

Article

Evaluation of Machine Learning Techniques for Hydrological Drought Modeling: A Case Study of the Wadi Ouahrane Basin in Algeria

Mohammed Achite ^{1,2}, Muhammad Jehanzaib ^{3,*}, Nehal Elshaboury ⁴ and Tae-Woong Kim ^{5,*}

- ¹ Laboratory of Water and Environment, Faculty of Nature and Life Sciences, Hassiba Benbouali University of Chlef, Chlef 02180, Algeria; m.achite@univ-chlef.dz
- ² National Higher School of Agronomy, ENSA, Algiers 16200, Algeria
- ³ Research Institute of Engineering and Technology, Hanyang University, Ansan 15588, Korea
- ⁴ Construction and Project Management Research Institute, Housing and Building National Research Centre, Giza 12311, Egypt; nehal.elshaboury@hbrc.edu.eg
- ⁵ Department of Civil and Environmental Engineering, Hanyang University, Ansan 15588, Korea
- * Correspondence: jehanzaib7@hanyang.ac.kr (M.J.); twkim72@hanyang.ac.kr (T.-W.K.)

Abstract: Forecasting meteorological and hydrological drought using standardized metrics of rainfall and runoff (SPI/SRI) is critical for the long-term planning and management of water resources at the global and regional levels. In this study, various machine learning (ML) techniques including four methods (i.e., ANN, ANFIS, SVM, and DT) were utilized to construct hydrological drought forecasting models in the Wadi Ouahrane basin in the northern part of Algeria. The performance of ML models was assessed using evaluation criteria, including RMSE, MAE, NSE, and R^2 . The results showed that all the ML models accurately predicted hydrological drought, while the SVM model outperformed the other ML models, with the average RMSE = 0.28, MAE = 0.19, NSE = 0.86, and $R^2 = 0.90$. The coefficient of determination of SVM was 0.95 for predicting SRI at the 12-months timescale; as the timescale moves from higher to lower (12 months to 3 months), R^2 starts decreasing.

Keywords: drought modeling; machine learning; support vector machine; Algeria



Citation: Achite, M.; Jehanzaib, M.; Elshaboury, N.; Kim, T.-W. Evaluation of Machine Learning Techniques for Hydrological Drought Modeling: A Case Study of the Wadi Ouahrane Basin in Algeria. *Water* **2022**, *14*, 431. <https://doi.org/10.3390/w14030431>

Academic Editor: Pankaj Kumar

Received: 4 January 2022

Accepted: 27 January 2022

Published: 30 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Drought is a climatic calamity that has far-reaching consequences for human society, ecosystems, agriculture, and water resources [1–3]. Extreme drought phenomena are anticipated to become more frequent and intense across the world [4]. Droughts have become more intense, frequent, and widespread in recent decades due to human influences on atmospheric dynamics [5]. Drought may result in desertification, water shortages, forest fires, and crop losses, among other consequences. Drought is classified into four categories based on the nature of the deficiency: hydrological, meteorological, socioeconomic, and agricultural [6]. Because of its catastrophic repercussions for groundwater and surface water resources, hydrological drought attracts the most attention from governments, stakeholders, scholars, and the general public. Long-term droughts have become common phenomena in Algeria in the last few decades [7]. Long periods of below-average rainfall in a semi-arid region have wreaked havoc on Algerian agriculture [8]. As a result, hydrological drought forecasting, monitoring, and mitigation are critical for successful water resource preservation and economic sustainability.

Hydrological and meteorological drought could be monitored using a variety of drought indices, including the standardized precipitation index (SPI), standardized precipitation evapotranspiration index (SPEI), standardized hydrological drought index (SHDI), standardized runoff index (SRI), surface water supply index (SWSI), soil moisture drought index (SMDI), and Palmer drought severity index (PDSI) [9]. All of these drought indices were developed with the goal of forecasting, quantifying, and monitoring drought at any

hydrometeorological station over multiple periods [10]. Because of their ease of use and versatility in assessing drought over a wide range of timeframes, with strong comparability in time and place and minimum data requirements, these standardized drought indices have been widely utilized [11,12]. Even though drought forecasting is a difficult issue owing to its uncertainty and high complexity [13], drought forecasting research is critical in providing relevant information for reducing the risks associated with drought occurrence.

Establishing long-term solutions to manage water shortages in the face of global warming and properly forecasting drought occurrences is critical [14,15]. In hydrological applications, two types of drought forecasting models are commonly used; physical-based models and data-driven models. Understanding the physical processes of a system underpins physical process-based models, whereas the most meaningful relationship between input and output data is established by data-driven models [16]. Because of the difficulties in obtaining parameter estimates, data-driven models have become more popular and developed than physical-based models. With computational intelligence methods, data-driven modeling that requires few inputs and develops rapidly is getting more powerful and versatile for forecasting meteorological time series prediction [17]. Because data-driven models offer a lot of potential for drought forecasting, they were utilized in this research study.

The most often used data-driven models in the domain of drought forecasting are machine learning (ML) models, which are a subset of artificial intelligence [18,19]. These models are effective for anticipating the onset of droughts with different frequencies, durations, and intensities that are not properly represented by empirical relationships. The models include support vector regression (SVR), artificial neural networks (ANN), and extreme learning machines (ELM). Dikshit et al. [20] undertook a study to forecast drought events in New South Wales, Australia. The SPEI was considered because it calculated both temperature and rainfall factors and had previously been shown to be more accurate in predicting droughts. Using a climate research unit dataset, the drought index was constructed at several periods (1, 3, 6, and 12 months). The study examined 13 factors to forecast the temporal aspect of the drought index, eight of which were sea surface temperature and climatic indicators, while the rest were diverse meteorological variables. ANN and SVR forecasting models were trained using the data between 1901 and 2010 and then evaluated for eight years (2011–2018). The models were tested using three distinct performance metrics; coefficient of determination (R^2), mean absolute error (MAE), and root-mean-square error (RMSE). The results showed that ANN ($R^2 = 0.86$) outperformed SVR ($R^2 = 0.75$) in forecasting temporal drought patterns. It was also revealed that the climatic indicator (Pacific decadal oscillation) and sea surface temperatures have minimal impact on temporal droughts.

The majority of artificial intelligence models are highly accurate, but they are complex and need computational resources to train. On the other hand, decision trees (DT), extreme gradient boosting (XGB), gradient boosting (GB), and random forests (RF) are becoming more popular since they are deemed more powerful, simpler, and more resilient prediction algorithms [21–23]. Zhang et al. [24] used meteorological measurements, drought indices, and climatic signals from 32 stations in Shaanxi Province, China, to develop a novel drought forecasting model. In order to choose the best predictors and determine their lag duration, cross-correlation function and distributed lag nonlinear model (DLNM) methods were utilized and compared. The performance of the DLNM, ANN, and XGB models was validated to forecast the SPEI for the next 1–6 months. With R^2 values of 0.68–0.82, 0.72–0.89, 0.81–0.92, and 0.84–0.95 for 3, 6, 9, and 12 months, the XGB model outperformed the DLNM and the ANN models with the lead time of 1–6 months. Furthermore, the XGB model exhibited the greatest forecast accuracy for overall droughts as well as for moderate, severe, and extreme droughts. Li et al. [25] used the antecedent SST fluctuation pattern (ASFP) and ML techniques (e.g., SVR, ELM, and RF) to develop a new meteorological drought prediction strategy. Four river basins that experience regular droughts on various continents were used as case studies, with 1- and 3-month-long lead periods for forecasting

SPEI. The results revealed that the ASFP–ELM model outperformed the other two models in predicting the space–time evolution of drought occurrences.

Some other research studies addressed the application of hybrid ML models that are trained using optimization algorithms for drought prediction. Mohamadi et al. [26] employed the support vector machine (SVM), radial basis function neural network (RBFNN), adaptive neuro-fuzzy inference system (ANFIS), and multilayer perceptron (MLP) models to anticipate climatic droughts in Iran using data from 1980 to 2014. The models were trained using a nomadic people algorithm (NPA), and this algorithm was compared to the bat, salp swarm, and krill algorithms. Soft computing models' accuracy and convergence speed were improved using these evolutionary approaches. The 3-month standardized precipitation index was forecast using hybrid ML models. The Nash–Sutcliffe efficiency (NSE) for the ANFIS, MLP, RBFNN, and SVM models coupled with the NPA algorithm was 0.93, 0.86, 0.85, and 0.83, respectively. The findings also showed that the hybrid models outperformed the standalone models. Nabipour et al. [27] predicted short-term hydrological droughts using a combination of ANN models trained using optimization algorithms. The algorithms included the grasshopper optimization algorithm (GOA), salp swarm algorithm (SSA), particle swarm optimization (PSO), and biogeography-based optimization (BBO). SHDI and SPI were calculated in aggregated months of one, three, and six. Then, using cross-correlation analysis, SHDI forecasting was carried out using three states and 36 input–output combinations. The results of the hybrid models were compared to those of the conventional ANN model. The hybrid ANN model coupled with the PSO algorithm outperformed the traditional ANN. The best models yielded $R^2 = 0.68$ and $RMSE = 0.58$ for SHDI1, $R^2 = 0.81$ and $RMSE = 0.45$ for SHDI3, and $R^2 = 0.82$ and $RMSE = 0.40$ for SHDI6.

