Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 27 (1) : 86-91   (March - 2025)
https://doi.org/10.54386/jam.v27i1.2731
https://journal.agrimetassociation.org/index.php/jam

Research Paper

Prediction of pan evaporation in Chhattisgarh using machine learning techniques

RUPESH NAIK[1], BABITA MAJHI[1*], and DIWAKAR NAIDU[2]

[1]Department of CSIT, Guru Ghasidas Vishwavidyalaya, Central University, Bilaspur 495009, Chhattisgarh, India
[2]Faculty of Agricultural Engineering, IGKV, Raipur 492012, Chhattisgarh, India
**Corresponding author email: babita.majhi@gmail.com*

ABSTRACT

Accurate measurement or estimation of evaporation loss is crucial for developing and successfully implementing water resource management strategies, irrigation planning, reservoir management etc. To predict the pan evaporation (EP) accurately for Raipur, Jagdalpur, and Ambikapur stations of Chhattisgarh, four deep learning models and three machine learning models were used and a hybrid model using Deep Neural Network (DNN) and Random Forest (RF) was proposed. Simulation results demonstrated that the hybrid model (DNN+RF) outperforms the rest with R^2 of 0.964, 0.920, 0.894 for Raipur, Jagdalpur and Ambikapur respectively. It has been observed that the hybrid DNN+RF model demonstrated faster convergence compared to other models with high accuracy, making it efficient and well-suited for real-time applications such as irrigation scheduling and water resource management.

Keywords: Pan evaporation, Machine learning, Deep learning, Deep neural network, Random Forest regressor.

Evapotranspiration is the sum of all processes by which water moves from the land surface to the atmosphere via evaporation and transpiration. It is an important agrometeorological parameter with various applications in agriculture and water resources management. There are various approaches to estimate evapotranspiration using climatic parameters (Allen *et al.,* 1998; Mehta and Pandey, 2015; 2018), as it is influenced by different climatic factors such as temperature, humidity, sunshine hours, and wind speed. It is an important factor in agrometeorology as it provides information about water loss due to evaporation from soil, which helps farmers and agricultural scientists in irrigation scheduling and water resource management. Additionally, it is used for drought monitoring, pest and disease management, crop yield prediction, and climate analysis. Pan evaporation (EP) provides crucial information about water demand and environmental conditions, which are valuable for decision-making in agriculture and, therefore, an important topic of research. Lakhawat *et al.,* (2024) investigated the effects of pan evaporation-based drip irrigation levels on guava performance in the Udaipur and Rewa regions. Sharma *et al.,* (2023) used sensor-based irrigation scheduling combined with pan evaporation and crop evapotranspiration (ETc) to optimize irrigation and fertigation practices in okra cultivation.

There is increasing interest in applying advanced machine learning and deep learning techniques (Moayedi *et al.,* 2022, Srivastava *et al.,* 2022; Abed *et al.,* 2022) to improve the accuracy of evaporation predictions and support more precise hydrological studies. Kumar *et al.,* (2024) demonstrated the use of Long Short-Term Memory (LSTM) and Artificial Neural Network (ANN) models for pan evaporation estimation. Dong *et al.,* (2021) developed a novel hybrid approach combining CatBoost (a gradient-boosting machine learning algorithm) with the Bat Algorithm (a nature-inspired optimization algorithm) to estimate pan evaporation in Northwest China. Majhi and Naidu (2021) used a Functional Link Artificial Neural Network (FLANN) to estimate daily EP, outperforming empirical and multilayer ANN models.

In this paper, an attempt has been made to develop a suitable model for predicting daily pan evaporation at three stations in Chhattisgarh using a deep hybrid model.

