

Article

Application of Artificial Intelligence to Forecast Drought Index for the Mekong Delta

Duong Hai Ha ^{1,*}, Phong Nguyen Duc ^{1,*}, Thuan Ha Luong ², Thang Tang Duc ³, Thang Trinh Ngoc ¹, Tien Nguyen Minh ¹ and Tu Nguyen Minh ¹

¹ Institute for Water and Environment, Vietnam Academy for Water Resources, Hanoi 100000, Vietnam; tunm1191@gmail.com (T.N.M.)

² Vietnam Water Resources Association, Hanoi 100000, Vietnam

³ The Southern Institute of Water Resources Research, Vietnam Academy for Water Resources, Ho Chi Minh 700000, Vietnam; tdthang.siwrr@gmail.com

* Correspondence: hahaiduong.cwe@gmail.com (D.H.H.); phongndtv@gmail.com (P.N.D.)

Abstract: Droughts have a substantial impact on water supplies, agriculture, and ecosystems worldwide. Agricultural sustainability and production in the Mekong Delta of Vietnam are being jeopardized by droughts caused by climate change. Conventional forecasting methods frequently struggle to comprehend the intricate dynamics of meteorological occurrences connected to drought, necessitating the use of sophisticated prediction techniques. This study assesses the effectiveness of various statistical models (ARIMA), machine learning, and deep learning models (Gradient Boosting, XGBoost, RNN, and LSTM) in forecasting the SPEI over different time periods (1, 3, 6, and 12 months) across six prediction intervals. The models were developed and evaluated using data from 11 meteorological stations spanning from 1985 to 2022. These models incorporated various climatic variables, including precipitation, temperature, humidity, potential evapotranspiration (PET), Southern Oscillation Index (SOI) Anomaly, and sea surface temperature in the NINO4 region (SST_NINO4). The results demonstrate that XGBoost and LSTM models exhibit outstanding performance, showcasing lower error metrics and higher R^2 values compared to Gradient Boosting and RNN. The performance of the model fluctuated depending on the forecast step, with error metrics often increasing with longer prediction horizons. The use of climatic indices improved the accuracy of the model. These findings are consistent with earlier research on drought episodes in the Mekong Delta and support studies from other areas that show the effectiveness of advanced modeling tools for predicting droughts. The work emphasizes the capacity of machine learning and deep learning models to enhance the precision of drought forecasting, which is vital for efficient water resource management and agricultural planning in places prone to drought.

Keywords: drought forecasting; standardized precipitation evapotranspiration index; machine learning algorithms; deep learning algorithms; Mekong delta; water resource management; Vietnam



Citation: Ha, D.H.; Duc, P.N.; Luong, T.H.; Duc, T.T.; Ngoc, T.T.; Minh, T.N.; Minh, T.N. Application of Artificial Intelligence to Forecast Drought Index for the Mekong Delta. *Appl. Sci.* **2024**, *14*, 6763. <https://doi.org/10.3390/app14156763>

Academic Editor: Athanasios Sfetsos

Received: 21 May 2024

Revised: 22 July 2024

Accepted: 23 July 2024

Published: 2 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Droughts are natural hazards that can occur in all climatic zones and have long-term economic and environmental impacts [1]. They can be defined in different ways, such as meteorological, hydrological, and agricultural droughts, depending on the time horizon and variables used [2]. Climate change has made drought one of the greatest natural hazards in Europe, affecting large areas and populations [3]. In the United States, precipitation deficits have been the primary drivers of past major drought events, with temperature as a secondary driver [4]. Droughts in South Africa have led to employment losses in the agricultural sector, affecting income generation [5]. Droughts adversely affect various environmental components, including soil processes, vegetation growth, wildlife, water quality, and aquatic ecosystems. They also limit access to water resources and can have an international impact.

Accurately predicting drought is essential for reducing the negative effects of drought in the Mekong Delta, Vietnam, which impact agriculture, water management, and community resilience. It facilitates strategic agricultural planning, optimizes water resource allocation, and improves early warning systems to prepare for difficulties associated with drought. Accurate predictions are essential for adjusting to climate change through guiding sustainable practices and policy development. This holistic method for predicting drought enhances agricultural output in the area, guarantees sustainable water resources, and strengthens resistance to climate-related challenges. Thus, accurate drought forecasting is the cornerstone of proactive and effective risk management. It empowers stakeholders to make informed decisions, implement timely interventions, and build resilience in the face of a changing climate, ultimately contributing to the sustainable development of regions vulnerable to drought in the Mekong Delta.

This study focuses on using Artificial Intelligence (AI) to anticipate the drought index in the Mekong Delta, with the goal of addressing and achieving critical objectives. Its primary goal is to overcome the limitations of traditional drought forecasting systems, which frequently fail to adequately describe the Mekong Delta's complex and fluctuating meteorological circumstances. The aim of this study is to dramatically increase the precision and dependability of drought forecasts using Artificial Intelligence. Furthermore, AI's better forecasting power is meant to assist in the delivery of timely warnings and the implementation of preventative actions. This proactive strategy enables communities, governments, and other stakeholders to appropriately prepare for and minimize the effects of impending water scarcity and related difficulties.

Current methods for forecasting drought involve a combination of statistical, probabilistic, and data-driven approaches. A study by [6] analyzed the spatiotemporal variability of meteorological droughts in the Mekong Delta area of Vietnam using the Standardized Precipitation Index (SPI) and found that the frequency of drought scales decreased while their spatial distribution tended to increase, with the main scales including moderate and severe droughts. The most extreme drought during the study period occurred in 1990–1992, with 11 out of 13 provinces experiencing extreme drought with a peak SPI value of -2.63 and a duration of 29 months. The study concluded that climate change was the major factor affecting drought in the study area rather than the El Niño phenomenon. The Mekong Delta has a long history of drought, with the 2015–2016 event being particularly severe, and The Mekong Delta suffered the worst historic drought and salinity intrusion occurrence on record [7]. This region has also experienced a shift in the spatial distribution of meteorological droughts, with a decrease in frequency and an increase in severity [8]. The impact of these droughts on agriculture, particularly on rice production, is significant [9]. Nguyen Thi Ngoc et al. evaluated meteorological droughts using the Standardized Precipitation Index (SPI) based on data from the Tropical Rainfall Measuring Mission (TRMM) [10]. Nguyen and Li analyzed the correlation between sea surface temperature anomalies (SSTA) and meteorological droughts in the Vietnam Mekong Delta [11]. These studies demonstrate the use of various methods and data sources for drought forecasting in the Mekong Delta. [12] investigated the spatiotemporal trends, intensity, duration, and frequency of meteorological droughts in the Vietnamese Mekong Delta (VMD) using the Standardized Precipitation Evaporation Index (SPEI) at multiple timescales (3, 6, and 12 months). The findings suggest that the intensity, duration, and frequency of drought events increased from 1985 to 2018, with extreme drought events from October 2013 to September 2016 being the most severe and prolonged during the study period.

Various research has demonstrated good results when applying the ARIMA model to anticipate the drought index, notably the Standardized Precipitation Evapotranspiration Index (SPEI). The ARIMA model has proven to be effective in analyzing and predicting drought conditions when applied to SPEI time-series data at various temporal scales, including 1-month, 3-month, and 6-month intervals. As a result, it improved the understanding of climatic patterns and assisted in decision-making and resource management. ARIMA models, employing historical SPEI data, demonstrated higher accuracy in the

Amman-Zarqa Basin compared to TBATS models. Performance measures indicated strong predicting capabilities [13]. In addition, the hybrid wavelet-ARIMA (W-ARIMA) model has been suggested as a means to enhance drought forecasting. This model combines wavelet transform with ARIMA and has demonstrated superior performance compared to the conventional ARIMA model, as evidenced by statistical metrics such as Root Mean Square Error, Mean Absolute Error, and the Mean Absolute Percentage Error [14]. The experiments demonstrate the strength and flexibility of ARIMA and its hybrid models in predicting drought indices such as SPEI. This offers useful insights for strategic planning and managing water resources.

Current approaches to drought forecasting in the Mekong Delta have limitations. The lack of observation stations reduces the reliability of the monitoring results, making it difficult to accurately identify droughts [10]. Additionally, current weather and climate conditions have negatively affected the accuracy and reliability of traditional prediction indicators used by small-scale farmers in the region [15]. These indicators, which are based on traditional environmental cues, may not be as effective in predicting drought events under the current conditions of climate uncertainty and variability [16]. These limitations highlight the need to enhance traditional prediction methods and develop new approaches that can better account for the changing environmental and climatic conditions in the Mekong Delta [17].

Drought index forecasting faces challenges in accurately predicting drought events due to deficiencies in capturing the complex interactions between meteorological and hydrological variables [18,19]. These deficiencies can lead to intensified drought characteristics in certain climatic regions, impacting the precision of drought forecasts. To address these issues, a novel approach utilizing the Informer model has been proposed, which outperforms traditional methods like machine learning and deep learning models by better capturing information over time and enhancing long-term prediction accuracy. The models adapt to different timescales effectively, significantly improving the precision of SPEI prediction and aligning forecasted trends with actual observations. By incorporating diverse climatological variables and advanced forecasting techniques, the proposal aims to overcome the limitations of existing SPEI forecasting methods and provide more reliable drought predictions.

Artificial Intelligence (AI) techniques have been increasingly used for drought forecasting. These models have been applied to improve current weather forecasts and as alternatives to conventional predictions of extreme events [20]. In the Mekong Delta of Vietnam, where drought has become more severe owing to climate change, ML-based models have been used to assess future drought hazards [21]. Luong Bang Nguyen and J. Lee demonstrated the effectiveness of this technology for predicting drought indices and rainfall, respectively [22,23]. The use of climate indices as input variables in these models further enhances their accuracy. A. Jalalkamali et al. compared the performance of various Artificial Intelligence models in drought forecasting, and the ARIMAX model showed the highest precision [24]. A. Kikon and P. C. Deka provided a comprehensive review of the role of Artificial Intelligence in drought assessment, monitoring, and forecasting, highlighting its significance in these areas [25].

Deep learning models have shown significant potential in forecasting drought indices, such as the Standardized Precipitation Evapotranspiration Index (SPEI), which is crucial for regions like the Mekong Delta in Vietnam [21,26–28]. These models, including Deep Neural Networks (DNNs), Multi-Layer Perceptron (MLP), and Convolutional Neural Networks (CNNs), can effectively handle complex datasets and extract relevant features for accurate predictions, aiding in mitigating drought impacts like crop failure and water shortages. By integrating deep learning methods like Long Short-Term Memory (LSTM) into forecasting systems, the accuracy and lead time of drought predictions can be improved, enabling policymakers to develop proactive risk management strategies and drought mitigation plans tailored to the long-term impacts of climate change in coastal regions like the Mekong

Delta. These studies collectively underscore the potential of Artificial Intelligence in improving drought forecasting in the Mekong Delta.

Deep learning models, particularly Long Short-Term Memory (LSTM), have been successfully applied in forecasting drought indices like the Standardized Precipitation Evaporation Index (SPEI) to predict drought occurrences with high accuracy [29,30]. These models utilize meteorological data, including precipitation and temperature, to make timely and precise predictions of drought conditions, which are crucial for mitigating the negative impacts on agriculture, water resources, and ecosystems [30,31]. LSTM models have shown superior performance in predicting drought at different time scales, such as 1 month and 3 months, outperforming other hybrid models like EMD-ELM. Additionally, LSTM models have been used to simulate vegetation dynamics and predict vegetation activities and stresses based on meteorological data, showcasing their versatility in various environmental forecasting applications [32].

The Standardized Precipitation Evapotranspiration Index (SPEI) is a popular index for evaluating drought conditions. It has been used in various studies to analyze drought patterns and severity [18,33–35]. The SPEI combines meteorological and hydrological variables, such as precipitation, evapotranspiration, and groundwater levels, to provide a comprehensive assessment of drought [36]. It has been found to accurately characterize severe drought events in different climatic regions. Additionally, the SPEI has been used to monitor drought conditions during critical phenological phases of crops, such as maize cultivation, and to assess the temporal and spatial variability of droughts. The SPEI is a drought index used to assess water balance and drought conditions. It calculates a standardized value based on a continuous probability distribution fitted to a water balance time series. Different probability distributions, such as generalized logistic (GLO), generalized extreme value (GEV), Pearson Type III (PE3), and normal (NOR) distributions, have been considered for SPEI analysis in various regions. Studies have recommended using PE3 or GEV distributions for SPEI analysis in Canada [37], whereas a new multiscale SPEI dataset has been provided for reference and future time horizons in Italy [38]. Regional drought analysis using SPEI has been performed in the Gediz Basin, Turkey, with different distributions found to be the best fit for different reference periods [35]. In China, the SPEI has been used to accurately monitor drought events, with spatiotemporal distribution and trends analyzed in various climatic sub-regions [39]. In Malaysia, the SPEI has been used to determine drought indices for the Pahang River Basin with the aim of mitigating the impact on water supply and economic development [40].

SPEI has several advantages: It is useful for assessing both drought and wetter-than-normal conditions and provides a comprehensive understanding of moisture variability [38]. SPEI is a reliable tool for drought prediction because it is simpler, faster, and requires fewer data points than dynamic models [30]. It can accurately determine the spatial and temporal dimensions of drought events, making it valuable for drought monitoring and risk assessments. The SPEI is particularly effective in predicting droughts, with higher overall accuracy and fewer mistakes compared to other indices, such as the Standardized Precipitation Index (SPI). Additionally, the SPEI can be used to estimate the impact of drought events on water availability, agriculture, and ecosystems, aiding in the mitigation of economic losses and damage to the quality of life [41]. The versatility of the SPEI allows for the development of ensemble PDFs, making it suitable for assessing drought projections throughout the 21st century. SPEI is a reliable drought index that can be used for accurate drought assessment and forecasting.

Artificial Intelligence plays a transformational role in drought forecasting, providing novel solutions to the constraints of traditional methodologies. Drought forecasting becomes increasingly accurate, flexible, and responsive as AI is used, ultimately aiding effective water resource management and increasing resilience in drought-prone areas.

2. Materials and Methods

2.1. Study Area

The Mekong Delta is a region in southwestern Vietnam where the Mekong River approaches and empties into the sea. It covers over 40,500 km² and is an important source of agriculture and aquaculture in the region. The area is also vulnerable to adverse impacts of climate change, including saltwater intrusion, coastal erosion, flooding, and drought, and efforts are being made to ensure greater productivity and climate resilience (Figure 1).

