Predicting SPI Drought Indicator Using Machine Learning Algorithms: Case study in Hiran Region, Somalia

Abdulhalim Hersi Farah
Somali National University

Abdikafi Elmi Abdishakur
Abrar University

Abdishakur Abdullahi Diiso
Jamhuriya University of Science & Technology

ISSN: 2096-3246
Volume 55, Issue 08, August, 2023

Predicting SPI Drought Indicator Using Machine Learning Algorithms: Case study in Hiran Region, Somalia

Mohamed Abukar Abdullahi[1,2], Abdulhalim H. Farah[2,3*], Abdikafi Elmi abdishakur[2,4], Abdulshakur Abdullahi Diiso[5], Mohamed Sheikh Abdisamad Sheikh Adam[6], Umer Abdulrahman Barkan[7]

Department of statistics and planning, Faculty of Economics SIMAD University, Mogadishu, Somalia[1]
Somali Institute for Water and Environmental Sustainability (SIWES), Mogadishu, Somalia[2]
Agricultural Engineering Department, Faculty of Agriculture and Environmental Science, Somalia National University, Mogadishu, Somalia[3]
Civil Engineering Department, College of Engineering and Technology, Abrar University, Mogadishu, Somalia[4]
Faculty of Engineering, Jamhuriya university for science and technology, Mogadishu, Somalia[5]
Netrich Technologies Center, Mogadishu, Somalia[6]
Crop production Departement, Faculty of Agriculture and Environmental Science, Somalia National University, Mogadishu, Somalia[7]

Corresponding Author: 2,3*

Keywords:
Standardized precipitation index, Drought forecast, Extreme learning machine, Random forest, Support vector regression

ABSTRACT

Drought is a recurring natural disaster that can cause significant damage to agricultural production, human livelihoods, and the environment. Drought forecasting is an important tool for managing and mitigating the impacts of drought. This study aimed to improve drought forecasting through the use of machine learning models. Specifically, the study evaluated the performance of three machine learning models, namely Extreme Learning Machine (ELM), Random Forest (RF), and Support Vector Regression (SVR), for forecasting Standardized Precipitation Index (SPI) drought. These models were trained using precipitation data of Hiran region, Somalia from 1980 to 2021, to evaluate their ability to accurately predict drought conditions. The results showed that the SVR model performed the best, with an R^2 value of 0.753, MAE of 0.344, and RMSE of 0.488. The ELM and RF models also performed well. The study highlights the potential of machine learning models to improve drought forecasting, and the importance of evaluating multiple models to select the one that performs best for a specific dataset.

1. INTRODUCTION

Drought is a natural disaster that occurs when there is a prolonged period of insufficient rainfall or other forms of precipitation [33]. It can be challenging to determine its length, intensity, and severity this causes a significant impact on the environment, agriculture, and communities [17], [48].

Drought has four main categories which are agricultural, meteorological, hydrological, and socioeconomic [45]. Precipitation is a typical indication of meteorological drought since it is connected to a lack of it [32]. A drought with insufficient groundwater and surface water supplies over the water system is called a hydrological drought. Streamflow, groundwater, and soil moisture can all give clues as to whether there is a hydrological drought [32], [60]. The emergence of meteorological and hydrological droughts that are related to a shortage of soil moisture and constrain plant growth are known as agricultural droughts [45]. A socioeconomic drought is described as a drought that influences both the supply and demand for economic products. The lines between various types of droughts are indistinct and occasionally interchangeable [45].

Drought and environmental degradation pose significant risks to the long-term development of the world special Africa, as a large portion of the African continent (12.933 million km^2) is classified as desert or dryland, making it particularly vulnerable to these drought issues [40]. These issues can lead to resource depletion and land degradation, which in turn can contribute to poverty, political instability, food insecurity, and economic downturns in the region [1]. The Horn of Africa is particularly prone to persistent droughts, which can harm the ecosystem and natural resources in the region [7].

For instance, Djibouti, Eritrea, Ethiopia, Somalia, and Kenya were all affected by one of the region's worst droughts in the summer of 2011 [23]. Because of this, annual agricultural production declined, the number of animals that died was high, and the price of food increased [20]. Somalia experienced famine as a result of the drought, unlike many East African countries, which were affected by the drought's effects on agricultural production. Ten million people were affected by severe food shortages as a result of the 2011 drought, and many of them perished from starvation [38].

The 2016–2017 drought in Somalia has had severe effects on the ecosystem. During the one-year drought, natural standing vegetation lost almost 68% of its area (113,282 km^2) or 18% of the country's entire landmass. Overall, the drought cost the environment around 564.8 million USD in losses and damages [56]

To overcome the negative impact of drought, forecasting and predicting drought becomes very essential since it plays a particularly vital role in risk reduction and management promptly [39], [55]. Accurate drought forecasting is a crucial aspect of drought management, as incorrect forecasting can lead to ineffective management and even harm the environment. So, there is a need for rapid, precise, and accurate drought forecasting models that can provide quantitative data on future drought risks [13], [47], [51]. These models can utilize a combination of input components or drought indices to accurately forecast droughts (R. [61]. 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 [52], [58]. ML models include artificial neural networks (ANN), Extreme Learning Machine (ELM), Random Forest (RF), and Support Vector Regression (SVR) among others. Several drought index (Dis) have been developed to monitor drought conditions (R. [61], among all, Standardized Precipitation Index (SPI) is a widely used drought indicator that is characterized by its statistical rigor and comprehensive approach [52]. Its simplicity and ease of use, as well as its independence from weather conditions, make it a valuable tool for drought prediction [27]. The SPI is calculated using a statistical analysis of precipitation data, and it is expressed as a standardized value on a scale from -3 to +3, with negative values indicating drought conditions and positive values indicating above-normal precipitation [27]. SPI is also often used in conjunction with other drought monitoring and forecasting tools, such as satellite imagery and hydrological models, to provide a more comprehensive view of drought conditions [33], [41], [44].

In the context of drought forecasting using ML, there have been several research papers that have compared the performance of different ML techniques using different DIS. [58] indicated that the ELM model forecasted better than RF, M5 Tree for predicting the SPI drought index at 3,6,9, and 12 months timescale. Other studies have found different results. [6] revealed that ANN is better than SVM in forecasting the drought index in Nigeria. Similarly, [63] in China compared several ML models and autoregressive integrated moving average (ARIMA) for predicting drought, the findings revealed that ARIMA statistical model forecasted better than wavelet neural network (WNN) and SVM. In contrast, [3] compared different ML adaptive neuro-fuzzy inference system ((ANFIS), ANN and SVN for drought forecasting in Algeria. The study indicated that SVM performed well compared to other models with R2 0.90. As we encountered yet, a few studies have investigated the occurrence of drought in Somalia [2], [46], Much previous research on drought in Somalia [30], [43] focused on crisis management rather than investigating drought occurrence patterns and geographical distribution. Others have investigated how drought affects fodder supply and drought risk reduction in urbanizing east African environments [26], [54]. So, it is necessary to develop drought forecasting and monitoring models that can provide timely measures to mitigate drought-related risks. Accurate drought forecasting is a crucial aspect of drought management, as incorrect forecasting can lead to ineffective management and even harm the environment. However, no study has been documented that utilizes and compares different ML approaches for drought forecasting in Somalia.

