Journal of Agrometeorology

(A publication of Association of Agrometeorologists)

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 27 (4) : 518-521 (December - 2025)
https://doi.org/10.54386/jam.v27i4.3148
https://journal.agrimetassociation.org/index.php/jam

Short communication

Rice yield prediction in Dakshina Kannada district using ensemble machine learning

A. SHAILESH RAO[1]* and ANJANA KRISHNAN[2]

[1]*Nitte (Deemed to be University), Nitte Meenakshi Institute of Technology (NMIT), Department of Mechanical Engineering, Bengaluru, India*
[2]*CSIR-National Aerospace Laboratories, Bengaluru, India*
**Corresponding author email id: shailesh.rao@nmit.ac.in*

Agriculture is the primary driver of Indian economy and country's GDP growth. Forecasting of agricultural production is an essential element for efficient resource management. Rice is primary crop grown in Dakshina Kannada District for three distinct seasons, each having its own set of climatic conditions. Furthermore, this region receives maximum rainfall and experiences fluctuations in monsoon intensity. This generates complex models in predicting rice crop yield. The machine learning (ML) techniques have proven effective in modelling using both historical and current data to predict crop yield (Aljuaydi and Wu 2022). ML algorithms like Random Forest (RF), Support Vector Machines (SVM), k-Nearest Neighbours (kNN) and Artificial Neural Networks (ANN) are used to predict crop growth (Shawon *et al.,* 2024). As example, ML models can analyze hidden relationships between soil health parameters and climatic variables (Zou and Okhrin 2024). Development of crop yield models for seasons-based agriculture is crucial (Paudel, 2020). In *rabi* season, the crop exhibits stability during its growth, while in *Kharif* seasons, the yields become erratic due to monsoon fluctuations (Mohapatra *et al.,* 2023).

The soil composition and NDVI are used to evaluate to forecast rice yield using RF regression (Sutha *et al.,* 2023). Also, SVM, kNN and RF are combined to create a stacking model for crop classification and yield prediction (Saravanan and Bhagavathiappan 2022). A study on use of Linear Regression, Random Forest and SVR models to increase paddy crop predictions (Sakthipriya and Chandrakumar 2024). A model for crop yield is established that considers salinity, waterlogging, and soil fertility during the heavy monsoon (Wani *et al.,* 2017). The extraction of datasets and machine learning models from remote sensing platforms such as Google Earth Engine are carried out to develop ML models. From MODIS dataset, RF models are used for predicting wheat crop yield using NDVI and precipitation (Pang *et al.,* 2022). GEE is also used to establish the paddy crop yield model based on monthly NDVI and precipitation values (Moriondo *et al.,* 2006). ML models combined with remote sensing data shows promising avenue for precise crop yield prediction.

The current research provides a solid basis in Dakshina Kannada district having high rainfall and fluctuated monsoons to predict rice crop yields. In this region, Belthangady and Bantwal Taluks contributes for highest crop yield for district. The rice yield for the district was collected from Government website (https://desagri.gov.in/) for years 2000 to 2020. Google Earth Engine (GEE) was used to derive maps and variables, wherein green colour region represents Dakshina Kannada district (Fig. 1). The blue (Bantwal) and red (Belthangady) colours represent taluk region. The precipitation and relative humidity were extracted monthly wise from the ERA5-Land dataset from GEE. The temperature index (GDD) was derived from MODIS (MOD13Q1) datasets aggregated

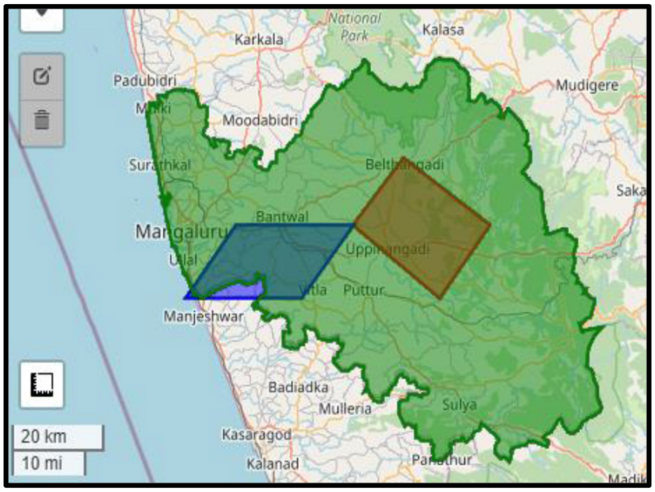

Fig. 1: Location map of Dakshina Kannda district and Belthangady-Bantwal taluk

Article info - DOI: *https://doi.org/10.54386/jam.v27i4.3148*
Received:6 August 2025; Accepted:18 September 2025; Published online : 1 December 2025

