Journal of Agrometeorology
ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 26 (1) : 37 - 44 (March - 2024)
https://doi.org/10.54386/jam.v26i1.2411
https://journal.agrimetassociation.org/index.php/jam

Research Paper

Multistage sugarcane yield prediction using machine learning algorithms

SHANKARAPPA SRIDHARA*, SOUMYA B. R and GIRISH R. KASHYAP

Centre for Climate Resilient Agriculture, Keladi Shivapa Nayaka University of Agricultural and Horticultural Sciences, Iruvakki, Shivamogga – 577201, Karnataka, India
**Corresponding author email: sridharas1968@gmail.com sridhars@uahs.edu.in*

ABSTRACT

Sugarcane is one of the leading commercial crops grown in India. The prevailing weather during the various crop-growth stages significantly impacts sugarcane productivity and the quality of its juice. The objective of this study was to predict the yield of sugarcane during different growth periods using machine learning techniques *viz.,* random forest (RF), support vector machine (SVM), stepwise multiple linear regression (SMLR) and artificial neural networks (ANN). The performance of different yield forecasting models was assessed based on the coefficient of determination (R^2), root mean square error (RMSE), normalized root mean square error (nRMSE) and model efficiency (EF). Among the models, ANN model was able to predict the yield at different growth stages with higher R^2 and lower nRMSE during both calibration and validation. The performance of models across the forecasts was ranked based on the model efficiency as ANN > RF > SVM > SMLR. This study demonstrated that the ANN model can be used for reliable yield forecasting of sugarcane at different growth stages.

Keywords: Neural Networks, Random Forest, Multistage Yield Forecast, Sugarcane, Support vector machines (SVM), Stepwise multiple linear regression

Artificial intelligence is increasingly becoming a technical solution for the agriculture sector's long-standing problems and helping farmers improve their products and mitigate unfavorable environmental effects. A subset of artificial intelligence is known as "machine learning." It aims to create systems that can learn from past data, identify patterns and reach logical conclusions without human involvement. Machine learning systems use automated optimization techniques to increase the output accuracy continuously.

Crop yield prediction is one of the challenging problems in precision agriculture. Crop yield depends on many factors such as climate, weather, soil, use of fertilizer, and seed variety (Xu *et al.,* 2019). For decision-makers, estimating agricultural production is vital since it allows for effective resource planning (Gyamerah *et al.,* 2020). In terms of economy, yield prediction can assist in acting appropriately to determine how much to export in the event of surplus or early judgments regarding quantities, contracts, agreements, and planning of imports in the event of a shortage. Contrarily, it can assist farmers in determining what and when to cultivate and organize their harvest and storage (Chergui *et al*., 2020). Interestingly, inter-annual crop yield fluctuation is influenced by climate variability (Ray *et al*., 2015; Kukal and Irmak, 2018) Weather variables affect every crop differently during different developmental phases (Ji *et al.,* 2007; Gupta *et al*., 2022). Sridhara *et al*., (2023) reported maximum temperature and relative humidity played a significant role in pigeon pea yield prediction and found that stepwise linear regression (SLR) was outperformed by support vector machine (SVM), random forest (RF), least absolute shrinkage and selection operator (LASSO), and elastic net (ENET). Kakati *et al*., (2022) reported that the ANN technique may be effectively used to produce accurate yield forecast models for rape seed and mustard in the districts of Brahmaputra valley of Assam. It was found that the maximum temperature and morning and evening relative humidity were the most significant weather variables positively affecting rapeseed and mustard yield at all three growth stages.

One of the leading commercial crops in India's agricultural landscape is Sugarcane. Sugarcane contributes about 1.1% of the Gross Domestic Product (GDP) in the Indian economy. India currently holds the second position in sugarcane production in the world, with a production of 399.25 million tonnes for the 2020-21 harvest, with 50% of this crop concentrated in the country's

Article info - DOI: *https://doi.org/10.54386/jam.v26i1.2411*
Received:17 October 2023; Accepted: 19 January 2024; Published online : 1 March 2024

Southern region. The impact of temperature and rainfall variations on sugarcane can have immediate and delayed consequences because it is a crop grown for an extended period (12 to 15 months) and is sensitive to the weather at various developmental stages. Even a slight departure from the long-term average weather pattern during the year can cause a significant loss of sugarcane yield and subsequent sugar production (Glasziou *et al*., 1965; Mali *et al*., 2014; Zhao and Li, 2015). High temperatures can accelerate abiotic disorders and transform the sucrose content into fructose and glucose. Due to increased photorespiration from high temperature, sugar buildup is decreased (Binbol *et al*., 2006; Gawander, 2007). Lack of rain during the elongation period causes a change to the heat regime, which changes the dynamics of disease and insect attacks impacting the cane and sugar yield (Bhardwaj *et al*., 1981; Berding and Hurney, 2000). In addition, Mathieson (2007) stated that high temperature directly impacted the maximum relative humidity, which could result in insect pest attacks. Drought during the early and mid-growth stages of sugarcane, lowers cane production, which results in a poor sucrose yield (Zhao and Li, 2015). Due to low adaptive capacity, high vulnerability to natural hazards, limited forecasting systems, and inadequate mitigating techniques, sugarcane yields vary significantly across years and regions.