The present study aims to address the issue of drought forecasting in Somalia, giving attention to the importance of long-term forecasting for the planning and management of natural droughts. The objective of this study is to analyze the accuracy and usefulness of the three machine learning models - ELM, RF, and SVR for predicting drought in Somalia.

2. Study area

The Hiran region of Somalia is located in the central part of the country and is known for its long periods of drought as well as flash floods (Fig. 1). The region is primarily rural and agricultural, with a large portion of the population dependent on farming and pastoral activities for their livelihood. The area's economy heavily relies on rain-fed agriculture and livestock, so the lack of sufficient and predictable rainfall is one of the main reasons for poverty and food insecurity in the region. The region also has a semi-arid and arid climate. The irregularity of rain, along with its high variability, makes drought and flash floods in the area common. This also results in a higher risk of failure of agricultural and pastoral production, which can lead to food insecurity.

The Hiran region is a typical example of the complexity of the drought-flood cycles that affects many semi-arid and arid regions, in which drought and flood hazards often co-occur, requiring an integrated approach that considers both.

3. METHODS AND MATERIALS

3.1 Data Sources

Statistical drought prediction needs long-term records spanning at least 30 years are necessary which does not accommodate data gaps. Remote sensing data was used in this study rather than using ground station data due to insufficient ground station data in Somalia. The current ground station data in Somalia was established in 2002 by Food and Agriculture Organization Somali Water and Land information management (FAO SWALIM). This data have no consistent temporal and spatial distribution throughout the country, this limited the usage of ground station data for training machine learning models. To tackle this problem, satellite data was used to forcest drought in the Hiran region, Somalia. The CHIRPS data is an

IR-based weather precipitation dataset having high-level spatial resolution (0.05° × 0.05°) with long-term records (1981–present), the dataset produces three outputs including gauge observations having different temporal resolutions namely, yearly, monthly, and daily, satellite estimates and global climatology [21]. Many studies indicated that the CHIRPS satellite dataset is suitable for describing the spatial distribution of precipitation and able to capture the occurrence and characteristics of drought events, suggesting that the CHIRPS dataset could be used as an alternative precipitation source for monitoring drought [22], [42], [50], [57]. Furthermore, there were a significant number of studies used CHIRPS as data set for machine learning models [5], [10], [24], [36], [37], [53], [59], [62]. In this study, the CHIRPS precipitation dataset for 1981–2021 was obtained from CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center - UC Santa Barbara (ucsb.edu).

Table 1 she classification of drought index SPI distribution [28].

SPI value	Drought category
(2.0, þ∞)	Extremely wet
(1.5, 2.0]	Severely wet
(1.0, 1.5]	Moderately wet
(–1.0, 1.0]	Normal
(–1.5, –1.0]	Moderately drought
(–2.0, –1.5]	Severely drought
(–∞, –2.0]	Extremely drought

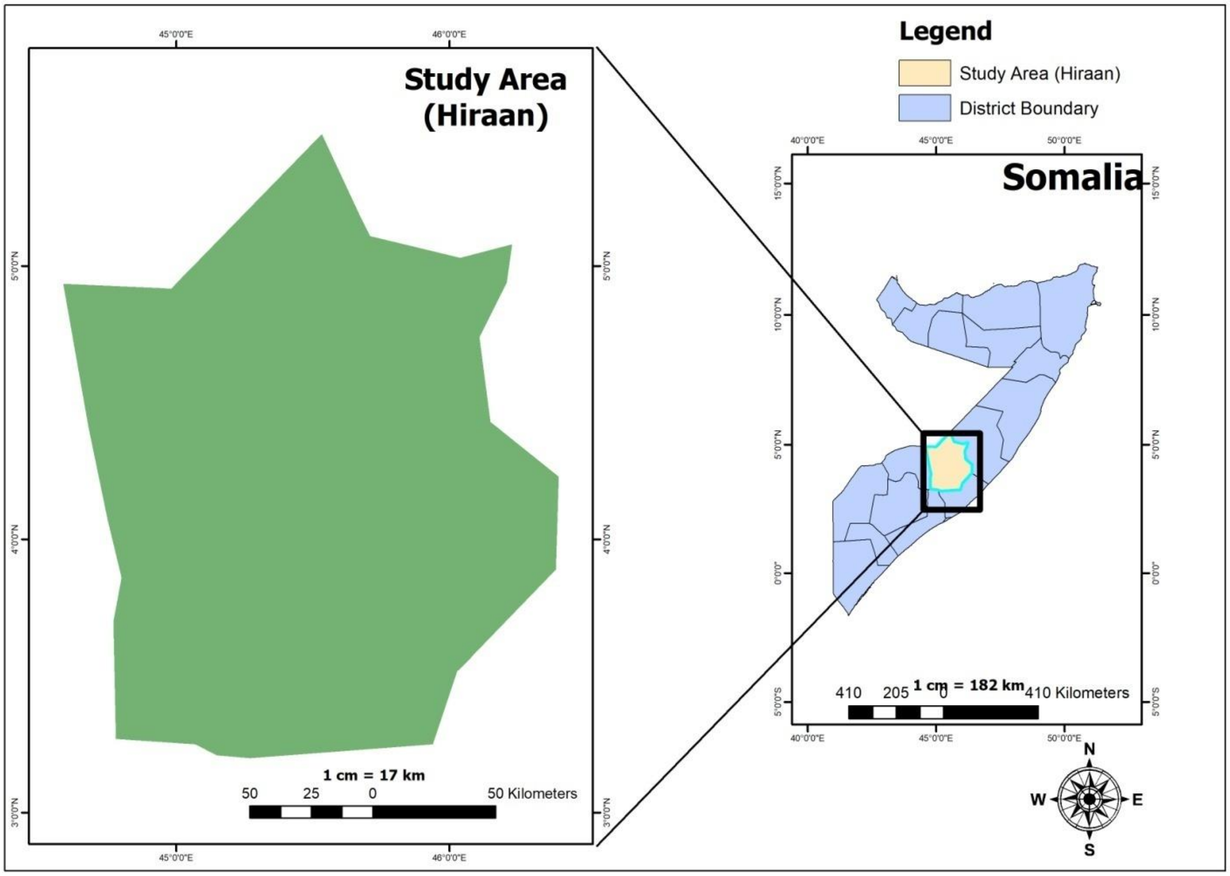


Fig. 1 Hiran region, Somalia

3.2 Methods

3.2.1 SPI Computation

The Standardized Precipitation Index (SPI) is a widely used method for measuring droughts and has been implemented in numerous drought forecasting scenarios using machine learning algorithms [28]. The World Meteorological Organization recommends the SPI as it provides a consistent measure of precipitation deficiency [31]. In this study, the SPI was used to measure droughts in the Hiran region, Somalia using the CHIRPS dataset over the past 40 years (1981-2021). The SPI is calculated by fitting the precipitation data series to a gamma probability density function and then transforming it to a regular distribution using an inverse normal function to yield the cumulative probability [14], [34]. This method allows for a consistent and reliable assessment of drought conditions in a specific region.