Table 1: Mean monthly precipitation, humidity and GDD for Dakshina Kannada district and Belthangady-Bantwal Taluk

Month	Precipitation (mm)		Humidity (%)		GDD (base 10 °C)	
	District	Taluka	District	Taluka	District	Taluka
January	8	15	58	60	430	440
February	15	25	60	62	500	510
March	40	65	65	67	570	580
April	90	150	70	73	610	620
May	180	280	75	77	640	650
June	1050	1250	85	87	450	460
July	1150	1400	90	91	420	430
August	950	1150	88	89	430	440
September	650	850	85	87	470	480
October	300	450	80	82	520	530
November	90	130	70	72	500	510
December	30	55	65	67	440	450

over district and taluk boundaries. The temporal range selected from the work is monthly data from June to October (*Kharif*), February to May (Summer) and November to March (*Rabi*).

The monthly mean values of precipitation, humidity and GDD are shown in Table 1. The fifteen variables (comprising five columns of each precipitation, humidity and GDD values for each season) were processed using feature importance ranking. The three variables, precipitation, humidity and GDD were considered initially to predict the rice yield. To improve robustness and interpretability three predictors such as precipitation (mm) that calculates monthly cumulative rainfall, humidity (%) derived from monthly relative humidity and GDD for three seasons were selected. The three machine learning algorithms were developed for comparing the crop yield. The SVM using both linear and polynomial kernels, Random Forest (RF) having 500 number of trees with number of variables measured at each split were selected as models for crop yield prediction. The model performance was evaluated using metrics such as R^2, RMSE, MAE and MSE; and their results were presented.

Selection of important predictors

Table 2 represents the Feature Importance values (%) for Dakshina Kannada district and Belthangady-Bantwal taluk data for different seasons. Here, the feature importance analysis using Random Forest regressor scores were performed to determine the significance value of each variable. It was computed by finding mean decrease in impurity that expressed as relative percentages. This provides a quantitative ranking of climatic variables for different month/seasons. Based on the ranked importance scores, top five predictors for each season were selected for model development.

During *Kharif* season, for Dakshina Kannada district, the significant predictors were found to be humidity in October and August, precipitation in June, GDD value in September and June. This understands that early rainfall (June precipitation) acts as foundation for crop establishment for *kharif* season. The later atmospheric conditions (October humidity) influence reproductive growth and grain filling. During the crop season, the temperature remains vital during sowing and active vegetation stages. Here GDD in June and September indicates its significance that relates with yield potential. Finally, humidity in the midseason (August) helps in moderating plant water demand during crop growth stage.

With the shift in data for taluk region, the top predictors shifted towards GDD in September and October, humidity in August and September, and precipitation in June, this imitates that the position of late monsoon heat and moisture during September strongly affects the vegetative vigour and tillering. The temperature at late season (October) shows its importance, that relates to heat stress affecting the reproductive success. Overall, the predictors defined in taluk-level models can capture better for localized climatic variability, particularly during mid-to-late *Kharif* season.

During *Rabi* season, in case of Dakshina Kannada district, the important predictors were GDD in November, precipitation in January and February, and humidity in November and December. The identified predictors indicated that the early season temperature accumulation (GDD November) and atmospheric humidity (November and December) have a greater influence in crop establishment and vegetative improvement. Also, the contribution of Precipitation in January and February underscores its role during mid-to-late season that helps in sustaining soil moisture during reproductive development. At taluk level, the predictor ranking underscores precipitation in December and January, humidity in December, and GDD in November and January. The early and mid-season precipitation becomes particularly critical for localized soil moisture management. The humidity during December supports a favourable microclimate for tillering. The temperature in November and January highlights that the thermal environment becomes important during establishment and mid-growth phases.

