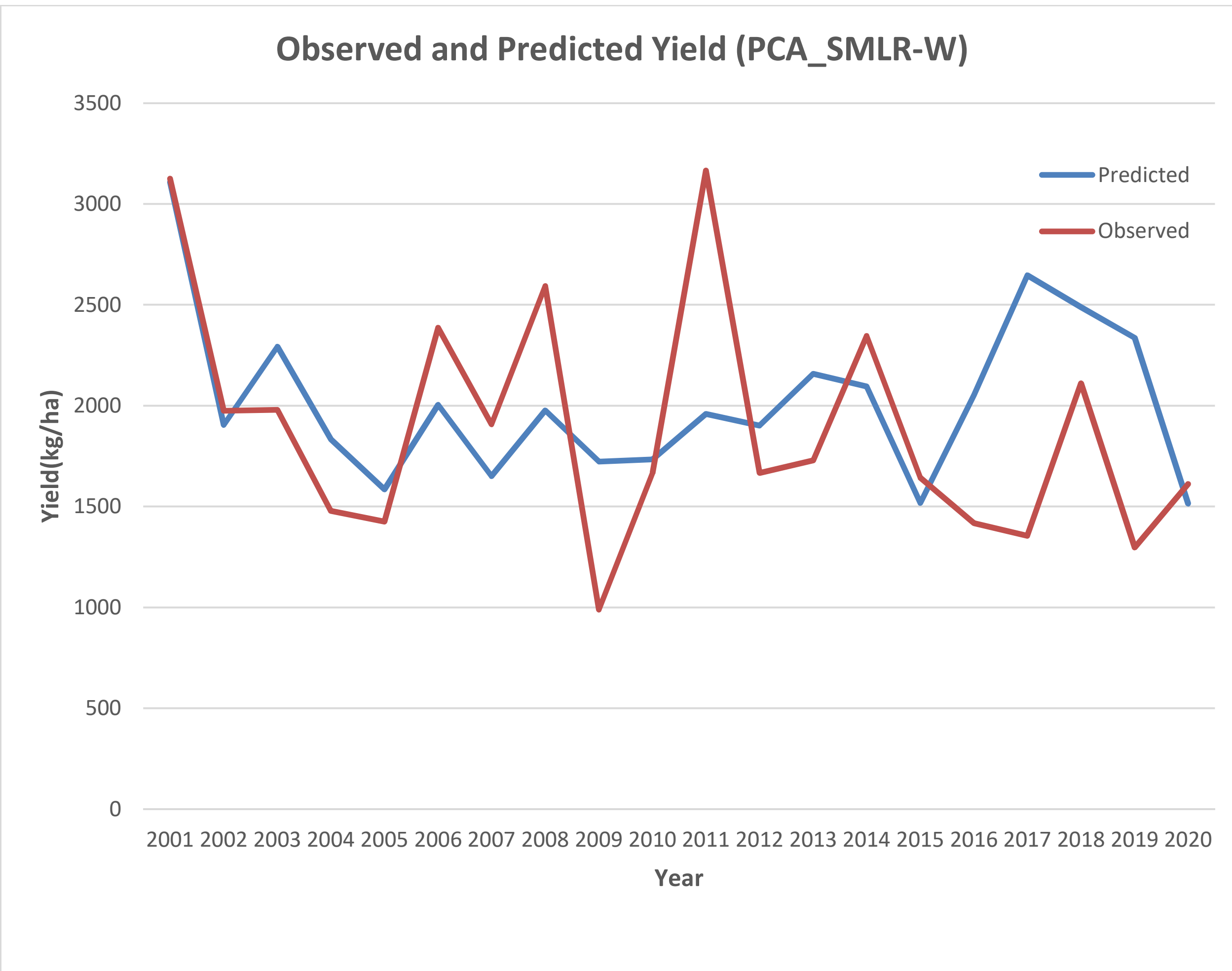

Fig. 7: Observed and Predicted Yield (PCA_SMLR-W)

Table 1: Quantitative measures obtained using SMLR models during calibration and validation

