International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 7 Issue 4, July-August 2023 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Estimating Evaporation using Machine Learning Based Ensemble Technique

R. S. Parmar[1], G. B. Chaudhari[2], S. H. Bhojani[3]

[1]Professor and Head, College of Agricultural Information Technology, AAU, Anand, Gujarat, India
[2]Assistant Professor, College of Agricultural Information Technology, AAU, Anand, Gujarat, India
[3]Assistant Professor, IT Center, AAU, Anand, Gujarat, India

ABSTRACT

Accurately estimating evaporation is necessary for calculating and scheduling irrigation water requirements. Current literature points to the use of individual machine learning models for better estimation of evaporation. However, such methods have not been used in the Indian framework. Moreover, given the diversity of climate, it is necessary to develop an ensemble technique incorporating a significant number of machine learning algorithms to have a better estimation of weekly evaporation. The purpose of this paper is to develop an ensemble technique that makes the machine learning models that have a better estimation of weekly evaporation. The results showed that the Bagging Random Forest model has a much better performance in estimating weekly evaporation compared to other fitted ensemble models.

***KEYWORDS**: Machine Learning, Ensemble, Bagging, Evaporation, Random Forest*

***How to cite this paper**:* R. S. Parmar | G. B. Chaudhari | S. H. Bhojani "Estimating Evaporation using Machine Learning Based Ensemble Technique" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-7 | Issue-4, August 2023, pp.985-990, URL: www.ijtsrd.com/papers/ijtsrd59847.pdf

IJTSRD59847

INTRODUCTION

Estimating evaporation is essential for managing water resources, optimizing irrigation schedules, and modeling agricultural production (Konapala *et al.*, 2020; Short Gianotti *et al.*, 2020). Besides, evaporation has significant importance in studying climate change because this parameter scatters a good proportion of the global precipitation (Ma, 2018; Wang *et al.* 2019). Few studies have been conducted to solve different water resource problems using different artificial intelligence approaches namely random forest, support vector machine, extreme learning machine, feed-forward neural network, Gaussian process regression, and, gradient boosting model (Hameed *et al.,2021*; Ghorbani *et al.,2020;* Ashrafzadeh *et al.,2018*). Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. They combine multiple algorithms to produce better classification and regression performance. Ensemble techniques improve the accuracy of fitted models by combining multiple models in place of using a single model. Bagging is an ensemble technique that helps to improve the performance and accuracy of machine learning models. It is used to reduce the variance of an estimation model. Bagging avoids the overfitting of data and is widely used for regression models. Ensemble techniques have been used to develop a system of crop yield estimation and fertilizer recommendation (Kumaravel *et al*., 2020). Machine learning-based classification and regression techniques are widely used in agriculture for estimating outcomes using datasets that can often comprise hundreds of features and observations (Kamani *et al*. 2019, Kamani *et al.* 2021, Parmar *et al*. 2022). Estimation of wheat crop yield using the machine learning techniques and the use of different activation functions have been used and recommended to choose the activation function consideting the research purpose and they type of datasets (Shital et al., 2021). The purpose of this paper is to determine the effectiveness of the machine learning-based ensemble technique for estimating weekly evaporation using weekly weather data of the Anand district of Gujarat.

OBJECTIVE

To develop an ensemble technique that makes the machine learning models for better estimation of weekly evaporation.

MATERIALS AND METHODS

The present study was undertaken to explore the possibility of estimating evaporation by using the effect of weekly weather variables. For judging the joint influence of weather variables, a week-wise approach was considered. The weekly weather data for Anand viz; temperature, sunshine hour, wind velocity, relative humidity, and evaporation for 43 years i.e. from 1980-2022 were collected from the Department of Agricultural Meteorology, Anand Agricultural University, Anand.

The ensemble is one of the most popular and successful techniques in machine learning. Ensemble learning is a learning method that consists of combining multiple machine learning models. A problem in machine learning is that individual models tend to perform poorly. The individual models are known as weak learners. Weak learners either have a high bias or high variance. Ensemble learning improves a model's performance in mainly three ways:

1. By reducing the variance of weak learners
2. By reducing the bias of weak learners,
3. By improving the overall accuracy of strong learners.