The SPI values can then be set as follows:

$$SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right);\ t = \sqrt{ln \frac{1}{(H(x))^2}}\ ;\ 0 < H(x) \le 0.5 \tag{1}$$

$$SPI = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3};\ t = \sqrt{ln \frac{1}{(1-H(x))^2}}\ ;\ 0.5 < H(x) \le 1 \tag{2}$$

where "x" represent monthly rainfall, the values of c_0, c_1 and c_2 are (2.515517, 0.802853, and 0.010328) respectively and the values of d_1, d_2 and d_3 are (1.432788, 0.189269 and 0.001308) respectively. "H(x)" represents the average probability of the series in the gamma distribution function [49]. The gamma distribution is expressed as follows:

$$g(x) = \frac{1}{\beta^{\alpha}\square(\square)} x^{\alpha-1} e^{-x/\beta^{-1}}; x > 0 \qquad (3)$$

where "α" and "β" represent the shape and the scale parameters respectively, and Γ(α) is the main function in the gamma distribution.

The study calculated several SPI time scales (1,3,6,9,12). The drought severity, moderateness and wetness SPI value are shown in Table 1 [14], [28].

3.2.2 Extreme Learning Machine (ELM)

Extreme learning machine (ELM) is a type of artificial neural network that was introduced by [25]. It is a single-hidden-layer feedforward neural network that has been shown to have good generalization performance and fast training times compared to other types of neural networks.

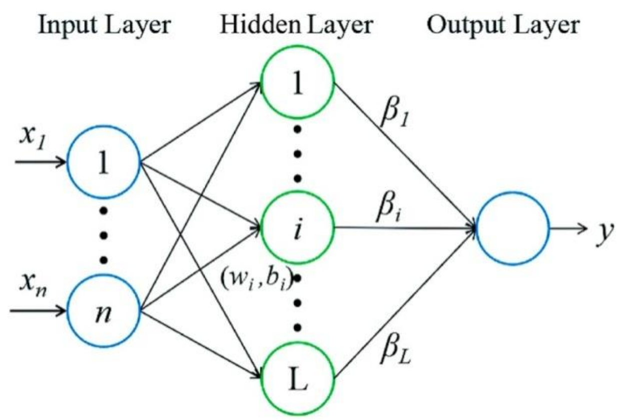


Fig. 2 ELM network structure [25]

ELM is a feedforward neural network that consists of an input layer, one or more hidden layers, and an output layer. It differs from traditional feedforward neural networks in that the weights of the hidden layer are randomly generated and not adjusted during the training process. Instead, the output weights are learned through a least-squares optimization process. This allows ELM to achieve faster training times compared to traditional neural networks, as the optimization process is simpler and requires fewer iterations [28]. One of the main advantages of ELM is that it can achieve good generalization performance with relatively small training datasets. It has also been shown to be robust to overfitting and can handle high-dimensional inputs effectively. However, ELM can be sensitive to the choice of hidden layer parameters and may not always outperform traditional neural networks in terms of accuracy [28].

The mathematical function of the ELM model may be represented as follows [28]:

$$\sum_{i=1}^{M} \beta i g(x_n,; bi, wi) = yn; n = 1, 2, 3, \ldots, \mathrm{N} \qquad (4)$$

According to the above ELM network equation, 'bi' represents the randomly assigned bias of the i^{th} hidden node, while 'wi' is the randomly assigned weight of the input vector connecting the ith hidden neuron to the input data. The weight vector connects the ith hidden node to the output neuron, and 'g(X_n ; bi , wi)' represents the output result of the x_n input sample associated with the ith hidden node. Each input is randomly distributed to the hidden nodes in the ELM networks. Then, it is possible to write Eq. (4) as:

$$\boldsymbol{H\beta = Y} \tag{5}$$

Where:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{g(x_1,;\ b_1, w_1)} & \dots \boldsymbol{g(x_1,;\ b_M, w_M)} \\ \boldsymbol{g(x_1,;\ b_1, w_1)} & \dots \boldsymbol{g(x_1,;\ b_M, w_M)} \end{bmatrix} \boldsymbol{N \times M} \tag{6}$$

$$\boldsymbol{H\beta} = (\boldsymbol{\beta_1^T, \beta_2^T, \dots, \beta_L^T})_{\boldsymbol{m \times M}}^{\boldsymbol{T}} \tag{7}$$

$$\boldsymbol{Y} = (\boldsymbol{t_1^T, t_2^T, \dots, t_L^T})_{\boldsymbol{m \times M}}^{\boldsymbol{T}} \tag{8}$$

According to [28], "H" represents the output matrix of the hidden layer and "T" represents the label matrix. is the output weights acquired by finding the least square solutions to the above-mentioned linear system [18]:

$$\boldsymbol{\beta = H^{+}Y} \tag{9}$$

[18] defines “H+“, as H’s matrix generalized inverse of Moore-Penrose.

3.2.3 Random Forest (RF)

Random forests are a type of machine learning algorithm that can be used for tasks involving classification and regression. They are made up of multiple decision trees that work together to make a prediction. This type of model is known as an ensemble model. The decision trees in a random forest are trained on a random subset of the data, and the final prediction is made by combining the predictions of all the trees through a voting process. Using multiple decision trees helps to reduce overfitting and improve the model's ability to generalize to new data. Essentially, each tree gets a "vote" on the final prediction, and the class or value that gets the most votes is chosen as the final prediction [11].

One of the main advantages of random forests is that they are very effective at handling large and complex datasets. They are also relatively simple to implement and can be used for both classification and regression tasks. They can also handle missing values and categorical variables, which makes them a useful tool for many different types of data [11]. However, one potential disadvantage of random forests is that they can be computationally intensive, particularly when the number of trees in the forest is large. This can make them less efficient than other models in some cases. Additionally, random forests can sometimes have difficulty with imbalanced datasets, where one class is significantly more prevalent than the other [11]. Overall, random forests are a powerful and widely-used machine learning technique that can be applied to a wide variety of problems.

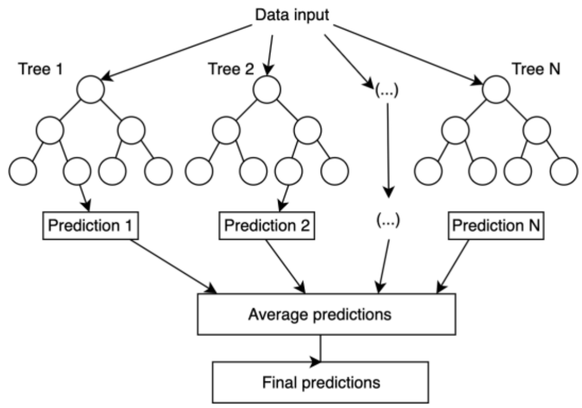


Fig. 3 Random Forest Sample Trees

3.2.4 Support Vector Machine (SVM)

Support vector machines (SVM) are a type of supervised machine learning algorithm that can be used for both classification and regression tasks [15]. When used for regression, the SVM algorithm is called support vector regression (SVR) [15]. SVR works by finding the hyperplane in a high-dimensional feature space that maximally separates the data points of different classes (in the case of classification) or that best fits the data points (in the case of regression) [15]. The distance of the data points from the hyperplane, called the margin, is maximized in the process of finding the optimal hyperplane. The data points that lie on the margin, or close to it, are called support vectors [15]. One of the main advantages of SVR is that it can handle non-linear relationships between the input features and the target variable [12]. This is achieved by using a kernel function, which transforms the data into a higher-dimensional space where a linear separation is possible. Commonly used kernel functions include the linear, polynomial, and radial basis function (RBF) kernels [12].