Model	Equation	R^2	MBE	RMSE	nRMSE	EF
	Calibration					
SMLR-WI	Y = - 15138.3 + 102.5 * Z_{31} + 1.5 * Z_{381}	0.87	-0.02	215.74	10.96	0.84
SMLR-W	Y = 7615.3 + 399.7 * Tmax41_PM - 655.9 * Tmin38_S - 31.907 * RHII37_S - 319.7 * Evap39_S + 203.2 * WV39_S -111.6 * Evap29_V + 0.9 * Rain30_V	0.99	0.02	41.20	2.09	0.99
PCA-SMLR-WI	Y = 1917.2 + 103.2 * PC1; No of PC's: 11	0.81	0.00	259.30	13.17	0.76
PCA-SMLR-W	Y = 2024.04 + 204.5 * PC15; No of PC's: 19	0.36	0.00	472.42	24.00	-0.73
	Validation					
SMLR-WI	Y= - 15138.3 + 102.5 * Z_{31} + 1.5 *Z_{381}	0.01	379.35	581.41	36.49	-2.08
SMLR-W	Y = 7615.3 + 399.7 * Tmax41_PM - 655.9 * Tmin38_S - 31.9 * RHII37_S - 319.7 * Evap39_S + 203.2 * WV39_S - 111.6 * Evap29_V + 0.9 * Rain30_V	0.01	443.32	583.07	36.59	-5.77
PCA-SMLR-WI	Y = 1917.2 + 103.2 * PC1; No of PC's: 11	0.35	118.86	423.18	26.56	-11.02
PCA-SMLR-W	Y = 2024 + 204.4 * PC15; No of PC's: 19	0.00	652.98	851.55	53.45	-2.81

Where, G = Germination, V = Vegetative, F = Flowering, Pod = Pod development, S = Seed development, PM = Plant maturity

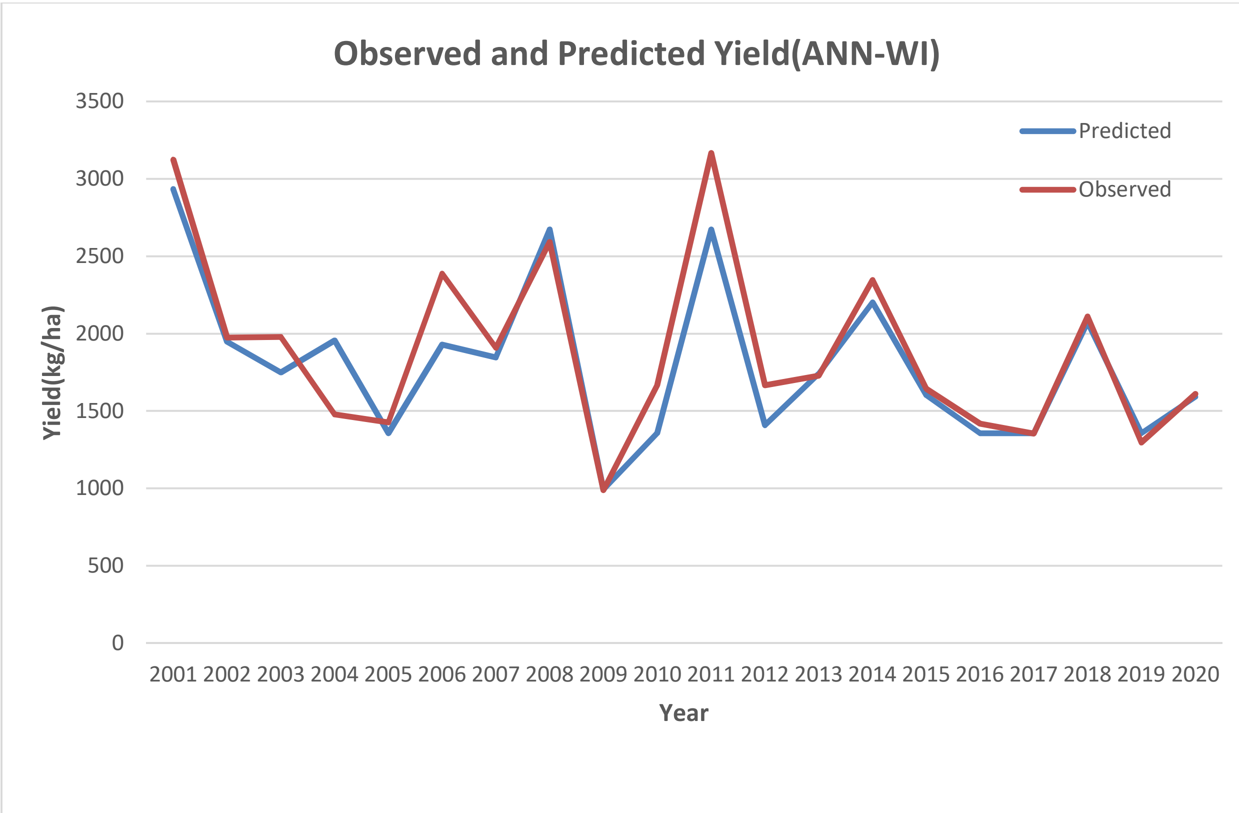

Fig. 8: Observed and Predicted Yield (ANN-WI)