Further changing rainfall and temperature in most developing countries, with high input costs, and low cane prices are incredibly typical, leaving sugarcane growers with limited income. The weather fluctuation during the growing season of crops is a significant factor in yield variability. Depending on the stage of development, different meteorological factors have varying effects on the crop. Hence, yield prediction at various stages of crop growth using machine learning algorithms must be accurate, scientifically sound, made as early as feasible for mill operation planning for sugarcane factories. So, the present study focuses on machine learning algorithms *viz*., random forest (RF), support vector machine (SVM), stepwise multiple linear regression (SMLR) and artificial neural network (ANN) for prediction of sugarcane yield at different growth stages.

MATERIALS AND METHODS

Location

The data for ten Districts of Karnataka *viz.,* Shivamogga, Bagalkot, Belagavi, Ballari, Dharwad, Kalaburagi, Mandya, Mysore, Chamarajanagara and Davanagere were selected for the study. These areas contribute 87.19% of total sugarcane production with an area of 84.27% in the state of Karnataka. Using the average sugarcane production in Karnataka over the previous year as a basis, districts were chosen to predict yield forecast models. Belagavi contributes the most area (45.51%) and output (48.08%) of all the districts taken into account (Table 1).

Karnataka is situated between latitudes 74° East to 78°30’ East and longitudes 11°30’ North to 18°30’ North. According to the Koppen-Geiger climatic classification, Karnataka has hot semi-arid (BSh), tropical savanna (Aw), tropical monsoon (Am), and humid subtropical climates (Cwa), with the former two being the most prevalent. The Maidan’s driest regions receive 500 mm of yearly precipitation, while the coastal plain’s wettest regions receive roughly 4000 mm. The state gets most of its annual rainfall between June and September, with the remaining amount coming from a weak northeast monsoon that blows in the months after the monsoon season. Wintertime, on the other hand, is mainly dry.

Data collection

Time series data of sugarcane yield (*Saccharum officinarum*) for ten major sugarcane growing districts of Karnataka viz., Shivamogga (1997-2020), Bagalkot (1998-2020), Belagavi (1998-2020), Ballari (1985-2020), Dharwad (1990-2020), Kalaburagi (1990-2020), Mandya (1985-2020), Mysore (1997-2020), Chamarajanagara (1999-2020) and Davanagere (1999-2020), has been obtained from Directorate of Economics and Statistics, Government of Karnataka.

The daily weather data on five weather variables, including maximum temperature (T_{max}), minimum temperature (T_{min}), rainfall (RF), maximum relative humidity (RH I), and minimum relative humidity (RH II), were collected for 52 weeks of sugarcane growth period (*i.e.,* 13th Standard Meteorological Week (SMW) to 12th SMW of next year) from India Meteorological Department (IMD). Later, the weekly average values of the daily weather variables T_{max}, T_{min}, RH I, and RH II were calculated and in case of Rainfall, weekly sum of rainfall was taken into account. The first 80% of the sample data collected was used to calibrate the model, and 20% was utilized for model validation.

Methodology

For each weather variable, simple and weighted weather indices were developed. Simple indices were derived by adding individual weather factors or their interactions. In contrast, weighted indices were calculated by taking sum product of individual weather variables or their interactions with their correlation to detrended sugarcane production (Ghosh *et al*., 2014). Overall, thirty indices were developed and used in the current study. The lists of formulas for calculating the simple and weighted weather indices are presented (Table 2). The forecast models were developed at different stages of the crop growth like germination phase (F1: 13th SMW to 30th SMW), tillering and grand growth phase F2: 13th SMW to 43rd SMW) and at the ripening (F3: 13th SMW to 12th SMW) stages of sugarcane.

$$Z_{ij} = \sum_{w=1}^{m} X_{iw} \quad (1)$$

$$Z_{ii'j} = \sum_{w=1}^{m} X_{iw}X_{i'w} \quad (2)$$

Weighted weather indices:

$$Z_{ij} = \sum_{w=1}^{m} r_{iw}^{j} X_{iw} \quad (3)$$

$$Z_{ii'j} = \sum_{w=1}^{m} r_{ii'w}^{j} X_{iw} X_{i'w} \quad (4)$$

Where,

X_{iw}/X –value of ith/i$'^{th}$ weather variable under study in wth week;

Table 1: Area and production statistics of districts under study

Sr. No.	State	Area	Percent contribution to total area	Production	Percent contribution to total production
1	Bagalkote	82529	18.64	7592668	17.94
2	Belagavi	201448	45.51	20346248	48.08
3	Ballari	5940	1.34	475200	1.12
4	Chamarajanagara	6844	1.55	718620	1.70
5	Davanagere	981	0.22	110853	0.26
6	Dharwad	5744	1.30	442288	1.05
7	Kalaburagi	19117	4.32	1338190	3.16
8	Mandya	38999	8.81	4328889	10.23
9	Mysuru	10721	2.42	1458056	3.45
10	Shivamogga	713	0.16	85560	0.20
	Total	373036	84.27	36896572	87.19
	Karnataka	442665	100	42317030	100

Table 2: Weather indices used for developing the different models

Parameter	Simple weather indices					Weighted weather indices				
	Tmax	Tmin	RF	RH I	RH II	Tmax	Tmin	RF	RH I	RH II
Tmax	Z10					Z11				
Tmin	Z120	Z20				Z121	Z21			
RF	Z130	Z230	Z30			Z131	Z231	Z31		
RH I	Z140	Z240	Z340	Z40		Z141	Z241	Z341	Z41	
RH II	Z150	Z250	Z350	Z450	Z50	Z151	Z251	Z351	Z451	Z51

Simple weather indices:

$r^j_{iw}/r^j_{ii'w}$ – Correlation coefficient of detrended yield with ith weather variable/product of ith and i'th weather variables in wth week;

m – Week of forecast

Stepwise multiple linear regression

Stepwise multiple linear regression (SMLR) is a linear feature selection technique. A model is built by successively adding or removing variables based on the p-value of the F statistic at each step (Draper and Smith, 1998). In the present study, for inclusion or removal of a weather index into the model, the p values were set at 0.05 and 0.10, respectively.