Deep learning models, subsets of ML models, are now commonly utilized for high accuracy in a variety of study domains. Convolutional neural network (CNN) and long short-term memory (LSTM) models are proven to be better suited for time series predictions. However, there is a scarcity of drought monitoring data utilizing deep learning [28]. For instance, Adikari et al. [29] compared three widely used artificial intelligence-based flood and drought forecasting models. In this research, fluvial floods and meteorological droughts were quantified using the change in river discharge and SPI, respectively. Five statistical performance criteria and indicators were used to compare the performance of the LSTM, CNN, and wavelet decomposition functions coupled with ANFIS (WANFIS) in arid and tropical climatic regions. The results revealed that independent of the environment of the region under investigation, the CNN and the WANFIS models exhibited the best performance in flood and meteorological drought forecasting, respectively. Mokhtar et al. [30] presented a combination of ML techniques for drought analysis in the Tibetan Plateau, China, during 1980–2019. Three-months (SPEI-3) and six-months (SPEI-6) aggregation timelines were investigated. To estimate SPEIs, four ML models were developed: RF, XGB, CNN, and LSTM. The models were built using seven scenarios with varying combinations of climatic variables (i.e., precipitation, temperature (average, maximum, and minimum), wind speed, relative humidity, sunshine hours, and solar radiation) as inputs. The first scenario was developed by considering the first two variables, and each subsequent scenario was built by accounting for an additional climatic input variable. For scenarios 4, 5, and 6, the best models for predicting SPEI-3 were LSTM, XGB, and RF, respectively. Meanwhile, the top models for SPEI-6 were XGB for scenario 5 and RF for scenario 7. The performance of the XGB and RF models produced satisfactory results for scenarios 4–7 for both timeframes based on the NSE index. For meteorological drought forecasting, the majority of research studies in the literature employed the ANFIS, ANN, and SVM models, but the relevance of newly developed ML approaches such as the DT has not been thoroughly examined.

Algeria, Africa's largest country, does not have many sources of drinkable water. As a result, approximately 40% of the total population suffer from water scarcity. The nation is 95% dry and 80% desert, with only a few inches of rain per year. Water shortage is a problem not just in the country's poorly inhabited areas. In addition, poor water infrastructure has led to inefficient water management: ruptures, theft, and leaks waste up to 30% of water in

transportation [31]. As a result, comprehensive drought forecasting is crucial for detecting and mitigating droughts early. The majority of research in the literature employed two or three prediction models for meteorological drought forecasting [20,24,25]. However, no study has been documented that applies and compares different ML approaches for hydrological drought forecasting in Algeria. The major goal of this research is to apply four ML techniques (i.e., ANN, ANFIS, SVM, and DT) to construct hydrological drought forecasting models. As a result, the proposed modeling techniques might lead to improved attempts to address limitations in drought prediction, which could then aid in mitigation strategies such as regulations for water supply system management and sustainable water use.

2. Study Area and Data Collection

The Wadi Ouahrane basin in northern Algeria, which is between 36°00' N and 36°24' N and between 01°00' E and 01°3' E, was the study area. This area is part of the Wadi Cheliff basin (Figure 1), and it extends over 270 km². The study area has a maximum altitude of 991 m and a minimum altitude of 165 m. The Wadi Ouahrane is a small tributary of the Wadi Cheliff. This basin (Figure 1) is controlled by six pluviometric stations and one hydrometric station. The Wadi Ouahrane basin is limited in the east by the basin of the Wadi Fodda, in the west by the Wadi Ras basin, in the north by the Wadi Allala basin, and in the south by the Wadi Sly basin. It is influenced by the Mediterranean climate with the interannual average rainfall of 333 mm over 1972–2018. The mean annual temperature is 18 °C.

The rainfall series database used in this study was gathered on a monthly scale at six stations from 1972 to 2018 (Figure 1, Table 1). These data were sourced from the National Agency of Meteorology and Hydrology (ANRH) and the National Office of Meteorology (ONM).

Table 1. Rainfall stations and hydrometric station characteristics.

Stations	ID	Name	Geographical Coordinates (° ' ")		Elevation (m)
			Longitude	Latitude	
Rainfall stations					
S1	012201	LARBAT OULED FARES	01°09'18"	36°16'20"	116
S2	012224	BOUZGHAIA	01°14'27"	36°20'15"	217
S3	012205	BENAIRIA	01°22'28"	36°21'04"	320
S4	012221	MEDJAJA	01°20'53"	36°16'39"	487
S5	012209	CHETIA	01°15'53"	36°12'56"	108
S6	NMO	Airport, Chlef	01°19'28"	36°13'31"	158
Hydrometric station					
S1	012201	LARBAT OULED FARES	01°13'5"	36°14'14"	173

The Thiessen polygons technique was used to calculate the monthly areal mean basin rainfall in this study. The weights of all the six stations were computed based on the Thiessen polygons method (Figure 1). To construct SPI, these monthly mean rainfalls were used.

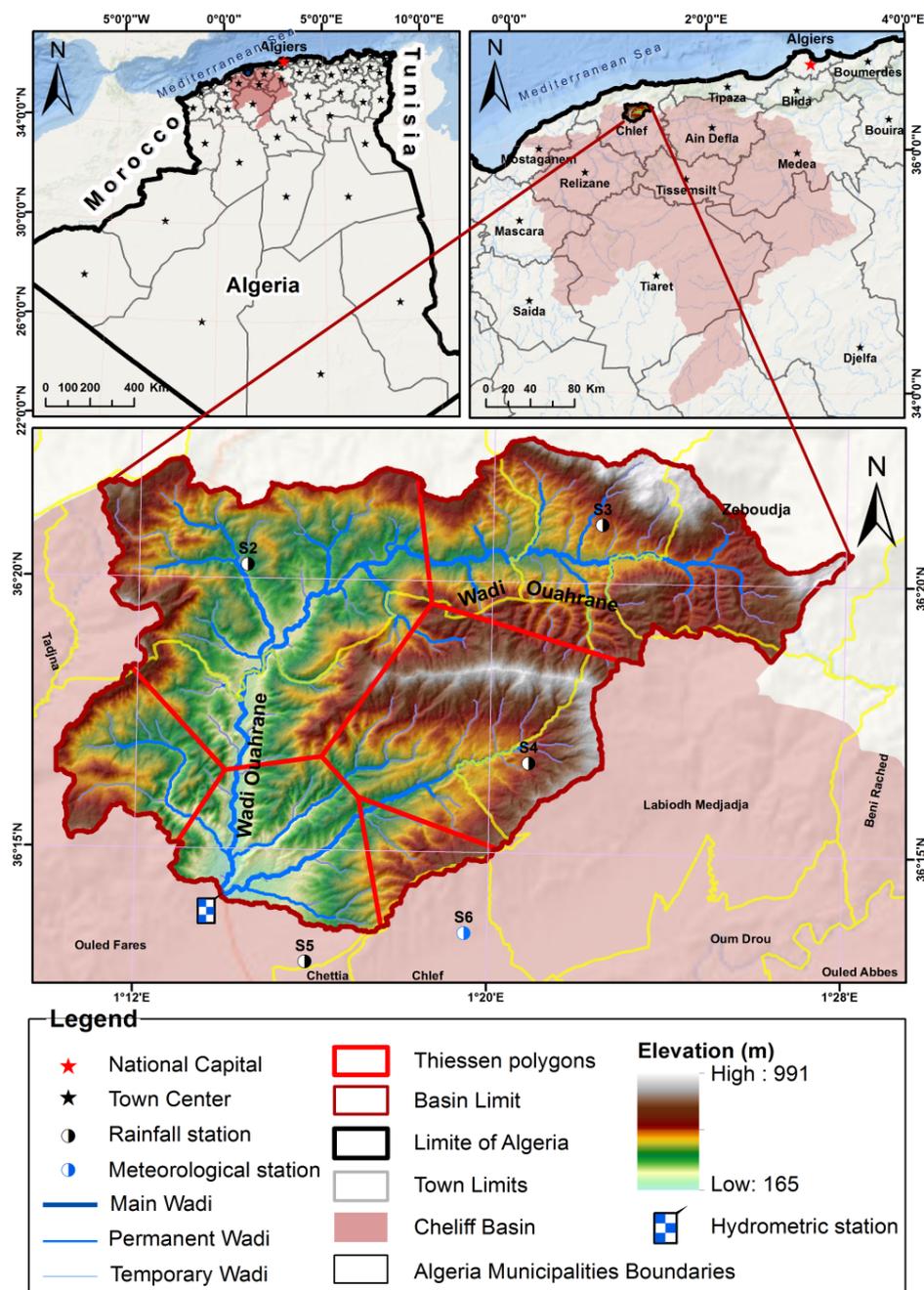


Figure 1. Location, topographic characteristics, pluviometric and hydrometric network of the study area.