MATERIALS AND METHODS

Study area and data

Three stations, viz., Ambikapur, Jagdalpur, and Raipur of Chhattisgarh state, were selected for this study. Climatic data

Article info - DOI: https://doi.org/10.54386/jam.v27i1.2731
Received: 9 September 2024; Accepted: 26 November 2024; Published online : 1 March 2025

"This work is licensed under Creative Common Attribution-Non Commercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) © Author (s)"

Table 1: Details of study location and datasets

Name of station	Latitude (°N)	Longitude (°E)	Training period	Testing period
Ambikapur	23.14	83.19	1996-2013	2014-2017
Jagdalpur	19.09	82.02	1993-2012	2013-2017
Raipur	21.25	81.63	1981-2010	2011-2017

on minimum temperature (Tmin), maximum temperature (Tmax), bright sunshine hours (BSS), wind speed (WS), morning relative humidity (RH1), afternoon relative humidity (RH2) and pan evaporation (EP) were collected from IMD-certified observatories located at these stations. These observatories are situated within three distinct agro-climatic zones (ACZs) of Chhattisgarh. The time periods covered by the datasets are provided in Table 1.

To address the missing data, the K-Nearest Neighbors (KNN) imputer (Zhang, 2012) was applied to fill the missing values for Ambikapur. Subsequently, the data were normalized to a range between 0 and 1 using min-max normalization. This normalization step ensured that all features were on a similar scale, improving model convergence, reducing bias, and enhancing overall model performance. The normalized data were then split into 80% for training and 20% for testing.

Methodology

In this study, various advanced machine learning (ML) techniques—Support Vector Regression (SVR) with three different kernels (radial basis function [RBF], polynomial, and linear) (Smola and Schölkopf, 2004), AdaBoost, XGBoost (Ji *et al.,* 2019) and deep learning techniques, including Deep Neural Networks (DNN) (LeCun *et al.,* 2015), Deep Recurrent Neural Networks (DRNN), Deep Long Short-Term Memory Networks (DLSTM), and Deep Bidirectional LSTM (DBiLSTM) (Hochreiter and Schmidhuber, 1997; Ibrahim and Elhafiz, 2023; Niknam *et al.,* 2023), were employed for daily pan evaporation rate prediction.

The inputs to the model consisted of different combinations of Tmax, Tmin, BSS, WS, RH1, and RH2, while the corresponding EP value served as the target variable. The model received the first input pattern and calculated the predicted output through its internal processes. The error was computed by comparing the predicted output with the corresponding target value. The model's parameters were then updated using the specific approach's update rules and the error value. This process continued until all 80% of the training inputs were used, marking the completion of the first experiment. The experiment was repeated until the RMSE reached its minimum value. Once the error was minimized, the model's parameters were frozen and used to evaluate the model with the testing set. During testing, the testing patterns were fed into the model, and various performance metrics were calculated to assess the model's effectiveness. The performance of these models, along with the proposed hybrid model, was evaluated and compared to identify the most effective approach for predictive analysis.

Proposed hybrid model using DNN and Random Forest

A hybrid model is proposed by combining a Deep Neural Network (DNN) and a Random Forest (RF) regressor (Breiman, 2001) in series. The DNN consists of four hidden layers with 64, 32, 16, and 8 neurons, respectively, each utilizing ReLU activation functions. The output of the fourth hidden layer is used as input to the RF regressor in a sequential manner. The Adam optimizer and Mean Squared Error (MSE) loss function are employed to train the combined model (DNN+RF). Initially, the DNN portion of the model is trained, and the output from the fourth hidden layer is passed to the RF regressor. Finally, the hybrid DNN+RF model is evaluated using the test data. This process is repeated for the specified number of epochs to ensure convergence and accuracy. To estimate the EP rate, simulations were conducted using Python 3.12.0, leveraging libraries such as pandas, sklearn, tensorflow, matplotlib, keras, and seaborn.

RESULTS AND DISCUSSION

The performance of various machine learning models, SVR (with RBF, Polynomial, and Linear kernels), AdaBoost, and XGBoost, across three stations (Raipur, Jagdalpur, and Ambikapur) using RMSE and R^2 as metrics for different input combinations are presented in (Table 2). XGBoost consistently delivered the best results across all stations, particularly with the input combination of all six inputs, achieving the lowest RMSE (0.67 for Raipur, 0.68 for Jagdalpur, and 1.03 for Ambikapur) and highest R^2 values (0.955, 0.879, and 0.912, respectively). While SVR with the RBF kernel showed competitive performance, XGBoost outperformed it in accuracy. AdaBoost generally lagged behind other models, especially with complex input combinations. Overall, XGBoost demonstrated robust and consistent performance across all stations and input features, making it the most effective model for these datasets (Table 2).

