



Contents lists available at SCCE

Journal of Soft Computing in Civil Engineering

Journal homepage: [www.jsoftcivil.com](http://www.jsoftcivil.com)



## Performance Evaluation of Machine Learning Models in Predicting Dry and Wet Climatic Phases

O. Aiyelokun<sup>1\*</sup>, A. Ojelabi<sup>1</sup>, O. Agbede<sup>1</sup>

1. Department of Civil Engineering, University of Ibadan, Ibadan, Nigeria

Corresponding author: [aiyelokuntobi@gmail.com](mailto:aiyelokuntobi@gmail.com)

 <https://doi.org/10.22115/SCCE.2020.213319.1154>

### ARTICLE INFO

#### Article history:

Received: 26 December 2019

Revised: 23 January 2020

Accepted: 17 February 2020

#### Keywords:

Water resources;

Support vector machine;

Artificial neural network;

Non-linear models.

### ABSTRACT

Water resource and environmental engineers need accurate information in harnessing water for diverse uses, therefore it is expedient to accurately predict dry and wet climatic phases in order to ensure optimum water resource planning and management. This study examined the applicability of machine learning models for the prediction of extreme dry and wet conditions in Minna, North Central Nigeria. Recorded rainfall, maximum temperature, minimum temperature, relative humidity, wind speed, sunshine hours and estimated potential evapotranspiration were used as predictors in the machine learning models, while hydrological extremes estimated from standardized precipitation index (SPI) served as a response variable. The performance of Support Vector Machine (SVM) based on different kernel types and Artificial Neural Network (ANN) based on different network structures were assessed for the prediction of the different phases of the climate of the study area. The study showed that while normal meteorological conditions occurred for about 74.8% of the study period, 8.9%, 4.6% and 1.9% of this period were moderately wet, severely wet and extremely wet respectively, and 4.7%, 3.4% and 1.7% of the study period were moderately dry, severely dry and extremely dry respectively. Furthermore, SVM based on Radial Basis Kernel with a coefficient of determination of 0.64 outperformed other SVM types and ANN with two hidden layers; with of coefficient of determination 0.68 was found to perform better than ANN with single layers. Generally, ANN was found to have higher accuracy than SVM in predicting dry and wet climatic phases in North Central Nigeria.

How to cite this article: Aiyelokun O, Ojelabi, A, Agbede O. Performance evaluation of machine learning models in predicting dry and wet climatic phases. J Soft Comput Civ Eng 2020;4(1):29–48. <https://doi.org/10.22115/scce.2020.213319.1154>.

2588-2872/ © 2020 The Authors. Published by Pouyan Press.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



## 1. Introduction

Water is necessary for agricultural, domiciliary, industrial, leisurely, and ecological purposes [1]. Its demand has already surpassed its supply in many parts of the world, and many more areas are anticipated to face this disparity in the near future [2]. Water resources management is of direct interest to all individuals [3]. Since the use of water is a part of everyday living. Such extensive concern in water is not exceptional, as often claimed by many professionals, because of its roles in other essential sectors such as transportation, environment, health, and energy [4].

The ability to predict dry and wet climatic phases is immensely important in order to moderate the effects of extreme events on surface water, groundwater and water resources management projects such as hydro-electric power plants, irrigation structures, flood control structures, and others. Droughts impacted 50 percent of the 2,8 billion people who suffered natural disasters between 1967 and 1991, and 3,5 million people who lost their lives due to natural disasters 35 percent were as a result of droughts [5]. More than 50 percent of the most populous areas in the world are susceptible to drought. Droughts trigger a compound set of effects that range across several parts of the economy and influences well beyond the physical space affected by them (Mishra and Desai, 2006); while on the other hand, flooding has caused tremendous losses to properties and sometimes life [6].

In the face of disputes between stakeholders and important decisions making as regards to water use, managers commonly rely on decision support tools such as models [7]. In the case of extremes, decision-support tools usually focus on their modeling, forecasting, and management. Hydrological issues such as drought and flood could be avoided if appropriate and accurate forecasting tools are in place [6].

Drought occurs when hydrological variables such as rainfall and streamflow fall below the barest minimum value; while flood occurs when rainfall and streamflow occur above the barest maximum value. While drought usually creates the impression of water shortage due to inadequate rainfall, excessive evapotranspiration, and overuse of water resources or a mixture of these variables [8], flood gives an impression of the direct opposite. The principal cause of a drought is the absence of precipitation over a large spatial extent for an extensive period of time; known as meteorological drought [9] While the primary cause of flood is too much available water than the draining capacity of an area over a period of time.

Extreme events affect nearly all regions of the world and lead to weighty economic, social, and environmental impacts [10], there optimization of informed decision making toward their risk reduction cannot be overemphasized. The decision-making procedure commences by coalescing and unifying data into pieces of information that are then modeled to create insights, which form the foundation on which decisions are made [10].

The ability to predict extreme weather events is immensely important in order to mitigate their effects [5] and since water resources and environmental engineers need accurate information in harnessing water for diverse uses. There is a need to emphasize the use of non-linear models that

could represent the complex water resources system, with minimum error for preparation and mitigation.

Past studies have highlighted some problems with water resources in Nigeria [11]. The country is labeled as being water-short and is likely to have water declined from 2,506 cubic meters per year in 1995 to 1,175 in 2025, if not managed properly [12]. By 2025, the Food and Agriculture Organization (FAO), estimated that about 1.8 billion persons will be residing in areas with complete scarcity of water, and about two-thirds of the world's population could be in water stress conditions. The current world population is about 7 billion people, with the potential of rising to 9 billion in the next 40 years [12]. Ayanshola et al. [11] identified climate variability as one of twenty-two (22) challenges of water resources development and management in the North Central zone of Nigeria.

One of the widely used methods for meteorological extremes assessment is the Standardized Precipitation Index (SPI) established by McKee et al. [13]. This method has the ability to compute the deficit of rainfall for multiple temporal scales while reflecting its impression on the obtainability of several water supplies. This resourcefulness enables the adaptability of SPI in monitoring long and short-range supplies of water.

Short term extremes are hard to diagnose, hence to assess them, it is expedient to use short time scales, for instance, drought decisions could be extensively analyzed when dealt with in short scales as events missed on a larger scale are caught in short ones. This study, therefore, seeks to model and predict SPI based on the one-month time scale using SVM and ANN, this will serve as decision support in water resources management in the study area.

## **2. Description of the study area**

The study lies at latitudes 9°37'N- 9°79'N and longitude 6°16'E - 6°65'E. Minna is the capital of Niger state, and one of the major growing states of North Central Nigeria [14]. Minna is situated about 150 kilometers from the capital of the Federal Republic of Nigeria. The state has a population of about 506,113 persons with an average population density of about 3448 persons per km<sup>2</sup> [15]. The population growth in Minna claimed to be greater than the average of the whole nation as a result of its closeness to Abuja, which is the capital of the Federal Republic of Nigeria [14].

The geologic formation of the study area is based on the undistinguishable basement complex of mainly gneiss and magnetite [16]. The climate of the study area lies within a region described as a tropical climate [16]. It has a tropical dry and wet climate. The region is characterized by double rainfall maxima. The study area has mean annual precipitation of 1300 mm [16]. The rainy season commences most of the time in April and lasts till October, with fluctuations in the amount of rainfall received per year. The highest mean monthly rainfall occurs in September with almost 300mm. Temperature is uniformly high throughout the year reaching the peaks of 40°C (Feb./March) and 30°C (Nov./Dec.) [16].



Fig. 1. Map of the Study Area.

### 3. Methodology

#### 3.1. Materials

Hydro-meteorological data for Minna rainfall station was obtained from the Nigeria Meteorological Agency (NIMET), which includes rainfall, maximum temperature, minimum temperature, sunshine hour, relative humidity, pan evaporation, and wind speed. The time series of the data covered 55 years covering 1961 to 2015.

