

Article

Analyzing Climate Dynamics and Developing Machine Learning Models for Flood Prediction in Sacramento, California

Dewasis Dahal, Binod Ale Magar, Abhiru Aryal, Bishal Poudel, Mandip Banjara and Ajay Kalra *

School of Civil, Environmental, and Infrastructure Engineering, Southern Illinois University, 1230 Lincoln Drive, Carbondale, IL 62901-6603, USA; dewasis.dahal@siu.edu (D.D.); binod.alemagar@siu.edu (B.A.M.); abhiru.aryal@siu.edu (A.A.); bishal.poudel@siu.edu (B.P.); mandip.banjara@siu.edu (M.B.)

* Corresponding author. E-mail: kalraa@siu.edu (A.K.); Tel.: +1-618-453-7008 (A.K.)

Received: 12 September 2024; Accepted: 25 October 2024; Available online: 31 October 2024

ABSTRACT: Climate change is leading to rapid environmental changes, including fluctuating precipitation and water levels, which raises the risk of flooding in coastal and riverine locations around the United States. This study focuses on Sacramento, California, a city significantly affected by these changes and recent severe flooding disasters. The ultimate goal is to understand the climate dynamics and create a more robust model to alert Sacramento and other communities to possible flooding and better prepare them for future climatic uncertainty. In this research, four classification machine learning models—Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM)—are examined for their capacity to predict the occurrence of floods using historical precipitation temperature and soil moisture data. Our results demonstrate that the LSTM model, with an accuracy of 89.99%, may provide better reliable flood predictions, possibly due to its ability to process complicated temporal data. SVM, RF, and ANN showed accuracies of 81.25%, 83.75%, and 85%, respectively. The study explores the correlation between increasing precipitation incidents and severe climate variations, such as the El Niño and La Niña cycles, which could have increased flooding risks. Significant rainfall peaks occurred in 1998 and 2007, indicating that external atmospheric circumstances might have considerably impacted local weather patterns. While LSTM models show potential, there remains room to improve their accuracy and adaptability in extreme flood scenarios. Given these findings, future research could combine multiple environmental data sources and hybrid modeling approaches to enhance predictions.

Keywords: SVM; Random Forest; ANN; LSTM; Flood



© 2024 The authors. This is an open access article under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The precipitation abnormality due to climate change has affected the rise of sea levels around the coastal areas. The main reasons behind the flooding in coastal areas are sea level rise, land subsidence, and increased urbanization [1]. In the twentieth century, the sea level has risen by 1.8 mm/year, and in a few decades, it has risen to 1.8 mm/year [2]. According to the National Oceanic and Atmospheric Administration, the sea level may rise approximately 254–304 mm in coastal areas of the United States (U.S.) by 2050, approximately the same as from 1920 to 2020 [3]. The increase in temperature due to global warming causes the warming of oceans and extreme rainfall events, causing floods around the coastal area. In addition, flooding in coastal areas results from land subsidence caused by groundwater pumping, oil and gas extraction, and an increase in impervious surfaces due to urbanization [4]. As global temperatures rise by 2 °C, the intensity of floods is predicted to increase considerably, potentially destroying places at a rate four to five times greater at 4 °C compared to 1.5 °C [5]. According to the IPCC Fifth Assessment Report (AR5) published in 2014, 20% of the population around coastal areas will be exposed to 100 coastal floods when the global mean sea level rises by 0.15 m [6]. The exposed population will double when the sea level rises by 0.75 m and triple when the sea level rises by 1.4 m, even when there is no change in the population of the coastal areas [6]. Consequently, understanding the environmental connections is essential to minimizing the impact of these changes and effectively controlling flood risks.

The United States coastal regions have some of the major cities in California, which are essential metropolises and central hubs for the country's foreign trade. Due to the rise of sea levels and their impact on marshy terrain, California

has recently had to deal with the issue of urban flooding and coastal danger [7]. Researchers estimate that by the year 2100, the average sea level will have risen by 1 to 1.4 m due to ongoing linear trends in global warming, based on forecasts of medium to medium-high emission scenarios [8]. The Mediterranean climate of California, which sees much precipitation in the winter, is another factor in the state's history of floods [9]. In recent years, California has seen a rise in the frequency and severity of flooding. California has experienced floods due to severe storms and heavy rain, seriously damaging the state's infrastructure and property. On 9 December 2022, there was a flood in the airport region, and research suggests that 90% of California's coastal airports are at risk owing to continual, worsening flood events [10]. Recent floods include floods on 9 January 2023, caused by an atmospheric river, and an extensive flood in Sacramento County on 31 December 2022 [9,11]. Given these challenges, alarming flood prediction systems are crucial for reducing possible damage in Sacramento, California. Improving flood prediction is critical in the Sacramento region, where its flat terrain and breaking levees enhance flood risk, requiring urgent state and municipal measures to strengthen defenses and protect against the growing impacts of climate change-induced floods [12].

The flood forecast makes the ability to protect oneself, one's property, and one's community from potential harm and disturbance possible, which is crucial for individuals, businesses, and government authorities [13]. It is the framework for developing an early warning system to lessen the potential severity of the flood. The warning system makes establishing an emergency response plan and activating safe shelters across the community possible [14]. Predictions of flooding are crucial for managing water resources. Water managers can release water from reservoirs in a controlled manner by giving a head up, lowering the possibility of a dam failure and floods downstream [15]. To minimize the effects on the communities and the state's economy, flood prediction is essential for safeguarding people, their possessions, and the environment [16]. Therefore, the significance of precise and timely flood predictions is expected to increase due to increased heavy rainfall and coastal area flooding caused by climate change.

Machine learning algorithms can be beneficial in flood prediction by delivering precise and timely flood risk estimates. In recent years, machine learning models have significantly increased, and they continue to be modified and updated to predict floods, providing the basis for developing early warning systems [17]. These models do extensive historical and real-time data analyses on various flood-influencing variables and forecast future flood episodes [18]. The ability of machine learning models to process enormous volumes of complex data is the main benefit of employing them to predict floods. Understanding the relationships among environmental characteristics such as temperature, rainfall, and sea level rise is crucial to minimizing the flooding effects on communities and improving prediction accuracy [19]. Several parameters, such as precipitation, temperature, soil moisture, and wind velocity, are critical for understanding and predicting flood events because they directly influence the hydrological processes within a watershed [20]. Conversely, other essential factors such as land topography, changes in land usage, and changes in land cover do not undergo considerable changes during short periods since significant changes in topography usually take several years to appear [21]. Therefore, it is necessary to understand the dynamic relationship between parameters like temperature, precipitation, and soil moisture to develop an early warning system that can predict the occurrence of floods as they offer timely, valuable data for flood prediction models [22].

Numerous research investigations have highlighted the need for and significance of developing "smart" flood forecasting models in response to some of these upcoming issues regarding prevention, adaptation, and mitigation. For instance, Fahad et al. (2022) created a framework to assist in decision-making when assessing whether the chosen method is appropriate for a portion of the coastal region that experiences the strongest storms [23]. Similarly, Nazari et al. (2022) investigated treatment facility concerns brought on by exceptionally severe floods, which highlights the need for models to be incorporated in forecasting flood frequency and intensity [24]. These papers highlight the current flood prediction techniques' more severe and extensive shortcomings. Recent advancements in flood forecasting have demonstrated the effective integration of machine learning techniques with hydrological models. For instance, Wang et al. (2023) recently presented a stable and understandable flood forecasting model by combining multiple linear regression with a multi-head attention mechanism, which improved the model's interpretability and prediction performance [25]. In the Lower Yellow River Basin, Wang et al. (2023) conducted a comparable analysis comparing the flood-predicting capabilities of LSTM and RNN at the HuaYuankou and LouDe stations [26]. To capture intricate hydrological patterns, the results may demonstrate that the LSTM model performs better than the RNN [26]. Additionally, the literature indicates that Wang et al. (2021) have significantly improved flood forecasting performance using spatially distributed velocity fields based on geomorphic unit hydrographs [27]. Wang et al. (2024) have presented a novel method for predicting flood flow that integrates spatiotemporal data through a two-dimensional hidden layer structure, showing great promise for improving predictive capabilities [28]. These recent studies provide worthy insights that could be useful in developing better models that provide more accurate flood forecasts. These new studies offer

valuable insights that may help design better models that produce more accurate flood forecasts. The models used to anticipate floods are typically built on the foundation of machine learning techniques. Nevertheless, there are several drawbacks. Lawal et al. (2021) used logistic regression, decision trees, and SVM to address flood forecasting challenges. Still, they were inaccurate because they could not address time sequence and non-linear issues [29]. However, Sarkar et al. (2024) have shown that while decision trees and random forests perform satisfactorily in estimating floods driven by rainfall, they do not perform better in other locations with damper weather patterns, such as Sacramento [30]. This research will use the Long Short-Term Memory (LSTM) model to recognize its effectiveness. It can recognize long-term relationships in historical climate data and is consequently more accurate at predicting floods in areas with unpredictable weather.

Our research is motivated by the need to develop a reliable classification model for predicting flood occurrence using machine learning techniques, which can serve as an early warning for flood disasters. This research study aims to analyze the historical interaction of climate dynamics with flood events in Sacramento, California, by collecting and processing historical precipitation and temperature data and identifying significant flood events during the study period. Additionally, our research aims to develop and compare four machine learning models—Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM) which are used for flood prediction model for Sacramento, California. The performance of these models is compared using metrics such as accuracy, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). This study also assesses the implications of employing machine learning models for flood prediction in the context of climate change.

2. Site Description and Data

2.1. Study Area

The study focuses on predicting floods in the Sacramento area. The particular study area is the downstream area of the watershed delineated at Station ID 11425500. The flooded areas downstream of this location are Laguna, Franklin, Thornton, and Sheldon. The location of the gauge station is Verona, with a latitude of 38.77 and a longitude of -121.598 . The watershed area covers approximately 64,796 km². While there are multiple precipitation gauge stations, we will focus on station ID 390135121261001 located in Wheatland, Yuba County. This station provides the necessary 11,680 data points collected at daily intervals for our machine learning models. Its coordinates are latitude 39.02 and longitude -121.43 . Figure 1 shows the location map of the watershed considered.

