Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 24(3) : 255-261 (September 2022)

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

Research Paper

Comparison of weather-based wheat yield forecasting models for different districts of Uttarakhand using statistical and machine learning techniques

PARUL SETIYA[1]*, ANURAG SATPATHI[1], AJEET SINGH NAIN[1] and BAPPA DAS[2]

[1]*Department of Agrometeorology, College of Agriculture, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India*
[2]*ICAR Central Coastal Agricultural Research Institute, Old Goa, India*
**Corresponding author's e-mail: parul.setiya@gmail.com*

ABSTRACT

The prediction of crop yield before harvest is crucial for facilitating the formulation and implementation of policies about food safety, transportation cost, and import-export, storage and marketing of agro-products. The weather plays a crucial role in crop growth and development. Therefore, models using weather variables can provide reliable forecasts for crop yield and choosing the right model for crop production forecasts can be difficult. Therefore in the present study, an attempt was made to find the best model for wheat yield forecast by using five different techniques viz. Stepwise Multiple Linear Regression (SMLR), Artificial Neural Network (ANN), Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ELNET) and Ridge regression. Historical wheat yield data (taken from the Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare) and weather data of past 18-20 years were collected for seven different districts of Uttarakhand. Analysis was carried out by fixing 80% of the data for calibration and remaining dataset for validation. The present study concluded that the performance of ANN was good for crop yield forecasting as compared to the other models based on the value of RMSE (0.005 - 0.474) and nRMSE (0.166 - 26.171).

Keywords : Stepwise multiple linear regression, ANN, least absolute shrinkage and selection operator, elastic net, ridge regression

The livelihood of major proportion of the Indian population depends upon agriculture. There are various factors that may affect the crop yield such as soil texture, genotype, weather conditions and management practices. Compared to the other factors, crop yield is highly influenced by weather conditions (Singh *et al*., 2014). In many countries, losses due to weather conditions account for up to 30% of the annual agriculture production (Attri and Rathore, 2003). Therefore, there is a high demand to develop models that give accurate yield prediction before harvest which can be used by the government, policymakers and farmers for making advance planning and strategies. India has a variety of food grain and amongst them; wheat (*Triticum aestivum* L.) is the second most important crop after rice and plays a significant role in food security. So, pre-harvest prediction of wheat yield may help to achieve food security. Traditionally, crop cutting experiment method was used to estimate the crop yield. But it consumes a lot of time and needs more human effort. The other alternate of this traditional method is the crop yield estimation by models developed using various statistical techniques.

In the current scenario, forecasting of crop yield using Artificial Neural Network (ANN), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ELNET) getting a great deal of attention (Das *et al*., 2020, Aravind *et al*., 2022). Various efforts have been made by the researchers to develop pre-harvest yield forecast models based on yield and weather-datasets. Das *et al*. (2018) have developed forecast models for rice yield for fourteen districts of West Coast using stepwise multiple linear regressions (SMLR), principal component analysis together with SMLR (PCA-SMLR), LASSO and ELNET, ANN and PCA-ANN based on the monthly weather indices. Singh *et al.* (2014) used the technique of SMLR to develop yield forecast model based on the weekly weather indices and yield dataset for the Amritsar, Bhatinda and Ludhiana districts of Punjab. Safa *et al.* (2015) used the approach of ANN to model the wheat production based on the dataset of 40 farms in Canterbury, New Zealand. The researchers found that the ANN model developed was capable of predicting wheat production. Vashisth *et al*. (2014) developed the pre-harvest yield forecast model for wheat crop based upon the weather indices. The researchers had also forecast the yield based on the developed model and found the results quite satisfactory.

Article info - DOI: *https://doi.org/10.54386/jam.v24i3.1571*
Received: 08 March 2022; Accepted: 27 June 2022; Published online: 31 August 2022

Sridhara *et al*. (2020) also applied LASSO, ELNET, PCA, ANN and SMLR modeling techniques to forecast district level yield of sorghum crop. Aravind *et al.* (2022) developed the weather indices based wheat yield forecasting model for five different locations viz. Hisar, Ludhiana, Amritsar, Patiala and New Delhi by applying SMLR, PCA-SMLR, PCA-ANN, ANN, LASSO and ELNET techniques. The researchers found that out of six different models, ELNET and LASSO was the best model followed by PCA-SMLR, SMLR, PCA-ANN and ANN respectively for wheat yield prediction. The performance of the models varied according to the region and datasets used. So, there is no universal single model identified for crop yield forecasting.

