FORUM GEOGRAFI
ISSN: 0852-0682 | E-ISSN: 2460-3945

HTTPS://JOURNALS.UMS.AC.ID/INDEX.PHP/FG/

Research article

Rainfall Prediction Using Machine Learning in Nineveh Governorate, Iraq

Khaled Raed Nafea[1,3*], Ali Mahdi Jawad Al-Dujaili[2]

[1] Al-Iraqia University, College of Arts, Department of Geography, Baghdad, 10054, Iraq; [2] Al-Iraqia University, College of Arts, Department of Geography, Baghdad, 10054, Iraq; [3] Al- Hamdaniya University, Laser and photonics center, Department of sensing and Nanophotonics, Nineveh , 41006, Iraq.

[*)] Correspondence: khaled.r.nafea@aliraqia.edu.iq

Citation:

Nafea, K. R., & Al-Dujaili, A. M. J. (2026). Rainfall Prediction Using Machine Learning in Nineveh Governorate, Iraq. Forum Geografi. 40(3), 367-390.

Article history:
Received: 30 March 2026
Revised: 14 June 2026
Accepted: 20 June 2026
Published: 24 July 2026

Abstract

Accurate rainfall forecasting is essential in arid and semi-arid regions such as Iraq due to increasing climate variability, water scarcity, and the growing need for sustainable water resource management. This study proposes an integrated spatiotemporal framework for rainfall forecasting in Nineveh Governorate, Iraq, combining non-parametric trend detection, multivariate correlation analysis, and ensemble machine learning. Monthly climate data from seven stations (1994-2024) were analysed using the Mann-Kendall test and Sen's slope estimator to assess long-term rainfall dynamics. Pearson and Spearman coefficients were applied to examine interrelationships among climatic variables and to mitigate multicollinearity prior to model construction. Four predictive models (XGBoost, CatBoost, Random Forest, and LSTM) were combined through a weighted stacking ensemble to improve robustness and reduce forecasting uncertainty. Results indicate no statistically significant long-term monotonic trend in annual rainfall; however, clear spatial heterogeneity and seasonal variability were observed. The stacking model outperformed the individual models, achieving an R^2 value of 0.981 (98.1%), indicating high predictive performance., with reduced RMSE and MAE values. Findings highlight the dominance role of temperature-humidity-pressure interactions in shaping rainfall behaviour in semi-arid environments. The proposed framework enhances both predictive reliability and structural interpretation of rainfall variability under regional climate stress conditions.

Keywords: Rainfall; Model; Standard Deviation; Machine Learning; Nineveh Governorate.

1. Introduction

Rainfall is considered a fundamental component in supporting water resources, as it represents a renewable natural resource that contributes to the replenishment of both surface water and groundwater on a periodic basis, making it a key factor in the sustainability of freshwater supplies (Al-Hussein *et al.,* 2024). Rain is also the main driver of a number of natural disasters, such as floods and droughts, due to an increase or decrease in its quantity or irregular frequency (Datta *et al.,* 2019) On the other hand, the intensity and frequency of rainfall are important indicators of climate change, particularly at the regional level (Ghosh *et al.,* 2023). Similarly, rainfall plays a crucial role in determining the temporal and spatial distributions of water resources, as rainfall patterns control the distributions of surface water and groundwater (Kõlõc, 2020). Furthermore, the irregularity and variability of rainfall can increase the risk of extreme events, such as floods, necessitating the development of accurate models to analyze these changes and improve rainfall forecasting (Bahrom *et al.,* 2026).

Rainfall forecasting is important because it supports the planning, protection, and development of water resources for irrigation, hydroelectric power generation, domestic supply, and disaster risk reduction, particularly flood mitigation (Kannan *et al.,* 2010; Fereshtehpour *et al.,* 2020; Gu *et al.,* 2022). In regions with unstable rainfall patterns, analyzing the quantitative and temporal variability of rainfall is scientifically important because rainfall directly influences water availability, agricultural activities, ecosystem stability, and climate-related hazards (Wang *et al.,* 2020). Historical rainfall patterns are shaped by multiple atmospheric and climatic parameters, including humidity, temperature, wind conditions, and antecedent rainfall, making rainfall prediction a complex but essential task in climate forecasting studies (Kundu *et al.,* 2023).

Recent studies show a strong movement from conventional statistical approaches toward machine learning, deep learning, and ensemble-based models for rainfall prediction. Machine learning methods are increasingly used because they can capture nonlinear relationships between rainfall and climatic predictors more effectively than many traditional models. Elshaboury *et al.* (2021), for example, demonstrated that artificial neural networks achieved better rainfall prediction performance than traditional statistical models in arid regions. Similarly, Uddin *et al.* (2025) applied linear regression, random forest, and artificial neural networks using climatic variables such as temperature, humidity, wind speed, and previous rainfall, showing improved accuracy through reduced RMSE and MAE values. Nirranjana *et al.* (2025) also reported high predictive accuracy and low error values from machine learning-based rainfall prediction models.

Beyond individual machine learning algorithms, recent research increasingly emphasizes hybrid and ensemble modeling strategies. El Hafyani et al. (2024) showed that a clustering-based machine learning model improved the accuracy of monthly rainfall forecasting and yielded lower RMSE values than individual models. Ghosh et al. (2023) demonstrated that ensemble learning techniques, including boosting and bagging, can reduce bias and variance, leading to better prediction performance and lower error rates. Baig et al. (2024) further showed that advanced models, such as XGBoost and LSTM, outperformed traditional approaches in extremely arid environments, especially when multiple climatic variables were incorporated. Similarly, Jumadi et al. (2025) reported that AI-based ensemble learning approaches combining random forest, XGBoost, and LSTM improved spatiotemporal rainfall forecasting and produced more reliable predictions in complex climatic environments.

Spatially explicit rainfall prediction has also become increasingly important, particularly for arid and semiarid regions where rainfall is unevenly distributed. Manaf et al. (2026) showed that integrating random forest with kriging improved the spatial representation of rainfall variability and outperformed traditional statistical methods in predictive accuracy. This indicates that rainfall forecasting is not only a temporal prediction problem but also a spatial modeling challenge, especially in regions where water availability, agricultural productivity, and hazard exposure vary considerably across space.

This advance is highly relevant to Nineveh Governorate, which is characterized by a semiarid climate, hot and dry summers, and cold, wet winters. The governorate's climatic characteristics are influenced by its geographical and geological position on the Arabian Plate, which affects rainfall distribution and temperature patterns (MileHacker, 2019; Rahi et al., 2019). Water-related challenges in Nineveh have intensified over time due to rainfall variability, natural hazards, and increasing pressure on water resources (Parmar et al., 2017). The region is already experiencing water scarcity, and further deterioration of surface and groundwater resources is expected under continued climatic stress and rising water demand (Voss et al., 2013; Al Hashimi et al., 2024). Therefore, the existing literature highlights the importance of robust rainfall forecasting approaches for supporting water resource management, agricultural planning, and disaster risk reduction in semiarid environments such as Nineveh.

Despite the large number of studies that have dealt with rainfall prediction in Iraq, only a limited number of them have dealt with it in a comprehensive and up-to-date manner Despite these global developments, studies that have addressed rainfall forecasting in Iraq, particularly in Nineveh Governorate, are limited. Hence, this study aims to fill the research gaps in previous studies, and among these studies are the following (Abdaki *et al.,* 2023), long-term forecasting of rainfall in Nineveh Governorate, northern Iraq, on the basis of the Prophet model. The study was based on daily rainfall data from five stations within Nineveh Governorate for the period (1981-2021), and future forecasts were produced up to the year 2030. The study predicted an annual decrease in rainfall amount. In this study, only the rainfall element was considered in the future forecasts. This study (Al-Hashimi *et al.,* 2024) predicted monthly rainfall in northern Iraq, including the Nineveh Governorate, using a bagging-based cluster learning model. The study relied on integrating fundamentals that included random trees, locally weighted learning, and k nearest neighbors.

This study aims to represent the nonlinear behavior of rainfall and improve forecast accuracy without considering the accompanying climatic effects on rainfall, such as temperature, humidity, wind speed and atmospheric pressure. (Al-Ozeer *et al.,* 2020) conducted a study that was not a direct forecasting objective but rather focused on estimating rainfall data in Nineveh to serve as a valid basis for subsequent predictive modeling. In Sulaymaniyah, a semiarid city, a hybrid SARIMA (seasonal autoregressive integrated moving average) artificial neural application (ANN) model achieved superior accuracy (coefficient of determination (R^2) = 0.98) compared with standalone artificial neural networks), highlighting the benefits of integrating statistical and ML techniques for rainfall prediction (Latif *et al.,* 2024). (Tahseen *et al.,* 2025) conducted a study on monthly rainfall forecasting in Erbil, northern Iraq, using a set of machine learning models, including ANN, RF, XGB and SVR, along with multiple linear regression, which was based on monthly climate data for the period of 2004-2023. (Shneishil, 2025) conducted a study with the aim of predicting rainfall amounts in northwestern Iraq up to 2035 via artificial intelligence techniques and statistical modeling, with a focus on the random forest model.

The novelty of this study lies in its integrated spatiotemporal modeling framework for rainfall analysis and prediction in a semi-arid region of Iraq. Unlike previous studies that mainly emphasized predictive accuracy or applied isolated statistical trend tests, this study combines long-term rainfall trend detection, spatial variability assessment, climate-variable interdependency analysis, and ensemble machine learning within a single analytical workflow. The framework first evaluates annual, seasonal, and monthly rainfall trends using robust non-parametric methods, then examines the correlation structure and multicollinearity among climatic variables before model development, and finally applies a weighted stacking ensemble to improve predictive stability and reduce model uncertainty. This approach advances rainfall forecasting from a purely accuracy-oriented exercise toward a more interpretable framework that explains how different climatic elements contribute to precipitation variability under semi-arid climatic stress.

This study aims to develop an integrated spatiotemporal rainfall modeling framework for Nineveh Governorate, Iraq, by combining statistical trend analysis, spatial rainfall variability assessment, multivariate climate-variable evaluation, and weighted stacking ensemble machine learning. Through this framework, the study provides both predictive and explanatory insights into rainfall dynamics in a semi-arid environment facing increasing climatic and water-resource stress.

2. Methods

2.1. Overview of the Study Area

Nineveh Province is located between latitudes 34° 52′ 25″ – 37° 06′ 3.5″ N and longitudes 41° 10′ 21″ – 43° 42′ 50″ E and has an area of 35,152 km2. It is bordered to the north by Duhok Governorate, east by Erbil and Kirkuk Governorates, south by Salah al-Din and Al-Anbar Governorates, and west by the Syrian Arab Republic, as shown in Figure 1 (Mohammed *et al.,* 2022). Iraq is located within the arid and semiarid climatic zone and is characterized by the succession of four climatic seasons: summer (July–October), hot and dry; autumn (October–November); winter (December–February), temperate and cold; and finally, rainy (rain). Approximately 90% of the total annual rainfall is concentrated during the winter and spring (Salman *et al.,* 2019). As shown in Table 1, the locations of the selected stations in the study area and their altitudes above sea level vary from station to station due to the topography of the study area.

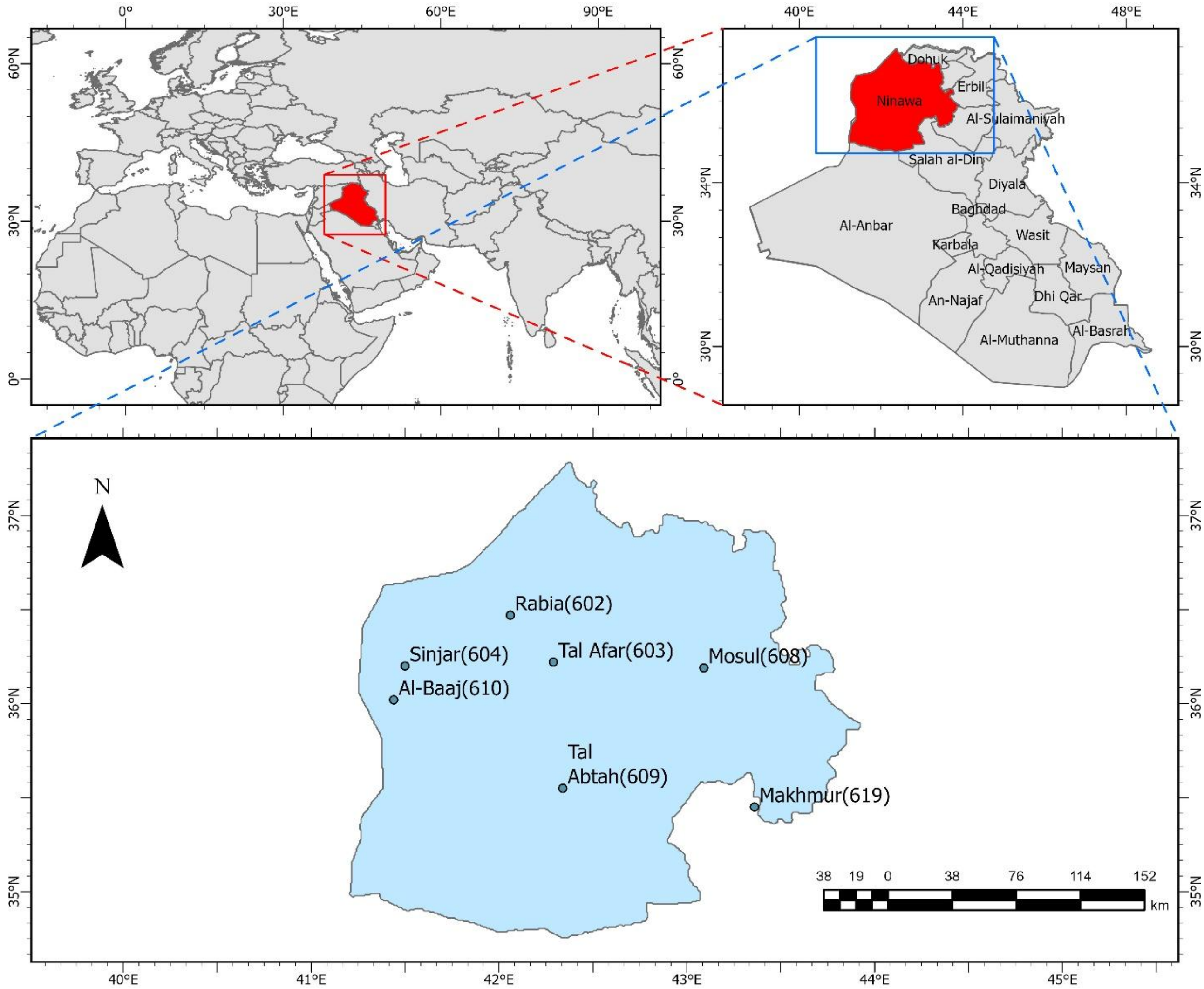

Figure 1. Location Map of the Study Area Showing Iraq within a Global Context and the Spatial Distribution of the Selected Climatic Observation Stations.