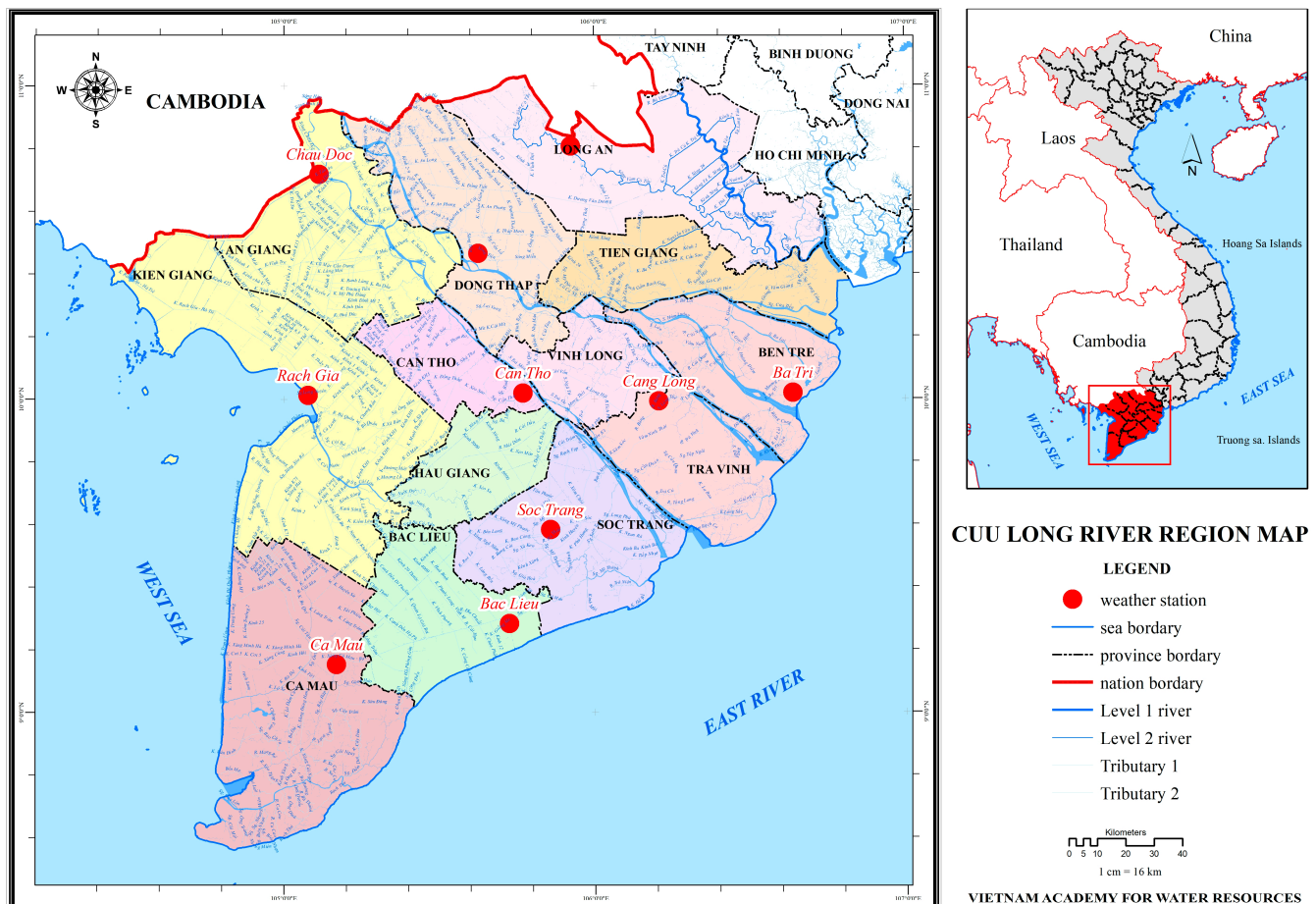


Figure 1. The geographical location and metrological stations in the Mekong Delta.

Drought is a prevalent concern in the Mekong Delta because of climate change impacts [42]. The intensity, duration, and frequency of meteorological droughts in the delta have been studied using various indices [43]. The Vietnamese Mekong Delta (VMD) has experienced droughts that affect agriculture and aquaculture [44]. These findings contribute to a better understanding of drought patterns and their impact on agricultural output in the Mekong Delta, providing valuable insights for policymakers and practitioners in water resource management.

From 1985 to 2022, the Mekong Delta in Vietnam had different levels of drought occurrences. Studies reveal that the area experienced several instances of drought, varying in intensity, throughout the years. The Mekong Delta in Vietnam had severe drought occurrences in 1988, 1998, 2010, 2016, and 2020, resulting in substantial economic losses as a consequence of crop devastation and detrimental impacts on the ecosystem and the livelihoods of farmers [12,45].

2.2. Data Sources

Drought forecasting relies on various data sources to accurately monitor and predict drought conditions. These data sources provide information on the meteorological, hydrological, and environmental variables that are critical for understanding and forecasting drought events.

For this study, we analyzed data from 11 meteorological stations (1985–2022): Chau Doc, Moc Hoa, Cao Lanh, Can Tho, My Tho, Ba Tri, Cang Long, Soc Trang, Bac Lieu, Ca Mau, and Rach Gia (Figure 1 and Table 1). Rainfall and temperature datasets from these meteorological stations were collected from the Southern Regional Hydrometeorological Center (Vietnam Meteorological and Hydrological Administration). Long-term rainfall records help identify trends and anomalies in precipitation patterns. Temperature data are essential for calculating potential evapotranspiration, which is a crucial component of drought assessment.

Table 1. Descriptive statistics of meteorological stations in the Mekong Delta.

Station Name	Geographical Locations		Annual Mean Rainfall (mm)	Annual Mean Temperature (°C)
	Latitude	Longitude		
Chau Doc	10°42′12.7″ N	105°07′58.7″ E	1360	27.0
Cao Lanh	10°28′16.6″ N	105°38′42.1″ E	1356	27.0
Moc Hoa	10°45′12.6″ N	105°56′00.5″ E	1564	27.3
Can Tho	10°01′33.9″ N	105°46′07.8″ E	1544	26.6
My Tho	10°21′03.3″ N	106°23′53.9″ E	1349	26.7
Cang Long	9°59′33.7″ N	106°12′11.3″ E	1672	26.8
Ba Tri	10°02′30.6″ N	106°35′37.3″ E	1473	26.8
Soc Trang	9°36′05.2″ N	105°58′24.9″ E	1859	26.8
Bac Lieu	9°17′43.5″ N	105°42′50.1″ E	1712	26.8
Ca Mau	9°10′28.5″ N	105°10′41.5″ E	2366	26.7
Rach Gia	10°00′44.5″ N	105°04′37.7″ E	2057	27.6

The yearly precipitation in the Mekong Delta typically exceeds 1350–2366 mm. The Ca Mau–Rach Gia region experiences the highest levels of rainfall, with measurements ranging from 2000 to 2366 mm or more. Approximately 30% of the weather stations record a rainfall of 1300 mm or more. In contrast, areas such as My Tho receive a lower amount of precipitation, with levels ranging from approximately 1300 mm. The rainfall distribution exhibits temporal and spatial irregularities. This matter is intricately connected to the equilibrium and utilization of water resources to support the objective of sustainable development of water resources in the Mekong Delta. Figure 2 clearly illustrates that the rainfall was predominantly concentrated between June and October.

In addition to monthly rainfall and temperature information, we incorporated four climatic variables, including soil moisture, humidity, Southern Oscillation Index (SOI), and Equatorial Pacific sea surface temperatures (SSTs), to develop machine learning models. Soil moisture and humidity data were acquired via The Enhanced POWER Data Access Viewer (NASA’s Prediction of Worldwide Energy Resources (POWER)) website at <https://power.larc.nasa.gov/data-access-viewer/> (5 May 2023). SOI and SST data were obtained from the National Oceanic and Atmospheric Administration (NOAA) website at <https://www.ncei.noaa.gov/access/monitoring/enso> (5 May 2023). The data sources were constantly maintained, and no post-processing was performed.

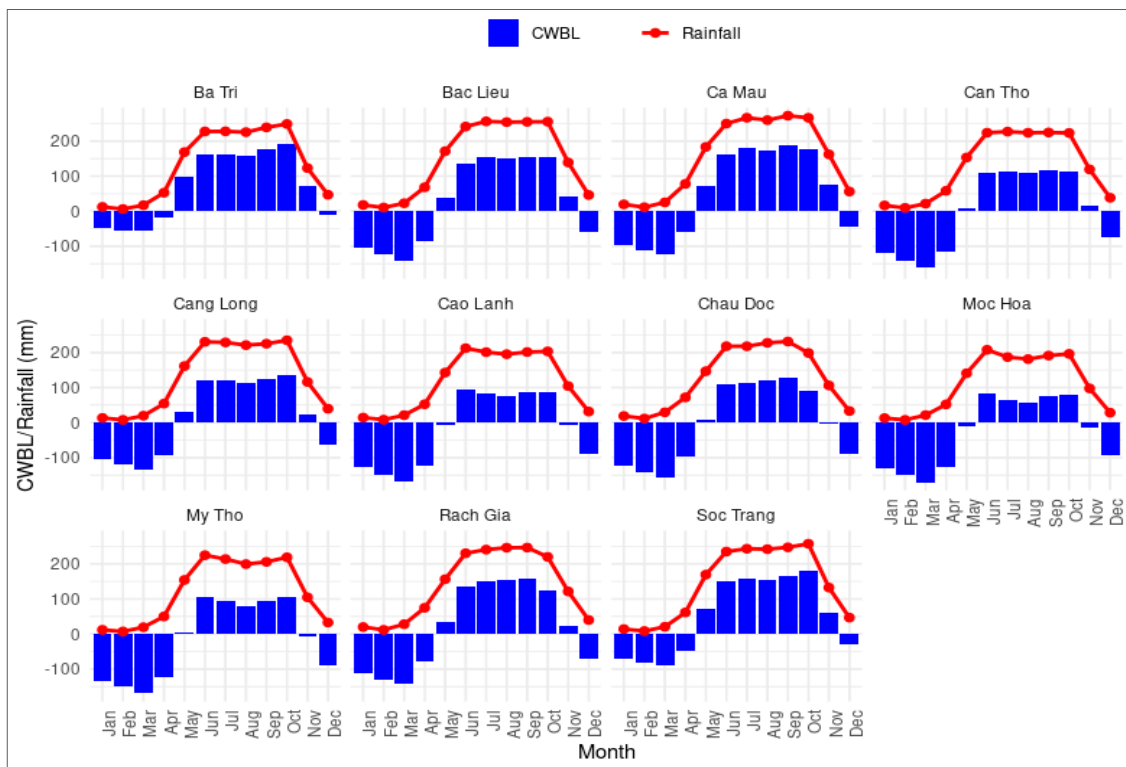


Figure 2. The average monthly precipitation and CWBL for eleven meteorological stations in the Mekong Delta from 1985 to 2022.

2.3. Methodology

2.3.1. Data Pre-Processing

Data pre-processing is an essential step to guarantee the integrity and dependability of the data utilized in modeling [46]. This entails the following:

1. Data cleaning:

- Handling missing values: Methods include replacing missing values with the mean, using regression to estimate missing values, or removing incomplete cases. The mean is commonly employed to replace missing values in numerical data when the data distribution is symmetrical and devoid of outliers. This approach is straightforward and maintains the average of the data. This method uses multiple imputation to generate several datasets with imputed values, which are subsequently merged to provide the final outcome. It takes into consideration the lack of certainty in the estimated values and is beneficial when there are more intricate connections between variables.
- Correcting inconsistencies and outliers: This involves identifying and resolving errors or outliers using data profiling and statistical methods. Outliers are identified as data points that fall far from the “cloud” of other points in a dataset and can significantly affect the results of a regression analysis. This study employs Huber regression to address outliers, as it is a robust technique that mitigates the impact of outliers by utilizing a combination of squared and absolute residuals. It exhibits greater robustness to outliers compared to ordinary least squares (OLS) regression.

2. Normalize data:

Ensures all variables are on the same scale, typically between 0 and 1, to improve algorithm accuracy. The formula used is as follows:

$$X_{norm} = \frac{X_0 - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X_{norm} is the normalized value, and X_0 , X_{min} , and X_{max} are the real value, minimum value, and maximum value of the same variable, respectively.

3. Split data: In machine learning, the data will be divided into two sets: a training set and a testing set. The training set will consist of 80% of the data, equivalent to 4012 records. The remaining 20% of the data, consisting of 1003 records, will be used to evaluate the model. In deep learning, the data are divided into three partitions: training, validation, and testing. The training set consists of 70% of the data, while the validation and testing sets each contain 15% of the data.

2.3.2. Method for Calculating Drought Index

In this study, the Hargreaves–Samani (HS) method was used to estimate reference evapotranspiration (ET₀) because this method has been used to estimate reference evapotranspiration (ET₀) in various regions. Studies have shown that the HS equation can accurately estimate ET₀ values when compared with the FAO Penman–Monteith (PM) method, which is considered the most accurate method for ET₀ estimation [47,48]. The HS equation performs well under different climatic conditions and environments, including regions with high altitudes [49]. In some cases, the HS equation outperforms the FAO56-PM method, particularly when meteorological data are limited or unavailable [50]. In general, the HS method has been demonstrated to be a viable option for calculating ET₀, especially in areas where the availability of data is limited, such as the Mekong Delta.

The Hargreaves–Samani (HS) approach calculates the reference evapotranspiration (ET₀) based solely on the highest and minimum temperatures [51,52], as shown in the following equation:

$$ET_0 = C_0 R_a (T_{\text{max}} - T_{\text{min}})^{0.5} (T + 17.8) \quad (2)$$

where

- R_a : extraterrestrial radiation (mm day⁻¹);
- C_0 : conversion parameter (=0.0023);
- T_{max} , T_{min} and T : the maximum, minimum and average temperature (°C).

Water balance, which refers to the excess or shortage of water, was computed accordingly.

$$D_i = P_i - E_{0i} \quad (3)$$

D_i values were then aggregated at different time scales. The log–logistic distribution $F(x)$ was applied to transform the original D series into standardized units at different time scales. Finally, the $F(x)$ distribution was used to calculate the SPEI [53].

The SPEI package for R [54] was used to compute the SPEI drought index. This tool serves as a valuable resource for the in-depth examination of drought conditions, both for research purposes and practical applications. Drought levels were classified according to the SPEI values, as indicated in Table 2.

Table 2. Characterization of drought using values of the standardized precipitation evapotranspiration index (SPEI).

SPEI	Drought Category
$\text{SPEI} \geq 2$	Extremely wet
$1.5 \leq \text{SPEI} < 1$	Severely wet
$1 \leq \text{SPEI} < 1.5$	Moderately wet
$-1 \leq \text{SPEI} < 1$	Near normal
$-1.5 \leq \text{SPEI} < -1$	Moderately dry
$-2 \leq \text{SPEI} < -1.5$	Severely dry
$\text{SPEI} < -2$	Extremely dry

2.3.3. Bayes Method (BMA)

The Bayes technique, also known as Bayesian Model Averaging (BMA), employs Bayesian statistics to address the tasks of model selection and averaging. The incorporation of the Bayes factor (BF) and the Bayesian Information Criterion (BIC) in this approach allows for the consideration of the trade-off between model complexity and predictive performance. These components aid in the assessment and choice of models that strike a balance between simplicity and complexity, hence enhancing the ability to generalize to new data [55,56].

Bayesian Model Averaging (BMA) effectively addresses the problem of duplication in multivariable linear regression by objectively identifying the variables that make a meaningful contribution to the model. By excluding factors that do not have a significant effect, the model's accuracy and interpretability are enhanced. The methodology utilizes probabilistic frameworks to calculate the average of many models, taking into account the uncertainty related to the parameters of each model [46].

