

Drought forecasting by ANN, ANFIS, and SVM and comparison of the models

Maryam Mokhtarzad¹  · Farzad Eskandari² · Nima Jamshidi Vanjani³ · Alireza Arabasadi⁴

Received: 5 February 2017 / Accepted: 16 October 2017 / Published online: 28 October 2017
© Springer-Verlag GmbH Germany 2017

Abstract Drought is a natural disaster that causes significant impact on all parts of environment and cause to reduction of the agricultural products. Other natural phenomena, for instance climate change, earthquake, storm, flood, and landslide, are also commonplace. In recent years, various techniques of artificial intelligence are used for drought prediction. The presented paper describes drought forecasting, which makes use of and compares the artificial neural network (ANN), adaptive neuro-fuzzy interface system (ANFIS), and support vector machine (SVM). The index that is used in this study is Standardized Precipitation Index (SPI). All of data from Bojnourd meteorological station (from January 1984 to December 2012) have been tested for 3-month time scales. The input parameters are as follows: temperature, humidity, and season precipitation, and the output parameter is SPI. This paper shows high accuracy of these models. The results indicated that the SVM model gives more accurate values for forecasting. On the other hand, we use the nonparametric inference to compare the proposal methods, and our results show that SVM model is more accurate than ANN and ANFIS.

Keywords Artificial neural network · Drought · Standardized precipitation index · Adaptive neuro-fuzzy interface system · Support vector machine · Forecasting Kolmogorov–Smirnov test

Introduction

Drought is an environmental disaster that occurs around the world, and it happens when precipitation is less than a specified amount for a period of time. It has widely negative impact on economy, agriculture, water resources, tourism, and ecosystems (Dai 2013; Maca and Pech 2015; Wambua et al. 2016). Some arid or semi-arid regions in Iran are among the most vulnerable to the impacts of weather variation and drought. On this matter, it is necessary to find out the most effective solution for exact drought prediction to reduce its harmful effects on the nature and environment. The aim of this research indicates that artificial intelligence techniques have been widely used for drought prediction. The standard precipitation index (SPI) is used extensively to forecast drought within certain time scales of precipitation, and this index is developed by (McKee et al. 1993). The SPI demonstrates the severity and probability of drought phenomenon, that the more negative values of SPI indicate to severe drought, while positive values illustrate wet condition (Lloyd-Hughes and Saunders 2002; Barker et al. 2016). Other indices like Palmer Drought Severity Index (PDSI) and Standardized Precipitation Evapotranspiration Index (SPEI) have been extended to observe, forecast, and evaluate the severity of drought (Palmer 1965; McKee et al. 1993; Paulo et al. 2012; Moreira et al. 2016).

The new techniques of artificial intelligence such as ANN, ANFIS, and SVM have been recently accepted as

✉ Maryam Mokhtarzad
mokhtarzad.maryam@yahoo.com

¹ Department of Computer Science, Kashani University, Qazvin, Iran
² Department of Statistics, Allameh Tabatabaee University, Tehran, Iran
³ Department of Computer Sciences, Tabriz University, Tabriz, Iran
⁴ Department of Computer Sciences, Shahid Beheshti University, Tehran, Iran

impressive alternative tools for drought forecasting (Shirmohammadi et al. 2013).

A functional method of drought forecasting is presented by artificial neural networks (ANNs). This model is a nonlinear algorithm and is used for solving the system modeling problems (Rezaeianzadeh et al. 2016; Sepahi et al. 2016). ANFIS is a type of artificial intelligence model that is classified as a system theoretical model and is able to create an acceptable simulation of complicated and nonlinear processes even when the data are infrequent (Kadhim 2011; Akbari and Vafakhah 2016). Also, SVM is a new machine learning method often claimed as a best model dealing with intricate classification problems (Ghosh et al. 2014; Suess et al. 2015).

Previous studies executed on these notions were as follows: Maca and Pech (2015) used two different models of artificial neural network and two drought indices SPI and SPEI based on two watersheds of the USA. The training of both neural network models was made by the adaptive version of differential evolution. The results showed that the integrated neural network model was superior to the feed-forward multilayer perceptron. Nguyen et al. (2015) demonstrated the correlations between sea surface temperature anomalies and used both indices SPI and SPEI with ANFIS model at the Cai River Basin in Vietnam. They found that the ANFIS forecasting model with long term was the best forecasting model. Keskin et al. (2009) applied the SPI for meteorological drought analysis at nine stations with different time scales, located around the Lakes District, Turkey. They used ANFIS and Fuzzy Logic models. Comparison of the observed values and the modeling results showed a better agreement with ANFIS models with long term than with fuzzy logic models. Shirmohammadi et al. (2013) used the ANFIS, ANN, Wavelet-ANN, and Wavelet-ANFIS models to forecast drought in the next 3 months on the basis of the SPI for Azerbaijan Province of Iran. The results of research demonstrated that all of the considered modeling methods were able to forecast SPI, but the hybrid Wavelet-ANFIS model demonstrated a better performance. Ustaoglu et al. (2008) used three different intelligent system methods in order to predict minimum, maximum, and dairy average temperature. Hosseinpour et al. (2011) used ANFIS model in order to forecast autumn droughts in eastern Iran with different input variables. The results showed that appropriate inputs were different for different delays and using a specific input could not lead to optimal modeling.