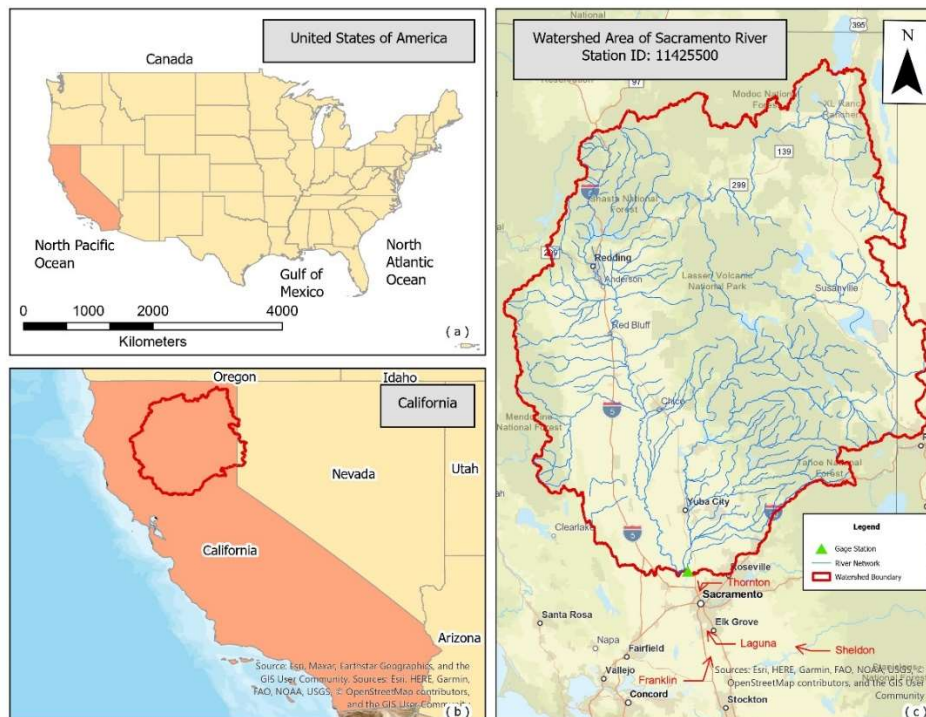


Figure 1. (a) Map of the United States; (b) Map of California with highlighted watershed area; (c) Detailed map of the Sacramento River watershed area.

The Sacramento River has a long history of flooding, with notable flood events recorded in 1986, 1995, 1997, 2006,

and 2017 [7]. The Sacramento Area Flood Control Agency identifies Sacramento as the nation’s greatest metropolitan flood risk. The flood in 1986 in Sacramento was due to the 10 inches of rainfall in 11 days [7]. The flood in 2006 was also due to the huge New Year’s Eve Storm, which lasted until the early week of 2006. The next catastrophic flood was recorded on 3 January 2017 and lasted till 24 January 2017; due to the recent flood news and its catastrophic historical events, the site was chosen to build the model that can be useful in predicting floods [10].

2.2. Data Sources

The ML models in this study utilized datasets about the study area, including boundary and rainfall data, as input parameters. The data span from 1990 to 2023 AD. Table 1 enumerates the various data sources employed. Regarding data sources, TerraClimate provides high-resolution precipitation data, which is critical for thorough flood modeling in Sacramento. NCEI maintains substantial long-term temperature records, which are critical for understanding temperature trends and their impact on flood dynamics. Streamstat provides high-quality hydrological data for establishing watershed boundaries and studying the factors influencing flood risk areas. The ERA5-Land dataset provides high-resolution hourly soil moisture data, critical for simulating the antecedent conditions that determine flood risk. These extensive datasets guarantee the accuracy and reliability of our flood prediction models.

Table 1. Sources of Study Area Data.

Data	Sources
Precipitation	TerraClimate from Climate Engine (https://www.climateengine.org/ (accessed on 1 March 2024))
Temperature	National Centers for Environmental Information (https://www.ncei.noaa.gov/cdo-web/ (accessed on 1 March 2024))
Boundary Shapefile	National Centers for Environmental Information (https://www.ncei.noaa.gov/cdo-web/ (accessed on 1 March 2024))
Soil Moisture	ECMWF-ERA-5 (ERA5-Land ECMWF (accessed on 1 May 2024))

3. Methodology

The study involves extracting data influencing flow values and using it to train and test various machine-learning models. The accuracy of these models is compared to determine the most effective one for predicting floods, which can be instrumental in creating an emergency response plan. The relevant data was collected and utilized to create the machine-learning models. The dataset was split into training and testing subsets to evaluate the models’ performance. Various evaluation metrics were applied to assess each model’s effectiveness. Finally, the model with the highest accuracy was selected as the flood prediction model. The comparison of monthly rainfall data with a 12-month moving average was used in the study in addition to the flood prediction model, and it effectively illustrates the long-term trends and patterns in the rainfall data. Figure 2 shows the methodology flowchart.

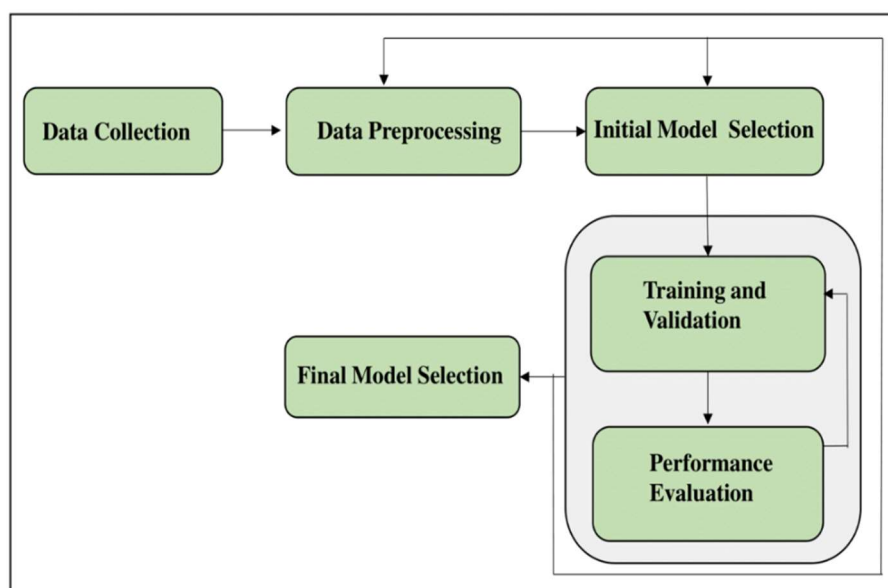


Figure 2. Diagrammatic Representation of Research Methodology.

3.1. Data Acquisition and Preprocessing

The daily rainfall records spanning thirty-two years (1 January 1990 to 31 December 2022) for Sacramento County, California, were obtained from Climate Engine, a web application powered by Google Earth Engine. This dataset consists of 11,680 data points collected at daily intervals. Prior to data preparation, the collected data was analyzed. Measures were taken during preprocessing to improve computation speed and accuracy. To simplify the dataset, monthly rainfall totals were first calculated. Calendar mean values were used to remedy the missing data points. Annual precipitation totals were also calculated to provide a more complete picture. The dataset was then divided into two subsets: training and testing. The train test split model selection approach assigned 80% of the data for model training and 20% for testing. The testing data will help us to understand the performance of the model. The test data input is precipitation, soil moisture, and temperature, which will predict the flow. The model was iterated over training and testing using 5-fold cross-validation techniques to ensure it generated well in new, unseen data while preventing overfitting.

3.2. Machine Learning Model Design

The study utilized a web-based interactive Jupyter Notebook environment to create a machine-learning model for flood prediction. The decision was motivated by its strong performance with machine learning and data analysis and its fast translation of scripts for real-time execution. It offers a dynamic interface appropriate for large-scale numerical and analytical processes [31]. Preparing the data was the first step, and then the model was tested to see how accurate it was predicted.

The average monthly rainfall is used as a threshold indicator in this baseline. Unlike the study by Lawal et al. in 2021, the model predicts no floods for that year if the monthly rainfall falls short of this average [29]. On the other hand, exceeding this threshold suggests a greater likelihood of flooding, a conclusion supported by our research and the recommendations of Lawal et al. [29]. A “Flood” characteristic was added to our dataset to improve data analysis. This attribute classified monthly rainfall numbers as above or below the average, with ‘YES’ (1) denoting more rainfall and ‘NO’ (0) lower. This approach is similar to the one used in the [30–33] studies, where the monthly, seasonal, and yearly rainfall measures are used as a baseline for predicting floods.

Moreover, the study rigorously investigated the effectiveness of four simple machine-learning techniques to enhance the accuracy of flood predictions. The selected underscored the model’s practicality in risk management and disaster planning techniques and demonstrated promising potential for predicting future flood disasters, thereby emphasizing the real-world significance of this research.

3.3. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning model that can be applied to classification and regression problems. Developed in the 1960s by Russian mathematician Vapnik, SVM identifies the optimal dividing line, known as the maximum-margin hyperplane, between classes in a dataset [34]. It seeks to increase the margin between the dataset’s classes using the widest separator available [35]. To accommodate non-linear data, SVM employs a kernel function that projects the input data into a space where a linear separator is possible [36]. This trait is beneficial when working with high-dimensional data or when there are fewer observations than features. SVM’s efficiency originates from using support vectors, which are data points closest to the decision focus and significantly decrease the determining load [31]. This research focuses on the SVM’s Support Vector Classification (SVC) part, which categorizes data using numerous strategies such as one-vs-one and one-vs-all. Our study uses the all-vs-all technique because of its robustness in binary classification across several classes. Previous research has shown that classification accuracy gains from well-separated classes. SVC performs in scenarios with complex feature connections, making it an effective tool for bioinformatics, image recognition, and text categorization applications. The fundamental goal of SVC in this study is to accurately separate case construct situations and a predictive model based on the contrasts drawn between these cases. The schematic diagram of the SVM is shown below in Figure 3.

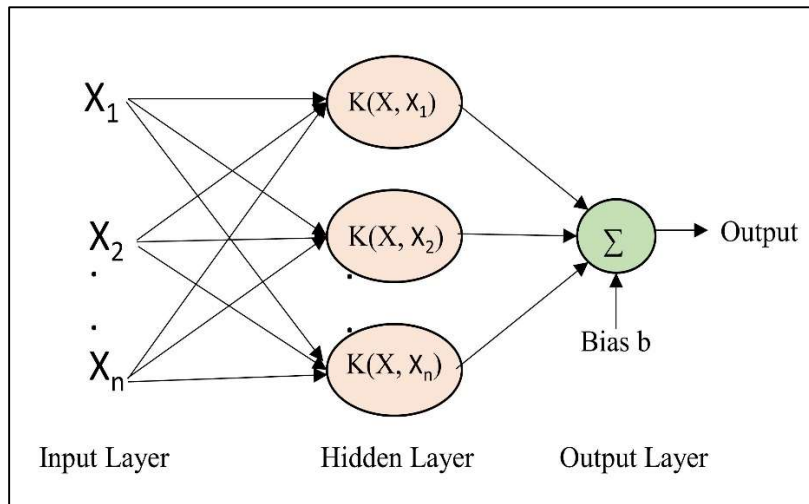


Figure 3. The schematic diagram of the Support Vector Mechanics.

