Research in Ecology
https://journals.bilpubgroup.com/index.php/re

ARTICLE

Drought Forecast Using Traditional and Custom Models for Dhaka, Bangladesh

Aunik Hasan Mridul [1*], *Tanumoy Bose* [2], *Swapneel Biswas* [3], *Nafi Ahmed* [3], *S. M. Hasan Kabir* [1], *Nebadeta Nath Tonney* [4], *Pooja Saha* [5]

[1] *Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh*

[2] *Department of Electrical and Electronics Engineering, Ahsanullah University of Science and Technology, Dhaka 1208, Bangladesh*

[3] *Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka 1208, Bangladesh*

[4] *Department of Computer Science and Engineering, American International, University-Bangladesh, Dhaka 1229, Bangladesh*

[5] *Department of Geography and Environment, University of Dhaka, Dhaka 1000, Bangladesh*

ABSTRACT

Water scarcity and climate change are two of the biggest worldwide concerns. A complicated and sometimes underappreciated occurrence, drought has an impact on many facets of human existence. Early drought forecasts are therefore essential for water resource management and strategic planning. In order to improve the accuracy of drought prediction, this work presents a unique hybrid model that combines the Autoregressive Moving Average (ARMA), Holt-Winters (Exponential Smoothing) model, Autoregressive Integrated Moving Average (ARIMA), and Random Forest Regressor model. We do a thorough analysis of the Dhaka Division, Bangladesh, daily precipitation data from January 1981 to March 2025. In contrast to other research that only examined standalone machine learning algorithms or conventional statistical models, our study combines the two and offers a comparative performance analysis of hybrid models in the context of drought prediction using SPI. Furthermore, the study uses these models in the understudied

*CORRESPONDING AUTHOR:
Aunik Hasan Mridul, Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh; Email: Aunik15-2732@diu.edu.bd

ARTICLE INFO
Received: 20 May 2025 | Revised: 27 May 2025 | Accepted: 18 June 2025 | Published Online: 11 November 2025
DOI: https://doi.org/10.30564/re.v7i5.10103

CITATION
Mridul, A.H., Bose, T., Biswas, S., et al., 2025. Drought forecast using traditional and custom models for Dhaka, Bangladesh. Research in Ecology. 7(5): 1–16. DOI: https://doi.org/10.30564/re.v7i5.10103

setting of Dhaka, Bangladesh, a place where little previous research has been done on drought forecasting. When examined side by side, our hybrid model Holt-Winters with LSTM model outperforms the hybrid approach. For SPI daily predictions, significant statistical parameters like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are especially crucial. This noteworthy enhancement highlights how much more accurate the innovative model is in forecasting droughts in Bangladesh's Dhaka Division. Our findings highlight the hybrid model's vital importance in tackling the problems caused by drought in the larger framework of climate change and water resource management.
Keywords: ARMA; ARIMA; Mean Absolute Error; LSTM; Drought

1. Introduction

Since droughts are natural disasters that almost always occur in all climates, researchers in a wide range of disciplines, including ecology, agriculture, meteorology, and environment, have been interested in researching them recently [1]. Human life is negatively impacted by drought, which is characterized as a time when there is insufficient soil moisture and a reduction in the amount of water available in surface and groundwater reservoirs. Given that drought is an unexplainable event that has particular detrimental impacts on civilization, one reasonable strategy to lessen its impact would be to forecast when it will occur [2]. Distinct areas may have distinct effects from drought. Drought indices (DIs) are commonly used to evaluate the effects of hydro-meteorological components, such as streamflow and precipitation, by showing a function of these variables. Four types of droughts may be evaluated using DI: hydrological, agricultural, socioeconomic, and meteorological. But there's no guarantee that there won't be a drought, which is why it's critical to monitor drought events with indexes. Given that drought varies both momentarily and geographically, it is crucial to evaluate the likelihood of DI as well as the availability of hydro-meteorological data [3]. In contrast to unexpected calamities, meteorological droughts occur gradually, giving time for education and planning. Even though these droughts don't first appear to be severe, they offer an opportunity for proactive water conservation and sustainable practices. As they develop, more individuals realize how important it is to use water, which encourages positive behavioral adjustments. Weather-related droughts also encourage scientific inquiry and result in improvements to forecast and resilience mechanisms. Most significantly, communities have more time to respond because of its slow start, which reduces the immediate hazards to people and property. While the advantages may vary, taking advantage of these opportunities contributes to societal resilience, sustainable practices, and the resolution of many drought-related problems. The meteorological perspective is important because it may be used to predict drought conditions before they happen [4].

Because of its inherent benefits, SPI was chosen for this inquiry. SPI will be quite beneficial as it just depends on precipitation, to start. This is particularly true in places where evaporation, moisture content, or temperature are unavailable. Second, SPI was made available as a tool for evaluating precipitation shortfalls across different time periods. Precipitation anomalies may cause different water resources to react at different times, with shorter or longer durations indicating these delays [5]. Lastly, SPI is a standardized index that is simpler to compute and versatile [6,7]. Time series event prediction is approached in a variety of ways. Because of its exceptional accuracy in predicting time-oriented events, the autoregressive integrated moving average (ARIMA) model is a widely used method. ARIMA has long been regarded as a reliable tool for time series forecasting, including streamflow and drought forecasting, due to some of its advantages over other approaches like neural networks and exponential smoothing [5]. ARIMA, for instance, may be able to adequately account for serial correlation, which is frequently seen in time series modeling. In order to choose an appropriate model, this model can also offer a searching stage that comprises identification, estimate, and diagnostic testing. Because of its prediction accuracy and adaptability, ARIMA is a commonly used approach for a variety of time series data sets [8]. Nevertheless, ARIMA can only capture a portion of nonlinear and nonstationary time-oriented data, and it is only very useful for linear and stationary datasets [9]. The literature has a wide range of time series forecasting techniques, including neural networks, ARMA, ARIMA, linear regression, and simple moving averages (MAs). These techniques use historical records to predict future occurrences; time series data, on the other hand, are not determinist series and are regarded by scholars as stationary series. One way to model time series is to think of them as a mix of white noise and deterministic functions. In any time, series, a de-nosing technique like the wavelet transform may reduce white noise and provide a better model [3]. In Dhaka, Bangladesh, the agriculture and water resource management sectors have a major challenge with drought predictions. Prompt and accurate forecasting of drought conditions can lessen its negative effects on agricultural productivity, water availability, and general economic stability. This work focuses on using daily Standardized Precipitation Index (SPI) data, a commonly used measure of meteorological drought, to anticipate drought conditions. The goal is to increase the accuracy and dependability of drought forecasts by utilizing sophisticated hybrid models that integrate machine learning techniques like Random Forest Regressor with statistical techniques like the Holt-Winters exponential smoothing technique. The resilience and sustainability of the region's agricultural and water resources will eventually be enhanced by improved forecasting skills, which will allow for better readiness and response measures.

