Preprints.org

Article

Application of Machine Learning to Forecast Drought Index for the Mekong Delta

Ha Hai Duong [*] , Nguyen Duc Phong [*] , Thuan Luong Ha , Thang Duc Tang , Thang Ngoc Trinh , Tien Minh Nguyen , Tu Minh Nguyen

Article

Application of Machine Learning 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] Vietnam Academy for Water Resources, Hanoi 100000, Vietnam; trinhthang1606@gmail.com (T.T.N.); nguyenminhtienhn@gmail.com (T.N.M.); tunm1191@gmail.com (T.N.M.)

[2] Vietnam Water Resources Association, Hanoi 100000, Vietnam; hathuancwe@gmail.com

[3] Southern Institute for Water and Environment, Hanoi, Ho Chi Minh 70000, Vietnam; tdthang.siwrr@gmail.com

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

Abstract: Droughts have a substantial effect on water resources, agriculture, and ecosystems on a worldwide scale. In the Mekong Delta of Vietnam, droughts exacerbated by climate change are significantly endangering the region's agricultural sustainability and output. Conventional forecasting techniques frequently do not capture the intricate dynamics of meteorological phenomena associated to drought effectively, prompting the exploration of more advanced methodologies. This work utilises artificial intelligence, particularly machine learning methods like Gradient Boosting and Extreme Gradient Boosting (XGBoost), to enhance drought pre-diction in the Mekong Delta. The study utilises data from 11 meteorological stations spanning from 1990 to 2022 to create and evaluate Machine Learning models based on several climatic factors. We utilise Gradient Boosting and XGBoost algorithms to estimate the Standardised Precipitation-Evapotranspiration Index (SPEI) and evaluate their effectiveness in comparison to con-ventional forecasting techniques. The results show that Machine Learning, particularly XGBoost, surpasses traditional approaches in predicting SPEI accuracy at various time scales. XGBoost demonstrates skill in understanding the complex relationships between climatic factors, with R^2 values falling between 0.90 and 0.94 for 1-month forecasts. The progress highlights the potential of Machine Learning in improving drought management and adaptation tactics, proposing the incorporation of Machine Learning forecasting models into decision-making processes to enhance drought resistance in susceptible areas.

Keywords: drought forecasting; machine learning algorithms; Mekong Delta; water resource management; Vietnam

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 conterminous 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 international impact.

Drought forecasting is important for several reasons, as follows. First, it allows early action to be taken to mitigate the impacts of drought events. This can include measures, such as early livestock destocking and water management strategies [6]. Second, accurate drought prediction can help in the

management of groundwater sources, agriculture, and ecosystems, thereby reducing the social and economic harm caused by drought [7]. Additionally, forecasting future drought conditions is crucial for preventing agricultural and hydrological resource damage in models that can be used to predict drought severity classes [8]. Furthermore, impact-based drought forecasting can provide critical information for disaster preparedness and adaptation, and increase community resilience [9]. Lastly, forecasting droughts on a seasonal timescale can provide useful insights into the increase in the frequency and intensity of extreme events and their location [10].

Accurate prediction of drought is essential for reducing the negative effects of drought in the Mekong Delta, Vietnam, impacting agriculture, water management, and community resilience. It facilitates strategic agricultural planning, optimises water resource allocation, and improves early warning systems to prepare for difficulties associated to 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 applying artificial intelligence (AI) to forecast the drought index in the Mekong Delta, aiming to address and achieve key objectives. Primarily, it seeks to surpass the constraints of conventional drought forecasting techniques, which frequently fail to accurately represent the intricate and changing climatic conditions unique to the Mekong Delta. Through the utilization of AI, the study aims to significantly improve the precision and dependability of drought forecasts. Furthermore, the enhanced forecasting capability provided by AI is intended to facilitate the issuance of timely warnings and the initiation of preemptive actions. This proactive approach empowers communities, governmental bodies, and other pertinent stakeholders to adequately prepare for and mitigate the repercussions of imminent water scarcities and related challenges.

Current methods for forecasting drought involve a combination of statistical, probabilistic, and data-driven approaches. The historical perspective of drought in the Mekong Delta reveals a complex interplay of climatic, hydrological, and anthropogenic factors that have shaped the region's vulnerability to water scarcity. Understanding the historical context provides insights into the recurring challenges faced by the Mekong Delta and sets the stage for innovative approaches, such as the application of artificial intelligence, to address contemporary drought issues. A study by [11] 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 [12]. This region has also experienced a shift in the spatial distribution of meteorological droughts, with a decrease in frequency and an increase in severity [13]. The impact of these droughts on agriculture, particularly on rice production, is significant [14]. The construction of mega-dams in the Mekong River has further exacerbated this situation, leading to reduced water levels and increased dry season droughts [15]. Nguyen Thi Ngoc et al. evaluated meteorological droughts using the standardized precipitation index (SPI) based on data from the Tropical Rainfall Measuring Mission (TRMM) [16]. Tran et al. used the Normalized Difference Water Index (NDWI) derived from Landsat satellite images to analyze drought severity and spatiotemporal dynamics. Tran et al. (EDSI) by integrating remote sensing data and spatiotemporal regression methods to assess the severity of agricultural drought severity [17]. Pal and Juddoo conducted a comprehensive drought risk assessment that considered climate change

impacts in the coastal provinces of the Mekong Delta [18]. Nguyen and Li analyzed the correlation between sea surface temperature anomalies (SSTA) and meteorological droughts in the Vietnam Mekong Delta [19]. These studies demonstrate the use of various methods and data sources for drought forecasting in the Mekong Delta. Quang et al. 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. El Niño was found to strongly influence extreme drought events in VMD, and adaptation measures are crucial for coping with drought disasters, particularly in the agricultural and aquaculture sectors [20].

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 [16]. 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 [21]. 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 [22]. Furthermore, the reduced number of elderly people in the community has led to a decline in the diversity and complexity of the interpretation of these indicators [23]. 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 [24].

Artificial Intelligence (AI) techniques, particularly Machine Learning (ML), 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 [25]. 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 [17]. Additionally, spatiotemporal regression methods and time-series biophysical data derived from remote sensing were integrated to develop a new drought index called the enhanced drought severity index (EDSI). These approaches have demonstrated the potential of AI and ML in drought forecasting and risk assessment in the Mekong Delta region.

Artificial intelligence, particularly in the form of artificial neural networks, has shown promise for drought forecasting. Luong Bang Nguyen and J. Lee demonstrated the effectiveness of this technology for predicting drought indices and rainfall, respectively [26,27]. 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 [28]. 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 [29]. These studies collectively underscore the potential of artificial intelligence in improving drought forecasting in the Mekong Delta.

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 [30–33]. The SPEI combines meteorological and hydrological variables, such as precipitation, evapotranspiration, and groundwater levels, to provide a comprehensive assessment of drought [34]. 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 [35], whereas a new multiscale SPEI dataset has been

provided for reference and future time horizons in Italy [36]. 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 [32]. In China, the SPEI has been used to accurately monitor drought events, with spatiotemporal distribution and trends analyzed in various climatic sub-regions [37]. 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 [38].

The Standardized Precipitation Evapotranspiration Index (SPEI) has several advantages. It is useful for assessing both drought and wetter-than-normal conditions, and provides a comprehensive understanding of moisture variability [36]. SPEI is a reliable tool for drought prediction because it is simpler, faster, and requires fewer data points than dynamic models [39]. 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) [40]. 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. The Standardized Precipitation Evapotranspiration Index (SPEI) is a reliable drought index that can be used for accurate drought assessment and forecasting.

The role of artificial intelligence in drought forecasting is transformative and offers innovative solutions to overcome the limitations of traditional approaches. By harnessing the power of AI, drought forecasting becomes more accurate, adaptive, and responsive, ultimately supporting effective water resource management and enhancing resilience in drought-vulnerable regions.