Bagging (Bootstrap aggregating) is used to reduce the variance of weak learners. Fig.1 depicts a conceptual view of the Bagging (Bootstrap aggregating) Process.

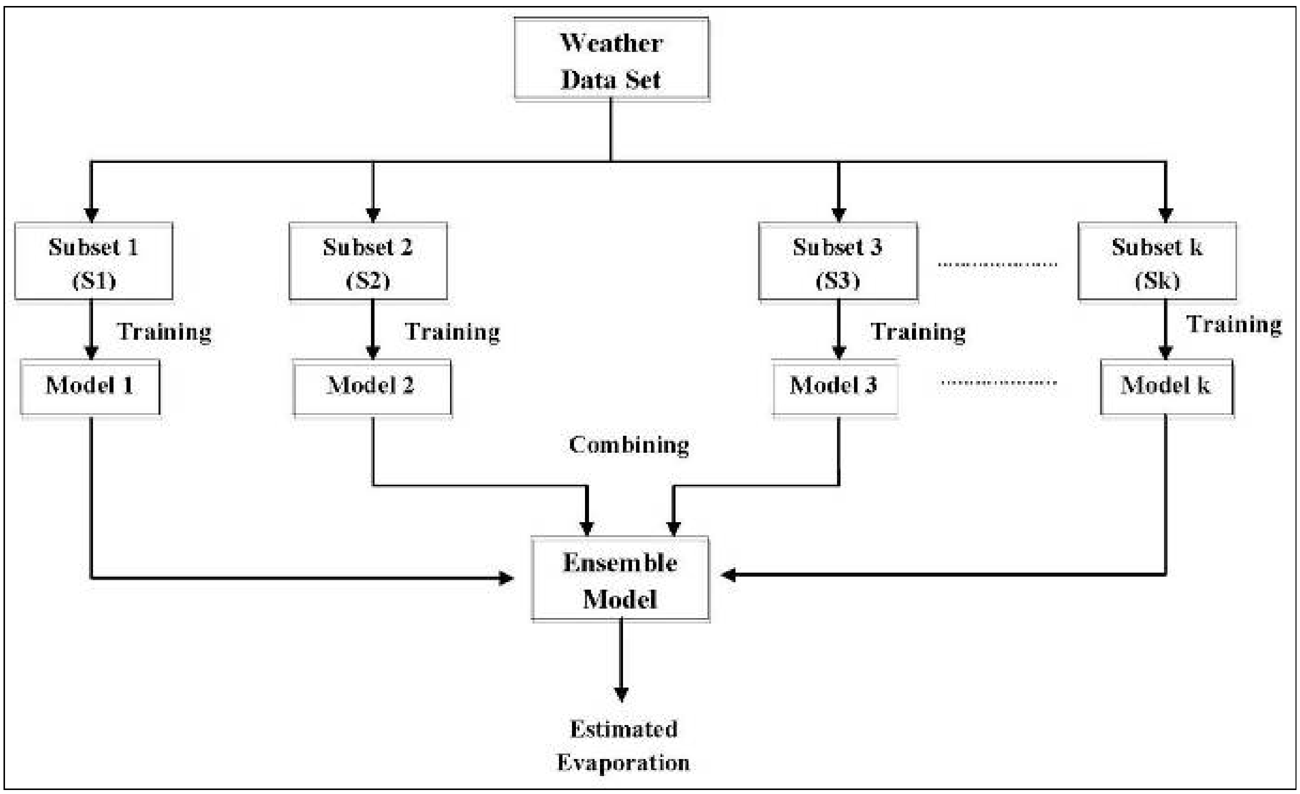

Fig. 1. Conceptual view of Bagging (Bootstrap aggregating) Process

RESULTS AND DISCUSSION

An open source weka version 3.8.5 is an ensemble toolkit for data regression and visualization. It was used to evaluate the performance and effectiveness of machine learning-enabled ensemble evaporation estimation models and 6 ensemble models were built from the bagging technique.