#### 3.2. Modified penman-monteith's method

The Penman-Monteith method, which was modified by the Food and Agricultural Organization (FAO) in 1963 was adopted for estimating reference evapotranspiration in the study area. The Modified Penman-Monteith method according to Doorenbos and Pruitt (1977)[17] and Jennifer, (2001)[18] can be mathematically expressed as follows:

$$ET_r = \frac{C}{\rho_w} \left[ W \frac{(R_n - G)}{\lambda} + (1 - W)f(u)(e_a - e_d) \right] \quad (1)$$

where  $ET_r$  is the reference evapotranspiration, mm/day;  $W$  is a temperature and the altitude of the area -related weighting factor;  $R_n$  is the net solar radiation,  $MJm^{-2}d^{-1}$ ;  $G$  is the soil heat flux in  $MJm^{-2}d^{-1}$ ;  $\lambda$  is the latent heat of evaporation,  $MJ kg^{-1}$ ;  $\rho_w = 1000 kgm^{-3}$ , is the density of water;  $f(u)$  is the wind-related function,  $kg hPa^{-1}m^{-2}d^{-1}$ ;  $e_a$  is the saturation vapor pressure at mean air

temperature, hPa;  $e_d$  is the mean actual vapour pressure of the air, hPa; and  $C$  is an adjustment factor to account for day and night weather conditions.

### 3.3. Standardized precipitation index (SPI)

The generally adapted distribution for SPI estimation is the two-parameter gamma distribution that has shape and scale parameters, and is expressed by the probability density function:

$$G(x) = \frac{1}{\delta^\sigma \tau(\sigma)} \int_0^x x^{\sigma-1} e^{-x/\delta} dx \quad \text{for } x > 0 \quad (2)$$

where  $\sigma$  is the shape parameter,  $\delta$  is the scale parameter,  $x$  is the rainfall series and  $\tau(\sigma)$  is the gamma function. The gamma distribution is indefinite for  $x = 0$ , but the rainfall could have zero value, consequently, the derived cumulative probability distribution (CPD) for a zero value is:

$$H(x) = q + (1-q) G(x) \quad (3)$$

Where  $q$  is the probability of the zero-rainfall value. The CPD is then changed into the standard normal distribution to compute SPI. The value of SPI specifies the potency of the irregularity in climatic phases, while the classification of dry/wet intensity as presented by Mckee et al. (1993) [13] can be found in table 1. In this study, a one-month time scale SPI was adopted.

**Table 1**  
Dry and Wet Climatic Phases based on SPI values.

Category	SPI values
Extremely wet	$\geq 2.0$
Severely wet	1.50 to 1.99
Moderate wet	1.00 to 1.49
Near normal	0.99 to -0.99
Moderately dry	-1.00 to -1.49
Severely dry	-1.50 to -1.99
Extremely dry	$\leq -2.0$

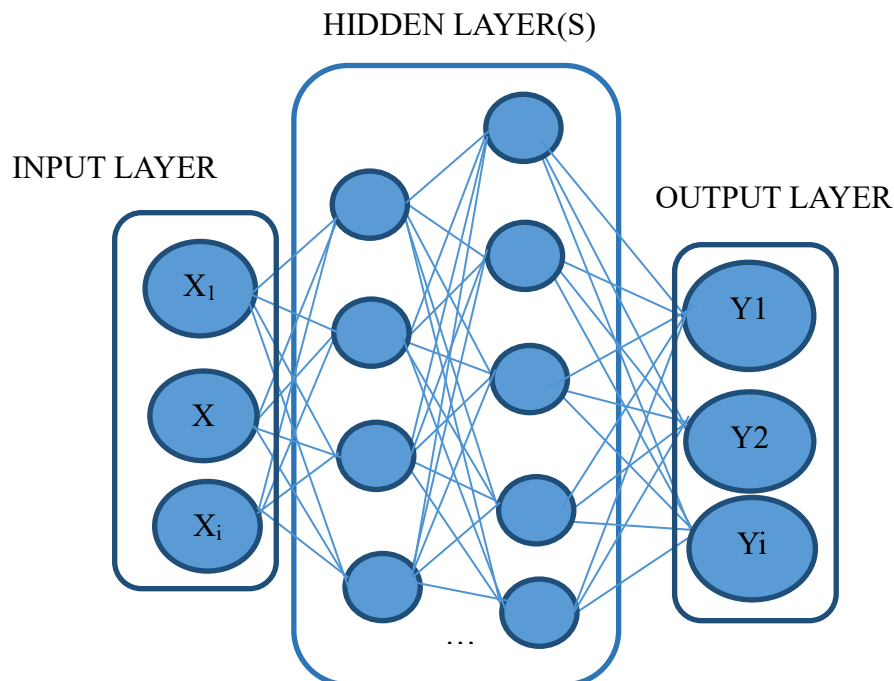
### 3.4. Machine learning models development

In the present study, three major steps were used for the development of non-linear models for predicting SPI. These steps include firstly the selection of input and target data for calibration and validation. Rainfall, estimated evapotranspiration, maximum and minimum temperature, relative humidity, evaporation, sunshine duration, and wind speed served as input, while computed one-month SPI served as output. A total of 660 months was used, where 75% was randomly selected as a training set, and the remaining 25% was used for the validation of models. Secondly, the model structure was selected, and parameters were estimated, while the third step involved the validation of the selected model and evaluation of the performance of models.

### 3.5. Artificial neural network (ANN)

According to Janaki (2013) [19], “an ANN is a non-linear mathematical model that has the ability to replicate detected properties of neuron systems and draws the analogies of adaptive biological learning system by using the learning rules and hybrid algorithm”.

A network may vary from single to multiple layers, but a network’s basic structure usually consists of three layers where data is given to the ANN network, the data is processed in hidden layers and the results of the input layer are generated in the output layer [19] as shown in figure 2. The backpropagation neural network with several numbers of neurons and hidden layers was adopted in this study.



**Fig. 2.** A typical Multilayered Neural Network Structure (Adapted from Saracoglu, 2008) [20].

### 3.6. Support vector machine (SVM)

SVM is a machine learning technique that has been widely used for the prediction of variables that involves non-linear relationships. It was introduced by Vapnik and Cortes in 1995. According to Cover’s theorem (1965), “a linear function  $f(.)$  can be formulated in the high dimensional feature space to represent a non-linear relation between the inputs ( $x_i$ ) and the output ( $y_i$ )”, this is further expatiated by [21]. This linear function is presented as follows:

$$y_i = f(x_i) = (w, \phi(x_i)) + b \quad (4)$$

where  $w$  and  $b$  are model parameters.

The type of SVM adopted for the present study is used for the regression problem, which is recognized as Support Vector Regression (SVR). According to Banafsheh and Mohsen (2011) [21], SVR finds an optimized solution to minimize equation 4.

$$\phi(w, \epsilon, \epsilon^*) = \frac{1}{2} w^T w + C \cdot \frac{1}{L} \sum_{i=1}^L (\epsilon_i + \epsilon_i^*) \tag{5}$$

Subject to  $((w, x_i) + b) - y_i \leq \epsilon + \epsilon_i$

$$y_i - ((w, x_i) + b) \leq \epsilon + \epsilon_i^*$$

$$\epsilon_i^* \geq 0$$

$$w \in X \quad \epsilon_i^* \in R \quad b \in R \quad i = 1, \dots, L$$

Where:

L: number of data points in the training dataset

C: model parameter

X<sub>i</sub>: feature space data points

w: optimization problem solution

ε<sub>i</sub>: model residuals (ε<sub>i</sub> = y<sub>i</sub> - f(x<sub>i</sub>))

ε<sub>i</sub> and ε<sub>i</sub><sup>\*</sup> “are positive slack variables and C is a positive real-valued and pre-specified constant” [21]. Zahraie *et al.* (2014) further posited that “the constant C determines the amount up to which deviations from ε are tolerated; while deviations above ε are denoted by ε<sub>i</sub>. whereas deviations below ε are denoted by ε<sub>i</sub><sup>\*</sup>. C which is always positive is the penalty parameter on the training error. Practically, selection of the kernel function K(x<sub>i</sub>, x<sub>j</sub>) = φ(x<sub>i</sub>)<sup>T</sup> · φ(x<sub>j</sub>) is enough for training the SVM”. The Kernel functions and their parameters, suggested by the previous researchers are displayed in table 2. A detailed explanation of SVM could be found in Chen (2015)[22]; Granata *et al.* (2016) [23] and Du *et al.* (2017)[24].