In the wake of climate change, population growth, and food demand, a timely, accurate, and reliable crop yield estimation is very essential (Cao *et al*., 2021). Hence, in the present study, an attempt was made to find the best forecast model, by taking the major spring wheat production regions of Uttarakhand, India. This study was carried out to assess the pre-harvest forecast of wheat crop for Udham Singh Nagar, Nainital, Haridwar, Dehradun, Champawat, Tehri-Garhwal and Pauri Garwal districts of Uttarakhand by using SMLR, ANN, LASSO, ELNET and Ridge regression.

MATERIAL AND METHOD

Among all the district of Uttarakhand, seven major districts that produce wheat crop has been considered in the study i.e. Udham Singh Nagar (28° 57′ N, 79° 30′ E), Nainital (29° 23′ N, 79° 27′ E), Haridwar (29° 56′ N, 78°09′E), Dehradun (30° 18′ N, 78° 01′E), Champawat (29° 20′ N, 80° 05′E), Tehri-Garhwal (30° 22′ N, 78° 28′E) and Pauri Garhwal (29°52′ N,78°50′E).

Weather data of twenty-two years (1998-2019) for Udham Singh Nagar, nineteen years for Tehri-Garhwal and eighteen years for Dehradun (2001-2018) were taken from the local observatory situated at the respective districts. For rest of the districts, the weather data were downloaded from the NASA POWER web portal (https://power.larc.nasa.gov/data-access-viewer/). Out of the total dataset, 80% data was used for calibration purpose, and remaining 20% data were used for the validation of the developed model (Li *et al*., 2017, Montaseri *et al.*, 2018, Rajaee *et al*, 2018).

Time-Series data of the wheat crop for these districts were taken from the Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare. Since the variability in the crop yield time series data over the long period may be influenced by various factors such as technology differences, and climatic variability etc. and this may lead to non-stationary trends. The variability should be removed before conducting the analysis (Sridhara *et al*., 2020, Das *et al*, 2020). Therefore, the yield data were detrended before performing the statistical analysis. Weekly average was calculated from the daily weather data. These average values are then used for the calculation of weighted and unweighted weather indices.

Here, n is the week of forecast, $X_{iw} / X_{i'w}$ is the value of i^{th}/ i'^{th} weather variable and $r^j_{iw} / r^j_{i'w}$ is the value of correlation coefficient of detrended yield with i^{th} weather variable/ product of i^{th} and i'^{th} weather variables in w^{th} week. By following the above procedure, 42 weather indices were generated as shown in Table 1.

The steps involved in the model development are illustrated in Fig. 1. At the first step, weekly average was calculated from the daily weather data and then by using this weekly data weather indices were computed. Then these weather indices along with the crop yield data were used to form the forecast models by using different multivariate techniques.

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

Parameter	Unweighted Weather indices						Weighted Weather indices					
	Tmax	Tmin	RF	SR	RHI	RHII	Tmax	Tmin	RF	SR	Tmax	Tmin
Tmax	Z_{10}						Z_{11}					
Tmin	Z_{120}	Z_{20}					Z_{121}	Z_{21}				
RF	Z_{130}	Z_{230}	Z_{30}				Z_{131}	Z_{231}	Z_{31}			
SR	Z_{140}	Z_{240}	Z_{340}	Z_{40}			Z_{141}	Z_{241}	Z_{341}	Z_{41}		
RH I	Z_{150}	Z_{250}	Z_{350}	Z_{450}	Z_{50}		Z_{151}	Z_{251}	Z_{351}	Z_{451}	Z_{51}	
RH II	Z_{160}	Z_{260}	Z_{360}	Z_{460}	Z_{560}	Z_{60}	Z_{161}	Z_{261}	Z_{361}	Z_{461}	Z_{561}	Z_{61}

Ghosh *et al*., (2014) and Das *et., al*. (2018) computed the weighted and unweighted weather indices by using the following equations (i) and (ii), respectively.