If we have two models M_1 and M_2 and assume that one of them is true, the posterior probability of M_1 is as follows:

$$P_j(M_1|y) = \frac{P(\gamma|M_1)P(M_1)}{P(y|M_1)P(M_1) + P(y|M_2)P(M_2)} \quad (4)$$

In fact, we can also compare the two models M_1 and M_2 through real evidence:

$$\frac{P(M_1|\gamma)}{P(M_2|\gamma)} = \frac{P(\gamma|M_1)}{P(\gamma|M_2)} \times \frac{P(M_1)}{P(M_2)} \quad (5)$$

This ratio is called the Bayes factor (BF). In the above interpretation, BF gives us information that the data are toward M_1 or M_2 . With the BMA, each study does not have only one model, but there can be many models that can also explain γ .

Various elements, such as climate and meteorology, influence the outcomes of calculating the drought index, known as the Standardized Precipitation Evapotranspiration Index (SPEI). This study employed the Bayesian method to discover influential elements that significantly impact the Standardized Precipitation Evapotranspiration Index (SPEI) in the Mekong Delta, in order to establish the typical parameters for machine learning models. The application of the Bayesian method (BMA) in statistical analysis enables the identification of influential elements that significantly impact the SPEI value. Consequently, these factors may be determined as the primary parameters affecting the SPEI and can then be utilized as input parameters for the SPEI.

2.3.4. ARIMA Model

The ARIMA (AutoRegressive Integrated Moving Average) model functions by detecting patterns in past data and extrapolating these trends into the future. The ARIMA model is used to forecast the SPEI, a drought index that captures the effects of both precipitation and temperature on water demand. The ARIMA(p,d,q) model combines autoregressive AR(p), differencing I(d), and moving average MA(q) components. The autoregressive AR(p) component of the model entails using past values of the variable to predict its current value. The integrated I(d) component involves transforming the data to make them stationary. The moving average MA(q) component models the error term as a composite of error terms that occurred at the same time and in the past. p, d, and q are the parameters corresponding to the three models [19]. Mathematically, it can be represented or described as follows:

$$C(t) = \varphi_0 + \sum_{i=1}^p \varphi_i C_{t-i} + \varepsilon_t + \sum_{i=1}^q \gamma_i \varepsilon_{t-i} \quad (6)$$

where $C(t)$ represents the reconstructed component time series formed after the SE algorithm; ε_t represents the current period random error disturbance; φ_i and γ_i represent the model parameters; p denotes the quantity of autoregressive terms; and q denotes the amount of terms in the moving average.

2.3.5. Artificial Intelligence Model Selection

The accuracy of the models utilized is crucial for forecasting the SPEI. Scientists have investigated many Artificial Intelligence (AI) methods, such as machine learning models and deep learning models, to make precise predictions of the SPEI. According to the results of the literature review, the machine learning methods widely employed for SPEI prediction include machine learning models (Gradient Boosting, Extreme Gradient Boosting) and deep learning models (RNN and LSTM).

Gradient Boosting algorithms such as XGBoost and Gradient Boosting can be effectively used to predict the Standardized Precipitation Evapotranspiration Index. These algorithms are powerful machine-learning methods that can handle complex relationships between input variables and the SPEI [57]. By utilizing the principles of Gradient Boosting, these algorithms can iteratively refine the predictions and incorporate the strengths of multiple weak models into a strong predictive model. Moreover, research studies have shown that the XGBoost and LightGBM outperform traditional machine learning algorithms, such as decision trees, neural networks, and random forests, in terms of prediction accuracy for the SPEI [58]. Additionally, the incorporation of specific characteristics of each variable through weighting distance based on sensitivity coefficients was found to further improve the performance of these algorithms in predicting the SPEI. Furthermore, these Gradient Boosting algorithms have shown promising results in forecasting different seasons and multi-month-ahead reference evapotranspiration. In summary, Gradient Boosting algorithms, specifically XGBoost and Gradient Boosting, were highly effective in predicting the SPEI.

1. Gradient Boosting algorithms

Gradient Boosting algorithms are a collection of strategies that improve the performance of weaker models (learners) by progressively combining them to decrease bias and variation in supervised learning situations [59]. Gradient Boosting leverages the advantages of several models to construct a resilient predictive model that outperforms any individual poor learner [46].

The Gradient Boosting algorithm is an iterative method for optimizing a predictive model by minimizing a loss function [60]. This section details the step-by-step process of applying Gradient Boosting, including the relevant mathematical Formulas (7)–(11):

$$F_0(x) = \operatorname{argmin}_c \sum_i^N L(y_i, c) \quad (7)$$

where L is the loss function, y_i is the target values, and N is the number of data points.

For $m = 1$ to M (number of boosting rounds):

$$r_{im} = - \left[\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)} \right] \text{ for } i = 1, \dots, n \quad (8)$$

where r_{im} represents the residual for the i -th data point at the m -th iteration.

Fit a base learner $h_m(x)$ to the residuals:

$$h_m(x) = \operatorname{argmin}_h \sum_i^N (r_{im} - h(x_i))^2 \quad (9)$$

Update the model with the new learner:

$$F_m(x) = F_{m-1}(x) + v \cdot h_m(x) \quad (10)$$

where v is the learning rate, controlling the contribution of each learner.

After M iterations, the final model is as follows:

$$F_M(x) = F_0(x) + \sum_{m=1}^M v \cdot h_m(x) \quad (11)$$

Below is a thorough analysis of the functioning of Gradient Boosting:

- Model initialization: The procedure commences by constructing an initial model utilizing the training data. The model generates predictions based on the training data, and subsequently calculates the residual errors, which represent the discrepancies between the actual values and the anticipated values.
- Sequential model addition: A novel model is trained to forecast the discrepancies between the preceding model's predictions and the actual values. The newly introduced model is incorporated into the ensemble, and the collective predictions of all existing models are utilized to revise the residuals.
- Weight adjustment and reweighting: The data points' weights are modified to prioritize the previously misclassified or poorly forecasted points. This procedure is iterated, wherein each subsequent model rectifies the inaccuracies of the collective ensemble of preceding models.
- Iterative process: Models are incrementally included until the training data are accurately predicted or a predetermined maximum number of models is attained. Every iteration has the objective of minimizing the total prediction error by dealing with the leftover residuals.

2. Extreme Gradient Boosting (XGBoost):

XGBoost is a machine learning technique that uses a Gradient Boosting framework to improve the accuracy of predictions. The algorithm lends greater importance to incorrectly classified data pieces, prioritizing their accurate prediction in subsequent iterations [61]. XGBoost enhances existing Gradient Boosting algorithms by integrating regularization approaches to reduce overfitting and employing advanced optimization techniques to improve computational performance [46].

The purpose of the model is simplification through the optimizations of the training loss (l) and regulations (Ω). f_k is the function of the K-tree. The objective function (J) in round t is given by Equation (12).

$$J^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (12)$$

Here is a concise summary of its functionality:

- Model generation: An initial decision tree is constructed using the initial data. The calculation involves determining the discrepancy between the projected values and the actual observations, which is referred to as residuals.
- Subsequent models: Additional trees are constructed to forecast the discrepancies from the preceding model. These algorithms prioritize the analysis of data points that were previously misclassified or inaccurately anticipated.
- Optimization involves the ongoing addition of new trees, where each tree aims to rectify the mistakes made by the preceding trees. The designated loss function, such as mean squared error, is optimized by utilizing the residuals obtained from each stage.
- Iteration and combination: This iterative process is carried out several times. The ultimate model is an amalgamation of all the separate trees, with each tree making a contribution to the overall prediction.

3. Recurrent Neural Networks (RNNs)

RNNs are characterized by their ability to maintain a memory of previous inputs through internal states, which makes them suitable for tasks involving sequences. Unlike traditional feedforward neural networks, RNNs have connections that form directed cycles, allowing information to persist.

Given an input sequence x_t , the hidden states h_t at time step t are updated as shown in the following equations:

$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_h) \quad (13)$$

where

- x_t is the input at time step t ;
- h_{t-1} is the hidden state from the previous time step;
- W_h and U_h are weight matrices;
- b_h is the bias vector;
- σ is an activation function, typically tanh ReLU.

The output y_t at time step t is given by

$$y_t = \sigma(W_y h_t + b_y) \quad (14)$$

where W_y is the weight matrix and b_y is the bias vector for the output layer.

4. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a specific type of Recurrent Neural Network that excel at learning order dependence in sequence prediction tasks. The output from the previous step is utilized as the input for the following step in a Recurrent Neural Network (RNN). It tackled the problem of long-term reliance in Recurrent Neural Networks (RNNs), where the RNN struggles to predict words that are stored in its long-term memory but can make more precise predictions using the most recent data. As the gap length increases, RNN's performance becomes less efficient. By default, the LSTM has the ability to retain information for an extended period. It is utilized for the processing, prediction, and classification of time-series data [62].

The LSTM is made up of four neural networks and numerous memory blocks known as cells in a chain structure. A conventional LSTM unit consists of a cell, an input gate, an output gate, and a forget gate. The flow of information into and out of the cell is controlled by three gates, and the cell remembers values over arbitrary time intervals. The LSTM algorithm is well adapted to categorize, analyze, and predict time series of uncertain duration.

The cells store information, whereas the gates manipulate memory. There are three entrances:

- Input gate: It determines which of the input values should be used to change the memory. The sigmoid function determines whether to allow 0 or 1 values through. The tanh function assigns weight to the data provided, determining their importance on a scale of -1 to 1 (Equations (15) and (16)).

$$i_t = \sigma(W_i[h_t - 1, x_t] + b_i) \quad (15)$$

$$C_t = \tanh(W_C[h_t - 1, x_t] + b_C) \quad (16)$$

where W_i is a weight of input gate, b_i is the bias value on the input gate, C_t is the cell state, b_C is the bias value on cell state, and σ is the sigmoid function. The sigmoid function is defined by Equation (17)

$$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (17)$$

- Forget gate: It finds the details that should be removed from the block. It is decided by a sigmoid function. For each number in the cell state C_{t-1} , it looks at the preceding state (h_{t-1}) and the content input (x_t) and produces a number between 0 (omit this) and 1 (keep this), as shown in Equation (18).

$$f_t = \sigma(W_f[h_t - 1, x_t] + b_f) \quad (18)$$

where f_t is a forget gate, σ is the sigmoid function, x_t is the input value in t order, W_f is the weight of forget gate, b_f is the bias value on the forget gate, and h_t is the output value in t order. The output of the t model is obtained from the output of the previous model or $t - 1$ model $t - 1$.

- Output gate: The block's input and memory are used to determine the output. The sigmoid function determines whether to allow 0 or 1 values through. The tanh function determines which values are allowed to pass through 0, 1. Furthermore, the tanh function assigns weight to the values provided, determining their relevance on a scale of -1 to 1 and multiplying it with the sigmoid output. (Equations (19)–(20)).

$$O_t = \sigma(W_o[h_t - 1, x_t] + b_o) \quad (19)$$

$$h_t = o_t \tanh(C_t) \quad (20)$$

where W_o is a weight of output gate, b_o is the bias value on cell state, C_t is the cell state, and o_t is the output gate.

The Recurrent Neural Network uses Long Short-Term Memory blocks to provide context for how the software accepts inputs and creates outputs. Because the program uses a structure based on short-term memory processes to build longer-term memory, the unit is dubbed a Long Short-Term Memory block.

Incorporating Bidirectional LSTM and LSTM with an Attention Layer in the study of SPEI forecasting can be advantageous as they provide advanced skills in capturing temporal relationships and focusing on key areas of the input sequence.

- A Bidirectional LSTM utilizes a Recurrent Neural Network architecture that processes data in both the forward and backward directions. This enables the model to acquire knowledge from preceding and subsequent states, enhancing its efficacy in comprehending temporal dynamics.
- The attention mechanism enables the model to choose and concentrate on particular segments of the input sequence that are more pertinent to the prediction task. This can enhance the model's capacity to acquire significant temporal patterns and interconnections. Incorporate an Attention Layer above the LSTM layer. This entails calculating attention scores for each time step and subsequently adjusting the LSTM outputs based on these values.

2.3.6. Model Evaluation Method

The evaluation of the accuracy of models forecasting the Standardized Precipitation Evapotranspiration Index (SPEI) in the Mekong Delta is based on the following indicators: Mean Absolute Error (MAE), the Mean Square Error (MSE), Root Mean Square Errors (RMSEs), the coefficient of determination (R^2), and the Mean Absolute Percentage Error (MAPE) [46]. The criteria for evaluating (calibrating) the models are presented in Equations (21)–(25):

$$MAE = \frac{1}{N} \sum_{i=1}^n |P_i - M_i| \quad (21)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (P_i - M_i)^2 \quad (22)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - M_i)^2} \quad (23)$$

$$R^2 = 1 - \left(\frac{ESS}{TSS} \right) \quad (24)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{M_i - P_i}{M_i} \right| \times 100 \quad (25)$$

3. Results

3.1. SPEI Calculation

Figure 3 shows the SPEI calculation results for 11 meteorological stations in the Mekong Delta region during various time periods (1, 3, 6, and 12 months). The Mekong Delta region has experienced droughts from 1985 to 1990 and from 2010 to 2016.

- Stations with an SPEI < −2 (extremely dry) include Can Tho. These stations received less rainfall than the other stations; hence, the SPEI was low.
- Stations with an SPEI ≥ 2 (extremely wet) included Ba Tri, Bac Lieu, Ca Mau, My Tho, and Soc Trang (1989). These coastal stations receive more rainfall than the other stations; hence, the SPEI is high.