Belayneh and Adamowski (2013) used artificial neural networks (ANN), wavelet neural networks (WNN), and support vector regression (SVR) for forecasting drought conditions in the Awash River Basin of Ethiopia. A 3-month standard precipitation index (SPI) and SPI 12 were forecasted over lead times of 1 and 6 months in each

sub-basin. The performance of all the models was assessed and compared using RMSE and coefficient of determination-R². WNN models revealed superior correlation between observed and predicted SPI compared to simple ANNs and SVR models. The main aim of this research is to investigate techniques of artificial intelligence such as ANN, ANFIS, and SVM to find proper model for drought forecasting in Bojnourd. The rest of the article is organized as follows. In “Materials and methods” section, related literature like drought indices, training method, datasets and model performance measures are briefly described. “Results and discussion” section is dedicated to the results and discussion of methodologies implementation and comparing the performances of these models. Conclusions are given in the last section.

Materials and methods

Study area and data

Bojnourd is the capital city of north Khorasan Province, located 701 km away from Tehran, between the Latitudes 37°28'30" Northern and Longitude 57°20'00 Eastern, is shown in Fig. 1. Bojnourd is located in semi-arid region. The classification of SPI is shown in Table 1.

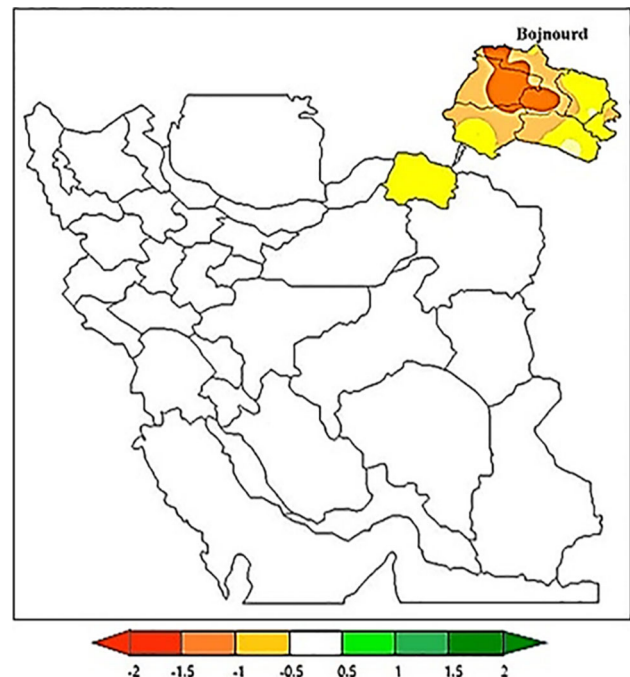


Fig. 1 Map of the Bojnourd containing SPI index

Table 1 Drought classification according to the SPI index (Jinal and Parekh 2013)

Index value	Class	Probability	ΔP
SPI ≥ 2.00	Extreme wet	0.977 - 1.00	0.023
1.50 ≤ SPI ≤ 2.00	Very wet	0.933 - 0.977	0.044
1.00 ≤ SPI ≤ 1.50	Moderate wet	0.841 - 0.933	0.092
-1.00 ≤ SPI ≤ 1.00	Near normal	0.159 - 0.841	0.682
-1.50 ≤ SPI ≤ -1.00	Moderate drought	0.066 - 0.159	0.092
-2.00 ≤ SPI ≤ -1.50	Severe drought	0.023 - 0.066	0.044
SPI < -2.00	Extreme drought	0.00 - 0.023	0.023

Artificial Neural Network (ANN)

An artificial neural network (ANN) is a computational model for information processing that is inspired of the human brain (Maier et al. 2010; Wambua et al. 2016). In this study, among the approaches of ANN, multilayer perceptron neural network is applied. Multilayer perceptron (MLP) can solve the math problem that needs nonlinear equations by defining proper weights (Scarselli and Tsoi 1998). The typical MLP consists of at least three layers. The first layer is called input layer, the last layer is called output layer, and the remaining layers are called hidden layers (Zhang et al. 2003). The structure of MLP is shown in Fig. 2.

Log sigmoid function is used as activation function between input layer and hidden layer, and linear activation function is used between hidden layer and output layer (Adam et al. 2016). These functions are given below:

$$Y = f(x) = 1 / (1 + e^{-x}) \tag{1}$$

$$Y = f(x) = x \tag{2}$$

Among the different training methods, Levenberg–Marquardt (LM) algorithm is one of the neural network training algorithms that is used to train the network with

the highest efficiency. The proposed method is the fastest method and provides a numerical solution to obtain mean square errors (Kayri 2016)

In this part, firstly, the input variables are divided into 2 parts, 85% of the dataset is used for training phase and 15% of the dataset is used for testing phase.

The prediction models require different metrics to measure the accuracy of the models. In this stage, we use the statistical parameters such as the correlation coefficient (R) and root square mean error (RSME) to measure the difference between estimated and observed values (Han and Kamber 2006; Gonzalez-Sanchez et al. 2014; Arabasadi et al. 2017).

$$R = \sqrt{1 - \frac{\sum_{i=0}^n (y_0 - y_p)^2}{\sum_{i=0}^n \left(y_0 - \frac{\sum y_o}{n} \right)^2}} \tag{3}$$

$$RSME = \sqrt{\frac{\sum_{i=0}^n (y_0 - y_p)^2}{N}} \tag{4}$$

That y_0 is referred to the observed data and y_p is referred to estimated data and N is the number of data. The best output is produced when the amount of RSME approaches to 0 and the amount of regression approaches to 1. The steps of artificial neural network are demonstrated as follows (Devi et al. 2012):

- Step 1: Preparation of the training and testing dataset.
- Step 2: Decide the number of nodes, and as initialization, set all weights and threshold value of the network to random number.
- Step 3: For every neuron in every layer $j = 1, 2, \dots, M$, from input to output layer, find the output from the neuron:

$$Y_{j,i} = f \left(\sum_{k=1}^N Y_{(j-1)k} W_{jik} \right) \text{ where } f(x) = \frac{1}{1 + \exp(-x)} \tag{5}$$

