Research article

A novel stabilized artificial neural network model enhanced by variational mode decomposing

Ali Danandeh Mehr [a,*], Sadra Shadkani [b], Laith Abualigah [c,d,e,f,g,h], Mir Jafar Sadegh Safari [i], Hazem Migdady [j]

[a] Civil Engineering Department, Antalya Bilim University, Antalya, 07190, Turkey
[b] Department of Water Engineering, University of Tabriz, Tabriz, Iran
[c] Computer Science Department, Al Al-Bayt University, Mafraq, 25113, Jordan
[d] MEU Research Unit, Middle East University, Amman, 11831, Jordan
[e] Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan
[f] Jadara Research Center, Jadara University, Irbid, 21110, Jordan
[g] Centre for Research Impact & Outcome, Chitkara University, Punjab, India
[h] Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Tabuk 71491, Saudi Arabia
[i] Department of Civil Engineering, Yasar University, Izmir, Turkey
[j] CSMIS Department, Oman College of Management and Technology, 320, Barka, Oman

ARTICLE INFO

ABSTRACT

Existing artificial neural networks (ANNs) have attempted to efficiently identify underlying patterns in environmental series, but their structure optimization needs a trial-and-error process or an external optimization effort. This makes ANNs time consuming and more complex to be applied in practice. To alleviate these issues, we propose a stabilized ANNs, called SANN. The SANN efficiently optimizes ANN structure via incorporation of an additional numeric parameter into every layer of the ANN. To exemplify the efficacy and efficiency of the proposed approach, we provided two practical case studies involving meteorological drought forecasting at cities of Burdur and Isparta, Türkiye. To enhance SANN forecasting accuracy, we further suggested the hybrid VMD-SANN that integrated variation mode decomposition (VMD) with SANN. To validate the new hybrid model, we compared its results with those obtained from hybrid VMD-ANN and VMD-Radial Base Function (VMD-RBF) models. The results showed superiority of the VMD-SANN to its counterparts. Regarding Nash Sutcliffe Efficiency measure, the VMD-SANN achieves accurate forecasts as high as 0.945 and 0.980 in Burdur and Isparta cities, respectively.

1. Introduction

Drought forecasting contributes to proactive decision-making, risk reduction, and the sustainable development of effective strategies to cope with the challenges associated with water scarcity. It is an essential tool for building resilience in the face of climate variability and change. Drought forecasting helps water resource managers anticipate and plan for potential water shortages. It allows for better management of surface water, groundwater, and allocation of water resources. From an agricultural perspective, farmers can

make in-formed decisions about crop planting, irrigation, and harvesting based on drought forecasts. Timely information helps mitigate the impact of drought on crops, reduces agricultural losses, and contributes to food security. With changing climate patterns, droughts may become more frequent or intense in certain regions [1]. Forecasting helps communities and policymakers adapt to these changes by implementing sustainable practices, developing resilient infrastructure, and planning for long-term water resource management.

Drought forecasting involves the use of various statistical/machine learning (ML) methods or climate/hydrologic models to predict the likelihood, severity, and duration of drought conditions [2]. ML and deep learning techniques, such as artificial neural network (ANNs), support vector machines (SVM), random forest (RF), M5-Tree, and extreme learning machines (ELM) can analyze large datasets/images and identify complex patterns behind data [3–5]. These methods are increasingly used for their ability to capture non-linear relationships in hydrological data. To improve the accuracy and reliability of forecasts, hybrid ML models are suggested in the recent literature [6]. They either combine the strengths of multiple individual models (i.e., ensemble modes) or integrate an ML model with advanced data pre/post processing techniques or fine-tuning optimization algorithms [6–8]. The recent review papers conducted by Sundararajan et al. [9] and Alawsi et al. [7] summarized preferred ML and hybrid ML techniques for drought forecasting, respectively.

Forecasting studies showed that data decomposition techniques played a crucial role in extracting relevant information and patterns from hydrological time series [10–15]. While the earlier drought forecasting models mostly employed Fourier Transform to de-composes drought series in frequency domain [16], Singular Spectrum Analysis to identify trends, oscillations, and noise [17], or Principal Component Analysis (PCA) to decom-pose multivariate drought indices and identify the principal components that capture the most significant variability in drought patterns [18,19], the recent studies proposed advanced decomposition techniques such as Empirical Mode Decomposition (EMD), Wave-let Transform (WT), and Variational Mode Decomposition (VMD) to decompose drought time series into different frequency components, which allow for the analysis of both high and low-frequency patterns, aiding in the identification of drought-related signals at different temporal scales [20–24]. For example, Khan et al. [22] introduced a discrete WT-based hybrid ANN model for meteorological drought forecasting and demonstrated the superiority of the new model to their vanilla versions. Özger et al. [21] compared the influence of WT and EMD on three ML-based models developed for self-calibrated Palmer Drought Severity Index (sc-PDSI) forecasting in cities of Antalya and Adana, Türkiye. The authors revealed that both decomposition techniques significantly improve the standalone ML models' accuracy. In addition, they found that VMD acts superior to EMD for the prediction of sc-PDSI in both study areas. Liu et al. [23] showed that the hybrid VMD-ELM provides more accurate forecasts than those of standalone ANN and ELM as well as the hybrid EMD-SVM models in several cities of China. Danandeh Mehr et al. [24] proposed the hybrid VMD-based genetic programming model for standardized precipitating evaporation index (SPEI) forecasting at ungagged catchments. Although the authors showed the superiority of the hybrid model to its vanilla benchmarks, they highlighted that the performance of VMD is sensitive to its hyper-parameters, such as the number of modes. More recently, Ekmekcioğlu [25] employed VMD to separate historical sc-PDSI time series into different modes and proposed VMD-XGBoost hybrid framework for sc-PDSI forecasting in Denizli, Türkiye. The authors compared the performance of VMD-XGBoost with the hybrid discrete WT-XGBoost and concluded that the former is superior, particularly for long lead time sc-PDSI forecasts.

Overall, our review proved that the hybrid models that integrate decomposition methods with ML approaches frequently yield more accurate forecasts. The choice of a specific decomposition technique depends on the characteristics of the observed drought time series and the forecasting horizon. In previous studies, VMD was found effective in capturing non-linear and non-stationary components in temperature and rainfall data [23], SPEI [24], and sc-PDSI [25]. However, proper tuning of VMD parameters, which is crucial to obtaining meaningful results, has not been explored yet. In addition, as of the current knowledge cut-off date, there has been no investigation into the efficiency of the VMD-based ANN model specifically for forecasting the Standardized Precipitation Index (SPI). This implies that there is a gap in the existing literature regarding the application of VMD in conjunction with ANN for SPI forecasting. Therefore, in this study, we aimed at filling this gap via introducing a new hybrid model integrating VMD with an enhanced neural network regressor. To address the optimum parameter tuning of VMD, the centre frequency method was used in this study for the first time. Furthermore, we introduced an additional numeric parameter, called stabilizer, into every layer of the network that enhances the stability of ANN training. The new hybrid model, named VMD-SANN, was applied to model and forecast SPI series at two meteorological stations in Türkiye. To verify the proposed model's efficiency, we compared its results with those attained by a classic ANN and radial base function (RBF) models trained by the same VMD-based inputs.

2. Material and methods

2.1. Overview of VMD

The VMD [26] is a relatively new decomposition technique that has been considered in a few hydrological studies so far (e.g., Seo et al. [27]; Zuo et al. [28]; Sibtain et al. [29]; Danandeh Mehr et al. [24]). It often overcomes the recursive-based decomposition techniques such as EMD, by demonstrating higher efficiency in removing high frequency noises without diminishing much of the signal amplitude [24,30]. The VMD algorithm concurrently extracts multiple band-limited intrinsic mode functions (IMFs) from an original time series signal, denoted as f(t), where each IMF oscillates around its central frequency. Each mode (uk) is most concentrated around its frequency center (ωk), a parameter determined during the decomposition process. For each drought index, VMD initially eliminates negative frequency components using the Hilbert transform. Subsequently, the frequency spectrum is shifted to the baseband by combining it with an exponential function tuned to its acquired frequency center [26]. The estimation of bandwidth relies on the Gaussian smoothness of the demodulated signal. Subsequently, a constrained optimization problem is defined (equation (1)) using the

squared $L2$-norm of the gradient. The solution to this problem can be obtained through the iterative alternate direction method of multipliers, as outlined in Ref. [26].

$$\underset{u_k, w_k}{\text{Min}} = \left\{ \sum_k \left\| \partial_t \left[\left(\delta\left(t\right) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2 \right\} \tag{1}$$

where k is the scale number that indicates total number of IMFs, u_k is the kth mode, ω is the frequency, δ is the Dirac distribution, the asterisk sign represents the convolution operator, and ∂_t stands for partial derivative respect to time t.