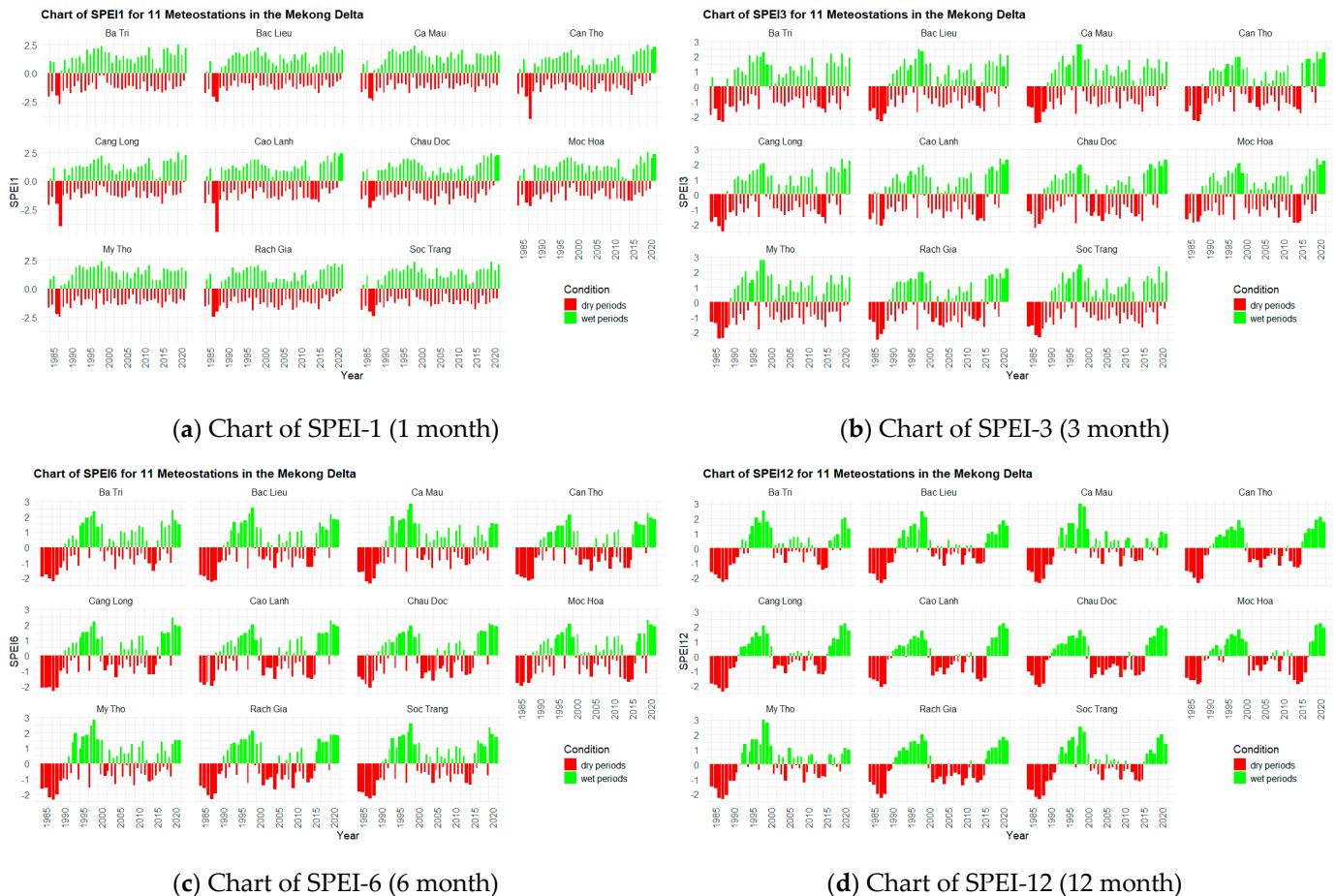


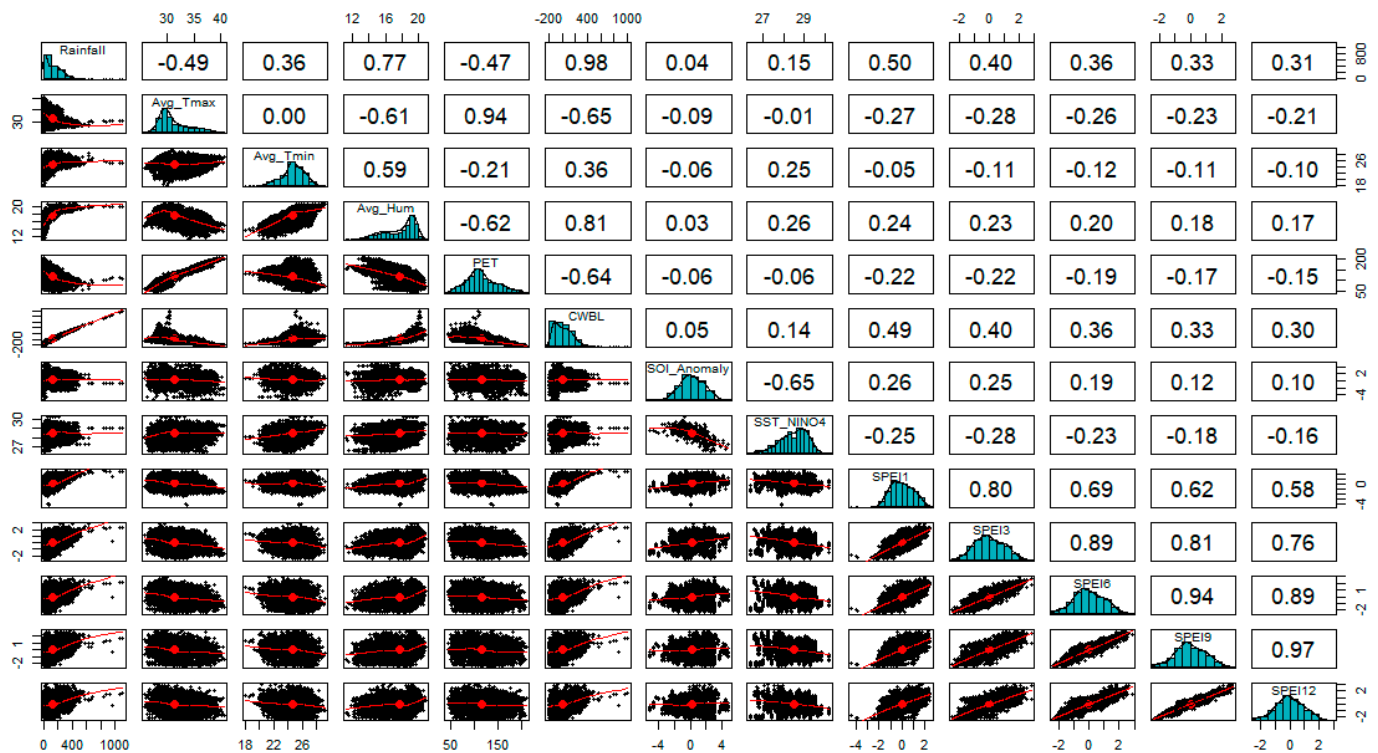
Figure 3. Chart of SPEI of 11 meteorological stations in the Mekong Delta (1985–2022).

3.2. Feature Selection Results by BMA

Feature selection is a technique that involves minimizing the number of input variables in a model by retaining only the pertinent data and eliminating irrelevant or noisy information. Based on the findings of the correlation study, it is evident that the correlation coefficients between the SPEI1, SPEI3, SPEI6, SPEI9, and SPEI12 indices were quite high, ranging from 0.57 to 0.96. The correlation coefficient values for the pairs of variables were as follows: 0.79 for SPEI1 and SPEI3, 0.88 for SPEI3 and SPEI6, 0.93 for SPEI6 and SPEI9, and 0.96 for SPEI9 and SPEI12. Thus, this study exclusively chose the indicators SPEI1, SPEI3, SPEI6, and SPEI12 to construct the models. The correlation coefficient between the SPEI and meteorological parameters exhibited a relatively low range, varying from 0.09 to 0.48. Hence, the task of selecting ideal parameters for calculating the SPEI using machine learning models is a challenge, necessitating the use of a method for identifying crucial factors (Figure 4).

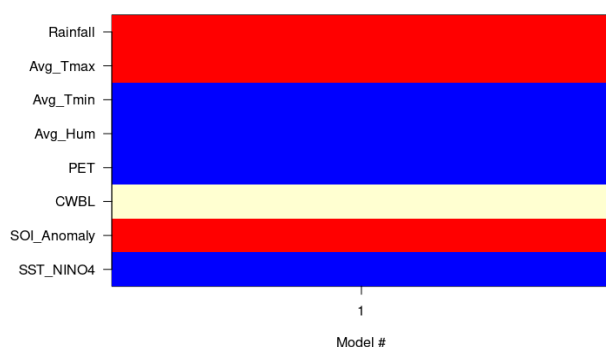
The nonlinear association between SPEI and climatic parameters is evident. This study utilized the Bayesian Model Averaging method (BMA) to choose the most suitable parameters. The statistical analysis results obtained using the BMA are presented in Figure 5. The BMA technique identified the five best models by selecting the essential parameters:

- The model for SPEI-1: seven parameters were selected: Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, and SST_NINO4 (posterior probability was 100%).
- The model for SPEI-3: four parameters were selected: Rainfall, Avg_Tmin, Avg_Hum, and SST_NINO4 (posterior probability was 92.5%).
- The model for SPEI-6: four parameters were selected: Rainfall, Avg_Tmin, Avg_Hum, and SST_NINO4 (posterior probability was 100%).

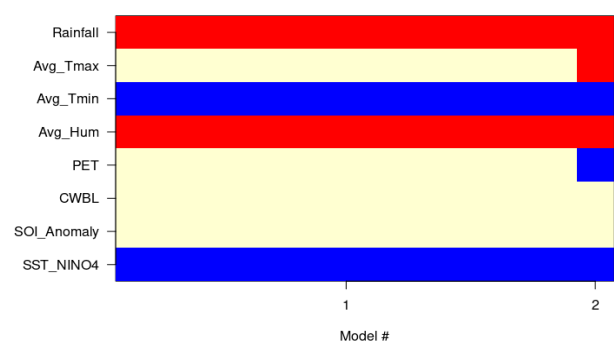


Note: The diagonal shows distribution histograms for each variable (blue). Off-diagonal elements display scatterplots (black and red dots) and correlation coefficients. Red dots indicate specific data points of interest.

Figure 4. Correlation chart of drought index and meteorological parameters.

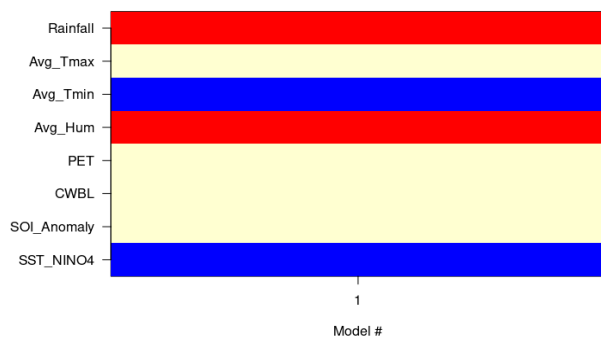


(a) Models SPEI-1 selected by BMA

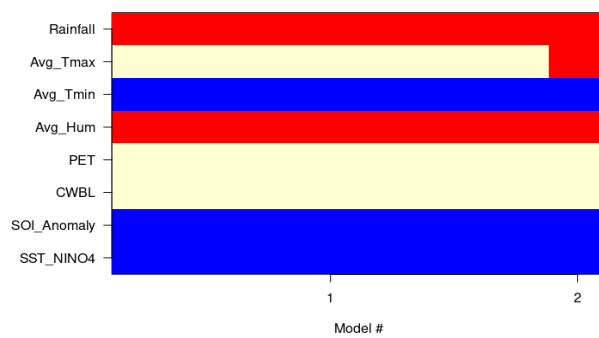


(b) Models SPEI-3 selected by BMA

Figure 5. Cont.



(c) Models SPEI-6 selected by BMA



(d) Models SPEI-12 selected by BMA

Note: Red bars indicate variables included with positive effects, blue bars show variables with negative effects, and the light yellow bar represents a variable with an effect close to zero.

Figure 5. Models selected by BMA (Bayesian Model Averaging).

The model for SPEI-12:5 parameters were selected as Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4 (posterior probability was 88.4%).

3.3. Results of Evaluating Machine Learning Models

This study established four models for the SPEI-1, SPEI-3, SPEI-6, and SPEI-12 indices using the BMA to predict only the SPEI. The models were developed based on various time periods and included ARIMA, Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM). Table 3 displays the outcomes of the hyperparameter-tweaking process.

Table 3. Table of results of hyperparameter tuning.

No.	Model Name	Hyperparameter Tuning
1	ARIMA	ARIMA(3,1,1) (1,0,1) [13] p = 3; d = 1; q = 1 P = 1; D = 0; Q = 1 s = 12 Distribution = "Gaussian".
2	Gradient Boosting (GB)	cv.folds = 10: shrinkage parameter = 0.01. Each terminal node should have at least 10 observations: n.minobsinnode = 10. n.trees = 1000.
3	eXtreme Gradient Boosting (XGBoost)	The number of trees (nround = 1000); The shrinkage parameter λ (eta in the params): 0.01; The number of splits in each tree: max.depth = 5.
4	Recurrent Neural Networks (RNNs)	- learning_rate = 0.001 - epochs = 1000 - activation = 'relu' - optimizer = 'adam'
5	Long Short-Term Memory (LSTM)	- learning_rate = 0.001 - epochs = 1000 - batch_size = 32 - validation_split = 0.2 - verbose = 1 - activation = 'relu' - optimizer = 'adam' - loss = 'mean_squared_error'

The table provides information about hyperparameter tuning for five different models. Every model possesses distinct hyperparameters that have been established or adjusted. The ARIMA model includes parameters for autoregressive (p), differencing (d), and moving average (q) components, as well as seasonal components. The machine learning models such as Gradient Boosting and XGBoost possess parameters associated with the quantity of trees, learning rate (shrinkage), and tree depth. The neural network models, specifically the Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM), are characterized by several adjustable parameters, including the learning rate, number of epochs, batch size, and activation functions.

The SPEI is calculated at different timescales (1, 3, 6, and 12 months) to capture short-term and long-term drought conditions (for six prediction steps). The forecast results for SPEI-1, SPEI-3, SPEI-6, and SPEI-12 are presented in Figures 6–10 along with a comparison chart illustrating the discrepancy between the predicted and computed SPEI values for the test data.

The performance of each model was assessed using four evaluation metrics: the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), and Mean Absolute Percentage Error (MAPE). These metrics are commonly used in regression analysis to quantify the accuracy, precision, and goodness-of-fit of the predictive models. Table 4 shows the results of the evaluation of the machine learning models (based on the four criteria) to predict the SPEI (for six prediction steps) in the study area (Table 4).

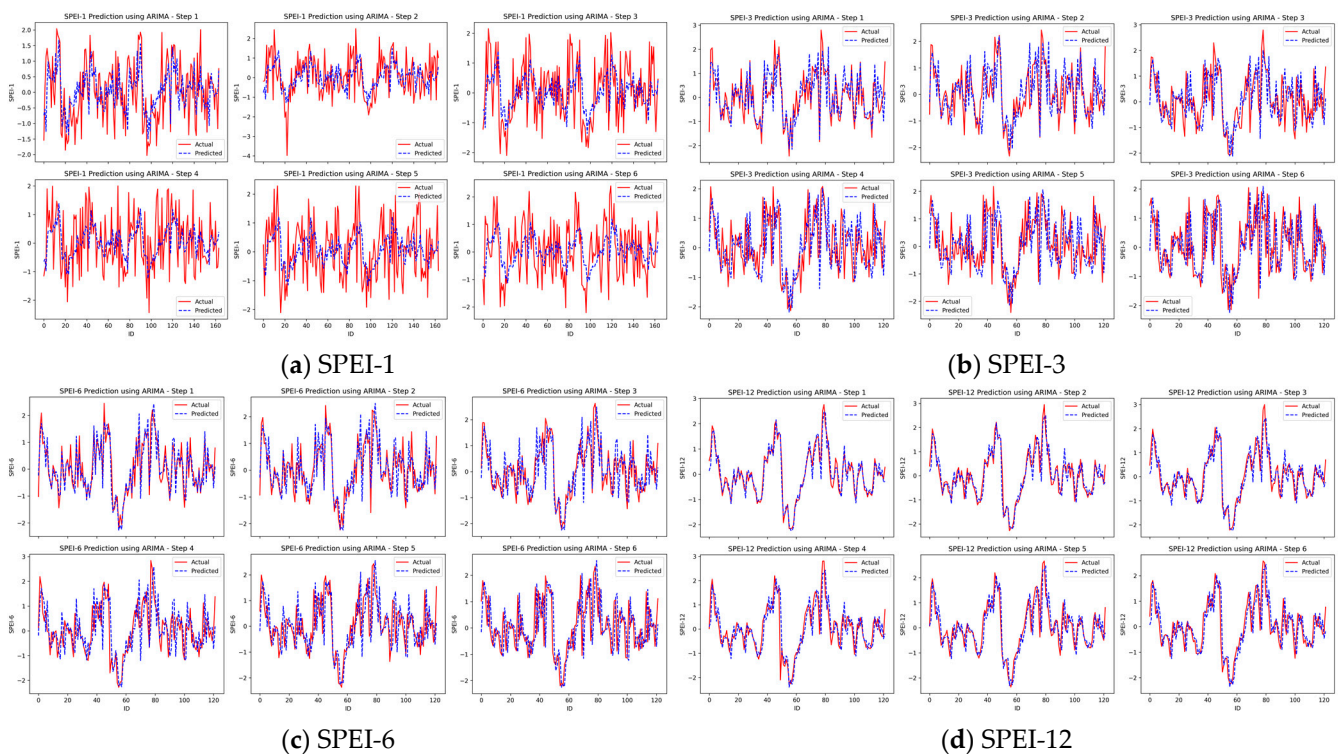


Figure 6. Comparison chart displaying the forecasted and actual SPEI values of the ARIMA model, with a prediction horizon of 6 steps.