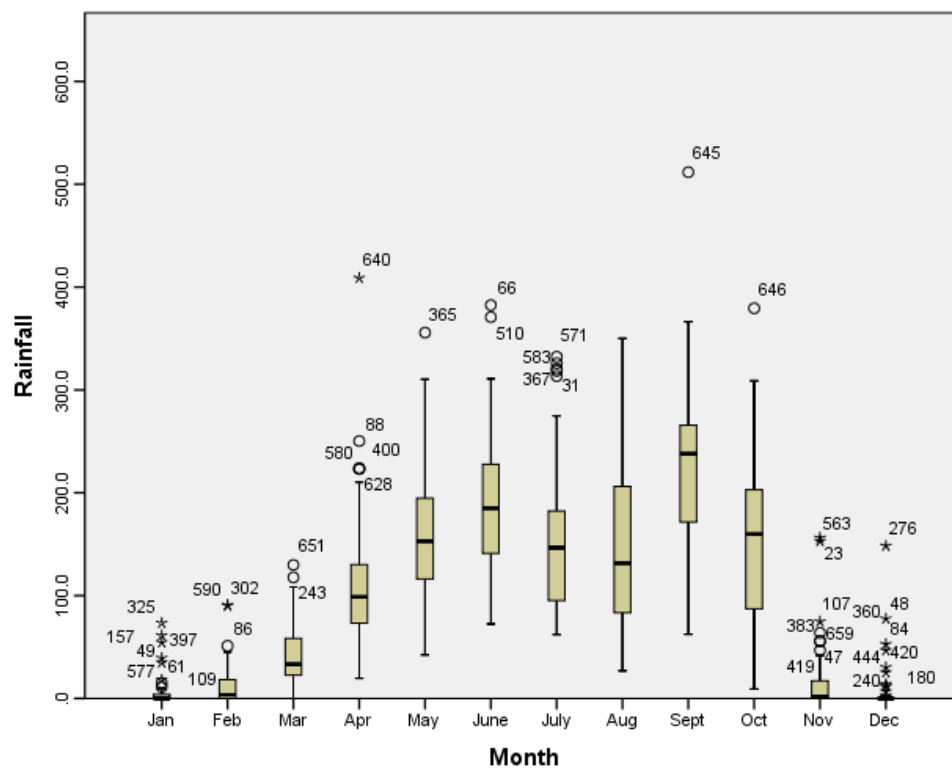
**Table 2**  
Kernel functions and Parameters (Zahraie *et al.*, 2014).

Kernel type	Kernel Function
Linear	K(x <sub>i</sub> , x <sub>j</sub> ) = (x <sub>i</sub> , x <sub>j</sub> )
Polynomial	K(x <sub>i</sub> , x <sub>j</sub> ) = ((x <sub>i</sub> , x <sub>j</sub> ) + C) <sup>p</sup>
Sigmoid	K(x <sub>i</sub> , x <sub>j</sub> ) = tan(b(x <sub>i</sub> , x <sub>j</sub> ) - c)
Radial Basis Function (RBF)	K(x <sub>i</sub> , x <sub>j</sub> ) = exp(-  x <sub>i</sub> , x <sub>j</sub>    γ)

## 4. Results

### 4.1. Descriptive analysis of hydro meteorological variables

The descriptive statistics of the hydro meteorological variables used in this study is represented by the box plots presented in figure 2 to figure 9. Figure 2 shows that rainfall was generally low between November and March (dry months) and substantial between April and October (wet months). As depicted in figures 3 and 4, the maximum temperature was generally experienced to be low in wet months and high in dry wet months, while the minimum temperature was lowest in the dry months. Figure 5 and 6 shows the distribution of relative humidity and sunshine duration, it could be observed that relative humidity was generally high in the wet months when sunshine duration was low due to high cloud cover and was found to be low in dry months when sunshine duration was relatively high due to low cloud cover and atmospheric water content. More so, it could be observed that wind speed does not follow the pattern of other hydro meteorological variables, however, its highest value was recorded in the month of May and lowest value in the month of January (figure 7). The recorded evaporation and the computed potential evapotranspiration followed similar patterns, with the lowest values in wet months (figure 8 and 9). This implies that more water is generally lost to the atmosphere by plants and open water bodies during the dry months in the study area.



**Fig. 2.** Box Plot of Monthly Rainfall Distribution.



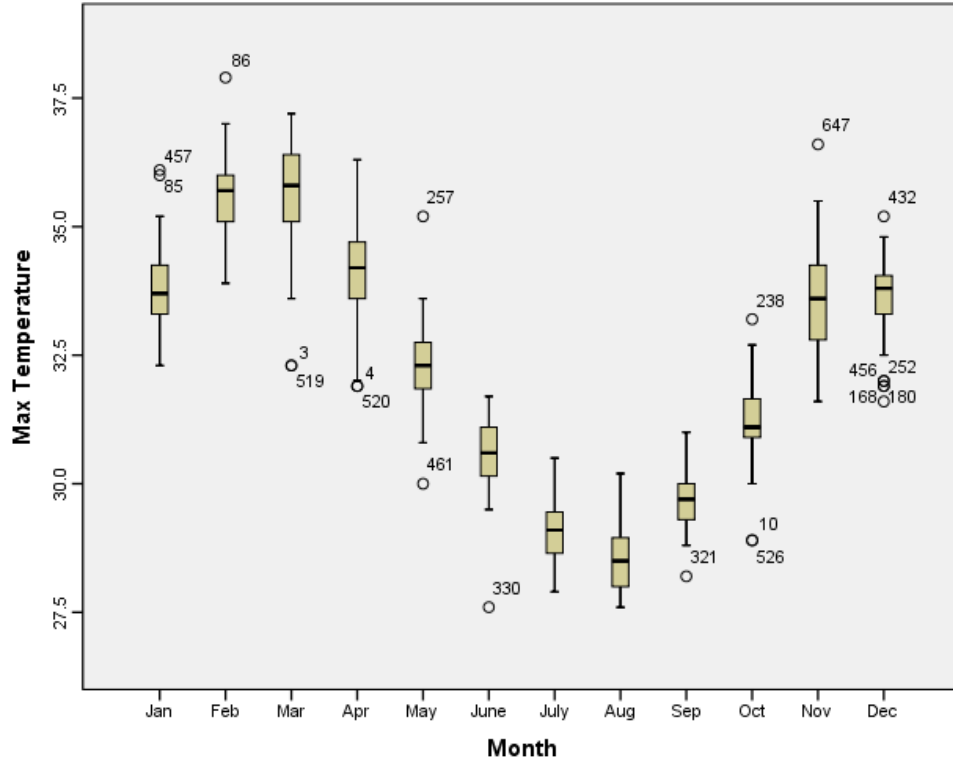


Fig. 3. Box Plot of Monthly Maximum Temperature Distribution.

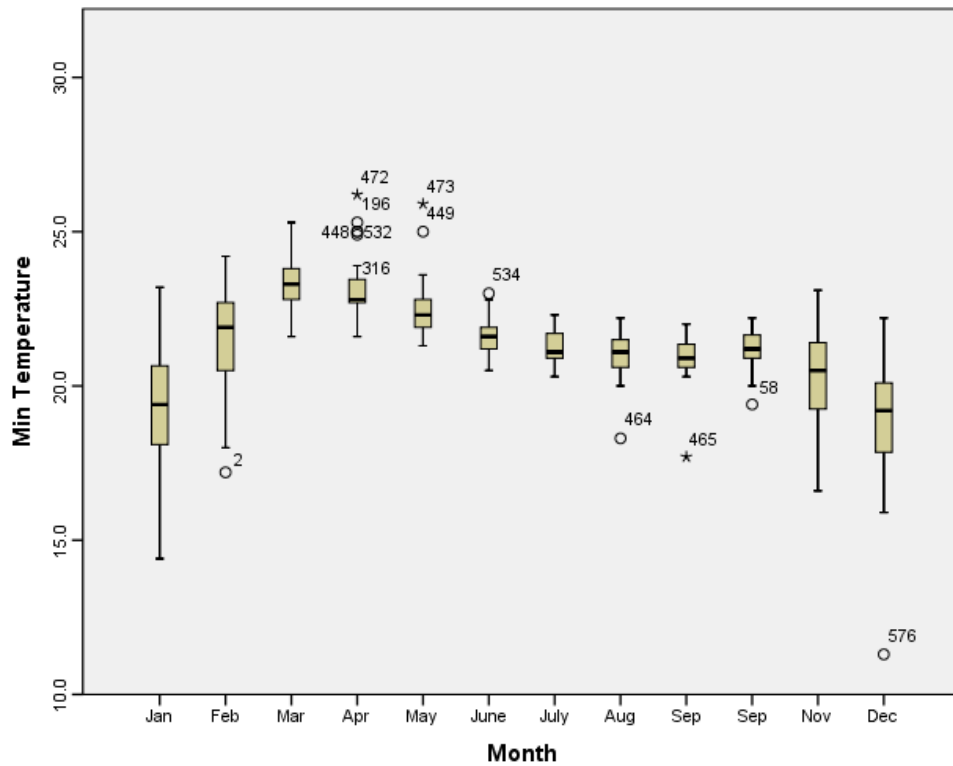


Fig. 4. Box Plot of Monthly Minimum Temperature Distribution.