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 Mekong Delta is known for its cultural diversity, with various ethnic groups and religions coexisting 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).

The main seasons in the Mekong Delta are dry and rainy. The dry season typically occurs from December to May, whereas the rainy season occurs from June to November [42]. The dry season is critical for rice production, as more than 80% of rice is planted during this time when water flow is at its lowest [43]. However, the wet season supplies the majority of the water needed for rice production and other agricultural activities in the delta [44]. The seasons in the Mekong Delta are affected by climate change, which leads to alterations in rainfall patterns and water availability [45]. These changes in weather patterns have implications for the crop yield and water quality in the region.

Drought is a prevalent concern in the Mekong Delta because of climate change impacts [18]. The intensity, duration, and frequency of meteorological droughts in the delta have been studied using various indices[46]. The Vietnamese Mekong Delta (VMD) has experienced droughts that affect agriculture and aquaculture [47]. The Standardized Precipitation Index (SPI) has been used to identify drought events and their impact on rice yields in the VMD [48]. VMD, particularly Ca Mau Province, experiences both dry and wet conditions, with prolonged drought events extending into the wet season [49]. 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.
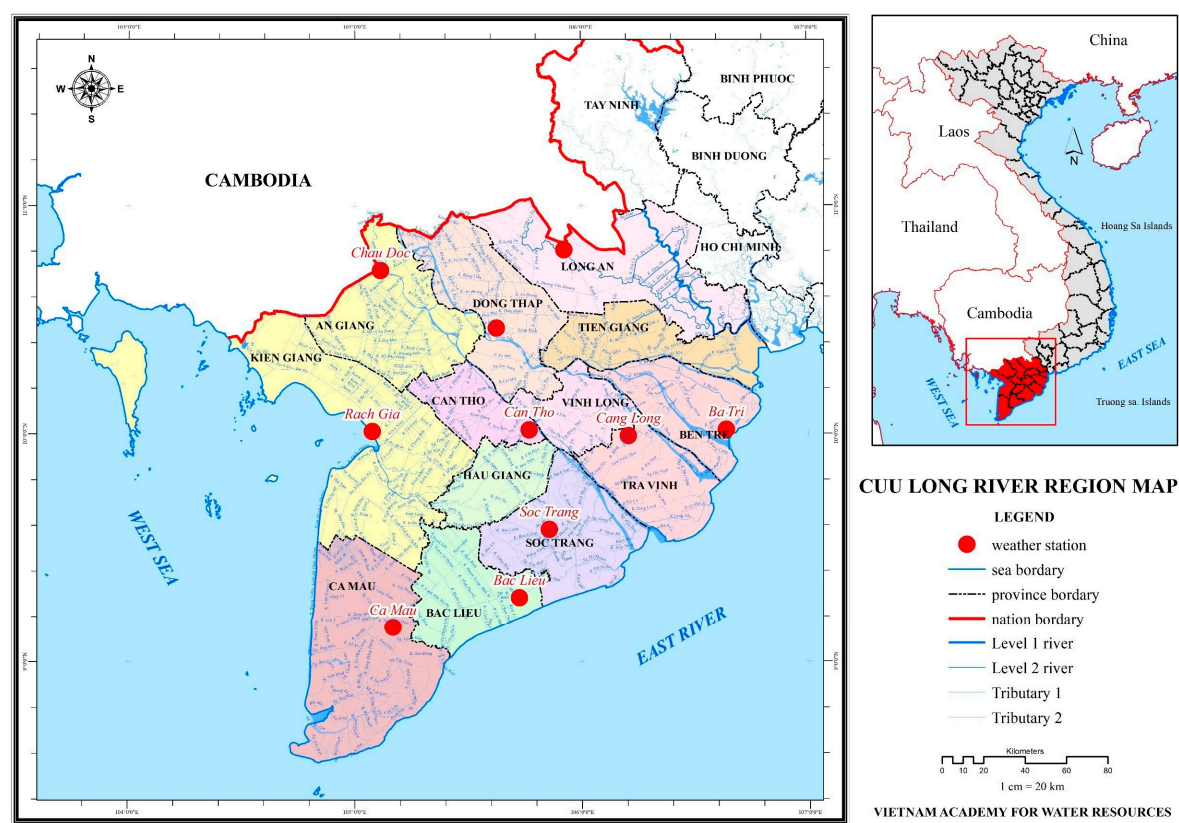
Figure 1. The geographical location and metrological stations in the Mekong Delta.

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 (1990 – 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 (^{0}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 1,350 - 2,366 mm. The Ca Mau - Rach Gia region experiences the highest levels of rainfall, with measurements ranging from to 2000-2,366 mm or more. Approximately 30% of the weather stations record a rainfall of 1,300 mm or more. In contrast, areas such as My Tho receive a lower amount of precipitation, with levels ranging from approximately 1,300 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 (SST), 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/. SOI and SST data were obtained from the National Oceanic and Atmospheric Administration (NOAA) website at https://www.ncei.noaa.gov/access/monitoring/enso. The data sources were constantly maintained, and no post-processing was performed.
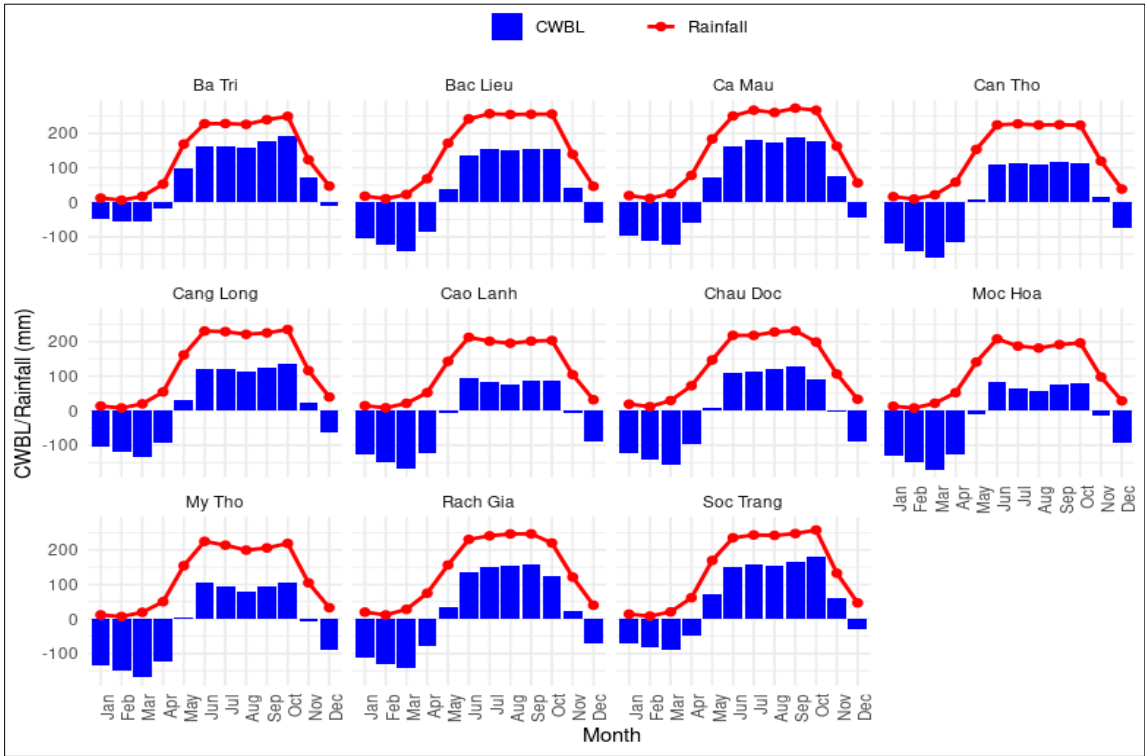

Figure 2. The average monthly precipitation and CWBL for eleven meteorological stations in the Mekong Delta from 1985 to 2022.