Table 1. Details of the Rain Stations within the Study Area.

Station name	Station ID	Elevation (m)	Latitude	Longitude
Mosul	608	223	36.19	43.09
Sinjar	604	583	36.2	41.5
Rabia	602	382	36.47	42.06
Al-Baaj	610	321	36.02	41.44
Tal Afar	603	373	36.22	42.29
Tal Abtah	609	200	35.55	42.34
Makhmur	619	270	35.45	43.36

Source: Republic of Iraq, Ministry of Transport. Iraqi Meteorological Organization/Climate Department, unpublished data.

Topographic features are among the most important natural factors affecting the climate of the study area. The diversity of topographic features, such as mountains, plateaus, hills, and valleys, and their lack of uniformity, create climatic variations from one region to another (Farouk *et al.,* 2005). Nineveh Governorate is characterized by diverse surface features. The topography and elevation above sea level are what control the elements of the climate. The most important topographical feature of the study area is the existence of two regions: the mountainous region and the undulating region and plains.

2.2. Tools and Methods

This study relies on the application of an integrated set of advanced technical and statistical tools to analyse climatic data and to forecast rainfall amounts with a high degree of accuracy, as these tools encompass a range of analytical and predictive modeling techniques.

2.2.1. Climatic Data

The data represent the recorded rainfall amounts over the period from 1994 to 2024 at a set of meteorological stations distributed within the study area. These data were adopted as the basis for the analysis of rainfall characteristics and the construction of predictive models. This period was adopted due to the limited availability and discontinuity of data in earlier years, which could affect the accuracy of the analysis. Accordingly, the relatively complete available records were used as a reliable basis for analyzing rainfall characteristics and developing predictive models. Rainfall is one of the most important forms of condensation in the upper layers of the atmosphere. It depends on the amount of water vapor present and the temperature of the dew point at which condensation occurs. Rainfall is one of the most important forms of atmospheric condensation. Its occurrence depends on the amount of water vapour in the atmosphere and the attainment of the dew point temperature.

Iraq is located within the arid and semi-arid climatic zone, and its rainfall regime is influenced by the Mediterranean rainfall system. Consequently, rainfall is concentrated during the cold season, while the hot season is generally dry (Al-Mousawi *et al.,* 2013). Table 2 shows that rainfall in Nineveh Governorate has a clear seasonal character, with precipitation concentrated from January to April, with a pronounced peak in January, February, and March across all stations (Al-Azzawi, 2017). Rainfall values reach monthly averages of approximately 60-70 mm at some stations, particularly at the Makhmour, Rabia, and Sinjar stations. Rainfall continues during April in moderate amounts and then begins to decrease sharply in May. The region enters a period of near-total drought during the summer (June-August), as the values are almost nonexistent at all stations. With the beginning of autumn (September-October), the rains gradually returned. For the annual total, the highest values were recorded at the Rabia (365.5 mm) and Makhmour (364.6 mm) stations, whereas the lowest values were recorded at Tel Abta (253.2 mm), reflecting spatial variation associated with geographical location and topography.

Table 2. Average Total Rainfall Amounts (mm) for the Study Stations for the Period 1994–2024.

Station	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual total
Mosul	60.9	47.3	57.9	40.8	13.4	0.8	0.2	0	0.6	12.2	38.7	54.5	327.2
Sinjar	71.7	47.3	57.4	37.9	22.3	1	0.1	0	0.5	15	31.7	61.5	346.3
TalAbtah	50.7	39.8	47.5	28.5	14.1	0.1	0.1	0	0.4	8.7	23.6	39.6	253.2
Tal Afar	65.3	46.1	58.8	36.5	19.9	0.5	0.2	0	0.9	10	30.9	49.4	318.5
Makhmur	76.7	57.3	63	42.3	10.8	0	0	0	0	17.9	41.3	55.3	364.6
Rabia	69.9	51.2	64.1	46.3	24.5	0.6	0.1	0	1.3	16.6	32.1	58.9	365.5
Al-Baaj	63.5	42.7	45.7	41.9	20.7	0	0	0	1.3	11.8	24.3	52.1	305.8

Source: Based on the Ministry of Transport, Iraqi Meteorological Organization and Seismological Monitoring, Climate Department, (unpublished data), 2024.

The construction of the predictive model involved incorporating a set of climatic variables that affect rainfall amounts, including as shown in Table 3 temperature, relative humidity, wind speed, and atmospheric pressure, as key explanatory variables in estimating rainfall behavior.

Table 3. Descriptive Statistics of Climate Data for the Period (1994–2024).

Station	Average maximum temperature (°C)	Average minimum temperature (°C)	Average Atmospheric Pressure	AverageRelative Humidity (%)	AverageWind Speed (m/s)
Mosul	28.7	13.9	1012	50	1.6
Sinjar	26.9	13.5	1012	45	1.8
TalAbtah	28.5	13.9	1011	48	1.9
Tal Afar	27.6	13.7	1012	46	1.7
Makhmur	28.5	14.3	1012	48	1.7
Rabia	27.2	13.7	1012	49	1.5
Al-Baaj	27.4	13.6	1013	46	1.9

Source: Based on the Ministry of Transport, Iraqi Meteorological Organization and Seismological Monitoring, Climate Department, (unpublished data), 2024.

2.2.2 Data Analysis Tools

The program was based on Python as the main programming language to utilize the artificial intelligence libraries and data analysis tools (Aurélien Géron, 2019). The pandas library was used for processing and organizing climate data (Reback *et al.,* 2020). Calculations and statistical analysis were also done using the NumPy library (Harris *et al.,* 2020). The predictive model performance was evaluated using the Scikit-learn library (Pedregosa *et al.,* 2011). In addition, the Matplotlib and Seaborn libraries were used to create the graphs, analyse climate patterns, and present the findings in an understandable graphic format (Hunter, 2007).

2.2.3. Data Processing

The first stage involved processing climate data by assessing data quality and filtering out extreme and illogical values, particularly anomalous or negative rainfall values, and addressing missing values to ensure the consistency and accuracy of the time series data (Han *et al.,* 2012). Owing to the prevailing climate in the study area, the summer months (June, July, and August) were excluded from the analysis because they represent a period of near-total drought. The monthly totals for the climatic variables, including temperature, relative humidity, wind speed, and atmospheric pressure, were then calculated. The relationships between independent climate variables and rainfall amounts were examined to identify the most influential variables and reveal patterns of correlation among them, thereby improving the efficiency of the model and enhancing the accuracy of the predictive results (Wilks, 2011).

The first stage involves processing climate data by assessing its quality and filtering out extreme and illogical values, especially anomalous or negative rainfall values, and addressing missing values to ensure the consistency and accuracy of time series data (Han *et al.,* 2012). Owing to the prevailing climate in the study area, the summer months (June, July, and August) were excluded from the analysis because they represent a period of near-total drought. The exclusion of the summer months (June-August) was primarily intended to avoid introducing noise associated with near-zero rainfall values, which could adversely affect model training and reduce predictive stability, particularly in data-driven approaches. The monthly totals for the climatic variables, including temperature, relative humidity, wind speed, and atmospheric pressure, were then calculated. The relationships between independent climate variables and rainfall amounts were examined to identify the most influential variables and reveal the patterns of correlation between them, which contributed to improving the efficiency of the model used and enhancing the accuracy of the predictive results (Wilks, 2011).

2.2.4. Climate Correlation Analysis

To investigate the relationships between rainfall and the associated climatic variables, correlation analysis was performed using both Pearson's and Spearman's correlation coefficients. Pearson's correlation coefficient was applied to evaluate the strength and direction of linear relationships among the climatic variables (Benesty *et al.,* 2009), as expressed in Equation (1). This method was used to identify variables that exhibit strong linear associations with rainfall patterns.

$$r = \frac{\sum Y_i X_i - \frac{(\sum Y_i)(\sum X_i)}{n}}{\sqrt{Y_i^2 - \frac{\sum (Y_i)^2}{n}}\sqrt{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}} \quad (1)$$

where (r) is Pearson's correlation coefficient, (Xi) and (Yi) represent the observed values of the climatic variables, and (n) is the total number of observations.

In addition, Spearman's rank correlation coefficient was employed to assess monotonic relationships that may not necessarily follow a linear pattern (Spearman, 1904), as shown in Equation 2. This approach improves the robustness of the statistical analysis by detecting non-linear associations among climatic variables before their integration into predictive machine learning models.

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} \tag{2}$$

Where (ρ) represents Spearman's rank correlation coefficient, (di) is the difference between the ranks of paired observations, and (n) is the number of observations.

2.3. Analysis and Processing Procedure

This study relied on a set of climatic variables to train and evaluate the models: temperature, relative humidity, wind speed, and atmospheric pressure. Given the correlation between rainfall amounts in the study area and these variables, rainfall data recorded at seven spatially distributed meteorological stations within the study area were utilized (Gu *et al.,* 2022). Approximately 90% of the data from each station were allocated for model training, while the remaining 10% were used to evaluate its predictive efficiency. The study was based on monthly averages of climatic elements. The climatic data spanning 31 years for each station were divided into two independent groups. The first group included training data for the period (1994-2014), whereas the second group included test data for the period (2015-2024) , The period (2015-2024) was used as an independent test phase to evaluate the ability of the trained models to predict observed rainfall before the validated model was applied to future forecasts during the period (2025-2034) (Hastie *et al.,* 2009).

The performance of the predictive models was evaluated on the basis of test data not included in the training process, ensuring the neutrality of the evaluation and the accuracy of measuring the predictive ability of the models. Through the trained models, future forecasts of rainfall amounts for the period from 2025-2034 were obtained (Kuhn *et al.,* 2013). The study methodology is illustrated in Figure 2.