Step 4: Calculate error value:

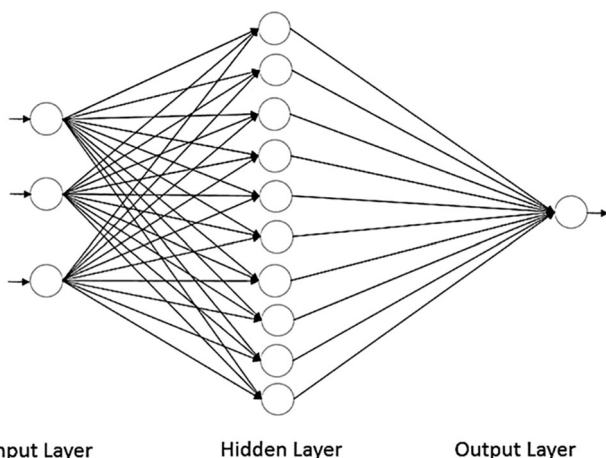


Fig. 2 Structure of multilayer perceptron

$$E(w) = \frac{1}{2} \sum_{d \in D} \sum_{k \in \text{outputs}} (t_{kd} - o_{kd})^2 \quad (6)$$

Step 5: For each network output unit k , calculate its error term:

$$\delta_k = y_k(1 - y_k)(t_k - y_k) \quad (7)$$

For each hidden unit h , calculate its error term:

$$\delta_h = \delta_h(1 - \delta_h) \sum_{k \in \text{outputs}} w_{kh} \delta_k \quad (8)$$

Step 6: Update each network weight w_{ji} :

$$w_{ji} = w_{ji} + \Delta w_{ji} \quad \text{where } \Delta w_{ji} = \eta \delta x_{ji} \quad (9)$$

Step 7: Update bias θ_j in network

$$\Delta \theta_j = (l) \text{Err}_j \quad (10)$$

$$\theta_j = \theta_j + \Delta \theta_j \quad (11)$$

Step 8: If termination condition is met then stop, else go to step 3 (Devi et al. 2012).

Fuzzy inference system (FIS)

The fuzzy logic theory proposed by Lotfi Zadeh aimed to solve the problems and ambiguous features that do not have precise mathematical solutions (Ramlan et al. 2016). Neural networks and fuzzy systems have some characteristic in common. They can solve the problems for instance pattern recognition, time series forecasting, or diagnostics if there is no mathematical solution. The comparison of the two methods is shown in Table 2 (Kruse 2008).

Adaptive neuro-fuzzy inference system (ANFIS)

The ANFIS model (Demyanova et al. 2017) has an approximating ability in real continuous function on a set to any degree of accuracy (Jang et al. 1997). The structure of neural-fuzzy network is organized by a combination of neural networks and fuzzy systems. This structure employs the capability of fuzzy systems which increases the power and trainable features of the neural networks and the

inference precision in uncertain conditions (Alipour et al. 2014).

A basic Sugeno inference system generates an output function f from input variables x and y by using Gaussian membership function (Patel and Parekh 2014; Demyanova et al. 2017). Assume the Sugeno-type of ANFIS model contains two fuzzy IF–THEN rules as follows (Patel and Parekh 2014):

$$\text{Rule 1 : If } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } B_1, \text{ then } f_1 \\ = a_1 x_1 + b_1 x_2 + c_1 \quad (12)$$

$$\text{Rule 2 : If } x_1 \text{ is } A_2 \text{ and } x_2 \text{ is } B_2, \text{ then } f_2 \\ = a_2 x_1 + b_2 x_2 + c_2 \quad (13)$$

where a_1, a_2 and b_1, b_2 are the membership of input variable x and y and the parameters of the output function are f_1 and f_2 . The ANFIS structure is composed of five layers. In the first layer or input layer, the amount of allocation of each input to different fuzzy areas is defined by the user. In the second layer, the weight of rules can be acquired by multiplying the input values in each node. In the third layer, the relative rule weights are computed. In layer four, each node computes the contribution of the rule to the entire output (Mohammadi et al. 2014). The last layer is the output layer of the network which aims to minimize the discrepancy between the acquired output and the actual output. The goal of training adaptive networks is to approximate unknown functions that are obtained from training data and meet accurate values. Suitable ANFIS structure is specified according to the input data, type of input and output membership functions, the number of functions and the IF–THEN rules (Alipour et al. 2014).

The ANFIS method is used via MATLAB toolbox (R2014b). The steps and the pseudocode of ANFIS are demonstrated in Fig. 3.

Support vector machine (SVM)

Support vector machine (SVM) is a supervised learning model that is presented by Vapnik (1999), and this model can be used for various processes like natural language processing, diagnostic, voice recognition, etc. The advantage of this method is more convenient than the other models in training phase and also has high efficiency. This algorithm designates the best separating line to classify the data with more safety margin. In this method, vectors are elected as a selection criterion, and these vectors are applied the best boundaries and categories for data. These vectors are termed as Support Vector (Vapnik and Chervonenkis 1991; Sujay Raghavendra and Deka 2014; Vieira et al. 2017).

