Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 77(2) 2023 185–192

Wavelet Extreme Learning Machine (W-ELM) Model for Drought Index Forecasting

K.N. Singh, Rajeev Ranjan Kumar and Mrinmoy Ray

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 06 December 2022; Revised 14 March 2023; Accepted 03 April 2023

SUMMARY

In an agriculturally depending country like India, accurate and reliable drought forecasting is very important to figure out how drought will affect water resources and agriculture. Data-driven machine learning forecasting techniques are promising approaches for drought forecasting since they take less development time, fewer inputs, and are less sophisticated than dynamic or physical models. Machine learning models for drought forecasting use drought indices that are more operational than raw climatic variables. In this study, the potential of wavelet-based extreme learning machine (W-ELM) model to forecast effective drought index has been explored for Sagar and Chattarpur districts of the Bundelkhand region of India.The performance of W-ELM model has been compared with the other competitive machine learning models like support vector machine (SVM), extreme learning machine (ELM), and artificial neural network (ANN). Observational outcomes reveals that the W-ELM model outperforms ELM, SVM, and ANN.

Keywords: Drought forecasting; ELM; SVM; ANN; Wavelet transformation.

1. INTRODUCTION

Drought is a multifaceted meteorological phenomenon of low rainfall regions that significantly impact agriculture, water resource sustainability, climate, and environmental management. According to Intergovernmental Panel on Climate Change (IPCC) report on catastrophic events, drought is a severe climatic event that must be managed to prevent its harmful consequences (Field, 2012). Hence, forecasting of droughts are very important for its management and mitigation (Mishra & Singh, 2010). Drought indices (DIs) are generally used to forecast drought, which are standardized measures of weather parameters like temperature, rainfall, evapotranspiration, etc. Numerically defined drought indices are more operational than basic climatic variables. Thus, they can be utilized as better triggers for recognizing the start and end of droughts, which is vital for contingency planning, prevention, and decision-making.

In the literature, many drought indices have been reported, like Effective Drought Index (EDI) (Byun & Wilhite, 1999), Standardized Precipitation Index (SPI) (McKee *et al.*, 1993; Yaseen *et al.*, 2021) etc. EDI is regarded as preferable for drought assessment among various DIs because it can identify the beginning and end of the drought (Jain *et al.*, 2015). It evaluates drought by utilising the idea of a water resource excess or deficit, so it is a reliable assessment metric (Jain *et al.*, 2015; Kim *et al.*, 2009; Malik & Kumar, 2020). The EDI was generated using effective precipitation (PE) and a time-dependent reduction function that determined the cumulative rainfall for the current and prior days. Therefore, in this study, EDI is employed to model and forecast drought. EDI's 'drought range' is defined as: (a) Extremely dry conditions at less than -2 (b) Drought is severe at-1.99 to-1.5 (c) Drought of moderate severity at -1.49 to -1 (d) Conditions are close to usual at -0.99 to 0.99. For EDI computation, *'R software'* package *'EDI'* has also been developed, which is accessible at https://github.com/rrk4910/EDI (Kumar *et al.* 2020).

Corresponding author: Rajeev Ranjan Kumar
E-mail address: rajeev.kumar4@icar.gov.in

The application of machine learning (ML) techniques in hydro-climatic investigations are expanding rapidly. However, there are minimal applications of machine learning algorithms for drought forecasting. The widely employed ML algorithm in the hydrological processes is an artificial neural network (ANN), since it is an efficient data-driven algorithm to capture complex nonlinear dynamics (Deo *et al.*, 2017).ANN is a very dynamic machine learning model as it requires very less assumptions about the processes under consideration. Due to this reason, ANN is suitable for the drought forecasting, as it is very difficult to define accurate variables that cause drought (Deo *et al.*, 2017; Deo & Şahin, 2015; Dikshit *et al.*, 2022). As ANN is supervised learning algorithm, input data and the corresponding output values are essential for the training. However, the ANN faces several challenges, including the need for iterative model parameter adjustment, time-consuming, and relatively low forecast performance compared to other advanced ML techniques (Acharya *et al.*, 2013). Apart from ANN, another major machine learning approach for complex system prediction is Support Vector Machine (SVM). When compared to gradient-based approaches, which are susceptible to the occurrence of local minima, an SVM model provides global error function solutions, which is a significant benefit (Deo *et al.*, 2016). In this study, we also used a support vector machine model to forecast effective drought index.