Determining the optimum number of IMFs and the ideal noise reduction thresholds are the most crucial aspects of the proposed hybrid models in this phase. However, there is no unified method to do the task [24]. In the present study, the task was done through adjusting the scale number (k) that in turns controls the number of IMFs. To this end, we employed the center frequency method in which a center frequency is computed for a varying decomposition layer in the predefined range [5 to 10]. To this end, from k = 5, the center, and the value of k is incrementally increased with each calculation. The criterion for determining the final value of k is based on the occurrence of similar center frequencies between the last components of two adjacent k values. When this occurs, the preceding k value is determined as the optimal.

2.2. Overview of ANNs

The ANNs, aka universal estimators, are of the well-known ML techniques with extensive implementations in hydrological modeling due to their ability to model a phenomenon without requiring extensive information on its underlying physical mechanisms. They consist of interconnected neurons arranged in layers. An ANN generally comprises three layers: an input layer, responsible for receiving information; one or several hidden layers, which process and transform the input data; and an output layer, responsible for producing the final output. Numerous neurons transmit multiple inputs to each neuron using weighted networks. The weighted inputs are aggregated to provide the basis for a transformation function, such as a linear, logistic, sigmoid, or hyperbolic tangent, which subsequently generates the result of the neuron [31,32]. In a broad sense, ANNs can be classified into two distinct categories, determined by data transmission and processing direction. These categories are commonly referred to as feed-forward networks and reverse networks. The multilayer perceptron network (MLP) is a fundamental component of feed forward ANN models characterized by its transformational and versatile nature. It comprises an input layer, hidden layers, and a single output layer. The MLP approach, which is used as a benchmark model in this study, can be mathematically represented by equation (2) as follows [33,34]:

$$Y = F\left(\sum_{j=1}^{m} W_{kj}.F\left(\sum_{i=1}^{n} W_{ji}X_i + B_j \right) + B_k \right) \tag{2}$$

where X_i is input variables, W_{ji} denotes weights connecting the hidden layers and input layers and W_{kj} denoted weights connecting the hidden and output. The letters m and n represent the number of neurons in the hidden and input layers, respectively. B_j represents the bias amount of the neurons in the hidden layer, while Bk represents the bias amount of the neurons in the output layer. F represents the activation function, and Y represents the output function.

As expressed by equation (3), a sigmoid activation function was used in the present study. This function squashes its input to a range between 0 and 1, making it useful for models that need to produce probabilities. It is a non-linear function, allowing neural networks to learn complex mappings between inputs and outputs. The sigmoid function has a smooth gradient, which facilitates gradient-based optimization algorithm used in this study during the training of all neural networks.

$$F(x) = \frac{1}{1 + exp\left(-x\right)} \tag{3}$$

2.3. Overview of radial basis function (RBF)

The RBF method is particularly well-suited for solving complex problems characterized by non-linear relationships and high-dimensional data. At its core, the RBF method relies on the concept of radial basis functions, which are mathematical functions defined in terms of their distance from a specific center point. These functions exhibit a radial symmetry, decaying as the distance from the center increases. This property makes them ideal for capturing spatial patterns and relationships in data. In this study, we first construct a set of RBFs by selecting appropriate centers, typically using a clustering algorithm or grid-based approach. These centers represent key locations or data points that are deemed significant for the problem at hand. Each RBF is associated with a specific center and captures the influence of that center on the surrounding data points. The next step involves determining the weights or coefficients associated with each RBF. These weights govern the contribution of each RBF to the overall approximation or interpolation process. We employed a gradient descent algorithm to find the optimal values for these weights. This optimization process aims to minimize the error between the approximated or interpolated values and the actual target values. Once the weights have been determined, the RBF method can be used to predict or estimate values at new, unobserved data points. By evaluating the RBFs at these points and combining their weighted contributions, accurate estimations can be obtained. It is important to note that the RBF method offers flexibility in terms of the choice of the basis functions. Commonly used radial basis functions include Gaussian, multiquadric, and inverse quadratic functions. Following Danandeh Mehr et al. [34], Gaussian kernel was employed in this study. For more information about kernels the

reader is referred to Luo [35] and Kisi [36].

2.4. Stabilized artificial neural networks (SANN)

Training an ANN is to iteratively adjust the values of the parameters W to optimize its goal function $\mathscr{L}$. The purpose of gradient descent algorithms is to maximize the objective function by iteratively updating the weight parameters in the direction opposite to the gradient, denoted as $\frac{\partial \mathscr{L}}{\partial w_t}$ and referred to as gt. The convergence of stochastic gradient descent in convex problems has been analytically demonstrated to be highly influenced by the choice of step size. The use of an effective learning rate, training cycles, momentum, error epsilon, and local random speed schedule has the potential to expedite the convergence process toward a more optimal local solution. There exist several approaches for implementing learning rate scheduling. Exponential or power scheduling systems are often employed in several domains. These strategies include the gradual reduction of the learning rate over time, following specific reducing functions such as $\eta_t = \eta_0(1 + \alpha t)^{-\beta}$ or $\eta_t = \eta_0 \times \exp^{-\beta t}$, where βt represents the learning rate at iteration t. Several methodologies exist that observe the behavior of the gradient and autonomously determine the appropriate learning rate, training cycles, momentum, error epsilon, and local random speed. Another approach, known as the learning rate auto-adjustment technique, determines the learning rate by evaluating the performance of the objective function. In this method, the learning rate is decreased by a certain amount if the aim function deteriorates on either the training or cross-validation set.

The SANN presents the incorporation of an additional numeric parameter into every layer of the ANN model. This technique aims to enhance the stability of the stochastic gradient descent training method. It is implemented by conducting joint training with the existing network parameters. The additional parameters can be understood as a stabilizer for each layer, with their values being adjusted based on the data to either drop or increase. This adjustment is aimed at making progress in decreasing the objective function $\mathscr{L}$. In its most basic configuration, the parameters of an ANN layer are enhanced by including a scalar β as expressed by equation (4).

$$y = \varphi \, (\, \beta \times W_x + b) \tag{4}$$

in the given context, the input vector is denoted as x, while an affine change is defined by the parameters W and b. The resulting output vector is represented by y. The value of β is set to 1 during initialization, resulting in an indiscoverable initial model, regardless of the inclusion of the additional parameter.

During training, the stabilizer parameters β are treated as additional parameters learned using the stochastic gradient descent algorithm. The derivation of the update rule for β is straightforward as shown in equations (5)–(8). In the backward loop, the computation of the gradient about the input vector x in layer i is performed.

$$\frac{\partial L}{\partial x} = \beta W^T \frac{\partial L}{\partial y} \tag{5}$$

The computation of the gradient about the parameter β is performed as:

$$\frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial \beta} = \frac{\partial L}{\partial y}^T W_x \tag{6}$$

The expression for the gradient of parameter β can be formulated as:

$$\frac{\partial L}{\partial \beta} = \frac{1}{\beta}\left(\frac{\partial L}{\partial x}\right)^T x = \frac{1}{\beta} < \frac{\partial L}{\partial x}, x > \tag{7}$$

According to the above information, the variable β is:

$$\beta_{t+1} = \beta_t - \frac{\eta}{\beta} < \frac{\partial L}{\partial x}, x > \tag{8}$$

The variation in β is proportional to how the input layer x is connected to the gradient of the goal function about said input. If the desired function is enhanced by increasing the scale of x, the value of β will grow. If the objective function exhibits improvement as x decreases, then the value of β will fall. If the relative direction between the input and its gradient is stochastic, then the parameter β will converge to a stable state. It is anticipated that this event will occur near convergence. The magnitude of the number of steps for variable w_{ij} is determined by the amount of β at the present iteration stage following equation (9).

$$g_{ij} = \frac{\partial L}{\partial \omega_{ij}} = \beta \frac{\partial L}{\partial y_i} x_j \tag{9}$$

in our experimental procedures, we employ the exponential function *exp* (β) as a substitute for the variable β. Furthermore, we begin the variable β with a value of 0.0. The condition imposes a restriction on the stabilizer's effectiveness, as expressed by equation (10), requiring it to have a positive value and a slower decay as it reaches zero.

$$y = \varphi(\exp(\beta) \times W_x + b) \tag{10}$$

2.5. Lagged mutual information (LMI)

Lagged Mutual Information (LMI) is a measure used to analyze the time series data and assess the relationship between variables at different time lags. It is useful in several applications, such as time series prediction, identifying causal relationships, and detecting patterns or dependencies within time-dependent data [37]. In this study, the LMI, defined by equation (11), considers the temporal delay between the target variable Y and lagged versions of the indicators X (t-k) is calculated X and itself Y (t-k) for various lag values (k).

$$LMI_{X,Y}(k) = \sum_{X,Y} P_{X,Y}(x_t, y_{t+k}) \log_2 \left[\frac{P_{X,Y}(x_t, y_{t+k})}{P_X(x_t) . P_Y(y_t)} \right] \tag{11}$$

2.6. Evaluation metrics

In this study, many indicators of efficiency were employed to evaluate the accuracy of the models and verify the precision of the predictions. These indicators encompassed the Pearson correlation coefficient (PCC, equation (12)), Root Mean Square Error (RMSE, equation (13)), and Nash Sutcliffe Efficiency (NSE, equation (14)). The precision of the model's predictions is enhanced if the values of the CC and NSE approach unity. In contrast, RMSE approaches zero. The formulas representing these parameters are listed below:

$$PCC = \frac{\sum_{i=1}^{n} O_i P_i - \frac{1}{n} \sum_{i=1}^{n} O_i \cdot \sum_{i=1}^{n} P_i}{\left(\sum_{i=1}^{n} O_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} O_i \right)^2 \right) \left(\sum_{i=1}^{n} P_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} P_i \right)^2 \right)} \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2} \tag{13}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (O_i - \overline{O})^2} \tag{14}$$

where the variables Pi and Oi represent the anticipated quantities and observed data for a given time point *i*, respectively. The symbol $\overline{O}$ denotes the mean observed data.