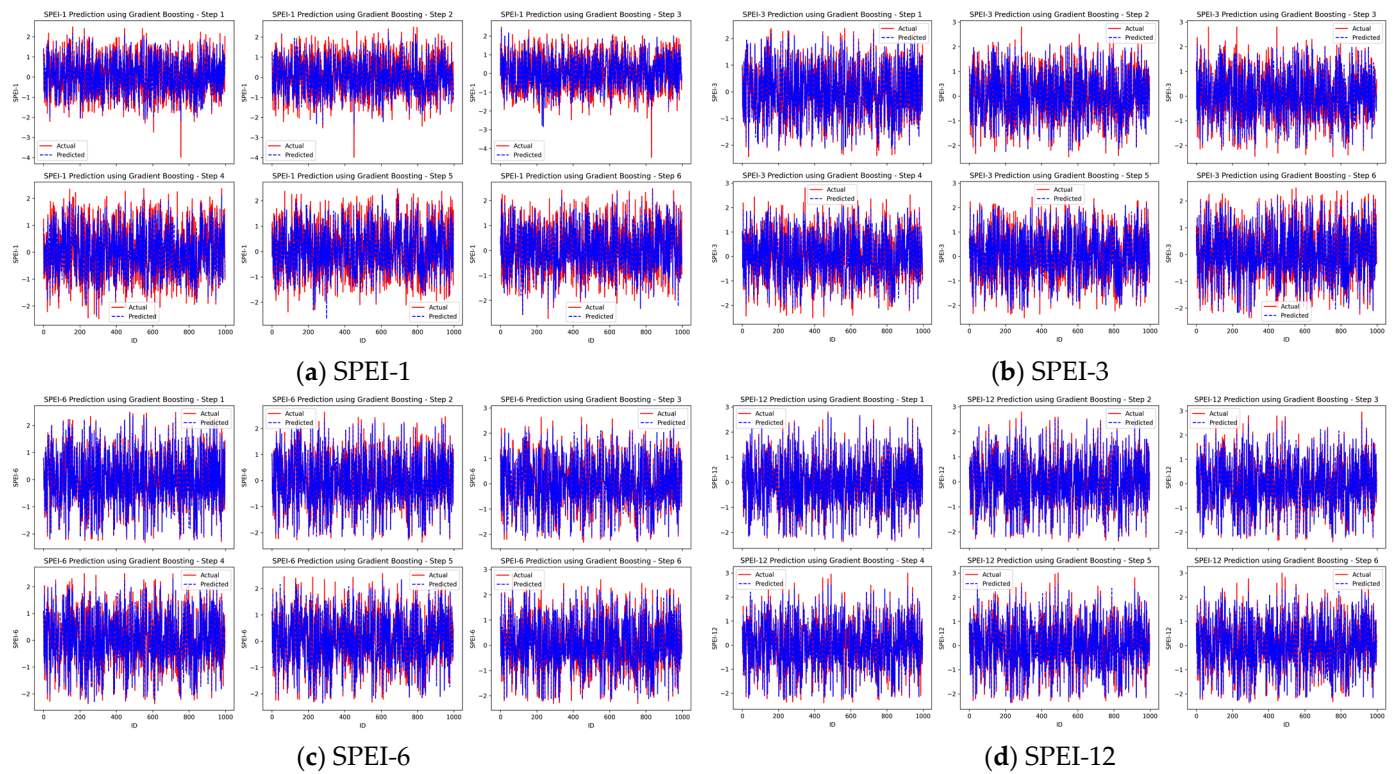


Figure 7. Comparison chart displaying the forecasted and actual SPEI values of the Gradient Boosting model, with a prediction horizon of 6 steps.

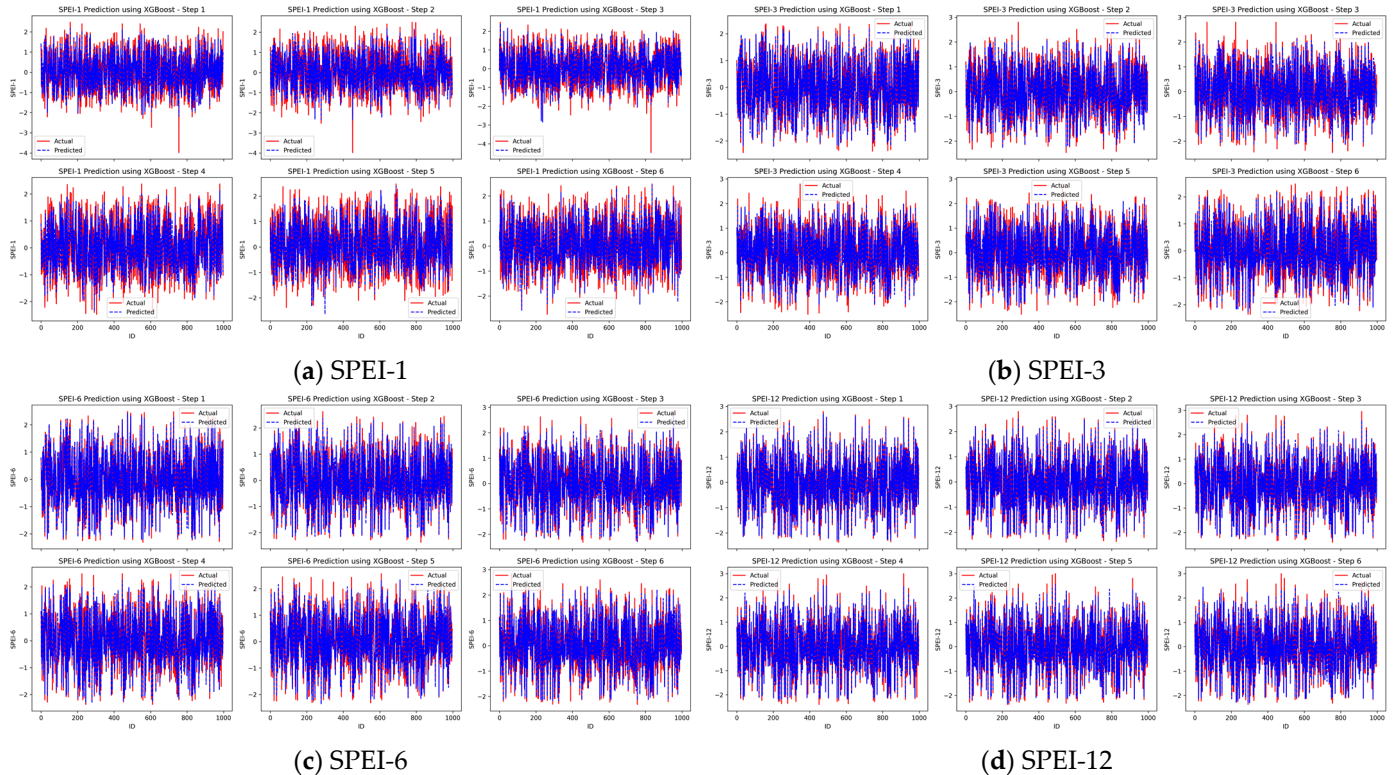


Figure 8. Comparison chart displaying the forecasted and actual SPEI values of the eXtreme Gradient Boosting model, with a prediction horizon of 6 steps.

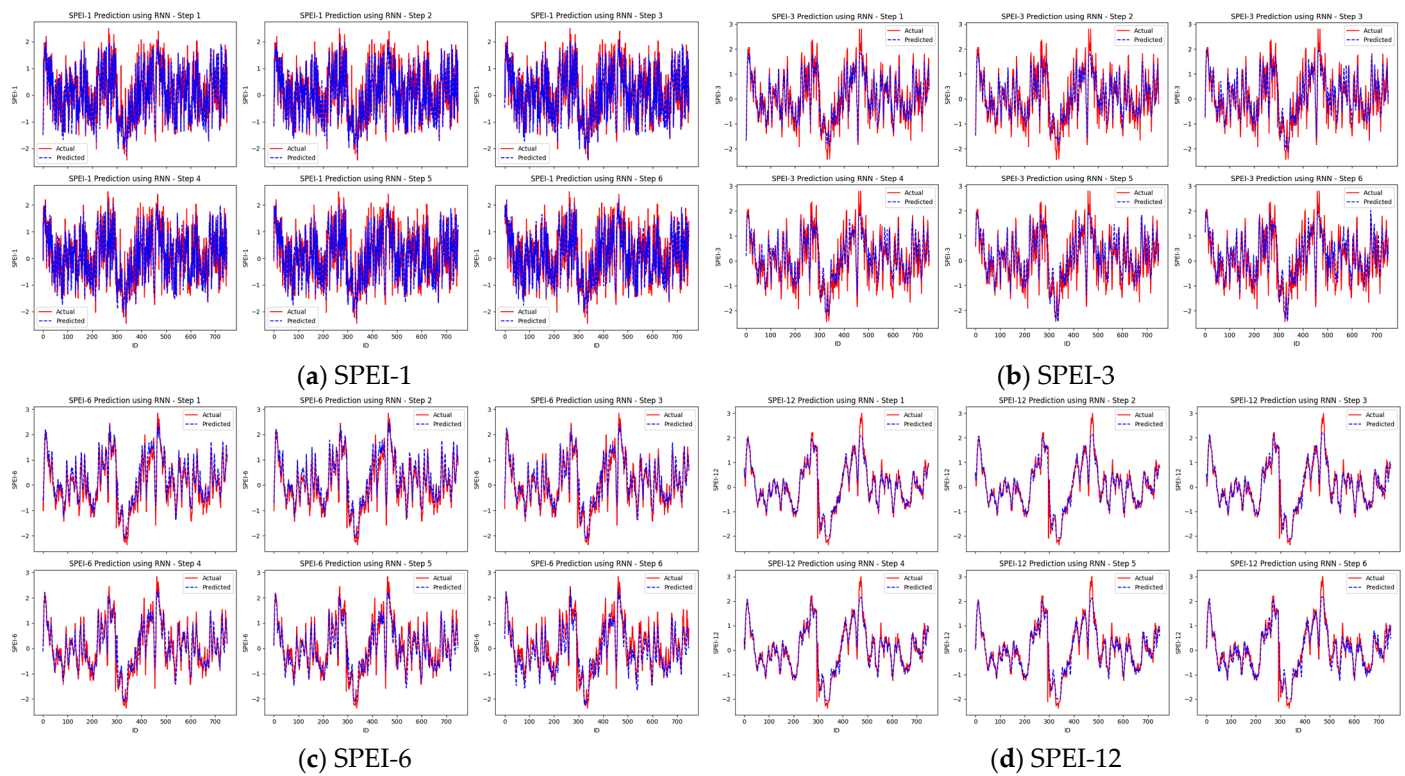


Figure 9. Comparison chart displaying the forecasted and actual SPEI values of the RNN model, with a prediction horizon of 6 steps.

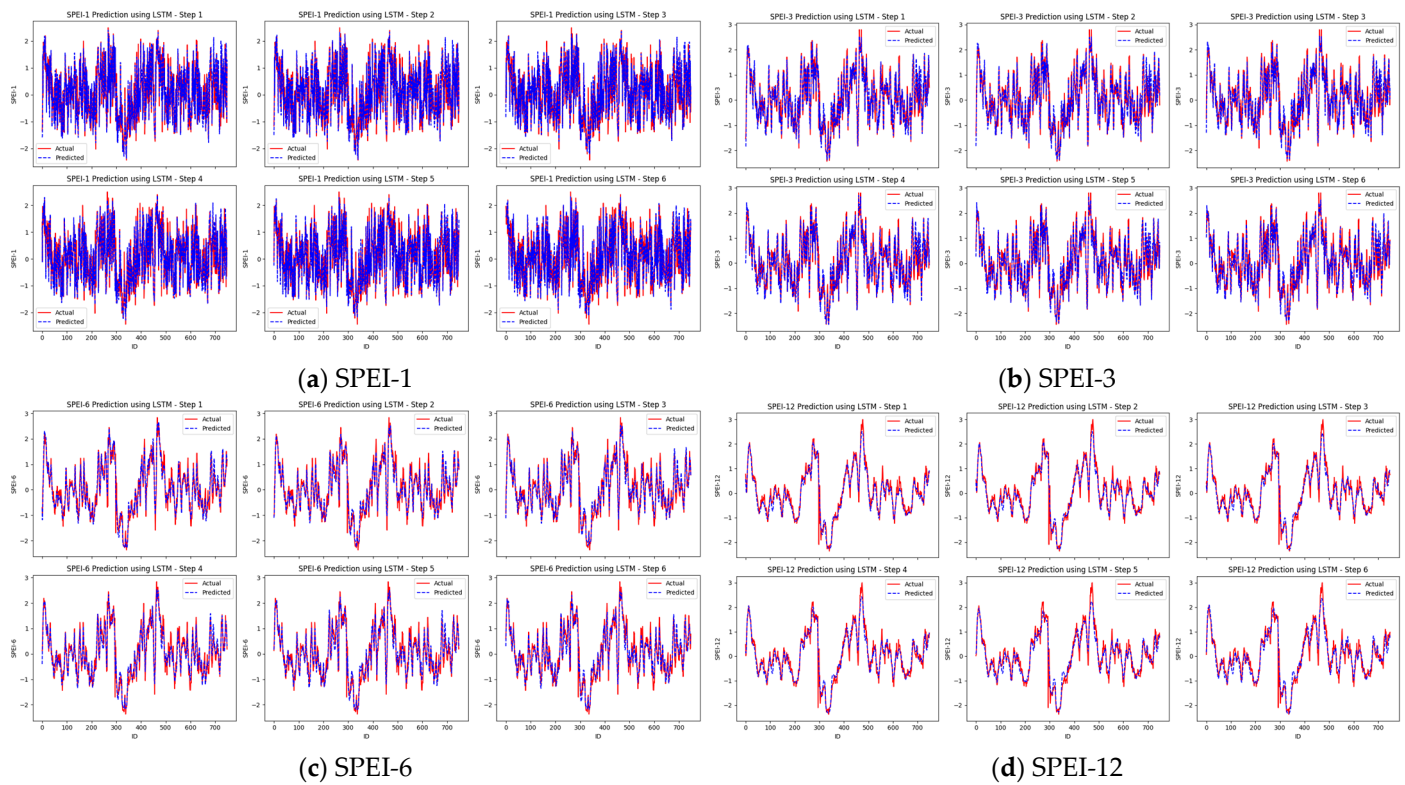


Figure 10. Comparison chart displaying the forecasted and actual SPEI values of the LSTM model, with a prediction horizon of 6 steps.