3. Materials and Methods

3.1. SPI and SRI

SPI and SRI are commonly used to detect meteorological and hydrological droughts. The cumulative probability of monthly precipitation quantity collected at the observation site is used to calculate SPI/SRI [32]. The parameters of the precipitation probability density function, assumed to be in the form of gamma distribution, were determined for the whole observation period at a meteorological station as per Equation (1):

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \tag{1}$$

where α and β are the shape and scale parameters, respectively. Meanwhile, x is consecutive precipitation and $\Gamma(\alpha)$ is the gamma function. The gamma function is defined by Equation (2).

$$\Gamma(a) = \int_0^{\infty} y^{a-1} e^{-y} dy \quad (2)$$

The precipitation time series was used to determine the alpha and beta parameters of gamma distribution as per Equation (3):

$$\alpha = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right), \quad A = \ln(\bar{x}) - \frac{\sum \ln(x_i)}{n}, \quad \beta = \frac{\bar{x}}{\alpha} \quad (3)$$

where \bar{x} is the mean precipitation quantity; n is the number of precipitation observations; x_i is the total amount of precipitation in a set of data. The cumulative probability can be presented in Equation (4).

$$G(x) = \int_0^x g(x) dx = \frac{1}{\beta^a \Gamma(a)} \int_0^x x^{a-1} e^{-x/\beta} dx \quad (4)$$

A mixed probability distribution was employed to account for the probability of zero precipitation, and the cumulative probability is indicated in Equation (5):

$$H(x) = q + (1 - q)G(x) \quad (5)$$

where q is the probability that the precipitation quantity equals zero. The calculation of SPI was presented based on Equation (6) [33]:

$$\text{SPI} = \begin{cases} -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), & 0 < H(x) \leq 0.5 \\ +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), & 0.5 < H(x) \leq 1.0 \end{cases} \quad (6)$$

where t was determined as shown in Equation (7):

$$t = \begin{cases} \sqrt{\ln\left(\frac{1}{H(x)^2}\right)}, & 0 < H(x) \leq 0.5 \\ \sqrt{\ln\left(\frac{1}{(1-H(x))^2}\right)}, & 0.5 < H(x) \leq 1.0 \end{cases} \quad (7)$$

where $c_0, c_1, c_2, d_1, d_2,$ and d_3 are coefficients whose values are:

$$c_0 = 2.515517. c_1 = 0.802853. c_2 = 0.010328$$

$$d_1 = 1.432788. d_2 = 0.189269. d_3 = 0.001308$$

Different classifications and estimated probabilities of wet and dry spells can be examined based on SPI for the studied timeframe, as shown in Table 2 [34]. Similarly, SRI is calculated by fitting log-normal probability to the hydrometric data, and cumulative probabilities are transformed into standard normal variates by following Equations (5)–(7).

3.2. Machine Learning Techniques

For drought forecasting, four machine learning approaches were used in this study: ANN, ANFIS, DT, and SVM. The following subsections offer a summary of each approach.

3.2.1. Artificial Neural Network

ANN is a powerful nonlinear modeling technique that mimics human brain activity. It is made up of basic processing neurons that are interconnected to conduct mathematical manipulation [35]. The network extracts the patterns between the input variables and

the predicted output values. Due to the stochastic nature of drought features and the complexity of the involved processes, it is excellent for drought forecasting [36]. There are many different types of ANNs, with the multilayer perceptron neural network being one of the most common. The most common ANN architecture comprises three layers and signals that are transmitted in a forward direction through the network. The first layer (input layer) receives input data but does not execute any numerical computations. The second layer (hidden layer) is responsible for carrying out the important numerical calculations. The output is finally produced and displayed by the last output layer. The back-propagation technique in the R-package *neuralnet* was used to train the network using normalized input and output data in this study.

Table 2. Categorization of different states of standardized drought indices.

State	SPI/SRI Range	
	Minimum	Maximum
Extremely wet		≥ 2.0
Severely wet	1.50	1.99
Moderately wet	1.00	1.49
Near normal	−0.99	0.99
Moderately dry	−1.49	−1.00
Severely dry	−1.99	−1.50
Extremely dry		≤ -2.0

3.2.2. Adaptive Neuro-Fuzzy Inference System

ANFIS is a soft computing approach for simulating nonlinearity functions [37]. It constructs input–output mapping using human knowledge [38,39]. The combination of neural networks with fuzzy logic is the foundation of ANFIS. Because fuzzy logic increases the learning properties of neural networks, ANFIS performs well in uncertain situations [40]. An ANFIS model consists of the input layer, the rules layer, the input/output membership function layers, and the output layer. The input parameters are fuzzified in the first layer, and the firing strength of the rules is calculated in the second layer. The firing strength of the rules is normalized in the third layer. The rule outputs are distributed in the fourth layer, and the system output is determined in the fifth layer by aggregating the rule outputs. The fundamental goal of training the ANFIS network is to obtain real output by estimating unknown parameters or weights from the training dataset. Using the MATLAB toolbox (R2019b), a hit-and-miss strategy was used to find the proper structure.

3.2.3. Decision Trees

Breiman et al. [41] introduced DT to tackle a variety of classification and regression problems. DT divide the data into partitioning variables based on the strongest association between the responses [42]. The following are the details of the conditional inference trees' working algorithms:

1. All response variables and covariates are subjected to a global test of independence. If the outcome is not rejected, the procedure must be halted. Otherwise, the procedure proceeds to find the covariate that has the greatest impact on the response variable.
2. In the adopted covariate, a binary split is performed.

This process is continued until the global independence test is passed. The null hypothesis of response variable independence underpins the termination criteria, with any covariate being natural and statistically justifiable. The minimal splitting criterion used in this study was 95%, and the test statistic distribution was calculated using the Monte Carlo test in R using the *party* package.

3.2.4. Support Vector Machine

SVM, introduced by Vapnik [43], is an efficient and simple ML method of handling noisy data. The regression aspect of an SVM includes SVR, which solves prediction problems. The main distinction between ANN and SVR is that the former aims to minimize training error while the latter strives to reduce generalization error. Furthermore, unlike ANN, SVR is based on the structural risk minimization principle [44]. This well-known supervised learning approach has been employed in numerous drought-predicting studies. The basic principle behind this method is to simplify the classification process by acquiring a high-dimensional space from the original dataset in the input space. In this approach, support vectors are employed as selection criteria, and they generate the best data categorization boundaries [45]. For SVR, the three parameters are epsilon (ϵ), cost (C), and gamma (γ). C is the capacity control parameter, ϵ is the loss function to define the regression vector, and γ monitors the output complexity and minimizes the model space. The radial basis kernel was employed in this study to develop the model, and the grid search technique was used to tune the model parameters using the *e1071* package in R [39,46].

3.3. Performance Evaluation Criteria

The performance of the developed models was assessed using various evaluation criteria: root-mean-square error (RMSE), mean absolute error (MAE), Nash–Sutcliffe coefficient of efficiency (NSE), and determination coefficient (R^2). The four criteria are calculated according to Equations (8)–(11) [47–50]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (SRI_{i(observed)} - SRI_{i(model)})^2} \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |SRI_{ti(observed)} - SRI_{ti(model)}| \quad (9)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^N (SRI_{i(observed)} - SRI_{i(model)})^2}{\sum_{i=1}^N (SRI_{i(observed)} - \overline{SRI_{i(observed)}})^2} \right] \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (SRI_{i(observed)} - SRI_{i(model)})^2}{\sum_{i=1}^N (SRI_{i(observed)} - SRI_{mean})^2} \quad (11)$$

where N is the number of observed SRI data, $SRI_{i(observed)}$ and $SRI_{i(model)}$ are the observed and modeled SRI estimations, respectively, and SRI_{mean} is the mean of the observed SRI.