The performance of four deep learning models, DNN, DRNN, DLSTM, and DBi-LSTM are presented in Table 3. Among these, the DRNN model demonstrated superior performance for Raipur with five input variables, achieving an RMSE of 1.05 and R^2 value of 0.891. Similarly, for Jagdalpur, DRNN produced the best results using six inputs, with an RMSE of 0.97 and R^2 value of 0.771. In contrast, the DLSTM model outperformed others for Ambikapur, delivering the lowest RMSE of 1.00 and R2 value of 0.751 with five input variables (Table 3).

Further, the hybrid DNN+RF model was trained, and the results of the model are displayed in Table 4 for different combinations of input variables. Table 4 reveals that for Raipur and Jagdalpur, the two-inputs and three-hidden-layers model performed better than the two-inputs and four-hidden-layers model, with RMSE values of 0.9297 and 0.869, and R^2 values of 0.9116 and 0.8154, respectively. The model with two inputs and three or four hidden layers yielded similar results for the Ambikapur station, with an RMSE of 0.93 and an R^2 of 0.770 (Table 4).

The hybrid DNN+RF model with four inputs and three or four hidden layers performs equally well for Raipur and Jagdalpur, with RMSE values of 0.73 and 0.67, and R^2 values of 0.946 and 0.893, respectively. For Ambikapur, the model with four inputs and four hidden layers performs better than the model with three hidden layers, with an RMSE value of 0.66 and an R^2 value of 0.886. For models with five inputs and three hidden layers, performance was

Table 2: Results of ML based models for all stations

Input Combinations	Model	Kernel	Raipur		Jagdalpur		Ambikapur	
			RMSE	R^2	RMSE	R^2	RMSE	R^2
Tmax,Tmin	SVR	RBF	1.45	0.784	1.17	0.641	1.22	0.635
		Polynomial	1.47	0.778	1.16	0.650	1.24	0.619
		Linear	1.54	0.756	1.21	0.619	1.34	0.556
	AdaBoost	-	1.53	0.762	1.20	0.621	1.49	0.452
	XGBoost	-	1.15	0.866	0.96	0.757	1.07	0.716
Tmax, Tmin, RH1 and RH2	SVR	RBF	1.34	0.815	1.04	0.719	1.12	0.693
		Polynomial	1.43	0.790	1.06	0.706	1.16	0.668
		Linear	1.46	0.783	1.05	0.713	1.28	0.598
	AdaBoost	-	1.65	0.720	1.09	0.687	1.61	0.361
	XGBoost	-	0.91	0.915	0.79	0.838	0.86	0.816
Tmax, Tmin, RH1, RH2 and WS	SVR	RBF	1.09	0.877	0.98	0.750	1.04	0.733
		Polynomial	1.13	0.869	1.00	0.740	1.10	0.701
		Linear	1.17	0.860	1.01	0.732	1.20	0.647
	AdaBoost	-	1.38	0.805	1.07	0.700	1.65	0.330
	XGBoost	-	0.75	0.942	0.72	0.866	0.79	0.848
Tmax, Tmin, RH1, RH2, WS and BSS	SVR	RBF	1.06	0.885	0.93	0.772	0.99	0.859
		Polynomial	1.10	0.876	0.95	0.764	1.07	0.905
		Linear	1.20	0.852	0.97	0.752	1.20	0.879
	AdaBoost	-	1.38	0.804	1.04	0.718	1.31	0.856
	XGBoost	-	0.67	0.955	0.68	0.879	1.03	0.912