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 utilised in modelling [50]. This entails:
1. Data cleaning:
   - Handling missing values: Methods include replacing missing values with the mean, using regression to estimate missing values, or removing incomplete cases.

- Correcting Inconsistencies and Outliers: This involves identifying and resolving errors or outliers using data profiling, statistical methods, or domain-specific knowledge.

2. Normalize data:

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

$$X_{norm} = \frac{X_0 - X_{min}}{X_{max} - X_{min}} \tag{1}$$

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: The data will be partitioned into training and testing sets, with 80% of the data used for training and 20% used for evaluating the model.

2.3.2. Method for Calculating Drought Index

In this study, the Hargreaves-Samani (HS) method was used to estimate reference evapotranspiration (ETo) because this method has been used to estimate reference evapotranspiration (ETo) in various regions. Studies have shown that the HS equation can accurately estimate ETo values when compared with the FAO Penman-Monteith (PM) method, which is considered the most accurate method for ETo estimation [51,52]. The HS equation performs well under different climatic conditions and environments, including regions with high altitudes [53]. In some cases, the HS equation outperforms the FAO56-PM method, particularly when meteorological data are limited or unavailable [54]. In general, the HS method has been demonstrated to be a viable option for calculating ETo, especially in areas where the availability of data is limited, such as the Mekong Delta.

The Hargreaves-Samani (HS) approach calculates the reference evapotranspiration (ETo) based solely on the highest and minimum temperatures [55], as shown in the following equation:

$$ET_0 = C_0 Ra'(T_{max} - T_{min})^{0.5}(T + 17.8) \tag{2}$$

Where:

- R_a': extraterrestrial solar radiation, in mm day^{-1}
- C_0: conversion parameter (=0.0023)
- $Tmax$, $Tmin$ and T: the maximum, minimum and average temperature (°C),

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

$$D_i = P_i - E_{0i} \tag{3}$$

Di 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 [56].

The SPEI package for R [57], 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 standardised precipitation evapotransporation index (SPEI).

SPEI	Drought Category
SPEI ≥ 2	Extremely wet
1.5 ≤ SPEI < 1	Severely wet
1 ≤ SPEI < 1.5	Moderately wet
-1 ≤ SPEI < 1	Near normal
-1.5 ≤ SPEI < -1	Moderately dry
-2 ≤ SPEI < -1.5	Severely dry

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 generalise to new data [58,59].

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 utilises probabilistic frameworks to calculate the average of many models, taking into account the uncertainty related to the parameters of each model [50].

If we have two models M_1 and M_2 and assume that one of them is true, the posterior probability of M1 is:

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

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

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

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 method, 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). The 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. 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, to make precise predictions of the SPEI index. According to the results of the literature review, the machine learning methods widely employed for SPEI prediction include Gradient Boosting, Extreme Gradient Boosting (XGBoost).

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 index [60]. 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 index [61]. 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 index.

Furthermore, these gradient boosting algorithms have shown promising results in forecasting different seasons and multimonth-ahead reference evapotranspiration. In summary, gradient boosting algorithms, specifically XGBoost and Gradient Boosting, were highly effective in predicting the SPEI index.

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 [62]. Gradient boosting leverages the advantages of several models to construct a resilient predictive model that outperforms any individual poor learner [50].

Given a training dataset D = {xi, yi}^{N_1}, the goal of gradient boosting is to find an approximation, $\hat{F}_{(x)}$, of the function F*(x), which maps instances x to their output values y by minimizing the expected value of a given loss function, L(y, F(x)). Gradient boosting builds an additive approximation of F*(x) as a weighted sum of functions (6):

$$F_m(x) = F_{m-1}(x) + \rho_m h_m(x)$$ (6)

where ρ_m is the weight of the mth function, h$_m$(x). These functions are the models of the ensemble (e.g., decision trees). The approximation is constructed iteratively.

Below is a thorough analysis of the functioning of gradient boosting:

- Model Initialization: The procedure commences by constructing an initial model utilising 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 utilised to revise the residuals.
- Weight Adjustment and Reweighting: The data points' weights are modified to prioritise 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 is accurately predicted or a predetermined maximum number of models is attained. Every iteration has the objective of minimising the total prediction error by dealing with the leftover residuals.

2. Extreme Gradient Boosting (XGBoost):

XGBoost is a machine learning technique that use a gradient boosting framework to improve the accuracy of predictions. The algorithm lends greater importance to incorrectly classified data pieces, prioritising their accurate prediction in subsequent iterations [63]. XGBoost enhances existing gradient boosting algorithms by integrating regularisation approaches to reduce overfitting and employing advanced optimisation techniques to improve computational performance [50].

The purpose of the model is simplification through 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 (7).

$$J^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$ (7)

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 prioritise the analysis of data points that were previously misclassified or inaccurately anticipated.

- Optimisation 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 optimised by utilising the residuals obtained from each stage.
- Iteration and Combination: This iterative process is done several times. The ultimate model is an amalgamation of all the separate trees, with each tree making a contribution to the overall prediction.

2.3.5. Model Evaluation Method

The evaluation of the accuracy of models forecasting the Standardised Precipitation Evapotranspiration Index (SPEI) in the Mekong Delta is based on the following indicators: Mean absolute error (MAE); The mean square error (MSE); The root mean square errors (RMSE); and The coefficient of determination (R^2) [50]. The criteria for evaluating (calibrating) the models are presented in Equations (8)–(11):

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|P_i - M_i| \tag{8}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2 \tag{9}$$

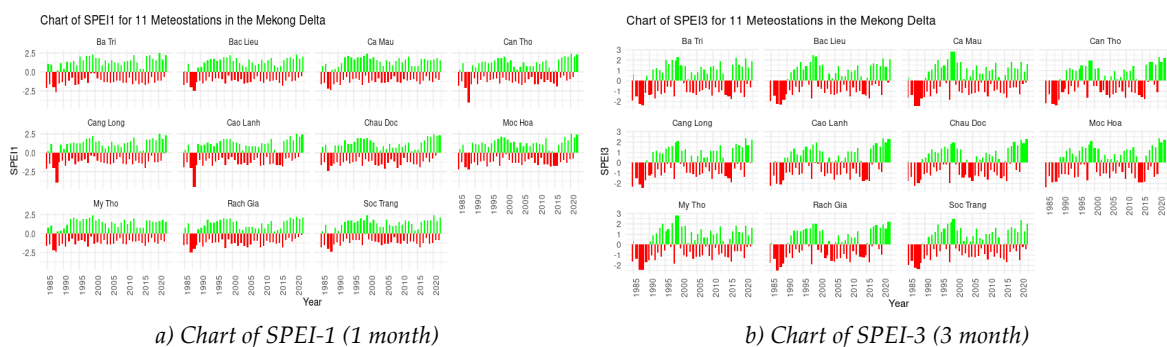$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - M_i)^2} \tag{10}$$

$$R^2 = 1 - \left(\frac{ESS}{TSS}\right) \tag{11}$$

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 index 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 index is high.
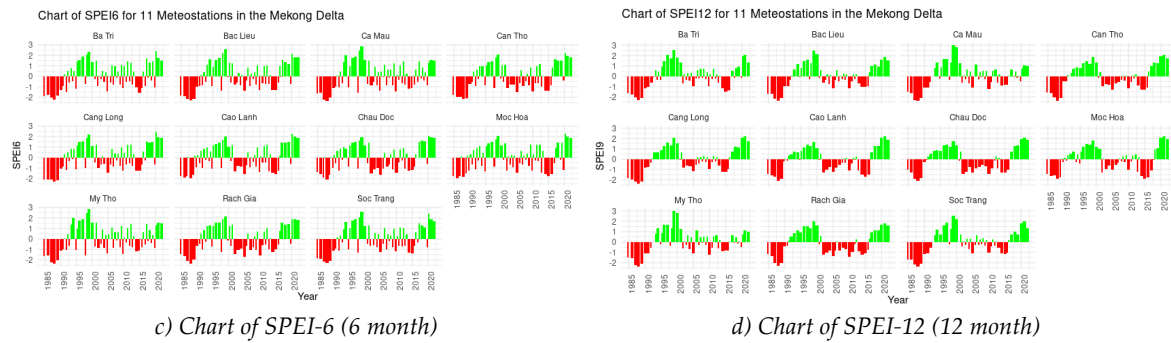