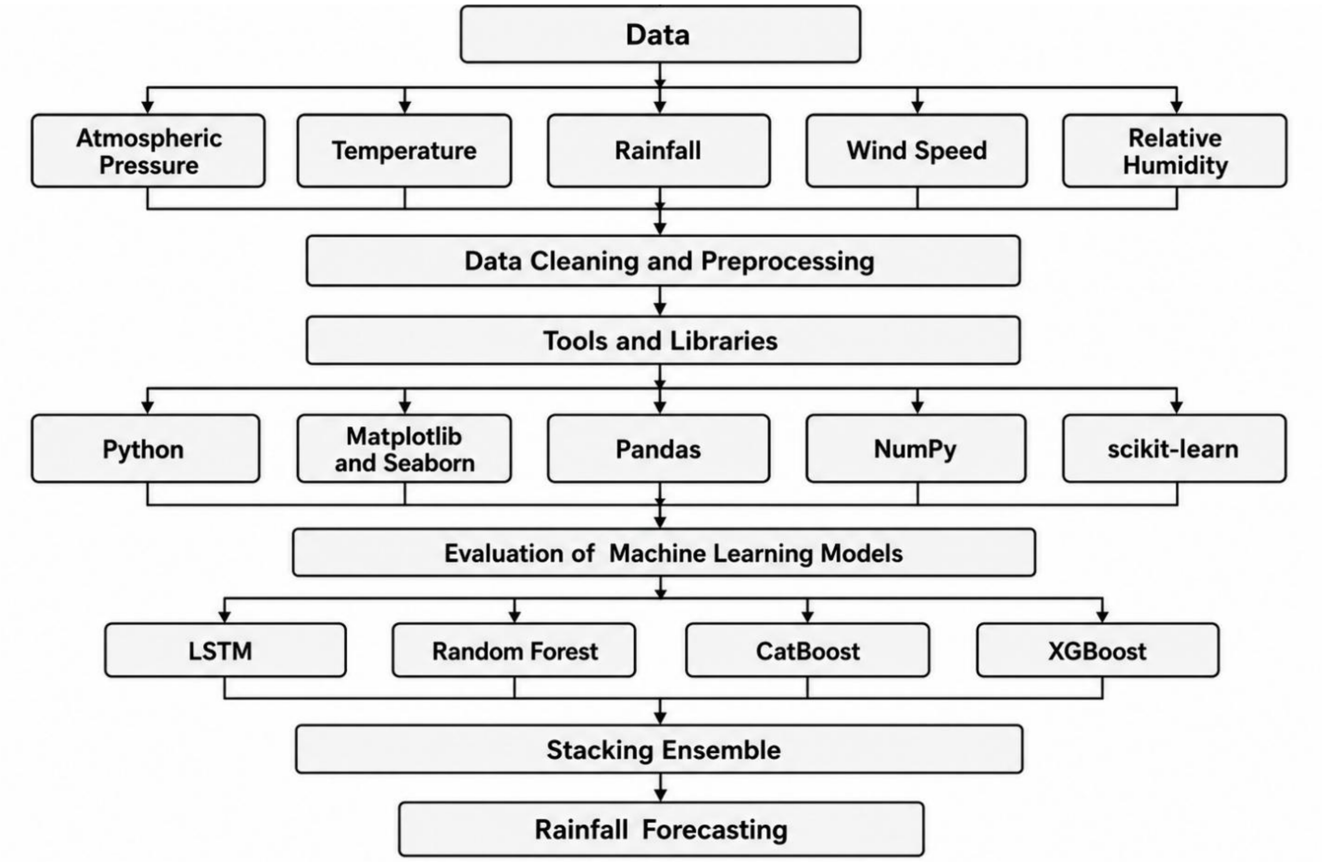

Figure 2. Study Methodology Outline.

2.3.1. Climate Trend Analysis

To detect monotonic long-term trends in rainfall time series (1994–2024), the non-parametric Mann–Kendall (MK) test was applied, as it is widely used in hydro-climatological trend detection studies for its robustness to non-normal data and outliers (Hamed, 2008).

The magnitude of change associated with the detected trend was quantified using Sen's slope estimator, which provides a median-based estimate of the annual rate of change (mm/year) without assuming a specific data distribution (Sen, 1968).

The combined use of the Mann–Kendall test and Sen's slope estimator has become a standard approach in contemporary climate trend assessments, particularly in semi-arid environments characterized by high rainfall variability (Yue and Wang, 2004; Hamed, 2008; Tabari *et al.,* 2018). A significance level of $\alpha = 0.05$ was adopted to determine the statistical significance of detected trends.

2.3.2. Machine Learning Methods

This study was premised on the creation of an extremely precise predictive model to approximate the rainfall quantities in the research site by implementing an integrated approach that incorporates statistical modeling and contemporary machine learning (Khan *et al.,* 2023). The models have had some success in explaining the temporal relationships among various climatic variables and rainfall quantities, which has given them a strong methodological basis to proceed to more sophisticated predictive approaches and improve the fidelity of rainfall forecasting (Ghosh *et al.,* 2023). The methodological framework of the study was extended with the use of advanced machine learning models that were selected as integrative models and incorporated the results of another set of predictive models . These included:

1. XGBOOST

The XGBoost algorithm is an advanced gradient boosting technique based on the Gradient Boosting Machine (GBM) framework. It was introduced by (Chen and Guestrin, 2016) and has demonstrated high performance in supervised learning tasks, including regression, classification, and ranking problems. The algorithm constructs an ensemble of decision trees iteratively, where each new tree minimises the prediction error of the previous trees through gradient optimisation. XGBoost also incorporates regularisation terms to improve model generalisation and reduce overfittingm, making it highly effective for rainfall prediction modeling.

In this study, the XGBoost model was applied to predict rainfall patterns by minimizing the objective function, which consists of a loss function and a regularisation component, as expressed in Equation 3:

$$L(\phi) = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

where $L(\phi)$ represents the objective function, $l\,(y_i, \hat{y}_i)$ denotes the loss function measuring the difference between observed and predicted rainfall values, f_k represents the individual decision trees, $\Omega(f_k)$ is the regularisation term used to control model complexity, and (K) is the total number of trees in the ensemble.

2. CatBoost

The CatBoost algorithm is a modern incremental boosting algorithm that was created by Yandex (Saber *et al.,* 2022). A key feature of this algorithm is that it can directly operate with categorical data, which is a frequent issue in most machine learning problems. Conventional approaches may need elaborate and time consuming preprocessing steps that may implicate data leakage. The CatBoost algorithm on the other hand directly works with categorical variables in its own architecture simplifying the process of data preparation to a large extent. The algorithm is efficient and simple to use since it is highly accurate and resistant even with default parameter settings. CatBoost is very efficient in most machine learning tasks, such as regression, classification, and ranking, thus it is an apt solution to working with complicated nonlinear data in predictions.

3. Random Forest

The Random Forest algorithm, introduced by Breiman (2001), is an effective ensemble machine learning algorithms for regression modeling and classification. This study used the random forest algorithm to predict rainfall amounts. The algorithm relies on creating many decision trees and standardizing their outputs. The final prediction is calculated on the basis of the average of the tree predictions in the case of regression (Sharma *et al.,* 2021), which contributes to improving accuracy and reducing the problem of overfitting. Random Forest is particularly effective in modeling nonlinear relationships between climatic variables and rainfall patterns. It also demonstrates strong performance when handling large datasets and missing values, making it suitable for climatic forecasting applications.

The Gini impurity criterion used in decision tree splitting is expressed in Equation 4.

$$Gini = 1 - \sum_{i=1}^{c} (Pi)^2 \tag{4}$$

where (Pi) represents the probability of class (i), and (c) denotes the total number of classes. Lower Gini values indicate higher node purity during the tree construction process.

4. LSTM Neural Network

Neural networks Advanced types of recurrent neural networks include long Short-Term memory(LSTM) networks. They were devised by Hochreiter and Schmidhuber 1997) to overcome the vanishing gradient issue that conventional models have when applied to long time series. This study paper integrated an LSTM network to predict precipitation levels in time series and model time series data by taking advantage of the capacity to represent time patterns of complex time variations and nonlinear interactions among climatic factors. Past research has shown that LSTM models are very efficient in hydrological and climatological applications, especially in predicting precipitation and surface runoff, as compared to conventional statistical models (Kratzert *et al.,* 2018).

2.3.3. Stacking Ensemble Learning

Wolpert (1992) introduced the stacking strategy for ensemble learning. It is one of the more complex approaches to machine learning that leverages the interdependence among a collection of underlying models to enhance predictive capacity and generalizability. In this study, four base models (XGBoost, CatBoost, random forest, and an LSTM neural network) were used to leverage each model's strengths in capturing nonlinear relationships and temporal dependencies in climate data and, subsequently, enhance the precision of rainfall predictions. The two stages in this model mechanism are dependent. The initial step involves training the four base models separately on climatic element data and extracting each model's predicted value. The second step entails integrating the outputs using a stacked ensemble model to produce the final rainfall forecast. To prevent the issue of overfitting, the meta-model is not trained with the initial outputs of the base models but instead is trained based on a cross-validation strategy through a leave-one-out approach. The results from the cross-validation folds are aggregated to create a new dataset, which is used to train the meta-model; this is why this strategy is known as stacking. Combining basic models is an essential step in creating a stacking model, as various machine learning models can serve as meta-models, including linear regression and random forests. The weighted combination method was adopted in this research. The contribution of each basic model was assigned different weights in the final rainfall prediction and can be mathematically represented by equation 5.

$$\hat{y}p,i = \sum_{m=1}^{M} w_m \, f_{m,i} \tag{5}$$

where ωm (m = 1, 2, …, M) is the weight assigned to each base model and fm,i represents the prediction of model m for the ith observation.

To obtain the optimal final prediction, the set of stacking weights was estimated by minimizing the mean squared linear regression. Thus, the objective function under two constraints is as in Equation 6.

$$\Omega = argmin \sum_{i=1}^{N} (\, yo,i - \sum_{m=1}^{M} W_m \, fm,i)^2 \tag{6}$$

where (Yo,i) is the observed rainfall value for observation (i), and (N) represents the total number of observations .

The optimization process was performed under the following Equations 7 and 8.

$$Wm \geq 0 \qquad m = 1,2 \ldots.., M \tag{7}$$

$$\sum_{m=1}^{M} W_m = 1 \qquad m = 1,2,\ldots\ldots,M \tag{8}$$

where Ω= [W1,W2,…, Mw]. The sum of the weights assigned to the base models is subject to two constraints: the weights must be nonnegative, and their sum must equal one. This leads to a quadratic optimization problem, which was solved via the qpsolvers package in Python. By calculating the weights of the base models, the final rainfall prediction was integrated into the stacking ensemble model.

2.4. Performance Evaluation

The performance of the above mentioned machine learning models was evaluated via the commonly used statistical measures (9) coefficient of determination (R^2), (10) mean absolute error (MAE), and (11) root mean square error (RMSE), as they are among the most commonly used measures in climate studies because of their ability to evaluate the extent to which the models represent the general trend and time variations in rainfall amounts. In addition to measuring the actual deviation between the predicted and observed values, this integration contributes to providing a comprehensive evaluation of the predictive model equation 9-11 used (Kim et al. 2010).

$$R^2 = \frac{\sum_{i=1}^{n} \left(y_i^{obs} - y_{obs}^{mean}\right)\left(y_i^{sim} - y_{sim}^{mean}\right)}{\sqrt{\sum_{i=1}^{n} \left(y_i^{obs} - y_{obs}^{mean}\right)^2 \left(y_i^{sim} - y_{sim}^{mean}\right)^2}} \tag{9}$$

$$MAE = \frac{\sum_{i=1}^{n} \left| y_i^{sim} - y_i^{obs} \right|}{n} \tag{10}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (e_i)^2}{n}} \tag{11}$$

3. Results

3.1. Model performance accuracy

A comparison of the accuracies of the models before and after parameter optimization in Table 4 reveals a clear improvement in the performance of the models used in rainfall prediction. The stacking ensemble (Stacking Ensemble) model achieves the highest accuracy after optimization, reaching 98.10% compared with 95.9% before optimization, which confirms the effectiveness of integrating the basic models in enhancing the predictive capability. The XGBoost and CatBoost models also significantly improved in accuracy after optimization, whereas the accuracy of the LSTM model markedly increased because of its efficiency in representing the time dependence of climate series. In contrast, the random forest model showed a slight decrease in accuracy after optimization. This indicates its sensitivity to parameter adjustment and, in general, confirms that the stacked ensemble model outperforms individual models in achieving more accurate rainfall predictions in the study area.

Table 4. Model Accuracy Before and After the Optimization Process.

Model	Accuracy before optimization%	Accuracy after optimization%
Stacking Ensemble	95.90%	98.10%
XGBoost	93.15%	97.02%
CatBoost	94.05%	97.32%
Random Forest	95.54%	94.14%
LSTM Neural Network	92.80%	96.10%

The performance evaluation results in Table S1 show the superiority of the stacking model over the individual models at all the study stations, as it has the lowest error metric values and the highest coefficient of determination values compared with the CatBoost, Random Forest, XGBoost, and LSTM models when applied individually. This superiority reflects the effectiveness of integrative integration in leveraging the different strengths of each algorithm and improving the predictive capability of the final model (Latif *et al.,* 2023).

In terms of the performance evaluation results, the stacking model achieved the best overall performance compared with the individual models at most stations. The Al-Baaj station recorded the lowest error values (RMSE=0.7 and MAE=0.5) with the highest explanatory power (R^2=0.8). similarly, at Tal Abtah , Rabia, and Makhmur , the stacking model showed notable improvement over the other models, achieving relatively high R^2 values reaching up (0.7,0.9,and 0.9) respectively , with comparatively lower metrics. Advanced results were also achieved at the Tal Abtah, Rabia, and Makhmur stations, with R^2 values of approximately 0.8 and a relative decrease in error measurements. In contrast, the Tal Afar station presented the highest error values and the lowest relative stability of the models, as the RMSE values for the stacking model reached approximately (1.2) and (MAE = 0.8), reflecting the complexity of rainfall behavior in this region. The average performance of the Sinjar and Mosul stations was also recorded, with RMSE values ranging from 0.9 and R^2 values ranging from 0.7.

3.2. Correlation Analysis Results

Table 5 displays the Pearson correlation coefficient of Pearson, which indicates depicts that the linear relationships existing between the climatic elements within the study area are very strong. The amount of rainfall had a strong negative relationship with maximum and average temperatures (r =-0.7) and a moderate negative relationship with minimum temperatures (r =-0.5) which means that higher temperatures are linked to a significant reduction in precipitation. Rainfall on the other hand had a strongly positive association with relative humidity (r = 0.7) and atmospheric pressure (r = 0.6) which is characteristic of the physical nature of the precipitation regime in semi arid conditions where cool and moist air masses are correlated with more rainfall. The outcomes also depicted a negative association that is apparent between rainfall and hours of sunshine (r = -0.6) and the converse of the cloud cover and the sunshine.