This study is based on machine learning, deep learning and some hybrid models combining for forecasting drought for Dhaka, Bangladesh. In this study we have presented the forecasting technique using standardized precipitation index (SPI) value for 1981 to 2023 for everyday intervals. The related works are discussed in this segment.

Wavelet transform has been extensively employed to enhance forecasting precision in various mathematical prediction domains, often in conjunction with stochastic and artificial intelligence-based techniques [10]. Wang and Ding et al., identified wavelet transform as a valuable tool for drought forecasting. In their study, the integration of wavelet transforms with an artificial neural network (ANN) demonstrated improved drought prediction accuracy by leveraging the strengths of both methodologies [11]. Similarly, Kriechbaumer et al. utilized wavelet transform as a preprocessing step to enhance the predictive performance of the ARIMA model in forecasting metal prices, validating its effectiveness [12].

Venkata Ramana et al., applied a wavelet-ANN hybrid approach to estimate monthly rainfall, where model calibration and validation were assessed using statistical criteria [13]. Their findings indicated that this hybrid model outperformed standalone ANN models, significantly improving rainfall prediction accuracy. Several researchers have recognized wavelet transform as a robust tool for time series analysis, particularly in identifying trends, periodicities, and fluctuations [14,15]. By applying wavelet transform, both time and frequency domain representations of a signal can be obtained, offering insights into the underlying processes. Research has demonstrated that wavelet-based methods are effective in analyzing time-oriented datasets, as each decomposed subseries reveals detailed information about data structure and periodicity [16–18].

Wavelet integration with stochastic models such as ARIMA and artificial intelligence models like ANN has gained popularity as a preprocessing technique for hydrologic time series, particularly in de-noising data and improving model performance [19–23]. This is particularly useful since single ANN models struggle with nonstationary time series.

Belayneh and Adamowski evaluated the effectiveness of three data-driven approaches—Support Vector Machines (SVM), ANN, and Wavelet-ANN (WANN)—for drought prediction in Ethiopia's Awash River Basin using Standardized Precipitation Index (SPI) values. Their results demonstrated that the hybrid WANN model provided superior prediction accuracy compared to the other methods, as measured by R^2, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) [19–21]. A subsequent study compared the performance of traditional stochastic ARIMA models, Support Vector Regression (SVR), and ANN in predicting drought occurrences within the same basin [19]. In this study, the wavelet transform was integrated into these models as a data preprocessing step. Notably, the wavelet-based SVR (WSVR) model showed superior predictive capabilities, particularly for longer forecasting horizons of 6 to 12 months.

In another approach, a hybrid model called wavelet linear genetic programming (WLGP) was developed using Palmer's modified drought index. Danandeh Mehr et al. found that WLGP outperformed standard genetic programming models for long-term drought prediction, as the latter struggled to model beyond a three-month lead time [24]. Similarly, Djerbouai [25] assessed drought forecasting accuracy in Algeria's Algerois Basin by comparing ANN models with stochastic ARIMA and SARIMA models using

SPI data for different lead times. Their study demonstrated that preprocessing data with wavelet transform significantly enhanced ANN model performance, as indicated by statistical metrics such as the Nash-Sutcliffe Efficiency (NSE) coefficient, RMSE, and MAE across lead times of 1 to 6 months [26–32].

For drought index forecasting, a novel hybrid model known as the wavelet-based extreme learning machine (WELM) was tested at three locations in Australia. This model was compared with Extreme Learning Machine (ELM), ANN, Least-Squares SVR (LSSVR), and their wavelet-enhanced counterparts (WANN and WLSSVR). Findings from Deo et al. revealed that WELM consistently outperformed the other models based on statistical performance metrics [26]. Additionally, WANN exhibited promising results with lower frequency error rates and computational efficiency, further demonstrating the advantages of wavelet-based models in drought forecasting.

This study aims to evaluate the predictive accuracy of hybrid models and ARIMA for forecasting SPI-based drought conditions. By comparing these models, the research assesses their effectiveness in predicting daily drought conditions at different time points. While ARIMA remains a popular choice for hydrological time series forecasting, it struggles with nonlinearity and nonstationarity—key characteristics of SPI data [33,34]. Despite the increasing application of drought forecasting models, there is a lack of research specific to Dhaka, Bangladesh, a region particularly vulnerable to the socioeconomic impacts of drought. This study addresses this gap by presenting both ARIMA and hybrid models to improve prediction accuracy and assess their comparative performance.

Recent years have witnessed a surge in research leveraging machine learning and deep learning approaches to enhance the accuracy and robustness of drought forecasting. Globally, numerous studies have demonstrated the effectiveness of hybrid and deep learning models in predicting drought indices such as SPI. For example, Gasmi et al. [35] and Agudelo et al. [36] explored hybrid ARIMA-LSTM architectures and neural feature fusion methods for rainfall and drought predictions, highlighting their advantages over standalone statistical models. Similarly, Zhang et al. [37,38] and Rezaiy and Shabri [39] proposed advanced neural networks such as STAT-LSTM and EEMD-ARIMA for capturing spatiotemporal patterns and short-term drought fluctuations. Otkin et al. [40] and Dikshit et al. [41] who noted significant improvements in early drought warning systems when integrating these approaches. Deep learning models such as LSTM, RBF, and quantum Mamba networks were investigated by Tang et al. [42] and Hossaini-Moghari et al. [43] showcasing their strong performance in nonlinear time series modeling. The importance of hybrid models combining statistical and machine learning techniques was further emphasized by studies like those of Brust et al. [44].

In the context of Bangladesh, a number of studies have focused on localized drought forecasting using SPI and related indices. Rahman and Azim [45], along with Rabby and Adhikary [46], examined the trends and uncertainty of meteorological droughts, while Mondol et al. [47] and Islam et al. [48] validated the performance of SPI, EDI, and other indices for regional drought monitoring. Several studies such as Paul et al. [49] and Hossain and Rahman [50] introduced machine learning-based tools using satellite and remote sensing data, advancing agricultural drought monitoring frameworks. In parallel, Khatun and Khan [51] used CMIP5 projections with SPI and EDI to forecast future drought conditions under climate change scenarios, while Rahman et al. [52] and Alam et al. [53] integrated geospatial and statistical indicators for drought mapping in northwestern Bangladesh [54]. Overall, these works collectively highlight a growing research trend toward integrating statistical methods with artificial intelligence to improve drought prediction, offering significant implications for water resource planning and disaster management, particularly in climate-vulnerable regions like Bangladesh.

2. Materials and Methods

At first, we have selected the topic as per daily life needs. Then we have justified the topic according to the problem solutions. We have collected the dataset. Then we have planned for the study. We have implemented Hybrid models with some regular forecasting models. The Box-Jenkins method and the theory of stochastic processes serve as the foundation for ARIMA models. Climate persistence causes temporal autocorrelation in drought indicators like the SPI, which makes them ideal for ARIMA modeling. The premise of this study is that previous climate patterns have an impact on future results (Granger causality). Therefore, by accounting for these lag-dependent effects, ARIMA may be used to statistically anticipate the onset of drought.