According to Fig. 4, two parallel hyper planes on either sides of “maximum margin hyper plan” are created to

Table 2 Comparison of neural control and fuzzy system (Kruse 2008)

Neural networks	Fuzzy systems
No mathematical model necessary	No mathematical model necessary
Learning from beginning	A priori knowledge essential
Several learning algorithms	Not capable to learn
Black-box behavior	Simple interpretation and implementation

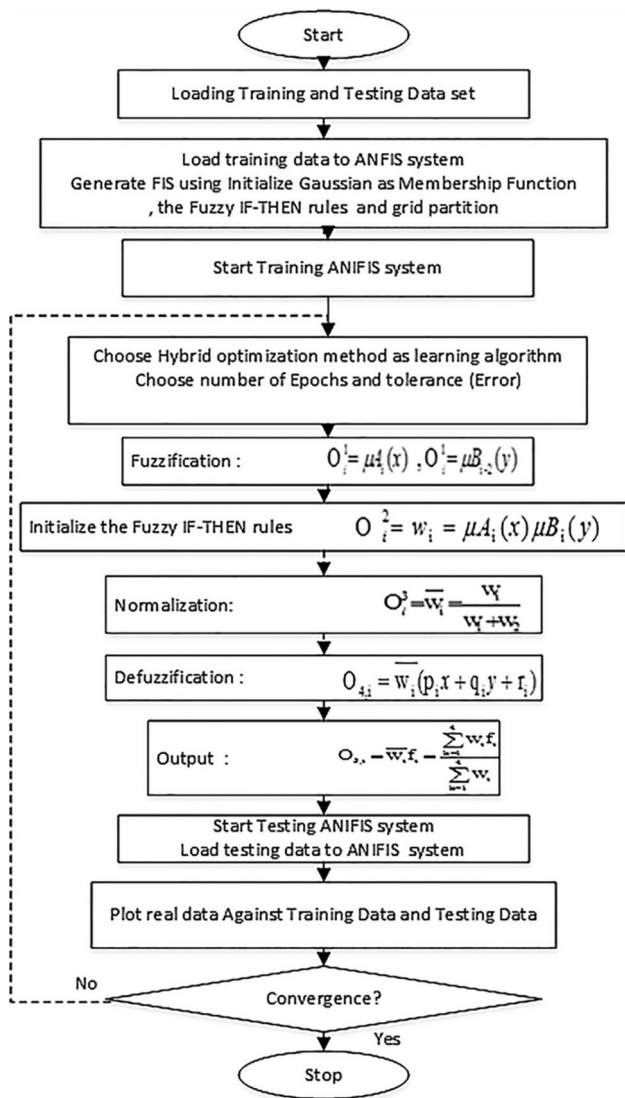
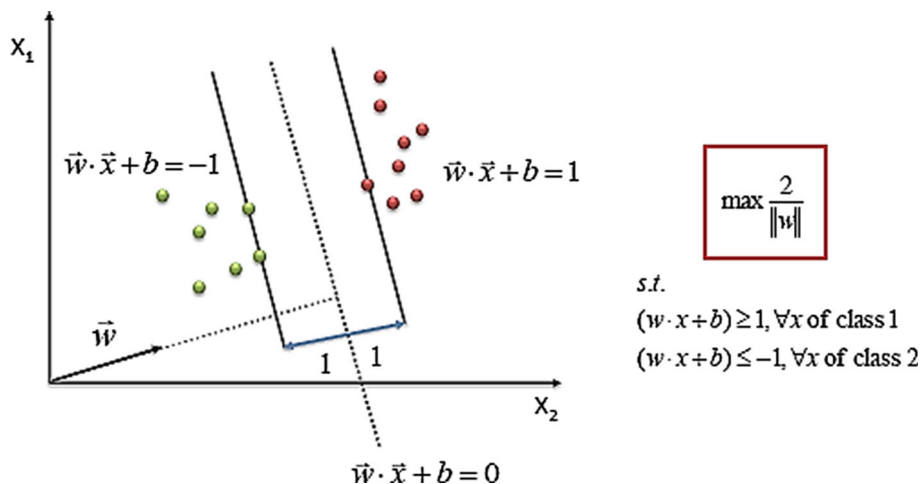


Fig. 3 Flowchart of ANFIS model

Fig. 4 Hyper plane with a maximum margin along with the separation boundaries for two classifications of data. The samples that are located on the borders are support vectors



separate the data which belong to each class (Chihaoui et al. 2016). Maximum margin hyper plane is hyper plane that maximizes the spacing between two parallel hyper planes. It is assumed that the classification error will reduce whatever boundary separating or the distance between the two parallel hyper planes increases (Vapnik 1999).

Overall, the hyper plane with linear decision boundary (Demyanova et al. 2017) can be defined as follows:

$$w \cdot x + b = 0 \tag{14}$$

In Eqs. (14, 15), x is a point on the decision boundary and w is an n -dimensional weight vector orthogonal to the hyper plane and b is called the bias.

If x_i is a support vector then $w \cdot x_i + b = 1$ (15)

Optimal decision boundary is a boundary that has the maximum margin. Optimal decision boundary is calculated by detection of the best hyper plane that maximizes the margin between two classes and minimizes the magnitude of the weights, shown as follows:

$$\max_{w,b} \min_{i=1..L} \left[y_i \frac{(wx_i + b)}{|w|} \right] \tag{16}$$

According to Eq. (15) and performing a series of mathematical operations, above equation is converted to below equation:

$$\min_{w,b} \frac{1}{2w \cdot b} |w|^2, \quad y_i (wx_i + b) - 1 \geq 0 \quad i = 1, \dots, L \tag{17}$$