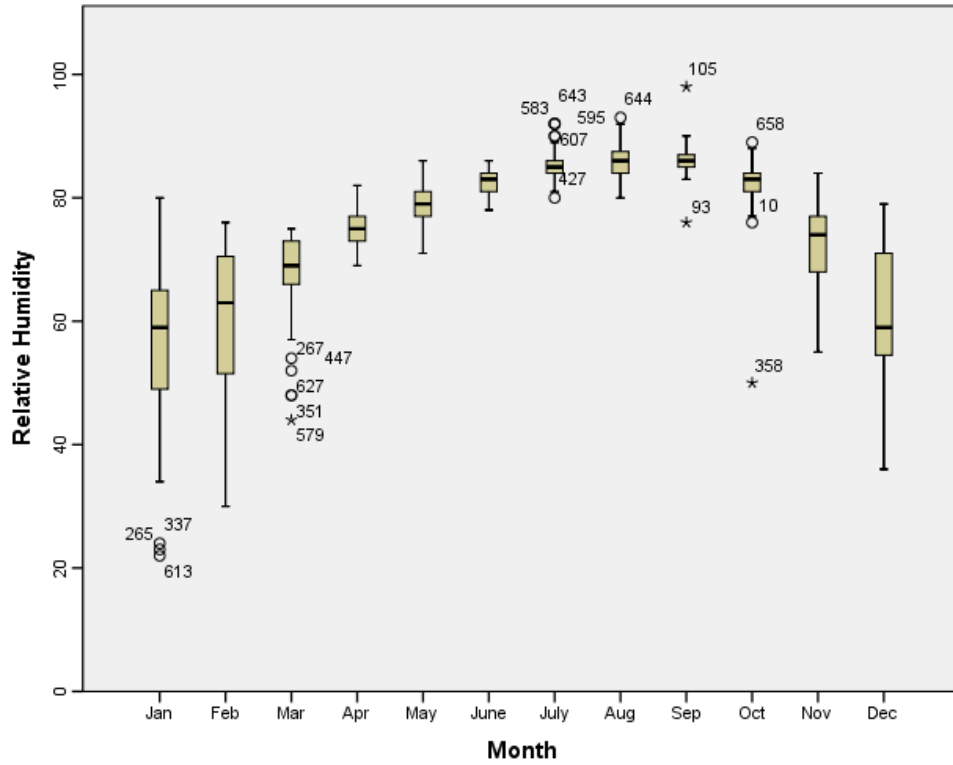


Fig. 5. Box Plot of Monthly Relative Humidity Distribution.

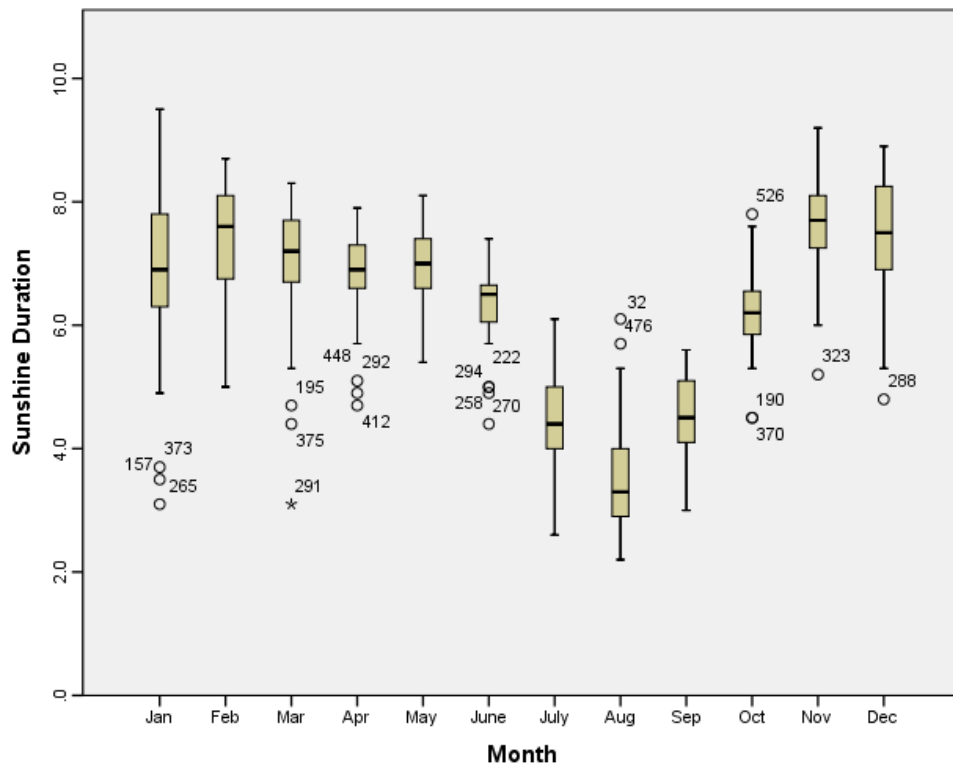


Fig. 6. Box Plot of Monthly Sunshine Hour Distribution.

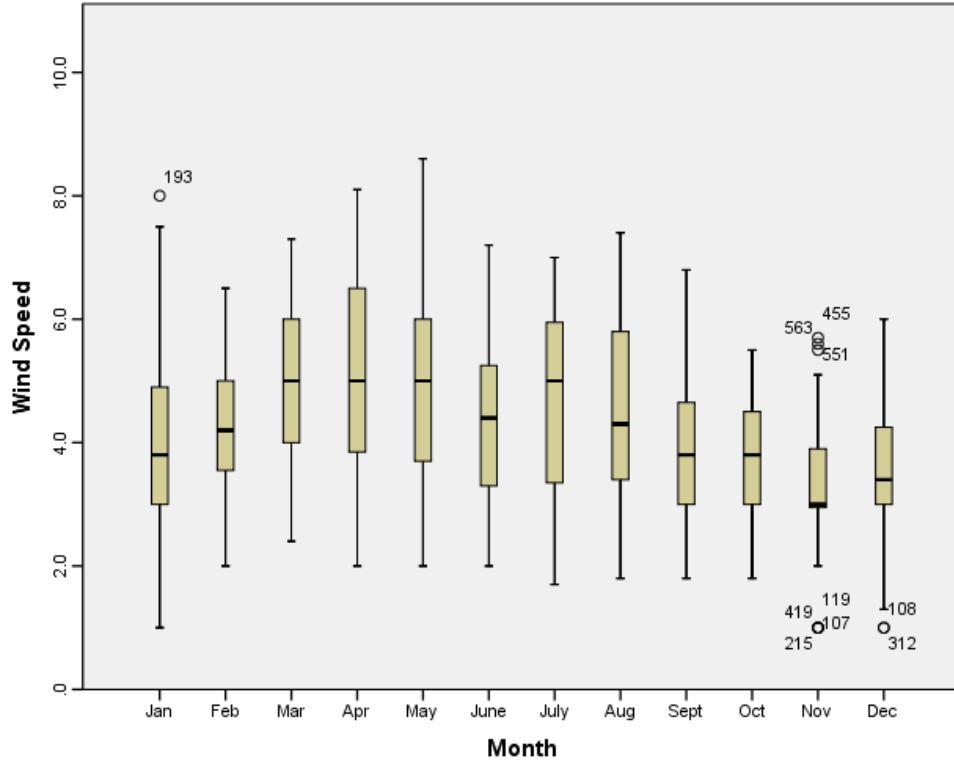


Fig. 7. Box Plot of Monthly Wind Speed Distribution.