Table 3: Results of four deep learning-based models for all stations

Input Combinations	Deep Learning models	Raipur		Jagdalpur		Ambikapur	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
Tmax,Tmin	DNN	1.41	0.804	1.11	0.697	1.14	0.577
	DRNN	1.19	0.859	1.14	0.684	1.18	0.592
	DLSTM	1.31	0.830	1.28	0.598	1.29	0.655
	DBiLSTM	1.53	0.767	1.16	0.673	1.31	0.679
Tmax, Tmin, RH1 and RH2	DNN	1.29	0.841	1.06	0.726	1.14	0.681
	DRNN	1.15	0.861	1.05	0.732	1.16	0.669
	DLSTM	1.27	0.840	1.19	0.653	1.41	0.507
	DBiLSTM	1.39	0.814	1.02	0.744	1.02	0.744
Tmax, Tmin, RH1, RH2 and WS	DNN	1.09	0.881	0.99	0.759	1.02	0.741
	DRNN	1.05	0.891	1.11	0.701	1.06	0.721
	DLSTM	1.14	0.870	1.10	0.704	1.00	0.751
	DBiLSTM	1.29	0.834	1.15	0.675	1.04	0.732
Tmax, Tmin, RH1, RH2, WS and BSS	DNN	1.15	0.870	0.98	0.766	1.06	0.721
	DRNN	1.11	0.878	0.97	0.771	1.15	0.674
	DLSTM	1.18	0.861	1.23	0.630	1.01	0.750
	DBiLSTM	1.95	0.622	0.99	0.760	1.19	0.652

better for Raipur and Jagdalpur, with low RMSE values of 0.56 and 0.63, respectively. However, for Ambikapur, the model with four hidden layers performs better, with an RMSE value of 0.65. The R^2 values for the five-input models with both three and four layers are almost identical across all three stations. The models with three hidden layers achieved the best results for the Jagdalpur station using all six inputs, with an RMSE of 0.54 and an R^2 of 0.929. In contrast, for Raipur and Ambikapur stations, the models with four hidden layers performed the best, with RMSE values of 0.63 and 0.60, and R^2 values of 0.959 and 0.904, respectively, using all six inputs.

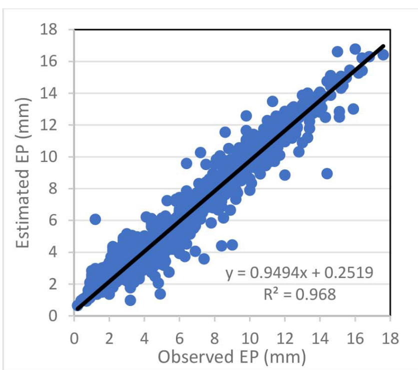
In summary, for the Raipur station, the proposed hybrid DNN+RF model with three hidden layers and five climate inputs performed better, with RMSE and R^2 values of 0.56 and 0.968, respectively. For Jagdalpur, the model with three hidden layers and

Table 4: Results of proposed hybrid DNN+RF model

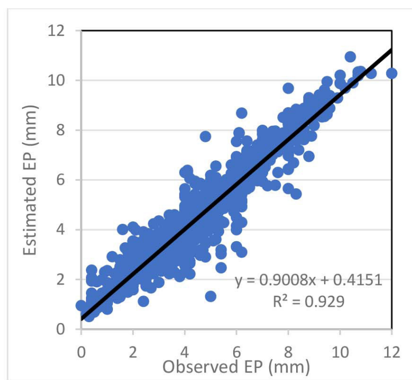
Inputs	Number of hidden layers	Raipur		Jagdalpur		Ambikapur	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
Tmax and Tmin	3	0.93	0.912	0.87	0.815	0.93	0.769
	4	0.97	0.905	0.93	0.788	0.93	0.770
Tmax, Tmin, RH1, and RH2	3	0.72	0.946	0.66	0.893	0.72	0.864
	4	0.73	0.946	0.67	0.889	0.66	0.886
Tmax, Tmin, RH1, RH2, and WS	3	0.56	0.968	0.63	0.904	0.67	0.883
	4	0.62	0.961	0.64	0.901	0.65	0.888
Tmax, Tmin, RH1, RH2, WS and BSS	3	0.65	0.886	0.54	0.929	0.66	0.885
	4	0.63	0.959	0.63	0.901	0.60	0.904

Table 5: Performance of all models with best input combinations for different regions.