3.3 Artificial Neural Network Models (ANN)

The values of prediction accuracy statistics of all ANN based models can be found in Table 2. The prediction accuracy indicated by the Coefficient of determination (R^2) value and RMSE value during calibration was found to be 0.86 and 246.95 kg/ha respectively by artificial neural network based on weather indices model (ANN-WI). R^2 value during validation was found to be 0.99 with RMSE value of 36.59 kg/ha. nRMSE value during calibration and validation was found to be 12.54% and 2.30% respectively. MBE at calibration stage and validation stage was found to be -110.82 kg/ha and 1.50 kg/ha respectively. Increase in R^2 value and decrease in errors (RMSE, nRMSE and MBE) during validation were observed. The performance of the model was good during calibration but excellent during validation. Error percentage ranged from -4.64% to 1.73%. Graphical analysis of predicted and observed values of yield for ANN-WI model is shown in figure 8. The variable importance of ANN-WI (10 most important indices) depicted in figure 9.

During the development of artificial neural network based on direct weather variables model (ANN-W), the value of coefficient of determination (R^2) and RMSE during calibration was found to be 0.73 and 315.75 kg/ha respectively. found to be -2.49 kg/ha and 24.08 kg/ha respectively.

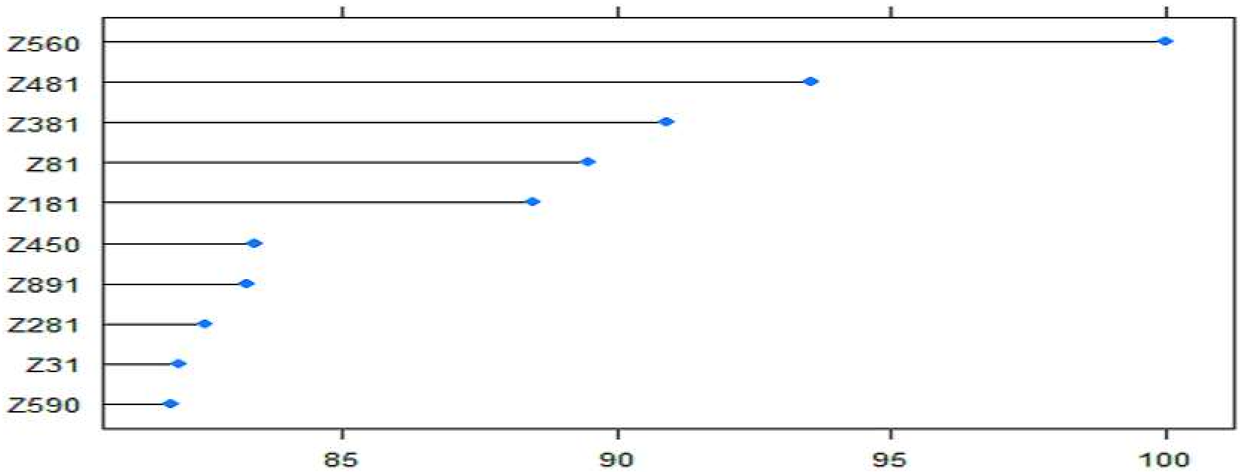

Fig. 9: Variable importance of ANN-WI model.

R^2 value during validation was found to be 0.95 with RMSE value of 83.52 kg/ha. nRMSE value during calibration and validation was found to be 16.04% and 5.24% respectively. MBE at calibration stage and validation stage was Increase in R^2 value and decrease in errors (RMSE, nRMSE and MBE) during validation were observed. The performance of model was good during calibration but excellent during validation. Error percentage ranged from -12.30% to 2.40%. Graphical analysis of predicted and observed values of yield for ANN-W model is shown in figure 10. The variable importance of ANN-W (10 most important indices) depicted in figure 11.