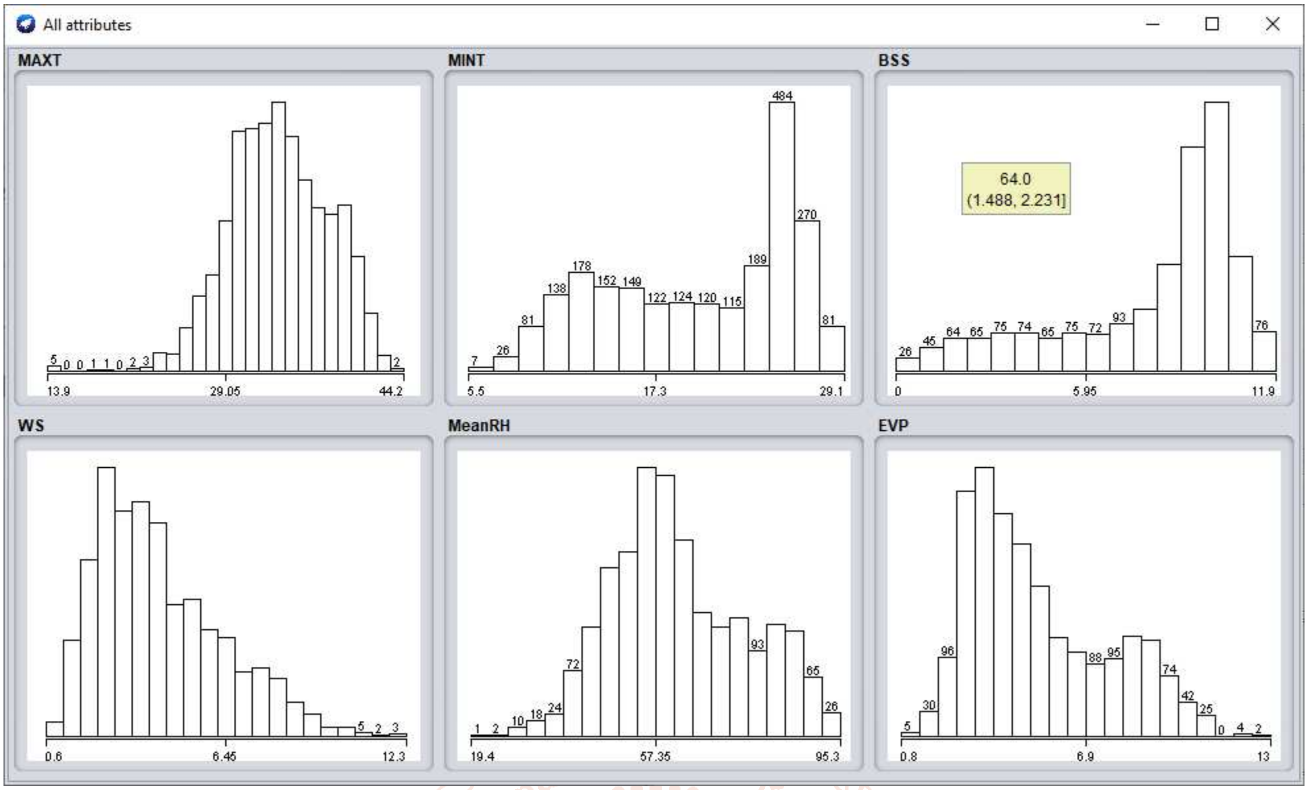

Fig. 2. Selected Weekly Weather Variables Distribution

The Bagging Linear Regression, Bagging Neural Network, Bagging REP Tree, Bagging Random Forest, Bagging KNN, and Bagging Support Vector Machines models were used to examine estimating evaporation. The result of each fitted ensemble model is checked in terms of R^2, MAE, RMSE, RAE, and RRSE. Fig. 2 demonstrates that the selected weekly weather variables have different distribution ranges.