The Taylor diagram was also employed to enhance comprehension of the models. The purpose of this diagram is to present a visual representation that allows for the comparison of several model results within a single chart. It showcases the extent of error, correlation, and standard deviation exhibited by the model outputs with respect to the real values [38]. The azimuth angle depicted in the graph corresponds to the PCC amount, whereas the radial distance from the data point and the radial distance from the coordinate center (0,0) reflect the RMSE and standard deviation amounts, respectively.
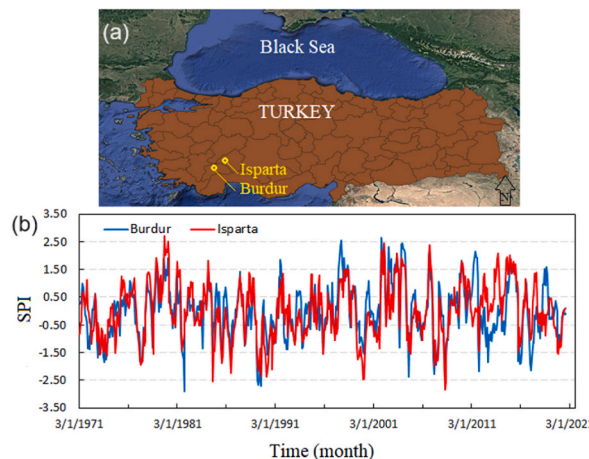

Fig. 1. Map of the study area including (a) Burdur and Isparta and (b) the SPI-6 time series calculated using historical precipitation.

2.7. SHAP (SHapley additive exPlanations) analysis

SHAP is an algorithm used for explaining the output of ML models. It provides a way to understand the contribution of each feature to the model's predictions. The main goal of the SHAP algorithm is to allocate the contribution of each feature to the prediction for a specific instance. The algorithm was first produced by Shapley [39] to determine the assistance of individual players to the result of a team game [40]. Lundberg and Lee [41] presented this idea to calculate the assistance and effect of parameters in formulating results of ML models. The SHAP values evaluate the importance and direction of each input parameter that impacts model results. The SHAP value is estimated by evaluating the average contribution over all possible parameter scenarios, as defined by equation (15).

$$\varnothing_i = \sum_{S \subseteq N(i)} \frac{|S|!(n - |S| - 1)!}{n!} [\vartheta(S \cup \{i\}) - \vartheta(S)] \tag{15}$$

where $\varnothing_i$ is the contribution of parameter i. The letter N is the set of all parameters. The letter n is the number of parameters in N. The letter S is the subset of N that does not include parameter i, and $\vartheta(S)$ is the base amount that defines the predicted result for each parameter in N.

3. Study area and data

The research encompasses the cities of Burdur and Isparta in the Mediterranean region of Turkey (Fig. 1). Renowned for its abundance of surface and groundwater resources, this area stands out as one of Turkey's most resource-rich regions. Over the last fifty years, the study area has witnessed various degrees of drought events—ranging from moderate to severe and extreme—according to reports [42]. This study utilized long-term observed precipitation datasets from 1971 to 2021 at the meteorological stations in Burdur and Isparta (Fig. 1a). These datasets were employed to calculate SPI time series over a 6-month accumulation period (i.e., SPI-6) as the case-study areas have Mediterranean climate commonly with six months dry (summer to fall) and six months wet (winter to spring) periods in a year. SPI and its advancements are used for meteorological drought monitoring and forecasting. The precipitation data were sourced from the Turkish State Meteorological Service, and their quality underwent thorough checks before SPI calculations. Fig. 1b depicted the attained SPI-6 time series and Table 1 summarized their main features at the training (the first 70 %) and the testing periods (the last 30 %).

4. Results

The forecasting process for a month ahead meteorological initiates with the data pre-processing phase. In this stage, the SPI-6 datasets undergo reshaping into an input/target format conducive to supervised learning. The sliding window method is employed for this purpose, generating potential predictors for the corresponding target series. Optimal size of inputs can be also determined via optimization algorithms such as particle swarm optimization [43]. Specifically, 12 preceding values (lags, i.e., SPIt-1, SPIt-2, …, SPIt-12) represent the raw potential features, while the current time step (SPIt) serves as the raw target variable. The LMI criterion was used to detect the most efficient lags as suggested by Danandeh Mehr and Gandomi [44]. Fig. 2 demonstrates the joint probability distribution between the SPI-6 at time t and its lags at both meteorological stations. The figure also shows the LMI attained, indicating a significant decreasing rate by lag three with a local minimum at lag five (see Fig. 2a). Thus, the first three lags were considered the most informative lags at both stations and the remaining ones were removed to avoid impeding in the subsequent calculations that may lead to more complex models.

4.1. Determination of the optimum scale number and IMFs

The data pre-processing process is continued via the decompositions of the most effective inputs (i.e., SPI-6t-3, SPI-6t-2, and SPI-6t-1) and the target (SPI-6t) variable into their associated IMFs and residuals. The center frequencies obtained for the decomposition of the SPI-6t series were tabulated in Table 2. It can be observed that the center frequency is stabilized at 0.42 (0.41) at Burdur (Isparta) when the number of decomposition level is 8 (7). Therefore, the optimum number of decomposition level was determined to be 8 for both experiments in this study. Accordingly, the most effective inputs and the target at each station were decomposed into their eight IMF and residual signals. For example, Fig. 3 demonstrates the associated signals attained for the target SPI-6t at Burdur (Fig. 3a) and

Table 1
Properties of the SPI-6 series acquired at the stations throughout the 1971–2021 timeframe.

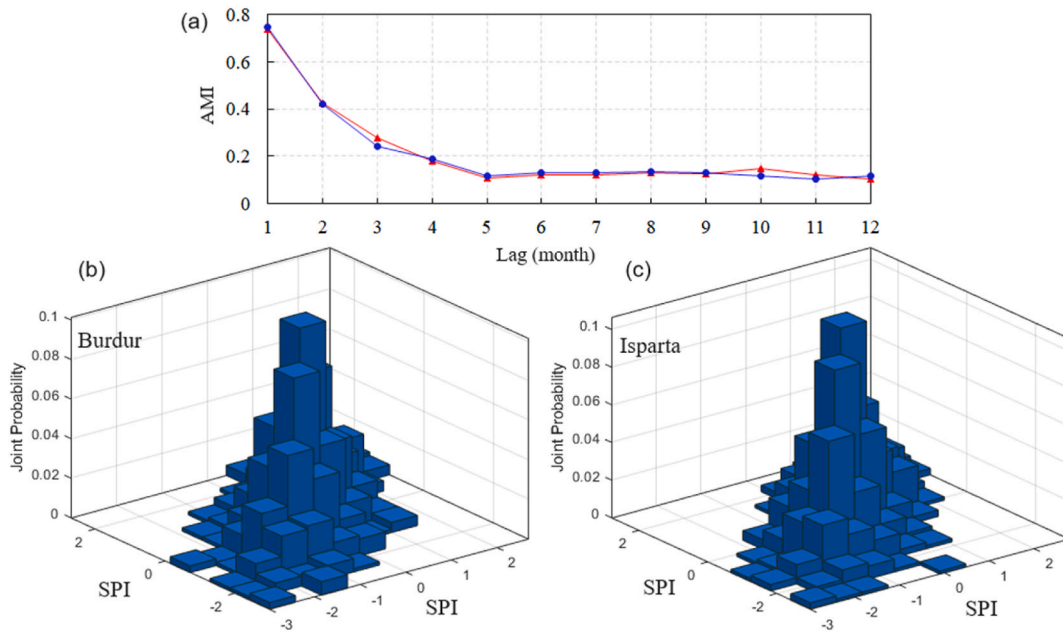| Station	Longitude	Latitude	dataset	Mean	Min.	Max.	S.D.
Burdur	30.29	37.72	Entire	0.00	−2.91	2.65	1.005
			Training	0.02	−2.91	2.65	0.987
			Testing	−0.04	−2.58	2.20	1.046
Isparta	30.57	37.78	Entire	0.00	−2.84	2.73	1.007
			Training	−0.08	−2.53	2.73	1.008
			Testing	0.17	−2.84	2.39	0.985

* Standard Deviation.