Wind speed, however, was strongly correlated with the rainfall (r = 0.1), which means that it is a rather weak factor that affects the precipitation levels directly. Very high correlation coefficients were obtained at the level of interaction of independent variables between Tmax and Tmean (r = 0.99) and between Tmin and Tmean (r = 0.99), between temperatures and hours of sunshine (r = 0.99), and at the same time, a strong negative correlation was observed between heat and humidity (r = -0.9). Such large values suggest that there is a high level of multicollinearity among the variables of the thermal variables, and these variables should be considered when constructing predictive models.

Table 5. Pearson Correlation Coefficient.

	Rain	Tmax	Tmin	Tmean	RH	Wind	Sunshine_hours	Pressure
Rain	1	-0.7	-0.5	-0.7	0.7	-0.1	-0.6	0.6
Tmax	-0.7	1	0.9	1	-0.9	0.3	0.9	-0.9
Tmin	-0.5	0.9	1	0.9	-0.8	0.3	0.8	-0.8
Tmean	-0.7	1	0.9	1	-0.9	0.3	0.9	-0.9
RH	0.7	-0.9	-0.8	-0.9	1	-0.3	-0.9	0.9
Wind	-0.1	0.3	0.3	0.3	-0.3	1	0.4	-0.4
Sunshine_hours	-0.6	0.9	0.8	0.9	-0.9	0.4	1	-0.9
Pressure	0.6	-0.9	-0.8	-0.9	0.9	-0.4	-0.9	1

Spearman's results in Table 6 confirmed the existence of strong, consistent, and even stronger relationships with Pearson's findings, indicating that the relationships are not only linear but also monotonic. Rainfall showed a strong negative correlation with all temperatures (ρ = −0.8). A very strong positive correlation was observed with relative humidity (ρ = 0.9) and atmospheric pressure (ρ = 0.8), reinforcing the hypothesis that humidity and pressure are crucial factors in shaping precipitation behavior. A strong negative correlation was also found between rainfall and hours of sunshine (ρ = −0.8), confirming that periods of high radiation coincide with low precipitation. Wind speed, however, had a relatively weak effect (ρ = −0.2). Spearman's coefficients show nearly perfect values (ρ = 1.0) between Tmax, Tmin and Tmean, indicating near-perfect monotonic homogeneity between the temperature variables, and confirming the existence of a strong intrinsic dependence between them.

In general, Pearson and Spearman's results reveal a strongly interconnected climate system governed by a clear thermal-humidity-pressure triad; precipitation decreases as temperatures rise and solar radiation increases. This increases with rising humidity and atmospheric pressure. The results also confirm a high degree of internal correlation among the thermal variables, necessitating caution when incorporating them into predictive models to avoid multicollinearity. The results also show very high correlations between maximum, minimum, and average temperatures (≈ 1.0), indicating strong linear overlap between them. Therefore, including these variables together in the predictive model may lead to multicollinearity, necessitating the selection of one variable or addressing the relationship between them to ensure model stability and accuracy.

Based on the results of correlation analysis (Pearson and Spearman), the climatic variables included in the forecasting models were systematically selected, taking into account the strength of the relationship with rainfall amounts and trends. The results showed very high correlations between temperature variables (maximum, minimum, and average). This suggests the possibility of multicollinearity among them. Therefore, these variables were carefully managed during model construction, either by selecting a representative variable or by addressing the interrelationships between them, to ensure model stability and improve its explanatory and predictive capabilities.

Table 6. Analysis of Spearman's Correlation Coefficient.

	Rain	Tmax	Tmin	Tmean	RH	Wind	Sunshine_hours	Pressure
Rain	1	-0.8	-0.8	-0.8	0.9	-0.2	-0.8	0.8
Tmax	-0.8	1	1	1	-0.9	0.4	0.9	-0.9
Tmin	-0.8	1	1	1	-0.9	0.4	0.9	-0.9
Tmean	-0.8	1	1	1	-0.9	0.4	0.9	-0.9
RH	0.9	-0.9	-0.9	-0.9	1	-0.4	-0.9	0.9
Wind	-0.2	0.4	0.4	0.4	-0.4	1	0.5	-0.5
Sunshine_hours	-0.8	0.9	0.9	0.9	-0.9	0.5	1	-0.9
Pressure	0.8	-0.9	-0.9	-0.9	0.9	-0.5	-0.9	1

3.3. Annual Trend Results

The outcomes of the Mann-Kendall test on the annual rainfall totals during the period (1994-2024) reveal that there is no statistically significant monotonic trend at a significance level of 0.05 at each of the studied stations, as presented in Table 7. The p-values were much larger than the critical value. This is indicative of the lack of a long-term, systematic change in the amount of precipitation per year in the governorate over the period of study. Even though the trend was not statistically significant, Sen's slope estimator values by Sen varied in signal and magnitude among the stations. The Tel Abtah and Sinjar stations had relative annual growth rates of about 2.8 and 2.76 mm/year, respectively. Whereas the Mosul and Al-Baaj stations showed slight negative tendencies of approximately −1.34 and −1.33 mm/year, respectively. the other stations portrayed a minimal positive tendency of about 1 mm/year.

These values indicate that the observed changes, although spatially different in direction, remain small in magnitude and insufficient to produce a structural shift in the annual rainfall pattern within the time frame studied. This reinforces the hypothesis of relatively stable annual rainfall in the semi-arid environment of Nineveh Governorate, with natural oscillations between years remaining the dominant feature of the time series.

Table 7. Analysis of the Annual Rainfall Trend in the Study Area for the Period (1994–2024).

Station	Number of Years	Trend	p-value	Sen's Slope (mm/year)
Tal Abta	31	increasing	0.248	2.8
Sinjar	31	increasing	0.362	2.762
Mosul	31	decreasing	0.5	-1.344
AL Baaj	31	decreasing	0.518	-1.333
Rabia	31	increasing	0.661	1.074
Makhmur	31	increasing	0.787	0.758
Tal Afar	31	increasing	1	0.252

3.4. Seasonal Trend Results

Table S2 shows the results of the Mann-Kendall test and Sen's slope estimator for seasonal rainfall trends at the study stations during the period (1994–2024), where the results generally show weak temporal trends and spatial instability, as the majority of values did not register statistical significance at the ($\alpha = 0.05$) level. Autumn was characterized by slight, mostly positive, but low values that do not reflect actual climate change, while spring showed relatively higher rates of positive trend at some stations without reaching the significance level Statistically, while the summer season was characterized by an almost complete absence of trend with a slope close to zero at all locations, the winter season saw some stations record decreasing slopes of varying magnitudes, but these were not statistically significant. Overall, the quarterly results reflect a natural annual fluctuation rather than a long-term, regular climatic shift in the rainfall regime of Nineveh Governorate.

3.4 Monthly Trend Results

Table S3 of the monthly analysis shows that statistically significant trends were concentrated to a limited extent in the spring months, particularly April and May at some stations. While the summer months were characterized by an almost zero trend, the winter and autumn months showed mostly inconsistent and insignificant trends, confirming that the observed change reflects a normal seasonal fluctuation rather than a long-term, regular monthly climate shift.

3.5. General Statistical Results of Rainfall

Comparisons of the historical rainfall record with the forecast outcomes of the period (2025 2034) at the Nineveh Governorate stations revealed that the overall seasonal trend of the rainfall was always the same, unlike the fact that there were no clear qualitative changes in monthly means and annual totals as indicated in Tables 2 and 8, respectively. In effect, the rainy season starts still in the autumn (October) and reaches its peak in winter (January and February) before slowly subsiding in the spring (March and April) and therefore comes to an end anyway in the first week of May. Conversely, the historical and the forecasting data have almost no rainfall in the summer (June to September), which proves the continued dry summer climate in the area of study.

Table 8. Monthly Average Expected Rainfall for the Period 2025–2034 in the Study Area.

Climatic stations	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual Total (mm)
Al-Baaj	55.7	40.1	41.1	40.9	23.6	0	0	0	1.2	16.2	20.3	52.9	291.9
Tal Abtah	59.8	56	38.8	45.6	25	0	0	0	1.4	20.8	34.1	54.7	336.1
Tal Afar	52.6	69.9	55.7	54.8	59.2	0	0	0	3.9	7.5	42.7	46.5	392.7
Rabia	68.1	86.2	70.3	77.1	48.6	0	0	0	1.3	15.2	44.9	52.6	464.5
Sinjar	77.9	80.6	58.3	69.4	49.9	0	0	0	1.9	10	11.5	34.7	394.1
Makhmur	49.9	41.9	53.2	47	19.3	0	0	0	0	16.2	34.1	55.6	317.2
Mosul	53.9	48.5	51	37.9	19.9	0	0	0	0	4.3	21.8	49	286.4

Source: Based on the results of the stacking rainfall forecasting model for the period (2025-2034) in Nineveh Governorate.

On a monthly basis, the forecast results indicate a relative increase in winter rainfall at a number of stations compared with historical values, particularly at the Rabia, Sinjar, Tal Afar, and Tal Abtah stations, which reflects a possible increase in the intensity of the winter season in the future. A slight extension of spring rains was also observed during April and May at some stations, which may indicate a limited lengthening of the rainy season compared with historical records. In contrast, other stations, such as Mosul and Al-Baaj, showed stability or a slight decrease in autumn rainfall, reflecting greater fluctuations at the beginning of the rainy season.

Comparing Figure 3 and the annual rainfall values between the predicted and historical data, one can find that a spatial difference exists between the stations. The values at Rabia, Sinjar, Tal Afar, and Tal Abtah were significantly higher than to the historical values. The Rabia station annual total increased from the historical 366 mm to a forecast value of more than 460 mm and Sinjar increased from the historical 346 mm to a forecast 394 mm, while Tal Afar increased from 319 mm to a forecast 393 mm. This increase is a relative positive variation of the future rainfall resources in the north and the western areas of the governorate. Other stations, such as Mosul, Makhmur, and Baaj, however, exhibited different trends with the stations being relatively stable or recording a decline in the total rainfall. At Mosul station, the annual total decreased to a projected under 290 mm /a by a traditional decrease of approximately 327 mm/a. At Makhmur station, the reduction was recorded at 365 to 317 mm Al-Baaj station was also close to being stationary with a slight lossThe difference indicates the dissimilarity in the spatial response to the future climate changes in the governorate. Having spatial comparisons meant that chances of increase in rainfall in the future are high in the northern and northern part of the Nineveh Governorate as compared to central and the south part which is likely to remain at a constant level or relatively lower amount of rainfall. The relevance of such a trend is the increasing spatial and temporal variability of rainfall that may directly apply to agricultural planning and water resource management in the study area.

Overall, the forecasting outcomes of the period (2025-2034) show that the rainfall regime in Nineveh Governorate will preserve the primary seasonal features of the onset and the conclusion of the rainy season, where quantitative changes of the monthly averages and annual totals are not identical in all of the stations. The findings indicate the relevance of integrating the use of sophisticated predictive models to aid in the comprehension of the future climate change and offer a concrete level of scientific foundation to inform decision making in areas of agriculture and water management.

Table 9 shows the results of the linear regression for rainfall forecasting at the Nineveh governorate, which reveals that the Mosul station records a weak downward trend of −1.97 mm/year with $R^2 = 0.93$ and $P = 0.422$, whereas the Baaj station shows almost complete rainfall stability with a slight increase of 0.44 mm/year with $R^2 = 0.7$ and $P = 0.818$. The Rabia station also records a limited upward trend of 2.23 mm/year, with $R^2 = 0.71$ and $P = 0.416$. The Tal Afar station shows an increase of 2.34 mm/year, with $R^2 = 0.82$ and $P = 0.494$, whereas the values at the Sinjar station increase to 3.20 mm/year, with $R^2 = 0.64$ and $P = 0.348$. The Makhmur station records an increase of 3.29 mm/year, with $R^2 = 0.85$ and $P = 0.240$. The Tel Abtah station recorded the highest annual change rate of 4.39 mm/year ($R^2 = 0.91$), which was close to statistical significance ($P = 0.087$), indicating that future rainfall changes in Nineveh Governorate are dominated by climatic fluctuations, with no significant long-term trend.