The relationship among the input and output is in the equation [29]:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{w} * \boldsymbol{xj} + \boldsymbol{b} \tag{10}$$

In this equation, w is the weight, b is the bias, and xj is the input features. To determine the values of w and b, the following optimization problem is solved:

$$\boldsymbol{minimize}\ \frac{1}{2}||\boldsymbol{w}||^2 + \boldsymbol{C}\sum_{i=1}^{l}(\boldsymbol{\xi}\,\mathbf{i} + \boldsymbol{\xi}\,\mathbf{i}\ *) \tag{11}$$

$$subject\ to \left\{ \begin{array}{l} \mathrm{yi} - < w, xj > - b \leq \varepsilon + \xi i \\ < w, xj > + b - yi \leq \varepsilon + \xi\mathrm{i} * \\ \boldsymbol{\xi}\,\mathbf{i}, \boldsymbol{\xi}\mathbf{i} * \geq \mathbf{0}, \mathbf{i} = \mathbf{1}, \mathbf{2}, \mathbf{3} \ldots \mathbf{N} \end{array} \right\} \tag{12}$$

In this equation, C is a constant, ξi and ξi* are slack variables, and ε is the boundary value. Several types of kernels can be used in Support Vector Regression, including linear, polynomial, radial basis function (RBF), and sigmoid. The present study used RBF kernel function.

ISSN: 2096-3246
Volume 55, Issue 08, August, 2023

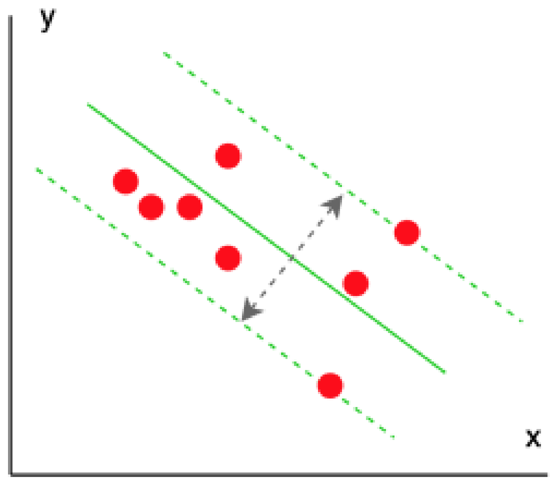


Fig. 4 Support Vector Machine (figure # is not in text)

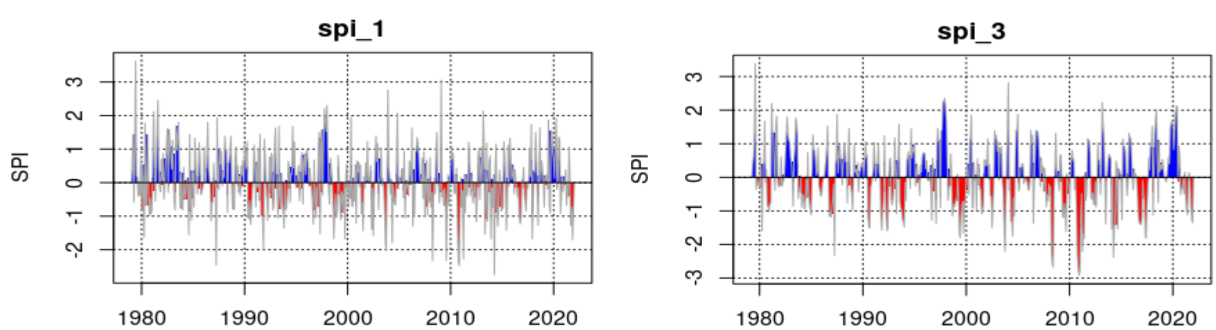
3.2.5 Criteria of Performance Evaluation

Various performance criteria were used in this study for machine learning models comparison including, root-mean-square error (RMSE), determination coefficient (R^2) and mean absolute error (MAE). These criteria are calculated according to following Equations [4], [16]

$$\mathbf{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{SPI12\ Observed} - \mathbf{SPI12\ Predicted})} \tag{13}$$

$$\mathbf{MAE} = \frac{1}{N}\sum_{i=1}^{N}|(\mathbf{SPI12\ Observed} - \mathbf{SPI12\ Predicted})| \tag{14}$$

$$\mathbf{R^2} = \mathbf{1} - \frac{\sum_{i=1}^{N}(\mathbf{SPI12\ Observed} - \mathbf{SPI12\ Predicted})}{\sum_{i=1}^{N}(\mathbf{SPI12\ Observed} - \mathbf{SPI12\ mean})} \tag{15}$$