Table 1 shows the characteristics of fitted machine learning-based ensemble models. Out of 6 formations of ensemble models in this research work Bagging Random Forest has achieved better performance than other fitted models. In general, it could be observed that Bagging Random Forest is the best-fitted ensemble model to estimate evaporation.

Table 1. Characteristics of Fitted Machine Learning based Ensemble Models

Ensemble Models	Parameters				
	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Relative Absolute Error (RAE)	Root Relative Squared Error (RRSE)	Coefficient of Determination (R^2)
Bagging Linear Regression	0.6419	0.8202	33.41 %	35.65 %	87.27 %
Bagging Neural Network	0.5475	0.7176	28.50 %	31.19 %	90.27 %
Bagging REP Tree	0.5604	0.7362	29.17 %	32.00 %	89.76 %
Bagging Random Forest	0.5345	0.7010	27.82 %	30.47 %	90.71 %
Bagging KNN	0.6406	0.8565	33.34 %	37.23 %	86.30 %
Bagging Support Vector Machines	0.6413	0.8213	33.38 %	35.70 %	87.25 %

Fig. 3 demonstrates the estimation accuracy of different fitted ensemble models. Bagging Random Forest has better estimation accuracy than other fitted ensemble models with 90.70 %, followed by Neural Network with 90.30 %. Bagging KNN has the lowest estimations accuracy with 86.30 %.

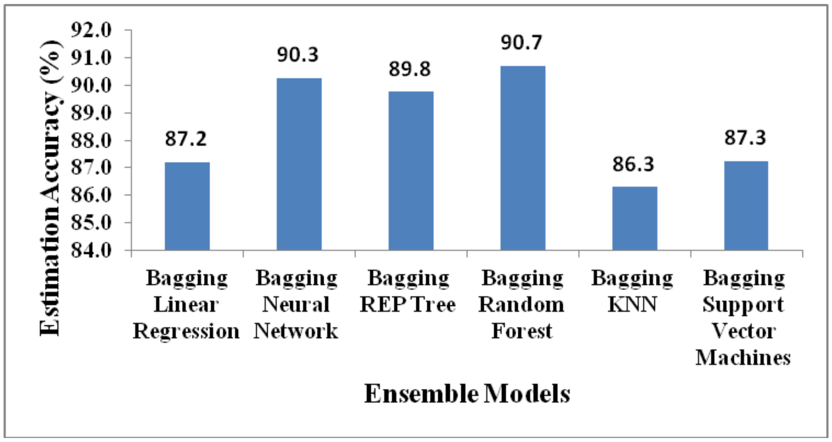

Fig. 3. Estimation Accuracy of Different Fitted Ensemble Models

Fig. 4 shows the error results of the different fitted ensemble models. Bagging Random Forest has the lowest Mean Absolute Error (MAE) of 0.53 and Root Mean Squared Error (RMSE) of 0.70. This reveals minimal error reported during the estimation of evaporation. Bagging KNN has the highest error with 0.64 and 0.86 of MAE and RMSE, respectively.