The simple mathematical formula of the SVM model is represented in Equation (1) [37].

$$\text{SVM}(C, \gamma, \text{kernel}) = \arg \max C, \gamma \{ \text{Accuracy}(\text{SVC}(C, \gamma, \text{kernel})) \} \tag{1}$$

Here C is the regularization parameter, γ is the kernel coefficient and kernel are the kernel type to be used in the algorithm.

3.4. Random Forest

The Random Forest approach is a composite method that combines many decision trees to improve the durability and precision of prediction models [38]. The present research uses it for classification and develops an aggregate model by training individual trees on different data subsets, using random feature subsets at every division. This technique reduces overfitting by averaging differences among trees. Figure 4 illustrates the schematic diagram of the Random Forest model.

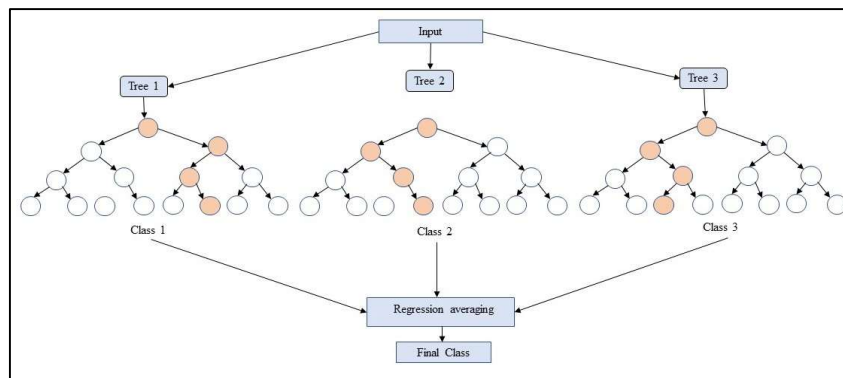


Figure 4. Schematic diagram of the Random Forest model.

Due to our selection of hyperparameters, the Random Forest model in our study is trained with a particular configuration of a hundred trees ($n_estimators = 100$), improving control over model complexity and ensuring an ideal balance between bias and variance. In the study, T is the set of decision trees, and t is a decision tree in T . Its performance is measured by comparing these predictions to the actual test labels y , resulting in a score for accuracy given as a percentage. Our research employed the Random Forest Classifier from the Sklearn library. In the model, the color circles highlight the nodes involved in the decision route. In Equation (2), y represents the final prediction for given input features (x). [39].

$$y = \text{majority_vote}(t) \text{ for } t \text{ in } T \tag{2}$$

where the majority vote($t(x)$) shows the class predicted by the majority of the decision trees.

3.5. Artificial Neural Network

An Artificial Neural Network (ANN) is a model for machine learning based on the structure and function of biological neural networks [40]. An ANN can be represented mathematically as a series of matrix multiplications, known as feedforward computation, followed by a non-linear activation function [41]. ANN constructs layers of artificial neurons that process and send information. The network input is routed through the input layer, which linearly transforms the input using a weight matrix. Data from the input layer is transmitted through one or more hidden layers that perform linear transformations and non-linear activations before reaching the output layer, where a final linear transformation and activation generate the network's final output [42]. A simple graphical representation of ANNs architecture is demonstrated in Figure 5.

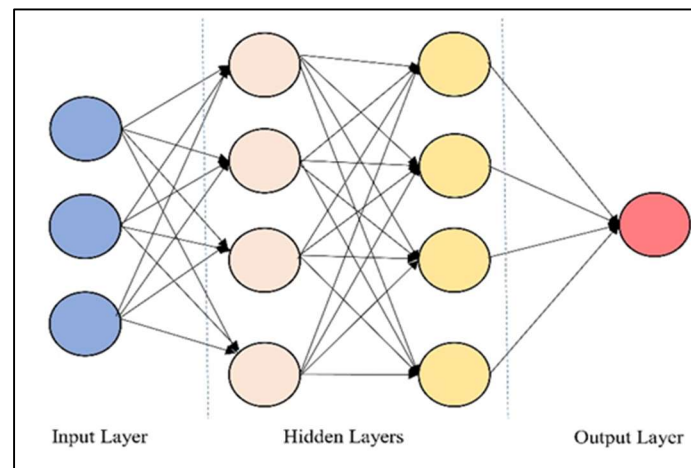


Figure 5. Graphical representation of ANNs architecture.

The ANN layout utilized in our research includes two hidden layers, each with 64 neurons, dropout layers to minimize overfitting, and ReLU activation functions. It was constructed using TensorFlow's Sequential API. This model is trained across 100 epochs with a batch size of 32, and a validation split of 20% is included to monitor and prevent overtraining. The model is constructed with the Adam optimizer and binary cross-entropy loss. Its mathematical representation is as Equation (3) [43].

$$f(x) = \left(\frac{1}{1 + \exp^{-x}} \right) \quad (3)$$

3.6. Long Short-Term Memory

The Long Short-Term Memory Model. LSTM can handle sequential data, including time series and plain language [44]. An LSTM network is a chain-like structure created by connecting numerous LSTM cells. In addition to three gates (input, forget, and output) that control the flow of information into and out of the cell, each LSTM cell has an internal state that can store data [45]. The input gate controls data entering a cell, the forget gate controls information leaving a cell, and the output gate controls data heading from one layer to the next [46]. This fundamental method enables LSTM to handle long-term dependencies while maintaining context over time by selectively preserving or forgetting information as it analyzes the sequence [47]. This study carefully configures the LSTM model to recognize patterns in time-sequenced data. The rainfall data used in the LSTM model consists of 11,680 daily data points collected over a thirty-two-year period. A StandardScaler is used to standardize the data before it is reshaped to meet the specifications of the LSTM input. A hyperparameter tuning method utilizing Keras Tuner, which assesses various LSTM and dense layer units and varying learning rates, optimizes the model's LSTM layer to increase its output efficiency. Dropout layers are built into this method to reduce overfitting and ensure an ideal structure. Training involves several epochs and validation splits to improve the model's performance. The resulting model is then evaluated on scale test data to verify prediction accuracy after the optimum hyperparameters have been identified and established. Figure 6 illustrates a visual representation of the LSTM model.

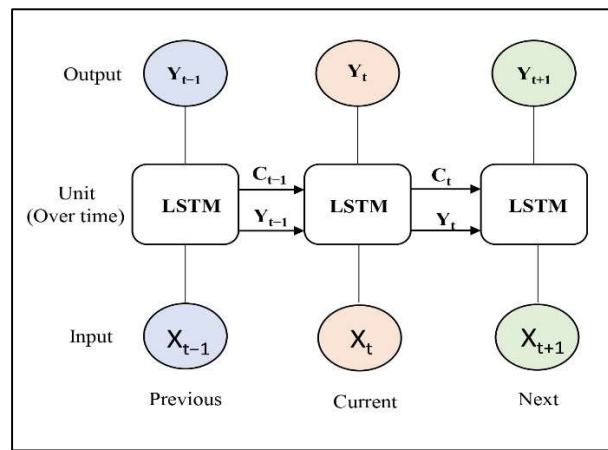


Figure 6. Graphical representation of LSTM model.

The simple mathematical formula of the LSTM model is represented in Equation (4) [48].

$$h_t = \text{LSTM}(x_t, h_{t-1}) \tag{4}$$

In Equation (4), h_t means output at time step t , as a function of the current input x_t and the previous output h_{t-1} .

3.7. Performance Evaluation

The assessment of the models involved a range of evaluation techniques. These metrics offer an in-depth analysis of the model’s performance, facilitating the identification of models that excel in accuracy, dependability, and predictive capability. Our research employed five critical metrics for model evaluation: root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE), percent bias (PBIAS), coefficient of determination (R^2), and normalized root mean squared error (NRMSE) to gauge the models’ effectiveness [49]. Additionally, recall, precision, F1-score, and the area under the curve (AUC) were computed for a more detailed evaluation [50].

3.7.1. Confusion Matrix

To further assess the model’s performance, this article used a confusion matrix presented in Table 2. The confusion matrix includes:

- True Positive (TP): Both actual and predicted classes are positive.
- False Positive (FP): Actual class is negative, but the predicted class is positive.
- False Negative (FN): Actual class is positive, but the predicted class is negative.
- True Negative (TN): Both actual and predicted classes are negative.

Table 2. The representation of the 2D confusion matrix.

	Predicted-Positive	Predicted-Negative
Actual-Positive	TP	FN
Actual-Negative	FP	TN

3.7.2. Receiver Operating Characteristics (ROC)

The Receiver Operating Characteristics (ROC) curve was also employed as an additional metric to evaluate model performance in this study. The ROC curve is a graphical representation that shows the performance of a binary classification model by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings [51]. The true positive rate, plotted on the Y-axis, measures the proportion of actual positives correctly identified by the model. The false positive rate, plotted on the X-axis, indicates the proportion of actual negatives incorrectly identified as positives.

The area under the ROC curve (AUC), a crucial metric in model performance evaluation, ranges from 0 to 1. A value of 1 represents perfect classification [51]. Higher AUC values, indicating better model performance, demonstrate a greater ability to discriminate between positive and negative classes, thus underlining the importance of this metric in this study.

3.7.3. Parameters Selection

Table 3 displays the specific model parameters for each technique, including the number of iterations, neurons, and depth of hidden layers. These criteria were established based on initial testing to identify each model's ideal performance for flood prediction.

Table 3. Model Parameters details.

S.N.	Models	Parameters
1	SVM	Kernel = rbf, C = 100, gamma = 0.1, epsilon = 0.1
2	RF	n_estimators = 100, random_state = 42
3	ANN	Hidden layer = 100, alpha = 0.0001, solver = adam, verbose = 10, random state = 42, tol = 0.000000001
4	LSTM	Neurons = 256, Optimizer = adam, loss = mse, epochs = 300

4. Results and Discussions

The main objective of this research was to develop a machine learning model for predicting possible flooding occurrences using several machine learning models. Additionally, the performance of the four distinct flood prediction models was compared to identify which model might have been the most effective. Furthermore, this section examined the relationship between rainfall temperature and flood incidents, as well as the changes in the county's climate during the previous 30 years. Determining the probability of flooding disasters requires understanding these patterns and correlations.

