Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 27 (2) :233 - 235 (June - 2025)
https://doi.org/10.54386/jam.v27i2.2888
https://journal.agrimetassociation.org/index.php/jam

Short communication

Sugarcane yield forecasting using machine learning techniques in Udham Singh Nagar district of Uttarakhand

NEHA CHAND* and RAJEEV RANJAN

Department of Agrometeorology, GB Pant University of Agriculture & Technology, Uttarakhand, India
**Corresponding author: nikitachandrajwar999@gmail.com*

Sugarcane (*Saccharum officinarum*), a key crop of the Poaceae family, is widely cultivated in tropical and subtropical regions, contributing significantly to the agricultural economy. India, the second-largest producer after Brazil, accounts for nearly 25% of global sugarcane production (Mishra *et al.,* 2021). The crop serves multiple purposes, providing sugar, jaggery, ethanol, and livestock feed, while its by-products, such as bagasse and molasses, hold commercial value. In Uttarakhand, sugarcane production stood at 80 tonnes per hectare in 2021–2022, ranking the state 10th in national production (IISR, 2022). Sugarcane yield is influenced by climatic factors like rainfall, temperature, and humidity, making its prediction complex. Traditional yield estimation methods were labour-intensive, whereas modern statistical and machine learning (ML) models offer more efficient alternatives. Techniques such as stepwise multiple linear regression (SMLR), LASSO, ridge regression, elastic net (ELNET), and artificial neural networks (ANN) have been employed to analyze meteorological data for yield forecasting (Rashid *et al.,* 2021; Setiya *et al*., 2022; Shahhosseini *et al*., 2020).

This study evaluates different statistical and ML-based regression models to develop an accurate sugarcane yield forecasting approach for the Udham Singh Nagar district situated in the *Tarai* region of Uttarakhand. Sugarcane yield data spanning 23 years (1998-99 to 2020-21) were sourced from the Directorate of Economics and Statistics, Department of Agriculture and Farmers Welfare, Government of India, for the Udham Singh Nagar district (https://eands.dacnet.nic.in/). Concurrently, daily weather data for six parameters—maximum temperature (°C), minimum temperature (°C), rainfall (mm), relative humidity (%), wind speed (m/s) and solar radiation (MJ/m²/day) for the sugarcane growing period were retrieved from the NASA POWER web portal (https://power.larc.nasa.gov/data-access-viewer/). The adjusted weekly weather variables were subsequently utilized to compute both weighted and unweighted weather indices, adhering to the approaches outlined by Ghosh *et al.,* (2014) and Satpathi *et al*. (2023).

We examined five models viz. Stepwise multiple linear regression (SMLR), LASSO, ELNET, Ridge regression and Artificial Neural Networks (ANN). The stepwise multiple linear regression (SMLR) can yield better results with large datasets (Singh *et al*., 2014). Ridge regression applies a small bias to the predictor variables, which helps mitigate overfitting and enhance results relative to traditional models (Montgomery *et al.*, 2021). It facilitates the estimation of coefficients, even when predictor variables are highly correlated. The Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENET) are regression methods designed to tackle multicollinearity by penalizing the size of the coefficients. These techniques involve two key parameters: lambda and alpha, which need to be optimized. The optimal lambda values for both LASSO and ENET were identified by minimizing the average mean squared error using leave-one-out cross-validation (Piaskowski *et al.,* 2016). For our study, the alpha parameter was set to 1 for LASSO and 0.5 for ENET. The 'glmnet' package in R was used to implement LASSO and ENET. The artificial neural network (ANN), which comprises input, hidden, and output layers, with neurons connected across layers were usd in which the number of nodes in the input and output layers is determined by the dataset. We used the 'train' function from the 'caret' package in R with the 'nnet' method, applying 10-fold cross-validation (Kuhn, 2008).

In the model development phase, seventy-five percent of the data was allocated for training, while the remaining twenty-five percent was set aside for testing. To evaluate the model performance during training and testing coefficient of dtermination (R^2), root mean square error (RMSE), normalized root mean square error (nRMSE) and mean bias error (MBE) are used in this study.