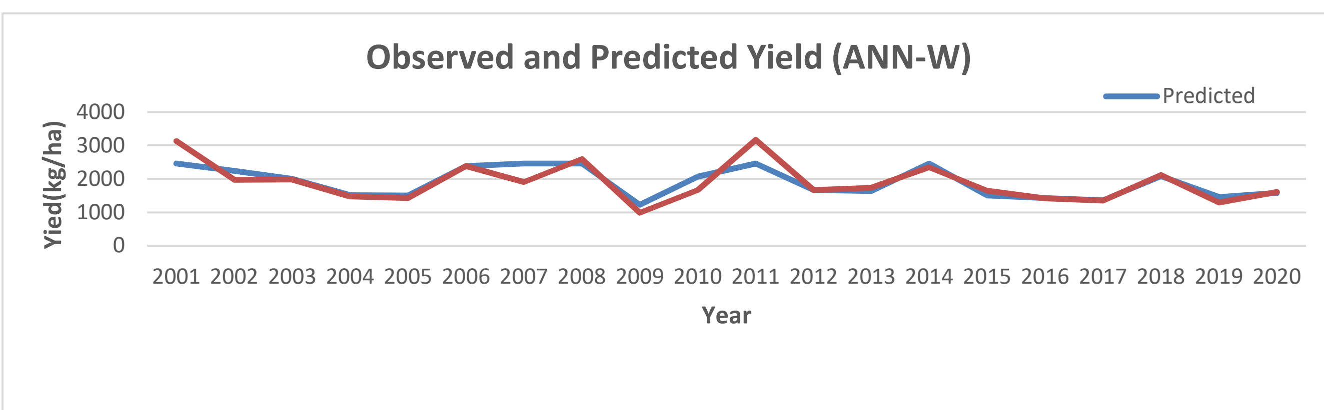

Fig. 10: Observed and Predicted Yield (ANN-W)