4.1. Analyzing Sacramento's LongTerm Rainfall and Climate Trends across Three Decades

Figure 7a compares monthly rainfall data with a 12-month moving average and reveals the complex patterns and aberrations in nearly a century of rainfall records. This study discusses moving average analysis because it clearly depicts the long-term trends and patterns in the rainfall data. Understanding the past interactions between flood occurrences and climate dynamics requires identifying the changing weather patterns. Individual monthly data were shown in light blue to highlight the inherent variability, while the moving average in red smoothed out the fluctuation to reveal underlying trends. The statistics revealed substantial seasonal changes, with an exceptionally high peak in 1998 when rainfall reached 312 mm. The unusual rise correlated with the El Niño event of that year, indicating that complex interactions between atmospheric and oceanic conditions could enhance precipitation [52]. The moving average indicated a gradual but steady increase in rainfall beginning in the early 2000s and peaking in the latter half of the decade. This gradual increase, which peaked at an average of 98 mm in 2015, might have indicated changing climatic conditions, potentially due to the overall influence of global [53]. While this trend highlighted changing rainfall patterns, it also showed the possibility of increased frequency and severity of rainfall events in the region. Years such as 1998 and 2007, which observed rainfall surges of 312 mm and 286 mm, respectively, stood out as examples of extreme weather events. These fluctuations might be attributed to global atmospheric phenomena, including the well-documented El Niño and La Niña cycles [54].

A 12-month moving average clarified the data by condensing it, providing a more visible perspective on regular and irregular precipitation trends. This analytical method did not just expose the unpredictability and seasonality associated with rainfall; it emphasized the significance of considering multiple aspects in interpreting such data, ranging from regional weather systems to global climate variations and the need for flood risk assessment and management.

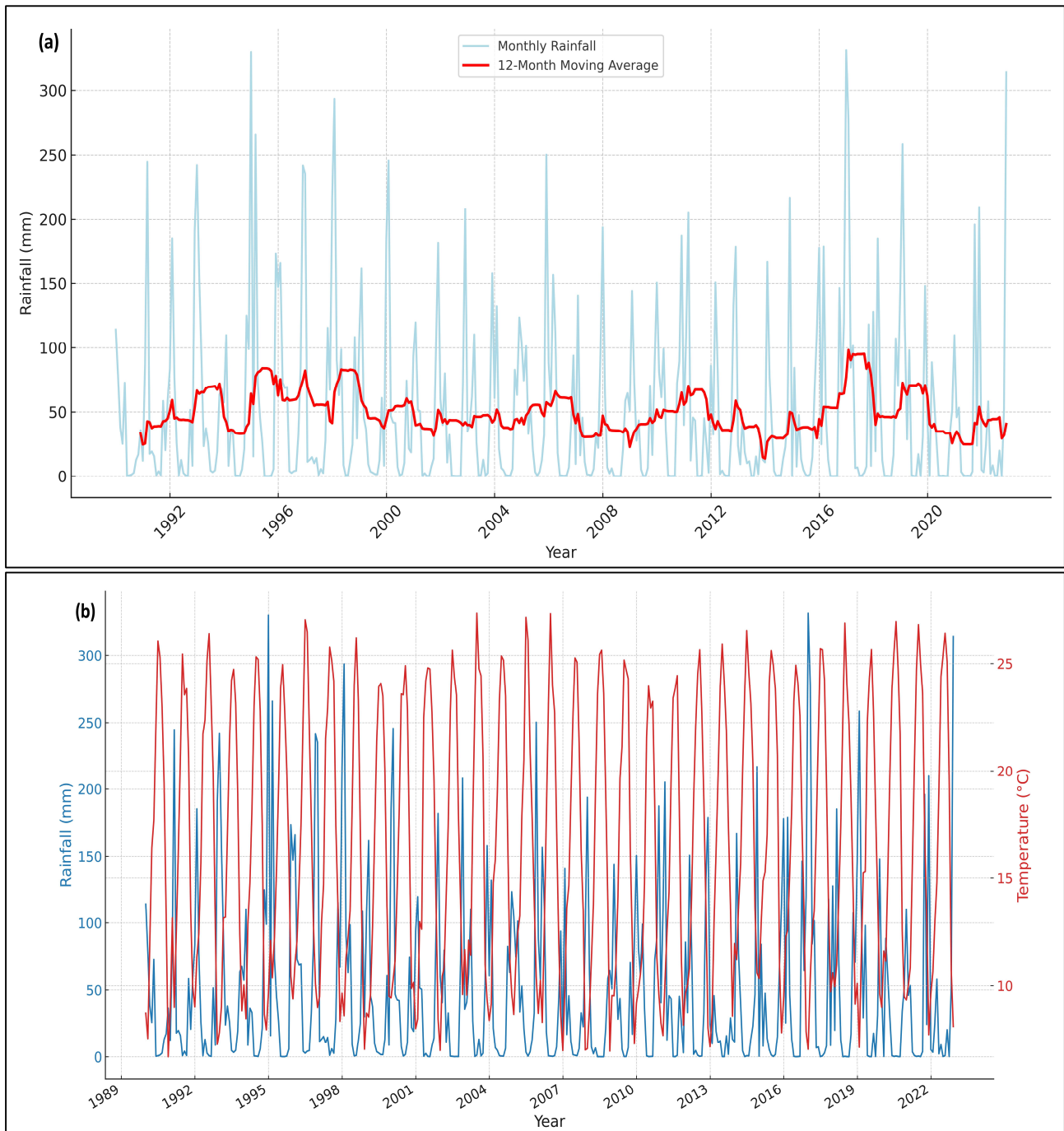


Figure 7. (a) Average rainfall vs year with a moving average of 12 months. (b) Time-series plot for temperature and rainfall for Sacramento County, California.

Long-term meteorological data analysis was essential for understanding the temporal dynamics of weather patterns and their potential impacts on regional hydrology. The visual representation in Figure 7b illustrates historical climate data, demonstrating potential relationships between rainfall (in millimeters) and temperature (in degrees Celsius) from 1990. Our observations indicated that January 2017 experienced considerable precipitation, with 331.39 mm. This occurrence appeared to be a coincidence compared to regular seasonal expectations. It also raised the possibility of climatic or environmental factors contributing to the recent increase in rainfall. Concurrently, the dataset indicated that July 2003 had the highest temperature, peaking at 27.34 °C. This contrasted with the heavy rainfall event, which did not occur during the warmer months as could have been predicted. Such data illustrated the variability of climatic patterns and the possibility of less clear temperature-precipitation connections.

The analysis of rainfall and temperature across time indicated a complex interplay between the two factors. The maximum rainfall in January might have indicated the influence of specific atmospheric events, such as a strong monsoon season or the effects of aquatic life that caused wetter winters in the area. Furthermore, the statistics demonstrated that extreme weather conditions did not always occur during the hottest months, as is commonly supposed.

This realization led us to examine additional meteorological elements, such as wind stream patterns or oceanic currents, that may have played a role in the unusual seasonal rainfall distribution. A complete statistical analysis was conducted to investigate the association between temperature and rainfall over the entire dataset. Such an analysis included determining correlation coefficients to quantify the degree of the link between the variables and regression models, helping to understand better the nature and importance of the trends seen across the research period. Integrating flood incidence data, especially for months with historically significant rainfall, provided a complete knowledge of the factors that caused flooding, which is critical for disaster preparedness and resource management [55]. It was critical to contextualize localized data within larger climatic models. If the observed trends, such as increased rainfall during traditionally drier seasons, align with existing climate change projections, this would reinforce the anticipated impacts of climate change. In that case, the findings would provide valuable evidence for our evolving understanding of regional climate patterns.

4.2. Correlation of Rainfall, Temperature and Soil Moisture with Flood Events

The heatmap in Figure 8 displayed the results of a correlation analysis to determine the links between meteorological variables—rainfall and temperature—and flooding incidents. These correlations were derived from Pearson’s coefficients and could have suggested potential relationships that had to be carefully interpreted within the broader context of climatic and hydrologic research studies.

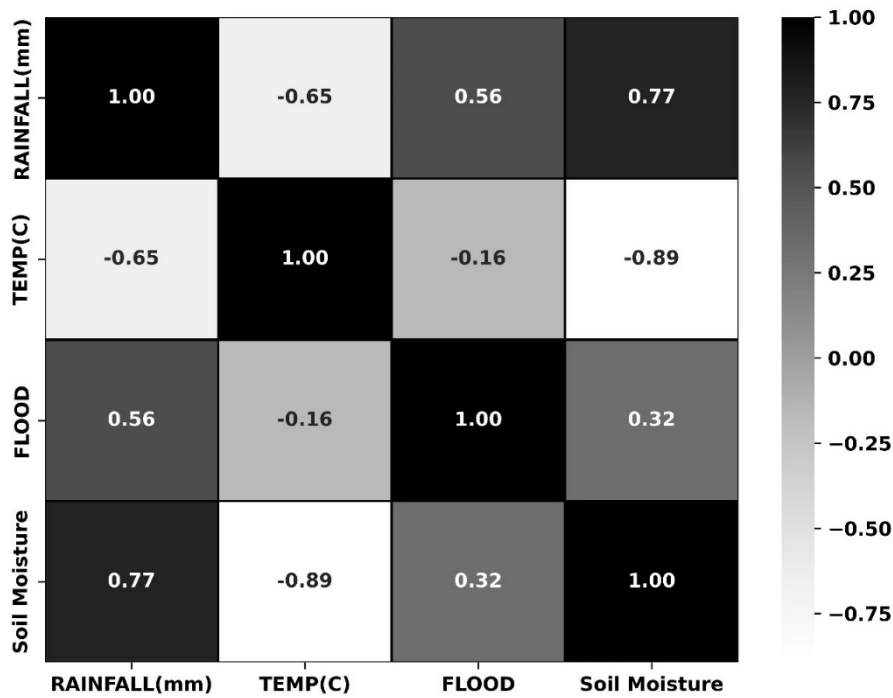


Figure 8. Correlation Heatmap of Rainfall, Temperature, Soil moisture and Flood Occurrence.

Our findings indicated a significant inverse correlation of -0.65 between rainfall and temperature. This may have reflected a climate trend in which more rainfall was frequently associated with lower temperatures, presumably due to evaporative cooling and cloud cover, which prevents solar heating. The result was consistent with the study’s reported results on temperature and precipitation patterns across the intermountain west in the United States, which supported the premise that higher precipitation correlated with lower temperatures [56]. Furthermore, the data showed a moderate positive association (0.56) between rainfall and floods. The reason for this is probably due to basin lag time, as there is a higher correlation between previous hours of rainfall and subsequent water flow. Hydrological studies supported this correlation, highlighting the essential role of precipitation variability in flooding events [22]. The research indicated a weak negative association (-0.16) between temperature and flood occurrences. This showed that warmer temperatures may have modestly reduced flood danger, maybe due to enhanced evaporation rates. However, this association could have been stronger and should have been used cautiously, as it showed only a minor influence of temperature on flooding.