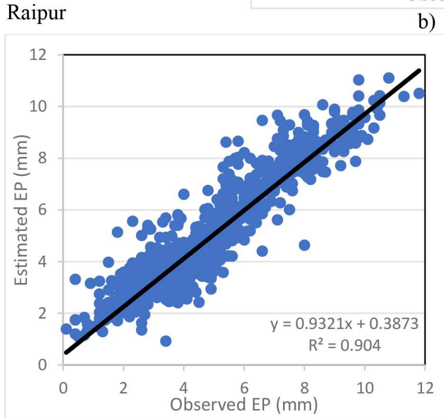
Model	Raipur		Jagdalpur		Ambikapur	
	Tmax, Tmin, RH1, RH2, and WS		Tmax, Tmin, RH1, RH2, WS and BSS		Tmax, Tmin, RH1, RH2, WS and BSS	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
DNN	1.09	0.881	0.98	0.766	1.06	0.721
DRNN	1.05	0.891	0.97	0.771	1.15	0.674
DLSTM	1.14	0.870	1.23	0.630	1.01	0.750
DBiLSTM	1.29	0.834	0.99	0.760	1.19	0.652
XGBoost	0.75	0.942	0.68	0.879	1.03	0.912
DNN+RF	0.56	0.968	0.54	0.929	0.60	0.904



a) Raipur



b) Jagdalpur



c) Ambikapur

Fig. 1: Comparison of observed and estimated EP (mm) during testing using the proposed hybrid DNN+RF model with the best combination of inputs and model parameters.

six inputs performed better, with an RMSE value of 0.54 and an R^2 of 0.929. For Ambikapur, the model with four hidden layers and six inputs performed better, with an RMSE of 0.60 and an R^2 of 0.904.

Table 5 shows the performance of various models based on optimal input combinations for Raipur, Jagdalpur, and Ambikapur. The proposed hybrid DNN+RF model achieves the best performance metrics for Raipur, with an RMSE of 0.56 and an R^2 of 0.968, significantly outperforming the other models. For Jagdalpur, the proposed hybrid DNN+RF model also excels, with an RMSE of 0.54 and an R^2 of 0.929, indicating strong predictive capability. In Ambikapur, the proposed hybrid DNN+RF model yields an RMSE of 0.60 and an R^2 of 0.904, again outperforming the other models. Overall, the results demonstrate that the proposed hybrid DNN+RF model effectively predicts EP across all three stations, particularly when using six inputs for Jagdalpur and Ambikapur, and five inputs for Raipur. This approach requires fewer epochs, reducing computation time while maintaining high predictive accuracy.

To exhibit the performance of the proposed hybrid DNN+RF model during testing, the observed EP and estimated EP values (obtained using the best combination of inputs and model parameters) are plotted and shown in Fig. 1 for Raipur, Jagdalpur, and Ambikapur. These figures show a very close mapping between the observed and estimated values for all three stations. Different combinations of inputs were considered to assess the performance of the models in the absence of one or more climatic inputs. When compared to other models from existing literature (Majhi and Naidu, 2021; Majhi *et al*., 2020), it is evident that the proposed model outperforms the existing models for all three stations.

CONCLUSION

This study evaluated different deep learning and machine learning models for estimating daily pan evaporation and proposed a hybrid DNN+RF model which demonstrated the best performance in terms of RMSE and R^2 values for all three stations (Raipur, Jagdalpur, and Ambikapur) of Chhattisgarh state. The deep learning models consistently outperformed machine learning models in EP prediction. The hybrid DNN+RF model proves to be well-suited for real-world EP estimation, due to its fast convergence and low computational time. Future research could focus on optimizing the hybrid model using swarm-based techniques, exploring multi-objective algorithms, and integrating quantum computing for enhanced performance.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Department of Agro-meteorology, Indira Gandhi Krishi Vishwavidyalaya, Raipur, for providing the necessary meteorological data for this study.

Conflict of Interests: The authors declare that there is no conflict of interest associated with this article.

Funding: Not applicable

Data availability: To be provided on request

Authors' Contributions: **R. Naik** – Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original draft; **B. Majhi** – Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – Review & editing; **D. Naidu** – Data collection, Formal analysis, Supervision, Investigation, Manuscript revision.