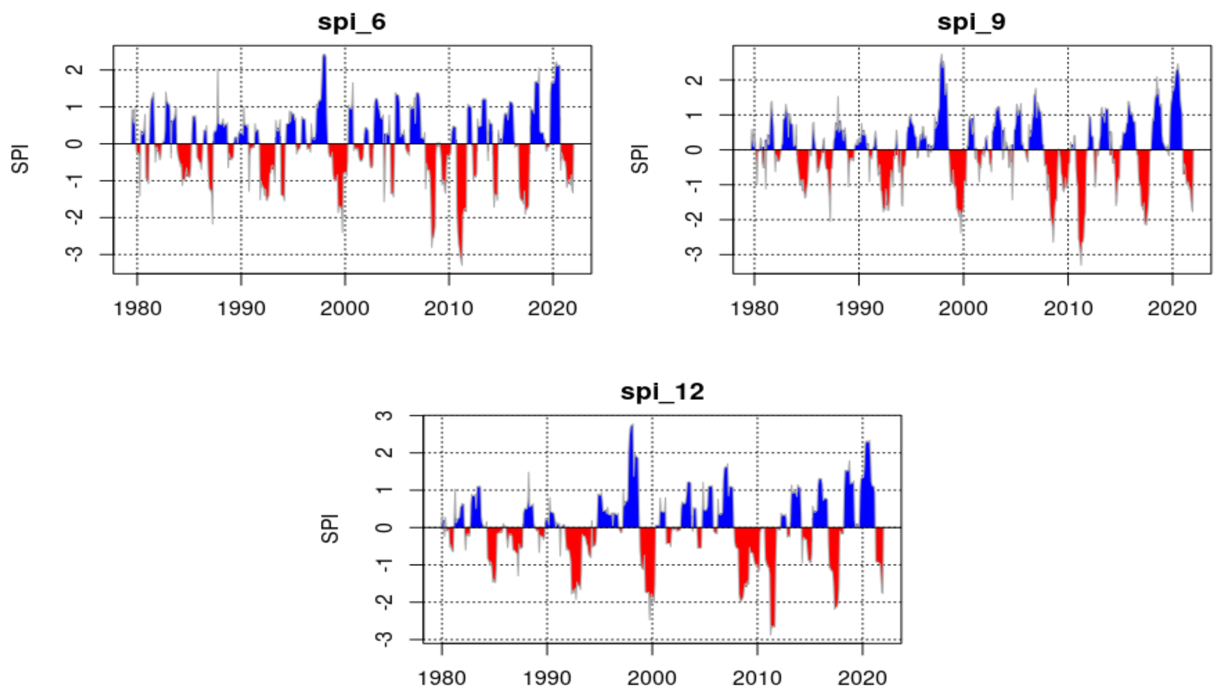


Fig. 5 (a) SPI_1, (b) SPI_3, (c) SPI_6 (d) SPI_9 (e) SPI_12

4. RESULTS AND DISCUSSION

4.1 Drought intensity, duration and frequency using SPI.

Drought index such as SPI are good for analysing drought distribution according to their location and time occurred. In this study, five timescales (in months) were chosen for drought evaluation i.e., SPI-1, SPI-3, SPI-6, SPI-9 and SPI-12. The various timescales taken quantify the precipitation deficit in the study area which indicate the effect of drought on available water resource forcing decision makers to take immediate mitigation actions. Fig. 5 (a to e) shows the duration, sequence, and severity of drought, the blue colour indicates wet events while red colour shows dry period. Fig 5(a & b) indicates SPI at 1and 3-month timescale. This timescale is called short-term timescale [9] the precipitation variability is high due to fewer months data compared with SPI-6,9,12 months Fig. 5(c,d and e). In SPI 6,9, and 12 months, the accumulation values of SPI are 6, 9 and 12 months, respectively which results smoother time-series with less variation. Generally, lead time or scale time is direct proportional with drought duration and inversely proportional with drought frequency and intensity [35]. Based on SPI-12, the extremely drought event occurred (SPI= -2.78) between April-September 2011, while in SPI-6 and SPI-9 the extremely drought event observed in November 2010 to April 2011 and January 2011 to July 2011, respectively with same SPI value (SPI= -3.2) (Fig 5, c-e). This indicates that detection of drought depends on timescale used. SPI managed to identify drought events. For instance, SPI-12 effectively identified April to September 2011 drought event as other studies indicated so [8], [19].

4.2 Machine learning models comparison

This study evaluates the performance of three machine-learning i.e. ELM, RF and SVR model in predicting monthly long-term standardized precipitation index SPI-12. The study used SPI time scales (SPI-1, SPI-3, SPI-6 and SPI-9) from 1980 to 2021 as input variables of the trained model to predict SPI-12. The monthly SPI data is partitioned into 70% and 30% for training and testing sets, respectively, the training set is

ISSN: 2096-3246
Volume 55, Issue 08, August, 2023

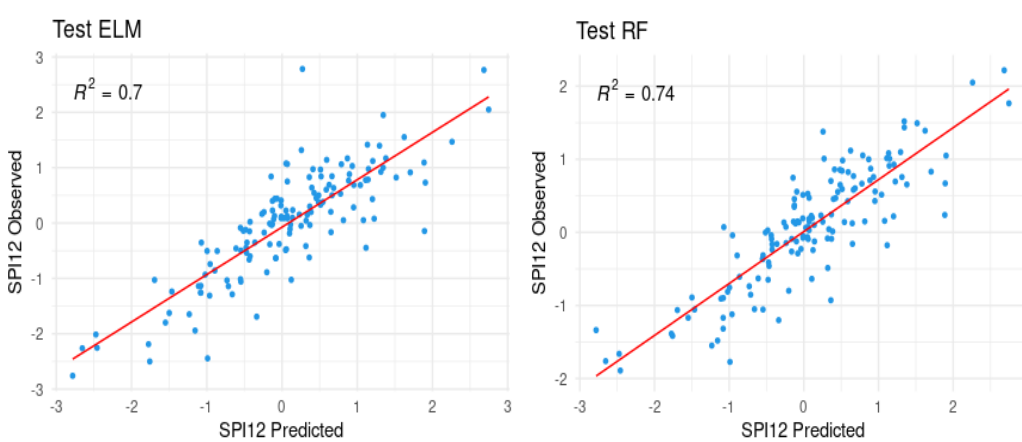
employed to train ELM, RF, and SVR models. In addition, the study assessed the performances of the three ML ELM, RF and SVR models in predicting SPI-12 using MAE, RMSE, and R^2 as indicator parameters. For SPI drought prediction, the ML model having R^2 close to 1 and RMSE and MAE near to 0 was chosen the best model.

Table 2, indicates the performance of the three ML models ELM, RF and SVR using several performance RMSE, MAE and R-squared. In the training set, The RMSE value of the ELM, RF and SVR models were found 0.417, 0.494, and 0.412, respectively. The mean absolute error (MAE) for the ELM, RF and the SVR models are found to be 0.410, 0.378, and 0.344, respectively. The coefficient-of-determination (R^2) values for the ELM, RF and the SVR models are calculated as 0.704, 0.740, and 0.753, respectively. Support Vector Machine (SVR) for regression showed that it has the highest coefficient of determination (R^2 = 0.753), lowest root mean square error (RMSE = 0.488), and lowest mean absolute error (MAE = 0.344). Therefore, Support Vector Machine (SVR) for regression outperformed Random Forest (RF) and Extreme Learning Machine (ELM) for predicting SPI_12.

Table 2 Presents the performance of ELM, RF and SVR models for predicting SPI_12

Model	**Training**			**Model**	**Testing**		
	RMSE	MAE	R^2		RMSE	MAE	R^2
ELM	0.417	0.304	0.822	ELM	0.568	0.410	0.704
RF	0.494	0.364	0.754	RF	0.500	0.378	0.740
SVR	0.412	0.284	0.826	SVR	0.488	0.344	0.753

Moreover, the scatter plot SPI_12 predicted and SPI_12 observed of the test data for the SVR model shows better fit line in using these three ML models. The scaterplots in Figure 6 a), b) and c) show how the three ML models ELM, RF and SVR respectively predict the SPI12.



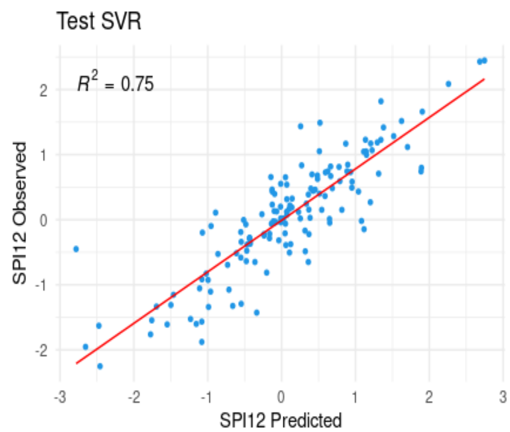


Fig. 6 (a), (b), (c) scatter plots of SPI_12 predicted and SPI_12 observed of the test data for the ELM, RF, and SVR respectively model

Table 3 below shows the evaluation performance of ELM, RF and SVR using the regression equation.

$$SPI12\_Predicted = m * SPI12\_Observed \quad (13)$$

The models listed in Table 3 have a slope of the best-fitted line 'm' of the estimated SPI-12, intercept 'c' of each model and the 'R^2' coefficient of determination of each model which shows the goodness of each model. The coefficient of determination ranges between 0 and 1. The closer R^2 of 1 indicates a high R^2 which implies a better fit. The three ML models ELM, RF and SVR show high R2 of 0.704, 0.740 and 0.753 respectively with slope line of 0.822, 1.042 and 0.952 respectively.