Fig. 2. (a) Power of LMI and joint probability between the SPI-6 and its antecedent values at (b) Burdur and (c) Isparta meteorology stations.

Table 2
Center frequencies at various scale number (k) attained for SPI-6 time series at each station.

Burdur						Isparta					
10	9	8	7	6	5	10	9	8	7	6	5
0.4202	0.4216	0.4212	0.4175	0.2527	0.2517	0.4147	0.4146	0.4122	0.4121	0.2526	0.2517
0.2886	0.2920	0.2891	0.2472	0.1240	0.1191	0.2883	0.2823	0.2483	0.2476	0.1202	0.1176
0.2457	0.2459	0.2390	0.1206	0.0910	0.0850	0.2467	0.2354	0.1187	0.1167	0.0892	0.0799
0.2035	0.2004	0.1180	0.0894	0.0690	0.0573	0.2020	0.1169	0.0892	0.0807	0.0709	0.0545
0.1280	0.1154	0.0882	0.0682	0.0504	0.0157	0.1157	0.0880	0.0719	0.0563	0.0519	0.0151
0.1052	0.0872	0.0677	0.0500	0.0153		0.0875	0.0713	0.0543	0.0229	0.0148	
0.0846	0.0672	0.0496	0.0153			0.0710	0.0541	0.0227	0.0064		
0.0665	0.0492	0.0152				0.0540	0.0226	0.0064			
0.0486	0.0152					0.0226	0.0063				
0.0152						0.0063					

Isparta (Fig. 3b) stations.

Table 3 summarized the main statistical features of the obtained IMFs and denoised SPI-6 (hereafter dSPI) series and their dispersion range were demonstrated for Burdur and Isparta in Fig. 4a and b, respectively. The dSPI series at each station was obtained via subtracting the residual signal from associated SPI_t. According to the figure, the dispersion of IMF8 and IM7 is the highest at Burdur and Isparta, respectively.

4.2. Creation of prediction scenarios

In the earlier works decomposed signals possessing the highest PCC values with the target series were used as the most effective inputs to evolve ML models [45,46]. Given an extra challenge in this study, two forecasting scenarios were developed based on correlation analysis (Fig. S1) between dSPI and associated IMFs. In scenario #1, the inputs were confined to those that show positive PCC values, and in Scenario #2, all the 24 generated IMFs were used as inputs. Accordingly, the inputs reduced to 14 and 18 IMFs in scenario 1 at Burdur and Isparta, respectively (See Table S1). Since the proposed scenarios are trained using denoised time series, the evolved solutions need to be modified using the stochastic component (residuals) of the original SPI time series detected via VMD. To this end, an appropriate probability distribution function that perfectly represents the residual pattern at each station must be added to the evolved models.

4.3. Forecasting results

Table 4 presented the performance measures of the evolved models under different scenarios for Burdur and Isparta. In the ANN
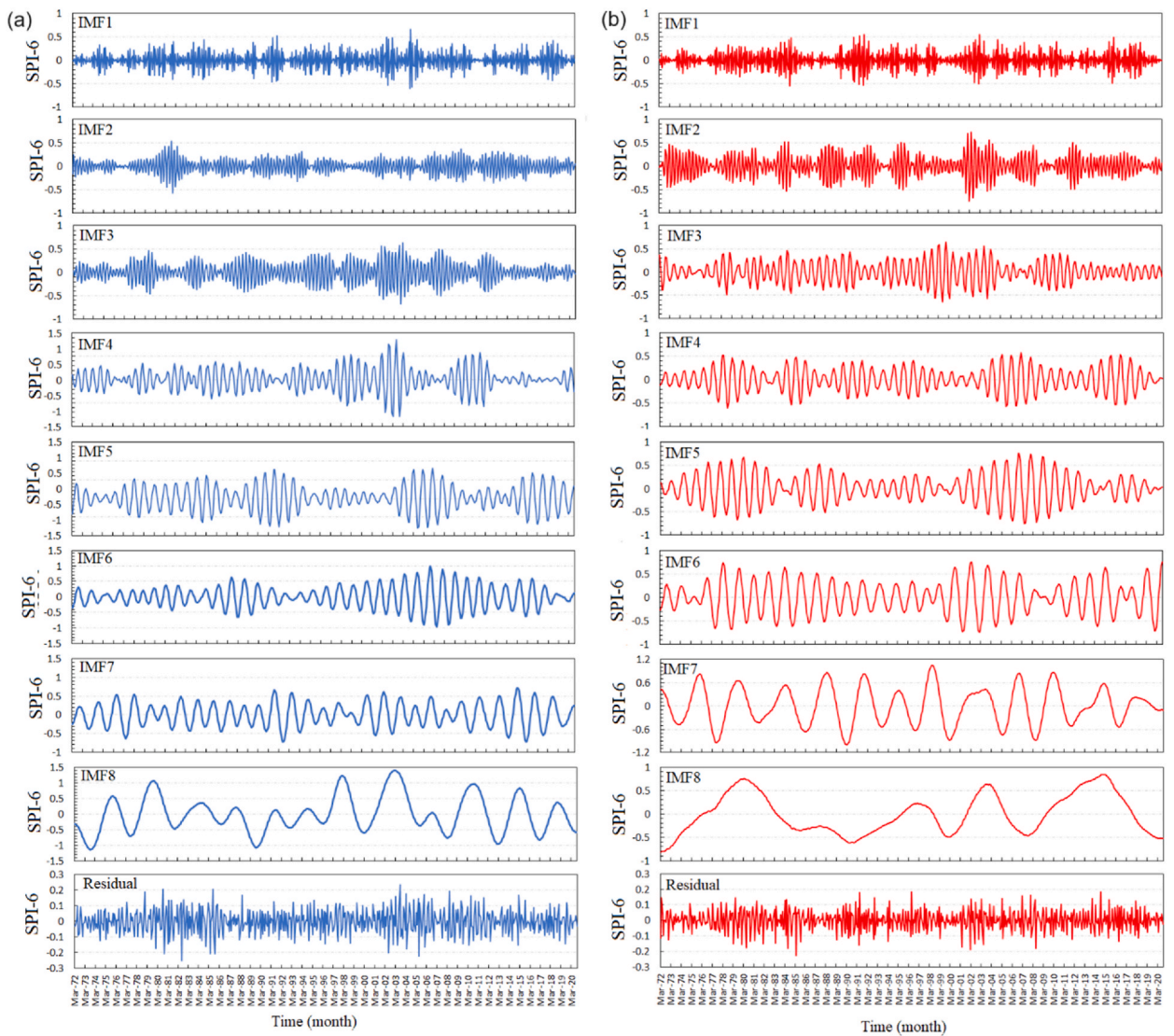
Fig. 3. The IMFs and residuals of the SPI-6 time series at (a) Burdur and (b) Isparta stations obtained by applying VMD.

method, the default parameters are 300 training cycles, 0.01 learning rate with a momentum of 0.95, 0.01 error epsilon, and a local random speed of 2000. After stabilizing the mentioned parameters, values were obtained as 87, 0.02, 0.9, 0.0003, and 0.87, respectively. Also, the optimal amounts of the adaptive learning rate, epochs, RHO, L1, and L2 parameters are 0.009, 26, 1.025, 0.00003, and 0.01, respectively. The results showed that VMD-RBF generally provide good performance, with higher PCC, lower RMSE, and moderate to high NSE values. Improvements in Scenario #2 are seen compared to Scenario #1. The VMD-ANN model resulted in strong performance across all metrics, especially in Scenario #2. It shows lower RMSE and higher NSE compared to the VMD-RBF. The proposed VMD-SANN provided exceptional performance, with very high PCC, low RMSE, and high NSE values. Significant improvements are also observed in Scenario #2. Therefore, the table indicates that the VMD-SANN is the most effective model, demonstrating superior performance in all scenarios and locations. The VMD-ANN also performs well and outperforms VMD-RBF in most aspects.