In order to solve the optimization problem by using the Lagrange multiplier λ_i , this problem can be converted to below optimization equation (Sujay Raghavendra and Deka 2014; Chao and Horng 2015; Wang et al. 2015). This algorithm is used when the classes are separable:

$$\max_{\lambda \dots \lambda_i} \left[\sum_{i=1}^L \lambda_i - \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L \lambda_i y_i (x_j x_j) y_j \lambda_j \right] \quad (18)$$

$$\lambda_i \geq 0 \quad i = 1, \dots, L \quad \sum_{i=1}^L \lambda_i y_i = 0$$

where λ_i is the Lagrangian multiplier, Eq. (18) provides linear boundary between two classes that are completely separated, but in non-separable cases, the error increases when the classes are separating by linear decision boundary which have overlapping (Demyanova et al. 2017). Consequently, the Karush–Kuhn–Tucker (KKT) conditions (Chao and Horng 2015) are used to solve the optimization problems that are stated as follows Eqs. (19, 20):

$$w = \sum_{i=1}^L \sum_{j=1}^L \lambda_i y_i x_j \quad (19)$$

$$\lambda_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1) = 0 \quad i = 1, \dots, L$$

$$\max_{\lambda \dots \lambda_i} \left[\sum_{i=1}^L \lambda_i - \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L \lambda_i y_i (\varphi(x_i) \varphi(x_j)) y_j \lambda_j \right] \quad (20)$$

$$C \gg \lambda_i \gg 0$$

$$\sum_{i=1}^L y_i \lambda_i = 0$$

In Eq. (20), the Lagrange multipliers for nonlinear separable data are limited to $0 \ll \lambda_i \ll C$ (Kumar 2016).

$$K(x_i \cdot x_j) = \varphi(x_i) \varphi(x_j) \quad (21)$$

By Eq. (21), x is mapped to a high-dimensional space $\varphi(x_i)$. Computing in non-separable or high dimensions case is fulfilled by using the kernel functions. The kernel functions like linear kernel function, polynomial kernel function, radial basis kernel function (RBF), and Sigmoid kernel function are used to calculate the inner products in the high-dimensional feature spaces (Hsu et al. 2013), that are shown as follows:

- Linear kernel

$$K(x_i \cdot x_j) = \varphi(x_i) \varphi(x_j) \quad (22)$$

- Polynomial kernel

$$K(x, y) = (xy + 1)^p \quad (23)$$

- Radial basis kernel (Gaussian)

$$K(x, y) = e^{-\|x-y\|^2/2\sigma^2} \quad (24)$$

- Sigmoid

$$K(x, y) = \tan h(kxy - \delta) \quad (25)$$

RBF kernel (radial basis kernel function) is selected in this study because it is appropriate for large or small data and various dimensions (Jiao et al. 2016).

Kolmogorov–Smirnov hypothesis test

This article uses analytical procedure and statistical hypothesis test in order to ensure the acceptable solution. Nonparametric statistical analysis methods like Chi square and Kolmogorov–Smirnov test are useful in this case. The two-sample Kolmogorov–Smirnov test is a nonparametric test that compares the cumulative distributions of two datasets and sees if they are meaningfully different. Kolmogorov–S–mirnov was named in honor of two Russian statisticians: A.n. Kolmogorov and N.v. Smirnov (Sahoo 2013).

$$D_s = \max |F(x) - F(x_0)| \quad (26)$$

$F(x)$ refers to the empirical distribution based on predicted data and $F(x_0)$ refers to the empirical distribution based on observed data. D_s is specified as the greatest difference between two cumulative distribution functions (CDF) (Hassani and Sirimal-Silva 2015). Hence, the two-sample K–S test hypothesis almost is illustrated as follows:

H₀ Demonstrates the rejection of the null hypothesis at the significance level.

H₁ Demonstrates a failure to reject the null hypothesis at the significance level.

In K–S test2, the decision to reject the null hypothesis is based on comparing the p value: If p value is > 0.05 , the null hypothesis is accepted, otherwise if p value is < 0.05 , the null hypothesis is rejected at a confidence level 0.95.

Results and discussion

In ANN approach, correlation coefficient in training phase is shown in Fig. 5. The most common way to determine the number of neurons in the hidden layer is the trial and error method.

This stage is part of the networks training and development phase. In general, the number of neurons in the hidden layer should be modified to produce a satisfactory answer. According to Table 3, the best structure of neural network is number four with 10 neurons in hidden layer.

In ANFIS approach, we applied ANFIS method to our datasets. The criterion chosen for the ANFIS model is shown in Table 4 is as follows:

- Membership function type
- Epoch size
- Data size

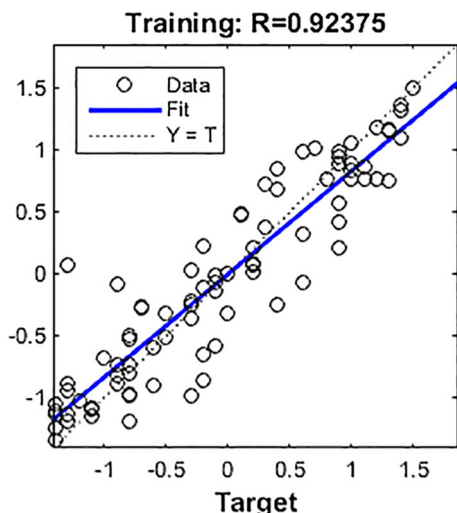


Fig. 5 Correlation coefficient output

- Learning algorithm
- Output type

Finding proper membership functions (MFs) to minimize the output error measure and maximize performance is essential. The Gaussian is used as a membership function in this research (Folorunsho et al. 2012). The number of training epoch in neuro-fuzzy system was set on 3000 epochs which show that error decreases by increasing the number of epochs. Fuzzy system is defined by linguistic variables.