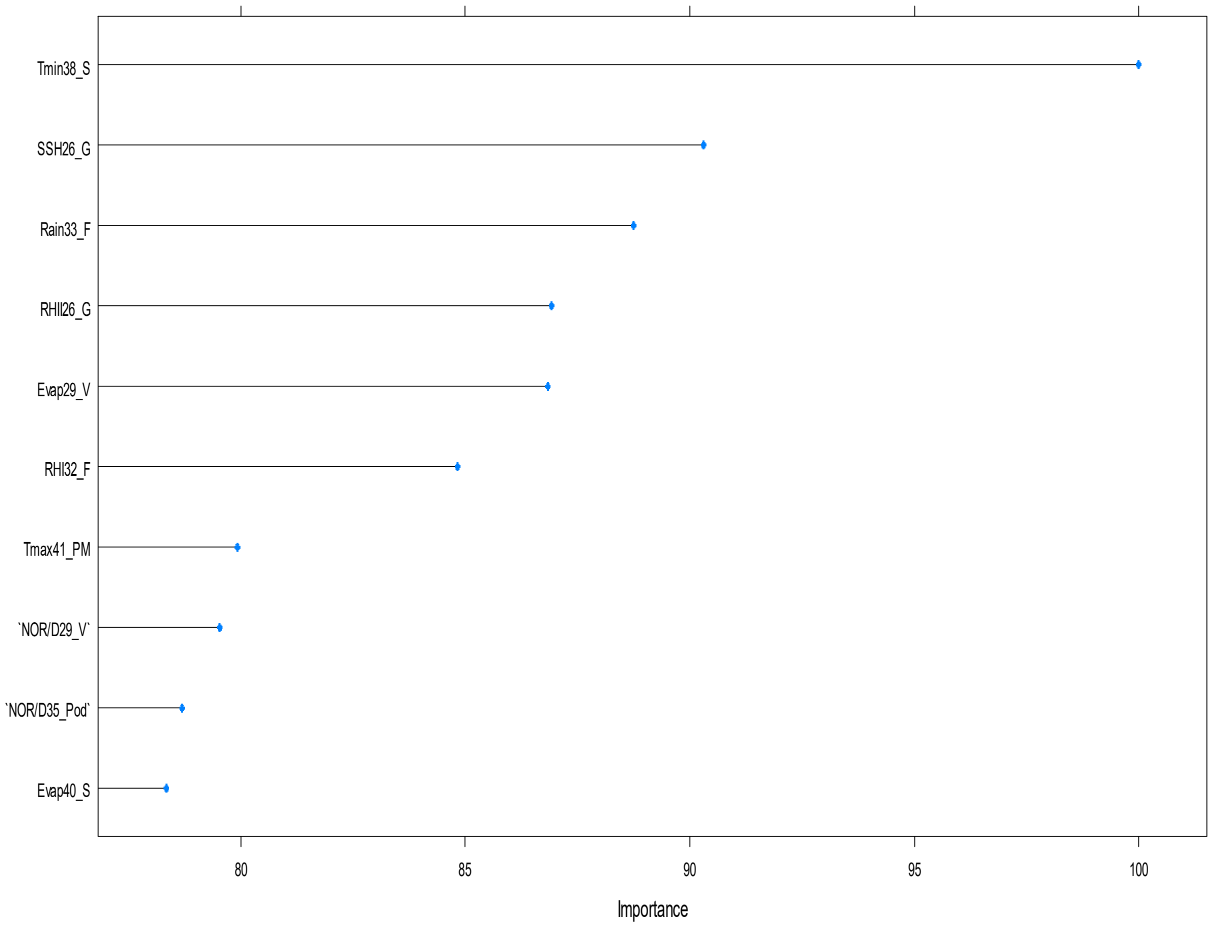

Fig. 11: Variable importance of ANN-W model

3.4 Principal Component Analysis-Artificial Neural Network Models (PCA-ANN): Principal Component Analysis-Artificial Neural Network Models based on weather indices (PCA-ANN-WI) model performance were in good approximation. The Coefficient of determination (R^2) value and RMSE value during calibration was found to be 0.74 and 360.69 kg/ha respectively. R^2 value during validation was found to be 0.88 with RMSE value of 172.83 kg/ha. nRMSE value during calibration and validation was found to be 18.32% and 10.85% respectively. MBE at calibration stage and validation stage was found to be -150.73 kg/ha and -109.75 kg/ha respectively. Increase in R^2 value and decrease in errors (RMSE, nRMSE and MBE) during validation were observed. Error percentage ranged from -3.43% to 16.70%. Graphical analysis of predicted and observed values of yield for PCA_ANN-WI model is shown in figure 12.The variable importance of PCA_ANN-WI (10 most important indices) depicted in figure 13.

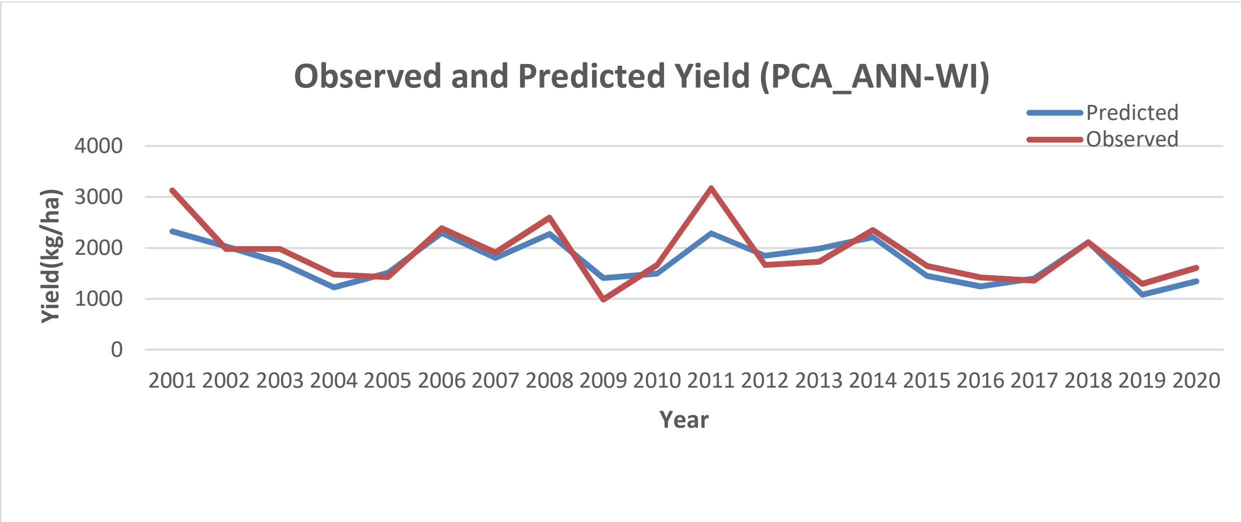

Fig. 12: Observed and Predicted Yield (PCA_ANN-WI)