According to Hettiarachchi et al., determining the impact of rising temperatures on flood risk is equally challenging, with significant regional variations [57]. Temperature variations may affect snowmelt and precipitation patterns,

changing the frequency of floods in specific areas. Similarly, Seo et al. (2024) note that while rising temperatures lessen the risk of flooding by snowfall, they may also raise the risk of flooding caused by rain and other extreme weather conditions [58]. Furthermore, when including soil moisture in our research revealed numerous exciting relationships. A substantial positive correlation (0.77) between rainfall and soil moisture can be seen, implying that increased rainfall raises soil moisture levels [59]. This relationship can be critical for understanding the pre-existing soil moisture conditions that can impact the possible risk of flooding events. The association between soil moisture and flooding (0.32) indicates that higher soil moisture levels can lead to an increased risk of flooding, as the saturated soils have a lower capacity to absorb further rainfall into the soil.

The negative correlation (−0.89) between temperature and soil moisture suggests that higher temperatures can lead to decreasing soil moisture levels, most likely due to greater evaporation [60] This interaction focuses on the need to include soil moisture dynamics in flood prediction models.

These correlations serve as a call for more in-depth research, particularly studies that encompass a broader spectrum of factors and take into account regional climatic and environmental conditions. These findings provide a strong impetus for developing comprehensive models that incorporate meteorological elements, thereby enhancing flood risk prediction and management. The research approach employed in our study was evidence-based, offering preliminary yet insightful information on the interactions between meteorological variables and flood episodes. It confirmed the idea with current literature, establishing the foundation for future studies to validate these findings and develop predictive models for flood control.

4.3. Flood Prediction Model Comparison

This results section includes Figure 9, which compares the prediction accuracy levels of various machine learning models, such as SVM, RF, ANN, and LSTM, when applied to predicting floods in the Sacramento region. The accuracy shown—81.25% for SVM, 83.75% for RF, 85% for ANN, and 89.9% for LSTM—provided a preliminary indication of each model's potential effectiveness in flood prediction. The observed trend of incremental accuracy obtained from SVM to LSTM may have indicated that models with greater complexity and temporal processing skills were more appropriate for managing the complexities of hydrological data. The performance of the LSTM model appeared to support the idea that recurrent neural networks have an inherent advantage in recognizing and exploiting sequential patterns that indicate the potential for flooding in meteorological and climatological data streams. These preliminary results could have had significant implications for flood risk management. The LSTM model, due to its ability to detect long-term relationships, might have provided more accurate forecasts, which could have helped disaster management organizations enhance their preparedness and response strategies. The LSTM model's substantially better accuracy suggested that it could improve the operational efficiency of flood warning systems.

This research could have suggested a tentative shift in the approaches used for predicting floods. Integrating sophisticated machine learning approaches, particularly LSTM networks, into established prediction frameworks may have improved their predictive accuracy. This, in turn, may have allowed for more nuanced decision-making in urban design and disaster management, increasing the resilience of communities in the Sacramento area to the adverse effects of flooding. Furthermore, the comparative analysis conducted in this study provided opportunities for future research. Despite the promising elements of the LSTM model, there may have been value in investigating hybrid models that combine the distinctive qualities of several models to overcome computing limits or shorten prediction timescales. Future research may also have benefited from investigating the combination of varied datasets, such as soil moisture or river flow data, to improve prediction precision even more.

In the broader context of environmental science research, these findings align with the work of Kim et al., who demonstrated the effectiveness of LSTM networks in environmental modeling [61]. However, it remained critical to approach these conclusions with a balanced perspective, acknowledging the need for more validation and the investigation of emergent computational strategies within this emerging field. Thus, cautious optimism is crucial to ensure our findings' robustness and reliability and inspire further research in this promising area.

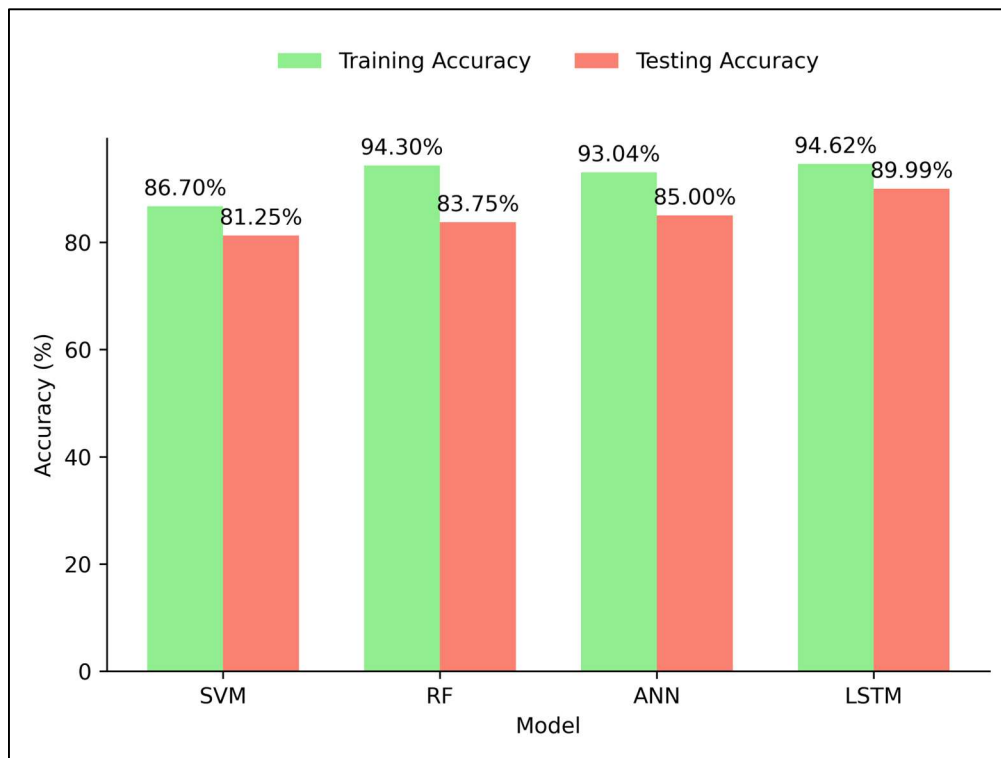


Figure 9. Accuracy Rates of Machine Learning Models.

4.4. Comparison of Model Accuracy Metrics

This section of the research paper comprehensively analyzed four predictive models—SVM, Random RF, ANN, and LSTM—to estimate the likelihood of flood occurrences in the Sacramento area.

Table 4 provides a detailed overview of the statistical metrics employed to assess the performance of each model, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Percent Bias (PBIAS), R-squared (R^2), and Normalized Root Mean Square Error (NRMSE).

Table 4. Comparison between SVM, RF, ANN, and LSTM Models for Flood Prediction.

Statistical Index	SVM Model	RF Model	ANN Model	LSTM Model
RMSE (Cfs)	0.43	0.40	0.40	0.31
MAE	0.19	0.16	0.17	0.14
PBIAS (%)	39.28	17.86	-8.57	-8.64
R^2	0.18	0.29	0.57	0.59
NRMSE	0.43	0.40	0.31	0.31

The analysis suggested that the LSTM model might have had an advantage over the other models in several key metrics, achieving the lowest RMSE and MAE. This implied higher accuracy and fewer predictive errors, which could have been crucial in practical applications. Interestingly, the LSTM model demonstrated minimal bias (PBIAS close to zero), suggesting that it did not systematically overestimate or underestimate flood events, unlike the SVM model, which exhibited a significant positive bias. The apparent higher accuracy of the LSTM model might have been attributed to its capacity to effectively capture and utilize temporal relationships in the precipitation data, an essential feature for accurate flood forecasting. These observations align with existing literature that has noted the proficiency of LSTM networks in handling complex sequential data across various fields, including hydrology [62]. Moreover, it has been reported that LSTMs generally perform better than traditional models in time-series prediction tasks [63], lending support for further validation of the findings of this study.

Despite its strengths in classification tasks, the SVM model showed a strong positive bias, suggesting a propensity to predict floods more frequently than they occurred. Such a tendency might have led to overly conservative forecasts that may not have always aligned with real-world conditions. Conversely, the balanced performance of the LSTM model underscores its potential suitability for integration into flood early warning systems, where both reliability and accuracy are paramount. Improvements in the R^2 values from SVM to LSTM were noted, with the LSTM model achieving the

highest R^2 . This indicates that it could explain a more substantial proportion of the variance in flood occurrences from the dataset. This aspect underscores the practical utility of the model in operational settings, as discussed in related studies [64].

Similarly, ROC curves were employed to evaluate the model's accuracy, which is demonstrated in Figure 10. An AUC of 0.5 indicated no discrimination, and 1.0 indicated perfect prediction. AUC values closer to 1.0 indicate a more dependable and accurate model for bias-free prediction of outcomes. Considering the complexity and variability of weather patterns that lead to floods, the precision of the LSTM model in making predictions might have been beneficial for enhancing flood risk management strategies in the Sacramento area. Such improvements could have mitigated floods' impacts on local communities and infrastructure.