During summer season for Dakshina Kannada district, precipitation in March, April and May, and humidity in March and April, were found as important predictors. The precipitation becomes strong during late summer rainfall that helps in maintaining soil moisture during dry season. The early season humidity in March and April also helps in vegetative development by reducing evapotranspiration stress and aiding yield growth. For Taluk level, important predictors included humidity in February, April and May, precipitation in March and GDD in April. This specifies that the

Table 2: Feature Importance values (%) for Dakshina Kannada District and Belthangady-Bantwal Taluk

Season	Month	Precipitation		Humidity		GDD	
		District	Taluka	District	Taluka	District	Taluka
Summer	February	5	5	5	15	5	10
	March	15	12	20	5	5	5
	April	18	8	15	18	5	12
	May	12	8	5	20	5	5
Kharif	June	25	22	10	10	15	12
	July	5	5	5	5	5	5
	August	5	5	13	12	5	5
	September	5	5	5	15	12	20
	October	5	5	18	10	5	15
Rabi	November	8	10	15	12	20	18
	December	6	12	18	20	5	5
	January	22	20	5	5	5	15
	February	12	8	5	5	5	5
	March	5	5	5	5	5	5

Table 3: Model evaluation district and taluk across three seasons

Season	Model	R^2		RMSE		MAE		MSE	
		District	Taluk	District	Taluk	District	Taluk	District	Taluk
Kharif	SVM (Linear)	0.340	0.426	0.339	0.286	0.285	0.214	0.115	0.082
	SVM (Polynomial)	0.523	0.654	0.241	0.171	0.165	0.120	0.045	0.031
	Random Forest	0.871	0.930	0.104	0.085	0.086	0.068	0.011	0.007
Rabi	SVM (Linear)	0.375	0.654	0.262	0.173	0.172	0.120	0.068	0.031
	SVM (Polynomial)	0.362	0.794	0.241	0.150	0.186	0.119	0.058	0.022
	Random Forest	0.898	0.910	0.094	0.092	0.093	0.072	0.008	0.008
Summer	SVM (Linear)	0.342	0.542	0.375	0.238	0.213	0.238	0.143	0.060
	SVM (Polynomial)	0.253	0.586	0.453	0.225	0.344	0.225	0.212	0.053
	Random Forest	0.818	0.841	0.197	0.189	0.165	0.153	0.039	0.036

localized moisture availability remains dominant factor for summer. The humidity in February, April and May directly helps in crop survival during water limited conditions. The GDD values in April focus on its importance during rice growth due to heat accumulation during critical growth during summer season.

With all the three different rice crop seasons, humidity, precipitation and GDD remains dominant predictors for both district and taluk scales. The district level models emphasized early season precipitation and late season humidity. The taluk level models highlighted localized mid-season conditions and temperature. This dissimilarity understands that the taluk level predictions have captured localized crop climate prediction that achieved a higher predictive accuracy.

Model evaluation

The different models are developed for rice yield prediction for district and taluk region across three seasons (*Kharif*, *Rabi*, and Summer) and the comparative results are presented in Table 3. For *Kharif* season, the linear SVM shown a modest performance (R^2 = 0.340, RMSE = 0.339), while polynomial with degree two kernel model has improved the predictive accuracy (R^2 = 0.523, RMSE = 0.241) for district region. Here the Random Forest model significantly outperformed both SVM model with R^2 as 0.871 and a very low error metrics. This demonstrates the RF model was able to capture nonlinear interactions between climate predictors and yield. In case of taluk region, some improvement in predictive model is observed for all three models. The linear SVM increased to R^2 as 0.426, while polynomial kernel has shown R^2 as 0.654 with RMSE as 0.171. the RF again delivered better results with higher R^2 value (0.930) and lower RMSE (0.085), that indicates the benefits of localized modelling, where variability in climate and soil conditions is better captured.

For *Rabi* season, and for district level, again SVM models performed moderately for both linear and polynomial kernels. The Random Forest developed improved models with higher R^2 (0.898) and lower RMSE (0.094). The model confirms its robustness across different seasonal regimes. In Taluk region, the SVM models showed substantial development compared to district-level performance. Here, the SVM with linear kernel achieved R^2 of 0.654, whereas polynomial kernel increased R^2 value to 0.794. The Random Forest maintained higher predictive capability and suggests that taluk-level improves model which reflects impact of climatic variability during *Rabi*.