a) Chart of SPEI-1 (1 month)   *b) Chart of SPEI-3 (3 month)*

c) Chart of SPEI-6 (6 month)   *d) Chart of SPEI-12 (12 month)*

Figure 3. Chart of SPEI of 11 meteorological stations in the Mekong Delta (1990 - 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 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 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).
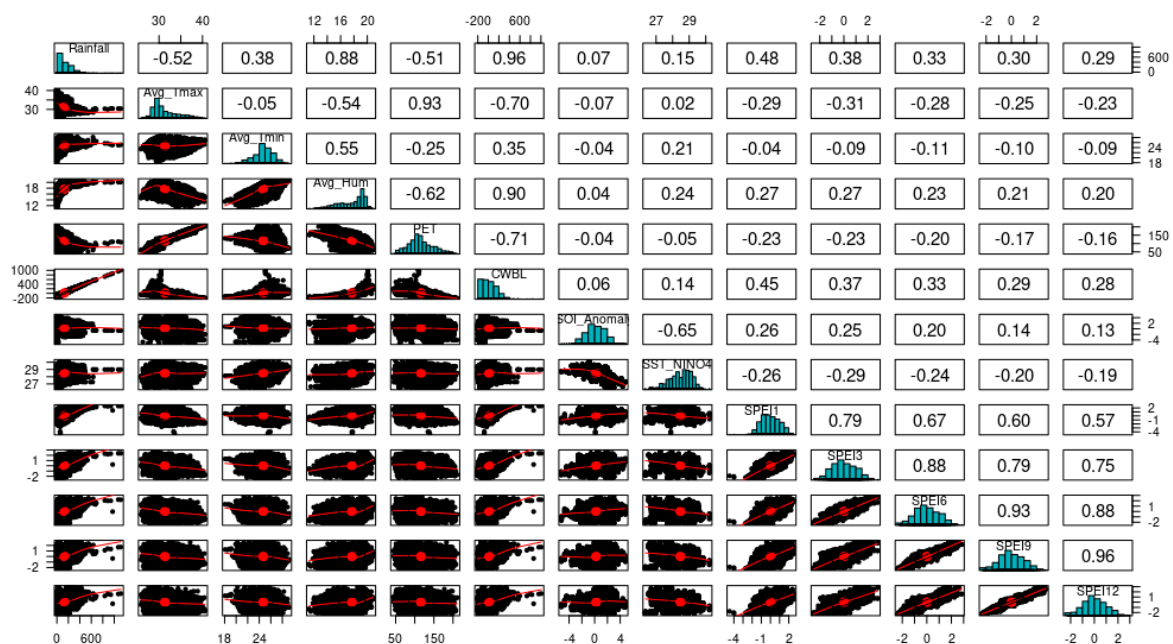

Figure 4. Correlation chart of drought index and meteorological parameters.

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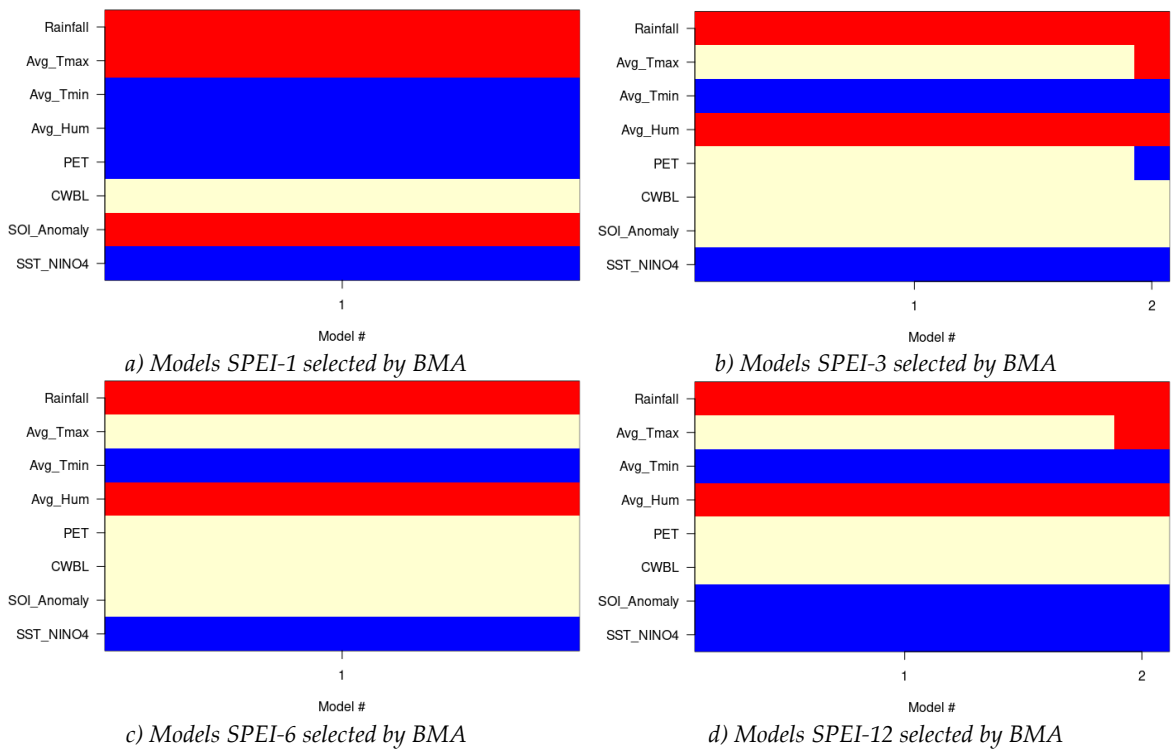

a) Models SPEI-1 selected by BMA

b) Models SPEI-3 selected by BMA

c) Models SPEI-6 selected by BMA

d) Models SPEI-12 selected by BMA

Figure 5. Graph of the selection of important parameters by BMA.

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 method to predict only the SPEI. The models were developed based on various time periods and included Gradient Boosting (GB) and Extreme Gradient Boosting (XGBoost). Table 3 displays the outcomes of the hyperparameter-tweaking process.

Table 3. Table of results of hyperparameter tuning.

No.	Model name	Hyperparameter tuning
1	Gradient Boosting (GB)	- Distribution = "Gaussian". - cv.folds = 10: - shrinkage parameter = 0.01. - Each terminal node should have at least 10 observations: n.minobsinnode = 10.
2	eXtreme Gradient Boosting (XGBoost)	- n.trees = 1000. - 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.