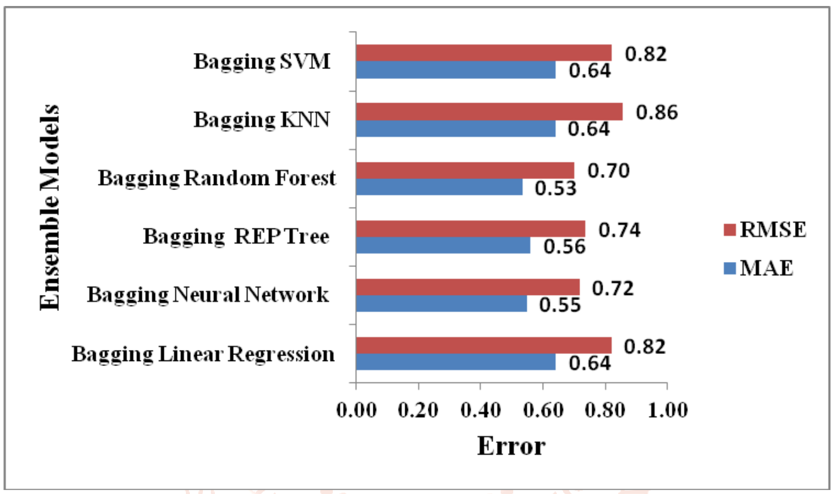

Fig. 4. Error Results of Different Fitted Ensemble Models

Fig. 5 depicts the Mean Absolute Error (MAE) of the Bagging Random Forest model. The measure of estimation accuracy is also called MAE and a low MAE suggests the fitted ensemble model is good at an estimation of evaporation. Multiple Correlation Coefficient (R) is a measure of how well evaporation can be estimated using a linear function of a set of weekly weather variables. Usually, a higher R-value indicates a better estimation of the evaporation from the selected weekly weather variables. MAE (0.53) and R (0.95) values were low and high respectively and thus indicated an excellent job by the fitted ensemble model.

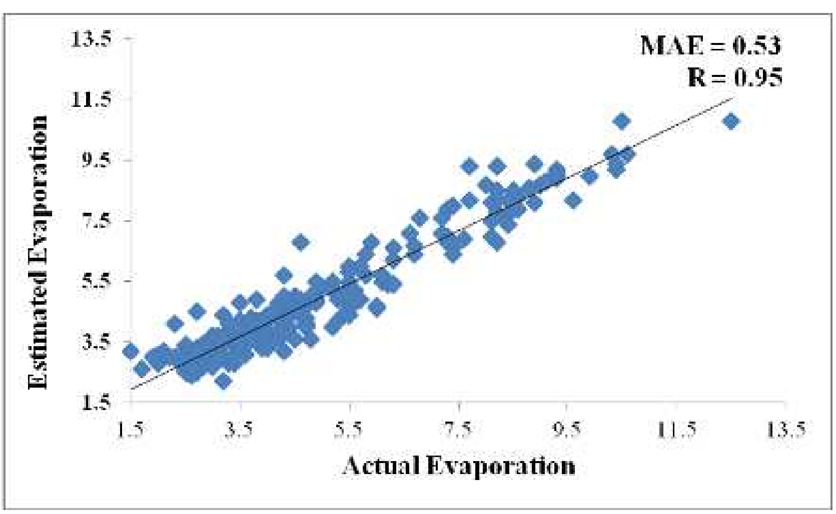

Fig. 5. Mean Absolute Error of Bagging Random Forest

The estimated evaporation is revealed in Fig. 6. It is observed that the actual evaporation and the estimated evaporation are very close to each other. The estimated evaporation showed deviations from actual evaporation ranging between –1.7 to 2.2.

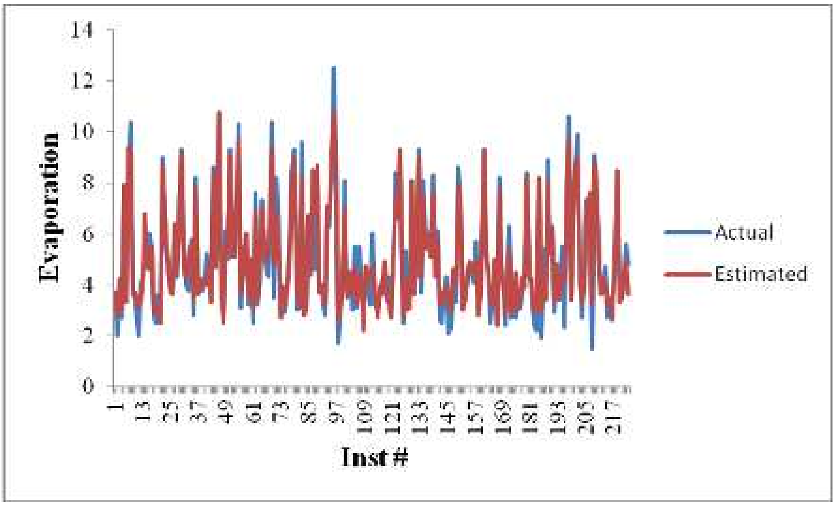

Fig. 6. Comparisons between Actual and Simulated estimation of evaporation using Bagging Random Forest