In hydrological forecasting, the extreme learning machine (ELM) model has gained traction since it requires less processing time and has a higher generalisation ability than traditional ANN models. First of all, Huang *et al.* (2006) proposed the concept of ELM (Huang *et al.*, 2006). In the extreme learning machine, hidden nodes are selected at random and output weights are calculated analytically. The implementation of ELM model has less complication than the traditional ANN model. It also avoids the problems of local minima. When compared to the ANN and SVM, the generalisation performance of ELM is significantly better.

The main contributions of the present paper consist of two parts. First, an attempt has been made to explore the potentiality of wavelet-based extreme learning machine model for the forecasting effective drought index in the Indian context. Second, we carry out an extensive evaluation of modern neural network architectures in effective drought index forecasting.

2. METHODOLOGY

2.1 Artificial Neural Network (ANN)

The ANN computing paradigm consists of massively connected networks with various weights that connect non-linear elements (neurons) functioning in parallel. To produce a single output known as the neuron's activation level, a single neuron sums up all of its inputs, adds a bias term, and then passes the result via a typically nonlinear activation function. Network architecture, neuron properties, and training or learning methods are used to define ANN models. These models have inputs, outputs, hidden layers, and hidden layers with inter connections. The fundamental processing unit is a neuron, which computes a weighted sum of its input signals, y_i, for i = 0,1,2,. . . , n, hidden layers, w_{ij} and then applies a nonlinear activation function to produce an output signal u_j.

A neuronal model consists of an externally applied bias, b_k which has the effect of increasing or decreasing the net input of the activation functions depending on whether it is positive or negative. Mathematically, a neuron k may be described by

$$u_k = \sum_{j=1}^{m} w_k x_j$$

$$y_k = \Phi(u_k + b_k)$$

where x_1, x_2, . . ., x_m are the inputs signals; w_{k1}, w_{k2}, . . ., w_{km} are the synaptic weights of neuron k; u_k is the linear combiner output due to input signals; b_k is the bias; Φ (.) is the activation function and y_k is the output signal of the neuron. Bias b_k has the effect of applying an affine transformation to the output u_k of the linear combiner in the model.

$$v_k = u_k + b_k$$

In particular, depending on whether b_k is positive or negative, the relationship between the induced local field or activation potential v_k of neuron k and linear combiner output u_k can be modified. The bias b_k is an external parameter of artificial neuron k.

$$v_k = \sum_{j=0}^{m} w_{kj} x_j$$

$$y_k = \Phi(v_k)$$

The tangent sigmoid, $\phi(x)$ logarithmic sigmoid, $\varnothing(x)$ and linear, $\chi(x)$ transfer function are described as follows.

$$\phi(x) = \frac{2}{1+e^{-2x}} - 1$$

$$\varnothing(x) = \frac{1}{1+e^{-x}}$$

$$\chi(x) = linear(x) = x$$

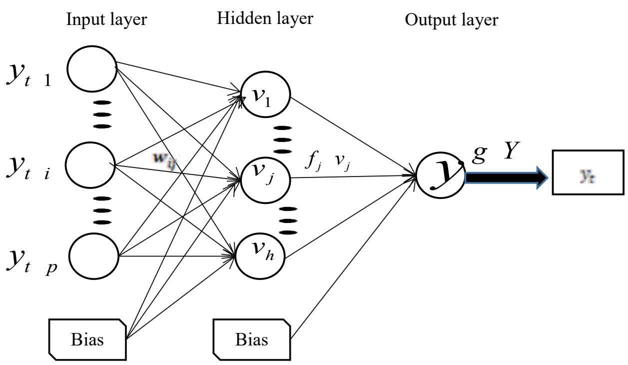

Fig. 1 represents the basic architecture of the ANN model.