4. Results and Discussion

The meteorological drought index (SPI) was computed at multiple timescales (1–12 months), whereas the hydrological drought index (SRI) was computed at the 3-, 6-, 9-, and 12-months timescales. Correlation analysis was performed between SPI at different timescales (1–12 months) and SRI at the 3-, 6-, 9-, and 12-months timescales as shown in Figure 2. The SPI timescale which best responds to SRI-3, -6, -9, and -12 months was chosen as an input for SRI prediction at the 3-, 6-, 9-, and 12-months timescales. Further, the optimal inputs (lags) were nominated through the partial autocorrelation function (PACF) at the 5% significance level for all the SRI timescales as shown in Figure 3. The brown dotted line in these figures represents the upper and lower critical limits at the 5% significance level. The lags where the PACF value crosses the limits are considered statistically significant and can be utilized as input of ML models. The best input combination for SRI prediction adopted in this research is presented in Table 3.

Table 3. Relationship between input and output for SRI prediction.

Station Name	Input Variables	Target Variable
LARBAT OULED FARES	SPI-5, SRI-3(t−1), SRI-3(t−2), SRI-3(t−3)	SRI-3
	SPI-6, SRI-6(t−1), SRI-6(t−2), SRI-6(t−11)	SRI-6
	SPI-10, SRI-9(t−1), SRI-9(t−2), SRI-9(t−9), SRI-9(t−11), SRI-9(t−12)	SRI-9
	SPI-12, SRI-12(t−1), SRI-12(t−2), SRI-12(t−9)	SRI-12

A wide range of ML models with varying complexity and robustness is available for drought prediction. In this research, the authors selected four state-of-the-art ML models (ANN, ANFIS, DT, and SVM) for hydrological drought prediction at 3-, 6-, 9-, and 12-months timescales. The data were divided into two phases: training and testing; 80% of the data were used to teach the models, while the remaining 20% were utilized to assess the learning efficacy of the models. In the ANN model, the hit-and-trial method was adopted to determine the most appropriate network structure. The structure of ANN with three hidden layers and neuron combinations in three hidden layers (6, 4, and 2) performed best and was used for SRI prediction. Similarly, the optimum structure of the ANFIS model was selected using the miss-and-hit strategy. A hybrid method was used to train the model, with the Gaussian membership function for inputs and the linear membership function for outputs. A DT model was also developed using the Monte Carlo test type and a 95% splitting condition. Finally, the SVM model was tuned by changing the cost parameter from 0.001 to 100 with steps of 10, the ϵ parameter—from 0 to 1 with steps of 0.01, and the gamma parameter—at 0.33. The normalized input data were applied for training ANN and ANFIS, whereas the DT and SVM models utilized the input data without normalization.

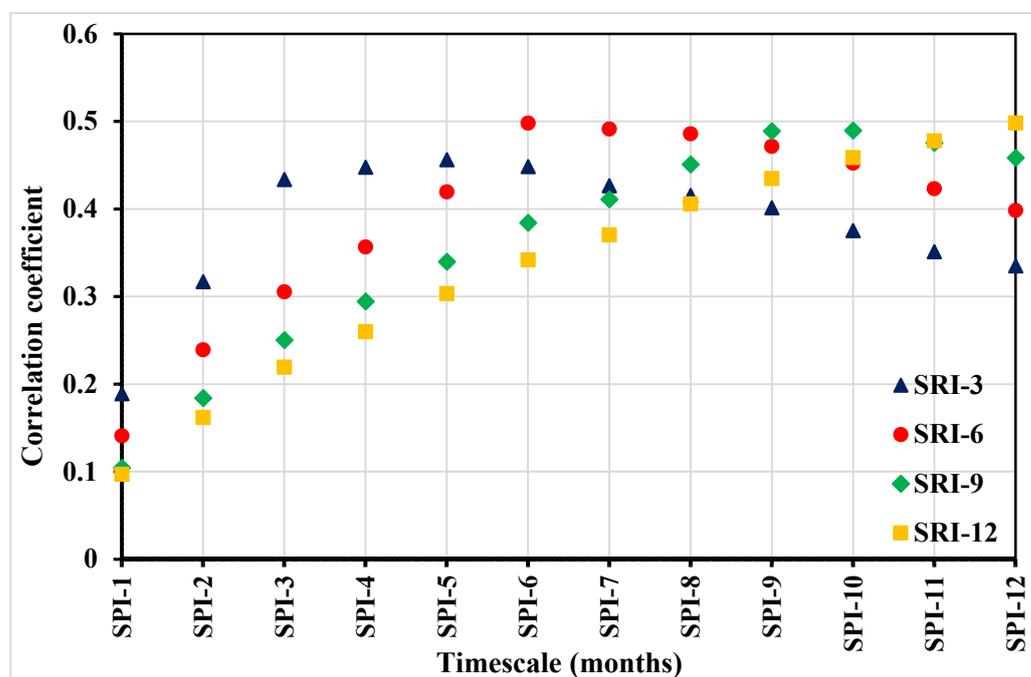


Figure 2. Correlation analysis between different timescales of SPI and SRI.

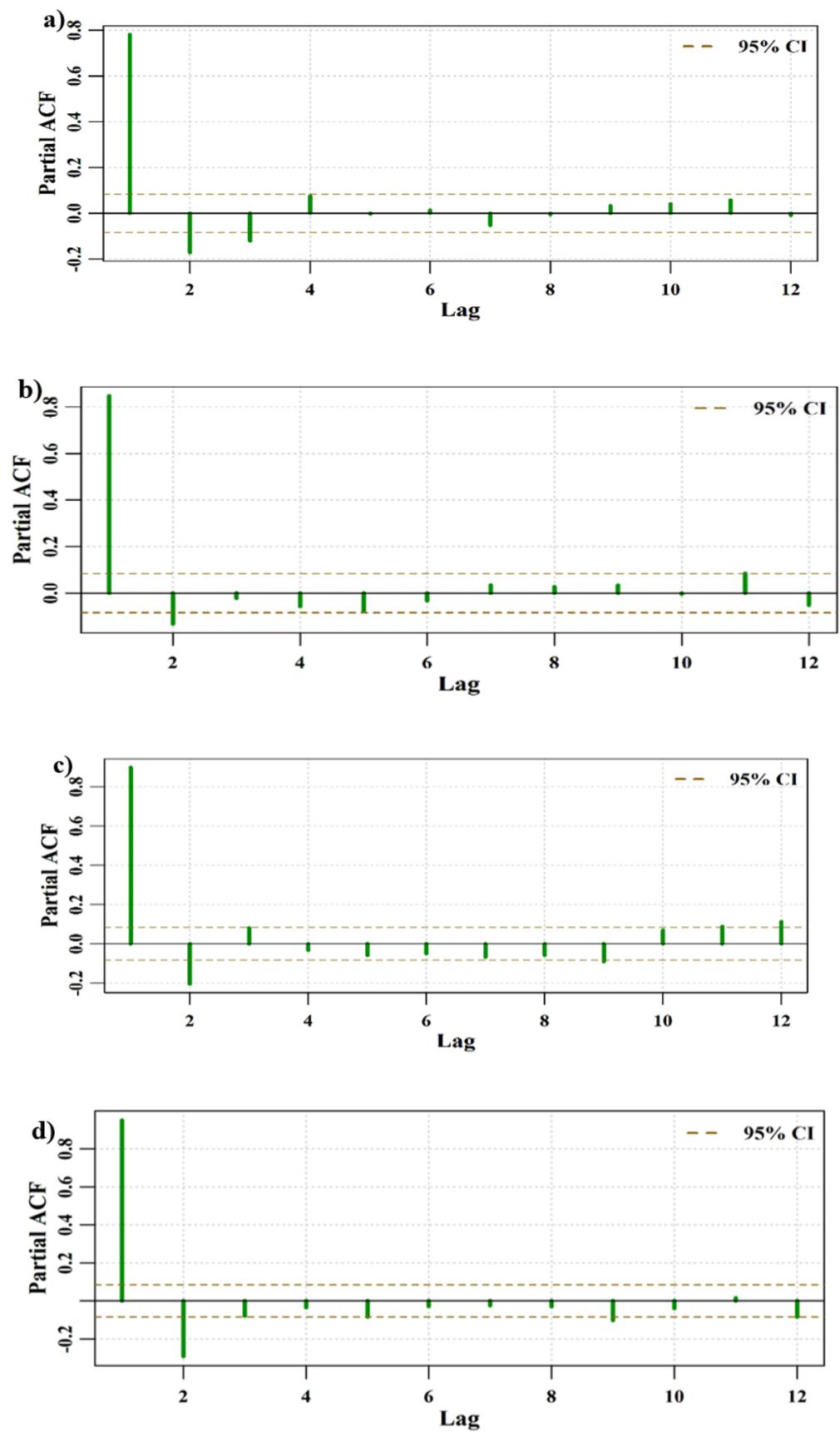


Figure 3. Partial autocorrelation function for (a) SPI-3, (b) SPI-6, (c) SPI-9, and (d) SPI-12.