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

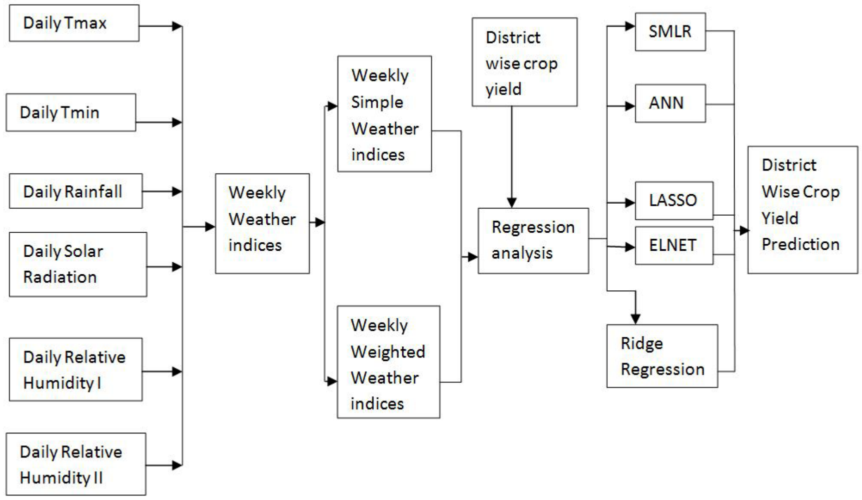

Fig. 1: Steps involved in model development

Multivariate techniques involved in model development

A comparison was made among different methods, i.e. SMLR, ANN, LASSO, ELNET and Ridge regression, by integrating station data and publicly available data of weather and crop yield.

Stepwise multiple linear regression (SMLR)

The SMLR technique is the simplest method to develop the yield forecast model based on the dataset of yield and weather parameters. This method allows choosing the best predictors among the large number of predictors (Singh *et al*., 2014; Das *et al*., 2018). In the present study, the *p*-value of 0.50 and 0.10 was considered for addition and removal of the variables respectively.

Least absolute shrinkage and selection operator (LASSO)

The LASSO is a powerful method that performs two major tasks: regularization and variable selection. This method helps to avoid over fitting by selecting LASSO applies a shrinking process, which penalizes the coefficients of the regression variables and shrinks some of the coefficients to zero. Thus, in this process the remaining input variables with non-zero coefficient after the shrinkage process get selected to be part of the model. The purpose of LASSO is to minimize the forecasting error (Kumar *et al*., 2019). LASSO regression uses L1 regularization technique. For LASSO the loss is defined as

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^{n}\left(y_i - x_i^j\hat{\beta}\right)^2 + \lambda\sum_{j=1}^{m}\left|\hat{\beta}_J\right| \qquad \text{(iii)}$$

Ridge regression

Ridge regression is a technique to reduce the over fitting of the data by incorporating little amount of bias to the regression estimates. The main purpose of applying the Ridge regression is to get more reliable estimates. Ridge regression perform consistently well on the training and testing dataset. It uses L2 regularization technique. For Ridge regression the loss is defined as

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}\left(y_i - x_i^j\hat{\beta}\right)^2 + \lambda\sum_{j=1}^{m}\beta_j^2 \qquad \text{(iv)}$$

Elastic net (ELNET)

In ELNET model, penalty of LASSO and Ridge regression get combined. ELNET regression uses both L1and L2 regularization techniques of LASSO and ridge regression to improve model performance (Abbas *et al*., 2020). L1 Regularization, also known as lasso regression, adds the "absolute value of magnitude" of the coefficient as a penalty term to the loss function. L2 Regularization, also known as ridge regression, adds the "squared magnitude" of the coefficient as the penalty term to the loss function. For ELNET regression the loss is defined as

$$L_{elne} = \frac{\sum_{i=1}^{n}\left(y_i - x_i^j\hat{\beta}\right)^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{m}\left|\hat{\beta}_j\right|\right) \qquad \text{(v)}$$

Artificial neural network (ANN)

Artificial neural network (ANN) is a type of non-linear machine learning technique. It consists of three layers i.e. input, hidden and output layer. In this technique, data from the input layer move through the hidden to the output layer (Kaul *et al*., 2005). Number of nodes in the input layer depends upon the number of independent predictors. The functioning of ANN is similar to that of the human brains biological neural process.

Evaluation of model performance

The statistical measures used to select the best model were R^2, root mean square error (RMSE), normalised root mean square error (nRMSE), Mean Biased Error (MBE) and modelling efficiency (EF). Formulas of the following measures are shown below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}, \quad nRMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \times \frac{100}{\bar{A}}$$

$$R^2 = \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sigma_y \sigma_{\hat{y}}}\right)^2, \quad EF = \left(1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right), \quad MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

Here y_i is the actual value and $\hat{y}_i$ is the predicted value for i=1, 2,……, n. σ_y and $\sigma_{\hat{y}}$ are the standard deviation of actual and predicted yield. The value of R^2 and EF close to 1 and the value of RMSE near 0 indicate better model performance. Whereas the model is considered as excellent, good, fair and poor based on the value of nRMSE lies between 0-10%,10-20%, 20-30% and >30%, respectively. Moreover, positive value of MBE indicates over estimation and negative value indicates underestimation.