The SPEI is calculated at different timescales (1, 3, 6, and 12 months) to capture short-term and long-term drought conditions. The forecast results for SPEI-1, SPEI-3, SPEI-6, and SPEI-12 are presented in Figures 6–9 along with a comparison chart illustrating the discrepancy between the predicted and computed SPEI values for the test data.
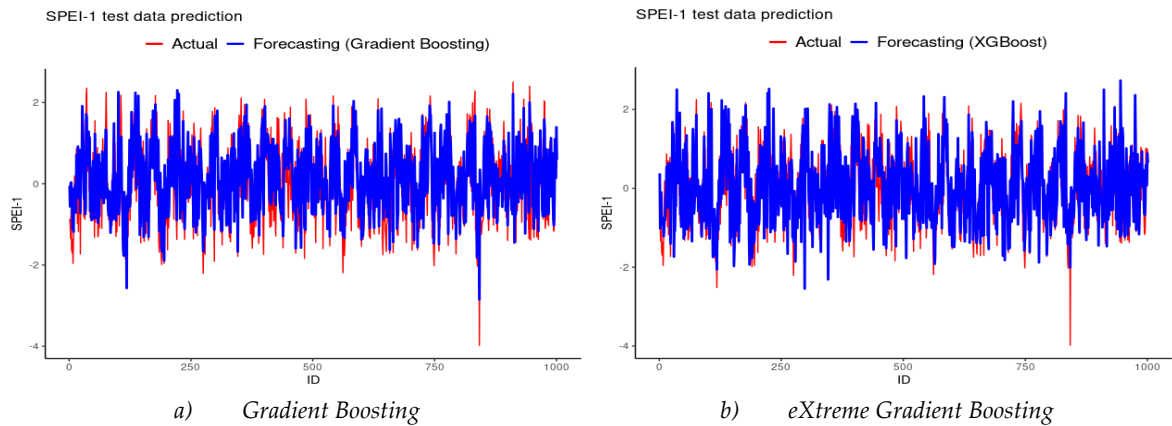
a)   *Gradient Boosting*   *b)*   *eXtreme Gradient Boosting*

Figure 6. Comparison chart between forecast and actual SPEI-1 for the test data.

a)   *Gradient Boosting*   *b)*   *eXtreme Gradient Boosting*
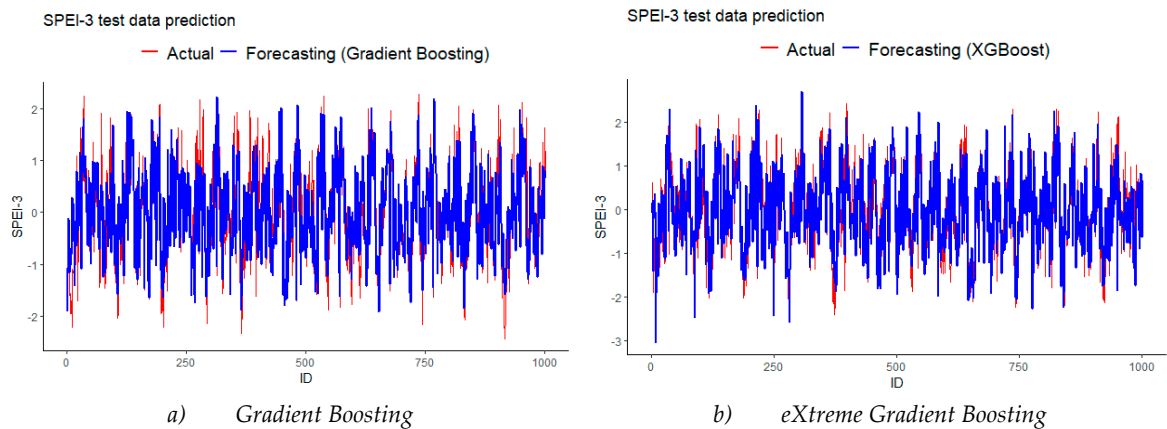
Figure 7. Comparison chart between forecast and actual SPEI-3 for the test data.

a)   *Gradient Boosting*   *b)*   *eXtreme Gradient Boosting*
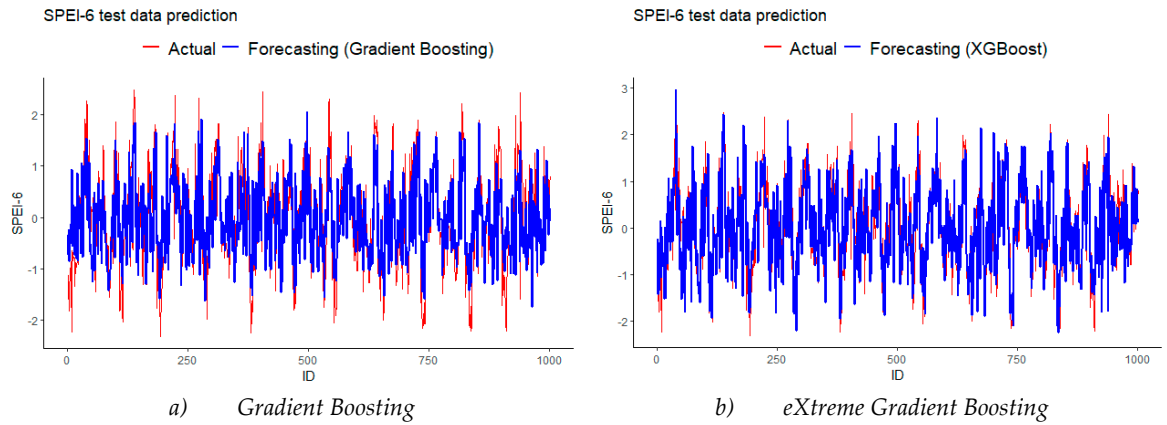
Figure 8. Comparison chart between forecast and actual SPEI-6 for the test data.

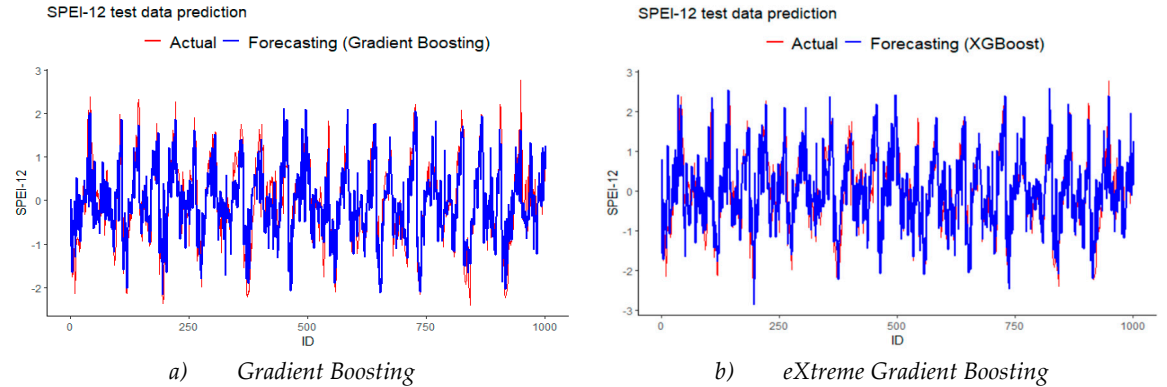*a)*   *Gradient Boosting*   *b)*   *eXtreme Gradient Boosting*

Figure 9. Comparison chart between forecast and actual SPEI-12 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), and coefficient of determination (R2). 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 in the study area.

Table 4. Statistical table of evaluation results of models to predict the SPEI.

Models	Input parameters	Output	Evaluation criteria			
			MAE	MSE	RMSE	R^2
Gradient Boosting	Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, SST_NINO4	SPEI-1	0.34	0.19	0.44	0.90
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-3	0.38	0.26	0.51	0.83
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-6	0.36	0.24	0.49	0.86
	Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4	SPEI-12	0.36	0.23	0.48	0.87
XGBoost	Rainfall, Avg_Tmax, Avg_Tmin, Avg_Hum, PET, SOI_Anomaly, SST_NINO4	SPEI-1	0.28	0.16	0.37	0.94
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-3	0.33	0.21	0.45	0.89
	Rainfall, Avg_Tmin, Avg_Hum, SST_NINO4	SPEI-6	0.32	0.19	0.41	0.90
	Rainfall, Avg_Tmin, Avg_Hum, SOI_Anomaly, SST_NINO4	SPEI-12	0.30	0.17	0.41	0.92

This table outlines the performance of the Gradient Boosting and XGBoost models trained to predict the Standardized Precipitation-Evapotranspiration Index (SPEI) over different time scales: 1, 3, 6, and 12 months. 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).