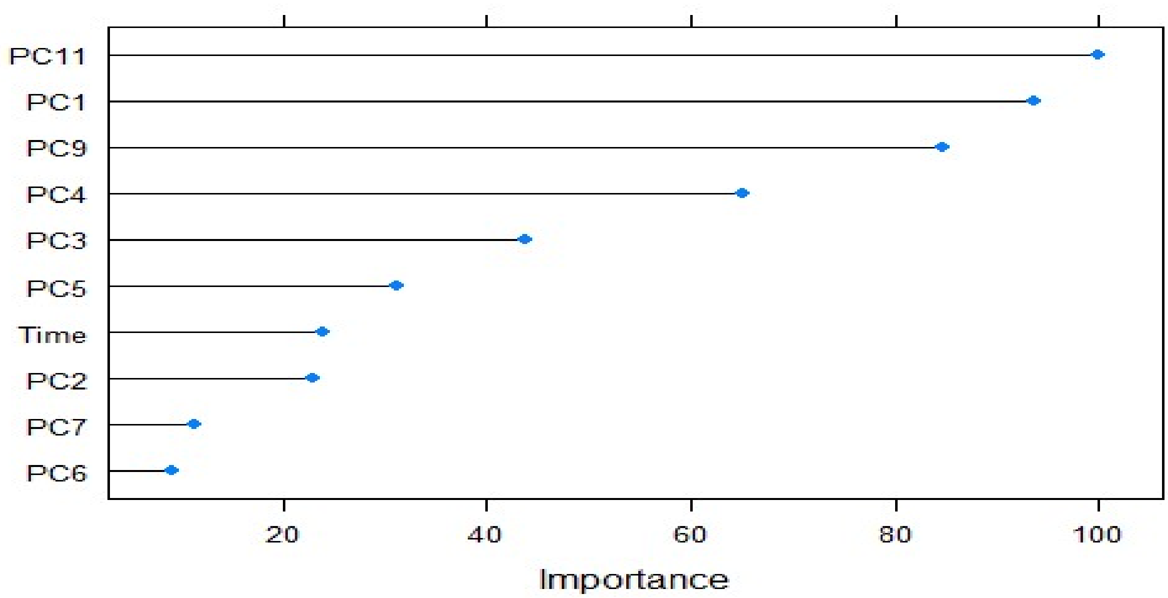

Fig. 13: Variable importance of PCA_ANN-WI model

The prediction accuracy indicated by the Principal Component Analysis-Artificial Neural Network Models based on direct weather variables (PCA-ANN-W) was found to be excellent. The coefficient of determination (R^2) and RMSE value during calibration was found to be 0.95 and 141.04 kg/ha respectively. R^2 value during validation was found to be 0.99 with RMSE value of 54.35 kg/ha. nRMSE value during calibration and validation was found to be 7.16% and 3.41% respectively. MBE at calibration stage and validation stage was found to be 98.29 kg/ha and 47.80 kg/ha respectively. Increase in R^2 value and decrease in errors (RMSE, nRMSE and MBE) during validation were observed. Error percentage ranged from -6.93% to 2.14%. Graphical analysis of predicted and observed values of yield for PCA_SMLR-W model is shown in figure 14.The variable importance of PCA_ANN-W (10 most important indices) depicted in figure 15.

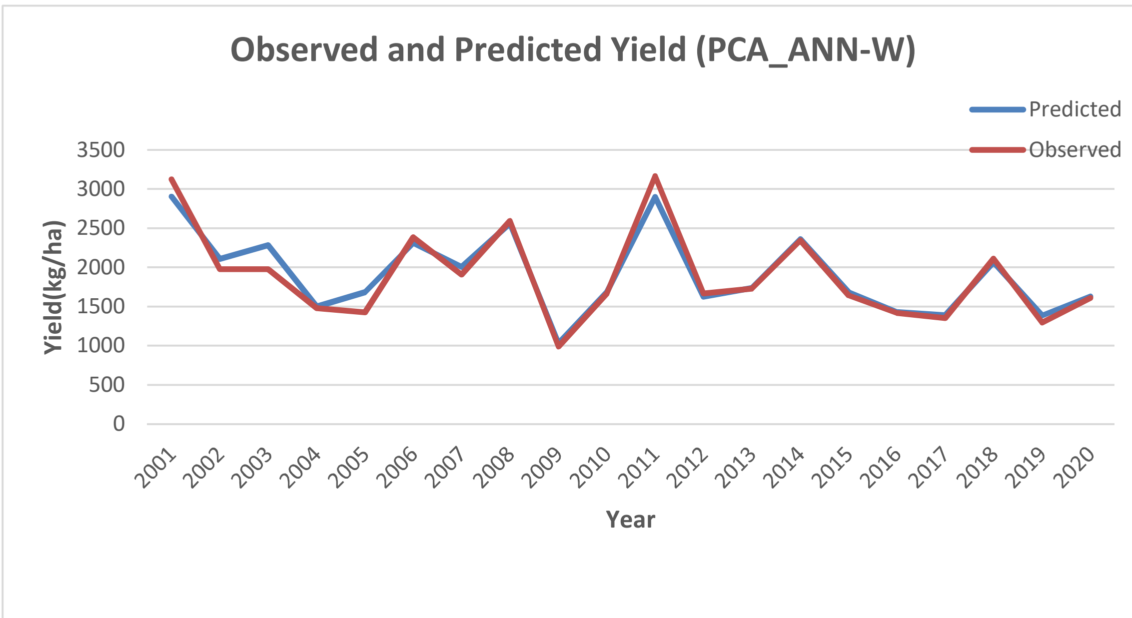

Fig. 14: Observed and Predicted Yield (PCA_ANN-W)