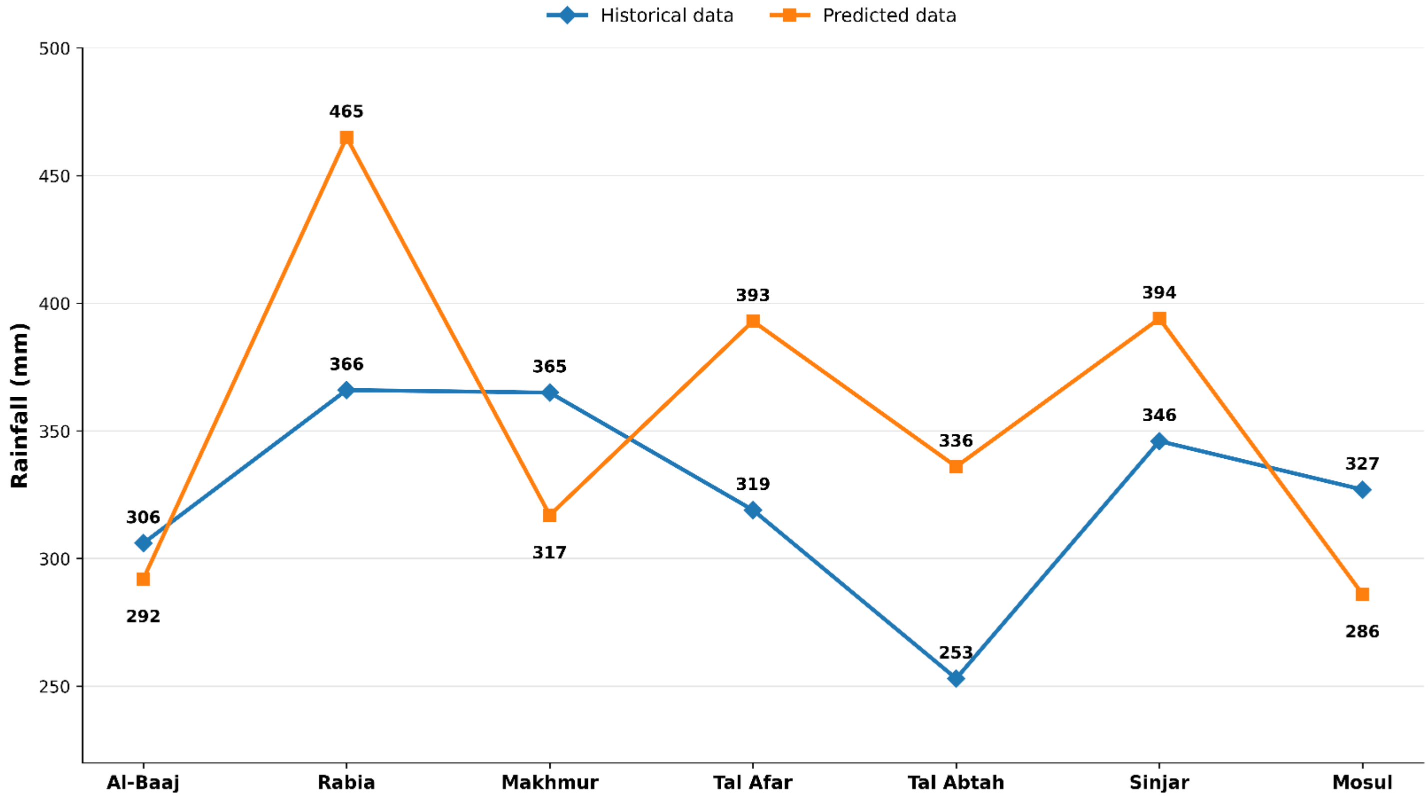

Figure 3. Comparison Between the Annual Totals of the Study Stations Between the Historical Data and the Predicted Data.

Table 9. Simple Linear Regression and Coefficient of Determination (R^2) Value of (P) for Stations in the Study Area.

Station	Slope	R^2	P value
Mosul	-1.97	0.93	0.422
Baaj	0.44	0.7	0.818
Rabia	2.23	0.71	0.416
Tal Afar	2.34	0.82	0.494
Sinjar	3.2	0.64	0.348
Makhmour	3.29	0.85	0.24
Tel Abta	4.39	0.91	0.087

Source: Based on the results of the stacking rainfall forecasting model for the period (2025-2034) in Nineveh Governorate.

3.6. Spatial Variability of Rainfall According to Standard Deviation

Table 10 shows the standard deviation values for the expected rainfall during the period (2025-2034). This value indicates an average spatial heterogeneity in rainfall volumes at stations within the Nineveh governorate, with values between 21.6 mm and 27.3 mm. This comparatively small area depicts a certain level of spatial homogeneity in the temporal variation pattern of precipitation. The differences in space exist and are associated with the geographical position of stations and topography of the area. The highest standard deviations were recorded at our stations Rabia (27.3 mm) and Sinjar (26.9 mm), which show that the north and northwest of the study area are more significantly exposed to the action of Mediterranean low-pressure systems and to variations in their routes and intensity. Conversely, expectant stability in the expected level of rainfall in the western and southern regions of the study area is observed by the fact that the standard deviation values of the Tal Afar (21.6 mm), the Al-Baaj (21.7 mm), and Makhmour (21.7 mm) stations are the lowest.

Table 10. Standard Deviation Values for Rainfall at Each Study Station During the 2025–2034 Period.

Station	SD (2025_2034)
Al-Baaj	21.7
Tal Abtah	22.6
Tal Afar	21.6
Rabia	27.3
Sinjar	26.9
Makhmur	21.7
Mosul	22.8

Source: Based on the results of the stacking rainfall forecasting model for the period (2025-2034) in Nineveh Governorate.

However, the Tel Abtah (22.6 mm) and Mosul (22.8 mm) stations recorded average values of standard deviation, indicating a gradual shift in rainfall characteristics between the high-variability areas in the north and the more stable study area in the south.

In general, these results confirm that the spatial variation in rainfall in Nineveh Governorate will remain relatively limited in the future, with clear spatial differences emerging that reflect the role of geographical location and influential weather systems, which is consistent with the results of the analysis of annual and monthly rainfall totals at stations in the study area (Al Azawi, 2017).

Figure 4 shows the annual changes in the standard deviation coefficient of rainfall at all the study stations for the period (2025–2034), where the highest values were recorded during the years 2026 and 2029, especially at the Mosul, Tal Afar, Makhmour and Sinjar stations, indicating high variability in rainfall during those years. In contrast, other years, such as 2028 and 2031, presented relatively lower values, reflecting a degree of relative stability in rainfall, although spatial variation between stations continued.

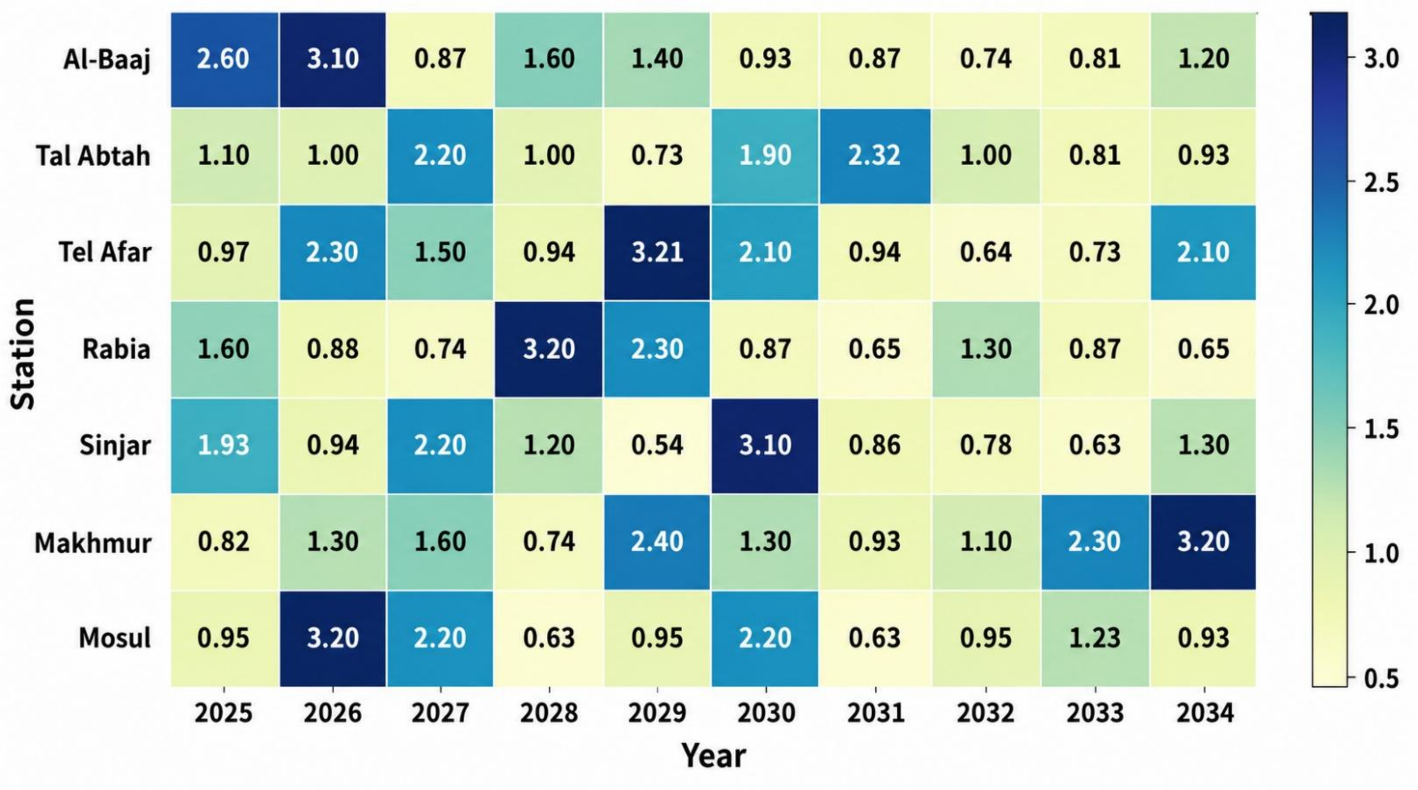

Figure 4. Annual Changes in the Standard Deviation Coefficient of Rainfall in the Study Area from 2025–2034.

Figure 5 shows monthly changes in the standard deviation coefficient of rainfall at the study stations, with the highest values being recorded during the winter months, which are the months of January and February, especially at the Tal Afar, Mosul, Al-Baaj, and Tal Abtah stations, meaning that the rainfall is highly variable during the rainy seasons. Conversely, rainfall levels were significantly reduced at all study sites, especially in July, indicating drought conditions and near-nonuse of rainfall throughout summer (June-August). The average rainfall levels in the transitional seasons, i.e. spring and autumn, indicate low variations between relative stability and variability of rainfall.

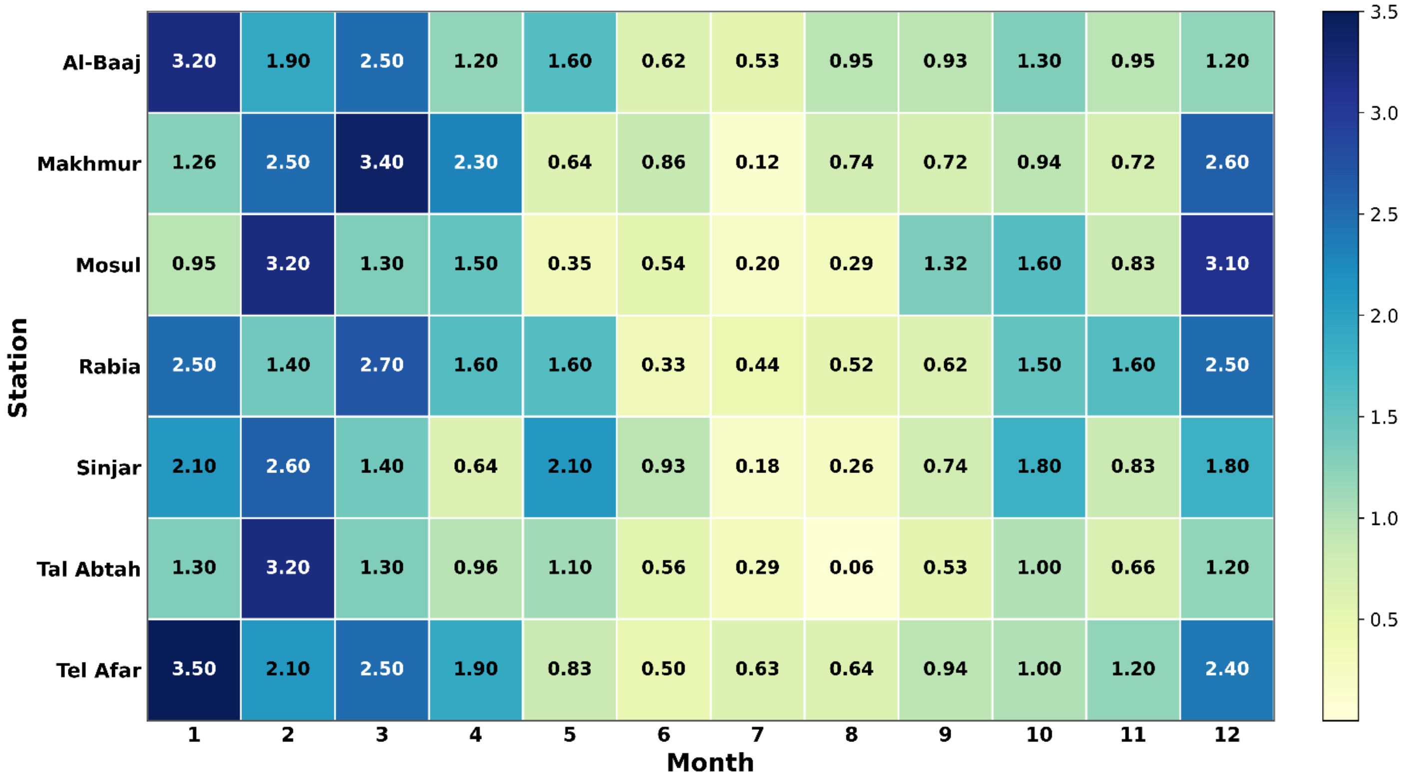

Figure 5. Seasonal Changes in Monthly Rainfall for Each Study Station for the Period 2025–2034.