The SPEI-1 model utilizes a comprehensive range of inputs to forecast SPEI on a monthly scale. The model's performance metrics included an MAE ranging from 0.28 to 0.34, MSE from 0.16 to 0.19, RMSE from 0.37 to 0.44, and an R² from 0.90 0.94. These values demonstrate high accuracy and indicate that XGBoost captures the intricate dynamics that influence monthly drought conditions.

The SPEI-3 model streamlined its inputs for the 3-month prediction by omitting the Avg_Tmax, PET, and SOI anomalies. The model demonstrates MAE ranging from 0.33 to 0.38, MSE from 0.21 to 0.26, RMSE from 0.45 to 0.51, and R² value between 0.83 and 0.89. It provides robust predictive capabilities, albeit significantly less precise than those of the 1-month model.

The SPEI-6 model, using identical inputs as the SPEI-3 model, produced a 6-month forecast with an MAE ranging from 0.32 to 0.36, MSE from 0.19 to 0.23, RMSE from 0.41 to 0.49, and an R² between 0.87 and 0.90. This suggests that the model is highly efficient in predicting mid-term droughts.

The SPEI-12 model incorporates the SOI Anomaly, in addition to the variables utilized in the SPEI-3 and SPEI-6 models for the 12-month forecast. The model achieves MAE ranging from 0.30 to 0.36, MSE ranging from 0.17 to 0.23, RMSE ranging from 0.41 to 0.48, and R² values ranging from 0.87 to 0.92. This indicates that including the SOI Anomaly leads to a substantial enhancement in the precision of the model for long-term drought forecasts.

4. Discussion

Artificial intelligence (AI) techniques, particularly machine learning models such as Gradient Boosting and Extreme Gradient Boosting (XGBoost), have demonstrated considerable promise in improving the accuracy of drought prediction in the Mekong Delta. These models overcome the constraints of conventional forecasting approaches by accurately reflecting the intricate dynamics of

meteorological variables that impact drought conditions. The study demonstrates that the XGBoost model outperforms other models in predicting droughts at different time intervals. It emphasises the model's capability to handle complex interactions between input variables, which is crucial for accurately predicting the diverse character of droughts.

AI-based methods provide in-depth analysis of drought vulnerability and advancement, assisting in the management of water resources, planning for agriculture, and conservation of ecosystems. Precise and timely predictions allow stakeholders to proactively take actions to reduce the negative effects of drought on vulnerable populations and their means of living. Nevertheless, there are still obstacles to overcome, including as dealing with imbalanced datasets, integrating various data sources, and improving the process of selecting models and modifying hyperparameters. Furthermore, the fluctuation in space and time of droughts requires ongoing enhancement and verification of models.

In order to tackle these difficulties, it is essential to improve the methods of gathering and exchanging data, encourage collaboration between different fields of study, and make use of the progress made in processing power and algorithms. Future study should investigate the integration of AI models with satellite and remote sensing technologies to enhance the comprehension of drought indicators and advance real-time monitoring capabilities. Creating hybrid models that combine machine learning with conventional forecasting methods or other artificial intelligence approaches could provide a strong and flexible framework for predicting droughts that is customised to the specific requirements of the Mekong Delta and other places susceptible to drought.

The utilisation of AI for drought prediction is a crucial measure in comprehending and alleviating the consequences of this devastating natural calamity. Despite ongoing obstacles, the capacity of AI to fundamentally transform drought management tactics is unquestionable. As these models are improved and extended, the goal of obtaining better resistance to drought in the Mekong Delta and other areas becomes more and more achievable.

5. Conclusions

The amalgamation of artificial intelligence (AI) and machine learning (ML) has brought about a significant transformation in drought prediction, especially in susceptible areas such as the Mekong Delta. This study demonstrates the effectiveness of Gradient Boosting and Extreme Gradient Boosting (XGBoost) models in predicting the Standardised Precipitation-Evapotranspiration Index (SPEI) with R^2 values ranging from 0.83 to 0.94. These models surpass traditional techniques by precisely reflecting the intricate interconnections of climate elements.

AI models utilise comprehensive climate data and advanced analytical tools to improve the precision, dependability, and effectiveness of drought predictions. The XGBoost model excels in its ability to effectively handle the complex interconnections among meteorological variables. This progress establishes a strong basis for making well-informed decisions in the fields of water resource management, agriculture, and ecosystem conservation. Artificial intelligence-powered drought forecasts enable preventive and adaptable measures to reduce the negative effects on societies, economies, and ecosystems worldwide.

The study emphasises the necessity for ongoing investment in AI research and the creation of integrated models that merge satellite and remote sensing data. Future endeavours should prioritise improving the comprehensibility of AI models to ensure that their findings are easily understood by a wide range of individuals involved. Collaboration among scientists, policymakers, and practitioners is essential for ensuring that research goals are in line with practical uses and for implementing AI solutions on a large scale.

Ultimately, the integration of AI in drought prediction represents a notable advancement in promoting environmental resilience and sustainability. As these technologies advance, it becomes more possible to tackle the issues caused by drought. This helps to safeguard our world and its residents from the escalating dangers of climate change.

Author Contributions: All authors contributed to the writing and editing of the manuscript. Duong Hai Ha led the investigation, conceptualization, methodology, statistics, visualization, and writing of the original and final

References

1. M. Merlo, M. Giuliani, Y. Du, I. Pechlivanidis, and A. Castelletti, "A pan-European analysis of drought events and impacts," Copernicus Meetings, EGU23-12961, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-12961.
2. G. J. McCabe *et al.*, "A hydrologic perspective of major U.S. droughts," *International Journal of Climatology*, vol. 43, no. 3, pp. 1234–1250, 2023. https://doi.org/10.1002/joc.7904.
3. I. R. Orimoloye, J. A. Belle, Y. M. Orimoloye, A. O. Olusola, and O. O. Ololade, "Drought: A Common Environmental Disaster," *Atmosphere*, vol. 13, no. 1, Art. no. 1, Jan. 2022. https://doi.org/10.3390/atmos13010111.
4. M. Eyvaz *et al.*, *Drought - Impacts and Management*. 2022. https://doi.org/10.5772/intechopen.97906.
5. A. P, a, N. Sahoo, B. Panigrahi, and D. M. Das, "Drought Assessment using Standardized Precipitation Index and Normalized Difference Vegetation Index," *Int.J.Curr.Microbiol.App.Sci.*, vol. 9, no. 7, pp. 1125–1136, Jul. 2020. https://doi.org/10.20546/ijcmas.2020.907.132.
6. T. S. Busker, H. de Moel, B. van den Hurk, and J. C. J. H. Aerts, "Impact-Based Seasonal Rainfall Forecasting to Trigger Early Action for Droughts." Rochester, NY, Oct. 14, 2022. https://doi.org/10.2139/ssrn.4247666.
7. Z. M. Yaseen and S. Shahid, "Drought Index Prediction Using Data Intelligent Analytic Models: A Review," in *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation: Theory and Practice of Hazard Mitigation*, R. C. Deo, P. Samui, O. Kisi, and Z. M. Yaseen, Eds., in Springer Transactions in Civil and Environmental Engineering. , Singapore: Springer, 2021, pp. 1–27. https://doi.org/10.1007/978-981-15-5772-9_1.
8. B. Son, J. Lee, J. Im, and S. Park, "Future drought prediction using time-series of drought factors and the US drought monitor data based on deep learning over CONUS," Copernicus Meetings, EGU23-11208, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-11208.
9. N. K. A. Rasheed *et al.*, "Automatic generation of impact-based drought forecast, implications for early warning and action in East Africa," Copernicus Meetings, EGU23-10940, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-10940.
10. T. D. Monte, A. Cherchi, A. Alessandri, and M. Gaetani, "Assessing the predictability of droughts through seasonal forecasts," Copernicus Meetings, EGU23-14731, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-14731.
11. L. Seung Kyu and D. Truong An, "Evaluating drought events under influence of El-Nino phenomenon: A case study of Mekong delta area, Vietnam," *Journal of Agrometeorology*, vol. 20, no. 4, Art. no. 4, Dec. 2018. https://doi.org/10.54386/jam.v20i4.565.
12. N. A. Nguyen, "Historic drought and salinity intrusion in the Mekong Delta in 2016: Lessons learned and response solutions," *Vietnam Journal of Science, Technology and Engineering*, vol. 59, no. 1, Art. no. 1, Mar. 2017. https://doi.org/10.31276/VJSTE.59(1).93.
13. S. K. Lee and T. A. Dang, "Spatio-temporal variations in meteorology drought over the Mekong River Delta of Vietnam in the recent decades," *Paddy Water Environ*, vol. 17, no. 1, pp. 35–44, Jan. 2019. https://doi.org/10.1007/s10333-018-0681-8.
14. P. Adamson and J. Bird, "The Mekong: A Drought-prone Tropical Environment?," *International Journal of Water Resources Development*, vol. 26, no. 4, pp. 579–594, Dec. 2010. https://doi.org/10.1080/07900627.2010.519632.
15. X. Lu and S. Chua, "River Discharge and Water Level Changes in the Mekong River: Droughts in an Era of Mega-Dams," Preprints, preprint, Jan. 2021. https://doi.org/10.22541/au.161189206.60684509/v1.
16. N. T. Ngoc *et al.*, "Meteorological drought assessment using satellite-based TRMM product in Vietnamese Mekong Delta," presented at the CAREES 2019, Hanoi, Vietnam: Publishing House for Science and Technology, Nov. 2019. https://doi.org/10.15625/vap.2019.000158.