Table 4. Statistical table of evaluation results of models to predict the SPEI (6 prediction steps).

Models	Input Parameters	Output	Evaluation Criteria				
			MAE	MSE	RMSE	R ²	MAPE
ARIMA	Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, SST_NINO4	SPEI-1	0.34–0.40	0.22–0.28	0.46–0.50	0.73–0.75	23.9–28.6
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-3	0.36–0.48	0.24–0.39	0.48–0.64	0.72–0.75	24.1–28.0
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-6	0.30–0.40	0.14–0.23	0.38–0.51	0.81–0.84	15.6–20.1
	Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4	SPEI-12	0.22–0.50	0.08–0.40	0.28–0.62	0.74–0.91	9.00–25.4
Gradient Boosting	Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, SST_NINO4	SPEI-1	0.28–0.33	0.14–0.17	0.39–0.43	0.82–0.84	14.4–17.4
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-3	0.19–0.26	0.07–0.11	0.26–0.35	0.88–0.91	6.30–10.3
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-6	0.13–0.22	0.04–0.09	0.19–0.29	0.88–0.92	3.53–8.13
	Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4	SPEI-12	0.08–0.17	0.03–0.06	0.18–0.26	0.91–0.93	3.28–6.21
XGBoost	Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, SST_NINO4	SPEI-1	0.28–0.32	0.13–0.16	0.36–0.40	0.84–0.87	13.2–16.3
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-3	0.19–0.25	0.06–0.10	0.25–0.32	0.90–0.93	6.20–10.0
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-6	0.13–0.22	0.03–0.08	0.18–0.28	0.92–0.96	3.40–7.70
	Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4	SPEI-12	0.08–0.16	0.03–0.06	0.17–0.24	0.94–0.97	3.00–5.70
RNN	Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, SST_NINO4	SPEI-1	0.24–0.27	0.10–0.12	0.31–0.35	0.87–0.89	10.7–13.0
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-3	0.26–0.33	0.12–0.18	0.34–0.42	0.81–0.87	12.7–19.3
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-6	0.24–0.30	0.09–0.15	0.30–0.38	0.83–0.89	10.2–16.4
	Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4	SPEI-12	0.14–0.19	0.04–0.07	0.21–0.26	0.92–0.95	4.80–7.60
LSTM	Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, SST_NINO4	SPEI-1	0.18–0.19	0.05–0.06	0.23–0.25	0.93–0.94	5.80–6.70
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-3	0.16–0.19	0.05–0.06	0.22–0.25	0.93–0.95	5.30–6.70
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-6	0.17–0.20	0.06–0.07	0.24–0.27	0.92–0.94	6.40–8.00
	Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4	SPEI-12	0.16–0.19	0.05–0.06	0.22–0.26	0.93–0.95	5.30–7.10

This table outlines the performance of the machine learning (Gradient Boosting and XGBoost models) and deep learning (RNN and LSTM) trained to predict the Standardized Precipitation Evapotranspiration Index (SPEI) over different time scales: 1, 3, 6, and 12 months (for six prediction steps). The input parameters for each model vary slightly but generally include measures of rainfall, temperature (maximum and minimum averages), humidity, potential evapotranspiration (PET), Southern Oscillation Index (SOI) Anomaly, and sea surface temperature in the NINO4 region (SST_NINO4).

As the prediction step grows from 1 to 6, there is a consistent decline in model performance across all metrics. The Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) generally exhibit an upward trend, whereas the R2 Score shows a decline. The ARIMA model exhibits the most prominent manifestation of this tendency, while the machine learning models, including XGBoost and Gradient Boosting, display a less severe impact. In the XGBoost model for SPEI3, the MAE increases from 0.18 at Step 1 to 0.24 at Step 6. Simultaneously, the R2 Score declines from 0.938 to 0.91. This trend indicates that as the forecast horizon increases, all models face greater difficulty in making precise forecasts. This behavior is commonly observed in time series forecasting.

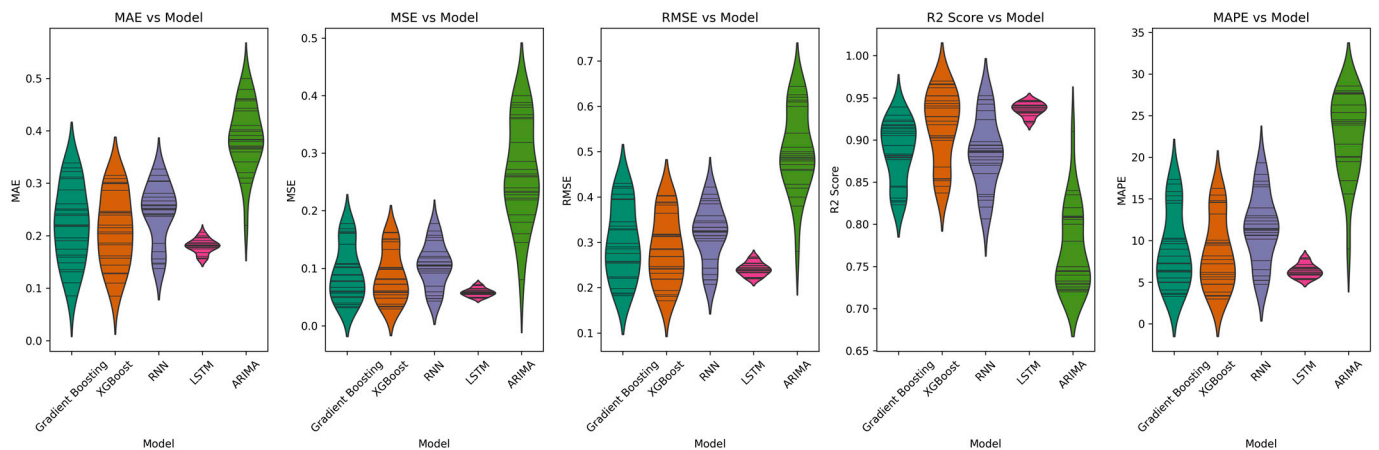
When comparing model performance, there is a noticeable pattern of improvement as the SPEI scale grows from 1 month to 12 months. This is clearly apparent in all measurements and frameworks. In the Gradient Boosting model, at Step 1, the MAE falls from 0.32 for SPEI1 to 0.085 for SPEI12. At the same time, the R2 Score increases from 0.84 to 0.94. The reason for this improvement is probably because longer-term drought indices, such as SPEI6 and SPEI12, are less influenced by short-term fluctuations and so exhibit greater predictability. The MAPE values exhibit a significant decline as the SPEI scale grows, suggesting that the percentage mistakes are considerably reduced for predictions made over longer periods.

Model comparison: XGBoost and LSTM consistently demonstrate superior performance compared to the other models across all criteria. They attain the minimum MAE, MSE, RMSE, and MAPE values, as well as the maximum R-squared (R2) scores. Gradient Boosting and RNN models demonstrate moderate performance; however, ARIMA consistently exhibits inferior performance. As an illustration, in the case of SPEI12 at Step 1, XGBoost obtains a Mean Absolute Error (MAE) of 0.084 and an R2 Score of 0.97, whereas ARIMA has an MAE of 0.22 and an R2 Score of 0.91. These findings indicate that tree-based machine learning and deep learning models are better at capturing the intricate relationships in SPEI data, regardless of the SPEI scales and steps.

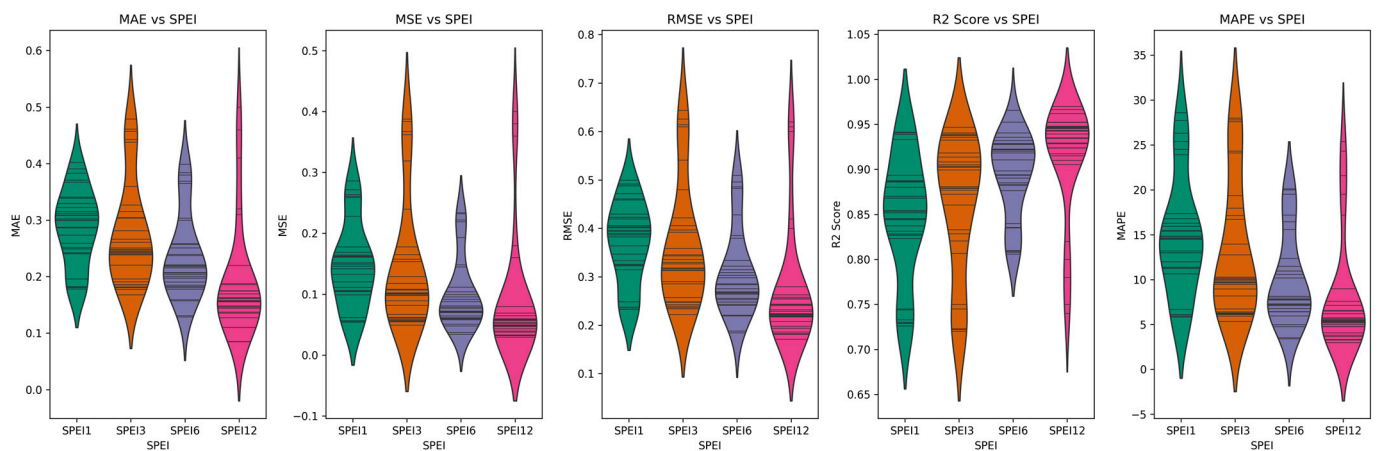
The interaction between Step, SPEI, and Model: the effect of increasing steps is more noticeable for shorter SPEI scales and less sophisticated models. ARIMA's performance deteriorates more rapidly over time for SPEI1 compared to SPEI12. In contrast, XGBoost and LSTM exhibit consistent performance across the iterations, particularly for longer SPEI scales. This interaction implies that more advanced models are more effective at maintaining accurate predictions over extended forecast periods, especially for long-term drought indices (Figure 11).

The extensive analysis of metrics across Steps, SPEI scales, and Models presents compelling evidence for the higher effectiveness of machine learning techniques, namely tree-based methods, in forecasting SPEI. Furthermore, it emphasizes the growing reliability of longer-term drought indices and the widespread difficulty of maintaining accurate predictions over long forecast periods.

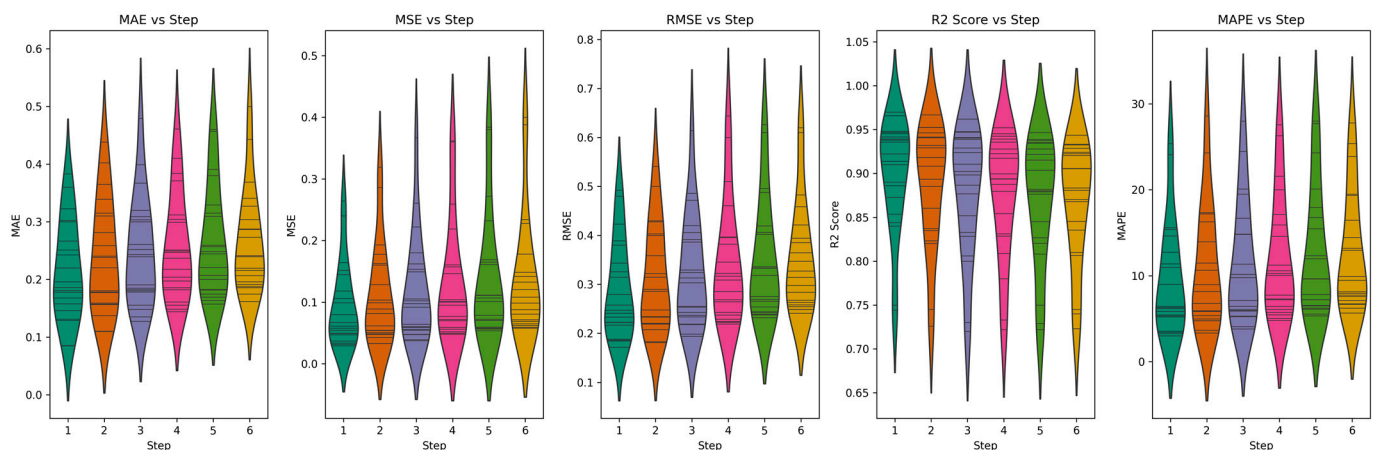
Through thorough comparisons, it becomes clear that LSTM routinely surpasses other models, particularly when it comes to making predictions for longer-term SPEI. The architecture of this model, specifically built for processing time series data, successfully captures temporal relationships, leading to reduced error metrics and improved R2 scores. Gradient Boosting exhibits strong performance, although it is marginally less precise than LSTM when making predictions over longer time intervals. XGBoost, although it performs well compared to other methods, often shows greater error metrics, suggesting that further optimization may be necessary. Recurrent Neural Network (RNN) models, while beneficial, typically exhibit elevated levels of errors, indicating that they may not be well-suited for this specific forecasting task. Overall, LSTM and Gradient Boosting models demonstrate notable efficacy, particularly for extended SPEI durations and prediction steps. These insights are essential for choosing the suitable model for drought forecasting applications, guaranteeing precise and dependable predictions over various time scales.



(a) Comparison chart displaying four evaluation measures based on different models



(b) Comparison chart displaying four evaluation measures based on SPEI



(c) Comparison chart displaying four evaluation measures based on prediction steps

Figure 11. Model performance comparison chart.

4. Discussion

This study evaluates the performance of various models (ARIMA, Gradient Boosting, XGBoost, RNN, and LSTM) across different prediction steps and SPEI durations using multiple error metrics: MAE, MSE, RMSE, R2 Score, and MAPE. Each metric provides unique insights into model accuracy and robustness, highlighting strengths and weaknesses in different contexts.

The performance of the models varies for each step, with the MAE and RMSE often being larger in the beginning steps and improving as the prediction horizons increase. The observed pattern indicates that models may demonstrate improved adaptation to data patterns as the forecast timeframe increases. For example, LSTM regularly demonstrates low MAE and RMSE values at each step, suggesting its effectiveness in capturing temporal relationships. Gradient Boosting exhibits strong performance; however, it is slightly inferior to LSTM, especially during the first steps.

Observations revealed a distinct pattern of enhanced model performance as the SPEI scale increased, encompassing all models and indicators. The enhancement was most noticeable while transitioning from SPEI1 and SPEI3 to SPEI6 and SPEI12. The observed trend indicates that it is easier to forecast long-term drought conditions compared to short-term variations. Multiple variables can potentially influence this phenomenon:

- Long-term indices are less influenced by short-term weather fluctuations, rendering them more stable and predictable.
- The process of collecting data over extended periods can reduce the impact of random variations and oscillations, enabling models to accurately identify and analyze underlying patterns.
- Prolonged drought conditions may have a stronger correlation with large-scale climatic trends, which often change at a slower and more predictable rate.
- These findings have significant ramifications for the prediction and control of droughts. They indicate that precisely anticipating short-term drought conditions is still difficult, but there is a higher level of confidence in predicting longer-term drought outlooks.