These spatial variability patterns also reflect the inherent uncertainty associated with rainfall prediction in semi-arid environments. The variability between stations, combined with the sensitivity of individual machine learning models to climatic fluctuations, indicates that model uncertainty cannot be neglected. However, the implementation of the weighted stacking ensemble approach helped reduce predictive uncertainty by integrating the strengths of individual models and minimizing their independent errors. This integration enhances the robustness and stability of the forecasting framework, providing more reliable projections under spatially heterogeneous climatic conditions.

3.7. Dynamics of Spatial Rainfall Distributions Across Four Climatic Periods

The spatial distribution maps of rainfall in the study area Figure 6 illustrate significant spatial and temporal differences in rainfall amounts among the four climatic cycles, reflecting the widespread influence of climatic factors and environmental changes in shaping rainfall patterns at the study area level.

For example, during the period of 1994–2003, the highest rainfall amounts were concentrated in the northern and northwestern parts of the governorate, particularly in the areas of Rabia, Sinjar and Al-Baaj, where high values exceed (330–360 mm annually) prevailed, whereas the southern and southeastern regions, such as Tel Abtah and Makhmour, recorded relatively lower amounts, which indicates a decreasing rainfall gradient towards the south. During the period of 2004–2013, a relative decline in the intensity of rainfall concentration was observed, with the continued dominance of average values (280–320 mm) over most parts of the governorate, in contrast to the decrease in areas with high rainfall and their confinement to limited areas in the north, which reflects the beginning of a change in the general rainfall pattern.

In the period of 2014–2024, clearer features of rainfall irregularity emerged as the ranges of medium and low rainfall expanded, and the areas with high values decreased, with the relative concentration of rainfall remaining in the north and northwest, especially in Sinjar and Rabia, compared with a noticeable decrease in the central and southern regions. This trend is expected to persist in the period 2025-2034, according to the forecast map, as evidenced by the dominance of average and low rainfall across most of Nineveh Governorate, with regions experiencing heavy rainfall decreasing and concentrated in thin strips in the northern governorate. When the spatial distribution of rainfall in previous climatic periods Figure 6 is compared with the expected distribution for the period (2025-2034) Figure 7, it becomes clear that the general spatial pattern of precipitation continues in Nineveh Governorate, where the northern and northwestern regions retain the most abundant rainfall. However, the predictive results indicate quantitative changes in rainfall intensity, characterized by a shrinking range of high values and an expansion of medium and low values, particularly in the central and southern regions.

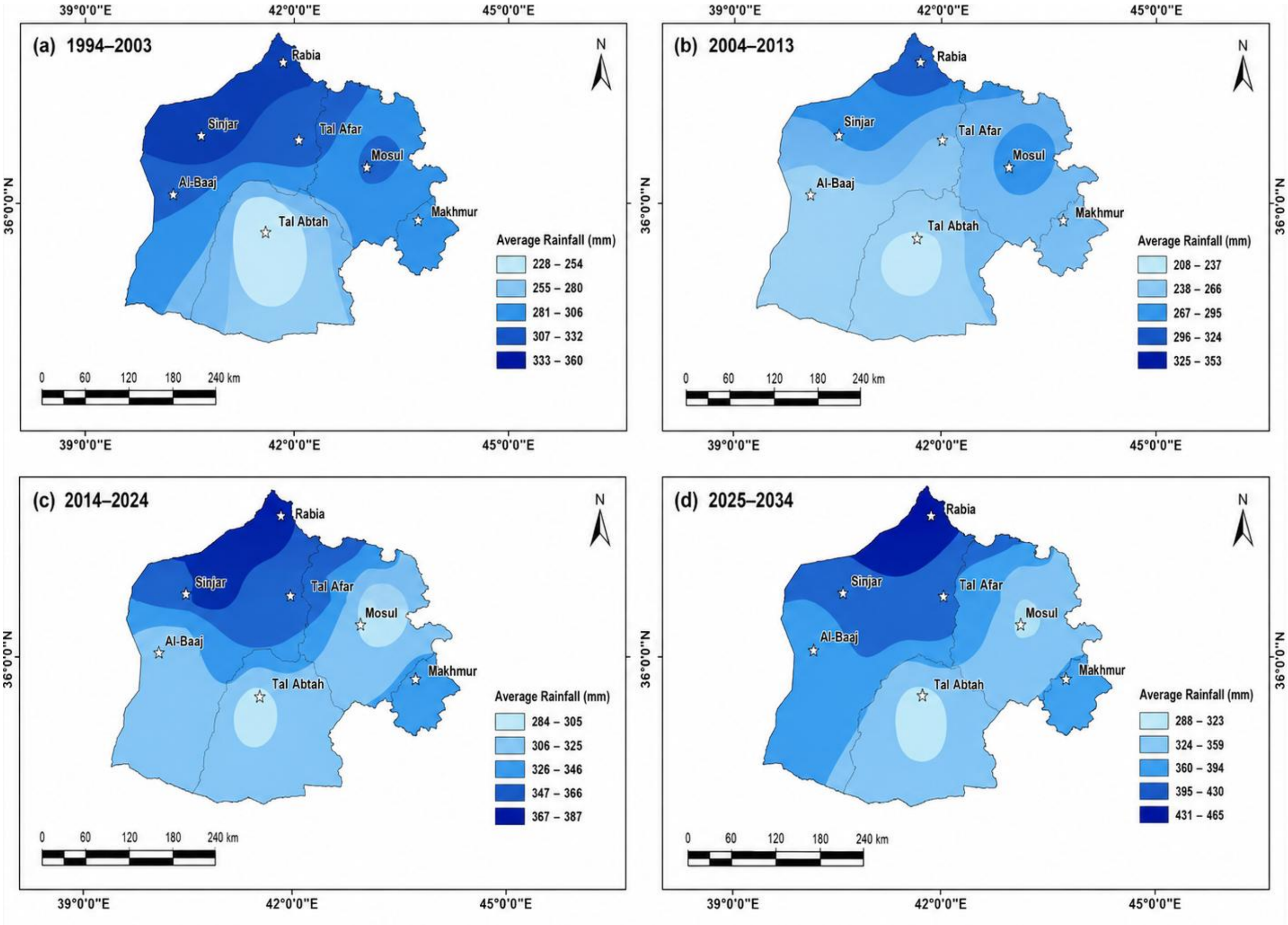

Figure 6. Geographical Distribution of Rainfall in the Study Area.

The spatial gradient from north to south continues, albeit with a relative decrease in intensity, reflecting increasing annual climatic variability. These data suggest that future changes represent an extension of the historical distribution with a relative redistribution of values, influenced by prevailing climatic factors. Overall, the results reflect a trend toward greater spatial variability and a decline in areas with high rainfall, reinforcing the prevalence of semiarid conditions and increasing pressure on water resources and agricultural stability.

Such outcomes suggest the reinforcement of arid and semiarid environmental features, since the significant increase in rainfall variability underscores the growing importance of climate change in reorganizing the spatial distribution of rainfall in the study area. Overall, rainfall trends across the climatic periods showed a downward shift, and variation and heterogeneity in spatial distribution were evident across the periods of investigation, as manifested by a gradual shift toward lower rainfall and a gradual reduction in high-rainfall areas. This rainfall pattern also demonstrates structural climatic change in the dominant hydroclimatic system, which implies further pressure on surface and groundwater resources and an increasing threat to the sustainability of agriculture and food security in Nineveh Governorate over the decades (Zelenakova *et al.,* 2022).

The combined results of the Mann–Kendall trend analysis and the correlation assessments (Pearson and Spearman) reveal that rainfall variability in Nineveh Governorate during the period (1994–2024) is primarily governed by interannual climatic oscillations rather than statistically robust long-term monotonic trends. Although certain stations exhibited weak positive or negative Sen's slope values, none reached a level sufficient to indicate structural climatic shifts.

The correlation structure indicates a tightly coupled thermo–humidity–pressure system, in which rainfall is strongly and negatively associated with temperature and solar radiation, and positively associated with relative humidity and atmospheric pressure. The near-perfect intercorrelation among thermal variables confirms multicollinearity, which was carefully addressed during model construction to ensure predictive stability.

Importantly, the absence of statistically significant long-term rainfall trends, combined with strong internal climatic coupling, suggests that rainfall forecasting performance depends more on dynamic inter-variable interactions than on persistent directional change. Therefore, the proposed modeling framework integrates statistically validated trend detection, robust dependence structure assessment, and cautious variable selection to enhance interpretability, reduce structural uncertainty, and improve forecasting reliability under semi-arid climatic conditions.

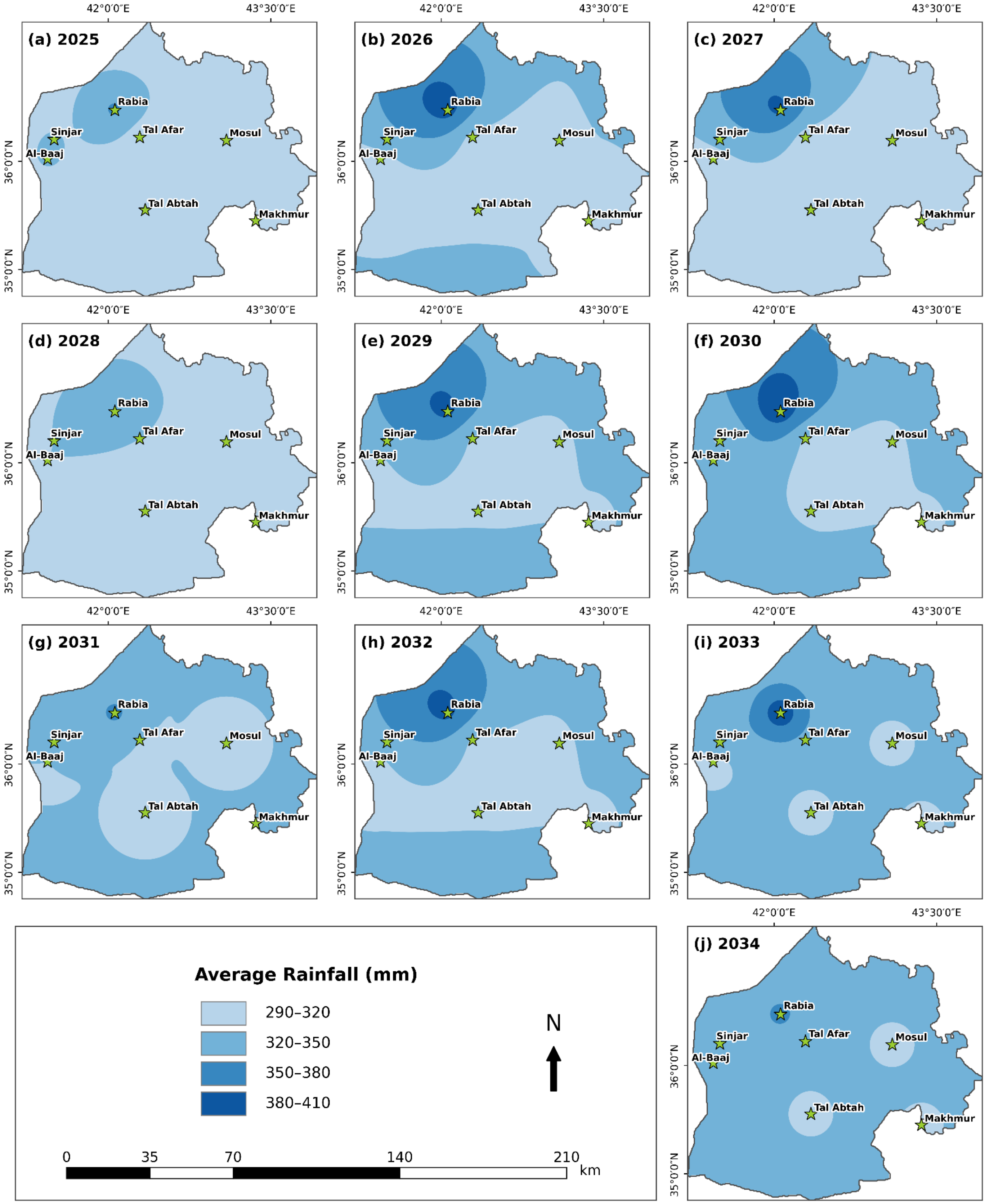

Figure 7. Spatial Distribution of Predicted Rainfall in Nineveh Governorate (2025–2034).

4. Discussion

The results of this study indicate that rainfall behavior in the Nineveh governorate is governed primarily by the dominance of annual interannual variability rather than by a statistically significant long-term trend. The Mann–Kendall and Sen estimators revealed that the observed changes, although spatially variable, remain limited in intensity and do not constitute a structural shift in the rainfall pattern. This is attributed to the prevailing climatic conditions in semiarid regions, which are characterized by high interannual variability and a lack of long-term climatic stability. However, the absence of a statistical trend does not indicate a stable climate system; rather, it reflects the system's complex dynamics. Correlation analysis revealed an interconnected climate system governed by clear relationships among temperature, humidity, and pressure. Rainfall was strongly inversely correlated with temperature and solar radiation, whereas it was directly correlated with relative humidity and atmospheric pressure. These results confirm that precipitation behavior in semiarid environments is not determined by a simple temporal trend but rather by the internal structure of relationships between climatic variables, which supports the findings of recent studies such as (Elshaboury *et al.,* 2021).