RESULTS AND DISCUSSION

Stepwise multiple linear regression (SMLR)

In stepwise multiple linear regression (SMLR) models, at calibration stage, the value of R^2 ranged between 0.58 to 0.95 (Table 2). The maximum R^2 was observed for Tehri-Garhwal (0.95) with RMSE value of 0.265 t/ha and minimum was observed for Udham Singh Nagar (0.58) with RMSE value of 0.795 t/ha. The results of validation indicated that the SMLR performance was excellent for Udham Singh Nagar with nRMSE value of 7.28% and good for Tehri-Garhwal district with nRMSE 12.47%. For rest of the districts the model performance was poor (nRMSE>30%). In all districts except Udham Singh Nagar, the MBE value at calibration stage shows little overestimation of crop yield.

Table 2: Performance of the model developed using SMLR technique

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF	R^2_{val}	$RMSE_{val}$	$nRMSE_{val}$	MBE_{val}
USNagar	0.58	0.795	18.630	0.757	-4.20	0.80	0.302	7.279	-0.063
Nainital	0.78	0.222	8.633	0.180	0.36	0.41	0.942	38.376	0.870
Haridwar	0.59	0.180	7.028	0.075	0.49	0.40	1.092	48.151	1.081
Dehradun	0.83	0.132	6.617	0.002	0.83	0.05	0.624	33.31	0.434
Champawat	0.75	0.234	18.028	-0.184	-0.31	0.65	0.615	37.933	0.577
Tehri-Garhwal	0.95	0.504	39.363	-0.491	-4.51	0.85	0.203	12.467	0.146
Pauri-Garhwal	0.82	0.265	24.895	0.242	-0.06	0.62	0.780	84.84	0.747

LASSO (Least absolute shrinkage and selection operator)

The data on LASSO analysis of wheat crop yield revealed that the value of R^2 ranged between 0.88 to 1.00 (Table 3). The highest R^2 was observed for Haridwar district (R^2= 1.00) with RMSE value of 0.014 t ha^{-1}, whereas the lowest R^2 was observed for Champawat (0.88) with RMSE value of 0.100 t ha^{-1}. On the other side, in validation stage, the R^2 value ranged from 0.36 to 0.92. The highest R^2 was observed for Tehri-Garhwal district (0.92) with RMSE value of 0.598 t ha^{-1}, whereas the lowest R^2 was observed for Dehradun (0.36) with RMSE value of 0.885 t ha^{-1}. The nRMSE value at calibration stage shows that the model performance was excellent for all the districts, on contrary at the validation stage performance of LASSO model was good for Udham Singh Nagar and Nainital, fair for Haridwar and poor for Dehradun, Champawat, Tehri-Garhwal and Pauri-Garhwal. The modelling efficiency ranged from 0.78 for Champawat to 1.00 for Haridwar district. The MBE value for all the districts indicate good model performance at calibration stage, though at the validation stage it showed the under estimation of the crop yield for all the district except for Udham Singh Nagar.

Table 3: Performance of the model developed using LASSO technique

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF	R^2_{val}	$RMSE_{val}$	$nRMSE_{val}$	MBE_{val}
USNagar	0.95	0.089	2.527	3.207e-16	0.94	0.84	0.834	20.127	0.000
Nainital	0.95	0.068	2.647	5.625e-16	0.94	0.61	0.514	15.470	-0.385
Haridwar	1.00	0.014	0.551	-6.217e-16	1.00	0.46	0.793	23.665	-0.781
Dehradun	0.95	0.081	4.045	-1.110e-16	0.94	0.36	0.885	38.354	-0.640
Champawat	0.88	0.100	7.698	-1.036e-16	0.78	0.62	0.906	41.201	-0.862
Tehri-Garhwal	0.93	0.062	4.860	-2.442e-16	0.92	0.92	0.598	33.663	-0.562
Pauri-Garhwal	0.96	0.052	4.859	-5.181e-17	0.96	0.65	0.572	34.336	-0.544