2.1. Data Collection

We utilized the NASA POWER website to obtain the daily temporal average of solar fluxes and related temperature, humidity/precipitation, wind/pressure, and solar fluxes from 1-1-1981 to 7-3-2025. The coordinates are: 23.873, 89.7573. We used the coordinates to pinpoint the site and then pulled all of the historical data into a CSV file [55]. **Figure 1** shows the map of Dhaka Division.

Figure 1. The Map of Dhaka Division.

2.2. Data Pre-Processing

Here we will explain the data pre-processing technique we have applied for our study. We have Converted the Year, Month, Day column into one Date column. We also dropped unnecessary columns from our dataset. We checked the missing values presented in our dataset. But there were no any missing value. We have evaluated the SPI value using equations. The SPI (Soil Precipitation Index) is a drought index that is calculated using the following formula:

SPI = (Rainfall − MonthlyAvgRainFall) / MonthlyStdRainFall

where:

Rainfall is the amount of rainfall that fell in a given month and year.

MonthlyAvgRainFall is the average amount of rainfall that falls in that month and year, based on a historical record.

MonthlyStdRainFall is the standard deviation of the rainfall in that month and year, based on a historical record.

The SPI is a standardized measure of drought, meaning that it has a mean of 0 and a standard deviation of 1. A positive SPI value indicates that the rainfall was above average for that month and year, while a negative SPI value indicates that the rainfall was below average.

The df.apply() function is used to apply the SPI formula to each row of the df DataFrame. The lambda function is used to define the SPI formula. The lambda function takes two arguments: the row of the DataFrame and the index of the row. The lambda function then calculates the SPI value for that row and returns it.

If the MonthlyStdRainFall value is 0, then the SPI value cannot be calculated. In this case, the lambda function returns the value -2. This value is used to indicate that the SPI value is not available. Here is a mathematical explanation of the SPI formula:

SPI = (Rainfall − MonthlyAvgRainFall) / MonthlyStdRainFall

The numerator of the fraction represents the difference between the actual rainfall and the average rainfall for that month and year. The denominator of the fraction represents the standard deviation of the rainfall for that month and year.

By dividing the difference between the actual and average rainfall by the standard deviation, we are standardizing the SPI value. This means that the SPI value has a mean of 0 and a standard deviation of 1.

A positive SPI value indicates that the rainfall was above average for that month and year. A negative SPI value indicates that the rainfall was below average.

- It then returns a classification label for the SPI value, based on the following criteria below.
- If the SPI value is greater than or equal to 2, the classification label is “Extremely wet”.
- If the SPI value is between 1.5 and 2, the classification label is “Severely wet”.
- If the SPI value is between 1 and 1.5, the classification label is “Moderately wet”.
- If the SPI value is between 0 and 1, the classification label is “Mild wet”.
- If the SPI value is between -1 and 0, the classification label is “Mild drought”.
- If the SPI value is between -1.49 and -1, the classification label is “Moderately drought”.
- If the SPI value is between -1.99 and -1.5, the classification label is “Severely drought”.
- If the SPI value is less than -1.5, the classification label is “Extremely drought”.

The **Figure 2** shows the graphical view of the dataset according to SPI. **Figure 3** shows the graphical view of the Seasonal Decompose for Monthly SPI value.

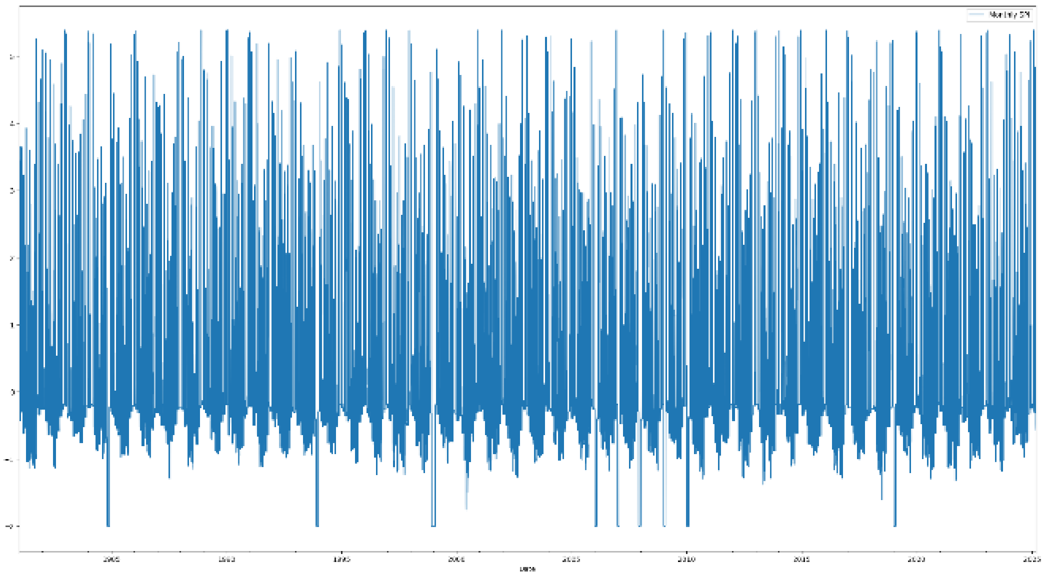

Figure 2. The Graphical View of the Dataset According To SPI.

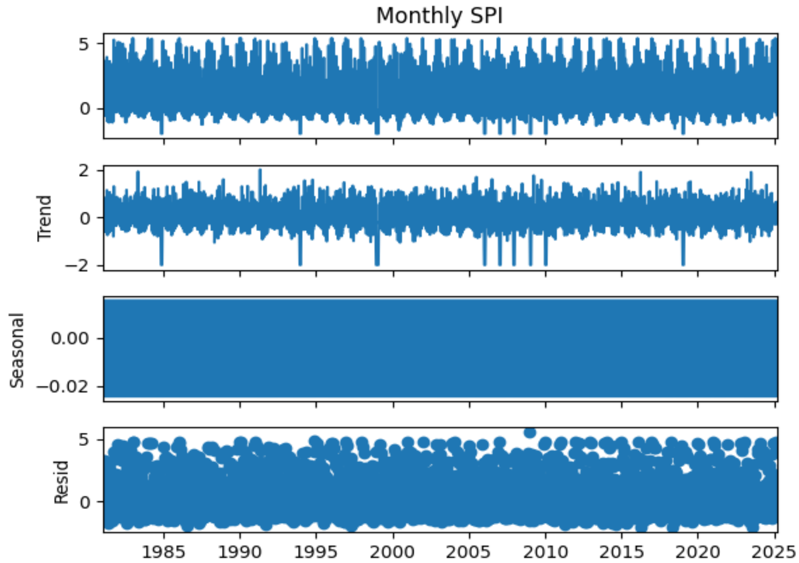

Figure 3. The Graphical View of the Seasonal Decompose for Monthly SPI Value.