Artificial Neural Network

The neural network employed in this study has three basic layers: an input layer, a hidden layer, and an output layer using a feed-forward algorithm. Each layer is connected to the next layer in the forward direction; there are no connections in the reverse direction (Dahikar and Rode, 2014). The number of nodes in the input and output layers is determined by the dataset used in the study. Choosing the ideal number of hidden neurons or nodes is the critical challenge in ANN implementation. The 'train' function of the 'caret' package and the 'nnet' method with 10-fold cross-validation in R software was used to calculate the number of hidden nodes (Kuhn, 2008).

Supervised machine learning algorithms

The current study uses supervised machine learning techniques *viz.*, support vector machines (SVM) and random forest (RF), primarily utilized to build models that can be either classification or regression. SVM balances model complexity and prediction errors by finding the optimal continuous-valued function. Saunders *et al*., (1998) reported that SVM uses a hyperplane as a decision boundary between the different classes. The RF algorithm implements Breiman's random forest approach for classification and regression. It produces decision trees using different data samples, forecasts the results from each subset, and then calculates the mean of all the decision trees. Using the R software's "caret" package with 10-fold cross-validation, the hyperparameters sigma, C for SVM, and mtry for RF were optimized (Kuhn, 2008).

Model performance analysis

The performance of different yield forecasting models used in the study was tested by computing co-efficient of determination (R^2), root mean square error (RMSE), normalized root mean square error (nRMSE) for calibration data and with RMSE and nRMSE for validation data and overall model efficiency (Jamieson *et al*., 1991).

RESULTS AND DISCUSSION

F1 forecast

Among the models tested for yield forecasting of sugarcane during F1, artificial neural networks (ANN) performed better, with model efficiency values ranging from 0.75 in the Ballari district to 1.00 in the Davanagere district. During calibration, better performance of the model was observed with a high coefficient of determination (0.78 in Ballari to 1.00 in Belagavi), lower RMSE (0.95 kg ha^{-1} in Belagavi to 8.76 kg ha^{-1} in Ballari) and nRMSE (0.94

in Mandya to 11.80 in Kalaburagi) values. Similar performance was also observed in the validation dataset. The RMSE and nRMSE values ranged between 0.63 to 8.90 kg/ha and 0.50 to 10.85% (Table 3). The random forest model was also better compared to SVM and SMLR models. The R^2 values ranged between 0.89 in Ballari to 0.96 in Belagavi. The RMSE values followed a similar trend. The performance of the model declined during the validation stage. Even though the calibration coefficient of determination (R^2) was high, model efficiency values were lower. Model efficiency values ranged from 0.73 (Shivamogga) to 0.88 (Belagavi) (Table 8). The lowest RMSE and nRMSE were observed during validation in the Dharwad district (2.60 kg ha^{-1} and 3.66%, respectively). The coefficient of determination (R^2) for SVM regression varied between 0.60 in Kalaburagi to 0.87 in Chamarajanagara and Dharwad districts. The nRMSE values during calibration were adjudged excellent, with values lying less than 10 percent in all the districts except Kalaburagi (19.22%). But the model performance sharply decreased during validation, with nRMSE values of the studied districts varying between 6.28 in Belagavi and 21.73 in the Ballari district (Table 5). The predictive performance of SMLR varied considerably among the study districts. The model efficiency values ranged from 0.49 (Kalaburagi) to 0.99 (Dharwad) (Table 8). During calibration, the RMSE and nRMSE values were the lowest in Bagalkot district (4.99 kg ha^{-1} and 4.95 %, respectively). During validation of model, lowest RMSE and nRMSE was noticed in Bagalkot district (4.22 t ha^{-1} and 5.13 %, respectively) while higher values were observed in Kalaburagi district (24.92 t ha^{-1} and 31.68%, respectively) (Table 6).

F2 Forecast

The F2 forecast of sugarcane yield revealed the same trend as the F1 forecast. Higher model efficiencies were noticed withs artificial neural network (ANN) model. The R^2 values during model calibration exceeded 0.96 for all the studied districts except Ballari (0.76). Similarly, nRMSE values were less than 10 percent for all the studied districts depicting an excellent fit of the model (Table 3). Overall model efficiency ranged from 0.56 in Mysore to 1.00 in Chamarajanagara and Shivamogga districts (Table 8). The predictive performance declined during the validation stage of the model. The improvement in the prediction performance of random forest model was noticed when compared to F1 forecast (Table 4). The R^2 values were in the range of 0.91 to 0.97 across the study districts. The nRMSE values observed during the model calibration suggested an excellent model fit with values less than 10 per cent for all the study districts. The validation performance was also better than the F1 forecast, with nRMSE values ranging between 4.40% (Mandya and Davanagere) and 24.94% (Ballari). The overall model efficiency was the highest in Dharwad (0.91) and lowest in Kalaburagi (0.75). In SVM regression, the highest R^2 during calibration was observed in Chamarajanagara (0.97). Four out of ten study districts showed an excellent model fit with an R^2 value > 0.90 during the calibration stage (Table 5). The overall model efficiency varied between 0.43 (Davanagere) and 0.87 (Chamarajanagara) (Table 8), with higher nRMSE values observed during model validation. The nRMSE values were greater than 10 per cent in seven districts out of ten during the validation stage. The performance of SMLR was better during the calibration stage. The lowest RMSE and nRMSE values were observed in the Bagalkot district (1.67 kg ha^{-1} and 1.66 %, respectively). The R^2 values were in the range of 0.65 (Ballari) to 0.98 (Bagalkot) (Table 6). The nRMSE values obtained during the validation stage of the model portrayed a higher error percentage in the yield prediction during validation. The overall model efficiency in yield prediction was in the range of 0.56 in Ballari district to 0.98 in Belagavi (Table 8).