Table 3 Regression equation (SPI12predicted = m*SPI12observed + c) performance of the ELM, RF and SVR models using test data.

	m	**R^2**	**r**	**C**
ELM	0.822	0.704	0.839	0.086
RF	1.042	0.740	0.860	0.011
SVR	0.952	0.753	0.868	0.031

The study also compared the performance of the three models. Figure 7 shows the performance comparison of ELM, RF and SVR in the test set data using monthly SPI-12 observed and SPI-12 predicted. The plot reveals that the predicted SPI-12 of the test data of all three models were smoothly fit to the observed monthly SPI-12.

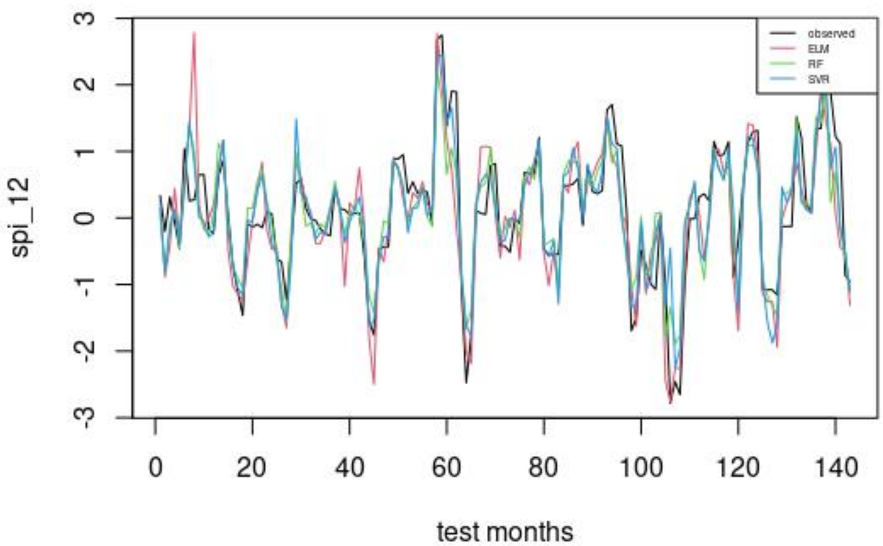


Fig. 7 Compares SPI_12 predicted for ELM, RF and SVR models versus observed SPI_12

5. CONCLUSION

The main purpose of this paper is to assess the performance of three machine learning algorithms to predict the Long- term standardized precipitation index (SPI-12) in Hiran, Somalia. Machine-learning models such as Extreme-Learning-Machine (ELM), Random-Forest (RF), and Support-Vector-Machine for Regression (SVR) were examined using monthly rainfall data from 1980–2021. Four SPI time scales (SPI1, SPI3, SPI6, and SPI9) were used as input variables for predicting SPI12. R programming software were used to predict SPI12 using the three ML models ELM, RF and SVR. It is observed that the Support Vector Machine (SVR) for regression has the highest value of R^2 of 0.753 and the lowest value of RMSE of 0.488. The results also showed that the three models were able to accurately predict drought events in the study area, with the SVM model performing slightly better than the ELM and RF models. The study observed low RMSE and high R^2 in prediction SPI- 12. However, the findings of the SVR show the good performance in the training models, with the highest R^2 (0.826) and minimal RMSE (0.412) values, and R^2 (0.753) and minimal RMSE (0.488) during testing set data in forecasting the SPI-12.

In conclusion, machine learning has shown to be a promising approach for drought forecasting in Hiran, Somalia, as it can effectively analyze complex meteorological data and make accurate predictions of drought events. However, further research is needed to improve the performance of machine learning models and to assess their potential for use in drought early warning systems in the region.

Supplementary Information

Author’s contribution

Funding

Data availability all detailed data are available on request.

Declarations

Conflict of Interest: The authors have no competing interests to disclose.

Ethical Approval and Participation Consent: Not applicable.

Publication Consent: All co-authors have given their consent for publication.

6. REFERENCE

[1] Ababa, A. (2007). Africa review report on agriculture and rural development. United Nations Economic and Social Council, August, 1–49.

[2] Abdulkadir. (2017). Assessment of Drought Recurrence in Somaliland: Causes, Impacts and Mitigations. Journal of Climatology & Weather Forecasting, 05(02), 1–12. https://doi.org/10.4172/2332-2594.1000204

[3] Achite, M., Jehanzaib, M., Elshaboury, N., & Kim, T. W. (2022a). Evaluation of Machine Learning Techniques for Hydrological Drought Modeling: A Case Study of the Wadi Ouahrane Basin in Algeria. Water (Switzerland), 14(3). https://doi.org/10.3390/w14030431

[4] Achite, M., Jehanzaib, M., Elshaboury, N., & Kim, T. W. (2022b). Evaluation of Machine Learning Techniques for Hydrological Drought Modeling: A Case Study of the Wadi Ouahrane Basin in Algeria. Water 2022, Vol. 14, Page 431, 14(3), 431. https://doi.org/10.3390/W14030431

[5] Ahmed, A. A. M., Deo, R. C., Ghahramani, A., Feng, Q., Raj, N., Yin, Z., & Yang, L. (2022). New Double Decomposition Deep Learning Methods for Stream-Flow Water Level Forecasting Using Remote Sensing Modis Satellite Variables, Climate Indices and Observations. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.4002418

[6] Aiyelokun, O., Ojelabi, A., & Agbede, O. (2020). Performance evaluation of machine learning models in predicting dry and wet climatic phases. Journal of Soft Computing in Civil Engineering, 4(1), 29–48. https://doi.org/10.22115/SCCE.2020.213319.1154

[7] Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. H. (Ted., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J. H., Allard, G., Running, S. W., Semerci, A., & Cobb, N. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. Forest Ecology and Management, 259(4), 660–684. https://doi.org/10.1016/J.FORECO.2009.09.001

[8] Anderson, W. B., Zaitchik, B. F., Hain, C. R., Anderson, M. C., Yilmaz, M. T., Mecikalski, J., & Schultz, L. (2012). Towards an integrated soil moisture drought monitor for East Africa. Hydrology and Earth System Sciences, 16(8), 2893–2913. https://doi.org/10.5194/HESS-16-2893-2012

[9] Anshuka, A., van Ogtrop, F. F., & Willem Vervoort, R. (2019). Drought forecasting through statistical models using standardised precipitation index: a systematic review and meta-regression analysis. Natural Hazards, 97(2), 955–977. https://doi.org/10.1007/S11069-019-03665-6/FIGURES/5

[10] Bouaziz, M., Medhioub, E., & Csaplovisc, E. (2021). A machine learning model for drought tracking and forecasting using remote precipitation data and a standardized precipitation index from arid regions.

Journal of Arid Environments, 189, 104478. https://doi.org/10.1016/J.JARIDENV.2021.104478

[11] Breiman, L. (2001). Random Forests (Vol. 45).