2.2 Support Vector Machine (SVM)

In this study we incorporated support vector machine for the comparison with the proposed model. SVM-based models adhere to the structural risk minimization principle in contrast to neural network models which are developed to reduce empirical risk.

Given an input–output data set, X and Y comprised of N training data samples, i.e., (x_t, y_t) for $t = 1, 2, \ldots, N$, where $x_t \in \mathbb{R}^d$ and $y_t \in \mathbb{R}^d$; the SVM model is minimized with respect to the loss function L defined by

$$L(W, e) = \frac{1}{2}W^T W + C\frac{1}{2}\sum_{t=1}^{N} e_t^2,$$

where

e_t^2 is the quadratic loss term,

W is the weight vector and

C is the regularization parameter

To solve for the SVM parameters, the Lagrangian multipliers method is used as follows:

$$L(W, k, e, \alpha) = L(W, e) - \sum_{t=1}^{N} \alpha_t \left\{ W^T \phi(x_t) + k + e_t - y_t \right\}$$

where

ϕ (x) is a nonlinear mapping function

$\alpha \in \mathbb{R}^N$ is the set of Lagrange multipliers and

$k \in \mathbb{R}$ is the bias term.

The conditions which prove to be optimal in solving the LSSVR parameters are determined by taking partial derivatives of the extended loss function [i.e., L($W,k,e\alpha$)] with respect to each term ($W,k,e\alpha$) as follows:

$$\frac{\delta L}{\delta W} = 0 \rightarrow W = \sum_{t=1}^{N} \alpha_t \phi(x_t)$$

$$\frac{\delta L}{\delta k} = 0 \rightarrow \sum_{t=1}^{N} \alpha_t = 0$$

$$\frac{\delta L}{\delta e_t} = 0 \rightarrow \alpha_t = Ce_t, \quad t = 1, \ldots, N$$

$$\frac{\delta L}{\delta \alpha_t} = W^T \phi(x_t) + k + e_t - y_t \ , \ t = 1, \ldots, N$$

The above conditions can be expressed in matrix form as:

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega + C^{-1} I \end{bmatrix} = \begin{bmatrix} K \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}$$

where

$\vec{1}$ is a vector of ones and

Ω is used to represent the kernel function (K) satisfying Mercer's theorem:

$$\Omega_{uv} = \phi(x_u)^T \phi(x_v) = K(x_u, x_v), \ u, v = 1, \ldots, N,$$

where the kernel function $\left[K(x_u, x_v) \in \mathbb{R}\right]$ may be represented by the commonly used RBF:

$$K(x_t, x) = e^{\left(-\frac{x - x^2}{2\sigma^2}\right)}$$

σ represent the RBF kernel width.

2.3 Extreme learning machine (ELM)

Extreme learning machine (ELM) developed by Huang *et al.* (2006) is the state-of-art novel machine learning algorithm for Single Layer Feed forward Neural Network (SLFN). Consequently, the ELM model has been widely used for the solution of estimation problems in many different fields and is now gaining attention within the financial time series. The ELM model is easy to use, and no parameters need to be tuned except the predefined network architecture, thus avoiding many complications faced by the gradient-based algorithms such as learning rate, learning epochs, and local minima. Importantly the ELM model has also been proven to be a faster algorithm compared with other conventional learning algorithms such as back propagation (BP) or support vector machines (SVM). In the ELM approach most of the training is accomplished in time span of seconds or at least in minutes in large complex applications which are not easily achieved by using the traditional neural network models.