Methodologically, the modeling results clearly demonstrated the superiority of the stacking ensemble model over individual models, achieving the highest coefficient of determination and the lowest error values. This reflects the superior ability of stacking models to represent nonlinear relationships and reduce random variance, which aligns with the findings of (El Hafyani *et al.,* 2024) regarding the effectiveness of stacking techniques in improving rainfall forecast accuracy. These results also corroborate the findings of (Manaf *et al.,* 2026), which showed that integrating statistical models with machine learning techniques contributes to improving the representation of spatial rainfall variability.

However, what distinguishes this study from others is its adoption of an integrated framework that combines climate trend analysis, understanding the interrelationships among variables, and the application of advanced machine learning models within a single, comprehensive model. Most previous studies focused either solely on prediction or on statistical analysis in isolation from modeling, whereas this study offers a systematic link between interpreting climate behavior and improving predictive performance, thus enhancing the scientific and practical value of the findings.

At the spatial level, a comparison between previous climatic periods and the projected results for the period 2025-2034 revealed clear continuity in the general pattern of rainfall distribution, with the northern and northwestern regions continuing to record the highest values due to the influence of Mediterranean low-pressure systems and topographical factors. However, the results indicate a spatial redistribution of rainfall values, characterized by a narrowing of the high-value range and an expansion of the average and low-value ranges, particularly in the central and southern regions.

This shift does not reflect a radical change in spatial patterns but rather indicates increased spatial heterogeneity and climate variability. In contrast to some studies that assume an overall decrease in rainfall, this study's results show that future change takes the form of a redistribution of values rather than an absolute decrease, reinforcing the importance of detailed spatial analysis for understanding the impacts of climate change. Furthermore, the decline in areas with high rainfall and the expansion of areas with low rainfall reflect a gradual trend toward strengthening semiarid characteristics in the region, which could increase pressure on water resources and exacerbate the risks associated with drought and agricultural instability. This necessitates adopting adaptation strategies based on a thorough understanding of the spatial and temporal variability of rainfall.

Overall, this study confirms that the accuracy of rainfall prediction in semiarid environments depends not on the existence of a clear climatic trend but rather on the ability to represent the complex interactions between climatic variables. Furthermore, integrating statistical analysis with machine learning techniques within an integrated framework provides a powerful tool for understanding rainfall dynamics and improving the reliability of future forecasts, thus supporting decision-making in water resource management and agricultural planning.

5. Conclusions

This study presented an integrated spatiotemporal framework for analyzing rainfall dynamics in Nineveh Governorate during the period (1994–2024) with future projections for the period (2025-2034), through the integration of climate trend analysis, statistical relationship analysis between climate variables, the building of multiple predictive models, and the comparison of their performance. The results of the Mann-Kendall test and Sen's slope estimator showed the absence of a statistically significant, uniform climatic trend in annual rainfall amounts across most stations, suggesting that the observed changes over the study period reflect interannual variability rather than a long-term structural climatic shift. At the seasonal and monthly levels, the trends were characterized by weakness and spatial instability, with values remaining within the normal fluctuation range for semi-arid environments. In contrast, correlation analysis (Pearson and Spearman) revealed an interconnected climate structure governed by clear thermo-hygroscopic-pressure dynamics; rainfall was strongly negatively correlated with temperature and sunshine hours, and strongly positively correlated with relative humidity and atmospheric pressure. Thermal variables also showed very high internal correlation. This underscores the importance of addressing multicollinearity when building predictive models.

The results obtained from the spatial distribution maps of rainfall confirm that future changes in rainfall in Nineveh Governorate will not be spatially uniform, as the relative improvement in rainfall amounts is concentrated within the northern and northwestern parts. In contrast to the widening range of average and low values in the central and southern regions, reflecting increasing aridity and semiarid conditions, comparisons of model accuracy before and after optimization revealed that ensemble learning models such as XGBoost and CatBoost achieved significant performance improvements after adjustment because of their ability to capture nonlinear relationships between climatic variables. However, the stacking ensemble model consistently outperformed the other models. Before optimization, it reached 95.9%; after optimization, it reached 98.1%, outperforming all other models. As shown by the quantitative performance evaluation results, the lowest RMSE and MAE values and the highest interpretation coefficients (R^2) were recorded at most stations. This confirms that integrating the outputs of basic models provides a more stable and reliable representation of future rainfall behavior. Accordingly, the study recommends relying on ensemble models, particularly the stacking ensemble. In practical terms, the Stacking Ensemble model has demonstrated a significant advantage over individual models in terms of accuracy indicators, showing that combining models contributes to reducing random error and enhancing predictive stability in complex climate systems. The importance of this advantage is underscored by the absence of a clear climate trend. The success of the prediction then depends on the model's ability to capture nonlinear patterns and interactions between variables.

Therefore, the scientific value of this study lies not only in improving predictive accuracy, but also in providing a comprehensive structural understanding of the rainfall system in a semi-arid environment subject to regional climate change pressures. Furthermore, the proposed analytical framework provides a scientific basis for supporting water resource management decisions and agricultural planning.

Acknowledgements

The authors would like to express their gratitude to mentors, data providers, and institutions for their guidance and support in completing this manuscript.

Author Contributions

Conceptualization: Nafea, K. R., & Al-Dujaili, A. M. J.; **methodology**: Nafea, K. R., & Al-Dujaili, A. M. J.; **investigation**: Nafea, K. R., & Al-Dujaili, A. M. J.; **writing—original draft preparation**: Nafea, K. R., & Al-Dujaili, A. M. J.; **writing—review and editing**: Nafea, K. R., & Al-Dujaili, A. M. J.; **visualization**: Nafea, K. R., & Al-Dujaili, A. M. J.. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

All authors declare that they have no conflicts of interest.

Data availability

Data is available upon Request.

Funding

This research received no external funding.

References

Abdaki, M. A., Al-Azzar, A. Z., Al-Obaidi, O., Al-Tiyawi, A. N. (2023). Predicting Rainfall in Nineveh Governorate in Northern Iraq using a Machine Learning Time-series Forecasting Algorithm. *Arab J Geosci*, 16, 11779. doi: 10.1007/s12517-023-11779-2.

Al-Azzawi, A. A. (2017). Analysis of the Spatial and Temporal Variability of Rainfall in Iraq using GIS and Remote Sensing Techniques

Al-Hashimi, M. M., Hayawi, H. A. A. (2024). Nonlinear Model for Precipitation Forecasting in Northern Iraq using Machine Learning Algorithms. *Int J Math Comput Sci*, 19(1), 171–179.

Al-Hussein, A. A. M., Khalil, S. A., Salman F. H., Mahmood, B. F. (2024). *Evaluation of accuracy for satellites rainfall datasets compared in ground stations: a case study of Duhok governorate, Northern Iraq.* Sustain Water Resour Manag. doi: 10.1007/s40899-024-01158-4

Al-Mousawi, A. S, Abu Rahil, A. H. M. (2013). [Title of the book], 1st edn. Al-Mizan Printing Press, Najaf Al-Ashraf.

Al-Ozeer, A. Z., Abdaki, M. A., Al-Iraqi, A. R., Al-Samman, S. H., Al-Hammadi, N. A. (2020). Estimation of mean areal rainfall and missing data by using GIS in Nineveh, northern Iraq. *Iraqi Geol J,* 53(1E), 93–103. doi: 10.46717/igj.53.1E.7Ry-2020-07.07

Géron, A. (2019). *Hands-on Machine Learning with Scikit-learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd edn. O'Reilly Media*. Retrieved From https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632

Bahrom, S., Abu, N., Abd Rahman, N. H., Zainal, N. A. (2026*). Forecasting of Rainfall in Malaysia using Time Series Analysis. Environ Sci Pollut Res*. Retrieved From https://doi.org/10.1007/s11356-026-37668-1

Baig, F., Ali, L., Faiz, M. A., Chen, H., Sherif, M. (2024). How Accurate are the Machine Learning Models in Improving Monthly Rainfall Prediction in Hyperarid Environment?. *J Hydrol*, 633, 131040. doi: 10.1016/j.jhydrol.2024.131040

Benesty, J., Chen, J., Huang, Y., Cohen, I. (2009). *Pearson correlation coefficient.* In: Noise reduction in speech processing. Springer.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.doi.org/10.1023/A:1010933404324

Chen T, Guestrin C (2016) Xgboost: a Scalable Tree Boosting System. *In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining,* 785–794. doi: 10.1145/2939672.2939785.

Datta, P., Das, S. (2019). Analysis of Long-term Precipitation Changes in West Bengal, India: an Approach to Detect monotonic Trends Influenced by Autocorrelations. *Dyn Atmos Oceans,* 88, 101118. doi: 10.1016/j.dynatmoce.2019.101118

El Hafyani, M, E. l., Himdi, K., E,. l., Adlouni, S. E. (2024). Improving Monthly Precipitation Prediction Accuracy using Machine Learning Models: a Multi-view Stacking Learning Technique. *Front Water,* 6, 1378598. doi: 10.3389/frwa.2024.1378598

Elshaboury, N., Abdelkader, E., Attia, A., Marzouk, M. (2021). Rainfall Forecasting in Arid Regions using an Ensemble of Artificial Neural Networks. *J Phys Conf Ser,* 1900, 012015. doi: 10.1088/1742-6596/1900/1/012015

Farouk, I. E., Ashour, M. A. (2005). *Tourism Geography: Development, Foundations, Approaches, and Applications.* Anglo-Egyptian Bookshop, Cairo

Fereshtehpour, M, Sabbaghian, R. J., Farrokhi, A., Jovein, E. B., Sarindizaj, E. E. (2020). Evaluation of Factors Governing the use of Floating Solar System: A Study on Iran's Important Water Infrastructures. *Renew Energy,* 171, 1171–1187. doi: 10.1016/j.renene.2020.12.005

Ghosh, S, Gourisaria, M. K., Sahoo, B., Das, H. (2023). A Pragmatic Ensemble Learning Approach for Rainfall Prediction. *Discov Internet Things,* 3l, 13. doi: 10.1007/s43926-023-00044-3

Gu, J., Liu, S., Zhou, Z., Chalov, S. R., Zhuang, Q. (2022). A stacking Ensemble Learning Model for Monthly Rainfall Prediction. *Water,* 14(3), 492. doi: 10.3390/w14030492

Hamed, K. H. (2008). Trend Detection in Hydrologic Data: the Mann–Kendall Trend Test Under the Scaling Hypothesis. *J Hydrol,* 349(3–4), 350–363. doi: 10.1016/j.jhydrol.2007.11.009

Han, J., Kamber, M., Pei, J. (2012). *Data Mining: Concepts and Techniques, 3rd edn. Elsevier.* Retrieved From https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P. (2020). Array programming with NumPy. *Nature,* 585, 357–362. doi: 10.1038/s41586-020-2649-2

Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer.* Retrieved From https://doi.org/10.1007/978-0-387-84858-7

Hochreiter and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Hunter, J. D. (2007). Matplotlib: a 2D Graphics Environment. *Comput Sci Eng,* 9(3), 90–95.

Jumadi, J., Danardono, D., Roziaty, E., Ulinuha, A., Supari, S., Choy, L. K., Sattar, F., Nawaz, M. (2025). AI-driven ensemble learning for spatio-temporal rainfall prediction in the Bengawan Solo River watershed, Indonesia. *Sustainability,* 17(20),9281. doi: 10.3390/su17209281

Kannan, S., Ghosh, S. (2010). Prediction of Daily Rainfall State in a River Basin using Statistical Downscaling from GCM Output. *Stoch Environ Res Risk Assess,* 25(4), 457–474. doi: 10.1007/s00477-010-0415-y

Kendall, M. G. (1975). *Rank correlation methods, 4th edn.* Charles Griffin, London

Khan, M. M. H., Mustafa, U., Shams. S., Hossain, M. S. (2023). Short-term and Long-term Rainfall Forecasting Using ARIMA Model. *Int J Environ Sci Dev,* 14(5), 292–298. doi: 10.18178/ijesd.2023.14.5.1447

Kim, J. W., Pachepsky, Y. A. (2010). Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *J Hydrol,* 394(3–4), 305–314. doi: 10.1016/j.jhydrol.2010.09.005

Kõlõc, Z. (2020). The Importance of Water and Conscious Use of Water. *Int J Hydrol,* 4(5), 239–241. doi: 10.15406/ijh.2020.04.00250

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M. (2018) Rainfall–Runoff Modeling Using Long Short-term Memory (LSTM) Networks. *Hydrol Earth Syst Sci,* 22, 6005–6022. doi: 10.5194/hess-22-6005-2018

Kuhn, M., Johnson, K. (2013). Applied Predictive Modeling. *Springer,* 26, 13. doi: 10.1007/978-1-4614-6849-3

Kundu, S., Biswas, S. K., Tripathi, D., Karmakar, R., Majumdar, S., Mandal, S. (2023). Forecasting Using Ensemble Learning Techniques. *E-Prime Adv Electr Eng Electron Energy,* 6, 100296. doi: 10.1016/j.prime.2023.100296

Latif, S. H., Koo, N. A. B., Ng, J. L., Chaplot, B., Huang, Y. F., El-Shafie, A., Ahmed, A. N. (2023). Assessing Rainfall Prediction Models: Exploring the Advantages of Machine Learning and Remote Sensing Approaches. *Alex Eng J,* 82, 16–25. doi : 10.1016/j.aej.2023.09.060

Manaf, M., Ali, Z., Scholz, M. (2026). Integrating Random Forest-Based Regression Kriging for Analyzing Spatial Variability of Rainfall in Arid and Semi-arid Regions. *Sci Rep,* 16, 5298. doi: 10.1038/s41598-026-36074-4.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica,* 13(3), 245–259. doi: 10.2307/1907187

MileHacker. (2019). *Best Time to Visit Mosul: Weather, Seasons, and Climate*. Retrieved From https://www.milehacker.com/travel/iraq/nineveh/mosul/best-time-to-visit-mosul-weather-seasons-climate/

Musa, M, Atiyah, K. S. (2022) Extract the effective rain in Nineveh Governorate. *J Tikrit Univ Humanit,* 29(5), 117–150. doi: 10.25130/jtuh.29.5.2022.07

Nirranjana, R., Aishwarya, R., Tejshree, S., Gayathri, K. S., Natarajan, S., Thirunavukkarasu, P. (2025) Rainfall forecasting model for Amaravathi basin using machine learning approach. *J Inst Eng India Ser A,* 106(4), 1067–1080. doi: 10.1007/s40030-025-00914-9

Parmar, A., Mistree, K., Sompura, M. (2017). *Machine learning techniques for rainfall prediction: a review. In: International conference on innovations in information embedded and communication systems.* Retrieved From https://doi.org/10.1109/ICIIECS.2017.8276023

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. (2011). Scikit-learn: machine learning in Python. J *Mach Learn Res,* 12, 2825–2830.