2.3. Algorithms

We have employed the Holt-Winters (Exponential Smoothing) model, the Autoregressive Moving Average (ARMA) model, the Autoregressive Integrated Moving Average (ARIMA) model, and the Random Forest Regressor model. Along with hybrid models ARIMA with Multi-Layer Perception Model (MLP), ARIMA with Long-Short Time Memory (LSTM) model, and ARIMA with Random Forest Regressor, Holt-Winters with Multi-Layer Perception Model, Holt-Winters with Long-Short Time Memory, and Holt-Winters with Random Forest Regressor were also used.

2.3.1. ARIMA Model

Time series data analysis and forecasting are commonly done using the ARIMA (AutoRegressive Integrated Moving Average) model. Our chosen ARIMA model, designated as ARIMA(2,0,1), has one moving average term, two autoregressive terms, and no differencing required to keep the time series stationary (d = 0). This model represents both the historical errors (moving average portion) and the connection between the series' current value and its prior values (autoregressive part). Our model's stepwise search method minimized the Akaike Information Criterion (AIC), a model quality metric that strikes a compromise between model complexity and goodness of fit, in order to determine the ideal parameters. Based on the lowest AIC value among the models that were taken into consideration, ARIMA(2,0,1) was deemed to be the best match for our data. Now that the model has been fitted, projections based on past data patterns may be made, providing insights into future trends. The links between the present value, previous values, and previous errors are captured by our model, ARIMA(2,0,1), which consists of one moving average term and two autoregressive components. A stepwise search was used to find the model by minimizing the Akaike Information Criterion (AIC), a metric that weighs complexity and fit quality. Now that the best-fitting model, ARIMA(2,0,1), has been identified, future values may be predicted using previous data patterns. It is expressed as:

$$y_t' = I + a_1 y_{t-1}' + a_2 y_{t-2}' + \ldots a_p y_{t-p}' + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

2.3.2. ARMA Model

A well-liked technique for time series forecasting, the ARMA (AutoRegressive Moving Average) model combines moving average (MA) and autoregressive (AR) components to identify relationships in the data. The best match for the Monthly SPI data, according to our study, was the ARMA(1,0) model, which has one autoregressive component and no moving average terms. Every value in the series is impacted by its recent past value, as the AR term illustrates. The auto_arima function, which finds the optimum compromise between model complexity and fit by doing a stepwise search to minimize the Akaike Information Criterion (AIC), was used to choose this model. It is now possible to use the final model to estimate future values since it has a substantial autoregressive coefficient, indicating a strong link between subsequent observations in the time series. An essential method for time series forecasting, the ARMA (AutoRegressive Moving Average) model combines moving average (MA) and autoregressive (AR) components to represent time-dependent data. The best model for the Monthly SPI data, according to our study, is the ARMA(1,0) model, which has one autoregressive term and no moving average terms. To ensure that the model offers a decent fit without needless complexity, the selection method made use of the auto_arima function, which uses a stepwise search to minimize the Akaike Information Criterion (AIC). According to the AR(1) concept, every value in the series is mostly impacted by the value that comes right before it. The autoregressive coefficient, which indicates a substantial temporal dependency in the data, was extremely significant when the model's parameters were calculated. AIC value of 42240.387 for the final model indicates a strong fit, and suitability of the residuals is shown by diagnostic tests. By utilizing the patterns seen in the historical data, this ARMA(1,0) model may now be utilized to predict future Monthly SPI values with accuracy. It is expressed as:

$$Y_n = \beta_0 W[n] + \beta_1 W[n-1] + \beta_2 W[n-2] + \ldots + \beta_q W[n-q]$$

2.3.3. Holt-Winters Model

In our research, we applied the Holt-Winters model, which is a widely used technique for time series forecasting, especially for data with seasonal trends. Three elements make up this model: seasonality, trend, and level. These elements are updated recursively to reflect the underlying patterns in the data. To accommodate for potential dampening effects over time, we have included a damped trend in addition to additive seasonal and trend components in our implementation of the model. Next, with optimization turned on for parameter estimation, the model is fitted to the training set of data using the Exponential Smoothing function. Predicted values are then acquired for the test data interval. All things considered, the Holt-Winters model offers a strong foundation for predicting time series data, using seasonal and trend data to provide precise forecasts. The Holt-Winters model is a powerful forecasting technique designed to capture trends and seasonal patterns in time series data. It incorporates three key components: level, trend, and seasonality. In our implementation, the model is configured with additive seasonal and trend components, indicating that the seasonal and trend effects are added to the level. Additionally, a damped trend is incorporated to account for potential dampening of the trend over time. It is expressed as:

$$S_t = \gamma * (y_t - l_t) + (1 - \gamma*) s_{t-m}$$

2.3.4. Random Forest Regressor Model

For regression problems, a flexible machine learning technique called the Random Forest Regressor is employed. In order to get a final result, it builds many decision trees during training and then averages their forecasts. To provide variety and lessen overfitting, every decision tree in the forest is trained using a random subset of features and a random portion of the training data. A reliable and accurate model is produced by averaging the predictions provided by each individual tree. Because of its versatility, scalability, and capacity to manage intricate datasets with high-dimensional features, the Random Forest Regressor is a well-liked solution for regression issues across a range of industries. A potent machine learning method for regression tasks involving the prediction of continuous outcomes is the Random Forest Regressor. It is based on the idea of an ensemble of decision trees, in which several separate trees are trained using different subsets of the feature set and data. In order to minimize overfitting and enhance generalization performance, each decision tree learns to generate predictions based on a distinct subset of the data and features during training. Furthermore, by taking into account just a random subset of characteristics at each split in the tree-building process, random forests provide randomization to the training process. This unpredictability keeps individual trees from being overly specialized to the training set and increases variability among them. It is expressed as:

$$g(x) = f_0(x) + f_1(x) + f_2(x)$$

2.3.5. MLP Model

A basic artificial neural network design for supervised learning tasks like regression and classification is the Multilayer Perceptron (MLP) model. MLPs are composed of several layers of networked nodes (neurons) that use forward propagation to transfer input data across the network. An activation function and the weighted sum of each neuron's inputs define each neuron's output. The model minimizes the discrepancy between expected and actual results during training by learning to modify the weights connecting neurons using backpropagation. Because of their ability to recognize intricate patterns in data through repeated optimization, multilevel perception (MLP) algorithms are useful for a variety of tasks in a variety of fields. An artificial neural network type called the Multilayer Perceptron (MLP) model is distinguished by its tiered design, which consists of an input layer, one or more hidden layers, and an output layer. Neurones, or networked nodes, make up each layer and process the input data through calculations. In order to reduce the discrepancy between expected and actual outputs, the MLP learns during training by modifying the weights and biases connected to each link between neurons. Gradients are used to update the model parameters, and backpropagation—which propagates mistakes backward through the network—helps with this process. The key strength of MLPs lies in their ability to learn complex nonlinear relationships within data. By employing activation functions at each neuron, MLPs can capture intricate patterns and dependencies, making them well-suited for tasks with high-dimensional input spaces and nonlinear decision boundaries.