ELM was proposed for "generalized" single-hidden layer feed forward networks where the hidden layer need not be neuron alike (Huang *et al.,* 2006). The output function of ELM for generalized SLFN is

$$f_L(X) = \sum_{i=1}^{L} \beta_i h_i(X) = h(X)\beta$$

where $\beta = [\beta_1, \ldots, \beta_L]^T$ is the output weight vector between the hidden layer of L nodes to the $m \geq 1$ output nodes, and $\boldsymbol{h}(\boldsymbol{X}) = [h_1(\boldsymbol{X}), \ldots, h_L(\boldsymbol{X})]$ is ELM nonlinear feature mapping, e.g., the output (row) vector of the hidden layer with respect to the input $\boldsymbol{X}$.

$h_i(\boldsymbol{X})$ is the output of the i^{th} hidden node output. The output functions of hidden nodes may not be unique. Different output functions may be used in different hidden neurons.

In particular, in real applications $h_i(\boldsymbol{X})$ can be

$$h_i(\boldsymbol{X}) = G(\boldsymbol{a}_i, b_i, \boldsymbol{X}), \boldsymbol{a}_i \in \boldsymbol{R}^d, b_i \in \mathrm{R}$$

where $G(\boldsymbol{a}, b, \boldsymbol{X})$ (with hidden node parameters (a, b)) is a nonlinear piecewise continuous function satisfying ELM universal approximation capability theorems (Huang, *et al.*, 2006).

2.4 Wavelet Extreme learning machine (W-ELM)

In this study, to improve the prediction performance of ELM model, discrete wavelet transformation (DWT) is used. Mathematically, wavelet transformation coefficients $W_f(r,s)$ of signal *f(t)*is defined as:

$$W_f(r,s)^* = |r|^{\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \left(\frac{t-s}{r}\right) dt$$

where, * denotes complex conjugate, r is the scale factor, contraction $(r < 1)$ and dilation $(r > 1)$, and $s$denotes time factor. The input variables used in the prediction of drought exhibit lolalized low and high-frequencies as well as nonlinear and nonstationary in nature, similar to other hydro-meteorological signals(Deo *et al.*, 2016). Hence, in this study, for efficient and reliable drought forecasting, a wavelet-based extreme learning machine (W-ELM) model has been proposed.

3. RESULTS AND DISCUSSIONS

This study utilized monthly precipitation data of Sagar and Chattarpur district of the Bundelkhand region, India, for 41 years (Jan-1980 – Dec-2020) to evaluate the forecasting performance of the models under consideration. The time plot of the precipitation is presented in Figures 2 and 3 for Sagar and Chattatpur districts, respectively. The average annual precipitation in the Sagar district was 1060 mm, while the average annual precipitation in the Chattarpur district was 1032 mm, with coefficients of variation of 0.211 and 0.257, respectively. In order to implement the machine learning models under consideration, by using the 'R' package *'EDI',* original precipitation data has been