Table 3 Output of neural networks after training network

Index	Num of nodes in hidden layer	RMSE	Regression
1	7	0.23238	0.87475
2	8	0.0960	0.91855
3	9	0.14829	0.90818
4	10	0.08125	0.9237
5	11	0.2201	0.90027

Table 4 ANFIS base modeling criterion

S/N	Parameters of ANFIS model	Values
1	Membership function type	Gaussian membership function
2	Epoch size	3000
3	Data size	17 data per 3 months
4	Learning algorithm	Hybrid learning algorithm
5	Output function	Linear

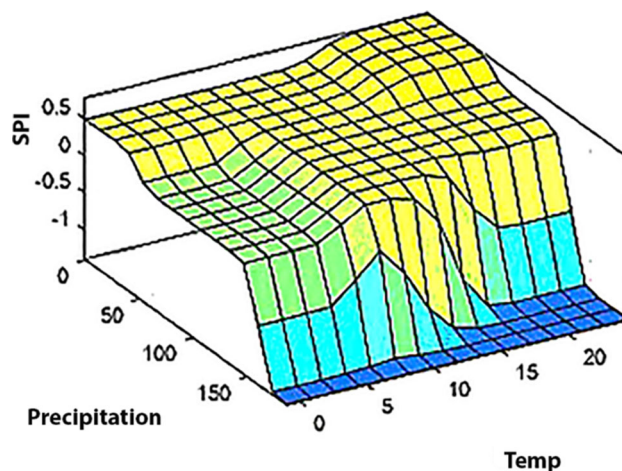


Fig. 6 Output surface for SPI output versus temperature and rain inputs

Table 5 Correlation coefficient for the predicted results

The models	Num. of data	R	K-S test
ANN	15% of data	0.9237	0.227
ANFIS	15% of data	0.9926	0.906
SVM	15% of data	0.9974	0.9303

The range for each input can be divided into three parts or subsets and converted to linguistic variables. These linguistic variables are defined as: min, average, and max. We insert IF-THEN rules to fuzzy system. Like this rule:

$$\text{IF } (Temp \text{ is average}) \text{ AND } (Rain \text{ is max}) \text{ AND } (Humid \text{ is max}) \text{ THEN } (SPI \text{ is min}) \tag{27}$$

Temp referred to temperature and Rain referred to precipitation. After adding the rules, the FIS surface presents relationship between certain inputs and SPI output.

In Fig. 6, we can see that there is a reasonable pattern such as when the temperature is extremely high, the SPI would be high. Also, it shows that the SPI tends to decrease when precipitation is high. The performance of these models for drought forecasting according to regression and KS test are presented in Table 5. Three methods of ANN, ANFIS, and SVM are applied to forecasting drought of Bojnourd in Fig. 7. The result shows that the predicted data by SVM are approaches to the actual values of meteorological data.

In SVM approach, like ANN model, 85% of data are used for model training and 15% are used to test the performance of this model. The result is shown in Table 5. According to this result, it is recommended to use SVM model to obtain suitable approximation of real dataset. Also, comparison the performance of the predicted data

Fig. 7 Comparing the calculated values using three methods with actual values

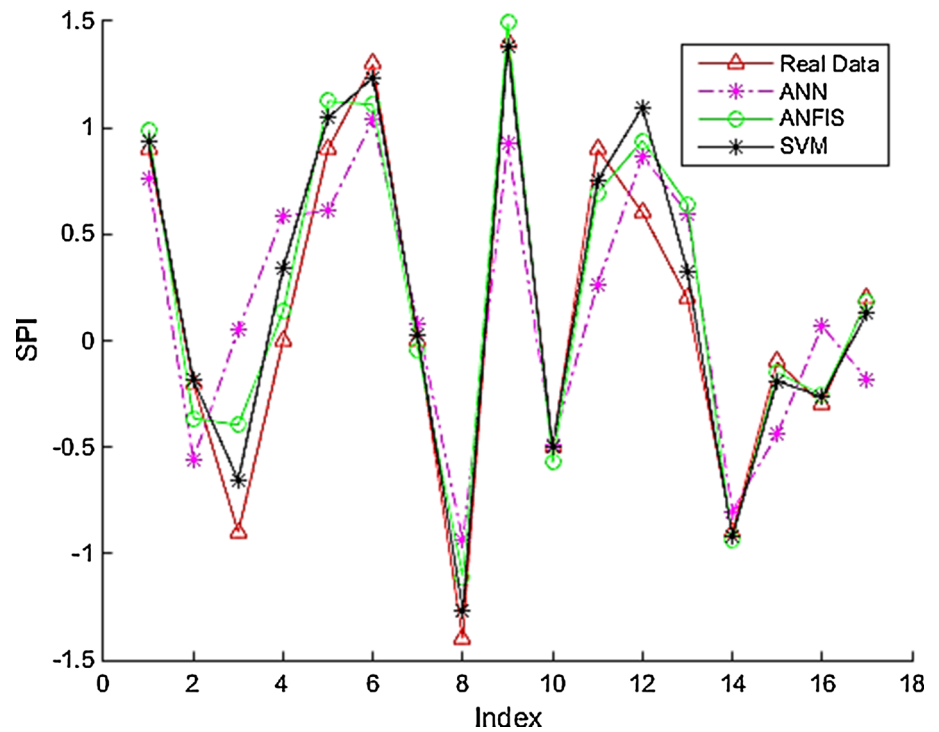
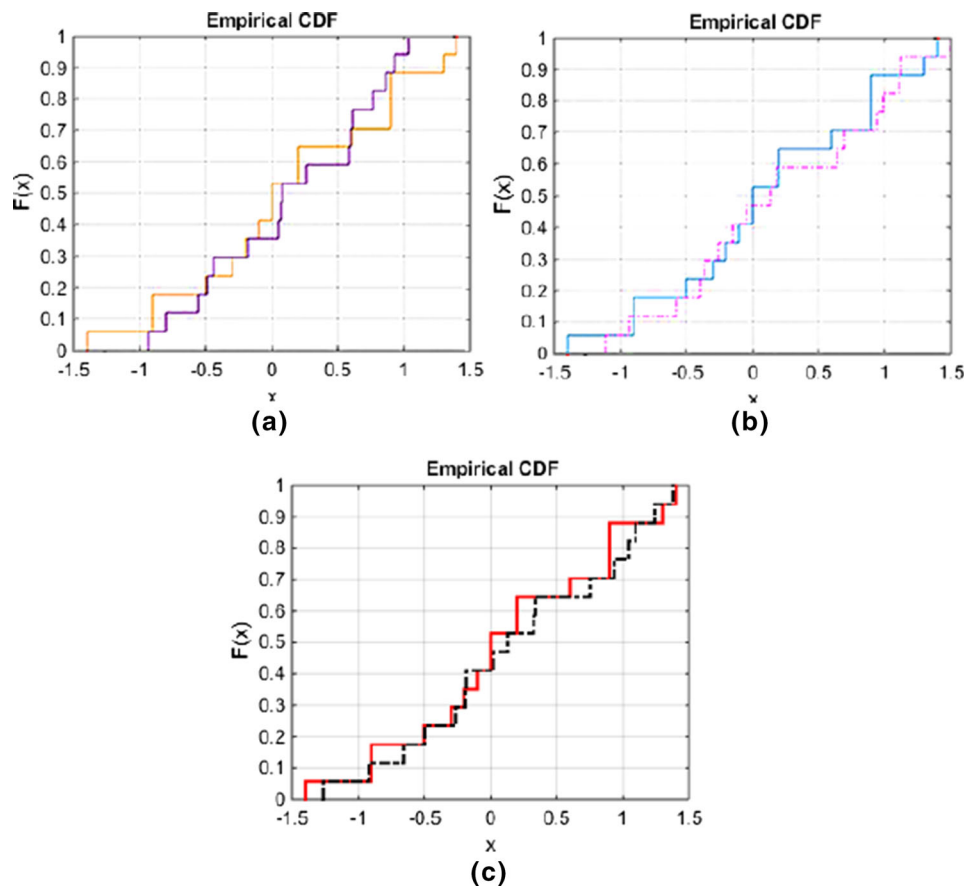