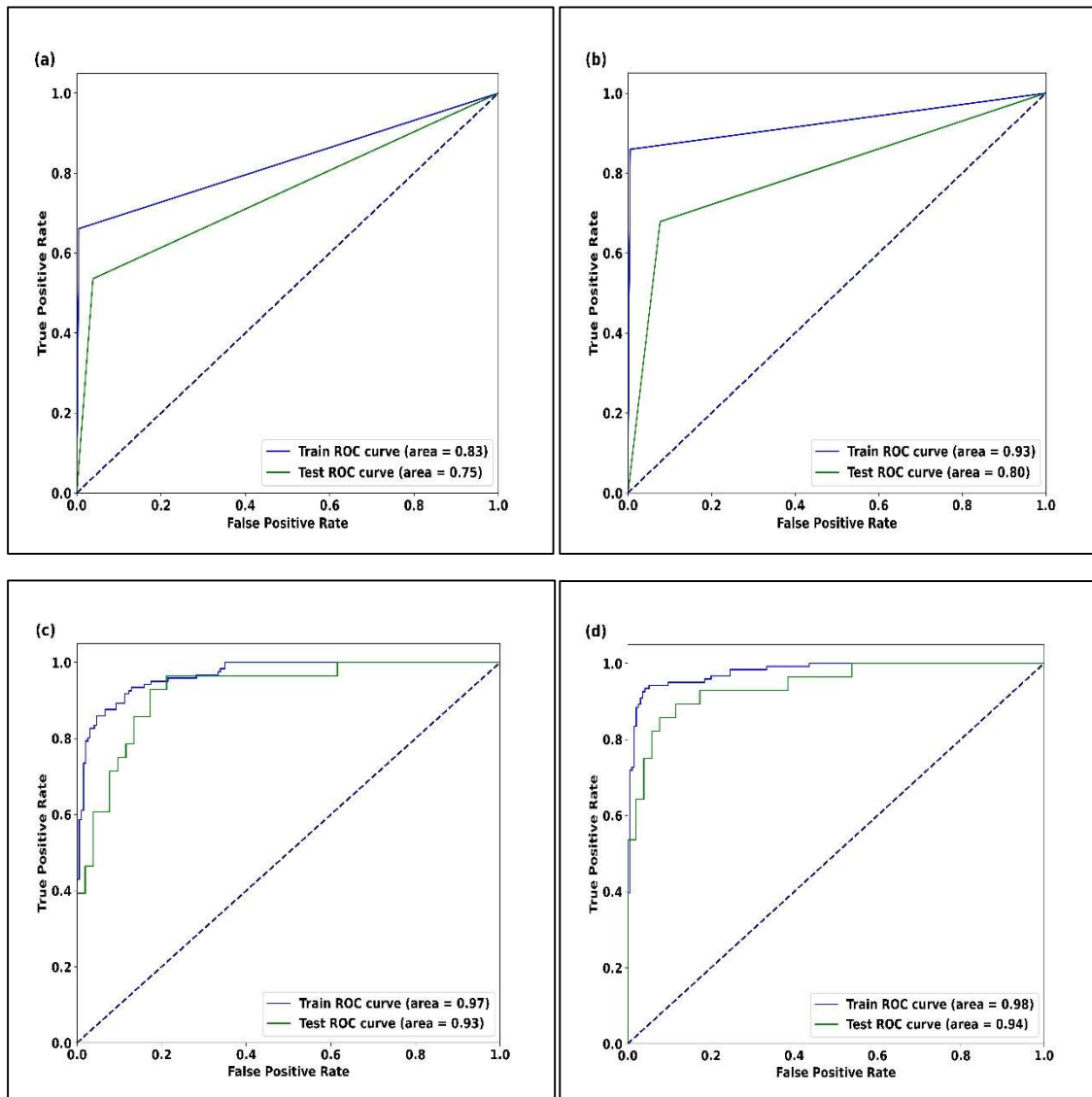


Figure 10. (a) ROC Curve for training and testing set for SVM Model (b) ROC Curve for training and testing set for RF Model (c) ROC Curve for training and testing set for ANN Model (d) ROC Curve for training and testing set for LSTM Model.

The confusion matrices of ANN, RF, ANN, and LSTM show different levels of accuracy in flood prediction models. Matrix (a) accurately detected 50 “No Flood” events and 15 “Flood” events but encountered 13 false negatives and 2 false positives, which demonstrates considerable difficulties in identifying floods. Matrix (b) enhanced flood detection with 48 “No Flood” and 19 “Flood” correct predictions by lowering false negatives to 9 while slightly increasing false positives to 4. Matrix (c) has performed well by correctly recognizing 45 cases of “No Flood” and 23 instances of “Flood,” showing a significant reduction in false negatives to 5 but having 7 false positives. Matrix (d) and (c) have the same confusion matrix result, with identical counts for correct and incorrect classifications. Overall, matrices (c) and (d) demonstrated the best balance and dependability in predictions by having fewer false negatives and a consistent ability to correctly identify both “Flood” and “No Flood” events. The graphical representation of this matrix is shown

in Figure 11. Matrices (a), (b), (c), and (d) correspond to the models ANN, RF, ANN, and LSTM, respectively. Similarly, Table 4 and Table 5 compare the performance of the four different machine-learning models used for flood prediction.

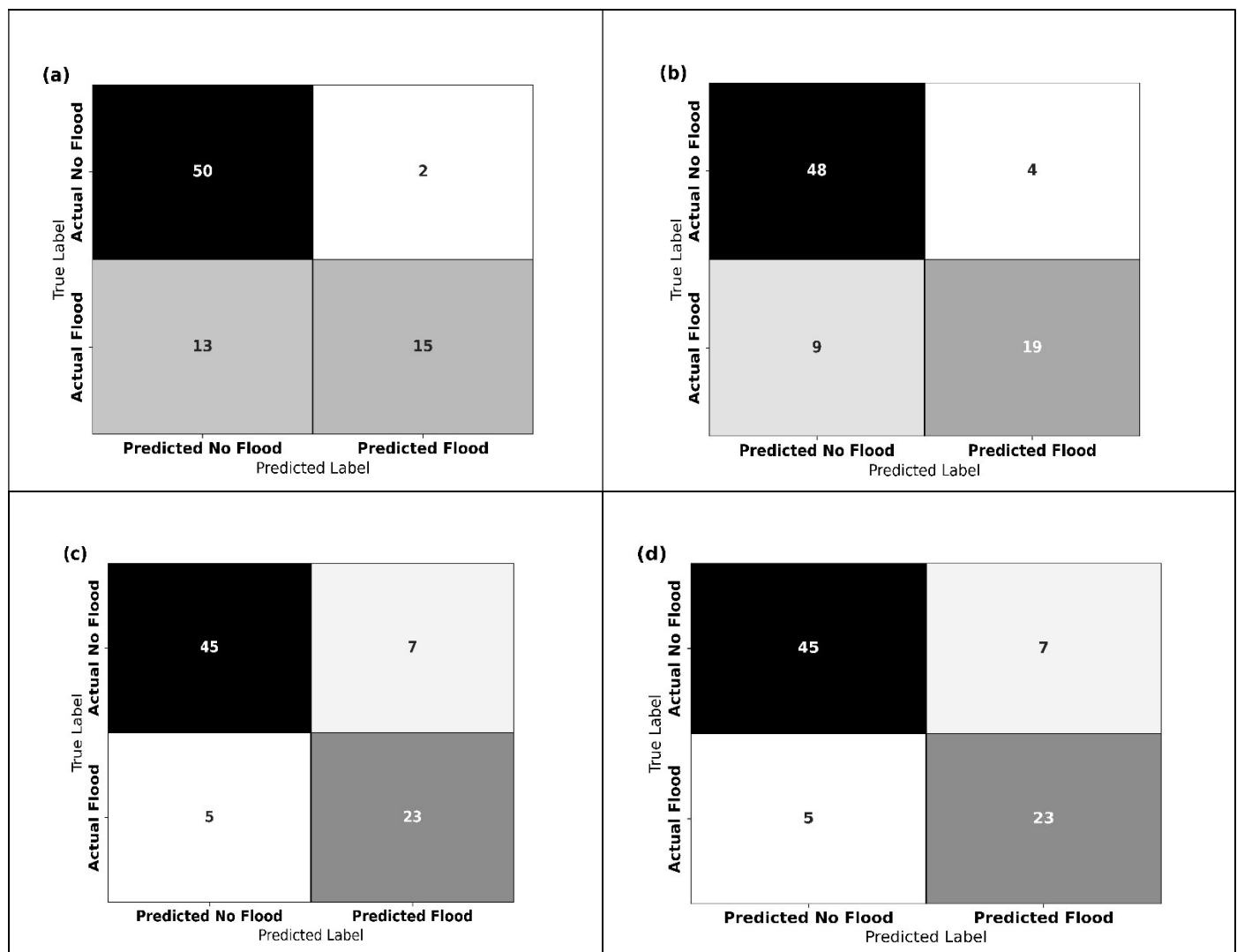


Figure 11. (a) Confusion matrix for SVM model (b) Confusion matrix for RF model (c) Confusion matrix for ANN model (d) Confusion Matrix for LSTM model.

Table 5. Comparison of performance matrix for SVM, RF, ANN, and LSTM model for Flood prediction.

	Recall	Precision	F1-Score	AUC
SVM	0.54	0.88	0.67	0.75
RF	0.68	0.83	0.75	0.80
ANN	0.82	0.77	0.79	0.94
LSTM	0.89	0.83	0.86	0.94

4.5. Related Research

Our results demonstrate that the Long Short-Term Memory (LSTM) model outperformed other machine learning models predicting floods in Sacramento, California. Similar studies in the Kebbi State in Nigeria compared Logistic Regression, Decision Tree, and Support Vector Classification model rainfall-based flood prediction, where Logistic Regression was more accurate among all [29]. Another study in Kerala, India, used rainfall indices to compare LSTM-RNN, Random Forest, SVM, Decision Tree, and Logistic regression models [32]. They found that logistic regression achieved the highest accuracy. Similarly, similar research conducted in Bihar, India, which used rainfall and temperature as the input parameters, showed that the Decision Tree performed better than the gradient boost and random forest algorithms [33]. However, it is crucial to remember that variations may influence model performance between data sets, regional climate variables, and model configurations. These results are consistent with prior research in similar climatic regions to Sacramento, California. Dettinger et al. studied how atmospheric rivers affected California's flood risks and

found that these rivers significantly raise streamflow and precipitation, essential for forecasting floods [65]. According to this study, atmospheric rivers account for 20–50% of the state’s streamflow and precipitation. This highlights the need for models that can handle temporal sequences, like LSTM models, which efficiently capture these nuances [65,66]. Additionally, research conducted in the Pacific Northwest, an area with a climate similar to California’s, by Hsu et al. discovered that LSTM models outperformed Random Forest and Support Vector Machine models regarding flood predictions’ accuracy and dependability [67]. The study showed that the prediction of flood incidents was improved with the ability of LSTM models to handle sequential data.

5. Conclusions and Future Scope

Our research utilized machine learning models to predict flood occurrences in the Sacramento area, focusing on evaluating the performance of SVM, RF, ANN, and LSTM algorithms. The meticulous research found that LSTM models dominated the group with an outstanding precision rate of 89.9%, followed by ANN at 85%, RF at 83.75%, and SVM at 81.25%. Despite specific variations seen during significant flood occurrences, neural network-based models, particularly ANN and LSTM, may have been more capable of recognizing the non-linear correlations within meteorological data. These results indicate that although the LSTM model had demonstrated a promising ability to track accurate data patterns, there was probably still room for improvement, particularly in engaging in severe flood events. The LSTM model may have been more accurate than SVM and RF because of its advanced temporal processing capabilities, which were probably required to handle the complex hydrological data in Sacramento. Additionally, the LSTM model’s reduced PBIAS and lower RMSE and MAE values supported its suitability for risk management and flood prediction applications. It appeared conceivable that integrating LSTM networks into Sacramento’s flood prediction models might have improved the precision and effectiveness of existing systems. This research was consistent with a broader body of literature that acknowledges the importance of advanced computational methods in environmental modeling. This study additionally evaluated the flood risks and how they are affected by climate cycles, such as El Niño and La Niña, by analyzing changes in rainfall patterns and climate variability over an extended period. This allowed us to identify the most critical hydrological patterns. The results of this study may suggest that local governments should adjust their approaches to emergency preparedness and flood response and utilize water management techniques to improve development planning in high-risk areas.