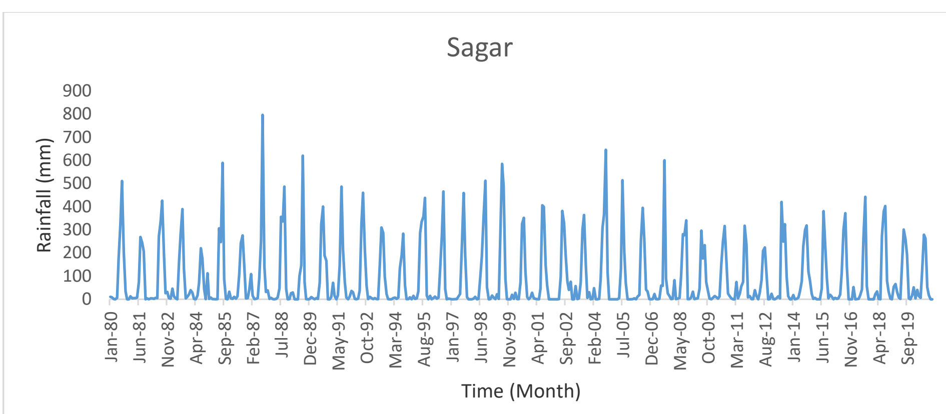

Fig. 2. Precipitation plot of Sagar district from Jan-1980 to Dec-2020

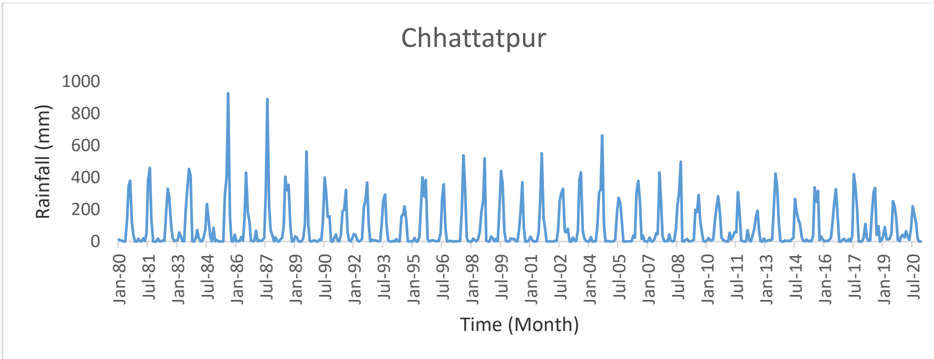

Fig. 3. Precipitation plot of Chhattarpur district from Jan-1980 to Dec-2020

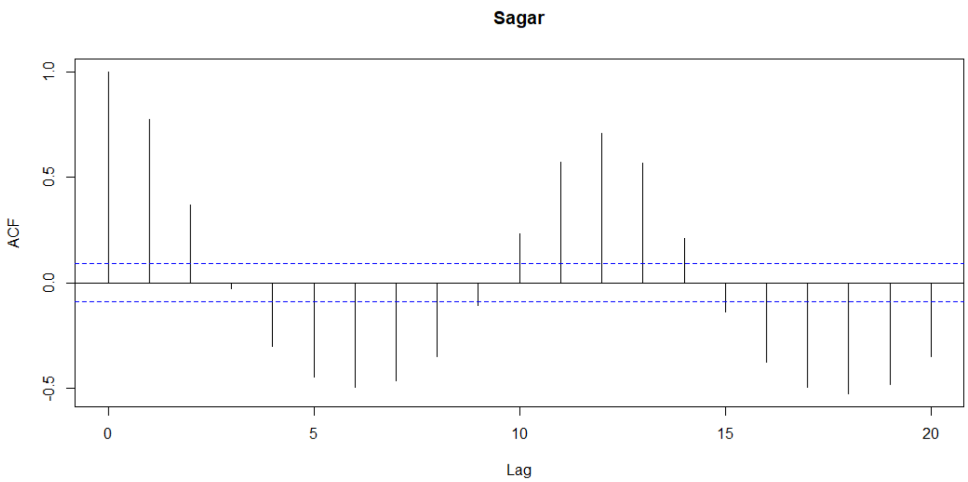

Fig. 4. ACF of Sagar district

converted to effective drought index series (Byun & Wilhite 1999, Kumar *et al.* 2020). The converted series is used as an input variable for ANN, SVM and ELM models. All the machine learning models have been implemented in R software.

For the implementation of ANN and ELM models, it is critical to think about the best combinations of predictor variables, selection of good training algorithm, appropriate number of hidden layer and hidden nodes etc. However, there is no standard algorithm for the selection of the input variables. It has been done by hit-and-trial and using the autocorrelation function (ACF) and partial autocorrelation function (PACF). Figs. 4 and 6 represents the ACF plots of Sagar and Chattarpur, respectively. Meanwhile, Figs. 5 and 7 depict the partial autocorrelation functions (PACF) for Sagar and Chattarpur, respectively.