$$x_j^{d+1} = \sum_i y_i^d w_{ij}^d - a_j^{d+1}$$

2.3.6. LSTM Model

The Long Short-Term Memory (LSTM) model is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem and capture long-term dependencies in sequential data. LSTM networks consist of memory cells with self-connected recurrent units, allowing them to selectively retain and update information over time. This architecture enables LSTMs to effectively learn from and remember patterns in sequential data, making them particularly well-suited for tasks such as time series prediction, natural language processing, and speech recognition. The Long Short-Term Memory (LSTM) model is a specialized type of recurrent neural network (RNN) designed to overcome the limitations of traditional RNNs in learning and retaining information over long sequences. Unlike standard RNNs, which struggle with the vanishing gradient problem and have difficulty capturing long-range dependencies in sequential data, LSTMs feature a more complex architecture with memory cells, input and output gates, and forget gates. The core components of an LSTM cell include a memory cell that stores information over time, an input gate that controls the flow of information into the cell, an output gate that controls the flow of information out of the cell, and a forget gate that determines which information to discard from the cell's memory. These gates, implemented using sigmoid and tanh activation functions, enable LSTMs to selectively update and retain information based on its relevance to the current task.

2.3.7. Hybrid Model

In this study we have used the combination of multiple forecasting models for best performance. We have used ARIMA with MLP, ARIMA with LSTM, ARIMA with Random Forest Regressor, Holt-Winter with MLP, Holt-Winter with LSTM model and Holt-Winter with Random Forest Regressor model.

In our analysis, we have employed various combinations of traditional time series models and advanced machine learning algorithms to forecast future values. The ARIMA models, coupled with Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and Random Forest Regressor, leverage autoregressive and moving average components to capture temporal patterns in the data. Similarly, Holt-Winters models, integrated with MLP, LSTM, and Random Forest Regressor, exploit seasonal and trend components to make predictions. By combining these time-tested models with powerful machine learning techniques, we aim to harness the strengths of both approaches, utilizing the flexibility and interpretability of traditional models alongside the capacity for capturing complex relationships and nonlinearities offered by machine learning algorithms. This comprehensive approach allows us to explore diverse modeling strategies and identify the most effective combination for accurate and robust forecasting. The proposed methodology is presented in **Figure 4**.

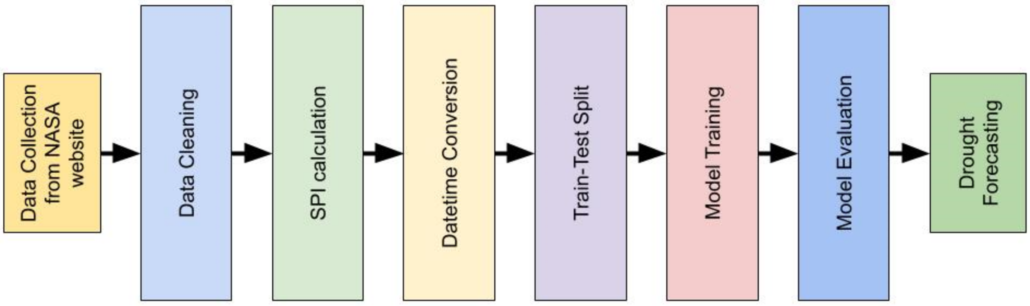

Figure 4. Proposed Methodology.

3. Results

The evaluation process for SPI (Standardized Precipitation Index) forecasting typically involves assessing the MSE and RMSE. The results are shown below **Figure 5** and **Table 1**.

The comparison of evaluation metrics for SPI forecasting across different models provides insights into their effectiveness. Among traditional time series models, ARIMA, ARMA, and Holt-Winters exhibit identical MSE (1.01) and RMSE (1.009) values, suggesting comparable predictive performance in capturing temporal patterns in the SPI data. In contrast, machine learning-based models demonstrate varying levels of accuracy. The Random Forest Regressor records the highest MSE (2.37) and RMSE (1.54), indicating potential difficulties in capturing the underlying time-dependent structures in the data. However, hybrid approaches combining ARIMA or Holt-Winters with machine learning models show promising improvements. ARIMA with MLP, LSTM, and Random Forest Regressor yield MSE values between 1.10 and 1.14 and RMSE values between 1.05 and 1.06, indicating slightly higher errors compared to traditional time series models. Holt-Winters with MLP, LSTM, and Random Forest Regressor generally perform better than ARIMA-based hybrids, with MSE scores ranging from 0.88 to 1.08 and RMSE scores between 0.93 and 1.04. Notably, Holt-Winters with LSTM achieves the lowest MSE (0.88) and RMSE (0.93) among all models, demonstrating strong capability in capturing both trend and seasonality patterns. While traditional models such as ARIMA and Holt-Winters provide consistent performance, hybrid models incorporating machine learning techniques, particularly Holt-Winters with LSTM, offer the best predictive accuracy for SPI forecasting. Future work could focus on further optimizing these hybrid approaches to enhance their ability to model complex time series behaviors.

The **Figure 6** shows the forecasting data according to the Test Data for Holt-Winters with LSTM model.

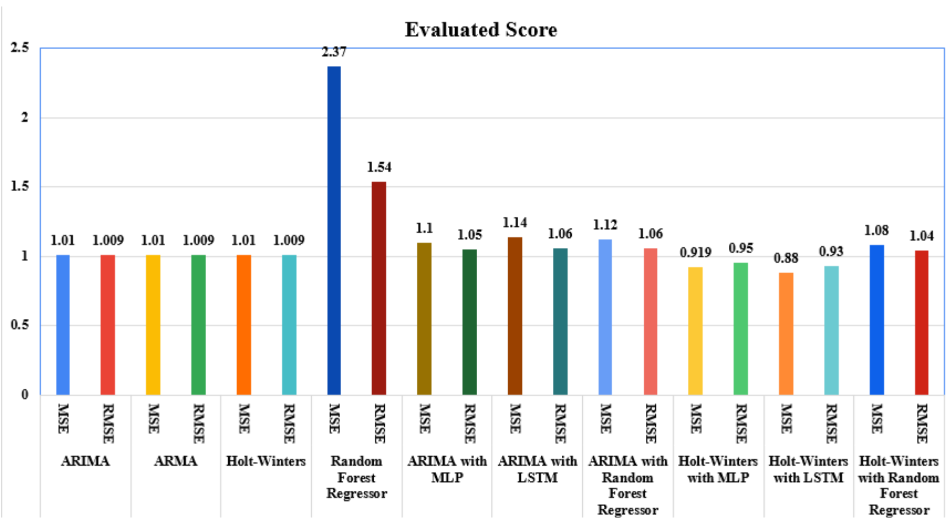

Figure 5. Evaluated Results.

Table 1. Result Analysis.