F3 forecast

The prediction accuracy of the ANN model increased during F3 forecast in most of the districts. The ANN model efficiency of all the districts was in the range of 0.95 (Chamarajanagara) to 1.00 (Shivamogga and Dharwad) except for the Mysore district (0.74) (Table 8). Random forest regression performance increased in F3 forecast. The R^2 values were greater than 0.94 in all the studied districts indicating an excellent model fit. Similarly, the RMSE values ranged from 3.00 t ha^{-1} to 8.31 t ha^{-1} (Table 4). The model efficiency was in the range of 0.76 (Ballari) to 0.92 (Dharwad) (Table 8). Support vector regression model performance was also better in F3 compared to F2 and F1 forecasts. The lowest R^2 value was observed in the Mandya district (0.73) and highest in Mysore (0.95) during calibration. But the nRMSE of prediction during validation increased in all the districts compared to nRMSE of calibration (Table 5). The model efficiency varied from 0.52 in Mandya to 0.87 in Dharwad (Table 8). The SMLR model had a varied performance across the districts with a large gap in model efficiency values. The model efficiency varied from 0.37 in the Mandya district to 0.91 in the Bagalkot district (Table 8).

Inter-comparison of models

To assess the accuracy of forecasted models, percent error (PE) was considered. For state-level forecasts, 5 percent accuracy is the ideal level of precision (Ghosh *et al*., 2014; Kakati *et al*., 2022). So, by using the ANN model apart from a few districts, the PE between actual and forecast yield significantly decreased over both validation years. More than 50% of the districts displayed less difference between the forecasted yield and observed yields at all phases of the development. Over the forecasts and districts, ANN performed well compared to all other models. Weather variables played a significant role in predicting yield in the ANN model in all the stages of sugarcane. Maximum temperature positively affected sugarcane yield in all districts except Dharwad and Kalaburagi, where morning relative humidity affected the yield in the F1 stage of sugarcane. In the F2 stage, crop yield in all districts was governed by maximum temperature. During the F3 stage of development, interaction effect between maximum temperature and relative humidity was an important parameter in predicting the yield. Over the forecasts, the model identified maximum temperature and relative humidity as the key determinants of yield. Recently, ANN has drawn a lot of interest since it is proving to be a reliable technique for treating complex issues. Many others have also reported the superiority of ANN in yield forecasting. Basir *et al*., (2021) said the excellent accuracy of the ANN model in predicting the yield of transplanted paddy with an R^2 value of 0.994 and RMSE of 4.577. Similarly, Yildirim *et al*., (2022) forecasted cotton yield four months before harvest with an accuracy of R^2> 0.80 using ANN.

Reliable yield predictions were obtained by random

Table 3: Comparison of yield forecast performance of the ANN model During F1, F2 and F3 growth stages of sugarcane

Districts	F1 Forecast					F2 Forecast					F3 Forecast				
	Calibration			Validation		Calibration			Validation		Calibration			Validation	
	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE
Bagalkot	0.85	5.06	5.03	5.28	6.41	0.99	1.37	1.36	0.30	0.36	0.95	3.74	3.72	0.49	0.60
Ballari	0.78	8.76	10.07	8.90	10.85	0.76	8.87	10.11	9.87	12.04	0.97	3.08	3.51	1.65	2.02
Belagavi	1.00	0.95	1.11	2.86	3.00	0.98	1.81	2.12	2.47	2.58	0.99	1.52	1.78	2.06	2.16
Dharwad	0.99	1.51	2.15	0.92	1.29	1.00	0.96	1.37	10.30	14.50	1.00	0.93	1.33	0.21	0.30
Kalaburagi	0.82	7.88	11.80	4.76	6.06	0.97	3.01	4.50	2.86	3.64	0.99	1.67	2.50	0.70	0.89
Mandya	0.99	1.11	0.94	0.56	0.50	0.99	1.10	0.91	0.56	0.50	1.00	1.15	0.96	1.14	1.02
Mysore	0.98	2.13	2.01	0.63	0.55	1.00	0.39	0.37	29.19	25.66	0.85	7.58	7.16	11.98	10.53
Davanagere	0.96	2.88	2.68	0.68	0.63	1.00	1.55	1.40	0.13	0.12	0.99	1.95	1.75	0.14	0.13
Chamarajanagara	0.98	2.06	2.16	1.22	1.28	1.00	0.40	0.42	0.15	0.15	0.97	3.85	3.75	1.66	1.68
Shivamogga	0.96	2.88	2.68	0.68	0.63	1.00	0.48	0.45	0.03	0.03	1.00	0.07	0.07	0.01	0.01