The district-level models exhibited lower performance compared to other seasons for summer seasons. The linear developed poor models which indicate the difficulty in capturing yield variability during moisture limited summer conditions. The Random Forest outperformed SVM models in developing rice crop yield models. At taluk level, the Random Forest finally performed better model than SVM. The developed model also emphasized that localized modelling during summer season, where small-scale climate variations significantly influence water accessibility and crop growth. For all seasons, the Random Forest consistently outperformed SVM models for all three seasons. The taluk level region has outperformed in model development by reducing spatial heterogeneity in climate and crop management conditions. The performance gap became important for *Kharif* and *Rabi* seasons, where rainfall distribution and temperature accumulation influence the rice yields.

To conclude, humidity, precipitation and temperature-derived GDD appeared as dominant predictors for rice crop yield for both district and taluk scales. In district level models, the early season precipitation and late season humidity are proved important during analysis in various ML models. In contrast, taluk level models emphasized on localized mid-season conditions and GDD that captures finer scale climatic variability. Finally, the taluk level models performed higher predictive accuracy than district level models. The results underscore the importance of combining seasonal climate variables with localized predictors for rice yield forecasting in Dakshina Kannada district.

ACKNOWLEDGEMENT

The authors thankfully acknowledge the use of Google Earth Engine platform for providing dataset for processing and analysis. We thank the Administrator, Principal and Department of Mechanical Engineering, Nitte Meenakshi Institute of Technology for providing necessary research facilities.

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

Funding: This research did not receive any specific funding.

Data Availability: The datasets analyzed during the current study are publicly available and can be accessed from: DES portal: https://desagri.gov.in; ERA5-Land (Copernicus Climate Data Store): https://cds.climate.copernicus.eu and MODIS NDVI (NASA Earthdata): https://earthdata.nasa.gov

Authors' contribution: **Shailesh Rao A**: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft; **A. Krishnan**: Supervision, Conceptualization, 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

Aljuaydi, F. and Wu, Y.H. (2022). Multivariate machine learning-based prediction models of freeway traffic flow under non-recurrent events. *Alexandria Engg. J.,* 65: 151-162. https://doi.org/10.1016/j.aej.2022.10.015

Mohapatra, S., Sahoo, D., Sahoo, A.K., Sharp, B. and Wen, L. (2023). Heterogeneous climate effect on crop yield and associated risks to water security in India. *Intern. J. Water Resource Develop.,* 40(3): 345-378. https://doi.org/10.1080/07900627.2023.2244086

Moriondo, M., Maselli, F. and Bindi, M. (2006). A simple model of regional wheat yield based on NDVI data. *Europ. J. Agron.,* 26(3): 266-274. https://doi.org/10.1016/j.eja.2006.10.007

Pang, A., Chang, M.W.L. and Chen, Y. (2022). Evaluation of Random Forests (RF) for regional and local-scale wheat yield prediction in Southeast Australia. *Sensors*, 22(3): 717. https://doi.org/10.3390/s22030717

Paudel, D., Boogaard, H., De Wit, A., Janssen, S., Osinga, S., Pylianidis, C. and Athanasiadis, I.N. (2020). Machine learning for large-scale crop yield forecasting. *Agric. Syst.,* 187: 103016. https://doi.org/10.1016/j.agsy.2020.103016

Sakthipriya, D. and Chandrakumar, T. (2024). Weather based paddy yield prediction using machine learning regression algorithms. *J. Agrometeorol.,* 26(3): 344-348. https://doi.org/10.54386/jam.v26i3.2598

Saravanan, K.S. and Bhagavathiappan, V. (2022). A comprehensive approach on predicting the crop yield using hybrid machine learning algorithms. *J. Agrometeorol.,* 24(2): 179-185. https://doi.org/10.54386/jam.v24i2.1561

Shawon, S.M., Niha, F.L. and Zubair, H. (2024). Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agric. Technol.,* 100718. https://doi.org/10.1016/j.atech.2024.100718

Sutha, K., Indumathi, N. and Shankari, S.U. (2023). Recommending and predicting crop yield using Smart Machine Learning Algorithm (SMLA). *Current Agric. Res. J.,* 11(2): 686-694. https://doi.org/10.12944/carj.11.2.30

Wani, S.P., Anantha, K. and Garg, K.K. (2017). Soil properties, crop yield, and economics under integrated crop management practices in Karnataka, southern India. *World Develop.,* 93: 43-61. https://doi.org/10.1016/j.worlddev.2016.12.012

Zou, J. and Okhrin, O. (2024). Data-driven determination of plant growth stages for improved weather index insurance design. *Agric. Finance Rev.,* 84(4/5): 297-319. https://doi.org/10.1108/afr-01-2024-0015