[12] Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. In Data Mining and Knowledge Discovery (Vol. 2).

[13] Buttafuoco, G., Caloiero, T., & Coscarelli, R. (2015). Analyses of Drought Events in Calabria (Southern Italy) Using Standardized Precipitation Index. Water Resources Management, 29(2), 557–573. https://doi.org/10.1007/S11269-014-0842-5/FIGURES/6

[14] Chen, S., Zhang, L., Zhang, Y., Guo, M., & Liu, X. (2020). Evaluation of Tropical Rainfall Measuring Mission (TRMM) satellite precipitation products for drought monitoring over the middle and lower reaches of the Yangtze River Basin, China. Journal of Geographical Sciences, 30(1), 53–67. https://doi.org/10.1007/s11442-020-1714-y

[15] Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-Vector Networks Editor. In Machine Leaming (Vol. 20). Kluwer Academic Publishers.

[16] Danandeh Mehr, A., Torabi Haghighi, A., Jabarnejad, M., Safari, M. J. S., & Nourani, V. (2022). A New Evolutionary Hybrid Random Forest Model for SPEI Forecasting. Water (Switzerland), 14(5). https://doi.org/10.3390/w14050755

[17] Deo, R. C., Kisi, O., & Singh, V. P. (2017). Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmospheric Research, 184, 149–175. https://doi.org/10.1016/J.ATMOSRES.2016.10.004

[18] Deo, R. C., & Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153, 512–525. https://doi.org/10.1016/j.atmosres.2014.10.016

[19] Dutra, E., Magnusson, L., Wetterhall, F., Cloke, H. L., Balsamo, G., Boussetta, S., & Pappenberger, F. (2013). The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products. International Journal of Climatology, 33(7), 1720–1729. https://doi.org/10.1002/JOC.3545

[20] Fredriksen, A. (2015). Crisis in 'a normal bad year': Spaces of humanitarian emergency, the Integrated Food Security Phase Classification scale and the Somali famine of 2011. Http://Dx.Doi.Org/10.1177/0308518X15597446, 48(1), 40–57. https://doi.org/10.1177/0308518X15597446

[21] Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data 2015 2:1, 2(1), 1–21. https://doi.org/10.1038/sdata.2015.66

[22] Gao, F., Zhang, Y., Ren, X., Yao, Y., Hao, Z., & Cai, W. (2018). Evaluation of CHIRPS and its application for drought monitoring over the Haihe River Basin, China. Natural Hazards, 92(1), 155–172. https://doi.org/10.1007/S11069-018-3196-0/FIGURES/9

[23] Ghebrezgabher, M. G., Yang, T., & Yang, X. (2016). Long-term trend of climate change and drought assessment in the Horn of Africa. Advances in Meteorology, 2016. https://doi.org/10.1155/2016/8057641

[24] Hinge, G., Mohamed, M. M., Long, D., & Hamouda, M. A. (2021). Meta-Analysis in Using Satellite Precipitation Products for Drought Monitoring: Lessons Learnt and Way Forward. Remote Sensing 2021, Vol. 13, Page 4353, 13(21), 4353. https://doi.org/10.3390/RS13214353

[25] Huang, G. Bin, Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. Neurocomputing, 70(1–3), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

[26] Kalantari, Z., Ferreira, C. S. S., Keesstra, S., & Destouni, G. (2018). Nature-based solutions for flood-drought risk mitigation in vulnerable urbanizing parts of East-Africa. Current Opinion in Environmental Science & Health, 5, 73–78. https://doi.org/10.1016/J.COESH.2018.06.003

[27] Kushwaha, N. L., Rajput, J., Shirsath, P. B., Sena, D. R., & Mani, I. (2022). Seasonal climate forecasts (SCFs) based risk management strategies: A case study of rainfed rice cultivation in India. Journal of Agrometeorology, 24(1), 10–17. https://doi.org/10.54386/JAM.V24I1.775

[28] Liu, Z. N., Li, Q. F., Nguyen, L. B., & Xu, G. H. (2018). Comparing machine-learning models for drought forecasting in vietnam's cai river basin. Polish Journal of Environmental Studies, 27(6), 2633–2646. https://doi.org/10.15244/pjoes/80866

[29] Malik, A., Tikhamarine, Y., Souag-Gamane, D., Rai, P., Sammen, S. S., & Kisi, O. (2021). Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. Meteorology and Atmospheric Physics, 133(3), 891–909. https://doi.org/10.1007/s00703-021-00787-0

[30] Maxwell, D., & Fitzpatrick, M. (2012). The 2011 Somalia famine: Context, causes, and complications. Global Food Security, 1(1), 5–12. https://doi.org/10.1016/J.GFS.2012.07.002

[31] McRoberts, D. B., & Nielsen-Gammon, J. W. (2012). The use of a high-resolution standardized precipitation index for drought monitoring and assessment. Journal of Applied Meteorology and Climatology, 51(1), 68–83. https://doi.org/10.1175/JAMC-D-10-05015.1

[32] Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. Journal of Hydrology, 391(1–2), 202–216. https://doi.org/10.1016/J.JHYDROL.2010.07.012

[33] Mishra, A. K., & Singh, V. P. (2011). Drought modeling – A review. Journal of Hydrology, 403(1–2), 157–175. https://doi.org/10.1016/j.jhydrol.2011.03.049

[34] Mondol, M. A. H., Ara, I., & Das, S. C. (2017). Meteorological Drought Index Mapping in Bangladesh Using Standardized Precipitation Index during 1981-2010. Advances in Meteorology, 2017. https://doi.org/10.1155/2017/4642060

[35] Musei, S. K., Nyaga, J. M., & Dubow, A. Z. (2021). SPEI-based spatial and temporal evaluation of drought in Somalia. Journal of Arid Environments, 184(September 2020), 104296. https://doi.org/10.1016/j.jaridenv.2020.104296

[36] Nakhaei, M., Mohebbi Tafreshi, A., & Saadi, T. (2023). An evaluation of satellite precipitation downscaling models using machine learning algorithms in Hashtgerd Plain, Iran. Modeling Earth Systems and Environment. https://doi.org/10.1007/S40808-022-01678-Y

[37] Neeti, N., Arun Murali, C. M., Chowdary, V. M., Rao, N. H., & Kesarwani, M. (2021). Integrated meteorological drought monitoring framework using multi-sensor and multi-temporal earth observation datasets and machine learning algorithms: A case study of central India. Journal of Hydrology, 601, 126638. https://doi.org/10.1016/J.JHYDROL.2021.126638

[38] Nicholson, S. E. (2014). A detailed look at the recent drought situation in the Greater Horn of Africa. Journal of Arid Environments, 103, 71–79. https://doi.org/10.1016/J.JARIDENV.2013.12.003

[39] Niranjan Kumar, K., Rajeevan, M., Pai, D. S., Srivastava, A. K., & Preethi, B. (2013). On the observed variability of monsoon droughts over India. Weather and Climate Extremes, 1, 42–50. https://doi.org/10.1016/J.WACE.2013.07.006

[40] Nyong, A., Adesina, F., & Osman Elasha, B. (2007). The value of indigenous knowledge in climate change mitigation and adaptation strategies in the African Sahel. Mitigation and Adaptation Strategies for Global Change, 12(5), 787–797. https://doi.org/10.1007/S11027-007-9099-0/METRICS