Table 4: Performance of the model developed using Ridge regression technique

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF	R^2_{val}	$RMSE_{val}$	$nRMSE_{val}$	MBE_{val}
USNagar	0.82	0.232	6.594	-4.935e-17	0.57	0.99	0.681	16.368	-0.677
Nainital	0.74	0.182	7.057	5.033e-16	0.57	0.88	0.783	23.536	-0.707
Haridwar	0.45	0.23	9.075	-4.441e-16	0.17	0.27	0.805	24.020	-0.785
Dehradun	0.92	0.155	7.719	-1.586e-16	0.78	0.16	0.743	32.189	-0.534
Champawat	0.80	0.118	9.075	-7.402e-17	0.67	0.80	0.873	39.700	-0.841
Tehri-Garhwal	0.81	0.125	9.726	-2.294e-16	0.66	0.65	0.575	32.423	-0.520
Pauri-Garhwal	0.87	0.126	11.763	3.701e-18	0.76	0.70	0.569	34.166	-0.544

Table 5: Performance of the model developed using ELNET techniques

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF	R^2_{val}	$RMSE_{val}$	$nRMSE_{val}$	MBE_{val}
USNagar	0.94	0.100	2.856	9.868e-16	0.91	0.82	0.750	18.028	-0.733
Nainital	0.97	0.049	1.902	3.552e-16	0.97	0.51	0.473	14.246	-0.326
Haridwar	0.99	0.018	0.714	-3.256e-16	0.99	0.46	0.799	23.862	-0.787
Dehradun	0.94	0.092	4.603	-3.013e-16	0.92	0.26	1.191	51.591	-0.782
Champawat	0.91	0.067	5.139	-1.480e-17	0.89	0.59	0.933	42.415	-0.892
Tehri-Garhwal	0.92	0.069	5.380	-7.179e-16	0.90	1.00	0.587	33.044	-0.548
Pauri-Garhwal	0.94	0.066	6.143	7.403e-18	0.94	0.70	0.565	33.905	-0.541

Ridge regression

The performance of the wheat yield forecasting model developed using Ridge regression was shown in Table 4. The maximum R^2 value was recorded for Dehradun (0.92) district with RMSE of 0.155 t ha^{-1}, while the minimum R^2 was recorded for Haridwar (0.45) with RMSE of 0.23 t ha^{-1}. While during validation, maximum R^2 was observed for Udham Singh Nagar (0.99) with RMSE of 0.681 t ha^{-1}, and the minimum R^2 was recorded for Dehradun (0.158) with RMSE of 0.743 t ha^{-1}. Additionally, the nRMSE statistic shows that the performance of ridge regression model was good for Udham Singh Nagar (nRMSE = 16.37%), fair for Nainital (nRMSE = 16.37%) and Haridwar (nRMSE = 16.37%), while model performed poor for Dehradun, Champwat, Tehri-Garhwal and Pauri-Garhwal, nRMSE ranging from 23.54 to 39.70%. The modelling efficiency ranged from 0.17 for Haridwar to 0.78 for Dehradun district. Despite a good model performance at the calibration stage, the MBE value for all the districts indicated underestimation of the crop yield at the validation stage.

ELNET (Elastic Net)

The data pertaining to ELNET model performance is shown in Table 5. Based on the value of R^2, RMSE and nRMSE and EF the performance of ELNET was excellent for all the districts considered in the study. The results showed that the maximum and minimum R2 were recorded in Haridwar (0.99) and Champawat (0.91) districts. The RMSE of calibration dataset ranged from 0.018 to 0.100 t ha^{-1}. During the validation stage, RMSE ranged from 0.473 to 1.191 t ha^{-1} and the value of nRMSE ranged between 14.25% and 51.59%. The modelling efficiency is close to 1 for all the districts, indicate good model performance. The MBE value indicates good model performance at the calibration stage, but at the validation stage it indicates that crop yields are underestimated in each district.

Artificial neural network (ANN)

Results of the analysis showed that the performance of Artificial Neural Network (ANN) was excellent for Haridwar and Dehradun district with R^2 value of 1.00 and 0.874 with RMSE 0.003 t ha^{-1} and 0.123 t ha^{-1} respectively during calibration, while during validation R^2 was 1.00 and 0.93 with RMSE of 0.005 t ha^{-1} and 0.121 t ha^{-1} respectively. In addition to this, the value of nRMSE for Haridwar and Dehradun district was less than 10%, indicating excellent model performance. Table 6 shows the performance of the ANN for all the districts. The modelling efficiency ranged from 0.12 for Champawat to 0.99 for Haridwar district. The MBE value close to zero for all the districts indicate good model performance at calibration stage, though at the validation stage it shows the under estimation of the crop yield for all the district.