Article info - DOI: *https://doi.org/10.54386/jam.v27i2.2888*
Presented in INAGMET 2024; Received: 7 July 2024; Accepted:9 April 2025; Published online : 1 June, 2025

Table 1: Evaluation criteria for various models for sugarcane yield

	Criteria/Model	SMLR	LASSO	Ridge	ENET	ANN
Calibration	R^2_{cal}	0.79	0.84	0.66	0.66	0.99
	$RMSE_{cal}$	1.45	1.23	1.89	1.89	0.07
	$nRMSE_{cal}$	18.9%	11.2%	17.1%	17.1%	0.6%
	MBE_{cal}	-0.55	0.00	-9.07	-1.10	0.05
Validation	R^2_{val}	0.03	0.09	0.09	0.19	0.99
	$RMSE_{val}$	11.9	11.3	11.3	12.2	0.28
	$nRMSE_{val}$	216%	74.8%	69.8%	80.2%	1.9%
	MBE_{val}	10.12	-9.63	-9.67	-1.10	-0.12

The evaluation of models for forecasting sugarcane yield in Udham Singh Nagar district, Uttarakhand, revealed varying predictive capabilities (Table 1). Stepwise multiple linear regression (SMLR) showed moderate predictive ability with an R^2 of 0.79 during calibration, but its validation R^2 dropped to 0.03, accompanied by higher RMSE and nRMSE values. LASSO regression improved performance with an R^2 of 0.84 in calibration and 0.09 in validation, demonstrating better accuracy than SMLR. Ridge regression and Elastic net regression (ENET) had similar calibration R^2 values of 0.66, but their validation performance was lower. The Artificial neural network (ANN) outperformed all models, achieving near-perfect R^2 values of 0.99 in both calibration and validation, with low RMSE and nRMSE values, indicating its superior predictive accuracy.

The Artificial Neural Network (ANN) exhibits exceptional performance with an R^2 of 0.999 in both calibration and validation, indicating strong learning and generalizability. Other models (ENET, Lasso, SMLR, Ridge Regression) show significant drops in R^2 during validation, suggesting overfitting. For example, ENET dropped from 0.66 to 0.19. During training, SMLR performs similarly to penalized regression models but falls short during testing, highlighting the superior predictive ability of penalized regression. The results of this study are consistent with earlier studies (Das *et al*., 2018; Kumar *et al*., 2021; Satpathi *et al*., 2023; Singh *et al*., 2019), which found that ANN performed better than SMLR. Additionally, LASSO, Ridge, and ENET demonstrated superior performance by employing feature selection techniques to reduce overfitting and streamline the model. Overall, artificial neural networks (ANNs) outperformed other models, achieving near-perfect R^2 values (0.99) in both calibration and validation, with low RMSE and nRMSE, confirming their accuracy and generalizability.

Machine learning techniques, especially ANNs, show significant potential for improving crop yield predictions over traditional models. While penalized regression methods provide practical solutions for some issues, machine learning excels in capturing complex interactions. Future research should explore hybrid models and incorporate more dynamic agricultural data to enhance model accuracy.

ACKNOWLEDGEMENT

Authors are grateful to the DACNET for supplying the agricultural data used in this study, as well as to my seniors for their continuous support and guidance throughout the research process.

Funding: This research received no external funding.

Conflict of interest: The authors declare no conflict of interest.

Data availability: The dataset used and analyzed in this study is available from DACNET and the Government of India for Udham Singh Nagar district.

Author's contribution: **N. Chand**: Conceptualization, Methodology, Formal analysis, writing original draft, writing review & editing; **R. Ranjan**: Conceptualization; Methodology, Writing review & editing.

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

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

REFERENCES

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.

Ghosh, K., Balasubramanian, R., Bandopadhyay, S., Chattopadhyay, N., Singh, K. K. and Rathore, L. S. (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

IISR (2022). Annual Report 2021–2022. ICAR—Indian Institute of Sugarcane Research, Lucknow.

Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Statist. Soft.*, 28: 1-26.

Kumar J, Devi M, Verma D, Malik DP, Sharma A (2021). Pre-harvest forecast of rice yield based on meteorological parameters using discriminant function analysis. *J. Agric. Food Res.,* 5:100194.

Mishra, P., Al Khatib, A. M. G., Sardar, I., Mohammed, J., Karakaya, K., Dash, A. and Dubey, A. (2021). Modeling and forecasting of sugarcane production in India. *Sugar Tech.,* 23(6):1317-1324.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.

Piaskowski, J. L., Brown, D. and Campbell, K. G. (2016). Near-infrared calibration of soluble stem carbohydrates for predicting drought tolerance in spring wheat. *Agron. J.,* 108(1): 285-293.

Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A. and Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE Access*, 9: 63406-63439.

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

Shahhosseini, M., Hu, G. and Archontoulis, S. V. (2020). Forecasting corn yield with machine learning ensembles. *Front. Plant Sci.*, 11: 1120.

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(2): 199-202.

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

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