Table 4: Comparison of yield forecast performance of the RF model During F1, F2 and F3 growth stages of sugarcane

Districts	F1 Forecast					F2 Forecast					F3 Forecast				
	Calibration			Validation		Calibration			Validation		Calibration			Validation	
	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE
Bagalkot	0.96	3.83	3.81	8.13	9.87	0.96	3.49	3.47	10.23	12.42	0.98	3.00	2.98	83.77	87.83
Ballari	0.89	6.48	7.38	19.97	24.36	0.91	5.90	6.73	20.45	24.94	0.95	4.64	5.29	19.14	23.34
Belagavi	0.93	4.54	5.30	8.31	8.70	0.91	5.11	5.97	4.32	4.52	0.94	4.04	4.72	6.23	6.52
Dharwad	0.93	4.92	7.00	2.60	3.66	0.96	3.72	5.30	4.14	5.84	0.96	3.77	5.36	1.77	2.49
Kalaburagi	0.94	5.07	7.59	21.17	26.92	0.94	4.74	7.09	13.11	16.67	0.96	4.59	6.87	14.30	18.18
Mandya	0.95	6.47	5.38	4.94	4.40	0.95	4.98	4.14	4.94	4.40	0.96	4.85	4.03	5.49	4.89
Mysore	0.93	6.53	6.17	12.84	11.29	0.94	4.77	4.51	16.65	14.64	0.96	4.15	3.92	14.65	12.88
Davanagere	0.94	4.98	4.63	72.38	67.57	0.96	8.53	7.70	4.84	4.40	0.95	8.31	7.50	1.97	1.79
Chamarajanagara	0.95	5.53	5.82	6.95	7.25	0.97	3.88	4.08	12.34	12.87	0.94	5.16	5.02	12.45	12.58
Shivamogga	0.94	4.98	4.63	14.22	13.27	0.95	5.32	4.95	10.28	9.60	0.95	4.15	3.86	5.80	5.42

Table 5: Comparison of yield forecast performance of the SVM model during F1, F2 and F3 growth stages of sugarcane

Districts	F1 Forecast					F2 Forecast					F3 Forecast				
	Calibration			Validation		Calibration			Validation		Calibration			Validation	
	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE
Bagalkot	0.80	6.98	6.93	13.50	16.40	0.92	5.11	5.07	11.92	14.48	0.95	3.74	3.72	87.50	91.66
Ballari	0.82	7.90	9.00	17.82	21.73	0.81	8.12	9.26	18.64	22.73	0.81	7.91	9.01	19.10	23.30
Belagavi	0.88	5.63	6.58	6.00	6.28	0.91	5.21	6.08	10.19	10.67	0.91	4.83	5.64	8.37	8.77
Dharwad	0.87	5.97	8.50	4.89	6.88	0.91	5.21	7.41	2.88	4.06	0.92	4.86	6.91	2.20	3.10
Kalaburagi	0.60	12.84	19.22	13.57	17.25	0.71	11.35	16.98	13.11	16.67	0.86	7.70	11.52	11.90	15.13
Mandya	0.83	8.10	6.74	8.66	7.71	0.72	10.19	8.47	8.66	7.71	0.73	10.33	8.59	5.32	4.74
Mysore	0.77	8.69	8.21	19.18	16.86	0.86	6.99	6.61	18.14	15.95	0.95	4.86	4.59	14.85	13.06
Davanagere	0.80	6.46	6.01	18.08	16.88	0.67	15.60	14.11	3.93	3.58	0.82	12.17	10.98	1.25	1.14
Chamarajanagara	0.87	6.31	6.63	10.08	10.52	0.97	4.03	4.24	9.64	10.06	0.89	6.56	6.39	9.73	9.83
Shivamogga	0.80	6.46	6.01	18.08	16.88	0.86	5.74	5.33	12.33	11.51	0.85	5.89	5.48	9.58	8.94

forest (RF) regression in all the three forecasts. Variable collinearity issues can be resolved by employing RF regression models, which are frequently sparked by using standard linear regression models. Prasad *et al*., (2021) forecasted the cotton yield before harvest across Maharashtra with a coefficient of determination (R^2) values of 69 per cent, 60 per cent and 39 per cent accuracy for September, December and February using Random Forest regression approach.

Support vector machine is a robust classification and regression tool. In SVM model, lowest PE was observed for Davanagere district at F1 (-2.03) and F3 (-0.74) stages and Shivamogga district at F2 stage (4.43) during 2018-2019. In 2019-2020, lowest PE was indicated for Ballari district at F1 (-0.13) stage, Davanagere district at F2 (0.05) and F3 (0.89) stages (Fig. 1, SVM). Lowest PE indicated for Mandya district at F1 (0.75), Shivamogga district at F2 (0.79) and Mysore

Table 6: Comparison of yield forecast performance of the SMLR model During F1, F2 and F3 growth stages of sugarcane