CONCLUSION

Bagging is an ensemble machine learning technique that helps to avoid overfitting data. It is a model averaging procedure that is often used with decision trees but can also be applied to other algorithms. It was observed that bagging random forest was the best fitted ensemble model for estimating weekly evaporation by achieving the highest coefficient of determination (R^2) of 90.71 % as compared with other fitted models. The fitted ensemble model has the lowest MAE of 0.53 and RMSE of 0.95. There is also a significant scope for using the ensemble ML technique in estimating the weekly evaporation of other data samples and extending the support of analysis. Hence, it can be concluded that the study helps the researcher in efficient ensemble model selection for estimating weekly evaporation. Scientific Community has recommended using the ensemble technique bagging for estimating weekly evaporation.

REFERENCES

[1] Ashrafzadeh, A., Malik, A., Jothiprakash, V., Ghorbani, M. A. and Biazar, S.M. (2018). “Estimation of daily pan evaporation using neural networks and meta-heuristic approaches” *ISH Journal of Hydraulic Engineering*, 26(4): 421–429.

[2] Ghorbani, M.A., Deo, R.C., Kim, Hasanpour Kashani, S.M., Karimi, V. and Izadkhah, M. (2020). “Development and evaluation of the cascade correlation neural network and the random forest models for river stage and river flow prediction in Australia,” *Soft Computing*, 24(16): 12079–12090.

[3] Hameed, M.M., Alomar, M.K. and Mohd Razali, S.F. (2021). “Application of artificial intelligence models for evapotranspiration prediction along the southern coast of Turkey” *Complexity*, 2021 Article ID 8850243, 20 pages.

[4] Kamani, G.J., Parmar, R.S., Ghodasara, Y.R. (2019). Data normalization in data mining using graphical user interface: a pre-processing stage, Guj. J. Ext. Edu, 30(2): 106-109.

[5] Kamani, G.J., Parmar, R.S., Ghodasara, Y.R. (2021). Machine learning models for productivity trend of Rice crop, Guj. J. Ext. Edu., 32(2): 189-196.

[6] Konapala, G., Mishra, A. K. Wada, Y. and Mann, M.E. (2020). “Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation,” *Nature Communications*, 11(3044): 1–10.

[7] Kumaravel, A. and Saranya, K.G. (2020). Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier. *Seventh Sense Research Group (SSRG) International Journal of Computer Science and Engineering*, 7(5): 1-4.

[8] Ma, W. (2018). “Research progresses of flash evaporation in aerospace applications”. *International Journal of Aerospace Engineering*, 2018. Article ID 3686802.

[9] Parmar, R.S., Vegad, N.M., Mehra, V.I. (2022). An attitude of PG scholars of agricultural

extension towards Application of mobile technology using artificial intelligence technique, Guj. J. Ext. Edu, 33(1): 136-143.

[10] Parmar, R.S., Mehra, V.I., Kamani, G.J. (2022). Analyzing Soil fertility using Data Mining Techniques, Guj. J. Ext. Edu, 34(2): 47-50.

[11] Short Gianotti, D.J., Akbar, R., Feldman, A.F., Salvucci, G.D. and Entekhabi, D. (2020). "Terrestrial evaporation and moisture drainage in a warmer climate," *Geophysical Research Letters*, 47(5), Article ID e2019GL086498.

[12] Wang, S., Lian, J., Peng, Y., Hu, B. and Chen, H. (2019). "Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China," *Agricultural Water Management*, 221: 220–230.

[13] Bhojani, S. H., and Bhatt, N. (2021). "Performance Analysis of Activation Functions for Wheat Crop Yield Prediction." In IOP Conference Series: Materials Science and Engineering (Vol. 1042, No. 1, p. 012015). IOP Publishing.