Table 4 shows the results of the performance evaluation of ML models using performance criteria (RMSE, MAE, NSE, and R^2). For hydrological drought prediction, the ML approach with RMSE and MAE close to 0 and NSE and R^2 close to 1 was deemed the best. These findings evidenced that all ML models can accurately predict hydrological drought. The average performance of ANFIS remained on the lower side in this analysis, while the performance of DT was better than that of ANFIS in predicting SRI at all the timescales. Similarly, the ANN model performed better as compared to ANFIS and DT at all the timescales. The performance of the SVM model was found robust in predicting SRI at all the timescales and, therefore, recommended for hydrological drought forecasting. These results correlate with the previous study on drought forecasting [39].

Table 4. Evaluation of the performance of various ML techniques of hydrological drought prediction.

ML Model	Timescale	Performance Indicators			
		RMSE	MAE	NSE	R^2
DT	3 months	0.45	0.31	0.57	0.73
	6 months	0.51	0.35	0.61	0.75
	9 months	0.46	0.31	0.66	0.74
	12 months	0.34	0.23	0.83	0.85
ANFIS	3 months	0.48	0.32	0.69	0.72
	6 months	0.55	0.39	0.67	0.72
	9 months	0.57	0.33	0.71	0.71
	12 months	0.4	0.23	0.79	0.8
ANN	3 months	0.39	0.27	0.64	0.83
	6 months	0.37	0.26	0.8	0.86
	9 months	0.35	0.24	0.8	0.86
	12 months	0.27	0.17	0.89	0.9
SVM	3 months	0.31	0.22	0.79	0.88
	6 months	0.34	0.24	0.85	0.89
	9 months	0.28	0.2	0.88	0.91
	12 months	0.19	0.12	0.95	0.95

The scatterplots of comparison between the observed and predicted SRI values at the 3-, 6-, 9- and 12-months timescales by all the ML models are shown in Figure 4, Figure 5, Figure 6, and Figure 7, respectively. Taylor diagram was also constructed to represent the performance of all the ML models graphically at all the timescales as shown in Figure 8. It is clearly shown in Figure 8 that the performance of the ML models was in the following order: SVM > ANN > DT > ANFIS. Further, the trained models were utilized for simulating droughts in the testing period at all the timescales. Figure 9 represents the simulation results of all the ML models. It is clear from Figure 9 that the simulations performed using the ANFIS and DT models showed a high deviation from the observed SRI, while the ANN and SVM models had very little deviation from the observed SRI.

The ML models were also assessed with respect to the computation time, as shown in Table 5. It was observed that the SVM model took less time to converge after tuning its parameters, while the convergence time of the other models was on the higher side. The models were arranged from the least to the most computational time as follows: SVM, DT, ANN, and ANFIS. This study was conducted using a machine with an Intel i7 8th generation processor and 16 GB of RAM. According to these findings, the SVM model surpassed all the other ML models in terms of performance and computation time. The ANFIS and DT models overpredicted droughts at all the timescales. Following a comprehensive evaluation, it was concluded that the SVM model can be used for hydrological drought forecasting and can be applied for drought prediction in Algeria as well as other parts of the globe.

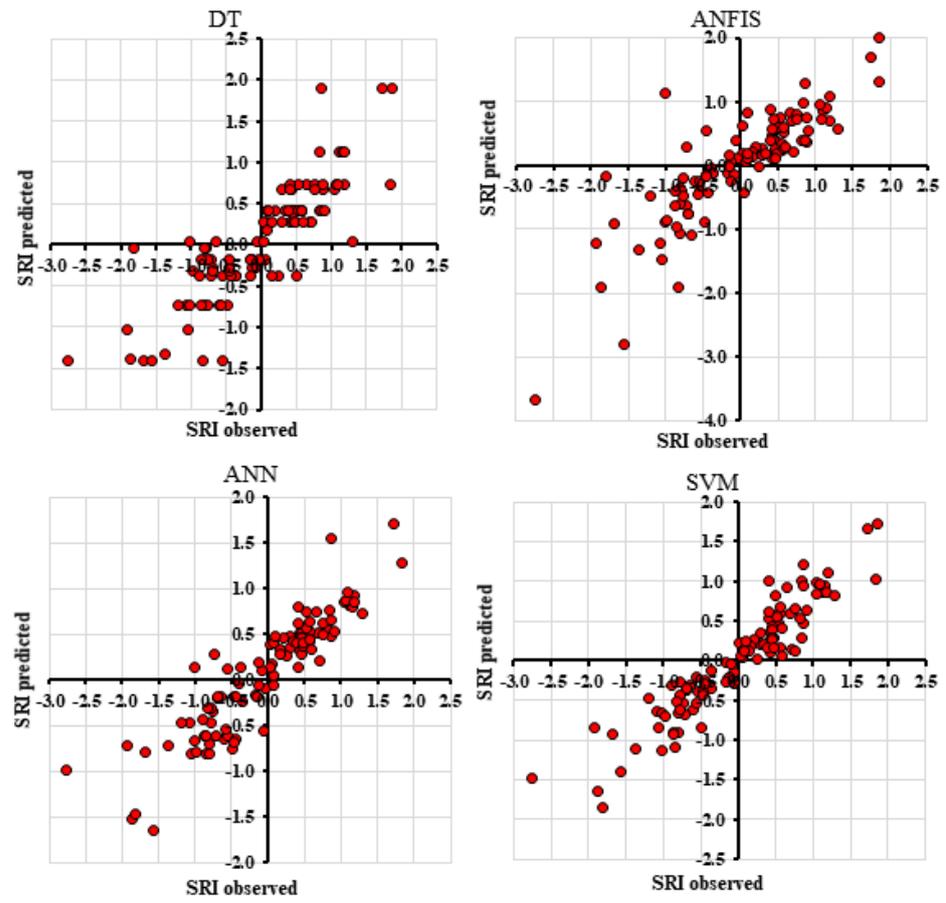


Figure 4. Scatterplots of predicted and observed SRI-3 using various ML models.

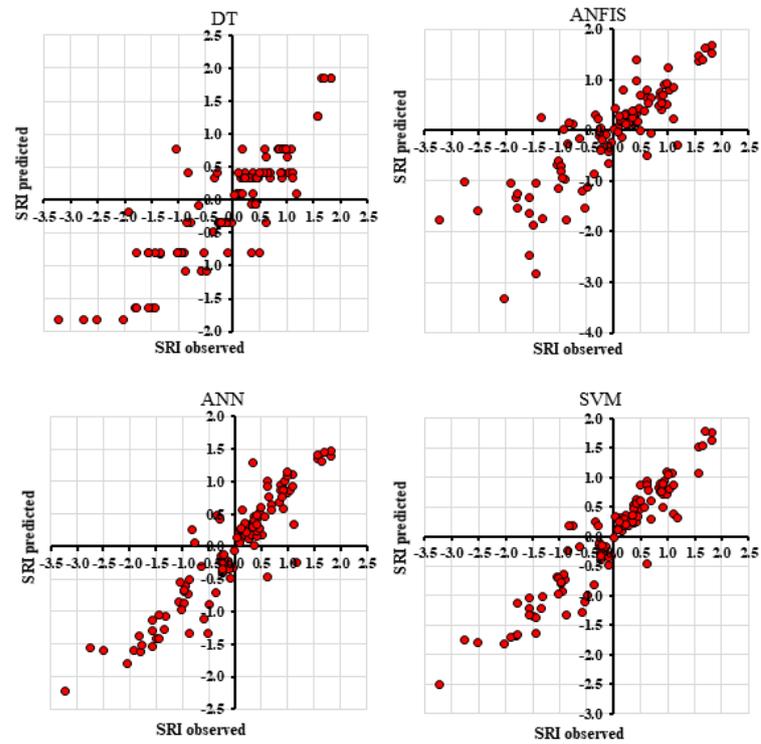


Figure 5. Scatterplots of predicted and observed SRI-6 using various ML models.