A three-layer network that contained the input, feature optimization and the output space was employed with the predictor dataset from training and the testing sets used for developing the present drought models. ELM model in this study was developed using the logarithmic sigmoid activation function. To identify the best ELM network architecture (i.e., number of hidden neurons), the number of hidden neurons was decided upon a priori. The architecture that performed best on a particular partition was then used to justify a particular architecture as optimal. Initially, the ELM model was randomly executed 50–1000 times to explore the effect of the variation of the randomized hidden layer weights and biases on the network's output. The objective was to obtain the smallest mean square error (MSE) for the weights and optimal nodes in the hidden layer. Finally, this resulted in 100 randomizations that were appropriate for a stable solution of the forecasted monthly EDI. For each station, the time to run drought models was also recorded.

A variety of activation functions were tested one by one in this study, like log-sigmoid, hyperbolic-

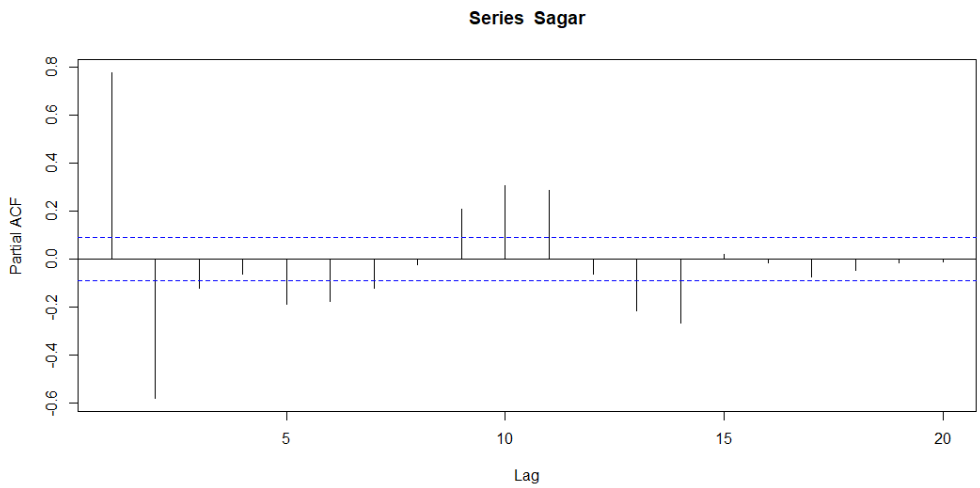

Fig. 5. PACF of Sagar district

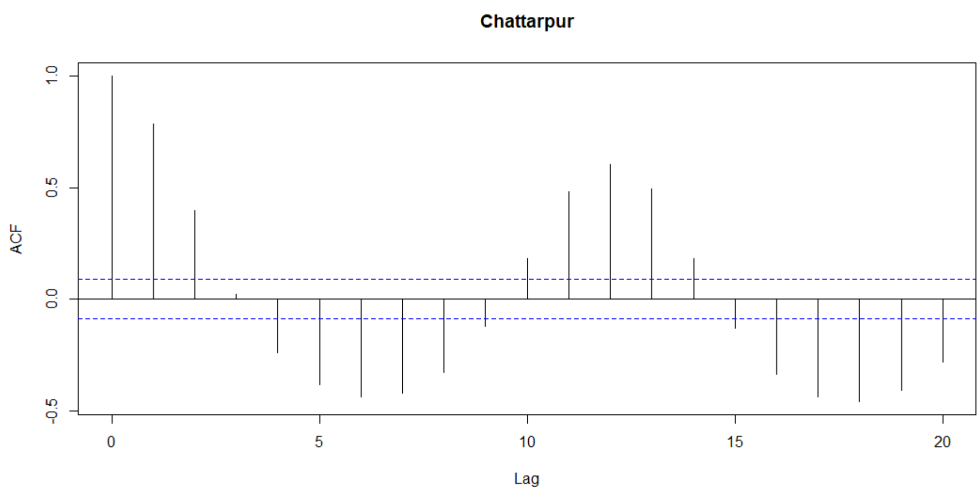

Fig. 6. ACF of Chhattarpur district

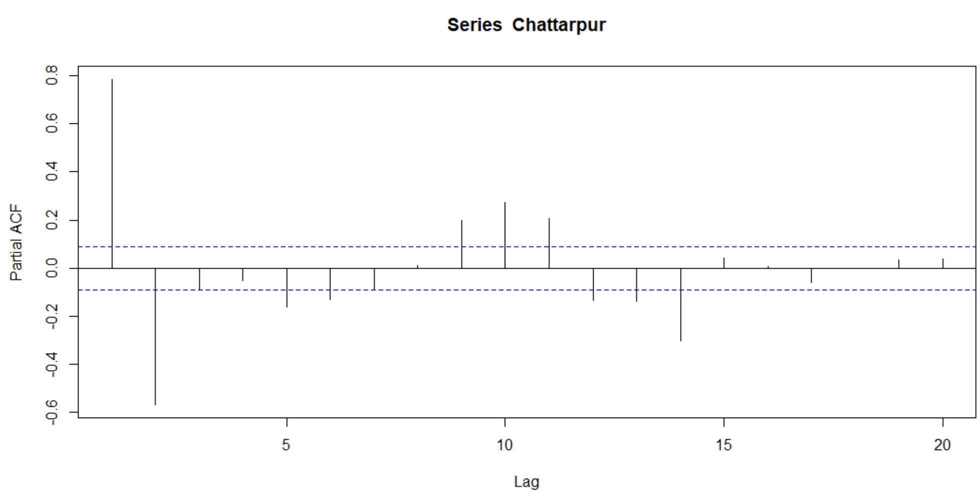

Fig. 7. PACF of Chhattarpur district

tangent, sigmoid, and hyperbolic-tangent sigmoid to develop ANN, ELM, and W-ELM models. Each trial progressively increased the hidden nodes in the middle layer by a factor of three. The log-sigmoid activation function is found to be the best activation function for all three models and data sets considered in the study. For fitting SVM model, radial basis kernel function (RBF) has been used as it is capable of mapping nonlinear input data to high dimensional space.

Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) were used to assess the prediction performance of the proposed W-ELM model. Table 1 compares the performance of each model for the data sets under consideration. Fig. 2 displays a graphical representation of the performance of the various models. The value of RMSE, MAE, and correlation coefficient of W-ELM model for the Sagar district is 0.632, 0.354 and 0.862, respectively, and for Chattarpur district, 0.673, 0.390 and 0.890, respectively. According to the detailed results, W-ELM has a lower RMSE and MAE value and a higher correlation coefficient than all other competing models. Hence, it can be concluded that the proposed W-ELM model outperformed the ELM, ANN, and SVM models on all performance measures comparing the expected and observed effective drought index.The utilization of wavelet transformation on input signals enables the extraction of time-frequency domain information, facilitating the identification of patterns in the stochastic nature of predictor variables. As such, wavelet transformation serves as a valuable auxiliary technique for the analysis of stochastic fluctuations, periodicity, and trends within hydrologic datasets. The W-ELM approach will be useful for predicting other nonlinear hydrological phenomenon.

Table 1. Performance evaluation of the W-ELM, ELM, SVR and the ANN models

Location	Predictive Model	RMSE	MAE
Sagar	W-ELM	0.632	0.354
	ELM	0.679	0.387
	ANN	0.690	0.390
	SVM	0.699	0.435
Chattarpur	W-ELM	0.673	0.390
	ELM	0.725	0.418
	ANN	0.743	0.429
	SVM	0.778	0.457

4. CONCLUSIONS

For the forecasting of drought index, machine learning techniques like, artificial neural networks, support vector machine, etc. have been widely used in the literature. However, these techniques face several challenges, like, need for iterative model parameter tuning, time-consuming, less generalization capability etc. To overcome these limitations, in this study, wavelet based extreme learning machine (W-ELM) model has been proposed for the forecasting of effective drought index. The forecasting performance of the proposed approach has been evaluated with other models for Sagar and Chattarpur district of Bundelkhand regions. It has been found that the proposed approach outperforms the ELM, SVM and ANN models in terms of forecasting accuracy for the considered locations.