Out of all the models, LSTM regularly performs better than the others, particularly when it comes to making predictions for longer time periods and SPEI durations. It demonstrates superior performance by achieving lower MAE, MSE, and RMSE values. Additionally, it exhibits higher R-squared (R^2) scores, indicating its ability to accurately capture intricate patterns in time series data. Gradient Boosting exhibits impressive performance, especially in regard to its R^2 score and MAPE, which demonstrates its resilience in generating precise predictions. XGBoost, although it is competitive, typically shows greater error metrics, indicating the necessity for more tuning. RNN models, while beneficial, generally exhibit elevated levels of errors, suggesting that they may not be as efficient for this particular task in comparison to LSTM and Gradient Boosting. The subpar performance of ARIMA underscores the constraints of linear statistical models in comprehending the intricacies of drought index dynamics.

This study emphasizes the need to choose suitable models and indicators that align with the unique forecasting needs. LSTM and Gradient Boosting models exhibit exceptional efficacy in long-term SPEI forecasts, rendering them well-suited for use in drought forecasting and climate studies. These findings emphasize the importance of employing comprehensive assessment metrics to measure many facets of model performance, thus guaranteeing strong and dependable predictions. This holds paramount importance in the realm of establishing timely cautionary measures and alleviating the repercussions of drought situations. The discoveries further indicate that the integration of supplementary meteorological and oceanographic indicators, such as the SOI Anomaly and SST NINO4, has the potential to bolster the prognostic capacities of the models, particularly for extended-term forecasts (SPEI-12). This underscores the significance of judiciously selecting input parameters that effectively encapsulate the fundamental climatic factors influencing drought occurrences.

The occurrence of droughts in the Mekong Delta is affected by a range of climatic factors, including fluctuations in precipitation, shifts in temperature, and the influence of large-scale atmospheric patterns like the Southern Oscillation Index (SOI). Prior research has emphasized the intricate characteristics of drought occurrences in the area, demonstrating substantial differences in space and time [63]. Our findings are consistent with these observations since the machine learning models accurately captured the complex dynamics of these climate variables. Moreover, the findings of this study align with prior research

findings about severe droughts in 2010 and 2019 (where $-2 \leq \text{SPEI} < -1.5$), as well as extreme droughts in 1987–1988, 1998, and 2015–2016 (where $\text{SPEI} < -2$) in the Mekong Delta in Vietnam [46,64,65].

The conventional approaches to drought prediction in the area have mainly depended on statistical models, which frequently fail to effectively forecast droughts due to their limited capacity to handle nonlinear associations among climatic variables. On the other hand, machine learning models, like XGBoost, have shown a greater ability to represent these complex interactions, resulting in more precise predictions. These results are consistent with findings from comparable studies conducted in other areas of the world, where the use of advanced modeling approaches has demonstrated enhanced accuracy in predicting droughts [19,66–71]. The table below provides a summary of the comparison between the results of this investigation and previous studies (Table 5).

Table 5. Table comparing the results of the study to the existing literature.

Related Work	Methods	Data Frequency	Prediction Steps	Metrics	Results
Application of Informer Model Based on SPEI for Drought Forecasting [19]	ARIMA, LSTM, and Informer models	monthly	1, 3, 6, 9, 12 and 24-month	MAE, RMSE and NSE	Informer model outperformed ARIMA and LSTM. NSE = 0.968–0.986. Informer model enhanced precision of SPEI prediction on different timescales
Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting [28]	ARIMA, SVR, LSTM, ARIMA-SVR, LS-SVR, and ARIMA-LSTM	monthly	1, 2, 6, 12 and 24-month	NSE	The ARIMA-LSTM model has the highest prediction accuracy at the 6-, 12-, and 24-month scales
Prediction of the standardized precipitation index based on the Long Short-Term Memory and empirical mode decomposition-extreme learning machine models: The Case of Sakarya, Türkiye [29]	LSTM and EMD-ELM hybrid model	monthly	1, 3, and 6-month	NSE, MAE and R^2	LSTM model yielded the best results for SPI-1 and SPI-3 month time scales.
A transparency fusion-based methodology for meteorological drought prediction [67]	XGBoost, RF, LightGBM Ensemble stacking model	monthly	1 and 12-month	R^2	The stacking model outperforms other models with an average R^2 value of 0.845. Extreme precipitation, soil moisture, runoff, and precedent SPEI explain over 80% of the prediction variance.
Drought Monitoring and Performance Evaluation Based on Machine Learning Fusion of Multi-Source Remote Sensing Drought Factors [68]	Remote Sensing, BRF, XGBoost and SVM	monthly	3-month	RMSE and R^2	The bias-corrected random forest (BRF) model outperforms XGBoost and SVM in estimating the Standard Precipitation Evapotranspiration Index (SPEI). The BRF model effectively monitors drought conditions in areas without ground observation data.

Table 5. Cont.

Related Work	Methods	Data Frequency	Prediction Steps	Metrics	Results
Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms [69]	RF, XGBoost, CNN and LSTM	monthly	3 and 6-month	NSE, MSE, MAE, MBE and R ²	The spatial extent of drought increased significantly. Mild drought showed a non-significant increase. The XGB and RF versions are considered to be highly effective.
Quantitative Assessment of Drought Impacts Using XGBoost based on the Drought Impact Reporter [70]	XGBoost, SHAP	monthly	1, 3, 6, 9 and 12-month	F2, Recall and Accuracy	The XGBoost model showed outstanding performance in predicting drought impacts in Texas. The Shapley additive explanation technique revealed the rules guiding the prediction.
Drought Forecasting in Alibori Department in Benin using the Standardized Precipitation Index and Machine Learning Approaches [71]	RF and XGBoost	monthly	1, 3, 6, 9 and 12-month	RMSE, MSE, MAE and R ²	XGBoost showed better performance in drought prediction models. XGBOOST had coefficients of determination 0.89, 0.83, 0.99.
Big data based architecture for drought forecasting using LSTM, ARIMA, and Prophet: Case study of the Jiangsu Province, China [72]	ARIMA, PROPHET, LSTM	monthly	1, 3, 6, 9 and 12-month	RMSE, MAE and R ²	LSTM outperformed other models in drought forecasting
Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network [73]	ARIMA, Holt-Winters and LSTM	monthly	1, 6 and 12-month	RMSE, MAE and R ²	LSTM outperforms ARIMA in long-term drought prediction accuracy.

This table provides a comprehensive comparison between the current study and the existing literature on drought prediction models. The results indicate that the LSTM and XGBoost models are consistently effective across various studies and regions, showcasing their robustness in predicting drought indices such as SPEI. Our study aligns with this trend, demonstrating high accuracy in forecasting drought conditions in the Mekong Delta with R² values ranging from 0.90 to 0.96.

The exceptional efficacy of LSTM and XGBoost models in all timeframes highlights their resilience in capturing the intricate interaction of climatic factors that affect drought conditions. The improved predictive capacity is of great importance for the Mekong Delta region. Accurate forecasts of drought can provide valuable information for appropriate actions, leading to a reduction in agricultural losses and the mitigation of negative social effects. Moreover, these models can be incorporated into current climate monitoring systems to offer uninterrupted, instantaneous predictions of drought.

This study has produced encouraging outcomes; however, it is imperative to acknowledge the existence of several constraints. The use of historical climate data suggests that the models might not fully encompass unprecedented climatic variations stemming from global warming. Future investigations ought to explore the inclusion of real-time data and the integration of supplementary environmental factors such as soil moisture and land-use

changes to enhance the accuracy of the model. Moreover, broadening the geographical scope beyond the Mekong Delta region could validate the applicability of the models to other areas facing similar climate-related challenges. Subsequent research should scrutinize the integration of more advanced feature selection techniques to further enhance the model's performance. Additionally, it would be valuable to investigate hybrid models that exploit the benefits of both machine learning and deep learning methodologies to boost the precision of SPEI forecasts. The implementation of these models in various geographic regions and the assessment of their adaptability to diverse environmental conditions could yield more profound insights into their robustness and importance.

5. Conclusions

The study evaluated ARIMA, Gradient Boosting, XGBoost, RNN, and LSTM models, among other predictive models for Mekong Delta drought forecasting. Using error measures, including MAE, MSE, RMSE, R2 Score, and MAPE, the models were evaluated over several prediction steps and SPEIs. Research evidence demonstrates that LSTM models frequently outperform other models, particularly for longer prediction intervals and SPEI durations. LSTM showed better R2 scores and reduced MAE, MSE, and RMSE values; therefore, this suggests its ability to capture intricate temporal patterns in drought time series data. Particularly for longer-term SPEIs, XGBoost also displayed excellent performance.

The models consistently demonstrated enhanced performance as the SPEI scale progressively raised from 1 month to 12 months, as observed across all models and indicators. This implies that long-term drought conditions are more predictable than short-term fluctuations. Nevertheless, the accuracy of the forecasts decreased for all models as the prediction steps grew from one to six, with ARIMA exhibiting the most significant deterioration. By incorporating climatic indices such as SOI Anomaly and SST NINO4, the model's performance was improved, especially for projections that extend over longer periods of time. This emphasizes the significance of including major climate factors in drought forecasting models for the Mekong Delta region.

These findings are consistent with earlier research on severe drought episodes in the Mekong Delta and are in line with similar studies conducted in other places, which have shown the effectiveness of advanced modeling tools for predicting droughts. The exceptional efficacy of LSTM and XGBoost models in comprehending intricate correlations among meteorological variables establishes a strong basis for making well-informed decisions in the management of water resources, agriculture, and the preservation of ecosystems in regions susceptible to drought.

Although this study presents encouraging findings, its limitations stem from the use of past climate data and the possible inability to accurately represent extraordinary climate fluctuations caused by global warming. Subsequent investigations should examine the integration of up-to-the-minute data, supplementary environmental variables, and more sophisticated methods for selecting features in order to further improve the performance of the model. Furthermore, conducting tests on these models in other geographical regions could yield more profound insights into their resilience and suitability in varied environmental settings.

To summarize, this study showcases the considerable capacity of Artificial Intelligence, specifically LSTM and XGBoost models, to improve drought forecasting capacities for the Mekong Delta. These advanced modeling tools enhance the precision, dependability, and effectiveness of drought predictions, thereby facilitating preemptive measures to alleviate the negative effects of droughts on communities, economies, and ecosystems in the region.

Author Contributions: D.H.H. led the investigation, conceptualization, methodology, statistics, visualization, and writing of the original and final drafts. P.N.D., T.H.L., T.T.D., T.T.N., T.N.M. (Tien Nguyen Minh) and T.N.M. (Tu Nguyen Minh) helped with conceptualization, writing, reviewing, visualization, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data collected for the study can be made available upon request from the corresponding author.

Acknowledgments: The author would like to thank the steering committee of the Project “Research and develop criteria and solutions to implement on-the-spot guidelines to ensure water source security for socio-economic development in the Mekong Delta” (ĐTĐL.CN-45/23) of Ha Luong Thuan provided data on water resources, hydrometeorology in Mekong Delta for this study.

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

References

- Merlo, M.; Giuliani, M.; Du, Y.; Pechlivanidis, I.; Castelletti, A. A Pan-European Analysis of Drought Events and Impacts. In Proceedings of the 25th EGU General Assembly, Vienna, Austria, 23–28 April 2023. Copernicus Meetings.
- McCabe, G.J.; Wolock, D.M.; Lombard, M.; Dudley, R.W.; Hammond, J.C.; Hecht, J.S.; Hodgkins, G.A.; Olson, C.; Sando, R.; Simeone, C.; et al. A hydrologic perspective of major U.S. droughts. *Int. J. Climatol.* **2023**, *43*, 1234–1250. [\[CrossRef\]](#)
- Orimoloye, I.R.; Belle, J.A.; Orimoloye, Y.M.; Olusola, A.O.; Ololade, O.O. Drought: A Common Environmental Disaster. *Atmosphere* **2022**, *13*, 111. [\[CrossRef\]](#)
- Eyvaz, M.; Albahnasawi, A.; Tekbaş, M.; Gürbulak, E.; Eyvaz, M.; Albahnasawi, A.; Tekbaş, M.; Gürbulak, E. *Drought—Impacts and Management*; IntechOpen: London, UK, 2022. [\[CrossRef\]](#)
- Panda, A.; Sahoo, N.; Panigrahi, B.; Das, D.M. Drought Assessment using Standardized Precipitation Index and Normalized Difference Vegetation Index. *Int. J. Curr. Microbiol. App. Sci.* **2020**, *9*, 1125–1136. [\[CrossRef\]](#)
- Seung Kyu, L.; Truong An, D. Evaluating drought events under influence of El-Nino phenomenon: A case study of Mekong delta area, Vietnam. *J. Agrometeorol.* **2018**, *20*, 275–279. [\[CrossRef\]](#)
- Nguyen, N.A. Historic drought and salinity intrusion in the Mekong Delta in 2016: Lessons learned and response solutions. *Vietnam J. Sci. Technol. Eng.* **2017**, *59*, 93–96. [\[CrossRef\]](#)
- Lee, S.K.; Dang, T.A. Spatio-temporal variations in meteorology drought over the Mekong River Delta of Vietnam in the recent decades. *Paddy Water Environ.* **2019**, *17*, 35–44. [\[CrossRef\]](#)
- Adamson, P.; Bird, J. The Mekong: A Drought-prone Tropical Environment? *Int. J. Water Resour. Dev.* **2010**, *26*, 579–594. [\[CrossRef\]](#)
- Ngoc, N.T.; Duong, B.D.; Chien, N.Q.; Darby, S.; Nga, P.T.T.; Thao, B.T.P.; Tai, N.V. *Meteorological Drought Assessment Using Satellite-Based TRMM Product in Vietnamese Mekong Delta*; Presented at the CAREES 2019; Publishing House for Science and Technology: Hanoi, Vietnam, 2019.
- Nguyen, L.; Li, Q. *Relationship between Pacific and Indian Oceans SST and Drought Trends in Vietnam Mekong Delta*; Presented at the Environment and Water Resource Management/813: Modelling and Simulation/814: Power and Energy Systems/815: Health Informatics; ACTA Press: Calgary, AB, Canada, 2014.
- Quang, C.N.X.; Hoa, H.V.; Giang, N.N.H.; Hoa, N.T. Assessment of meteorological drought in the Vietnamese Mekong delta in period 1985–2018. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *652*, 012020. [\[CrossRef\]](#)
- Hasan, N.A.; Dongkai, Y.; Al-Shibli, F. SPI and SPEI Drought Assessment and Prediction Using TBATS and ARIMA Models, Jordan. *Water* **2023**, *15*, 3598. [\[CrossRef\]](#)
- Rezaei, R.; Shabri, A. Drought forecasting using W-ARIMA model with standardized precipitation index. *J. Water Clim. Chang.* **2023**, *14*, 3345–3367. [\[CrossRef\]](#)
- Polpanich, O.-u.; Bhatpuria, D.; Santos Santos, T.F.; Krittasudthacheewa, C. Leveraging Multi-Source Data and Digital Technology to Support the Monitoring of Localized Water Changes in the Mekong Region. Available online: <https://www.mdpi.com/2071-1050/14/3/1739> (accessed on 27 January 2024).
- Zhang, X.; Qu, Y.; Ma, M.; Liu, H.; Su, Z.; Lv, J.; Peng, J.; Leng, G.; He, X.; Di, C. Satellite-Based Operational Real-Time Drought Monitoring in the Transboundary Lancang–Mekong River Basin. *Remote Sens.* **2020**, *12*, 376. [\[CrossRef\]](#)
- Salite, D. Traditional prediction of drought under weather and climate uncertainty: Analyzing the challenges and opportunities for small-scale farmers in Gaza province, southern region of Mozambique. *Nat. Hazards* **2019**, *96*, 1289–1309. [\[CrossRef\]](#)
- Kumari, P.; Rehana, S.; Singh, S.K.; Inayathulla, M. Development of a new agro-meteorological drought index (SPAEL-Agro) in a data-scarce region. *Hydrol. Sci. J.* **2023**, *68*, 1301–1322. [\[CrossRef\]](#)
- Shang, J.; Zhao, B.; Hua, H.; Wei, J.; Qin, G.; Chen, G. Application of Informer Model Based on SPEI for Drought Forecasting. *Atmosphere* **2023**, *14*, 951. [\[CrossRef\]](#)
- Bertini, C.; van Andel, S.J.; Perez, G.C.; Werner, M. AI-enhanced drought forecasting: A case study in the Netherlands. In Proceedings of the 24th EGU General Assembly, Vienna, Austria, 23–27 May 2022. Copernicus Meetings.
- Tran, T.V.; Tran, D.X.; Myint, S.W.; Latorre-Carmona, P.; Ho, D.D.; Tran, P.H.; Dao, H.N. Assessing Spatiotemporal Drought Dynamics and Its Related Environmental Issues in the Mekong River Delta. *Remote Sens.* **2019**, *11*, 2742. [\[CrossRef\]](#)