Fig. 8 Comparison of two cumulative distribution functions (CDF), **a** real data and neural networks output, **b** real data and ANFIS, **c** real data and SVM



and observed data by using two-sample K–S test is shown in Table 5, and the cumulative distribution function plot is depicted in Fig. 8. The K–S test reports the maximum difference between these cumulative distributions. The results confirm that the K–S test of SVM model is more powerful than the other models and SVM model created the best result.

Conclusions

The results of the predicted data by these models indicate low errors and more accuracy of these methods to predict. On the other hand, achieve to the regression with 0.9974 of SVM model is shown that this model can be applied in other meteorological stations. Moreover, the results indicate that the high flexibility and accuracy of SVM model can be used as a powerful tool in simulating and forecasting. The meteorological factors such as precipitation, temperature, and humidity are the most effective factors in increasing the accuracy of prediction. At the end, the 2-sample K–S test accepts the null hypothesis because the p value of SVM model is 0.9303, and this value is greater than the default value of the level of significance compared to ANN and ANFIS, can be reached an acceptable response.

References

- Adam SP, Magoulas GD, Karras DA, Vrahatis MN (2016) Bounding the search space for global optimization of neural networks learning error: an interval analysis approach. *J Mach Learn Res* 17:1–40
- Akbari MH, Vafakhah M (2016) Monthly river flow prediction using adaptive neuro-fuzzy inference system (a case study: Gharasu Watershed, Ardabil Province-Iran). *ECOPERSIA* 3(4):1175–1188
- Alipour Z et al (2014) Comparison of three methods of ANN, ANFIS and Time Series Models to predict ground water level: (case study: North Mahyar plain). *Bull Environ Pharmacol Life Sci* 3(Special Issue V):128–134
- Arabasadi Z et al (2017) Computer aided decision making for heart disease detection using hybrid neural network—genetic algorithm. *Comput Methods Progr in Biomed* 141:19–26
- Barker LJ et al (2016) From meteorological to hydrological drought using standardized indicators. *Hydrol Earth Syst Sci* 20:2483–2505
- Belayneh A, Adamowski J (2013) Drought forecasting using new machine learning methods. *J Water Land Dev* 18:3–12
- Chao C-F, Horng M-H (2015) The construction of support vector machine classifier using the firefly algorithm. *Comput Intell Neurosci* 2015:8. doi:10.1155/2015/212719
- Chihaoui M, Elkefi A, Bellil W, Ben Amar C (2016) A survey of 2D face recognition techniques. *Computers* 5:21. doi:10.3390/computers5040021
- Dai A (2013) Increasing drought under global warming in observations and models. *Nat Clim Change* 3(1):52–58
- Demyanova Y et al (2017) Empirical software metrics for benchmarking of verification tools. *Form Methods Syst Des* 50:289–316
- Devi CJ et al (2012) ANN approach for weather prediction using back propagation. *Int J Eng Trends Technol* 3(1):19–23
- Folorunsho JO et al (2012) Application of adaptive neuro fuzzy inference system (Anfis) in river kaduna discharge forecasting. *Res J Appl Sci Eng Technol* 4(21):4275–4283
- Ghosh A et al (2014) A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. *Int J Appl Erath Obs Geoinf* 26:49–63
- Gonzalez-Sanchez A, Frausto-Solis J, Ojeda-Bustamante W (2014) Attribute selection impact on linear and nonlinear regression models for crop yield prediction. *Sci World J*. Article ID 509429
- Han J, Kamber M (2006) *Data mining: concepts and techniques*, 2nd edn. Morgan Kaufmann, Los Altos
- Hassani H, Sirimal-Silva E (2015) A Kolmogoro–v–Smirnov based test for comparing the predictive accuracy of two sets of forecasts. *Econometrics* 3:590–609
- Hosseinpour NH et al (2011) Drought forecasting using ANFIS, drought time series and climate indices for next coming year (Case study: Zahedan). *Water Wastewater Consult Eng Res Dev* 2:42–51
- Hsu CH-W, Chang CH-CH, Lin CH-J (2013) *A practical guide to support vector classification*. Department of Computer Science National Taiwan University, Taipei, p 106
- Jang J-SR, Sun C-T, Mizutan E (1997) *Neuro-fuzzy and soft computing*. Prentice Hall, Englewood Cliffs (Cited by 102)
- Jiao G et al (2016) A new hybrid forecasting approach applied to hydrological data: a case study on precipitation in Northwestern China. *Water* 8:367
- Jinal JD, Parekh F (2013) Assessment of drought using standardized precipitation index and reconnaissance drought index and forecasting by artificial neural network. *Int J Sci Res (IJSR) Index Copernic Value* 6:1665–1668
- Kadhim HH (2011) Self learning of ANFIS inverse control using iterative learning technique. *Int J Comput Appl* 21(8):24–29
- Kayri M (2016) Predictive abilities of bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: a comparative empirical study on social data. *Math Comput Appl* 21:20
- Keskin M et al (2009) Meteorological drought analysis using data-driven models for the Lakes District, Turkey. *Hydrol Sci* 54(6):1114–1124
- Kruse R (2008) *Fuzzy neural network*. Institute for Information and Communication Systems Otto von-Guericke-University of Magdeburg, Magdeburg
- Kumar KS (2016) Performance variation of support vector machine and probabilistic neural network in classification of cancer datasets. *Int J Appl Eng Res* 11(4):2224–2234
- Lloyd-Hughes B, Saunders MA (2002) A drought climatology for Europe. *Int J Climatol* 22:1571–1592
- Maca P, Pech P (2015) Forecasting SPEI and SPI drought indices using the integrated artificial neural networks. *Comput Intell Neurosci* 17. Article ID 3868519
- Maier AR, Jain A, Dandy GC, Sudheer KP (2010) Methods used for development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *J Environ Model* 25(8):891–909
- McKee TB et al (1993) The relationship of drought frequency and duration to time scales. In: *Proceedings of 8th conference on applied climatology*, California, pp 17–22
- Mohammadi A et al (2014) Predicting product life cycle using fuzzy neural network. *Manag Sci Lett* 4:2057–2064