Fig. 5 compares time series (see Fig. 5a and c) and scatter plots (see Fig. 5b and d) of the evolved models and observed dSPI series in Scenario #2. The figure shows that all models can capture nonlinear and fluctuating features of dSPI series. Also, the high prediction power of the VMD-SANN model for peak and taught values at both stations. According to the figure, VMD-SANN model (purple star points) has a distribution of more points around the bisector line (X = Y), the trend line is closer to the bisector, and the smaller confidence band is the best model.

For further evaluation of the models at Burdur and Isparta stations, Taylor's diagram is presented in Fig. 6. The diagram compares the RMSE, PCC, and standard deviation of the models with observed values at Burdur (Fig. 6a) and Isparta (Fig. 6b) stations. The red arc lines show the error values in the center of which the observed data is placed. The orange points (i.e., VMD-SANN) are the closest to

Table 3

Statistical features of the IMFs and denoised SPI (dSPI) series used for drought forecasting at Burdur and Isparta stations.

| Time	Signal	Burdur				Isparta			
		Min	Max	S.D.[a]	Skewness	Min	Max	S.D.	Skewness
t	dSPI	−2.942	2.689	1.007	0.041	−2.844	2.727	1.008	0.034
t-1	IMF1	−0.582	0.641	0.191	0.000	−0.481	0.471	0.172	−0.002
	IMF2	−0.558	0.520	0.151	−0.029	−0.504	0.488	0.195	−0.013
	IMF3	−0.656	0.622	0.212	0.004	−0.446	0.464	0.157	0.002
	IMF4	−0.779	0.843	0.229	0.018	−0.656	0.626	0.243	−0.011
	IMF5	−0.811	0.785	0.318	−0.030	−0.960	0.946	0.336	−0.001
	IMF6	−0.944	0.974	0.344	0.010	−0.843	0.787	0.369	0.000
	IMF7	−0.753	0.738	0.321	0.002	−1.013	1.085	0.495	0.009
	IMF8	−1.104	1.406	0.583	0.388	−0.638	0.872	0.428	0.388
t-2	IMF1	−0.576	0.635	0.189	0.000	−0.485	0.475	0.173	−0.002
	IMF2	−0.551	0.512	0.151	−0.031	−0.502	0.483	0.191	−0.014
	IMF3	−0.653	0.618	0.211	0.004	−0.465	0.475	0.157	0.001
	IMF4	−0.787	0.856	0.230	0.018	−0.654	0.625	0.245	−0.012
	IMF5	−0.809	0.797	0.325	−0.029	−0.964	0.953	0.336	−0.002
	IMF6	−0.945	0.983	0.344	0.008	−0.838	0.782	0.373	0.000
	IMF7	−0.736	0.732	0.319	0.000	−1.018	1.091	0.494	0.006
	IMF8	−1.103	1.405	0.582	0.389	−0.751	0.877	0.435	0.352
t-3	IMF1	−0.578	0.637	0.190	0.000	−0.493	0.485	0.175	−0.002
	IMF2	−0.548	0.509	0.150	−0.031	0.380	0.408	0.150	−0.026
	IMF3	−0.649	0.615	0.211	0.004	−0.666	0.650	0.207	0.000
	IMF4	−0.789	0.857	0.229	0.018	−0.662	0.639	0.246	−0.010
	IMF5	0.810	0.800	0.327	−0.029	−0.980	0.958	0.339	−0.004
	IMF6	−0.943	0.983	0.342	0.009	−0.840	0.789	0.378	0.000
	IMF7	−0.729	0.728	0.316	0.007	−1.002	1.074	0.483	0.008
	IMF8	−1.123	1.404	0.581	0.384	−0.774	0.857	0.426	0.335

[a] Standard Deviation.
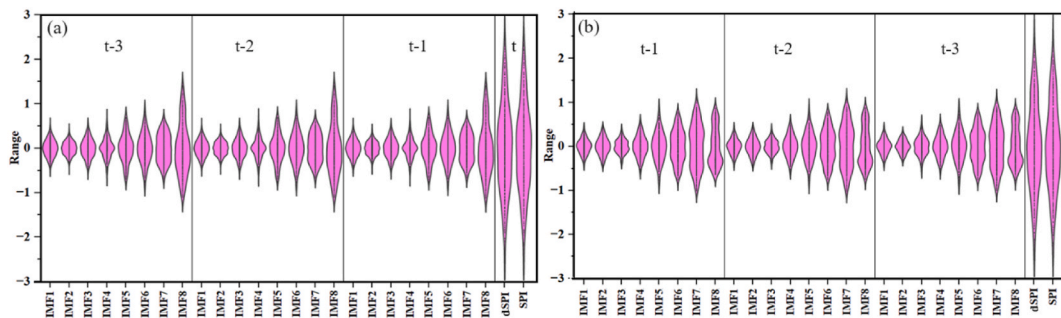
Fig. 4. Violin plots of distribution of data sets at (a) Burdur and (b) Isparta stations.

Table 4

Performance analysis of the hybrid drought forecasting models for dSPI prediction at Burdur station.

Models	Performance measure	Burdur		Isparta	
		Scenario#1	Scenario #2	Scenario #1	Scenario #2
VMD-RBF	PCC	0.784	0.911	0.741	0.849
	RMSE	0.680	0.431	0.702	0.572
	NSE	0.568	0.832	0.551	0.710
VMD-ANN	PCC	0.844	0.923	0.869	0.920
	RMSE	0.572	0.420	0.506	0.408
	NSE	0.703	0.840	0.735	0.831
VMD-SANN	PCC	0.927	0.975	0.945	0.989
	RMSE	0.406	0.252	0.334	0.141
	NSE	0.850	0.945	0.885	0.980

the base point (observational dSPI), which shows the effectiveness of this model among all models.

Feature extraction is a vital task in developing machine learning models [47]. To figure out the sensitivity of input parameters of the VMD-SANN model, visual and average absolute SHAP values were depicted in Figs. 7 and 8. By using SHAP, we can gain insights
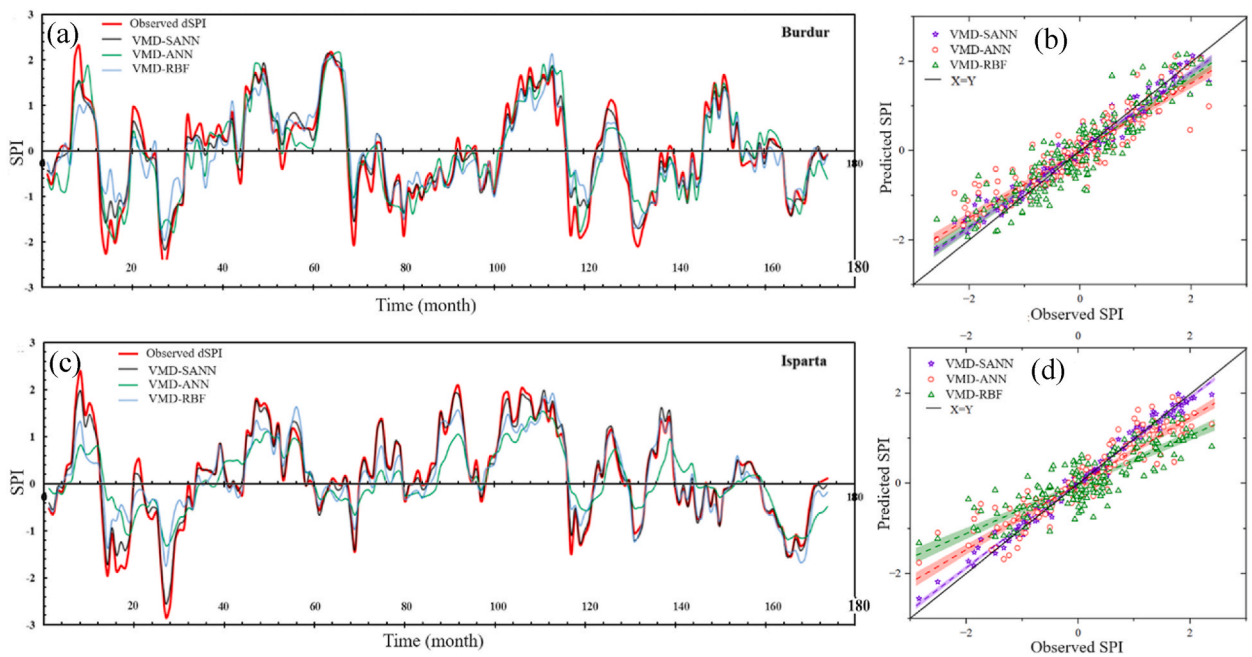
Fig. 5. Time series (a and c) plots of observed and predicted SPI series and (b and d) the associated scatter plots at Burdur and Isparta stations.
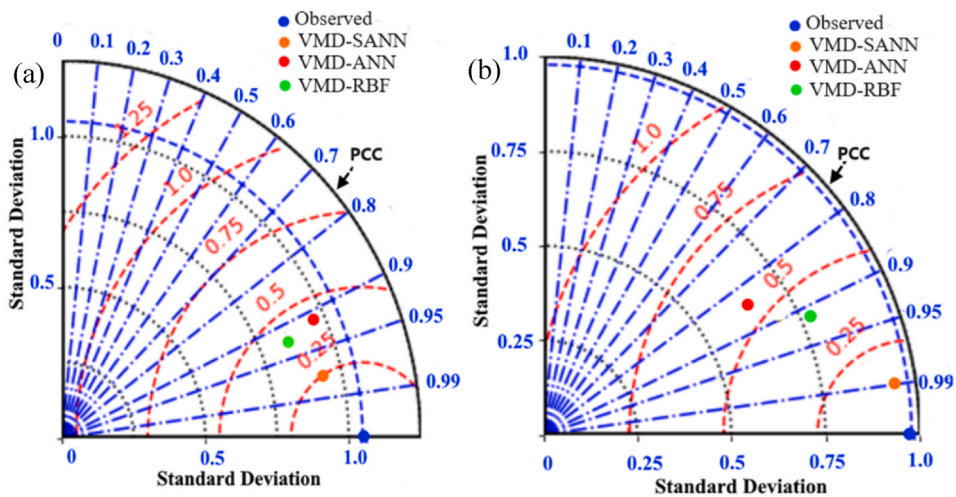