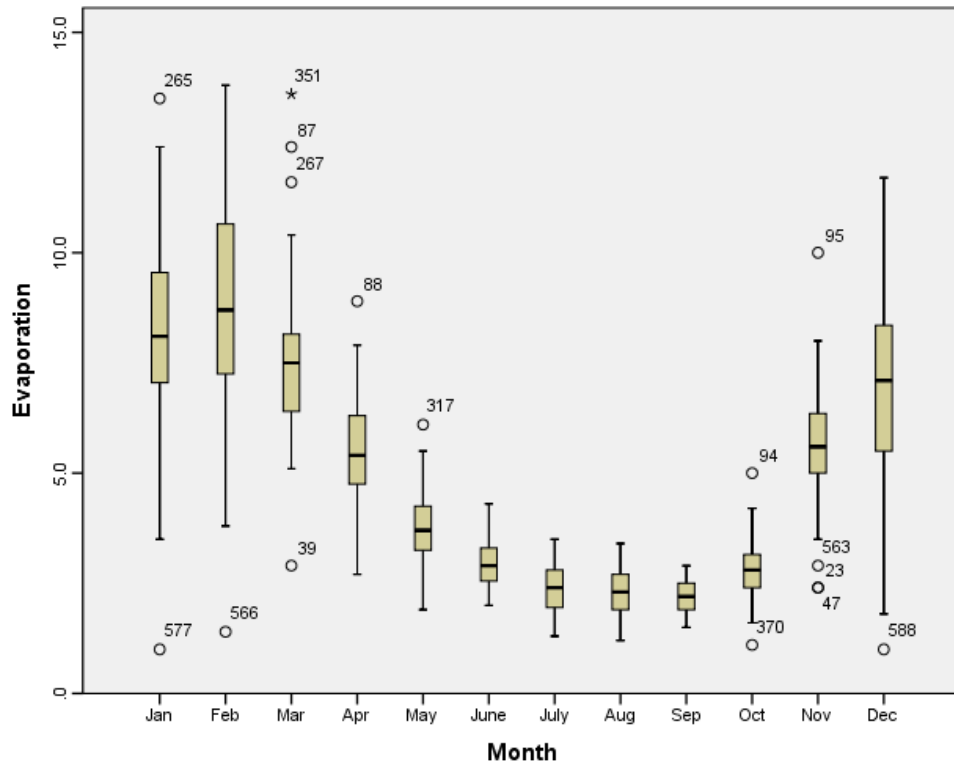
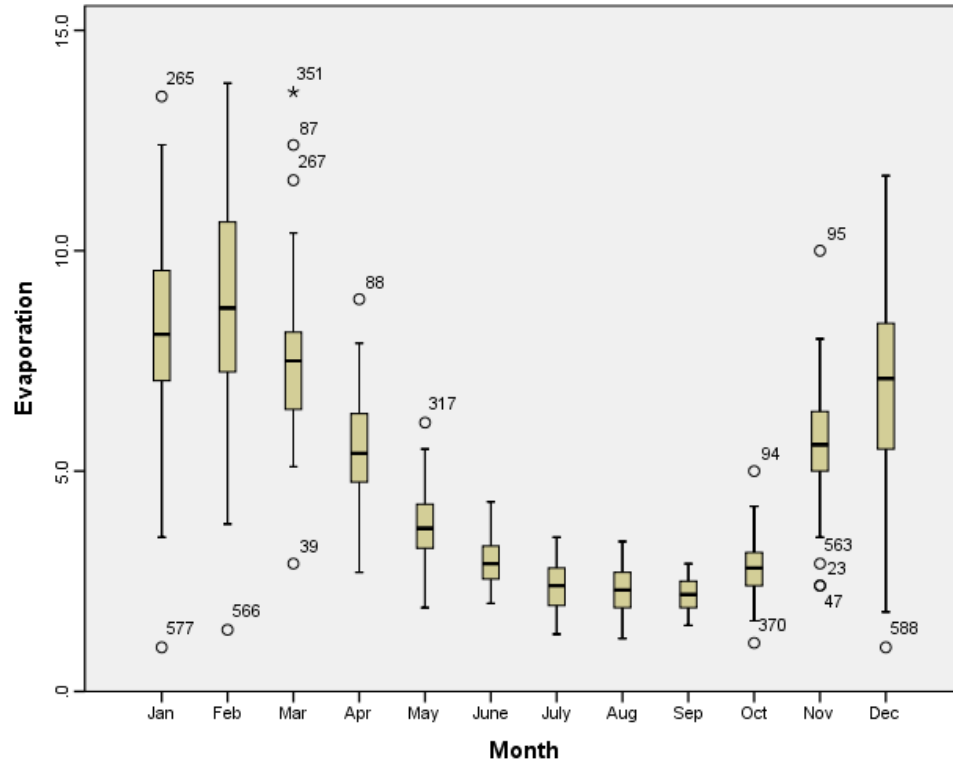


Fig. 8. Box Plot of Monthly Evaporation Distribution.



**Fig. 9.** Box Plot of Monthly Maximum Potential Evapotranspiration Distribution.

#### 4.2. Analysis of standardized precipitation index

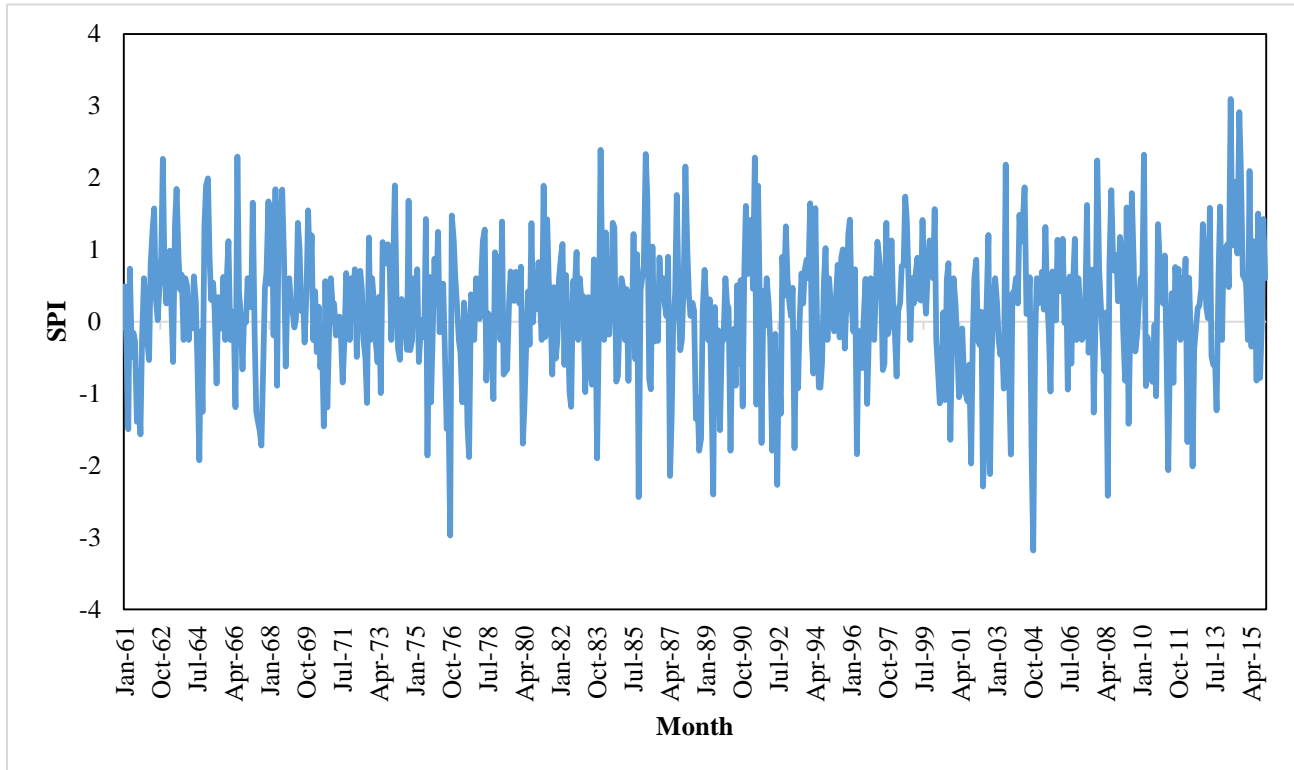
Table 3 shows the frequency of occurrence of the SPI category during the study period. The table shows that the majority of the months in the study period were near normal SPI (74.8%), while the remaining 25.2% of the study period were moderate to extreme conditions. As presented in table 3, only thirteen and eleven months experienced extreme wet and dry conditions respectively throughout the study period, indicating that extreme conditions are scarce and could cause havoc to water resource systems in the study area if not monitored.