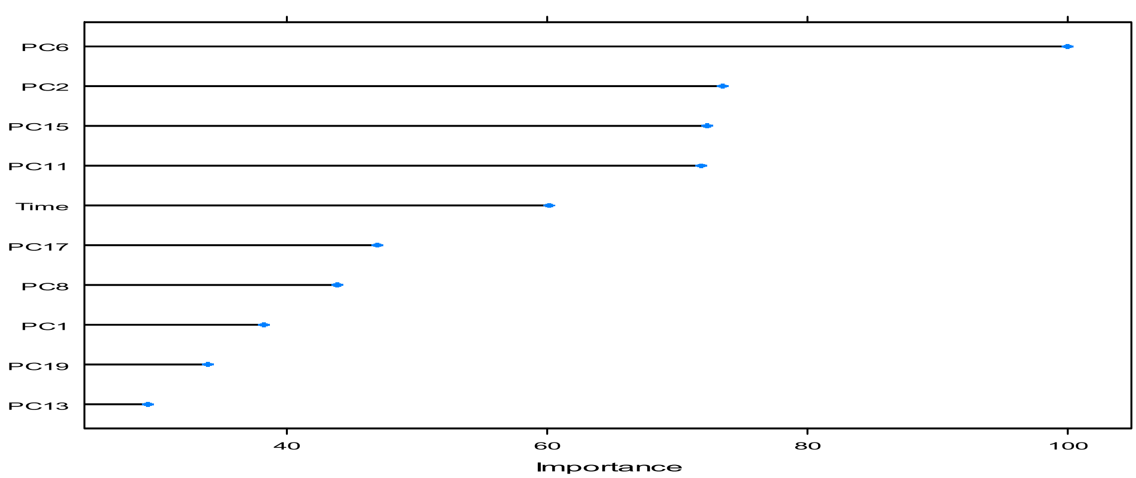

Fig. 15: Variable importance of PCA_ANN-W model

Table 2: Quantitative measures obtained using ANN models during calibration and validation

Model	No. of hidden Neurons	R^2	MBE	RMSE	nRMSE	EF
		Calibration				
ANN-WI	3	0.86	-110.82	246.95	12.54	0.83
ANN-W	4	0.72	-2.49	315.75	16.04	0.72
PCA-ANN-WI	3; No of PC's: 11	0.74	-150.73	360.69	18.32	0.63
PCA-ANN-W	13; No of PC's: 19	0.95	19.32	141.04	7.16	0.94
		Validation				
ANN-WI	3	0.99	1.50	36.59	2.30	0.99
ANN-W	4	0.95	24.08	83.52	5.24	0.93
PCA-ANN-WI	3; No of PC's: 11	0.88	-109.75	172.83	10.85	0.71
PCA-ANN-W	13; No of PC's: 19	0.99	25.18	54.35	3.41	0.97

IV. Discussion

All the developed models were compared based on the R^2 and nRMSE values provided in Table 3. Based on the table values SMLR-W and PCA-ANN-W model were found to be excellent during calibration, while SMLR-WI, PCA-SMLR-WI and ANN-WI performed good. During the validation stage PCA-ANN-W again performed excellent, becoming the best model for soybean prediction compared to other models in the study region. The overall ranking based on the performances of the models can be given as: PCA-ANN-W > ANN-WI > SMLR-W > SMLR-WI ≈ PCA-SMLR-WI > ANN-W > PCA-ANN-WI > PCA-SMLR-W. The study results indicated that PCA-ANN-W and ANN-WI model performed well for the study region. 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. 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 was better as compared to other models.