Fig. 6. Taylor diagrams of estimated dSPI at (a) Burdur and (b) Isparta stations.

into the process of the model, which is especially valuable where model interpretability is crucial, and decisions have significant consequences.

Fig. 7 reveals that specific features exhibit a broad spectrum of SHAP values, indicating their diverse influence on the model's predictive outcomes across various data points. For example, the features IMF8C and IMF7C demonstrate extensive variability in SHAP values, suggesting their significant and inconsistent effects on the predictive results at the Burdur (see Fig. 7a) and Isparta (see Fig. 7b) stations, respectively. Conversely, attributes like IMF5B in Burdur and IMF2A in Isparta display SHAP values clustered near zero, denoting a minimal or consistent impact. Fig. 8 presents the mean absolute SHAP values for various characteristics, quantifying their influence on the output of the machine learning model. At Burdur station (Fig. 8a), the feature IMF8C exhibits the most substantial impact with a SHAP value proximate to +0.26, succeeded by IMF7C and IMF6C with values around +0.16 and + 0.12, respectively. Other features such as IMF8B, IMF3B, IMF1C, IMF7B, and IMF8A show lower impact values, ranging from +0.11 to +0.04. Collectively, an additional aggregate of 15 features contributes approximately +0.2 to the model. At Isparta station (Fig. 8b), IMF7C emerges as the most impactful, with a SHAP value around +0.25. This is followed by IMF8C and IMF6C with SHAP values of approximately +0.16 and + 0.15, respectively. Features including IMF5C, IMF8A, IMF7B, IMF4C, IMF8B, and IMF2B range between SHAP values of
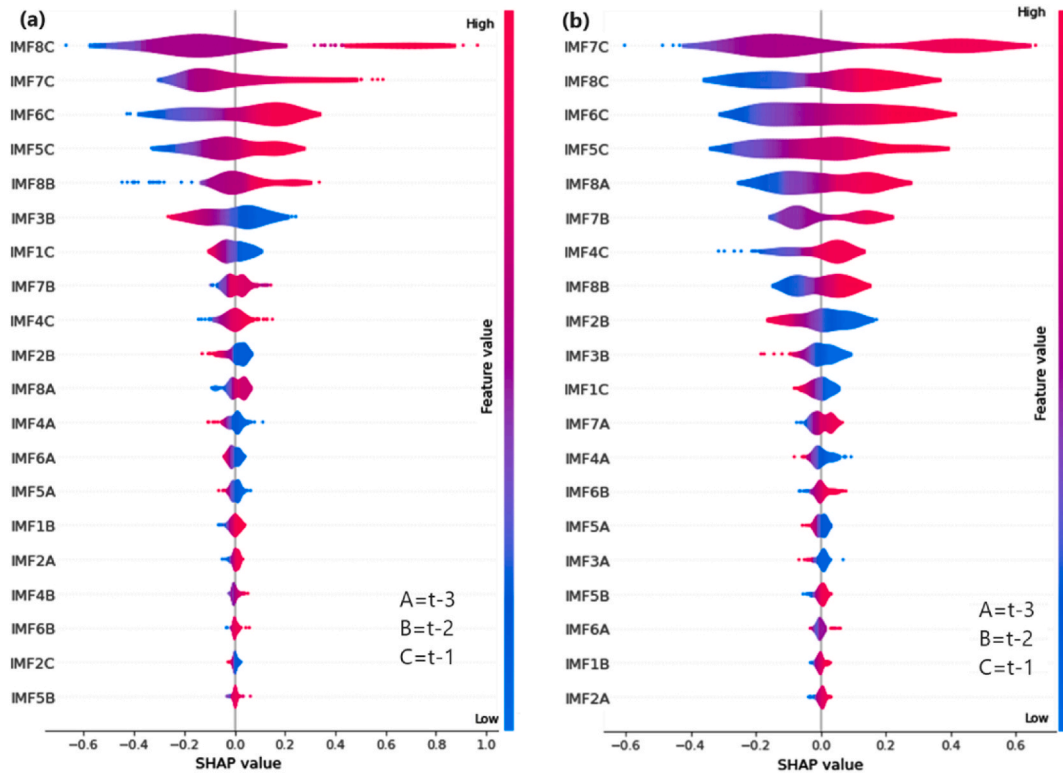
Fig. 7. The individual feature importance of the best dSPI prediction model at (a) Burdur and (b) Isparta stations.
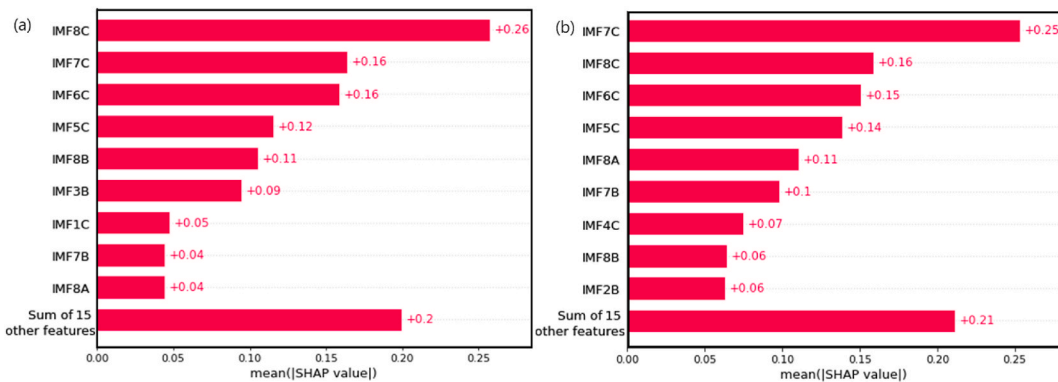

Fig. 8. The SHAP average values for the best SPI-noise prediction model at (a) Burdur and (b) Isparta station.

+0.14 and + 0.06. The cumulative impact of another 15 features in Isparta station is marginally greater than in Burdur, with a combined SHAP value just above +0.21.

5. Discussion

Precise forecasting of drought events holds substantial significance in water resources management. Conventional time series forecasting approaches like ARMA or ARIMA are deemed unreliable to do such a task [48,49]. On the other hand, extensively studied ML methods such as ANNs, ELM, and SVR have demonstrated inadequate accuracy rates for predicting SPI and SPEI, particularly at 3- and 6-month time scales [24,50,51]. Although deep neural networks are suggested recently to enhance accuracy of ML models [52–54], they are criticized for their complex network and less explainability [5,55]. Our results demonstrated that the proposed hybrid model consistently provides superior performance over the benchmark models and some hybrid/nonlinear models suggested for the index in the literature. For example, the VMD-ANN model outperforms the RF, LSTM, ANN, SVR, and Wavelet-based ANN models developed by Tian et al. [56]. It is also superior to the GP, Gene expression programming, and the enhanced variational mode

decomposed-GP models suggested by Danandeh Mehr et al. [24] for drought prediction in Erbil, Iraq. Focusing on the southern Türkiye, our new model provided more accurate predictions than those of the ANN and multi-objective multi-gene GP models respectively developed by Keskin et al. [57] and Reihanifar et al. [58]. The authors reported the RMSE values of their best models equal to 0.439 at Isparta [57] and 0.542 at Burdur [58]. These are significantly higher than those of VMD-SANN introduced in this study.