17. T. V. Tran *et al.*, "Assessing Spatiotemporal Drought Dynamics and Its Related Environmental Issues in the Mekong River Delta," *Remote Sensing*, vol. 11, no. 23, Art. no. 23, Jan. 2019. https://doi.org/10.3390/rs11232742.

18. N. Bui, I. Pal, and N. Chollacoop, "Drought risk assessment under climate change impacts utilizing CMIP6 climate models in the coastal zone of the Mekong Delta," Copernicus Meetings, EGU23-328, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-328.

19. L. Nguyen and Q. Li, "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, Sep. 2014. https://doi.org/10.2316/P.2014.812-017.

20. C. N. X. Quang, H. V. Hoa, N. N. H. Giang, and N. T. Hoa, "Assessment of meteorological drought in the Vietnamese Mekong delta in period 1985-2018," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 652, no. 1, p. 012020, Feb. 2021. https://doi.org/10.1088/1755-1315/652/1/012020.

21. Orn-uma Polpanich, Dhyey Bhatpuria, Tania Fernanda Santos Santos, and Chayanis Krittasudthacheewa, "Leveraging Multi-Source Data and Digital Technology to Support the Monitoring of Localized Water Changes in the Mekong Region." Accessed: Jan. 27, 2024. [Online]. Available: https://www.mdpi.com/2071-1050/14/3/1739

22. X. Zhang *et al.*, "Satellite-Based Operational Real-Time Drought Monitoring in the Transboundary Lancang–Mekong River Basin," *Remote Sensing*, vol. 12, no. 3, Art. no. 3, Jan. 2020. https://doi.org/10.3390/rs12030376.

23. A. Shabri, "A hybrid wavelet analysis and adaptive neuro-fuzzy inference system for drought forecasting," *ams*, vol. 8, pp. 6909–6918, 2014. https://doi.org/10.12988/ams.2014.48263.

24. D. Salite, "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*, vol. 96, no. 3, pp. 1289–1309, Apr. 2019. https://doi.org/10.1007/s11069-019-03613-4.

25. C. Bertini, S. J. van Andel, G. C. Perez, and M. Werner, "AI-enhanced drought forecasting: a case study in the Netherlands," Copernicus Meetings, EGU22-11408, Mar. 2022. https://doi.org/10.5194/egusphere-egu22-11408.

26. Luong Bang Nguyen and Manh Hung Le, "Application of Artificial Neural Network and Climate Indices to Drought Forecasting in South-Central Vietnam." Accessed: Jan. 27, 2024. [Online]. Available: http://www.pjoes.com/Application-of-Artificial-Neural-nNetwork-and-Climate-Indices-to-Drought-nForecasting,105972,0,2.html

27. J. Lee, C.-G. Kim, J. E. Lee, N. W. Kim, and H. Kim, "Application of Artificial Neural Networks to Rainfall Forecasting in the Geum River Basin, Korea," *Water*, vol. 10, no. 10, Art. no. 10, Oct. 2018. https://doi.org/10.3390/w10101448.

28. A. Jalalkamali, M. Moradi, and N. Moradi, "Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index," *Int. J. Environ. Sci. Technol.*, vol. 12, no. 4, pp. 1201–1210, Apr. 2015. https://doi.org/10.1007/s13762-014-0717-6.

29. A. Kikon and P. C. Deka, "Artificial intelligence application in drought assessment, monitoring and forecasting: a review," *Stoch Environ Res Risk Assess*, vol. 36, no. 5, pp. 1197–1214, May 2022. https://doi.org/10.1007/s00477-021-02129-3.

30. Z. N. Qaisrani *et al.*, "Drought variability assessment using standardized precipitation index, reconnaissance drought index and precipitation deciles across Balochistan, Pakistan," *Braz. J. Biol.*, vol. 84, p. e261001, Jun. 2022. https://doi.org/10.1590/1519-6984.261001.

31. F. D. Kartika and P. Wijayanti, "Drought disaster modeling using drought index: a systematic literature review," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1190, no. 1, p. 012026, Jun. 2023. https://doi.org/10.1088/1755-1315/1190/1/012026.

32. M. Öney and A. Anli, "Regional Drought Analysis with Standardized Precipitation Evapotranspiration Index (SPEI): Gediz Basin, Turkey," *J Agr Sci-Tarim Bili*, vol. 29, no. 4, Art. no. 4, Nov. 2023. https://doi.org/10.15832/ankutbd.1030782.

33. P. Kumari, S. Rehana, S. K. Singh, and M. Inayathulla, "Development of a new agro-meteorological drought index (SPAEI-Agro) in a data-scarce region," *Hydrological Sciences Journal*, vol. 68, no. 9, pp. 1301–1322, Jul. 2023. https://doi.org/10.1080/02626667.2023.2206969.

34. B. Kobulniczky, I.-H. Holobâcă, Z. Črepinšek, T. Pogačar, A.-M. Jiman, and Z. Žnidaršič, "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)," Copernicus Meetings, EGU23-499, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-499.

35. B. Y. Tam, A. J. Cannon, and B. R. Bonsal, "Standardized precipitation evapotranspiration index (SPEI) for Canada: assessment of probability distributions," *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, vol. 48, no. 3, pp. 283–299, Jul. 2023. https://doi.org/10.1080/07011784.2023.2183143.

36. M. Santini, S. Noce, M. Mancini, and L. Caporaso, "A Global Multiscale SPEI Dataset under an Ensemble Approach," *Data*, vol. 8, no. 2, Art. no. 2, Feb. 2023. https://doi.org/10.3390/data8020036.

37. X. Shi, Y. Yang, H. Ding, F. Chen, and M. Shi, "Analysis of the Variability Characteristics and Applicability of SPEI in Mainland China from 1985 to 2018," *Atmosphere*, vol. 14, no. 5, Art. no. 5, May 2023. https://doi.org/10.3390/atmos14050790.