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 about jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Abed, M., Imteaz, M. A., Ahmed, A. N., and Huang, Y. F. (2022). Modelling monthly pan evaporation utilising Random Forest and deep learning algorithms. *Sci. Rep.,* 12(1): 13132. https://doi.org/10.1038/s41598-022-17263-3.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M. (1998). CropEvapotranspiration: Guidelines for computing crop waterrequirements. Irrigation and Drainage Paper 56, FAO ofthe United Nations, Rome. 300 pp

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1): 5-32. https://doi.org/10.1023/A:1010933404324

Dong, L., Zeng, W., Wu, L., Lei, G., Chen, H., Kumar Srivastava, A. and Gaiser, T. (2021). Estimating the pan evaporation in northwest China by coupling catboost with bat algorithm. *Water (Switz.),* 13(3): https://doi.org/10.3390/w13030256.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comp.*, 9(8):1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Ibrahim, M., and Elhafiz, R. (2023). Modeling an intrusion detection using recurrent neural networks. *J. Eng. Res.*, 11(1), 100013. https://doi.org/10.1016/j.jer.2023.100013

Ji, C., Zou, X., Hu, Y., Liu, S., Lyu, L., and Zheng, X. (2019). XG-SF: An XGBoost classifier based on shapelet features for time series classification. *Procedia Comput.Sci.*, 147: 24-28. https://doi.org/10.1016/J.PROCS.2019.01.179

Kumar, A., Sarangi, A., Singh, D. K., Mani, I., Bandyopadhyay, K. K., Dash, S. and Khanna, M. (2024). Evaluation of soft-computing techniques for pan evaporation estimation. *J. Agrometeorol.,* 26(1): 56-62. https://doi.org/10.54386/jam.v26i1.2247

Lakhawat, S. S., Sharma, V., Singh, T. K., Patil, P., Priyadevi, S. and Gutam, S. (2024). Effects of pan evaporation-based drip irrigation levels on the performance of guava grown in Udaipur and Rewa regions of India. *J. Agrometeorol.,* 26(1): 69-73. https://doi.org/10.54386/jam.v26i1.2306

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature,* 521(7553): 436-444. https://doi.org/10.1038/nature14539.

Majhi, B., Naidu, D., Mishra, A. P. and Satapathy, S. C. (2020). Improved prediction of daily pan evaporation using Deep-LSTM model. *Neural Comput. Appl.,* 32(12): 7823-7838. https://doi.org/10.1007/s00521-019-04127-7

Majhi, B. and Naidu, D. (2021). Pan evaporation modeling in different agroclimatic zones using functional link artificial neural network. *Inf. Process. Agric.,* 8(1): 134-147. https://doi.org/10.1016/j.inpa.2020.02.007

Mehta, Rashmi and Vyas Pandey. (2015). Reference evapotranspiration (ETo) and crop water requirement (ETc) of wheat and maize in Gujarat. *J. Agrometeorol.,* 17(1): 107-113. https://doi.org/10.54386/jam.v17i1.984

Mehta, Rashmi and Vyas Pandey (2018). "Crop Water Requirement-Evapotranspiration and Estimation" Educationist Press, Write and Print Publication New Delhi. Pp 106. ISBN 97893862863702

Moayedi, H., Ghareh, S. and Foong, L. K. (2022). Quick integrative optimizers for minimizing the error of neural computing in pan evaporation modeling. *Eng. Comput.*, 38: 1331-1347. https://doi.org/10.1007/s00366-020-01277-4

Niknam, A., Zare, H. K., Hosseininasab, H., and Mostafaeipour, A. (2023). Developing an LSTM model to forecast the monthly water consumption according to the effects of the climatic factors in Yazd, Iran. *J. Eng. Res.,* 11(1): 100028. https://doi.org/10.1016/j.jer.2023.100028

Sharma, V., Singh, P. K., Bhakar, S. R., Yadav, K. K., Lakhawat, S. S. and Singh, M. (2023). Pan evaporation and sensor-based approaches for irrigation scheduling for crop water requirement, growth, and yield of okra. *J. Agrometeorol.,* 23(4): 389-395. https://doi.org/10.54386/jam.v23i4.142

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Stat. Comp.*, 14(3):199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Srivastava, A. K., Naidu, D., Bhan, M. and Bal, L. M. (2022). Neural network based predictors for evaporation estimation at Jabalpur in central India. *J. Sci. Ind. Res.,* 81(3): 319-328. https://doi.org/10.56042/jsir.v81i03.58166

Zhang, S. (2012). Nearest neighbor selection for iteratively kNN imputation. *J. Syst. Softw.,* 85(11): 2541-2552. https://doi.org/10.1016/j.jss.2012.05.073