Model	**Evaluation Matrices**	**Evaluated Score**
ARIMA	MSE	1.01
	RMSE	1.009
ARMA	MSE	1.01
	RMSE	1.009
Holt-Winters	MSE	1.01
	RMSE	1.009
Random Forest Regressor	MSE	2.37
	RMSE	1.54
ARIMA with MLP	MSE	1.1
	RMSE	1.05
ARIMA with LSTM	MSE	1.14
	RMSE	1.06
ARIMA with Random Forest Regressor	MSE	1.12
	RMSE	1.06
Holt-Winters with MLP	MSE	0.919
	RMSE	0.95
Holt-Winters with LSTM	MSE	0.88
	RMSE	0.93
Holt-Winters with Random Forest Regressor	MSE	1.08
	RMSE	1.04

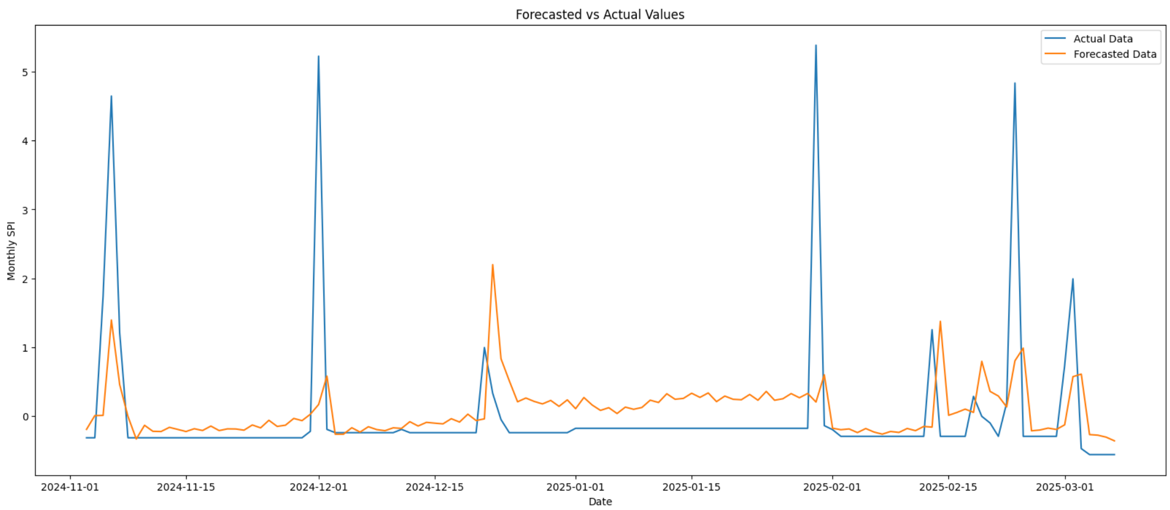

Figure 6. The Forecasting Data According to Test Data.

4. Discussion and Future Works

A decline in rainfall leads to increased drought severity. This temporal causality is captured by autoregressive components in ARIMA, which implies that past SPI values statistically 'cause' current drought conditions under the Granger framework. This study may help the facts like crop planning and irrigation scheduling. Risk mitigation in agriculture-dependent economies. Early warning systems to reduce socioeconomic losses. In future work, several avenues can be explored to enhance the accuracy and robustness of SPI forecasting models. Firstly, incorporating additional meteorological and hydrological variables, such as temperature, humidity, and soil moisture, into the modeling process can provide a more comprehensive understanding of the factors influencing drought dynamics. This multi-variable approach may improve the predictive capability of the models by capturing a broader range of environmental conditions that contribute to drought development and persistence. Secondly, leveraging advanced machine learning techniques, such as deep learning architectures like Convolutional Neural Networks (CNNs) and

Transformer models, could offer further improvements in SPI forecasting accuracy. These models have shown promise in capturing complex spatiotemporal patterns in various environmental datasets and may be particularly effective in modeling the intricate relationships within SPI time series data. Additionally, exploring ensemble modeling approaches, which combine predictions from multiple individual models, can help mitigate the inherent uncertainties in SPI forecasting and enhance overall prediction accuracy. Ensemble methods, such as model averaging and stacking, integrate diverse modeling perspectives and exploit the complementary strengths of different forecasting techniques, thereby providing more robust and reliable predictions. Furthermore, there is a need for enhanced data collection and preprocessing techniques to address issues related to data quality, missing values, and temporal inconsistencies in SPI datasets. Utilizing advanced data assimilation methods and remote sensing technologies can facilitate the integration of diverse data sources and improve the spatial and temporal resolution of SPI observations, leading to more accurate and timely drought assessments. Moreover, developing tailored decision support systems and visualization tools for SPI forecasting can facilitate the interpretation and communication of forecasting results to stakeholders, including policymakers, agricultural practitioners, and water resource managers. These tools can help translate SPI forecasts into actionable insights and inform proactive drought mitigation strategies, such as water resource allocation, crop planning, and disaster preparedness measures. Finally, conducting comprehensive validation studies and real-time testing of SPI forecasting models in diverse geographical regions and climatic conditions is essential to assess their reliability and generalizability. Collaborative efforts between researchers, government agencies, and stakeholders can facilitate the validation process and promote the adoption of validated forecasting models for operational use in drought monitoring and management initiatives. Overall, by addressing these future research directions, we can advance the state-of-the-art in SPI forecasting and contribute to more effective drought preparedness, response, and resilience-building efforts at local, regional, and global scales.

5. Conclusions

In this study, we investigated various forecasting models for predicting the Standardized Precipitation Index (SPI), a critical indicator of drought conditions, across different regions and time scales. Through a comprehensive evaluation process, we compared the performance of traditional time series models including ARIMA, ARMA, and Holt-Winters, with machine learning algorithms such as Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and Random Forest Regressor. Our analysis revealed valuable insights into the strengths and limitations of each modeling approach in capturing the complex temporal patterns inherent in SPI data. The results of our study demonstrated that traditional time series models, particularly Holt-Winters with LSTM, exhibited competitive performance in SPI forecasting, achieving relatively low Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) scores across different evaluation matrices. These models effectively captured the underlying seasonal and trend components of the SPI time series, making them suitable choices for drought prediction tasks. However, the ARMA model showed slightly higher error metrics, indicating a relatively weaker fit to the SPI data compared to Holt-Winters with LSTM. On the other hand, machine learning-based approaches, including MLP, LSTM, and Random Forest Regressor, presented mixed results in SPI forecasting. While Random Forest Regressor demonstrated promising performance, showcasing its ability to capture nonlinear relationships in the data, the combinations of ARIMA and machine learning algorithms, as well as Holt-Winters with machine learning algorithms, exhibited higher MSE and RMSE values. This suggests that further refinement and optimization of these hybrid models are needed to improve their predictive accuracy for SPI forecasting tasks.