[41] Park, S., Im, J., Jang, E., & Rhee, J. (2016). Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agricultural and Forest Meteorology, 216, 157–169. https://doi.org/10.1016/J.AGRFORMET.2015.10.011

[42] Prakash, S. (2019). Performance assessment of CHIRPS, MSWEP, SM2RAIN-CCI, and TMPA precipitation products across India. Journal of Hydrology, 571, 50–59. https://doi.org/10.1016/J.JHYDROL.2019.01.036

[43] Qu, C., & Hao, X. (2018). Agriculture Drought and Food Security Monitoring Over the Horn of Africa (HOA) from Space. 2018 7th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 1–4. https://doi.org/10.1109/Agro-Geoinformatics.2018.8476128

[44] Qutbudin, I., Shiru, M. S., Sharafati, A., Ahmed, K., Al-Ansari, N., Yaseen, Z. M., Shahid, S., & Wang, X. (2019). Seasonal Drought Pattern Changes Due to Climate Variability: Case Study in Afghanistan. Water 2019, Vol. 11, Page 1096, 11(5), 1096. https://doi.org/10.3390/W11051096

[45] Rahmati, O., Falah, F., Dayal, K. S., Deo, R. C., Mohammadi, F., Biggs, T., Moghaddam, D. D., Naghibi, S. A., & Bui, D. T. (2020). Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. Science of The Total Environment, 699, 134230. https://doi.org/10.1016/J.SCITOTENV.2019.134230

[46] Said, A. A., Cetin, M., & Yurtal, R. (2019). Drought assessment and monitoring using some drought indicators in the semi-arid puntland state of Somalia. Fresenius Environmental Bulletin, 28(11 A), 8765–8772.

[47] Sakaa, B., Elbeltagi, A., Boudibi, S., Chaffaï, H., Islam, A. R. M. T., Kulimushi, L. C., Choudhari, P., Hani, A., Brouziyne, Y., & Wong, Y. J. (2022). Water quality index modeling using random forest and

improved SMO algorithm for support vector machine in Saf-Saf river basin. Environmental Science and Pollution Research, 29(32), 48491–48508. https://doi.org/10.1007/S11356-022-18644-X/METRICS

[48] Samarah, N. H. (2005). Effects of drought stress on growth and yield of barley. Agronomy for Sustainable Development, 25(1), 145–149. https://doi.org/10.1051/agro:2004064

[49] Santos, C. A. G., Brasil Neto, R. M., Passos, J. S. A., & da Silva, R. M. (2017). Drought assessment using a TRMM-derived standardized precipitation index for the upper São Francisco River basin, Brazil. Environmental Monitoring and Assessment, 189(6). https://doi.org/10.1007/s10661-017-5948-9

[50] Shrestha, N. K., Qamer, F. M., Pedreros, D., Murthy, M. S. R., Wahid, S. M., & Shrestha, M. (2017). Evaluating the accuracy of Climate Hazard Group (CHG) satellite rainfall estimates for precipitation based drought monitoring in Koshi basin, Nepal. Journal of Hydrology: Regional Studies, 13, 138–151. https://doi.org/10.1016/J.EJRH.2017.08.004

[51] Singh, B., Sihag, P., & Singh, K. (2017). Modelling of impact of water quality on infiltration rate of soil by random forest regression. Modeling Earth Systems and Environment, 3(3), 999–1004. https://doi.org/10.1007/S40808-017-0347-3/TABLES/5

[52] Sundararajan, K., Garg, L., Srinivasan, K., Bashir, A. K., Kaliappan, J., Ganapathy, G. P., Selvaraj, S. K., & Meena, T. (2021). A Contemporary Review on Drought Modeling Using Machine Learning Approaches. Computer Modeling in Engineering & Sciences, 128(2), 447–487. https://doi.org/10.32604/CMES.2021.015528

[53] Tian, D., He, X., Srivastava, P., & Kalin, L. (2022). A hybrid framework for forecasting monthly reservoir inflow based on machine learning techniques with dynamic climate forecasts, satellite-based data, and climate phenomenon information. Stochastic Environmental Research and Risk Assessment, 36(8), 2353–2375. https://doi.org/10.1007/S00477-021-02023-Y/METRICS

[54] Vrieling, A., Meroni, M., Mude, A. G., Chantarat, S., Ummenhofer, C. C., & de Bie, K. C. A. J. M. (2016). Early assessment of seasonal forage availability for mitigating the impact of drought on East African pastoralists. Remote Sensing of Environment, 174, 44–55. https://doi.org/10.1016/J.RSE.2015.12.003

[55] Wang, Y., & Witten, I. H. (1996). Induction of model trees for predicting continuous classes. https://researchcommons.waikato.ac.nz/handle/10289/1183

[56] World Bank Group. (2018). Rebuilding resilient and sustainable agriculture in Somalia. In International Bank for Reconstruction and Development (Vol. 1).

[57] Wu, W., Li, Y., Luo, X., Zhang, Y., Ji, X., & Li, X. (2019). Performance evaluation of the CHIRPS precipitation dataset and its utility in drought monitoring over Yunnan Province, China. Http://Www.Tandfonline.Com/Action/JournalInformation?Show=aimsScope&journalCode=tgnh20#.VsXodSCLRhE, 10(1), 2145–2162. https://doi.org/10.1080/19475705.2019.1683082

[58] Yaseen, Z. M., Ali, M., Sharafati, A., Al-Ansari, N., & Shahid, S. (2021). Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh. Scientific Reports

2021 11:1, 11(1), 1–25. https://doi.org/10.1038/s41598-021-82977-9

[59] Yeditha, P. K., Rathinasamy, M., Neelamsetty, S. S., Bhattacharya, B., & Agarwal, A. (2022). Investigation of satellite precipitation product driven rainfall-runoff model using deep learning approaches in two different catchments of India. Journal of Hydroinformatics, 24(1), 16–37. https://doi.org/10.2166/HYDRO.2021.067

[60] Yihdego, Y., Vaheddoost, B., & Al-Weshah, R. A. (2019). Drought indices and indicators revisited. Arabian Journal of Geosciences, 12(3), 1–12. https://doi.org/10.1007/S12517-019-4237-Z/TABLES/5

[61] Zhang, R., Chen, Z. Y., Xu, L. J., & Ou, C. Q. (2019). Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. Science of The Total Environment, 665, 338–346. https://doi.org/10.1016/J.SCITOTENV.2019.01.431

[62] Zhang, Ying, Hou, J., & Huang, C. (2022). Integration of Satellite-Derived and Ground-Based Soil Moisture Observations for a Precipitation Product over the Upper Heihe River Basin, China. Remote Sensing 2022, Vol. 14, Page 5355, 14(21), 5355. https://doi.org/10.3390/RS14215355

[63] Zhang, Yuhu, Yang, H., Cui, H., & Chen, Q. (2020). Comparison of the Ability of ARIMA, WNN and SVM Models for Drought Forecasting in the Sanjiang Plain, China. Natural Resources Research, 29(2), 1447–1464. https://doi.org/10.1007/s11053-019-09512-6