Table 3: Cross comparison of the model performances based on R^2 values

Model/Performance	R^2_{cal}	R^2_{val}	$nRMSE_{cal}$	$nRMSE_{val}$
SMLR-WI	Good	Poor	Good	Poor
SMLR-W	Excellent	Poor	Excellent	Poor
PCA-SMLR-WI	Good	Poor	Good	Poor
PCA-SMLR-W	Poor	Poor	Fair	Poor
ANN-WI	Good	Excellent	Good	Excellent
ANN-W	Fair	Excellent	Good	Excellent
PCA-ANN-WI	Fair	Good	Good	Good
PCA-ANN-W	Excellent	Excellent	Excellent	Excellent

V. Conclusion

In the present study, eight multivariate models were examined for soybean yield prediction based on different weather variables. The results revealed that the performance of PCA-ANN-W model was found to be best compared to other multivariate models considered in this study. The next best model was ANN-WI. Thus, it can be concluded from the present findings that PCA-ANN-W and ANN-WI were the best model for yield prediction of soybean in Pantnagar compared to SMLR-WI, SMLR-W, PCA-SMLR-WI, PCA-SMLR-W, ANN-W and PCA-ANN-WI models.

Conflict of Interests

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

Data availability

To be provided on request.

References

[1] A REFERENCE GUIDE TO IMPORTANT SOYBEAN FACTS & FIGURES, SOYSTATS. Accessed online (2022). http://soystats.com/wp-content/uploads/Soy-Stats-2022-Web.pdf (Accessed on 20 April 2023).

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

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

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

[5] Basso, B., Cammarano, D. and Carfagna, E. “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, Vol. 18, pp. 19, FAO Headquarters, Rome, Italy (2013).

[6] Bhartiya, A., Chandra, N., Pal, R. S., Aditya, J. P. and Bajeli, J. “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 (2017).

[7] Brejda, J. J., Moorman, T. B., Karlen, D. L. and Dao, T. H. “Identification of Regional Soil Quality Factors and Indicators, I. Central and Southern High Plain. Soil Sci. Soc. Am. J.”. 64, 2115–2124 (2000).

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

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

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

[11] Das, B., Nair, B., Arunachalam, V., Reddy, K. V., Venkatesh, P., Chakraborty, D., & Desai, S. “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 (2020).

[12] Das, B., Nair, B., Reddy, V. K. and Venkatesh, P. “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 (2018).

[13] FAO “Climate change and food security: risks and responses, Rome, Available online” (2015). https://www.fao.org/3/i5188e/I5188E.pdf (Accessed on 20 April 2023).

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

[15] Ihsan, M. Z., El-Nakhlawy, F. S., Ismail, S. M., Fahad, S. and Daur, I. “Wheat Phenological Development and Growth Studies As Affected by Drought and Late Season High Temperature Stress under Arid Environment. Frontiers”. (2016).

[16] Ji, Z., Yaozhong, P., Xiufang, Z., Jinyun, W. and Qiannan L. "Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. Sensors". 21(4), 1406 (2021).

[17] Kaul, M., Hill, R. L. and Walthall, C. "Artificial neural networks for corn and soybean yield prediction. Agric. Syst.". 85(1), 1-18 (2005).

[18] Khan, A., Ahmad, M., Ahmed, M. and Iftikhar H. M. "Rising atmospheric temperature impact on wheat and thermotolerance strategies. Plants". 10(1), 43 (2020).

[19] Khan, Y., Kumar, V., Setiya, P. and Satpathi, A. "Comparison of phenological weather indices based statistical, machine learning and hybrid models for soybean yield forecasting in Uttarakhand. J. Agrometeorol.". 25(3), *2583-2980 (2023)*. https://doi.org/10.54386/jam.v25i3.2232.

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

[21] Kumar, S. "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.". (2019).

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

[23] Mann, M. L., Warner, J. M. and Malik, A. S. "Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia. Clim. Change". 154, 211–227 (2019).

[24] Mishra, M., Bhavyashree, S. and Sharma, H. "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 (2017).

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

[26] Ransing, D. M., Verma, A. K., Meshram, M. R. and Paikra, P. R. "Nodulation, yield and thermal requirement of mungbean (Vigna radiata L.) genotypes as influenced by date of sowing. Int. J. Agric. Sci.". 10(2), 638-641 (2014).

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

[28] Seo, B., Lee, J., Lee, K. D., Hong, S. and Kang, S. "Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA. Field Crops Res.". 238, 113-128 (2019).

[29] Setiya, P., Satpathi, A., Nain, A. S. and Das, B. "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 (2022). https://doi.org/10.54386/jam.v24i3.1571.

[30] Singh R. S., Patel C, Yadav M. K. and Singh K. K. "Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. J. Agrometeorol.". 16: 199–202 (2014). https://doi.org/10.54386/jam.v16i2.1521.

[31] Sridhara, S., Manoj, K. N., Gopakkali, P., Kashyap, G. R., Das, B., Singh, K. K. and Srivastava, A. K. "Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. International Journal of Biometeorology". 67(1), 165-180 (2023).

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

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