**Table 3**

Frequency of Standardized Precipitation Index (SPI) Class (1961 -2015).

Category	SPI values	Frequency (months)	% Occurrence
Extremely wet	>2.0	13	1.9
Severely wet	1.50 to 1.99	30	4.6
Moderate wet	1.00 to 1.49	58	8.9
Near normal	0.99 to -0.99	490	74.8
Moderately dry	-1.00 to -1.49	31	4.7
Severely dry	-1.50 to -1.99	22	3.4
Extremely dry	< -2.0	11	1.7

Figure 10 shows the variability of computed one-month SPI of the study area, it could be observed that extreme wet conditions were experienced in 1962, 1966, 1983, 1986, 1988, 1991, 2003, 2007, 2010, 2014 and 2015, while extreme dry conditions were experienced in 1976, 1985, 1987, 1989, 1992, 2002, 2004, 2008, 2011, 2012.



**Fig. 10.** SPI Values Based on a One-month Time Scale.

#### 4.3. Mapping of hydro-meteorological variables to SPI using machine learning models

Eight hydro-meteorological variables that include rainfall, maximum temperature, minimum temperature, relative humidity, wind speed, sunshine duration, evaporation, and potential evapotranspiration were considered for simulation in the SVM and ANN model since these hydro-meteorological indices influence water resources management decisions.

After a series of trials and errors in the training and testing of multiple ANN models with different hidden layers, a most suitable model with the lowest mean square error was selected (figure 11). The ANN model consists of two hidden layers with eight and four nodes in each hidden layer (figure 11) was selected due to its ability to perform well with test data. Table 4 shows the performance of the models, by comparing the predicted SPI using the machine learning models and the computed SPI. It could be observed in the table that SVM with radial kernel performed better in predicting the SPI values of the training set than SVM with other kernels and ANN because of its lower mean square error (MSE) and root mean square error (RMSE) of 0.2151 and 0.4638 respectively; and high correlation and coefficient of determination (0.8563 and 0.7333 respectively). While ANN with two hidden layers performed best in

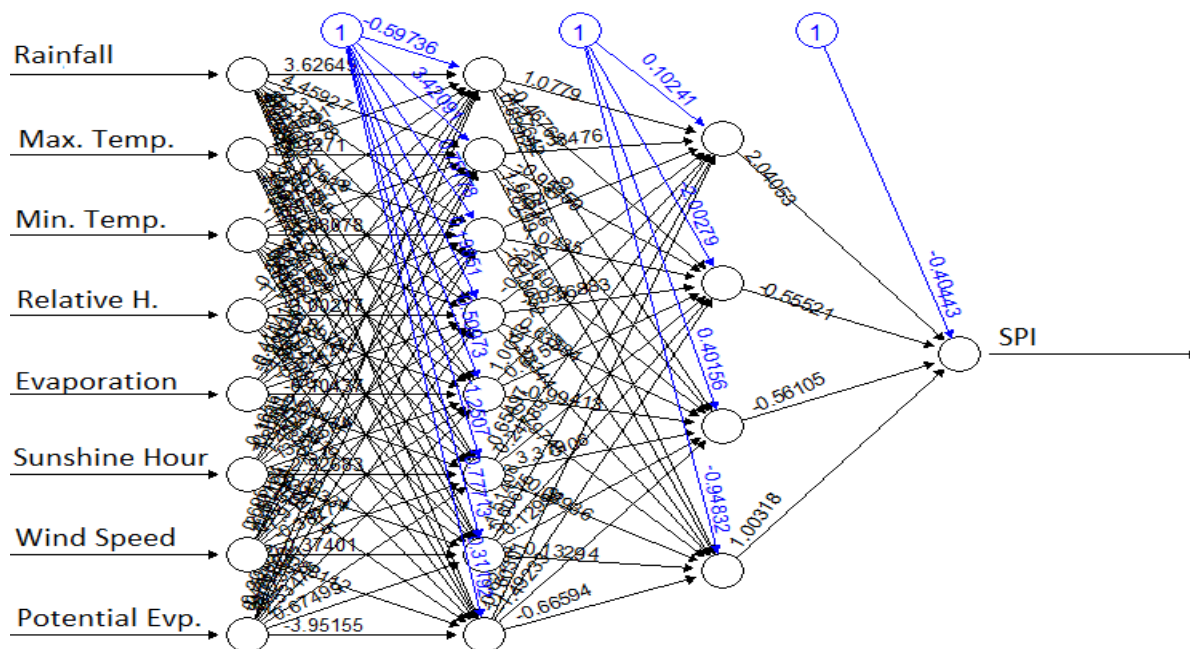
predicting the testing test than other models since it has the lowest MSE and RMSE of 0.3221 and 0.5675 respectively, and correlation and 0.8315 and coefficient of determination of 0.6914.

The fact that predicted and the computed SPI has a high collinear relationship (correlation is greater than 0.8) in both models annul the claim of superiority of one model over the other (table 4) in predicting dry and wet climate phases on the study area. Figure 12 and figure 13 show the relationship between the computed climate phases and those predicted by SVM with radial kernel and ANN with two hidden layers respectively. Both models were able to explain more than 60% of the computed SPI, whether for calibration or validations.

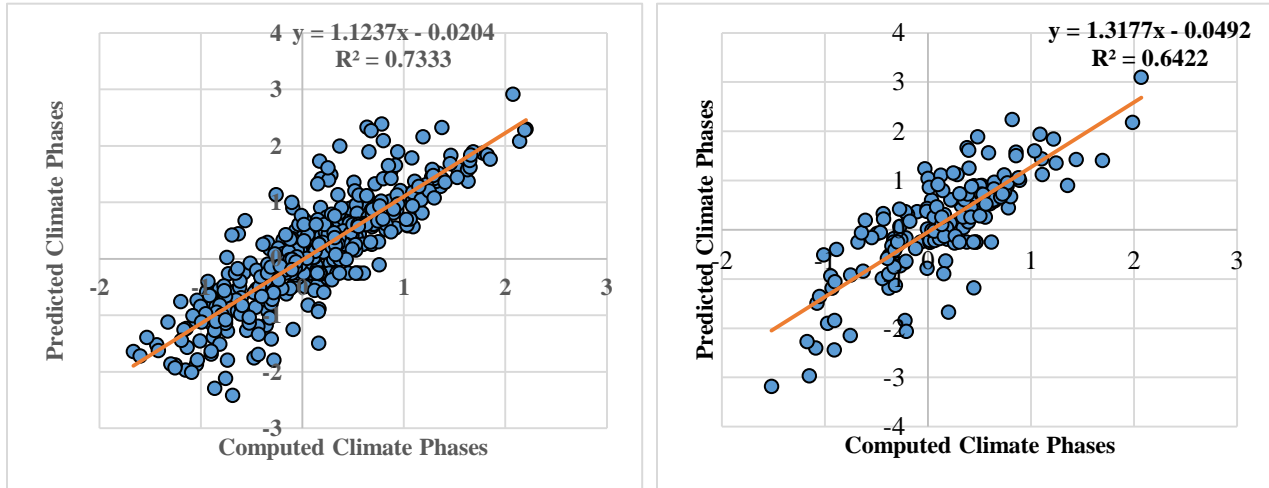
**Table 4**

Performance Results of Machine Learning Algorithms for Dry and Wet Climate Phases.