38. R. M. N. R. Azman, N. a. M. Noor, S. Abdullah, and M. M. Ideris, "Analysis of Drought Index in Sub-Urban Area Using Standard Precipitation Evapotranspiration Index (SPEI)," *International Journal of Integrated Engineering*, vol. 14, no. 9, Art. no. 9, Dec. 2022.

39. P. Sandhya Krishna, B. Yamini Krishna, Shaik. Nafisa, T. Ratna Sravani, J. Ragha Madhuri, and Ch. Vanditha, "Prediction of Droughts using SPEI," in *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, Apr. 2023, pp. 839–845. https://doi.org/10.1109/CSNT57126.2023.10134742.

40. R. M. N. R. Azman, N. a. M. Noor, S. Abdullah, and M. M. Ideris, "Analysis of Drought Index in Sub-Urban Area Using Standard Precipitation Evapotranspiration Index (SPEI)," *International Journal of Integrated Engineering*, vol. 14, no. 9, Art. no. 9, Dec. 2022.

41. J. A. M. Careto, P. M. M. Soares, R. M. Cardoso, A. Russo, and D. C. A. Lima, "A new ensemble-based SPI and SPEI index to depict droughts projections for the Iberia Peninsula with the EURO-CORDEX," Copernicus Meetings, EGU22-12405, Mar. 2022. https://doi.org/10.5194/egusphere-egu22-12405.

42. V. H. Nguyen and H. P. H. Yen, "Seasonal variation and its impacts in rice-growing regions of the Mekong Delta," *International Journal of Climate Change Strategies and Management*, vol. 13, no. 4/5, pp. 483–491, Jan. 2021. https://doi.org/10.1108/IJCCSM-05-2020-0048.

43. Thanh Nguyen Giao and Thi Hong Huynh Nhien, "Evaluating Water Quality Variation in the Vietnamese Mekong Delta Area Using Cluster and Discriminant Analysis | Applied Environmental Research." Accessed: Jan. 30, 2024. [Online]. Available: https://ph01.tci-thaijo.org/index.php/aer/article/view/240912

44. T. K. H. Dinh and T. A. Dang, "Potential risks of climate variability on rice cultivation regions in the Mekong Delta, Vietnam," *Rev. bras. eng. agríc. ambient.*, vol. 26, pp. 348–355, Feb. 2022. https://doi.org/10.1590/1807-1929/agriambi.v26n5p348-355.

45. T. L. Cosslett and P. D. Cosslett, "The Mekong Delta Provinces," in *Water Resources and Food Security in the Vietnam Mekong Delta*, T. L. Cosslett and P. D. Cosslett, Eds., in Natural Resource Management and Policy. , Cham: Springer International Publishing, 2014, pp. 23–74. https://doi.org/10.1007/978-3-319-02198-0_2.

46. Minh Huynh, Pankaj Kumar, and Nguyen Van Toan, "Deciphering the relationship between meteorological and hydrological drought in Ben Tre province, Vietnam." Accessed: Jan. 26, 2024. [Online]. Available: https://www.researchsquare.com

47. K. Zhou, X. Shi, and F. Renaud, "Understanding precipitation moisture sources of the Vietnamese Mekong Delta and their dominant factors during recent drought events," Copernicus Meetings, EGU23-11377, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-11377.

48. K. Lavane *et al.*, "Assessing the Effects of Drought on Rice Yields in the Mekong Delta," *Climate*, vol. 11, no. 1, Art. no. 1, Jan. 2023. https://doi.org/10.3390/cli11010013.

49. H. V. T. Minh *et al.*, "Understanding Dry and Wet Conditions in the Vietnamese Mekong Delta Using Multiple Drought Indices: A Case Study in Ca Mau Province," *Hydrology*, vol. 9, no. 12, Art. no. 12, Dec. 2022. https://doi.org/10.3390/hydrology9120213.

50. D. P. Nguyen, H. D. Ha, N. T. Trinh, and M. T. Nguyen, "Application of artificial intelligence for forecasting surface quality index of irrigation systems in the Red River Delta, Vietnam," *Environmental Systems Research*, vol. 12, no. 1, p. 24, Jul. 2023. https://doi.org/10.1186/s40068-023-00307-6.

51. V. Rattayová, M. Garaj, M. Kandera, and K. Hlavčová, "Evaluation of Hargreaves method for calculation of reference evapotranspiration in selected stations of Slovakia," Copernicus Meetings, EGU23-1712, Feb. 2023. https://doi.org/10.5194/egusphere-egu23-1712.

52. K. Al-Asadi, A. A. Abbas, A. S. Dawood, and J. G. Duan, "Calibration and Modification of the Hargreaves–Samani Equation for Estimating Daily Reference Evapotranspiration in Iraq," *Journal of Hydrologic Engineering*, vol. 28, no. 5, p. 05023005, May 2023. https://doi.org/10.1061/JHYEFF.HEENG-5877.

53. D. L. Koç and M. E. Can, "Reference evapotranspiration estimate with missing climatic data and multiple linear regression models," *PeerJ*, vol. 11, p. e15252, Apr. 2023. https://doi.org/10.7717/peerj.15252.

54. N. A. Elagib and A. A. Musa, "Correcting Hargreaves-Samani formula using geographical coordinates and rainfall over different timescales," *Hydrological Processes*, vol. 37, no. 1, p. e14790, 2023. https://doi.org/10.1002/hyp.14790.

55. G. Hargreaves and Z. Samani, "Reference Crop Evapotranspiration From Temperature," *Applied Engineering in Agriculture*, vol. 1, Jan. 1985. https://doi.org/10.13031/2013.26773.

56. G. M. Mulualem and Y.-A. Liou, "Application of Artificial Neural Networks in Forecasting a Standardized Precipitation Evapotranspiration Index for the Upper Blue Nile Basin," *Water*, vol. 12, no. 3, Art. no. 3, Mar. 2020. https://doi.org/10.3390/w12030643.

57. S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index," *Journal of Climate*, vol. 23, no. 7, pp. 1696–1718, Apr. 2010. https://doi.org/10.1175/2009JCLI2909.1.

58. Max Hinne, Quentin F. Gronau, and Eric-Jan Wagenmakers, "A Conceptual Introduction to Bayesian Model Averaging." Accessed: Feb. 02, 2024. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/2515245919898657

59. Tuan N. V., "Regression Models and Scientific Discovery," Sách Khai Minh - Tri Thức Là Sức Mạnh. Accessed: Feb. 02, 2024. [Online]. Available: https://www.sachkhaiminh.com/mo-hinh-hoi-quy-va-kham-pha-khoa-hoc-gs-nguyen-van-tuan

60. J. Fan, X. Ma, L. Wu, F. Zhang, X. Yu, and W. Zeng, "Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data," *Agricultural Water Management*, vol. 225, p. 105758, Nov. 2019. https://doi.org/10.1016/j.agwat.2019.105758.

61. J. Fan *et al.*, "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 Conversion and Management*, vol. 164, pp. 102–111, May 2018. https://doi.org/10.1016/j.enconman.2018.02.087.

62. R. E. Schapire, "The Boosting Approach to Machine Learning: An Overview," in *Nonlinear Estimation and Classification*, D. D. Denison, M. H. Hansen, C. C. Holmes, B. Mallick, and B. Yu, Eds., in Lecture Notes in Statistics. , New York, NY: Springer, 2003, pp. 149–171. https://doi.org/10.1007/978-0-387-21579-2_9.

63. Z. Ali, Z. Abduljabbar, H. Tahir, A. Sallow, and S. Almufti, "Exploring the Power of eXtreme Gradient Boosting Algorithm in Machine Learning: a Review," vol. 12, pp. 320–334, May 2023. https://doi.org/10.25007/ajnu.v12n2a1612.