- Moreira EE et al (2016) SPI drought class predictions driven by the North Atlantic Oscillation Index using log-linear modeling. *Water* 8:43
- Nguyen LB et al (2015) Adaptive neuro-fuzzy inference system for drought forecasting in the Cai River Basin in Vietnam. *J Fac Agric Kyushu Univ* 60(2):405–415
- Palmer WC (1965) Meteorological drought. US Department of Commerce, Washington, DC
- Patel J, Parekh F (2014) Forecasting rainfall using adaptive neuro-fuzzy inference system (ANFIS). *Int J Appl Innov Eng Manag (IJAIEEM)* 3(6):262–269
- Paulo AA et al (2012) Climate trends and behavior of drought indices based on precipitation and evapotranspiration in Portugal. *Nat Hazards Earth Syst Sci* 12:1481–1491
- Ramlan R et al (2016) Implementation of fuzzy inference system for production planning optimization. In: Proceedings of the 2016 international conference on industrial engineering and operations management, Kuala Lumpur, p 8
- Rezaeianzadeh M et al (2016) Drought forecasting using Markov chain model and artificial neural networks. *Water Resour Manag* 30:2245–2259
- Sahoo P (2013) Probability and mathematical statistics. Department of Mathematics, University of Louisville, Louisville
- Scarselli F, Tsoi AC (1998) Universal approximation using feedforward neural networks: a survey of some existing methods, and some new results. *Neural Netw* 11:15–37
- Sepahi S et al (2016) Prediction of cell density of polystyrene/nanosilicafoams by artificial neural network. In: 12th International seminar on polymer science and technology
- Shirmohammadi B et al (2013) Forecasting of meteorological drought using wavelet-ANFIS hybrid model for different time steps (case study: southeastern part of east Azerbaijan province, Iran). *Nat Hazards* 69(1):389–402
- Suess S et al (2015) Using class probabilities to map gradual transitions in shrub vegetation from simulated EnMAP data. *Remote Sens* 7:10668–10688
- Sujay Raghavendra N, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput* 19:372–386
- Ustaoglu B, Cigizoglu H, Karaca M (2008) Forecast of daily minimum, maximum and mean temperature time series by three artificial neural networks. *Meteorol Appl* 15:431–445
- Vapnik V (1999) The nature of statistical learning theory, 2nd edn. Springer, Berlin
- Vapnik V, Chervonenkis A (1991) The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognit Image Anal* 1(3):283–305
- Vieira LM et al (2017) PlantRNA_Sniffer: a SVM-based workflow to predict long intergenic non-coding RNAs in plants. *Non Coding RNA* 3:11
- Wambua RM, Mutua BM, Raude JM (2016) Prediction of missing hydro-meteorological data series using artificial neural networks (ANN) for Upper Tana River Basin. *Kenya Am J Water Resour* 4(2):35–43
- Wang Y et al (2015) Improved reliability-based optimization with support vector machines and its application in aircraft wing design. *Mathemat Prob Eng* 2015:14. doi:[10.1155/2015/569016](https://doi.org/10.1155/2015/569016)
- Zhang QJ, Gupta KC, Devabhaktuni VK (2003) Artificial neural networks for RF and microwave design from theory to practice (IEEE, Kuldip C. Gupta, Fellow, IEEE, and Vijay K. Devabhaktuni, Student Member)