Table 6: Performance of the model developed using ANN techniques

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF	R^2_{val}	$RMSE_{val}$	$nRMSE_{val}$	MBE_{val}
USNagar	0.74	0.202	5.736	0.040	0.67	0.81	0.392	9.410	-0.384
Nainital	0.76	0.155	6.026	0.075	0.69	0.96	0.465	13.992	-0.401
Haridwar	1.00	0.003	0.122	0.001	0.99	1.00	0.005	0.166	-0.002
Dehradun	0.87	0.123	6.142	0.021	0.82	0.93	0.121	5.283	-0.099
Champawat	0.49	0.192	14.739	0.107	0.12	0.88	0.445	20.249	-0.423
Tehri-Garhwal	0.69	0.143	11.165	0.076	0.56	0.87	0.368	20.723	-0.306
Pauri-Garhwal	0.77	0.155	14.400	0.085	0.62	0.69	0.474	26.171	-0.455

The study results indicated that based on the range of nRMSE, the model developed by SMLR and ANN performed excellently for Udham Singh Nagar (USN), while LASSO, ELNET and ridge regression showed good model performance for USN. For Nainital location, the performance of LASSO, ELNET and ANN was good, while ridge regression perform fairly for Nainital, though SMLR performed worst for Nainital. This finding was similar to the finding of Singh *et al.* (2019), who observed that the performance of LASSO was better as compared to SMLR. The study also revealed that the performance of LASSO, ELNET and ridge regression was fair for district Haridwar, as compared to SMLR. Addition to these, it was also observed that ANN performed better for Dehradun, Champawat, Tehri-Garhwal and Pauri-Garhwal districts as compared to the other methods, included in the study. This finding was in line with the study done by Arvind *et al.* (2022), which concluded that the performance of ANN was better as compared to SMLR, PCA-SMLR, LASSO and ELNET for Patiala district. Uno *et al.,* (2005) also found ANN as a potential tool for developing in-season yield mapping and forecasting systems for corn in eastern Canada. They found that ANN yield models achieved better prediction accuracy (about 20% validation RMSE) than conventional models.

CONCLUSION

In the present study, five different methods viz, SMLR, LASSO, ELNET, Ridge regression and ANN was used to study the relationship of yield with weather parameters for six districts of Uttarakhand, India. The overall ranking based on RMSE and nRMSE value during validation revealed that ANN performed the best as compared to other models. The study also showed that the performance of SMLR, LASSO, ELNET, ridge regression was good during calibration but not good during validation, which may be due to the overfitting of the data. So, ANN can be used to forecast the wheat yield for the studied region.

ACKNOWLEDGEMENT

The authors acknowledge NASA Langley Research Center POWER Project funded through the NASA Earth Science Directorate Applied Science Program.

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

Disclaimer: The contents, opinions and views expressed in the research article published in 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

Abbas, F., Afzaal, H., Farooque, A. A., and Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agron.*, 10(7): 1046.

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. https://doi.org/10.54386/jam.v24i1.1002

Asseng, S., Foster, I.A.N. and Turner, N.C. (2011). The impact of temperature variability on wheat yields. *Glob. Chang. Biol.,* 17: 997–1012.

Attri, S. D., and Rathore, L. S. (2003). Pre-harvest estimation of wheat yield for NW India using climate and weather forecast. *Mausam*, 54(3): 729-738.

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.

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*., 1-13.

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.

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

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

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

Montaseri, M.; Zaman Zad Ghavidel, S.; Sanikhani, H. (2018). Water quality variations in different climates of Iran: Toward modeling total dissolved solid using soft computing techniques. *Stoch. Environ. Res. Risk Assess*., 32: 2253–2273.

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.

Safa, M., Samarasinghe, S., and Nejat, M. (2015). Prediction of wheat production using artificial neural networks and investigating indirect factors affecting it: case study in Canterbury province, New Zealand.

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

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

Singh, K. N., Singh, K. K., Sudheer, K., Sanjeev, P., and Bishal, G. (2019). Forecasting crop yield through weather indices through LASSO. *Indian J.Agric.Sci.,* 89(3): 540-544.

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.

Uno, Y., Prasher, S.O., Lacroix, R.,Goel, P.K., Karimi, Y., Viau, A.,and Patel,R.M. (2005). Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Comput. Electron. Agric.*, 47:149–161.

Vashisth, A., Singh, R., and Choudary, M. (2014). Crop Yield Forecast at Different Growth Stage of Wheat Crop using Statistical Model under Semi Arid Region. *J. Agroecol. Nat. Resour. Manag.,* 1:1-3.