Our study was limited to the use of ground truth data. Integration of remote sensing data, satellite imagery, and reanalysed meteorological data with VMD-SANN to enhance the spatial and temporal resolution of drought forecasting could be considered the topic for future studies. In addition, current literature shows the frequent occurrence of flash drought which has not received enough attention among hydrological modellers [59]. Therefore, efficiency of VMD-SANN for short-term meteorological drought forecasting could be investigated in future studies.

6. Conclusion

Early awareness of impending drought conditions allows communities and authorities to implement drought preparedness plans. This may include water conservation measures, emergency response planning, and public awareness campaigns. In this article, a new ML model, named VMD-SANN, was introduced and used for meteorological drought forecasting. The model was demonstrated and trained using antecedent SPI-6 values obtained from two meteorology stations in Türkiye. The term VMD in this model implies the separation of the input and output signals into their IMFs and residuals, respectively. Such decomposition enhanced the accuracy of the proposed SANN model. The experimental results showed that the SANN outperforms the counterpart models (i.e., ANN and RBF) owing to the introduced stabilizer. The performance of the VMD-SANN model in the test set was best among the models, and it was followed by VMD-ANN and VMD-RBF. There were significant differences in model performance, with generally higher accuracy in Scenario #2 in which all 24 generated IMFs were used as inputs. The VMD-SANN model had the highest accuracy (NSE >0.9) for both case study areas, indicating that the VMD-SANN model is suitable for one-month ahead meteorological drought forecasting.

It was also found that the complexity of the VMD-SANN is in line with the number of IMFs. Thus, the proposed hybrid model uses LMI and central frequency to optimize the number of inputs. This strategy not only simplifies the model structure but reduces computation time. Moreover, through the comparative assessment between two scenarios at each station, we concluded that use of PCC to reduce the number of inputs may diminish forecasting accuracy.

Data availability statement

The SPI data used in this study will be made available by corresponding author on request.

CRediT authorship contribution statement

Ali Danandeh Mehr: Writing – review & editing, Writing – original draft, Validation, Supervision, Formal analysis, Conceptualization. **Sadra Shadkani:** Writing – original draft, Software, Formal analysis, Conceptualization. **Laith Abualigah:** Writing – review & editing, Software, Formal analysis. **Mir Jafar Sadegh Safari:** Writing – review & editing, Validation. **Hazem Migdady:** Writing – review & editing, Validation, Software.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGP T3.5 in order to improve the language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e34142.

References

[1] G. Naumann, L. Alfieri, K. Wyser, L. Mentaschi, R.A. Betts, H. Carrao, J. Spinoni, J. Vogt, L. Feyen, Global changes in drought conditions under different levels of warming, Geophys. Res. Lett. 45 (2018) 3285–3296, https://doi.org/10.1002/2017GL076521.
[2] A.K. Mishra, V.P. Singh, Drought modeling – a review, J. Hydrol. (Amst.) 403 (2011) 157–175, https://doi.org/10.1016/J.JHYDROL.2011.03.049.
[3] K.A. Tareke, A.G. Awoke, Hydrological drought forecasting and monitoring system development using artificial neural network (ANN) in Ethiopia, Heliyon 9 (2023) e13287, https://doi.org/10.1016/j.heliyon.2023.e13287.

[4] R. Gautam, A. Sinha, H.R. Mahmood, N. Singh, S. Ahmed, N. Rathore, M.S. Raza, Enhancing handwritten alphabet prediction with real-time IoT sensor integration in machine learning for image, Journal of Smart Internet of Things 2022 (1) (2023) 53–64.

[5] A.K. Shakir, Optimal deep learning driven smart sugarcane crop monitoring on remote sensing images, Journal of Smart Internet of Things 2022 (1) (2023) 163–177.

[6] L. Yin, L. Wang, B.D. Keim, K. Konsoer, Z. Yin, M. Liu, W. Zheng, Spatial and wavelet analysis of precipitation and river discharge during operation of the Three Gorges Dam, China, Ecol. Indicat. 154 (2023) 110837, https://doi.org/10.1016/j.ecolind.2023.110837.

[7] M.A. Alawsi, S.L. Zubaidi, N.S.S. Al-Bdairi, N. Al-Ansari, K. Hashim, Drought forecasting: a review and assessment of the hybrid techniques and data pre-processing, Hydrology 9 (2022), https://doi.org/10.3390/HYDROLOGY9070115. Page 115 2022, 9, 115.

[8] M.M. Hameed, S. Fatin, M. Razali, W.H. Melini, W. Mohtar, M. Omar, A. Alsaydalani, Z.M. Yaseen, Deep learning versus hybrid regularized extreme learning machine for multi-month drought forecasting: a comparative study and trend analysis in tropical region, Heliyon 0 (2023) e22942, https://doi.org/10.1016/J.HELIYON.2023.E22942.

[9] K. Sundararajan, L. Garg, K. Srinivasan, A.K. Bashir, J. Kaliappan, G.P. Ganapathy, S.K. Selvaraj, T. Meena, A contemporary review on drought modeling using machine learning approaches, Comput. Model. Eng. Sci. 128 (2021) 447–487, https://doi.org/10.32604/CMES.2021.015528.

[10] S. Guellal, Y. Cherruault, M.J. Pujol, P. Grimalt, Decomposition method applied to hydrology, Kybernetes 29 (2000) 499–504, https://doi.org/10.1108/03684920010322244/FULL/XML.

[11] Y.F. Sang, A review on the applications of wavelet transform in hydrology time series analysis, Atmos. Res. 122 (2013) 8–15, https://doi.org/10.1016/J.ATMOSRES.2012.11.003.

[12] V. Nourani, A. Hosseini Baghanam, J. Adamowski, O. Kisi, Applications of hybrid wavelet–artificial intelligence models in hydrology: a review, J. Hydrol. (Amst.) 514 (2014) 358–377, https://doi.org/10.1016/J.JHYDROL.2014.03.057.

[13] W. Fang, S. Huang, K. Ren, Q. Huang, G. Huang, G. Cheng, K. Li, Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models, J. Hydrol. (Amst.) 568 (2019) 534–550, https://doi.org/10.1016/J.JHYDROL.2018.11.020.

[14] E. Gomaa, B. Zerouali, S. Difi, K.A. El-Nagdy, C.A.G. Santos, Z. Abda, S.S.M. Ghoneim, N. Bailek, R.M. da Silva, J. Rajput, et al., Assessment of hybrid machine learning algorithms using TRMM rainfall data for daily inflow forecasting in Três Marias Reservoir, Eastern Brazil, Heliyon 9 (2023) e18819, https://doi.org/10.1016/j.heliyon.2023.e18819.

[15] S. Adarsh, M. Janga Reddy, Evaluation of trends and predictability of short-term droughts in three meteorological subdivisions of India using multivariate EMD-based hybrid modelling, Hydrol. Process. 33 (1) (2019) 130–143.

[16] G. Incerti, E. Feoli, L. Salvati, A. Brunetti, A. Giovacchini, Drought estimation through a neural network approach, Int. J. Biometeorol. 51 (2022) 253–263, https://doi.org/10.1007/S00484-006-0071-6.

[17] Q.B. Pham, T.C. Yang, C.M. Kuo, H.W. Tseng, P.S. Yu, Coupling singular spectrum analysis with least Square support vector machine to improve accuracy of SPI drought forecasting, Water Resour. Manag. 35 (2021) 847–868, https://doi.org/10.1007/S11269-020-02746-7/FIGURES/15.

[18] R. Arabzadeh, M.M. Kholoosi, J. Bazrafshan, Regional hydrological drought monitoring using principal components analysis, J. Irrigat. Drain. Eng. 142 (2015) 04015029, https://doi.org/10.1061/(ASCE)IR.1943-4774.0000925.

[19] L. Tadić, T. Brleković, K. Potočki, M. Leko-Kos, Application of principal component analysis to drought indicators of three representative Croatian regions, Electronic J. Fac. Civ. Eng. Osijek-e-GFOS 12 (2021) 41–55, https://doi.org/10.13167/2021.22.4.

[20] M. Ali, R.C. Deo, T. Maraseni, N.J. Downs, Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms, J. Hydrol. (Amst.) 576 (2019) 164–184, https://doi.org/10.1016/J.JHYDROL.2019.06.032.

[21] M. Özger, E.E. Başakın, Ö. Ekmekcioğlu, V. Hacısüleyman, Comparison of wavelet and empirical mode decomposition hybrid models in drought prediction, Comput. Electron. Agric. 179 (2020) 105851, https://doi.org/10.1016/J.COMPAG.2020.105851.

[22] M.M.H. Khan, N.S. Muhammad, A. El-Shafie, Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting, J. Hydrol. (Amst.) 590 (2020) 125380, https://doi.org/10.1016/J.JHYDROL.2020.125380.