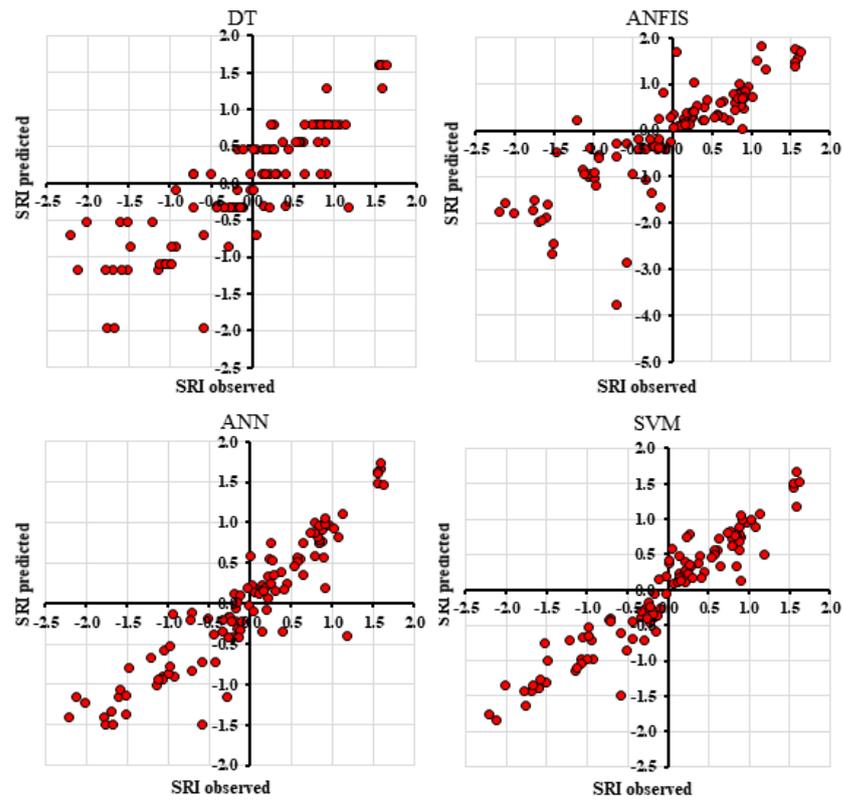


Figure 6. Scatterplots of predicted and observed SRI-9 using various ML models.

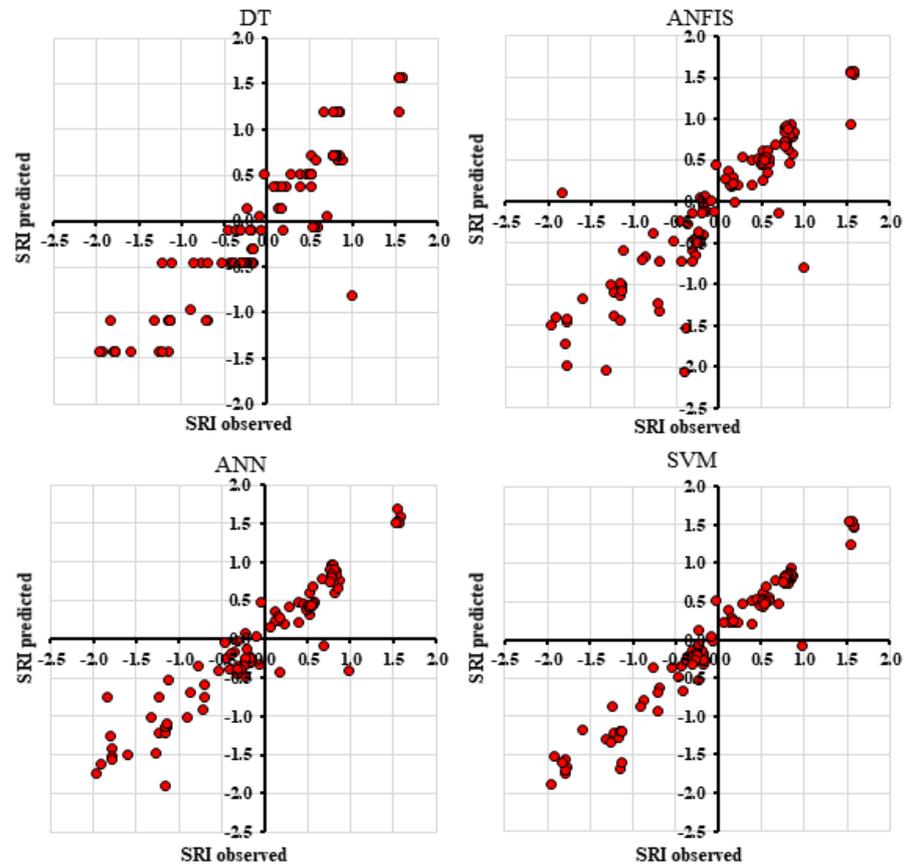


Figure 7. Scatterplots of predicted and observed SRI-12 using various ML models.

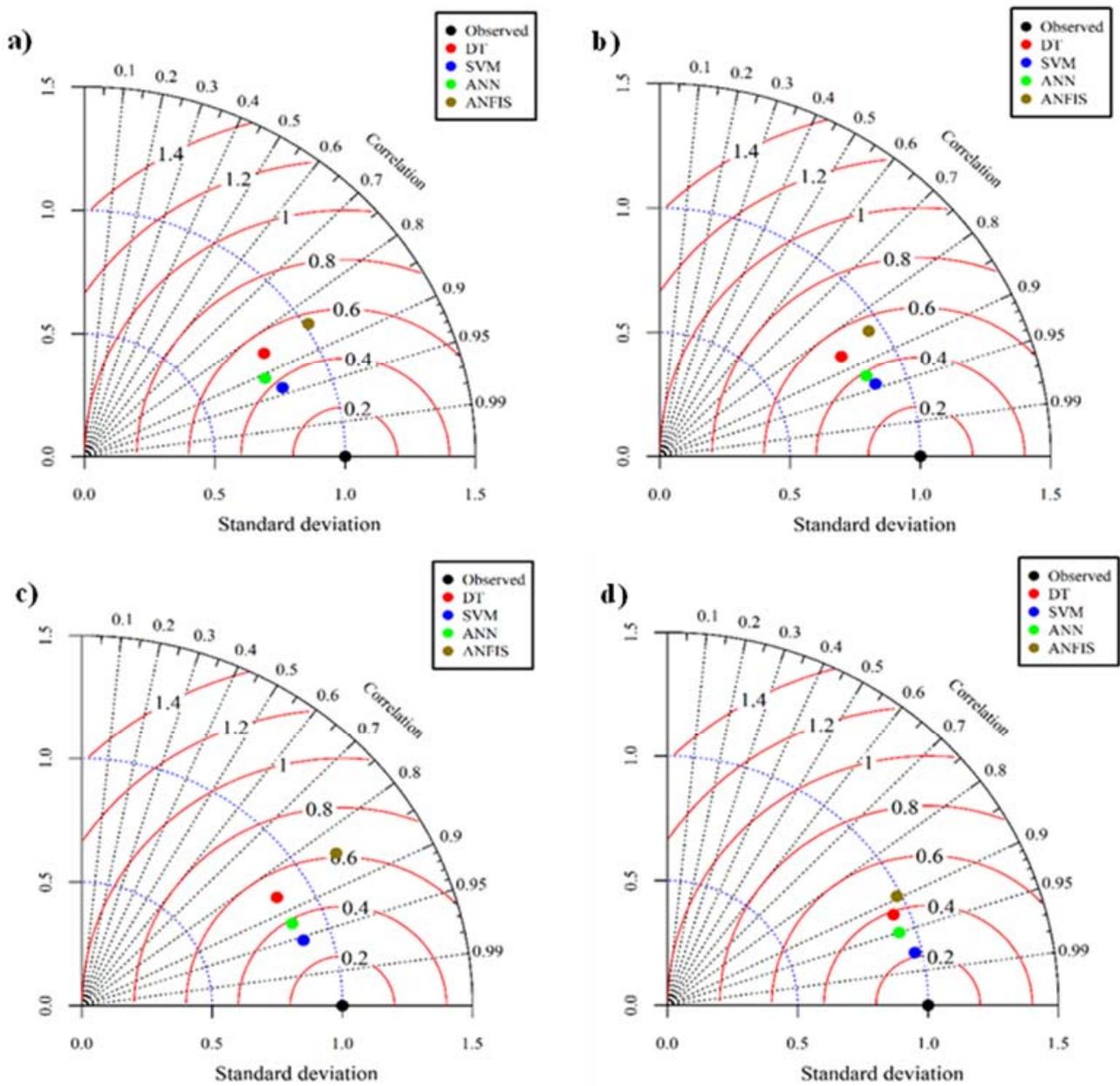


Figure 8. Taylor diagram of the predicted and observed SRI using various ML models: (a) SPI-3, (b) SPI-6, (c) SPI-9, and (d) SPI-12.

Table 5. Comparison of techniques based on computation time.