Data availability: The data can be available on request to the corresponding author.

ACKNOWLEDGEMENTS

The authors are grateful to the editors and the reviewers for their constructive suggestions/comments that helped to improve the quality of the manuscript.

REFERENCES

Acharya, N., Shrivastava, N.A., Panigrahi, B.K. and Mohanty, U.C. (2013). Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. *Climate Dynamics, 43*(5), 1303–1310. https://doi.org/10.1007/S00382-013-1942-2

Byun, H.R. and Wilhite, D.A. (1999). Objective quantification of drought severity and duration. *Journal of Climate*, *12*(9), 2747–2756. https://doi.org/10.1175/1520-0442(1999)012<2747:OQODSA>2.0.CO;2

Deo, R.C., Byun, H.R., Adamowski, J.F. and Begum, K. (2017). Application of effective drought index for quantification of meteorological drought events: a case study in Australia. *Theoretical and Applied Climatology*, *128*(1–2), 359–379. https://doi.org/10.1007/S00704-015-1706-5/TABLES/7

Deo, R.C., and Şahin, M. (2015). Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. *Atmospheric Research*, *161–162*, 65–81. https://doi.org/10.1016/J.ATMOSRES.2015.03.018

Deo, R.C., Tiwari, M.K., Adamowski, J.F. and Quilty, J.M. (2016). Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stochastic Environmental Research and Risk Assessment, 31*(5), 1211–1240. https://doi.org/10.1007/S00477-016-1265-Z

Dikshit, A., Pradhan, B. and Santosh, M. (2022). Artificial neural networks in drought prediction in the 21st century–A scientometric analysis. *Applied Soft Computing*, *114*, 108080. https://doi.org/10.1016/J.ASOC.2021.108080

Field. (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation — IPCC*. https://www.ipcc.ch/report/managing-the-risks-of-extreme-events-and-disasters-to-advance-climate-change-adaptation/

Huang, G. Bin, Zhu, Q.Y. and Siew, C.K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501. https://doi.org/10.1016/J.NEUCOM.2005.12.126

Jain, V.K., Pandey, R.P., Jain, M.K. and Byun, H.R. (2015). Comparison of drought indices for appraisal of drought characteristics in the Ken River Basin. *Weather and Climate Extremes*, *8*, 1–11. https://doi.org/10.1016/J.WACE.2015.05.002

Kim, D.W., Byun, H.R. and Choi, K.S. (2009). Evaluation, modification, and application of the Effective Drought Index to 200-Year drought climatology of Seoul, Korea. *Journal of Hydrology*, *378*(1–2), 1–12. https://doi.org/10.1016/J.JHYDROL.2009.08.021

Kumar, R.R., Singh, K.N., Mishra,D.C. and Budhlakoti, N. (2020). EDI: Calculation of Effective Drought Index (EDI), GitHub repository. URL: https://github.com/rrk4910/EDI

Malik, A. and Kumar, A. (2020). Meteorological drought prediction using heuristic approaches based on effective drought index: a case study in Uttarakhand. *Arabian Journal of Geosciences*, *13*(6). https://doi.org/10.1007/s12517-020-5239-6

McKee, T.B., Doesken, N.J. and Kleist, J. (1993). The relationship of drought frequency and duration to time scales. *Eighth Conference on Applied Climatology*, 179–184.

Mishra, A.K. and Singh, V.P. (2010). A review of drought concepts. In *Journal of Hydrology,* 391, 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Yaseen, Z.M., Ali, M., Sharafati, A., Al-Ansari, N. and Shahid, S. (2021). Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh. *Scientific Reports, 11*(1), 1–25. https://doi.org/10.1038/s41598-021-82977-9