[23] Y. Liu, L.H. Wang, L.B. Yang, X.M. Liu, Drought prediction based on an improved VMD-OS-QR-ELM model, PLoS One 17 (2022) e0262329, https://doi.org/10.1371/JOURNAL.PONE.0262329.

[24] A. Danandeh Mehr, M. Reihanifar, M.M. Alee, M.A. Vazifehkhah Ghaffari, M.J.S. Safari, B. Mohammadi, VMD-GP: A new evolutionary explicit model for meteorological drought prediction at ungauged catchments, Water 15 (2023) 2686, https://doi.org/10.3390/W15152686. Page 2686 2023, 15.

[25] Ö. Ekmekcio, Drought forecasting using integrated variational mode decomposition and extreme gradient boosting, Water 15 (2023) 3413, https://doi.org/10.3390/W15193413. Page 3413 2023, 15.

[26] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (2014) 531–544, https://doi.org/10.1109/TSP.2013.2288675.

[27] Y. Seo, S. Kim, V.P. Singh, Comparison of different heuristic and decomposition techniques for river stage modeling, Environ. Monit. Assess. 190 (2018) 1–22, https://doi.org/10.1007/S10661-018-6768-2/METRICS.

[28] G. Zuo, J. Luo, N. Wang, Y. Lian, X. He, Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting, J. Hydrol. (Amst.) 585 (2020) 124776, https://doi.org/10.1016/J.JHYDROL.2020.124776.

[29] M. Sibtain, X. Li, S. Saleem, Qurat Ul-Ain, M.S. Asad, T. Tahir, H. Apaydin, A multistage hybrid model ICEEMDAN-SE-VMD-RDPG for a multivariate solar irradiance forecasting, IEEE Access 9 (2021) 37334–37363, https://doi.org/10.1109/ACCESS.2021.3062764.

[30] U. Maji, S. Pal, Empirical Mode Decomposition vs. Variational Mode Decomposition on ECG Signal Processing: A Comparative Study. 2016 International Conference on Advances in Computing, Commun, Inform., ICACCI, 2016 2016, pp. 1129–1134, https://doi.org/10.1109/ICACCI.2016.7732196.

[31] R. Luo, Z. Peng, J. Hu, B.K. Ghosh, Adaptive optimal control of affine nonlinear systems via identifier–critic neural network approximation with relaxed PE conditions, Neural Network. 167 (2023) 588–600, https://doi.org/10.1016/j.neunet.2023.08.044.

[32] R. Fei, Y. Guo, J. Li, B. Hu, L. Yang, An improved BPNN method based on probability density for indoor location, IEICE Trans. Info Syst. 106 (5) (2023) 773–785.

[33] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Network. 2 (1989) 359–366, https://doi.org/10.1016/0893-6080(89)90020-8.

[34] A. Danandeh Mehr, E. Kahya, A. Şahin, M.J. Nazemosadat, Successive-station monthly streamflow prediction using different artificial neural network algorithms, Int. J. Environ. Sci. Technol. 12 (2015) 2191–2200, https://doi.org/10.1007/S13762-014-0613-0/METRICS.

[35] C.T. Luo, KELL: A kernel-embedded local learning for data-intensive modeling, Artif. Intell. Appl. 2 (1) (2024) 38–44, https://doi.org/10.47852/bonviewAIA32021381.

[36] Ö. Kişi, Stream flow forecasting using neuro-wavelet technique, Hydrol. Process. 22 (2008) 4142–4152, https://doi.org/10.1002/HYP.7014.

[37] C. Granger, J. Lin, -L using the mutual information coefficient to identify lags in nonlinear models, J Time Ser Anal 15 (1994) 371–384, https://doi.org/10.1111/J.1467-9892.1994.TB00200.X.

[38] K.E. Taylor, Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. Atmos. 106 (2001) 7183–7192, https://doi.org/10.1029/2000JD900719.

[39] L.S. Shapley, A Value for N-Person Games, 1953, pp. 307–317.

[40] A. Dikshit, B. Pradhan, Explainable AI in drought forecasting, Mach. Learn. Appl. 6 (2021) 100192, https://doi.org/10.1016/J.MLWA.2021.100192.

[41] S.M. Lundberg, P.G. Allen, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[42] D. Soylu Pekpostalci, R. Tur, A. Danandeh Mehr, Spatiotemporal variations in meteorological drought across the Mediterranean region of Turkey, Pure Appl. Geophys. 180 (2023) 3089–3104, https://doi.org/10.1007/S00024-023-03312-Z/METRICS.

[43] Y.A.L. Al-Zuhairy, F.Q. Mohammed, Tilos Island's ideal microgrid size for wind, solar, and batteries, Edison Journal for electrical and electronics engineering 1 (2023) 11–16.

[44] A.D. Mehr, A.H. Gandomi, MSGP-LASSO: an improved multi-stage genetic programming model for streamflow prediction, Inf. Sci. 561 (2021) 181–195, https://doi.org/10.1016/J.INS.2021.02.011.

[45] C. Xiujia, Y. Guanghua, G. Jian, M. Ningning, W. Zihao, Application of WNN-PSO model in drought prediction at crop growth stages: a case study of spring maize in semi-arid regions of northern China, Comput. Electron. Agric. 199 (2022) 107155, https://doi.org/10.1016/J.COMPAG.2022.107155.

[46] Z. Zhou, H. Shi, Q. Fu, Y. Ding, T. Li, Y. Wang, S. Liu, Characteristics of propagation from meteorological drought to hydrological drought in the Pearl river basin, J. Geophys. Res. Atmos. 126 (2021) e2020JD033959, https://doi.org/10.1029/2020JD033959.

[47] Z.S. Ageed, H.M. Yasin, Z.N. Rashid, S.R. Zeebaree, Leveraging high resolution remote sensing images for vehicle classification using sea lion optimization with deep learning model, Journal of Smart Internet of Things 2022 (1) (2023) 97–113.

[48] M.M. Moghimi, A.R. Zarei, M.R. Mahmoudi, Seasonal drought forecasting in arid regions, using different time series models and RDI index, J. Water Clim. Chang. 11 (2020) 633–654, https://doi.org/10.2166/WCC.2019.009.

[49] D. Xu, Q. Zhang, Y. Ding, D. Zhang, Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting, Environ. Sci. Pollut. Res. 29 (2022) 4128–4144, https://doi.org/10.1007/S11356-021-15325-Z/METRICS.

[50] S. Yalçın, M. Eşit, Ö. Çoban, A new deep learning method for meteorological drought estimation based-on standard precipitation evapotranspiration index, Eng. Appl. Artif. Intell. 124 (2023) 106550, https://doi.org/10.1016/J.ENGAPPAI.2023.106550.

[51] J.H. Lee, J.S. Kim, H.W. Jang, J.C. Lee, Drought forecasting using the multi-layer perceptron (MLP) artificial neural network model, J. Korea Water Resour. Assoc. 46 (2013) 1249–1263, https://doi.org/10.3741/JKWRA.2013.46.12.1249.

[52] B. Hrnjica, A.D. Mehr, Energy demand forecasting using deep learning, Smart cities performability, cognition, & security (2020) 71–104.

[53] C. Zhu, X. Li, C. Wang, B. Zhang, B. Li, Deep learning-based coseismic deformation estimation from InSAR interferograms, IEEE Trans. Geosci. Rem. Sens. 62 (2024) 5203610, https://doi.org/10.1109/TGRS.2024.3357190.

[54] E. Foroumandi, V. Nourani, J.J. Huang, H. Moradkhani, Drought monitoring by downscaling GRACE-derived terrestrial water storage anomalies: a deep learning approach, J. Hydrol 616 (2023) 128838.

[55] K. Cheng, N. Wang, M. Li, Interpretability of deep learning: a survey, in: The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, Springer International Publishing, Cham, 2020, August, pp. 475–486.

[56] W. Tian, J. Wu, H. Cui, T. Hu, Drought prediction based on feature-based transfer learning and time series imaging, IEEE Access 9 (2021) 101454–101468, https://doi.org/10.1109/ACCESS.2021.3097353.

[57] M.E. Keskin, Ö. Terzi, E. Dilek Taylan, D. Küükyaman, Meteorological drought analysis using data-driven models for the Lakes District, Turkey, Hydrol. Sci. J. 54 (2009) 1114–1124, https://doi.org/10.1623/HYSJ.54.6.1114.

[58] M. Reihanifar, A. Danandeh Mehr, R. Tur, A.T. Ahmed, L. Abualigah, D. Dąbrowska, A new multi-objective genetic programming model for meteorological drought forecasting, Water 15 (2023) 3602, https://doi.org/10.3390/W15203602. Page 3602 2023, 15.

[59] Q. Zhang, C. Miao, J. Gou, H. Zheng, Spatiotemporal characteristics and forecasting of short-term meteorological drought in China, J. Hydrol. 624 (2023) 129924.