Sr. No.	ML Models	Average Computation Time (s)
1	ANN	11.65
2	ANFIS	28.35
3	DT	1.37
4	SVM	0.15

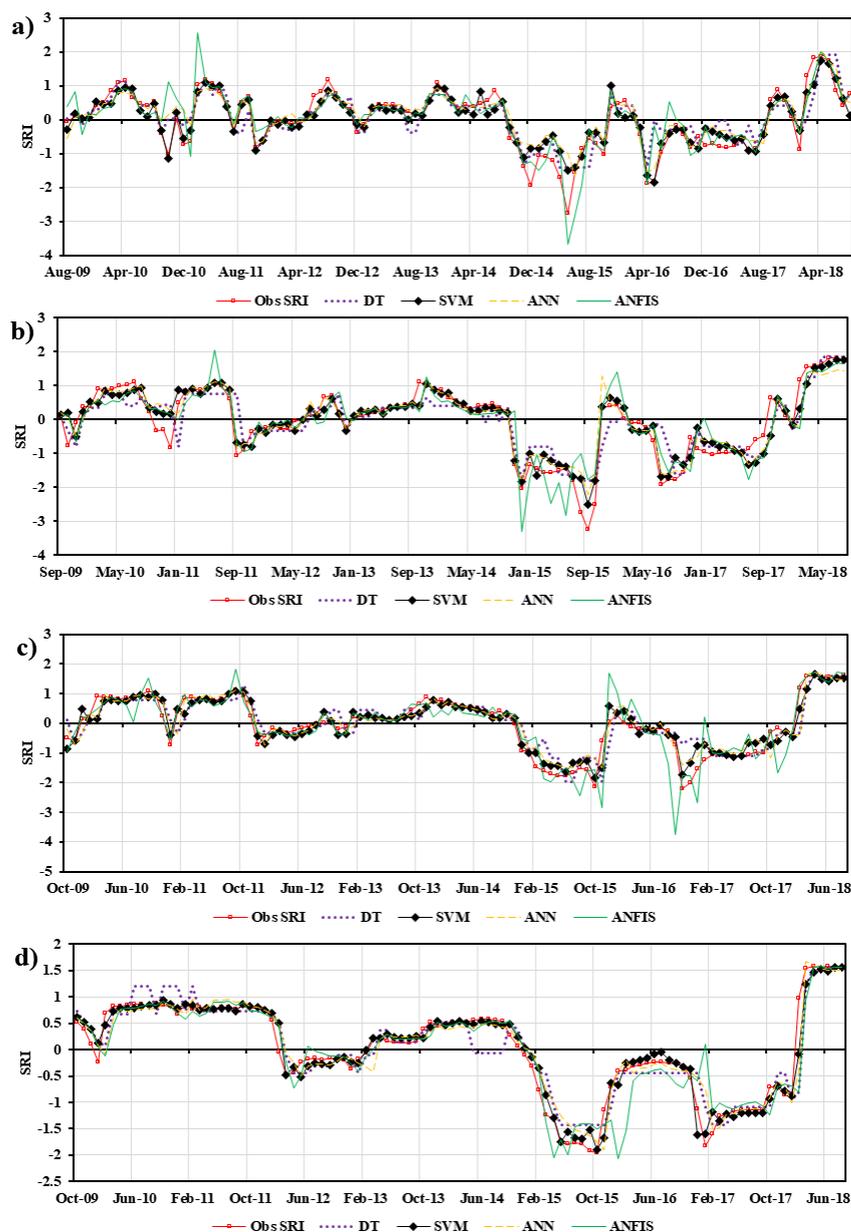


Figure 9. Comparison of the predicted and observed SRI values using different machine learning techniques: (a) SPI-3, (b) SPI-6, (c) SPI-9, and (d) SPI-12.

5. Conclusions

The efficacy of several machine learning approaches to forecast hydrological droughts was investigated in this study. SPI was estimated at various timescales (1–12 months), while SRI was calculated at the 3-, 6-, 9-, and 12- months timescales. Correlation analysis between all the timescales of SPI and SRI was performed to select the timescale of SPI which corresponds to SPI3-, 6-, 9-, and 12 months. The optimal inputs (lags) were identified using PACF at the 5% significance level for all the SRI timescales. The SPI timescale which showed the highest correspondence with SRI and statistically significant SRI lags were utilized as input of ML models for hydrological drought forecasting. The performance of the ML models was assessed using several evaluation criteria, including RMSE, MAE, NSE, and R^2 . The results showed that all the ML models accurately predicted hydrological droughts, while the SVM model outperformed the other ML models, with the average RMSE = 0.28, MAE = 0.19, NSE = 0.86, and $R^2 = 0.90$. The coefficient of determination of SVM was 0.95 for predicting SRI at the 12-months timescale; as the timescale moves

from higher to lower (12 months to 3 months), R^2 starts decreasing. This is because a small timescale results in a large number of minor droughts as compared to a large timescale. The performance of the ANFIS and DT models was found to be on the lower side throughout the analysis because these models overpredicted the drought. Finally, the performance of the ML models was also compared based on computation time. The primary drawbacks of using the ML techniques are the time and resources involved in their computations. Because the SVM's convergence rate is faster and the approach requires less computational resources, it successfully overcomes this problem.

The application of SVM to drought prediction is strongly recommended after conducting a comprehensive assessment of various ML approaches. The inclusion of additional climatic and hydrological factors is recommended in future studies. The findings of this study would be useful in making early decisions and developing preparedness strategies to deal with the effects of impending calamities.

Author Contributions: Conceptualization, M.A.; methodology, M.J. and M.A.; software, M.J.; validation, M.J., M.A. and T.-W.K.; formal analysis, N.E.; investigation, M.A. and M.J.; resources, M.A., M.J. and N.E.; data curation, M.A.; writing—original draft preparation, M.A., M.J. and N.E.; writing—review and editing, M.A., M.J., N.E. and T.-W.K.; visualization, M.A.; supervision, T.-W.K.; project administration, M.J.; funding acquisition, T.-W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Lower-Level and Core Disaster Safety Technology Development Program funded by the Ministry of Interior and Safety (grant No. 2020-MOIS33-006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all the subjects involved in the study.

Data Availability Statement: Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: Thanks to peer reviewers who improved this manuscript. We thank the Lower-Level and Core Disaster Safety Technology Development Program funded by the Ministry of Interior and Safety (grant No. 2020-MOIS33-006). We also thank the ANRH agency for the collected data and the General Directorate of Scientific Research and Technological Development of Algeria (DGRSDT).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Williams, A.P.; Cook, E.R.; Smerdon, J.E.; Cook, B.I.; Abatzoglou, J.T.; Bolles, K.; Livneh, B. Large contribution from anthropogenic warming to an emerging North American megadrought. *Science* **2020**, *368*, 314–318. [[CrossRef](#)] [[PubMed](#)]
- Jiang, W.; Wang, L.; Zhang, M.; Yao, R.; Chen, X.; Gui, X.; Cao, Q. Analysis of drought events and their impacts on vegetation productivity based on the integrated surface drought index in the Hanjiang River Basin, China. *Atmos. Res.* **2021**, *254*, 105536. [[CrossRef](#)]
- Li, J.; Wang, Z.; Wu, X.; Zscheischler, J.; Guo, S.; Chen, X. A standardized index for assessing sub-monthly compound dry and hot conditions with application in China. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1587–1601. [[CrossRef](#)]
- Vicente-Serrano, S.M.; Van der Schrier, G.; Beguería, S.; Azorin-Molina, C.; Lopez-Moreno, J.I. Contribution of precipitation and reference evapotranspiration to drought indices under different climates. *J. Hydrol.* **2015**, *526*, 42–54. [[CrossRef](#)]
- Khan, N.; Sachindra, D.A.; Shahid, S.; Ahmed, K.; Shiru, M.S.; Nawaz, N. Prediction of droughts over Pakistan using machine learning algorithms. *Adv. Water Resour.* **2020**, *139*, 103562. [[CrossRef](#)]
- Heim, R.R., Jr. A review of twentieth-century drought indices used in the United States. *Bull. Am. Meteorol. Soc.* **2002**, *83*, 1149–1166. [[CrossRef](#)]
- Achite, M.; Wałęga, A.; Toubal, A.K.; Mansour, H.; Krakauer, N. Spatiotemporal characteristics and trends of meteorological droughts in the Wadi Mina Basin, Northwest Algeria. *Water* **2021**, *13*, 3103. [[CrossRef](#)]
- Habibi, B.; Meddi, M.; Torfs, P.J.; Remaoun, M.; Van Lanen, H.A. Characterisation and prediction of meteorological drought using stochastic models in the semi-arid Chélif–Zahrez basin (Algeria). *J. Hydrol. Reg. Stud.* **2018**, *16*, 15–31. [[CrossRef](#)]
- Mishra, A.K.; Singh, V.P. Drought modeling—A review. *J. Hydrol.* **2011**, *403*, 157–175. [[CrossRef](#)]
- Jain, V.K.; Pandey, R.P.; Jain, M.K.; Byun, H.R. Comparison of drought indices for appraisal of drought characteristics in the Ken River Basin. *Weather Clim. Extrem.* **2015**, *8*, 1–11. [[CrossRef](#)]