Districts	F1 Forecast					F2 Forecast					F3 Forecast				
	Calibration			Validation		Calibration			Validation		Calibration			Validation	
	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE	R^2	RMSE	nRMSE	RMSE	nRMSE
Bagalkot	0.84	4.99	4.95	4.22	5.13	1.67	0.98	1.66	8.38	10.18	3.21	0.93	3.18	7.34	8.91
Ballari	0.72	8.91	10.15	22.49	27.42	10.0	0.65	11.39	21.18	25.83	9.47	0.69	10.79	18.40	22.44
Belagavi	0.73	6.91	8.07	12.06	12.63	6.07	0.79	7.08	8.09	8.47	4.26	0.90	4.97	8.13	8.51
Dharwad	0.67	7.61	10.83	8.45	11.91	4.70	0.88	6.69	8.03	11.32	5.32	0.84	7.57	9.79	13.78
Kalaburagi	0.70	9.11	13.63	24.92	31.68	8.33	0.75	12.47	24.53	31.19	7.48	0.80	11.19	19.49	24.77
Mandya	0.73	7.72	6.42	11.82	10.52	7.65	0.74	6.36	11.82	10.52	6.36	0.82	5.29	12.95	11.53
Mysore	0.72	8.32	7.86	16.63	14.62	7.06	0.80	6.68	19.27	16.94	3.33	0.96	3.15	34.03	29.92
Davanagere	0.73	11.39	10.27	11.75	10.69	9.64	0.80	8.70	9.94	9.03	8.86	0.81	7.99	21.54	19.59
Chamarajanagara	0.70	8.46	8.89	11.92	12.43	6.49	0.83	6.82	10.47	10.92	10.71	0.65	10.43	12.00	12.12
Shivamogga	0.77	6.19	5.63	18.45	17.22	6.18	0.80	5.75	18.02	16.82	6.35	0.79	5.91	6.57	6.13

Table 7 : Yield equations generated for SMLR model at F1, F2 and F3 stages of sugarcane

Districts	Regression equation		
	F1 Forecast	F2 Forecast	F3 Forecast
Bagalkot	Y = 107.08 + 6.34Z21 -1.24Time + 0.001Z451	Y = 7.17 + 0.01Z251 - 1.5Time + 1.9Z21 + 0.03Z141 - 0.05Z241 + 0.27Z20	Y = 153.34 + 1.55Z21 -1.044Time + 0.001Z451
Ballari	Y = 83.57 + 2.34Z11 + 0.01Z150 + 0.05Z151	Y= 68.89 + 0.01Z131 + 0.76Time	Y = 54.13 + 0.01Z131 + 0.13Z41
Belagavi	Y = 48.38 + 0.009Z231 + 0.01Z141	Y = 34.83 + 0.01Z231 + 0.01Z241	Y = -9.78 + 0.003Z231 + 0.01Z241 + 0.0007Z150
Dharwad	Y = 7.65 + 0.004Z451 + 0.01Z131	Y = 117.80 + 4.06Z21 + 0.02Z131 - 0.48Z31	Y = 55.65 + 0.01Z131 + 0.003Z451
Kalaburagi	Y = 37.46 -1.27Time + 0.01Z131 + 0.53Z51	Y = 8.53 + 0.21Z31 - 1.21Time + 0.39Z51	Y = 30.31 + 0.19Z31 - 1.09Time + 0.002Z451
Mandya	Y = 168.60 + 0.01Z131 + 0.01Z141 + 0.73Time	Y = 109.70 + 0.05Z131 -0.07Z231 + 0.02Z251 + 0.38Time	Y = 219.48 + 0.38Z51 + 0.61Time + 1.13Z11
Mysore	Y = 107.20 + 7.79Z11 + 0.01Z131	Y = 215 + 3.27Z11	Y = 3411 - 1.78Time + 328Z21 + 0.01Z120 + 0.01Z251
Davanagere	Y = 226.50 + 0.01Z351 + 1.03Z41	Y = 407.88 - 0.12Z40 + 0.01Z131	Y = -81.59 + 5Z21 + 0.04Z121
Chamarajanagara	Y = 135.60 + 0.10Z121	Y = - 6.54 + 0.17Z51 + 0.01Z131	Y = - 16.64 + 0.01Z151
Shivamogga	Y = 275.59 + 3.48Z11 - 1.08Time	Y = 27.31 - 1.49Time + 0.06Z31 + 0.64Z41	Y = 263.70 + 3.89Z11

Table 8: Model efficiency of SMLR, ANN, SVM and RF models at F1, F2 and F3 stages of Sugarcane

Districts	ANN			RF			SVM			SMLR		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Bagalkot	0.86	0.99	0.99	0.88	0.86	0.86	0.68	0.78	0.82	0.87	0.93	0.91
Ballari	0.75	0.73	0.97	0.76	0.77	0.76	0.71	0.69	0.83	0.61	0.56	0.62
Belagavi	0.99	0.75	0.98	0.84	0.85	0.88	0.81	0.77	0.82	0.63	0.98	0.85
Dharwad	0.94	0.93	1.00	0.80	0.91	0.92	0.79	0.84	0.87	0.99	0.84	0.79
Kalaburagi	0.78	0.97	0.99	0.75	0.75	0.85	0.36	0.49	0.74	0.49	0.54	0.67
Mandya	0.98	0.99	0.99	0.81	0.88	0.88	0.51	0.51	0.52	0.69	0.69	0.75
Mysore	0.99	0.56	0.74	0.78	0.78	0.83	0.57	0.67	0.81	0.64	0.65	0.37
Davanagere	1.00	0.99	0.99	0.65	0.84	0.85	0.84	0.43	0.69	0.68	0.77	0.67
Chamarajanagara	0.98	1.00	0.95	0.85	0.84	0.85	0.78	0.87	0.83	0.62	0.76	0.59
Shivamogga	0.96	1.00	1.00	0.73	0.79	0.89	0.56	0.73	0.77	0.54	0.58	0.78

district at F3 (3.20) stages during 2018-2019, while Davanagere district showed lowest PE at F1 (0.85) stage, Dharwad district at F2 (-1.34) and Kalaburagi district at F3 (0.10) stage in 2019-2020 by utilizing RF model (Fig. 1, RF). Validation of the SMLR model indicated lowest PE for Davanagere district at F1 (0.17) and F2 (-0.66) and Mandya district at F3 (0.54) during 2018-2019, while Bagalkot district showed lowest PE at F1 (2.01), Davanagere district at F2 (0.20) stages and Belagavi district at F3 (1.81) stage during