22. Nguyen, L.B.; Le, M.-H. Application of Artificial Neural Network and Climate Indices to Drought Forecasting in South-Central Vietnam. Available online: <http://www.pjoes.com/Application-of-Artificial-Neural-nNetwork-and-Climate-Indices-to-Drought-nForecasting,105972,0,2.html> (accessed on 27 January 2024).
23. Lee, J.; Kim, C.-G.; Lee, J.E.; Kim, N.W.; Kim, H. Application of Artificial Neural Networks to Rainfall Forecasting in the Geum River Basin, Korea. *Water* **2018**, *10*, 1448. [\[CrossRef\]](#)
24. Jalalkamali, A.; Moradi, M.; Moradi, N. Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. *Int. J. Environ. Sci. Technol.* **2015**, *12*, 1201–1210. [\[CrossRef\]](#)
25. Kikon, A.; Deka, P.C. Artificial intelligence application in drought assessment, monitoring and forecasting: A review. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 1197–1214. [\[CrossRef\]](#)
26. Gyaneshwar, A.; Mishra, A.; Chadha, U.; Raj Vincent, P.M.D.; Rajinikanth, V.; Pattukandan Ganapathy, G.; Srinivasan, K. A Contemporary Review on Deep Learning Models for Drought Prediction. *Sustainability* **2023**, *15*, 6160. [\[CrossRef\]](#)
27. Loukas, A.; Vasiladias, L. A spatiotemporal deep learning forecasting model for long-term drought prediction. In Proceedings of the EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022. Copernicus Meetings. [\[CrossRef\]](#)
28. Xu, D.; Zhang, Q.; Ding, Y.; Zhang, D. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. *Environ. Sci. Pollut. Res.* **2022**, *29*, 4128–4144. [\[CrossRef\]](#)
29. Coşkun, Ö.; Citakoglu, H. Prediction of the standardized precipitation index based on the long short-term memory and empirical mode decomposition-extreme learning machine models: The Case of Sakarya, Türkiye. *Phys. Chem. Earth Parts A/B/C* **2023**, *131*, 103418. [\[CrossRef\]](#)
30. Sandhya Krishna, P.; Yamini Krishna, B.; Nafisa, S.; Ratna Sravani, T.; Ragha Madhuri, J.; Vanditha, C. Prediction of Droughts using SPEI. In Proceedings of the 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 8–9 April 2023; pp. 839–845.
31. Son, B.; Lee, J.; Im, J.; Park, S. Future drought prediction using time-series of drought factors and the US drought monitor data based on deep learning over CONUS. In Proceedings of the 25th EGU General Assembly, Vienna, Austria, 23–28 April 2023. Copernicus Meetings.
32. Sun, Y.; Lao, D.; Ruan, Y.; Huang, C.; Xin, Q. A Deep Learning-Based Approach to Predict Large-Scale Dynamics of Normalized Difference Vegetation Index for the Monitoring of Vegetation Activities and Stresses Using Meteorological Data. *Sustainability* **2023**, *15*, 6632. [\[CrossRef\]](#)
33. Qaisrani, Z.N.; Nuthammachot, N.; Techato, K.; Asadullah; Jatoti, G.H.; Mahmood, B.; Ahmed, R. Drought variability assessment using standardized precipitation index, reconnaissance drought index and precipitation deciles across Balochistan, Pakistan. *Braz. J. Biol.* **2022**, *84*, e261001. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Kartika, F.D.; Wijayanti, P. Drought disaster modeling using drought index: A systematic literature review. *IOP Conf. Ser. Earth Environ. Sci.* **2023**, *1190*, 012026. [\[CrossRef\]](#)
35. Öney, M.; Anli, A. Regional Drought Analysis with Standardized Precipitation Evapotranspiration Index (SPEI): Gediz Basin, Turkey. *J. Agric. Sci.* **2023**, *29*, 1032–1049. [\[CrossRef\]](#)
36. Kobulniczky, B.; Holobacă, I.-H.; Črepinšek, Z.; Pogačar, T.; Jiman, A.-M.; Žnidaršič, Z. Comparison of Standardized Precipitation Index (SPI) and Standardized Potential Evapotranspiration Index (SPEI) applicability for drought assessment during the maize growing period between Bărăgan (Romania) and Prekmurje (Slovenia) regions (1991). In Proceedings of the 25th EGU General Assembly, Vienna, Austria, 23–28 April 2023. Copernicus Meetings.
37. Tam, B.Y.; Cannon, A.J.; Bonsal, B.R. Standardized precipitation evapotranspiration index (SPEI) for Canada: Assessment of probability distributions. *Can. Water Resour. J./Rev. Can. Des Ressour. Hydr.* **2023**, *48*, 283–299. [\[CrossRef\]](#)
38. Santini, M.; Noce, S.; Mancini, M.; Caporaso, L. A Global Multiscale SPEI Dataset under an Ensemble Approach. *Data* **2023**, *8*, 36. [\[CrossRef\]](#)
39. Shi, X.; Yang, Y.; Ding, H.; Chen, F.; Shi, M. Analysis of the Variability Characteristics and Applicability of SPEI in Mainland China from 1985 to 2018. *Atmosphere* **2023**, *14*, 790. [\[CrossRef\]](#)
40. Azman, R.M.N.R.; Noor, N.A.M.; Abdullah, S.; Ideris, M.M. Analysis of Drought Index in Sub-Urban Area Using Standard Precipitation Evapotranspiration Index (SPEI). *Int. J. Integr. Eng.* **2022**, *14*, 157–163. [\[CrossRef\]](#)
41. Careto, J.A.M.; Soares, P.M.M.; Cardoso, R.M.; Russo, A.; Lima, D.C.A. A new ensemble-based SPI and SPEI index to depict droughts projections for the Iberia Peninsula with the EURO-CORDEX. In Proceedings of the EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022. Copernicus Meetings.
42. Bui, N.; Pal, I.; Chollacoop, N. Drought risk assessment under climate change impacts utilizing CMIP6 climate models in the coastal zone of the Mekong Delta. In Proceedings of the 25th EGU General Assembly, Vienna, Austria, 23–28 April 2023. Copernicus Meetings.
43. Huynh, M.; Kumar, P.; Van Toan, N. Deciphering the Relationship between Meteorological and Hydrological Drought in Ben Tre Province, Vietnam. Available online: <https://www.researchsquare.com> (accessed on 26 January 2024).
44. Zhou, K.; Shi, X.; Renaud, F. Understanding precipitation moisture sources of the Vietnamese Mekong Delta and their dominant factors during recent drought events. In Proceedings of the 25th EGU General Assembly, Vienna, Austria, 23–28 April 2023. Copernicus Meetings.

45. Lavane, K.; Kumar, P.; Meraj, G.; Han, T.G.; Ngan, L.H.B.; Lien, B.T.B.; Van Ty, T.; Thanh, N.T.; Downes, N.K.; Nam, N.D.G.; et al. Assessing the Effects of Drought on Rice Yields in the Mekong Delta. *Climate* **2023**, *11*, 13. [CrossRef]
46. Nguyen, D.P.; Ha, H.D.; Trinh, N.T.; Nguyen, M.T. Application of artificial intelligence for forecasting surface quality index of irrigation systems in the Red River Delta, Vietnam. *Environ. Syst. Res.* **2023**, *12*, 24. [CrossRef]
47. Rattayová, V.; Garaj, M.; Kandra, M.; Hlavčová, K. Evaluation of Hargreaves method for calculation of reference evapotranspiration in selected stations of Slovakia. In Proceedings of the 25th EGU General Assembly, Vienna, Austria, 23–28 April 2023. Copernicus Meetings.
48. Al-Asadi, K.; Abbas, A.A.; Dawood, A.S.; Duan, J.G. Calibration and Modification of the Hargreaves–Samani Equation for Estimating Daily Reference Evapotranspiration in Iraq. *J. Hydrol. Eng.* **2023**, *28*, 05023005. [CrossRef]
49. Koç, D.L.; Can, M.E. Reference evapotranspiration estimate with missing climatic data and multiple linear regression models. *PeerJ* **2023**, *11*, e15252. [CrossRef]
50. Elagib, N.A.; Musa, A.A. Correcting Hargreaves–Samani formula using geographical coordinates and rainfall over different timescales. *Hydrol. Process.* **2023**, *37*, e14790. [CrossRef]
51. Hargreaves, G.; Samani, Z. Reference Crop Evapotranspiration From Temperature. *Appl. Eng. Agric.* **1985**, *1*, 96–99. [CrossRef]
52. Althoff, D.; dos Santos, R.A.; Bazame, H.C.; da Cunha, F.F.; Filgueiras, R. Improvement of Hargreaves–Samani Reference Evapotranspiration Estimates with Local Calibration. *Water* **2019**, *11*, 2272. [CrossRef]
53. Muluaem, G.M.; Liou, Y.-A. Application of Artificial Neural Networks in Forecasting a Standardized Precipitation Evapotranspiration Index for the Upper Blue Nile Basin. *Water* **2020**, *12*, 643. [CrossRef]
54. Vicente-Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* **2010**, *23*, 1696–1718. [CrossRef]
55. Hinne, M.; Gronau, Q.F.; Wagenmakers, E.-J. A Conceptual Introduction to Bayesian Model Averaging. Available online: <https://journals.sagepub.com/doi/full/10.1177/2515245919898657> (accessed on 2 February 2024).
56. Tuan, N.V. Regression Models and Scientific Discovery. Sách Khai Minh—Tri Thức Là Sức Mạnh. Available online: <https://www.sachkhaiminh.com/mo-hinh-hoi-quy-va-kham-pha-khoa-hoc-gs-nguyen-van-tuan> (accessed on 2 February 2024).
57. Fan, J.; Ma, X.; Wu, L.; Zhang, F.; Yu, X.; Zeng, W. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric. Water Manag.* **2019**, *225*, 105758. [CrossRef]
58. Fan, J.; Wang, X.; Wu, L.; Zhou, H.; Zhang, F.; Yu, X.; Lu, X.; Xiang, Y. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Convers. Manag.* **2018**, *164*, 102–111. [CrossRef]
59. Schapire, R.E. The Boosting Approach to Machine Learning: An Overview. In *Nonlinear Estimation and Classification*; Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B., Yu, B., Eds.; Lecture Notes in Statistics; Springer: New York, NY, USA, 2003; pp. 149–171.
60. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot* **2013**, *7*, 21. [CrossRef] [PubMed]
61. Ali, Z.; Abduljabbar, Z.; Tahir, H.; Sallow, A.; Almufti, S. Exploring the Power of eXtreme Gradient Boosting Algorithm in Machine Learning: A Review. *J. Nawroz Univ.* **2023**, *12*, 320–334. [CrossRef]
62. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
63. Ahmad, I.; Ahmad, T.; Rehman, S.U.; Mufrah Almanjahie, I.; Alshahrani, F. A detailed study on quantification and modeling of drought characteristics using different copula families. *Heliyon* **2024**, *10*, e25422. [CrossRef] [PubMed]
64. Nguyen, M.N.; Nguyen, P.T.B.; Van, T.P.D.; Phan, V.H.; Nguyen, B.T.; Pham, V.T.; Nguyen, T.H. An understanding of water governance systems in responding to extreme droughts in the Vietnamese Mekong Delta. *Int. J. Water Resour. Dev.* **2021**, *37*, 256–277. Available online: <https://www.tandfonline.com/doi/abs/10.1080/07900627.2020.1753500> (accessed on 11 June 2024). [CrossRef]
65. Minh, H.V.T.; Kumar, P.; Van Ty, T.; Duy, D.V.; Han, T.G.; Lavane, K.; Avtar, R. Understanding Dry and Wet Conditions in the Vietnamese Mekong Delta Using Multiple Drought Indices: A Case Study in Ca Mau Province. *Hydrology* **2022**, *9*, 213. [CrossRef]
66. Musonda, B.; Jing, Y.; Iyakaremye, V.; Ojara, M. Analysis of Long-Term Variations of Drought Characteristics Using Standardized Precipitation Index over Zambia. *Atmosphere* **2020**, *11*, 1268. [CrossRef]
67. Zhang, H.; Sauter, T.; Loaiciga, H. A transparency fusion-based methodology for meteorological drought prediction. In Proceedings of the EGU General Assembly 2023, Vienna, Austria, 24–28 April 2023. Copernicus Meetings. [CrossRef]
68. Zhao, Y.; Zhang, J.; Bai, Y.; Zhang, S.; Yang, S.; Henchiri, M.; Seka, A.M.; Nanzad, L. Drought Monitoring and Performance Evaluation Based on Machine Learning Fusion of Multi-Source Remote Sensing Drought Factors. *Remote Sens.* **2022**, *14*, 6398. [CrossRef]
69. Mokhtar, A.; Jalali, M.; He, H.; Al-Ansari, N.; Elbeltagi, A.; Alsafadi, K.; Abdo, H.G.; Sammen, S.S.; Gyasi-Agyei, Y.; Rodrigo-Comino, J. Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms. *IEEE Access* **2021**, *9*, 65503–65523. [CrossRef]
70. Zhang, B.; Salem, F.K.A.; Hayes, M.J.; Tadesse, T. Quantitative Assessment of Drought Impacts Using XGBoost based on the Drought Impact Reporter. *arXiv* **2022**. [CrossRef]
71. Vodounon, R.B.W.; Soude, H.; Mamadou, O. Drought Forecasting in Alibori Department in Benin using the Standardized Precipitation Index and Machine Learning Approaches. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2022**, *13*, 987–994. [CrossRef]

72. Balti, H.; Abbas, A.B.; Mellouli, N.; Sang, Y.; Farah, I.R.; Lamolle, M.; Zhu, Y. Big data based architecture for drought forecasting using LSTM, ARIMA, and Prophet: Case study of the Jiangsu Province, China. In Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 4–5 July 2021; pp. 1–8.
73. Poornima, S.; Pushpalatha, M. Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft Comput.* **2019**, *23*, 8399–8412. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.