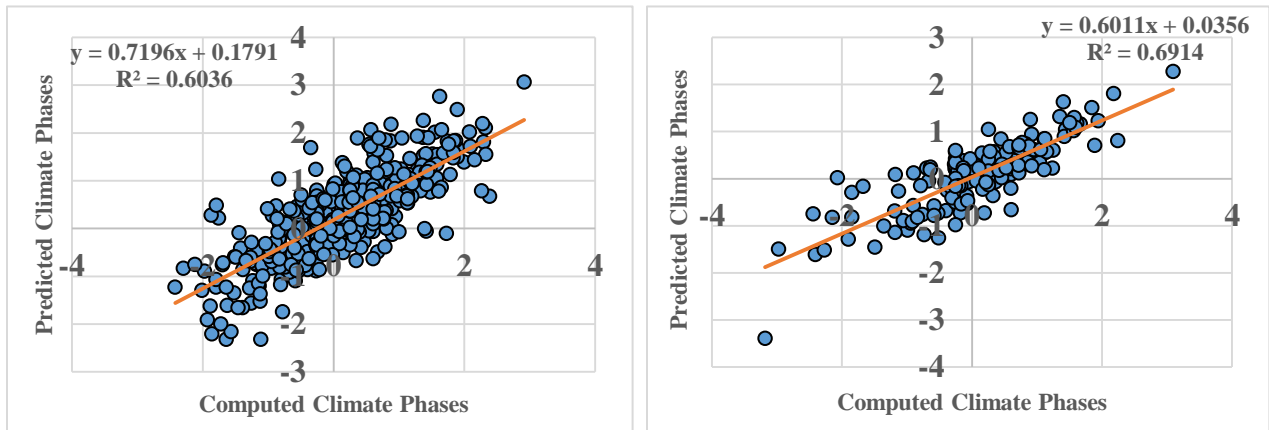
Model Type	Training				Testing			
	R	R <sup>2</sup>	MSE	RMSE	R	R <sup>2</sup>	MSE	RMSE
SVM-Radial	<b>0.8563</b>	<b>0.7333</b>	<b>0.2151</b>	<b>0.4638</b>	0.8014	0.6422	0.3970	0.6300
SVM-Linear	0.7125	0.5076	0.3849	0.6204	0.7331	0.5375	0.4795	0.6925
SVM-Sigmol	-0.1407	0.0198	24.1201	4.9112	-0.1318	0.0174	34.8234	5.9011
SVM-Polynomial	0.7003	0.4904	0.4057	0.6370	0.6005	0.3606	0.6549	0.8092
ANN (8-4-1)	0.7081	0.5015	1.5193	1.2326	0.8074	0.6519	0.3564	0.5970
ANN (8-8-4-1)	0.7769	0.6036	1.0317	1.0157	<b>0.8315</b>	<b>0.6914</b>	<b>0.3221</b>	<b>0.5675</b>



**Fig. 11.** Graphical Representation of ANN Model (Authors' Experiment).



**Fig. 12.** Performance of SVM in Predicting Climate Phases based on Calibration (Right) and Validation (Left).



**Fig. 13.** Performance of ANN in Predicting Climate Phases based on Calibration (Right) and Validation (Left).

## 5. Discussion

The suitability of extreme events prediction based on non-linear models is presented in this study. A time series of the data of 55 years covering 1961 to 2015, which constitutes 660 months was employed. Descriptive statistics of the hydro-meteorological variables reveals that rainfall was generally low between November and March (dry months) and substantial between April and October (wet months) when the temperature is lower due to due cloud cover, high atmospheric water content and low water loss from evaporation and transpiration. This corroborates with the findings of Olayemi et al. (2014), that the magnitude of the climatic variables, recorded in the area, falls squarely within ranges reported for the Tropics as a whole.

In many water resources engineering applications, linear empirical modeling has been identified to have many short comes. This is because these linear empirical models are established to predict future events that are stochastic in nature, and since they are developed from observed

data representing past events, the models become less reliable. In many instances, the observed data are finite and non-uniform forming only a sparse distribution in the input space (Yonas et al., 2000). Linear empirical models have been confirmed to offer acceptable predictive performance when dealing with linear or close to linear, but are usually unable to properly explain nonlinear patterns that are hidden in hydro-meteorological data. The need to develop reliable models for predicting water resources variables has aroused the interest of researchers in machine learning models, which was why the present study endeavors to assess the performance of ANN of different structures and SVMs of difference kernels in predicting dry and wet climate phases of the study area.

Machine learning models have taken the forefront of soft computing in field of hydrology and civil engineering; Sahai et al. (2000)[25] employed ANN in the prediction of seasonal rainfall in India, Toth et al. (2000)[26] combined ARMA model, ANN, and KNN to approximate runoff using rainfall, and discovered that ARMA outperformed other models, in contrast, Somvanshi et al. (2006) [27] noted that ANN outperformed ARIMA, and not ARMA for predicting rainfall series, while Karamouz et al. (2009)[28] compared ANN and Statistical Down-Scaling Model (SDSM) for predicting rainfall and concluded that the SDSM performance is better, although, is found to more data exhaustive model than ANN. Khalili et al. (2011) [29] got satisfactory results after applying ANN for rainfall prediction in Iran, while Geetha and Selvaraj (2011)[30] emphasized the inability of ANN to predict sharp peaks of monthly rainfall. More recently, Belayneh et al (2016) [31] emphasized that wavelet neural network was able to predict SPI more accurately than SVM and ANN,

Recently, due to Support Vector Machine's (SVM's) admirable features of being robust [24], water resources and environmental engineers have considered the SVM method for the modeling of water resources systems. For instance, Lu and Wang (2011)[32], Nayak and Ghosh (2013)[33], Ortiz-Garcia (2014)[34], Sanchez-Monedero (2014)[35], Jinglin et al. (2017) [24], Sehad (2016)[36] applied SVM in rainfall studies; while, Young (2017)[37] integrated physically-based models such as Hydrologic Modeling System (HEC-HMS) and data-driven (Support vector regression, SVR) models for the prediction of runoff in Taiwan.

In the present study it was unraveled that of all the kernel types considered for the SVM model, SVR with SVM with Radial kernel ( $R^2=0.64$ , RMSE=0.63) outperformed the SVM with linear, signal and polynomial kernels, while ANN with two hidden layers ( $R^2=0.68$ , RMSE=0.57) outperformed ANN model with a single layer ( $R^2=0.66$ , RMSE=0.64) in predicting dry and wet climatic phases. The inability of the models to perform more than 68% could be as a result of overfitting in ANN, and sensitivity to parameter section [38], which major shortcomings of the data-driven model. More-so, the hydrology system is influenced by many factors such as weather, land cover, infiltration, evapotranspiration, so it includes a good deal of stochastic dependent component, multi-time scale, and highly non-linear characteristics. Apart from the challenges of overfitting and parameter selection that could reduce the performance of the models, the models could be affected by hydrologic data signals that are usually nonstationary and their seasonal irregularity [39]. Despite this shortcomings, the models were found to perform better than those presented by Aiyelokun et al. (2017)[40], who used ANN to predict SPI for



drought investigation in part of the Tropics and unraveled that the ANN models had RMSE of 2.126 to 0.971, although the predictors were limited to just rainfall, mean temperature, relative humidity and evapotranspiration; in contrast to the present study that used eight predictors, which implies that the characteristics and number of input variables impact the performance of machine learning models.

The study further revealed that the performance of ANN and Radial kernel-based SVM were close to each other, this was found to be similar to the result obtained in Belayneh and Adamowski (2013), who stated that ANN and SVR had similar performance when predicting SPI of different time scales.

In this study, both SVM and ANN have been applied for the prediction of extremes in Minna, North Central Nigeria. However, based on the scope of this study, both SVM and ANN are adequate for the prediction of extreme events and are able to explain more than 60% of the computed SPI.

## 6. Conclusion

Water resource and environmental engineers need accurate information in harnessing water for diverse uses, therefore it expedient to predict extreme events in order to ensure water resource planning and management. This study assessed the dimensions of short-term dry and wet phases using SPI as well as predicting those using machine-learning techniques (SVM and ANN).

Based on the result of the study, the following key insights drawn from the study;

- i. Rainfalls were generally low between November and March and were high between April and October; this pattern was found to be exhibited by relative humidity. While other parameters were found to exhibit a reverse pattern, except wind speed;
- ii. extreme wet conditions were experienced in 1962, 1966, 1983, 1986, 1988, 1991, 2003, 2007, 2010, 2014 and 2015, while extreme dry conditions were experienced in 1976, 1985, 1987, 1989, 1992, 2002, 2004, 2008, 2011, 2012;
- iii. while normal meteorological conditions occurred for about 74.8% of the study period, 8.9%, 4.6% and 1.9% of this period were moderately wet, severely wet and extremely wet respectively, and 4.7%, 3.4% and 1.7% of the study period were moderately dry, severely dry and extremely dry respectively;
- iv. SVM with radial kernel performed better than SVM with Linear, Sigmoid and Polynomial kernels; while ANN with two hidden layers performed better than ANN with a single hidden layer.
- v. SVM with radial kernel performed best in predicting calibration or training data, while ANN performed best in Validation or Test data set.

Although, both SVM and ANN showed good predictability; future studies may focus on improving the performance of these models. This study serves as a baseline for the advancement of the application of machine-learning techniques in water resources management studies in Nigeria.