11. Kim, T.W.; Jehanzaib, M. Drought risk analysis, forecasting and assessment under climate change. *Water* **2020**, *12*, 1862. [CrossRef]
12. Jehanzaib, M.; Bilal Idrees, M.; Kim, D.; Kim, T.W. Comprehensive evaluation of machine learning techniques for hydrological drought forecasting. *J. Irrig. Drain. Eng.* **2021**, *147*, 04021022. [CrossRef]
13. Durbach, I.; Merven, B.; McCall, B. Expert elicitation of autocorrelated time series with application to e3 (energy-environment-economic) forecasting models. *Environ. Model. Softw.* **2017**, *88*, 93–105. [CrossRef]
14. Bordi, I.; Sutera, A. Drought monitoring and forecasting at large scale. In *Methods and Tools for Drought Analysis and Management*; Springer: Dordrecht, The Netherlands, 2007; pp. 3–27.
15. Soh, Y.W.; Koo, C.H.; Huang, Y.F.; Fung, K.F. Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia. *Comput. Electron. Agric.* **2018**, *144*, 164–173. [CrossRef]
16. Sari, Y.D.; Zarlis, M. Data-driven modelling for decision making under uncertainty. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *300*, 012013.
17. Adamowski, J.F. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *J. Hydrol.* **2008**, *353*, 247–266. [CrossRef]
18. Sundararajan, K.; Garg, L.; Srinivasan, K.; Bashir, A.K.; Kaliappan, J.; Ganapathy, G.P.; Meena, T. A contemporary review on drought modeling using machine learning approaches. *CMES-Comput. Model. Eng. Sci.* **2021**, *128*, 447–487. [CrossRef]
19. Yaseen, Z.M.; Ali, M.; Sharafati, A.; Al-Ansari, N.; Shahid, S. Forecasting standardized precipitation index using data intelligence models: Regional investigation of Bangladesh. *Sci. Rep.* **2021**, *11*, 1–25. [CrossRef] [PubMed]
20. Dikshit, A.; Pradhan, B.; Alamri, A.M. Temporal hydrological drought index forecasting for New South Wales, Australia using machine learning approaches. *Atmosphere* **2020**, *11*, 585. [CrossRef]
21. Feng, Y.; Cui, N.; Zhang, Q.; Zhao, L.; Gong, D. Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. *Int. J. Hydrog. Energy* **2017**, *42*, 14418–14428. [CrossRef]
22. Hassan, M.A.; Khalil, A.; Kaseb, S.; Kassem, M.A. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Appl. Energy* **2017**, *203*, 897–916. [CrossRef]
23. Papadopoulos, S.; Azar, E.; Woon, W.L.; Kontokosta, C.E. Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. *J. Build. Perform. Simul.* **2018**, *11*, 322–332. [CrossRef]
24. Zhang, R.; Chen, Z.Y.; Xu, L.J.; Ou, C.Q. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. *Sci. Total Environ.* **2019**, *665*, 338–346. [CrossRef]
25. Li, J.; Wang, Z.; Wu, X.; Xu, C.Y.; Guo, S.; Chen, X.; Zhang, Z. Robust meteorological drought prediction using antecedent SST fluctuations and machine learning. *Water Resour. Res.* **2021**, *57*, e2020WR029413. [CrossRef]
26. Mohamadi, S.; Sammen, S.S.; Panahi, F.; Ehteram, M.; Kisi, O.; Mosavi, A.; Al-Ansari, N. Zoning map for drought prediction using integrated machine learning models with a nomadic people optimization algorithm. *Nat. Hazards* **2020**, *104*, 537–579. [CrossRef]
27. Nabipour, N.; Dehghani, M.; Mosavi, A.; Shamshirband, S. Short-term hydrological drought forecasting based on different nature-inspired optimization algorithms hybridized with artificial neural networks. *IEEE Access* **2020**, *8*, 15210–15222. [CrossRef]
28. Shen, R.; Huang, A.; Li, B.; Guo, J. Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *79*, 48–57. [CrossRef]
29. Adikari, K.E.; Shrestha, S.; Ratnayake, D.T.; Budhathoki, A.; Mohanasundaram, S.; Dailey, M.N. Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. *Environ. Model. Softw.* **2021**, *144*, 105136. [CrossRef]
30. Mokhtar, A.; Jalali, M.; He, H.; Al-Ansari, N.; Elbeltagi, A.; Alsafadi, K.; Rodrigo-Comino, J. Estimation of SPEI meteorological drought using machine learning algorithms. *IEEE Access* **2021**, *9*, 65503–65523. [CrossRef]
31. Future Directions International Pty Ltd. Water Protests in Algeria Are Giving Cause for Concern about Its Long-Term Stability. Available online: <https://www.futuredirections.org.au/publication/water-protests-in-algeria-are-giving-cause-for-concern-about-its-long-term-stability/> (accessed on 6 December 2021).
32. Awange, J.L.; Mpelasoka, F.; Goncalves, R.M. When every drop counts: Analysis of droughts in Brazil for the 1901–2013 period. *Sci. Total Environ.* **2016**, *566*, 1472–1488. [CrossRef]
33. Naresh Kumar, M.; Murthy, C.S.; Sessa Sai, M.V.R.; Roy, P.S. On the use of Standardized Precipitation Index (SPI) for drought intensity assessment. *Meteorol. Appl.* **2009**, *16*, 381–389. [CrossRef]
34. McKee, T.B.; Doesken, N.J.; Kliest, J. The Relationship of Drought Frequency and Duration to Time Scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993.
35. See, L.; Openshaw, S. Applying soft computing approaches to river level forecasting. *Hydrol. Sci. J.* **1999**, *44*, 763–778. [CrossRef]
36. Mishra, A.K.; Desai, V.R. Drought forecasting using stochastic models. *Stoch. Environ. Res. Risk Assess.* **2005**, *19*, 326–339. [CrossRef]
37. Jang, J.S.R.; Sun, C.T.; Mizutani, E. Neuro-fuzzy and soft computing—a computational approach to learning and machine intelligence [Book Review]. *IEEE Trans. Automat. Contr.* **1997**, *42*, 1482–1484. [CrossRef]
38. Gholami, A.; Bonakdari, H.; Zaji, A.H.; Ajeel Fenjan, S.; Akhtari, A.A. Design of modified structure multi-layer perceptron networks based on decision trees for the prediction of flow parameters in 90 open-channel bends. *Eng. Appl. Comput. Fluid Mech.* **2016**, *10*, 193–208. [CrossRef]
39. Mokhtarzad, M.; Eskandari, F.; Vanjani, N.J.; Arabasadi, A. Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environ. Earth Sci.* **2017**, *76*, 1–10. [CrossRef]

40. Alipour, Z.; Ali, A.M.A.; Radmanesh, F.; Joorabyan, M. Comparison of three methods of ANN, ANFIS and time series models to predict ground water level:(Case study: North Mahyar plain). *BEPLS* **2014**, *3*, 128–134.
41. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.
42. McClean, S.I. Data mining and knowledge discovery. In *Encyclopedia of Physical Science and Technology*; Academic Press: Cambridge, MA, USA, 2003; pp. 229–246.
43. Vapnik, V. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer Science & Business Media: New York, NY, USA, 1999.
44. Belayneh, A.; Adamowski, J.; Khalil, B.; Ozga-Zielinski, B. Long-term SPI drought forecasting in the Awash river basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J. Hydrol.* **2014**, *508*, 418–429. [[CrossRef](#)]
45. Deka, P.C. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput.* **2014**, *19*, 372–386.
46. Belayneh, A.; Adamowski, J. Drought forecasting using new machine learning methods. *J. Water Land Dev.* **2013**, *18*, 3–12. [[CrossRef](#)]
47. Elshaboury, N.; Marzouk, M. Comparing Machine Learning Models for Predicting Water Pipelines Condition. In Proceedings of the 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 24–26 October 2020; pp. 134–139.
48. Abdelkader, E.M.; Al-Sakkaf, A.; Elshaboury, N.; Alfalah, G. On the Implementation of Machine Learning Models for Emulating Daily Electricity Consumption in Hotel Facilities. In Proceedings of the 6th World Congress on Civil, Structural, and Environmental Engineering (CSEE'21), Lisbon, Portugal, 21–23 June 2021.
49. Elshaboury, N. Assessment of Different Artificial Neural Networks for Predicting Bridge Deck Condition. In Proceedings of the 4th Smart Cities Symposium (SCS21), Zallaq, Bahrain, 21–23 November 2021.
50. Idrees, M.B.; Jehanzaib, M.; Kim, D.; Kim, T.W. Comprehensive evaluation of machine learning models for suspended sediment load inflow prediction in a reservoir. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1805–1823. [[CrossRef](#)]