In conclusion, our study highlights the importance of selecting appropriate modeling techniques based on the specific characteristics of the SPI data and the desired forecasting objectives. Traditional time series models such as Holt-Winters with LSTM remain robust choices for SPI forecasting, particularly in capturing seasonal and trend patterns. However, the integration of machine learning algorithms offers opportunities for enhancing predictive performance, especially in handling nonlinearities and complex dependencies within the SPI time series. Future research should focus on refining and optimizing hybrid modeling approaches to leverage the strengths of both traditional and modern forecasting methods for more accurate and reliable SPI predictions, ultimately aiding in effective drought monitoring and management efforts. In climate-vulnerable areas, this study offers a workable strategy to increase the accuracy of drought predictions. The results can help make well-informed decisions on catastrophe risk reduction, water resource management, and agricultural planning. This study provides important insight for future forecasting frameworks in environmental analytics by comparing many models and highlighting the benefits of hybrid models over conventional ones.

Recommendations

Based on the findings of this study, it is recommended that future efforts in SPI forecasting prioritize the use of traditional time series models, particularly Holt-Winters with LSTM, due to their strong ability to capture seasonal and trend components effectively. However, given the potential of machine learning approaches like Random Forest Regressor to model nonlinear relationships, further research should focus on refining and optimizing hybrid models that combine the strengths of both traditional and machine learning techniques. Enhancing these hybrid approaches could lead to more accurate and reliable SPI predictions, ultimately improving drought monitoring, early warning systems, and resource management strategies.

Author Contributions

T.B. and S.B. conceptualized the research framework and led the development of the hybrid modeling approach. N.A. and S.M.H.K. handled data acquisition, preprocessing, and statistical analysis of the precipitation dataset for the Dhaka Division. N.N.T. contributed to the implementation and fine-tuning of machine learning models, particularly the Random Forest and LSTM components. P.S. assisted in the integration and comparative evaluation of ARIMA and Holt-Winters-based hybrid models. A.H.M. supervised the entire project, ensured the methodological rigor, and contributed to the writing, editing, and final approval of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

There was no outside support for this study. The authors were the only ones who organized and supplied all financial assistance, including the tools, resources, and materials needed for the study.

Data Availability Statement

The daily temporal average of solar fluxes and associated temperature, humidity/precipitation, wind/pressure, and solar fluxes from 1-1-1981 to 7-3-2025 were obtained via the NASA POWER website. The location is 23.873, 89.7573. After locating the location using the coordinates, we extracted all of the historical data into a CSV file [35].

Acknowledgments

We would like to sincerely thank our academic colleagues, mentors, and supervisors for their tremendous assistance, encouragement, and direction during this project. We are also grateful to the organizations and scientists that made resources and datasets accessible to the general public, which helped us create and validate our model for detecting rice leaf disease. We would especially want to thank our friends, family, and peers for their unwavering support and understanding during the whole research. Their combined assistance has been crucial to this work's successful completion.

Conflict of Interest

The authors declare no conflicts of interest.

References

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

[2] Wilhite, D.A., 2005. Drought and Water Crises. CRC Press: Boca Raton, FL, USA. DOI: https://doi.org/10.1201/9781420028386

[3] Khan, Md.M.H., Muhammad, N.S., El-Shafie, A., 2020. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. Journal of Hydrology. 590, 125380. DOI: https://doi.org/10.1016/j.jhydrol.2020.125380

[4] Wu, H., Hubbard, K.G., Wilhite, D.A., 2004. An agricultural drought risk-assessment model for corn and

soybeans. International Journal of Climatology. 24(6), 723–741. DOI: https://doi.org/10.1002/joc.1028

[5] Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment. 19(5), 326–339. DOI: https://doi.org/10.1007/s00477-005-0238-4

[6] Guttman, N.B., 1998. Comparing the palmer drought index and the standardized precipitation index. Journal of the American Water Resources Association. 34(1), 113–121. DOI: https://doi.org/10.1111/j.1752-1688.1998.tb05964.x

[7] Zargar, A., Sadiq, R., Naser, B., et al., 2011. A review of drought indices. Environmental Reviews. 19, 333–349. DOI: https://doi.org/10.1139/a11-013

[8] Zhang, G.P., 2003. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing. 50, 159–175. DOI: https://doi.org/10.1016/s0925-2312(01)00702-0

[9] Nourani, V., Alami, M.T., Aminfar, M.H., 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Engineering Applications of Artificial Intelligence. 22(3), 466–472. DOI: https://doi.org/10.1016/j.engappai.2008.09.003

[10] Nourani, V., Hosseini Baghanam, A., Adamowski, J., et al., 2014. Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. Journal of Hydrology. 514, 358–377. DOI: https://doi.org/10.1016/j.jhydrol.2014.03.057

[11] Wang, W., Ding, J., 2003. Wavelet Network Model and Its Application to the Prediction of Hydrology. 1(1), 67–71.

[12] Kriechbaumer, T., Angus, A., Parsons, D., et al., 2014. An improved wavelet–ARIMA approach for forecasting metal prices. Resources Policy. 39, 32–41. DOI: https://doi.org/10.1016/j.resourpol.2013.10.005

[13] Venkata Ramana, R., Krishna, B., Kumar, S.R., et al., 2013. Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. Water Resources Management. 27(10), 3697–3711. DOI: https://doi.org/10.1007/s11269-013-0374-4

[14] Seo, Y., Choi, Y., Choi, J., 2017. River Stage Modeling by Combining Maximal Overlap Discrete Wavelet Transform, Support Vector Machines and Genetic Algorithm. Water. 9(7), 525. DOI: https://doi.org/10.3390/w9070525

[15] Quilty, J., Adamowski, J., Boucher, M., 2019. A Stochastic Data-Driven Ensemble Forecasting Framework for Water Resources: A Case Study Using Ensemble Members Derived From a Database of Deterministic Wavelet-Based Models. Water Resources Research. 55(1), 175–202. DOI: https://doi.org/10.1029/2018wr023205

[16] Kim, T.-W., Valdés, J.B., 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering. 8(6), 319. DOI: https://doi.org/10.1061/(ASCE)1084-0699(2003)8:6(319)

[17] Partal, T., Kişi, Ö., 2007. Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. Journal of Hydrology. 342(1–2), 199–212. DOI: https://doi.org/10.1016/j.jhydrol.2007.05.026

[18] Rathinasamy, M., et al., 2014. Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models. Water Resources Research. 50(12), 9721–9737. DOI: https://doi.org/10.1002/2013wr014650

[19] Adamowski, J., Fung Chan, H., Prasher, S.O., et al., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resources Research. 48(1). DOI: https://doi.org/10.1029/2010wr009945

[20] Shabri, A., 2015. A Hybrid Model for Stream Flow Forecasting Using Wavelet and Least Squares Support Vector Machines. Jurnal Teknologi. 73(1), 101–111. DOI: https://doi.org/10.11113/jt.v73.3380

[21] Belayneh, A., Adamowski, J., Khalil, B., 2015. Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustainable Water Resources Management. 2(1), 87–101. DOI: https://doi.org/10.1007/s40899-015-0040-5

[22] Belayneh, A., Adamowski, J., 2012. Standard Precipitation Index Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Support Vector Regression. Applied Computational Intelligence and Soft Computing. 2012, 1–13. DOI: https://doi.org/10.1155/2012/794061