## Acknowledgment

The authors are grateful to the anonymous reviewers for the great contributions.

## References

- [1] Wang M, Zhang L, Baddoo TD. Hydrological modeling in a semi-arid region using HEC-HMS. *J Water Resour Hydraul Eng* 2016;5:105–15.
- [2] Vairavamoorthy K, Gorantiwar SD, Pathirana A. Managing urban water supplies in developing countries – Climate change and water scarcity scenarios. *Phys Chem Earth, Parts A/B/C* 2008;33:330–9. doi:10.1016/j.pce.2008.02.008.
- [3] Okeke OC, Uzoh OF. Towards achieving sustainable water resources management in Nigeria. *Glob J Geol Sci* 2009;7.
- [4] Akpor OB, Muchie B. Environmental and public health implications of wastewater quality. *African J Biotechnol* 2011;10:2379–87.
- [5] Gayan L, Upul S. Forecasting Droughts using Artificial Neural Networks. *Natl. Symp. Disaster Risk Reduct. Clim. Chang. Adapt.*, 2009.
- [6] Agbede OA, Aiyelokun OO. Establishment of a Stochastic Model for Sustainable Economic Flood Management in Yewa Sub-Basin, Southwest Nigeria. *Civ Eng J* 2016;2:646–55.
- [7] ConsortiumDolcine L, Prével C, Brahm A, Er-Raji A, Qaimi A. Implementation of an Integrated Decision Support System for the Souss-Massa Watershed, Morocco, in UNESCO/IHP, Application of Satellite Remote Sensing to Support Water Resources Management in Africa: Results From the TIGER Initiative, IHP-VII Technical 2010.
- [8] Bhuiyan C. Various drought indices for monitoring drought condition in Aravalli terrain of India. *Proc. XXth ISPRS Congr. Istanbul, Turkey, 2004*, p. 12–23.
- [9] Tallaksen LM, Van Lanen HAJ. Hydrological drought: processes and estimation methods for streamflow and groundwater. vol. 48. Elsevier; 2004.
- [10] Goddard S, Harms SK, Reichenbach SE, Tadesse T, Waltman WJ. Geospatial decision support for drought risk management. *Commun ACM* 2003;46. doi:10.1145/602421.602442.
- [11] Ayanshola A, Sule B, Salami A. Evaluation of Willingness to Pay for Reliable and Sustainable household Water Use in Ilorin, Nigeria. *Ethiop J Environ Stud Manag* 2013;6:754. doi:10.4314/ejesm.v6i6.6S.
- [12] UN-Water Report. *World Water Development Report 4. (2012). Managing Water Under Uncertainty and Risk (Vol. 1), Knowledge Base (Vol. 2) and Facing the Challenges (Vol. 3)*. Paris: UNESCO. 2012.
- [13] McKee TB, Doesken NJ, Kleist J. The relationship of drought frequency and duration to time scales. *Proc. 8th Conf. Appl. Climatol.*, vol. 17, Boston; 1993, p. 179–83.
- [14] Musa JJ. Environmental Appraisal Environmental Appraisal Appraisal of Drainage System Drainage System Drainage System in Minna. Minna, Minna, Niger State Niger State Niger State-Nigeria Niger Niger Int J Appl Biol Res 2012;4:85–93.
- [15] UNDP/NISEPA. Niger state framework for integrated sustainable waste management. Niger State Strategic Waste Management Framework (Unpublished) 2009.

- [16] Ishiaku I, Emigilati MA, Kuta GI, Usman BY, Hassan AB. An Assessment of Alternative Water Source for Domestic Used in Minna Metropolis, Niger State, Nigeria. *J Environ Earth Sci* 2014;4:19–23.
- [17] Doorenbosm J, Pruitt WO. Crop water requirement, irrigation and drainage. Pap No 1977.
- [18] Jennifer MJ. Faculty Investigator Evaluation of reference evapotranspiration methodologies and afsirs crop water use simulation model (Final Report). Department of Civil and Coastal Engineering University of Florida, Gainesville, Florida 2001.
- [19] Janaki B. Hydrological Modeling of a Typical Ungauged Basin Of Odisha 2013.
- [20] Saracoglu Ö. An Artificial Neural Network Approach for the Prediction of Absorption Measurements of an Evanescent Field Fiber Sensor. *Sensors* 2008;8:1585–94. doi:10.3390/s8031585.
- [21] Zahraie B, Nasser M. Basin scale meteorological drought forecasting using support vector machine (SVM). *Int. Conf. drought Manag. Strateg. Arid semi Arid Reg. Muscat, Oman, December, 2011.*
- [22] Chen S-T. Mining Informative Hydrologic Data by Using Support Vector. *Entropy* 2015;17:1023–41. doi:10.3390/e17031023.
- [23] Granata F, Gargano R, de Marinis G. Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with the EPA's Storm Water Management Model. *Water* 2016;8:69. doi:10.3390/w8030069.
- [24] Du J, Liu Y, Yu Y, Yan W. A Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms. *Algorithms* 2017;10:57. doi:10.3390/a10020057.
- [25] Sahai AK, Soman MK, Satyan V. All India summer monsoon rainfall prediction using an artificial neural network. *Clim Dyn* 2000;16:291–302. doi:10.1007/s003820050328.
- [26] Toth E, Brath A, Montanari A. Comparison of short-term rainfall prediction models for real-time flood forecasting. *J Hydrol* 2000;239:132–47. doi:10.1016/S0022-1694(00)00344-9.
- [27] Somvanshi VK, Pandey OP, Agrawal PK, Kalanker N V, Prakash MR, Chand R. Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J Ind Geophys Union* 2006;10:141–51.
- [28] Karamouz M, Rasouli K, Nazif S. Development of a Hybrid Index for Drought Prediction: Case Study. *J Hydrol Eng* 2009;14:617–27. doi:10.1061/(ASCE)HE.1943-5584.0000022.
- [29] Khalili. Rainfall prediction using artificial neural networks. Thesis of Master degree of Irrigation and drainage engineering. Faculty of Agriculture, Ferdowsi university of Mashhad (In Persian).2008. (6) (PDF) Daily Rainfall Forecasting for Mashhad Synoptic Stati. 2009.
- [30] Geetha G, Selvaraj RS. Prediction of monthly rainfall in Chennai using back propagation neural network model. *Int J Eng Sci Technol* 2011;3.
- [31] Belayneh A, Adamowski J, Khalil B. Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. *Sustain Water Resour Manag* 2016;2:87–101. doi:10.1007/s40899-015-0040-5.
- [32] Lu K, Wang L. A Novel Nonlinear Combination Model Based on Support Vector Machine for Rainfall Prediction. 2011 Fourth Int. Jt. Conf. Comput. Sci. Optim., IEEE; 2011, p. 1343–6. doi:10.1109/CSO.2011.50.

- [33] Nayak MA, Ghosh S. Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. *Theor Appl Climatol* 2013;114:583–603. doi:10.1007/s00704-013-0867-3.
- [34] Ortiz-García EG, Salcedo-Sanz S, Casanova-Mateo C. Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data. *Atmos Res* 2014;139:128–36. doi:10.1016/j.atmosres.2014.01.012.
- [35] Sánchez-Monedero J, Salcedo-Sanz S, Gutiérrez PA, Casanova-Mateo C, Hervás-Martínez C. Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal–ordinal support vector classifier. *Eng Appl Artif Intell* 2014;34:199–207. doi:10.1016/j.engappai.2014.05.016.
- [36] Sehad M, Lazri M, Ameer S. Novel SVM-based technique to improve rainfall estimation over the Mediterranean region (north of Algeria) using the multispectral MSG SEVIRI imagery. *Adv Sp Res* 2017;59:1381–94. doi:10.1016/j.asr.2016.11.042.
- [37] Young C-C, Liu W-C, Wu M-C. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Appl Soft Comput* 2017;53:205–16. doi:10.1016/j.asoc.2016.12.052.
- [38] Wu CL, Chau KW. Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *J Hydrol* 2011;399:394–409. doi:10.1016/j.jhydrol.2011.01.017.
- [39] Deka PC, Prahlada R. Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Eng* 2012;43:32–42. doi:10.1016/j.oceaneng.2012.01.017.
- [40] Oluwatobi A, Gbenga O, Oluwafunbi F. An Artificial Intelligence Based Drought Predictions In Part Of The Tropics. *J Urban Environ Eng* 2017;11:165–73.