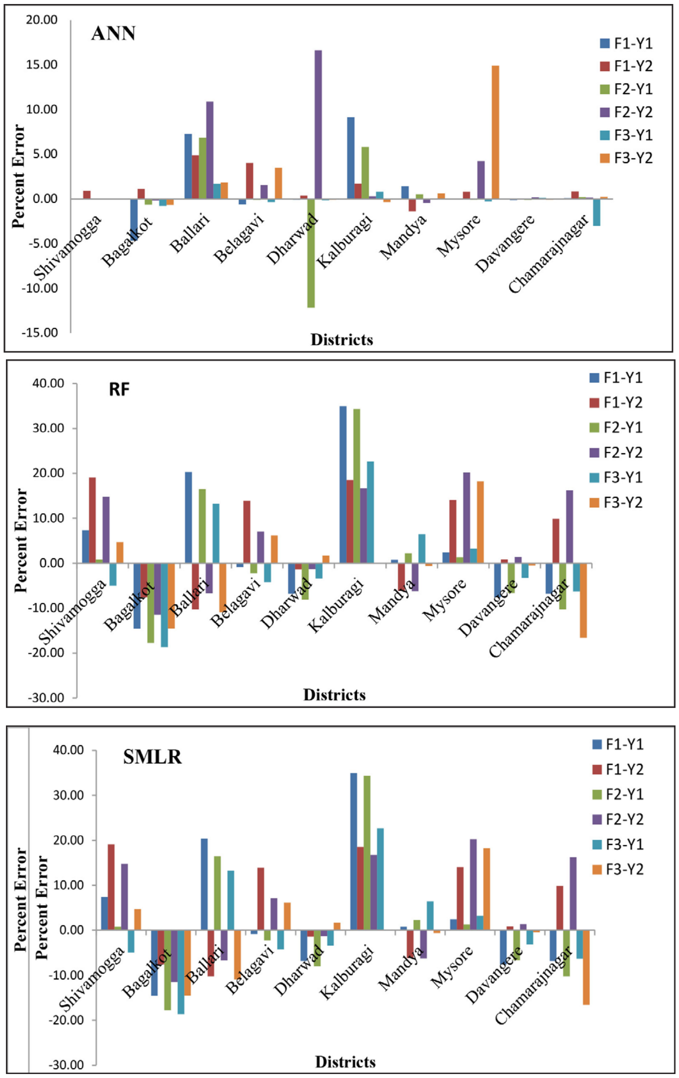

Fig. 1: Percent Error of yield forecast models of sugarcane at F1 (13-30th SMW), F2 (13th -43rd SMW) and F3 (13th -12th SMW) stages using A) ANN, B) RF, C) SVM and D) SMLR techniques (Y1=2019–2020 and Y2=2020–2021)

2019-2020 (Fig. 1, SMLR). Overall, non-linear models were more precise in prediction of yield compared to linear SMLR models (Table 7) indicating the non-linear relationship between weather and sugarcane yield. Sugarcane being grown throughout the year is affected significantly by weather elements. Using of weather indices approach helped to glean out the interactions between the weather elements in deciding the crop yield.

CONCLUSION

In this study district wise sugarcane yield prediction models were developed at three growth phases of sugarcane using four multivariate techniques and monthly weather variables as input for the state of Karnataka. Among the models ANN model forecasted the yield with greater accuracy during all the three forecasts. Inclusion of weighted weather variables improved the prediction performance of all the models. The major drawback of using machine learning algorithms in yield prediction is the black box nature of their learning process. Future direction implies on building more robust models with help of boosting, ensembling and deep learning approaches to improve the precision of forecasts.

ACKNOWLEDGEMENT

The authors acknowledge and thank Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, for their guidance and support.

Source of funding: India Meteorological Department, New Delhi through FASAL Project

Conflict of interest statement: The authors declare that there is no conflict of interest.

Data availability statement: The data is available from corresponding author upon reasonable request

Author contribution statement: **S. Sridhara**: Planning the study, interpretation of data, and manuscript editing; **Soumya B.R**: Collection of data and working out the indices and writing of draft manuscript; **Girish R Kashyap**: Formal analysis of data, interpretation and coding and drafting the manuscript

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

Basir, M.S., Chowdhury, M., Islam, M.N. and Ashik-E-Rabbani, M. (2021). Artificial neural network model in predicting yield of mechanically transplanted rice from transplanting parameters in Bangladesh. *J. Agric. Food Res*., 5:100186. https://doi.org/10.1016/j.jafr.2021.100186.

Berding, N. and Hurney, A. P. (2000). Suckering: A faced of ideotype selection and declining CCS in the wet tropics. In: Proceedings of the Australian Society of Sugar Cane Technology. 22:153-162.