[23] Soh, Y.W., Koo, C.H., Huang, Y.F., et al., 2018. Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia. Computers and Electronics in Agriculture. 144, 164–173. DOI: https://doi.org/10.1016/j.compag.2017.12.002

[24] Danandeh Mehr, A., Kahya, E., Özger, M., 2014. A gene–wavelet model for long lead time drought forecasting. Journal of Hydrology. 517, 691–699. DOI: https://doi.org/10.1016/j.jhydrol.2014.06.012

[25] Djerbouai, S., Souag-Gamane, D., 2016. Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Stochastic Models: Case of the

Algerois Basin in North Algeria. Water Resources Management. 30(7), 2445–2464. DOI: https://doi.org/10.1007/s11269-016-1298-6

[26] Deo, R.C., Tiwari, M.K., Adamowski, J.F., et al., 2016. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stochastic Environmental Research and Risk Assessment. 31(5), 1211–1240. DOI: https://doi.org/10.1007/s00477-016-1265-z

[27] Hayes, M., Svoboda, M., Wall, N., et al., 2011. The Lincoln declaration on drought indices: universal meteorological drought index recommended. Bulletin of the American Meteorological Society. 92(4), 485–488. DOI: https://doi.org/10.1175/2010BAMS3103.1

[28] Koycegiz, C., Buyukyildiz, M., 2019. Calibration of SWAT and Two Data-Driven Models for a Data-Scarce Mountainous Headwater in Semi-Arid Konya Closed Basin. Water. 11(1), 147. DOI: https://doi.org/10.3390/w11010147

[29] Koycegiz, C., Buyukyildiz, M., 2022. Investigation of precipitation and extreme indices spatiotemporal variability in Seyhan Basin, Turkey. Water Supply. 22(12), 8603–8624. DOI: https://doi.org/10.2166/ws.2022.391

[30] Koycegiz, Cihangir, Buyukyildiz, Meral, 2022. Investigation of spatiotemporal variability of some precipitation indices in Seyhan Basin, Turkey: monotonic and sub-trend analysis. Natural Hazards. 116(2), 2211–2244. DOI: https://doi.org/10.1007/s11069-022-05761-6

[31] Mahmud, I., Bari, S.H., Rahman, M.T.U., 2016. Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method. Environmental Engineering Research. 22(2), 162–168. DOI: https://doi.org/10.4491/eer.2016.075

[32] Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 11(7), 674–693. DOI: https://doi.org/10.1109/34.192463

[33] Mckee, T.B., Doesken, N.J., Kleist, J., 1995. Drought monitoring with multiple time scales. 9, 233–236.

[34] Souther, S., Colombo, S., Lyndon, N.N., 2023. Integrating traditional ecological knowledge into US public land management: Knowledge gaps and research priorities. Frontiers in Ecology and Evolution. 11, 988126. DOI: https://doi.org/10.3389/fevo.2023.988126

[35] Gasmi, A., Rahoma, Y.T., Soliman, M.M., 2024. Time Series Forecasting Using Hybrid ARIMA-LSTM for Rainfall Prediction. Journal of Hydrology and Environment Research. 13(2), 101–111.

[36] Agudelo, M., Paredes, L., Gutierrez, R.G., 2025. Drought forecasting using a hybrid neural architecture for integrating time series and static data. Environmental Modelling & Software. 167, 105534.

[37] Zhang, J., Chen, W., Liu, Q., 2025. STAT-LSTM: A multivariate spatiotemporal feature aggregation model for SPEI-based drought prediction. Knowledge-Based Systems. 284, 111035.

[38] Zhang, P., Liu, X., Wang, J., 2022. A review of deep learning applications for drought forecasting. Water Resources Management. 36(3), 983–1001.

[39] Rezaiy, H., Shabri, A.B., 2023. Short-term drought forecasting using hybrid EEMD-ARIMA model based on SPI. Stochastic Environmental Research and Risk Assessment. 37, 453–469.

[40] Otkin, J.A., Hain, C.R., Svoboda, M., 2021. DroughtCast: A machine learning approach for forecasting drought categories. Journal of Applied Meteorology and Climatology. 60(8), 1173–1187.

[41] Dikshit, S., Li, H., Waters, D.M., 2021. Long-term drought forecasting using LSTM for NSW, Australia. Remote Sensing. 13(14), 2721–2738.

[42] Tang, K., Yuan, L., Zhang, T., 2025. Quantum-empowered drought forecasting using a Mamba Network for SPEI analysis. AI for Earth Science Journal. 2, 1–14.

[43] Hossaini-Moghari, M., Khosravi, A., 2021. Application of LSTM and RBF in short-term drought prediction based on SPI. Journal of Hydrology. 602, 126748.

[44] Brust, C., Abolafia-Rosenzweig, A., Mote, T.L., 2021. Flash Drought Forecasting: Progress and Challenges in the Deep Learning Era. Drought Research Quarterly. 8, 25–36.

[45] Rahman, M., Azim, M.A., 2022. Assessment of Meteorological Drought in Bangladesh Using SPI. Journal of Climate Research. 14(2), 67–78.

[46] Rabby, T., Adhikary, S.K., 2024. Uncertainty-based drought assessment in Bangladesh using SPI. Environmental Systems Research. 13(1), 11.

[47] Mondol, S., Hasan, M., Islam, M.S., 2021. Observed Trends of Meteorological Drought in Bangladesh using EDI and SPI. Theoretical and Applied Climatology. 143(3–4), 1233–1246.

[48] Islam, N., Ahmed, R.U., Jahan, M., 2021. Comparative Evaluation of SPI and EDI for Drought Monitoring in Bangladesh. Climate Dynamics and Sustainability. 6(2), 209–220.

[49] Paul, A., Kabir, R., Khatun, M.A., 2025. Enhanced drought forecasting using satellite and machine learn-

ing data for Bangladesh. Remote Sensing of Environment. 302, 112949.

[50] Hossain, T., Rahman, S.M., 2021. Evaluating agricultural drought using EDI over Bangladesh's northwestern belt. Agricultural Water Management. 250, 106827.

[51] Khatun, M., Khan, A.M., 2021. CMIP5-Based Drought Projections for Bangladesh Using SPI and EDI. Climatic Change Letters. 159(1), 83–95.

[52] Rahman, A., Baten, S.A., Nahar, T., 2020. An integrated drought risk mapping using TVDI and climatic indicators in NW Bangladesh. Natural Hazards. 104, 1201–1216.

[53] Alam, M.U., Jahan, S.S., Kawsar, F.M., 2020. Pre-Monsoon Water Scarcity in Agriculture: A Parametric SPI Approach in Bangladesh. Hydrology and Earth System Sciences Discussions. 25(5), 2401–2418.

[54] Ali, M.M., Mahmud, T., Barua, R.S., 2021. Volatility and Predictability of SPI-Based Droughts in Northwestern Bangladesh. Climate Risk Management. 34, 10037.

[55] NASA, 2018. ArcGIS Web Application. Nasa.gov. Available from: https://power.larc.nasa.gov/data-access-viewer/ (cited 17 March 2025)