Rahi A, Al-Madhhachi AAT, Al-Hussaini SN (2019) Assessment of surface water resources of eastern Iraq. *Hydrology,* 6(3), 57. doi: 10.3390/hydrology6030057

Reback, J., McKinney, W. (2020). *Pandas: Python data analysis library.* Retrieved From https://pandas.pydata.org/docs/

Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K. I., Kantoush, S. A., Sumi, T., Boutaghane, H., Nohara, D., Mabrouk E. (2022). Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. *Geocarto Int,* 37(25), 7462–7487. doi: 10.1080/10106049.2021.1974959

Salman, S. A., Shahid, S., Ismail, T., Ahmed, K., Chung, E. S., Wang, X. J. (2019). Characteristics of annual and seasonal trends of rainfall and temperature in Iraq. Asia-Pac J Atmos Sci 55:429–438. https://doi.org/10.1007/s13143-018-0073-4

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc,* 63(324), 1379–1389. doi: 10.1080/01621459.1968.10480934

Sharma, A. K., Ahmad, R., & Chen, R. (2021). *Predicting the default borrowers in P2P platform using machine learning models. In: Artificial intelligence and sustainable computing for smart city, 267–281.* Springer International Publishing

Shneishil, B. S. (2025). *Rainfall prediction using statistical modeling and artificial intelligence in arid and semiarid environments of northwestern Iraq: an applied case study.* Ramah J Res Stud

Spearman, C. (1904). The proof and measurement of association between two things. *Am J Psychol,* 15(1), 72–101. doi: 10.2307/1412159

Tabari, H., AghaKouchak, A., Willems, P. (2018). A perturbation approach for assessing trends in precipitation extremes across Europe. J Hydrol 519, 2019–2033. doi: 10.1016/j.jhydrol.2014.08.047

Tahseen, D., Ali, R. (2025). Advancing an innovative machine learning model for monthly rainfall forecasting in a hot-summer Mediterranean climate region, Erbil, Iraq. *Iraqi Geol J,* 58(2D), 1–20. doi: 10.46717/igj.2025.58.2D.1

Uddin, M. T. M., Rahman, M. F., Rahib, M. A., Arun, N., Venkatesh, J. (2025). Rainfall prediction using machine learning. *Int J Res Anal Rev,* 12,(1).

Voss, K. A., Famiglietti, J. S., Lo, M., De, Linage, C., Rodell, M., Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris–Euphrates–Western Iran region. Water Resour Res 49(2), 904–914. doi: 10.1002/wrcr.20078

Wang, Y., Xu, Y., Tabari, H., Wang, J., Wang, Q., Song, S., Hu, Z. (2020). Innovative Trend Analysis of Annual and Seasonal Rainfall in the Yangtze River Delta, Eastern China. Atmos Res, 231, 104673. doi: 10.1016/j.atmosres.2019.104673

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences, 3rd edn. Academic Press*. Retrieved From https://doi.org/10.1016/B978-0-12-385022-5.00021-X

Wolpert, D. H. (1992). Stacked generalization. *Neural Netw,* 5(2), 241–259. doi: 10.1016/S0893-6080(05)80023-1

Yue, S., Wang, C. Y. (2004). The Mann–Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resour Manag,* 18, 201–218. doi: 10.1023/B:WARM.0000043140.61082.60

Zelenakova, M., Abd-Elhamid, H., Krajnikova, K., Smetankova, J., Purcz, P. (2022). Spatial and temporal variability of rainfall trends in response to climate change: a case study of Syria. *Water,* 14(10),1670. doi: 10.3390/w14101670

Attachment

Table S1. Performance Evaluation Forms.

Station	Model	RMSE	MAE	R²
Al-Baaj	Stacking	0.7	0.5	1
Al-Baaj	CatBoost	0.9	0.6	0
Al-Baaj	Random Forest	0.9	0.6	0
Al-Baaj	XGBoost	0.9	0.6	0
Al-Baaj	LSTM	1.9	1.6	0
Tal Abtah	Random Forest	0.7	0.5	1
Tal Abtah	XGBoost	0.7	0.5	0
Tal Abtah	CatBoost	0.8	0.6	0
Tal Abtah	Stacking	0.8	0.6	1
Tal Abtah	LSTM	1.9	1.6	0
Tal Afar	XGBoost	1.1	0.6	0
Tal Afar	CatBoost	1.1	0.6	0
Tal Afar	Random Forest	1.1	0.7	0
Tal Afar	Stacking	1.2	0.8	1
Tal Afar	LSTM	2.1	1.7	0
Rabia	Random Forest	0.9	0.6	0
Rabia	CatBoost	1	0.7	1
Rabia	XGBoost	1	0.7	0
Rabia	Stacking	1	0.7	1
Rabia	LSTM	2	1.6	1
Sinjar	XGBoost	0.8	0.5	0
Sinjar	Random Forest	0.9	0.6	1
Sinjar	CatBoost	0.9	0.7	0
Sinjar	Stacking	0.9	0.7	1
Sinjar	LSTM	2	1.6	1
Makhmur	CatBoost	0.7	0.4	1
Makhmur	XGBoost	0.7	0.4	1
Makhmur	Random Forest	0.7	0.5	0
Makhmur	Stacking	0.7	0.6	1
Makhmur	LSTM	1.9	1.6	0
Mosul	CatBoost	0.7	0.5	1
Mosul	XGBoost	0.8	0.5	1
Mosul	Random Forest	0.9	0.6	0
Mosul	Stacking	0.9	0.7	1
Mosul	LSTM	0.9	0.6	0

Table S2. Analysis of Seasonal Rainfall Trends in the Study Area for the Period (1994-2024).

Station	Season	Trend	p-value	Sen's Slope (mm/year)
Makhmur	Autumn	increasing	0.362	0.469
Tal Abta	Autumn	increasing	0.414	0.338
Baaj	Autumn	decreasing	0.61	-0.214
Rabia	Autumn	increasing	0.711	0.343
Sinjar	Autumn	increasing	0.838	0.089
Mosul	Autumn	increasing	0.84	0.226
Tal Afar	Autumn	increasing	0.946	0.044
Sinjar	Spring	increasing	0.061	2.748
Tal Abta	Spring	increasing	0.144	1.667
Rabia	Spring	increasing	0.341	1.429
Makhmur	Spring	increasing	0.399	1.039
Baaj	Spring	increasing	0.434	0.818
Tal Afar	Spring	increasing	0.458	1.286
Mosul	Spring	decreasing	0.973	-0.016
Rabia	Summer	no trend	0.001	0
Tal Abta	Summer	no trend	0.049	0
Mosul	Summer	decreasing	0.05	0
Sinjar	Summer	no trend	0.061	0
Tal Afar	Summer	no trend	0.076	0
Baaj	Summer	no trend	0.118	0
Makhmur	Summer	no trend	0.197	0
Baaj	Winter	decreasing	0.153	-1.615
Mosul	Winter	decreasing	0.163	-1.729
Makhmur	Winter	decreasing	0.344	-0.933
Tal Afar	Winter	decreasing	0.362	-1.267
Tal Abta	Winter	increasing	0.575	0.333

Table S2. Continued.

Station	Season	Trend	p-value	Sen's Slope (mm/year)
Sinjar	Winter	decreasing	0.787	-0.164
Rabia	Winter	decreasing	0.893	-0.322

Table S3. Analysis of the Monthly Trend of Rainfall in the Study Area for the Period (1994-2024).

Station	Month	Trend	p-value	Sen's Slope (mm/year)
Baaj	1	decreasing	0.261	-0.688
Baaj	2	decreasing	0.276	-0.5
Baaj	3	no trend	0.932	0
Baaj	4	decreasing	0.76	-0.2
Baaj	5	increasing	0.143	0.462
Baaj	6	no variation	1	0
Baaj	7	no variation	1	0
Baaj	8	no trend	0.118	0
Baaj	9	no trend	0.083	0
Baaj	10	decreasing	0.357	-0.111
Baaj	11	no trend	0.892	0
Baaj	12	decreasing	0.486	-0.353
Tal Abta	1	increasing	0.747	0.295
Tal Abta	2	increasing	0.586	0.438
Tal Abta	3	increasing	0.575	0.357
Tal Abta	4	increasing	0.262	0.535
Tal Abta	5	increasing	0.083	0.421
Tal Abta	6	no trend	0.118	0
Tal Abta	7	no trend	0.219	0
Tal Abta	8	no variation	1	0
Tal Abta	9	no trend	0.115	0
Tal Abta	10	decreasing	0.099	-0.167
Tal Abta	11	increasing	0.138	0.412
Tal Abta	12	increasing	0.838	0.057
Tal Afar	1	increasing	0.762	0.273
Tal Afar	2	decreasing	0.264	-0.514
Tal Afar	3	decreasing	0.787	-0.35
Tal Afar	4	increasing	0.198	0.887
Tal Afar	5	increasing	0.114	0.342
Tal Afar	6	no trend	0.176	0
Tal Afar	7	no trend	0.12	0
Tal Afar	8	no trend	0.205	0
Tal Afar	9	no trend	0.06	0
Tal Afar	10	decreasing	0.332	-0.111
Tal Afar	11	increasing	0.711	0.228
Tal Afar	12	decreasing	0.683	-0.309
Rabia	1	increasing	0.262	0.7
Rabia	2	decreasing	0.418	-0.479
Rabia	3	increasing	0.812	0.15
Rabia	4	increasing	0.734	0.2
Rabia	5	increasing	0.103	0.594
Rabia	6	no trend	0.003	0
Rabia	7	no trend	0.005	0
Rabia	8	no trend	0.264	0
Rabia	9	no trend	0.474	0
Rabia	10	decreasing	0.696	-0.096
Rabia	11	increasing	0.163	0.733
Rabia	12	decreasing	0.671	-0.317
Sinjar	1	increasing	0.541	0.575
Sinjar	2	no trend	0.986	0
Sinjar	3	increasing	0.786	0.27
Sinjar	4	increasing	0.049	1.21
Sinjar	5	increasing	0.008	1
Sinjar	6	no trend	0.131	0
Sinjar	7	no trend	0.134	0
Sinjar	8	no trend	0.779	0
Sinjar	9	no trend	0.013	0
Sinjar	10	decreasing	0.946	-0.004
Sinjar	11	increasing	0.415	0.514
Sinjar	12	decreasing	0.622	-0.525
Makhmur	1	decreasing	0.734	-0.158

Table S3. Continued.

Station	Month	Trend	p-value	Sen's Slope (mm/year)
Makhmur	2	decreasing	0.518	-0.3
Makhmur	3	increasing	0.973	0.038
Makhmur	4	increasing	0.634	0.246
Makhmur	5	increasing	0.045	0.338
Makhmur	6	no trend	0.197	0
Makhmur	7	no variation	1	0
Makhmur	8	no variation	1	0
Makhmur	9	no variation	1	0
Makhmur	10	decreasing	0.434	-0.14
Makhmur	11	increasing	0.236	0.442
Makhmur	12	decreasing	0.878	-0.075
Mosul	1	decreasing	0.61	-0.45
Mosul	2	decreasing	0.438	-0.6
Mosul	3	decreasing	0.21	-0.769
Mosul	4	decreasing	0.84	-0.062
Mosul	5	increasing	0.174	0.229
Mosul	6	no trend	0.437	0
Mosul	7	no trend	0.029	0
Mosul	8	no trend	0.718	0
Mosul	9	no trend	0.723	0
Mosul	10	decreasing	0.563	-0.067
Mosul	11	increasing	0.589	0.286
Mosul	12	decreasing	0.292	-0.883