Our study, despite its limitations, has yielded promising results that are of significant importance to the field of climate science and modeling:

1. Our research utilized machine learning models, and the availability of time series data and stations can be inconsistent, as our research used data from 1990. Some stations might not have complete or continuous datasets, which can impact the model’s accuracy and reliability, and any inaccuracies in the data can affect the prediction outcomes.
2. While this article used precipitation, temperature, and soil moisture data in the models, various other parameters, like wind speed and humidity, can also impact the flooding scenario, and they were not considered. Including a broader range of climate variables, topography, and land use data might improve the model’s predictive capabilities.

In the future, researchers could work in the following areas:

1. This article revealed a significant correlation between precipitation, temperature, and flood events, leading to the design of algorithms that consider these factors in future studies. The inclusion of other environmental elements, such as wind patterns, soil properties, and topographical data, could significantly improve flood prediction models, potentially revolutionizing our understanding of flood dynamics.
2. This study used four different individual models. Combining machine learning models like LSTM and ANN may provide a solid foundation for improving predicted accuracy and reliability. This hybrid strategy may better manage the intricate relationships within flood-related data, thereby increasing real-time and long-term flood prediction.
3. Although the study location is a small watershed county in Sacramento County, California, there is a clear need for more comprehensive research in a more dynamic, heterogeneous, and meteorologically unique basin that experiences a relatively high frequency of flooding. Comprehensive knowledge of how regional and global climate trends could influence local weather events may be obtained by investigating comprehensive climate models, improving prediction abilities, and developing more effective mitigation techniques.

Acknowledgments

The authors would like to thank the reviewers for their valuable suggestions. The authors are grateful to the Office of the Vice-Chancellor for Research at Southern Illinois University, Carbondale for providing research support. Publicly available datasets are used for the analysis.

Author Contributions

Conceptualization, A.K. and D.D.; methodology, A.K., D.D. and B.A.M.; software, M.B., B.P., A.A.; validation, A.K. and D.D.; formal analysis, D.D. and A.K. writing—original draft preparation, D.D., B.A.M., M.B., A.A., B.P. and A.K.; writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Ethics Statement

Not applicable.

Informed Consent Statement

Not Applicable

Funding

This research received no external funding.

Declaration of Competing Interest

The authors declare no conflict of interest.

Data Availability Statement

All data used in the study are available on public websites, and the links are provided in the data section of the manuscripts.

References

1. Wang J, Yi S, Li M, Wang L, Song C. Effects of Sea Level Rise, Land Subsidence, Bathymetric Change and Typhoon Tracks on Storm Flooding in the Coastal Areas of Shanghai. *Sci. Total Environ.* **2018**, *621*, 228–234. doi:10.1016/j.scitotenv.2017.11.224.
2. Yadava PK, Kumar H, Singh A, Kumar V, Verma S. Chapter 3—Impact of Climate Change on Water Quality and Its Assessment. In *Visualization Techniques for Climate Change with Machine Learning and Artificial Intelligence*; Srivastav A, Dubey A, Kumar A, Kumar Narang S, Ali Khan M, Eds.; Elsevier: Amsterdam, The Netherlands, 2023; pp 39–54. doi:10.1016/B978-0-323-99714-0.00002-9.
3. Rizzardi KW. Sea Level Lies the Duty to Confront the Deniers. *Stetson L. Rev.* **2014**, *44*, 75.
4. Malekmohammadi B, Rezaei M, Balist J, Andarabi A. Y. 5—Effects of Floods on the Oil, Gas, and Petrochemical Industries: Case Study in Iran. In *Crises in Oil, Gas and Petrochemical Industries*; Rahimpour MR, Omidvar B, Shirazi NA, Makarem MA, Eds.; Elsevier: Amsterdam, The Netherlands, 2023; pp 103–133. doi:10.1016/B978-0-323-95154-8.00014-1.
5. Jevrejeva S, Jackson LP, Grinsted A, Lincke D, Marzeion B. Flood Damage Costs under the Sea Level Rise with Warming of 1.5 °C and 2 °C. *Environ. Res. Lett.* **2018**, *13*, 074014. doi:10.1088/1748-9326/aacc76.
6. Haasnoot M, Winter G, Brown S, Dawson RJ, Ward PJ, Eilander D. Long-Term Sea-Level Rise Necessitates a Commitment to Adaptation: A First Order Assessment. *Clim. Risk Manag.* **2021**, *34*, 100355. doi:10.1016/j.crm.2021.100355.
7. Parker VT, Boyer KE. Sea-Level Rise and Climate Change Impacts on an Urbanized Pacific Coast Estuary. *Wetlands* **2019**, *39*, 1219–1232. doi:10.1007/s13157-017-0980-7.
8. Heberger M, Cooley H, Herrera P, Gleick PH, Moore E. Potential Impacts of Increased Coastal Flooding in California Due to Sea-Level Rise. *Clim. Chang.* **2011**, *109* (Suppl. 1), 229–249. doi:10.1007/s10584-011-0308-1.
9. Cordeira J, Neureuter M, Kelleher L. Atmospheric Rivers and National Weather Service Watches, Warnings, and Advisories Issued Over California 2007–2016. *J. Oper. Meteorol.* **2018**, *6*, 87–94. doi:10.15191/nwajom.2018.0608.
10. Habel S, Fletcher CH, Barbee MM, Fornace KL. Hidden Threat: The Influence of Sea-Level Rise on Coastal Groundwater and the Convergence of Impacts on Municipal Infrastructure. *Annu. Rev. Mar. Sci.* **2024**, *16*, 81–103. doi:10.1146/annurev-marine-020923.

11. Schubert SD, Chang Y, DeAngelis AM, Lim Y-K, Thomas NP, Koster RD, et al. Insights into the Causes and Predictability of the 2022/23 California Flooding. *J. Clim.* **2024**, *37*, 3613–3629. doi:10.1175/JCLI-D-23-0696.1.
12. Burton C, Cutter SL. Levee Failures and Social Vulnerability in the Sacramento-San Joaquin Delta Area, California. *Nat. Hazards Rev.* **2008**, *9*, 136–149. doi:10.1061/ASCE1527-698820089:3136.
13. Wilby RL, Keenan R. Adapting to Flood Risk under Climate Change. In *Progress in Physical Geography*; SAGE Publications Ltd.: Thousand Oaks, CA, USA, 2012, pp 348–378. doi:10.1177/0309133312438908.
14. Alexander D. Disaster and Emergency Planning for Preparedness, Response, and Recovery. In *Oxford Research Encyclopedia of Natural Hazard Science*; Oxford University Press: Oxford, UK, 2015. doi:10.1093/acrefore/9780199389407.013.12.
15. Banjara M, Bhusal A, Ghimire AB, Kalra A. Impact of Land Use and Land Cover Change on Hydrological Processes in Urban Watersheds: Analysis and Forecasting for Flood Risk Management. *Geosciences* **2024**, *14*, 40. doi:10.3390/geosciences14020040.
16. Bichard E, Kazmierczak A. Are Homeowners Willing to Adapt to and Mitigate the Effects of Climate Change? *Clim. Chang.* **2012**, *112*, 633–654. doi:10.1007/s10584-011-0257-8.
17. Wagenaar D, Curran A, Balbi M, Bhardwaj A, Soden R, Hartato E, et al. Invited Perspectives: How Machine Learning Will Change Flood Risk and Impact Assessment. In *Natural Hazards and Earth System Sciences*; Copernicus GmbH: Göttingen, Germany, 2020; pp. 1149–1161. doi:10.5194/nhess-20-1149-2020.
18. Ighile EH, Shirakawa H, Tanikawa H. A Study on the Application of GIS and Machine Learning to Predict Flood Areas in Nigeria. *Sustainability* **2022**, *14*, 5039. doi:10.3390/su14095039.
19. Mcleod E, Poulter B, Hinkel J, Reyes E, Salm R. Sea-Level Rise Impact Models and Environmental Conservation: A Review of Models and Their Applications. *Ocean. Coast. Manag.* **2010**, *53*, 507–517. doi:10.1016/j.ocecoaman.2010.06.009.
20. Hapuarachchi HAP, Wang QJ, Pagano TC. A Review of Advances in Flash Flood Forecasting. *Hydrol. Process* **2011**, *25*, 2771–2784. doi:10.1002/hyp.8040.
21. Liu J, Kuang W, Zhang Z, Xu X, Qin Y, Ning J, et al. Spatiotemporal Characteristics, Patterns, and Causes of Land-Use Changes in China since the Late 1980s. *J. Geogr. Sci.* **2014**, *24*, 195–210. doi:10.1007/s11442-014-1082-6.
22. Wasko C, Nathan R. Influence of Changes in Rainfall and Soil Moisture on Trends in Flooding. *J. Hydrol.* **2019**, *575*, 432–441. doi:10.1016/j.jhydrol.2019.05.054.
23. Fahad MGR, Nazari R, Motamedi MH, Karimi M. A Decision-Making Framework Integrating Fluid and Solid Systems to Assess Resilience of Coastal Communities Experiencing Extreme Storm Events. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108388. doi:10.1016/j.ress.2022.108388.
24. Nazari R, Vasiliadis H, Karimi M, Fahad MGR, Simon S, Zhang T, et al. Hydrodynamic Study of the Impact of Extreme Flooding Events on Wastewater Treatment Plants Considering Total Water Level. *Nat. Hazards Rev.* **2022**, *23*, 04021056. doi:10.1061/(asce)nh.1527-6996.0000531.
25. Wang YY, Wang W, Chau KW, Xu DM, Zang HF, Liu CJ, et al. A New Stable and Interpretable Flood Forecasting Model Combining Multi-Head Attention Mechanism and Multiple Linear Regression. *J. Hydroinform.* **2023**, *25*, 2561–2588. doi:10.2166/hydro.2023.160.
26. Wang Y, Wang W, Zang H, Xu D. Is the LSTM Model Better than RNN for Flood Forecasting Tasks? A Case Study of HuaYuankou Station and LouDe Station in the Lower Yellow River Basin. *Water* **2023**, *15*, 3928. doi:10.3390/w15223928.
27. Wang WC, Zhao YW, Xu DM, Chau KW, Liu CJ. Improved Flood Forecasting Using Geomorphic Unit Hydrograph Based on Spatially Distributed Velocity Field. *J. Hydroinform.* **2021**, *23*, 724–739. doi:10.2166/hydro.2021.135.
28. Wang Y, Wang W, Xu D, Zhao Y, Zang H. A Novel Strategy for Flood Flow Prediction: Integrating Spatio-Temporal Information through a Two-Dimensional Hidden Layer Structure. *J. Hydrol.* **2024**, *638*, 131482. doi:10.1016/j.jhydrol.2024.131482.
29. Lawal ZK, Yassin H, Zakari RY. Flood Prediction Using Machine Learning Models: A Case Study of Kebbi State Nigeria. In Proceedings of the 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Brisbane, Australia, 8–10 December 2021; pp. 1–6.
30. Sarkar A, Kulkarni AM, Khodaskar MR, Tidake SP, Diwate RB. A Predictive Model for Occurrence of Floods Using Machine Learning Techniques. *Res. Mil.* **2023**, *13*, 5054–5072.
31. Bhandari S, Thakur B, Kalra A, Miller WP, Lakshmi V, Pathak P. Streamflow Forecasting Using Singular Value Decomposition and Support Vector Machine for the Upper Rio Grande River Basin. *J. Am. Water Resour. Assoc.* **2019**, *55*, 680–699. doi:10.1111/1752-1688.12733.
32. Vidya S. Rainfall Based Flood Prediction in Kerala Using Machine Learning. *Int. J. Intell. Syst. Appl. Eng.* **2024**, *12*, 23s.
33. Kunverji K, Shah K, Banu Shah N. A Flood Prediction System Developed Using Various Machine Learning Algorithms. Available online: <https://ssrn.com/abstract=3866524> (accessed on 15 March 2024).
34. Cristianini N, Schölkopf B. Support Vector Machines and Kernel Methods: The New Generation of Learning Machines. *Ai Mag.* **2002**, *23*, 31.

35. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine Learning for Identifying Randomized Controlled Trials: An Evaluation and Practitioner’s Guide. In *Research Synthesis Methods*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2018; Volume 9, pp. 602–614. doi:10.1002/jrsm.1287.
36. Ahmad S, Kalra A, Stephen H. Estimating Soil Moisture Using Remote Sensing Data: A Machine Learning Approach. *Adv. Water Resour.* **2010**, *33*, 69–80. doi:10.1016/j.advwatres.2009.10.008.
37. Nanda MA, Seminar KB, Nandika D, Maddu A. A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection. *Information* **2018**, *9*, 5. doi:10.3390/info9010005.
38. Bhusal A, Parajuli U, Regmi S, Kalra A. Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in DuPage River Basin, Illinois. *Hydrology* **2022**, *9*, 117. doi:10.3390/hydrology9070117.
39. Belgiu M, Drăgu L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. doi:10.1016/j.isprsjprs.2016.01.011.
40. Yang GR, Wang XJ. Artificial Neural Networks for Neuroscientists: A Primer. In *Neuron*; Cell Press: Cambridge, MA, USA, 2020, pp. 1048–1070. doi:10.1016/j.neuron.2020.09.005.
41. Koçak Y, Üstündağ Şiray G. New Activation Functions for Single Layer Feedforward Neural Network. *Expert. Syst. Appl.* **2021**, *164*, 113977. doi:10.1016/j.eswa.2020.113977.
42. Dubey SR, Singh SK, Chaudhuri BB. Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *Neurocomputing* **2022**, *503*, 92–108. doi:10.1016/j.neucom.2022.06.111.
43. Shang W, Sohn K, Almeida D, Lee H. Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units. In *Proceedings of the 33rd International Conference on Machine Learning*; Balcan MF, Weinberger KQ, Eds.; PMLR: New York, NY, USA, 2016; Volume 48, pp. 2217–2225.
44. Song X, Liu Y, Xue L, Wang J, Zhang J, Wang J, et al. Time-Series Well Performance Prediction Based on Long Short-Term Memory (LSTM) Neural Network Model. *J. Pet. Sci. Eng.* **2020**, *186*, 106682. doi:10.1016/j.petrol.2019.106682.
45. Landi F, Baraldi L, Cornia M, Cucchiara R. Working Memory Connections for LSTM. *Neural Netw.* **2021**, *144*, 334–341. doi:10.1016/j.neunet.2021.08.030.
46. Gao M, Shi G, Li S. Online Prediction of Ship Behavior with Automatic Identification System Sensor Data Using Bidirectional Long Short-Term Memory Recurrent Neural Network. *Sensors* **2018**, *18*, 4211. doi:10.3390/s18124211.
47. Md AQ, Kapoor S, AV CJ, Sivaraman AK, Tee KF, Sabireen H, et al. Novel Optimization Approach for Stock Price Forecasting Using Multi-Layered Sequential LSTM. *Appl. Soft Comput.* **2023**, *134*, 109830. doi:10.1016/j.asoc.2022.109830.
48. Liu D, Jiang W, Mu L, Wang S. Streamflow Prediction Using Deep Learning Neural Network: Case Study of Yangtze River. *IEEE Access* **2020**, *8*, 90069–90086. doi:10.1109/ACCESS.2020.2993874.
49. Althoff D, Rodrigues LN. Goodness-of-Fit Criteria for Hydrological Models: Model Calibration and Performance Assessment. *J. Hydrol.* **2021**, *600*, 126674. doi:10.1016/j.jhydrol.2021.126674.
50. Naidu G, Zuva T, Sibanda EM. A Review of Evaluation Metrics in Machine Learning Algorithms. In *Artificial Intelligence Application in Networks and Systems*; Silhavy R, Silhavy P, Eds.; Springer International Publishing: Cham, Switzerland, 2023; pp. 15–25.
51. Reda HT, Anwar A, Mahmood A, Chilamkurti N. Data-Driven Approach for State Prediction and Detection of False Data Injection Attacks in Smart Grid. *J. Mod. Power Syst. Clean. Energy* **2023**, *11*, 455–467. doi:10.35833/MPCE.2020.000827.
52. Young AP, Guza RT, Flick RE, O’Reilly WC, Gutierrez R. Rain, Waves, and Short-Term Evolution of Composite Seacliffs in Southern California. *Mar. Geol.* **2009**, *267*, 1–7. doi:10.1016/j.margeo.2009.08.008.
53. Dettinger M, Anderson J, Anderson M, Brown LR, Cayan D, Maurer E. Climate Change and the Delta. *San. Fr. Estuary Watershed Sci.* **2016**, *14*, 1–26. doi:10.15447/SFEWS.2016V14ISS3ART5.
54. LaDochy S, Witiw M. Climate Change in California: Past, Present, and Future. In *Fire and Rain: California’s Changing Weather and Climate*; Springer Nature Switzerland: Cham, Switzerland, 2023; pp. 163–183. doi:10.1007/978-3-031-32273-0_11.
55. Avalon-Cullen C, Caudill C, Newlands NK, Enenkel M. Big Data, Small Island: Earth Observations for Improving Flood and Landslide Risk Assessment in Jamaica. *Geosciences* **2023**, *13*, 64. doi:10.3390/geosciences13030064.
56. Blay ES, Schwabedissen SG, Magnuson TS, Aho KA, Sheridan PP, Lohse KA. Variation in Biological Soil Crust Bacterial Abundance and Diversity as a Function of Climate in Cold Steppe Ecosystems in the Intermountain West, USA. *Microb. Ecol.* **2017**, *74*, 691–700. doi:10.1007/s00248-017-0981-3.
57. Hettiarachchi S, Wasko C, Sharma A. Increase in Flood Risk Resulting from Climate Change in a Developed Urban Watershed—The Role of Storm Temporal Patterns. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 2041–2056. doi:10.5194/hess-22-2041-2018.
58. Seo SB, Jee HW, Cho J, Oh C, Chae Y, Jo S, et al. Assessment of the Flood Damage Reduction Effect of Climate Change Adaptation Policies under Temperature Increase Scenarios. *Mitig. Adapt. Strateg. Glob. Chang.* **2024**, *29*, 8. doi:10.1007/s11027-024-10105-9.
59. Zhou S, Williams AP, Lintner BR, Berg AM, Zhang Y, Keenan TF, et al. Soil Moisture–Atmosphere Feedbacks Mitigate Declining Water Availability in Drylands. *Nat. Clim. Chang.* **2021**, *11*, 38–44. doi:10.1038/s41558-020-00945-z.

60. Baldocchi DD, Keeney N, Rey-Sanchez C, Fisher JB. Atmospheric Humidity Deficits Tell Us How Soil Moisture Deficits Down-Regulate Ecosystem Evaporation. *Adv. Water Resour.* **2022**, *159*, 104100. doi:10.1016/j.advwatres.2021.104100.
61. Kim K, Kim DK, Noh J, Kim M. Stable Forecasting of Environmental Time Series via Long Short Term Memory Recurrent Neural Network. *IEEE Access* **2018**, *6*, 75216–75228. doi:10.1109/ACCESS.2018.2884827.
62. Tripathy KP, Mishra AK. Deep Learning in Hydrology and Water Resources Disciplines: Concepts, Methods, Applications, and Research Directions. *J. Hydrol.* **2024**, *628*, 130458. doi:10.1016/J.JHYDROL.2023.130458.
63. Latif SD, Ahmed AN. Application of Deep Learning Method for Daily Streamflow Time-Series Prediction: A Case Study of the Kowmung River at Cedar Ford, Australia. *Int. J. Sustain. Dev. Plan.* **2021**, *16*, 497–501. doi:10.18280/IJSDP.160310.
64. Rahimzad M, Moghaddam Nia A, Zolfonoon H, Soltani J, Danandeh Mehr A, Kwon HH. Performance Comparison of an LSTM-Based Deep Learning Model versus Conventional Machine Learning Algorithms for Streamflow Forecasting. *Water Resour. Manag.* **2021**, *35*, 4167–4187. doi:10.1007/s11269-021-02937-w.
65. Dettinger MD, Ralph FM, Das T, Neiman PJ, Cayan DR. Atmospheric Rivers, Floods and the Water Resources of California. *Water* **2011**, *3*, 445–478. doi:10.3390/w3020445.
66. Dettinger M. Climate Change, Atmospheric Rivers, and Floods in California—A Multimodel Analysis of Storm Frequency and Magnitude Changes. *J. Am. Water Resour. Assoc.* **2011**, *47*, 514–523. doi:10.1111/j.1752-1688.2011.00546.x.
67. Le XH, Ho HV, Lee G, Jung S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11*, 1387. doi:10.3390/w11071387.