Bhardwaj, S.C., Gupta, J. N., Jain, B. K. and Yadav, S. R. (1981). Comparative incidence of stalk borer, *Chilo auricilius* Ddgn. in autumn and spring planted and ratoon crops of sugarcane. *Indian J. Agric. Res.,* 15:135-140.

Binbol, N. L., Adebayo, A. A. and Kwon-Ndung, E. H. (2006). Influence of climatic factors on the growth and yield of sugar cane at Numan, Nigeria. *Clim. Res*., 32:247-252. https:// doi.org/10.3354/cr032247.

Chergui. N., Kechadi, T. and McDonnell, M. (2020). The impact of data analytics in digital agriculture: a review. In: the 2020 IEEE International multi-conference on: organization of knowledge and advanced technologies (OCTA)-International society for knowledge organization. https://doi.org/10.1109/OCTA49274.2020.9151851

Dahikar, S. S. and Rode, S.V. (2014). Agricultural crop yield prediction using artifcial neural network approach. *Int. J. Inov. Res. Electr. Electron. Instrum. Control Eng*., 2(1):683–686.

Draper, N.R. and Smith, H. (1998). Applied regression analysis, John Wiley & Sons, New York.

Gawander, J. (2007). Impact of climate change on sugarcane production in Fiji. WMO Bulletin, 56: 34-39.

Ghosh, K., Balasubramanian, R., Bandyopadhyay, S.N., Chattopadhyay and Singh, K.K. (2014). Development of crop yield forecast models under FASAL- a case study of kharif rice in West Bengal. *J. Agrometeorol.,* 16(1):1-8. https://doi.org/10.54386/jam.v16i1.1479.

Glasziou, K.T., Bull, T.A., Hatch, M.D. and Whiteman, P.C. (1965). Physiology of Sugar-Cane VII. Effects of temperature, photoperiod duration, and diurnal and seasonal temperature changes on growth and ripening. *Australian J. Bio. Sci.*, 18(1):53-66.

Gupta, S., Vashisth, A., Krishnan, P., Lama, A., Prasad, S. and Arvind, K. S. (2022). Multistage wheat yield prediction using hybrid machine learning techniques. *J. Agrometeorol.,* 24(4):373-379. https://doi.org/10.54386/jam.v24i4.1835.

Gyamerah, S.A., Ngare, P. and Ikpe, D. (2020). Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov Kernel function. *Agric. Forest Meteorol*., 280, p.107808. https://doi.org/10.1016/j.agrformet.2019.107808.

Jamieson, P.D., Porter, J.R. and Wilson, D.R. (1991). A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. *Field Crops Res*., 1991, 27(4): 337-350. https://doi.org/10.1016/0378-4290(91)90040-3.

Ji, B., Sun, Y., Yang, S. and Wan, J. (2007). Artificial neural networks for rice yield prediction in mountainous regions*. J. Agric. Sci*., 145(3):249-261.

Kakati. N., Deka, R. L., Das, P., Goswami, J., Khanikar, K. G. and Saikia, H. (2022). Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in the Brahmaputra valley of Assam, North East India. *Theor. Appl. Climatol*., 150:1201–1215. https://doi.org/10.1007/s00704-022-04220-3.

Kuhn, M. (2008). Building predictive models in R using caret package. *J. Stat. Softw*., 28:1–26.

Kukal, M.S. and Irmak, S. (2018). Climate driven crop yield variability and climate change impacts on the Great Plains agricultural production. *Sci. Rep.,* 8:3450. https://doi.org/10.1038/s41598-018-21848-2.

Mali, S.C., Shrivastava, P.K. and Thakare, H.S. (2014). Impact of weather changes on sugarcane production. *Res. Environ. Life Sci*., 2014, **7**(4), p.4.

Mathieson, L. (2007). Climate change and the Australian Sugar Industry: Impacts, adaptation and R & D opportunities. Sugar Research and Development Corporation. Australia.

Prasad, N. R., Patel, N. R. and Danodia, A. (2021). Crop yield prediction in cotton for regional level using random forest approach. *Spat. Inf. Res*., 29(2):195–206. https://doi.org/10.1007/s41324-020-00346-6.

Ray, D. K., Gerber, S.J., MacDonald, K. G. and West, C.P. (2015). Climate variation explains a third of global crop yield variability. *Nat. Commun*., 6:5989. https://doi.org/10.1038/ncomms6989.

Saunders, C., Stitson, M.O., Weston, J., Bottou, L. and Smola, A. (1998). Support vector machine-reference manual. Issue number CSD-TR-98-03, Department of Computer Science, Royal Holloway, University of London.

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. https://doi.org/10.1007/s00484-022-02396-x

Xu, X., Gao, P., Zhu, X., Guo, W. Ding, J. Li, C. and Wu. X. (٢٠١٩). Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China. *Ecol. Ind.,* 101: 943-953. https://doi.org/10.1016/j.ecolind.2019.01.059.

Yildirim, T., Moriasi, D. N., Starks, P. J., Chakraborty, D. (2022). Using Artificial Neural Network (ANN) for Short-Range Prediction of Cotton Yield in Data-Scarce Regions. *Agronomy* ١٢(4): 828; https://doi.org/10.3390/agronomy12040828.

Zhao, D. and Li, R. (2015). Climate Change and Sugarcane Production: Potential Impact and Mitigation Strategies, *Int. J. Agron.,* 1-10.