POLYTECHNIC JOURNAL

Polytechnic Journal

Polytechnic Journal

Volume 15 | Issue 1 Article 3

Utilizing Machine Learning and Deep Learning for Precise Intensity-Duration- Frequency (IDF) Curve Predictions

Sheeraz M. Ameen
Petroleum Technology-Petrochemical Department, Koya Technical Institute, Erbil Polytechnic University, Erbil, Iraq.

Shuokr Qarani Aziz
Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq

Anwer Hazim Dawood
Department of Geotechnical Engineering, Faculty of Engineering, Koya University, Koya, Kurdistan Region, Iraq, anwer.hazim@koyauniversity.org

Azhin Tahir Sabir
Department of Software Engineering, Faculty of Engineering, Koya University, Koya, Kurdistan Region, Iraq.

Dara Muhammad Hawez
Department of Civil Engineering, University of Raparin, Ranya, Sulaymani, Kurdistan Region, Iraq

How to Cite This Article

Ameen, Sheeraz M.; Aziz, Shukr Qarani; Dawood, Anwer Hazim; Sabir, Azhin Tahir; and Hawez, Dara Muhammad (2025) "Utilizing Machine Learning and Deep Learning for Precise Intensity-Duration- Frequency (IDF) Curve Predictions," *Polytechnic Journal*: Vol. 15: Iss. 1, Article 3.
DOI: https://doi.org/10.59341/2707-7799.1848

Utilizing Machine Learning and Deep Learning for Precise Intensity-Duration-Frequency (IDF) Curve Predictions

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI Usage Declaration

The authors declare that the content of this work was not generated using AI.

Acknowledgments

The authors sincerely acknowledge the Directorate of Agriculture in Koya for their valuable data support and cooperation.

ORIGINAL ARTICLE

Utilizing Machine Learning and Deep Learning for Precise Intensity-Duration-Frequency (IDF) Curve Predictions

Sheeraz Majeed Ameen [a], Shuokr Qarani Aziz [b], Anwer Hazim Dawood [c,*], Azhin Tahir Sabir [d], Dara Muhammad Hawez [e]

[a] Petroleum Technology-Petrochemical Department, Koya Technical Institute, Erbil Polytechnic University, Erbil, Iraq
[b] Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq
[c] Department of Geotechnical Engineering, Faculty of Engineering, Koya University, Koya, Kurdistan Region, Iraq
[d] Department of Software Engineering, Faculty of Engineering, Koya University, Koya, Kurdistan Region, Iraq
[e] Department of Civil Engineering, University of Raparin, Ranya, Sulaymani, Kurdistan Region, Iraq

Abstract

Intensity-Duration-Frequency (IDF) curves are crucial for the design and management of engineering infrastructure, including storm sewers, retention ponds, dams, and flood mitigation systems. This study adopts a comparative approach to estimate IDF curves using a combination of traditional statistical methods, machine learning techniques, and advanced deep learning models. Rainfall data from Koya City, Iraq (2005−2022), was used, with the 2005−2015 period for training and 2016−2022 for validation. The models evaluated include the Gumbel Distribution, Linear Regression, Support Vector Regression (SVR), and Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM), assessed based on three metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). Among these, the RNN-LSTM model demonstrated the lowest RMSE (1.44 mm/hr), lowest MAE (0.81 mm/hr), and highest R^2 (0.99), outperforming the Gumbel Distribution (RMSE: 9.13 mm/hr), Linear Regression (RMSE: 10.76 mm/hr), and SVR (RMSE: 6.19 mm/hr). This establishes RNN-LSTM as the most reliable approach for IDF curve prediction.

Leveraging the RNN-LSTM model, rainfall trends for 2023−2043 were forecasted, revealing an expected increase in short-duration, high-intensity rainfall events, heightening flood risks, and emphasizing the need for adaptive stormwater management strategies. The findings underscore the significant potential of deep learning models like RNN-LSTM in enhancing IDF curve predictions and guiding the development of resilient hydraulic infrastructure, particularly in regions like Koya City, where complex topography exacerbates flood challenges during intense rainfall events.

Keywords: Rainfall intensity-duration-frequency curves, RNN-LSTM, Flood risk management, Machine learning, Koya City, Iraq

1. Introduction

The Intensity-Duration-Frequency (IDF) curve is a key tool in hydrological engineering, offering essential insights into the relationship between rainfall intensity, duration, and occurrence frequency [1]. Derived from historical rainfall data, IDF curves are vital for forecasting extreme rainfall events, especially in regions vulnerable to severe weather. These curves enable engineers to determine the intensity and duration of rainfall that stormwater systems, dams, and drainage infrastructure must withstand to prevent flooding and structural failure. By quantifying rainfall extremes, IDF curves act as critical benchmarks for assessing the stability and resilience of hydraulic infrastructure under varying weather conditions. Accurate IDF predictions support more resilient infrastructure

Received 1 October 2024; accepted 12 January 2025.
Available online 12 February 2025

* Corresponding author.
E-mail address: anwer.hazim@koyauniversity.org (A.H. Dawood).

https://doi.org/10.59341/2707-7799.1848

designs, particularly in areas facing increased storm intensity and frequency due to climate change. Thus, IDF curves are indispensable in developing safe, efficient, and sustainable flood control systems that safeguard both urban and rural communities from flood-related damage [2]. In regions like Koya City, where intense rainfall events are frequent, these curves are critical for designing stormwater infrastructure and flood control systems. They help mitigate urban flooding, minimize property damage, and protect lives by guiding the development of resilient hydraulic structures tailored to local conditions. Nevertheless, the local civil engineers are in need of having up-to-date hydrological formulae, relationships, and information in order to be able to adequately design the flood control hydraulic structures. Accurate IDF curves are essential for predicting extreme rainfall events, which are crucial in flood risk management, and planning of infrastructure, particularly in the face of increasing climate variability and extreme weather events [3,4].

Traditionally, the development of IDF curves has relied on statistical methods, with the Gumbel distribution being a widely used approach due to its effectiveness in modeling extreme rainfall events [5]. The Gumbel distribution, along with other methods like the Log-Pearson Type III distribution, has been foundational in estimating the probability of extreme hydrological events [6]. However, these traditional methods often assume stationarity in the climatic data, which may not hold true under the current trends of global climate change [7]. This limitation underscores the need for more sophisticated approaches that can better capture the complex, non-linear relationships inherent in rainfall data.

In recent years, advancements in Machine Learning (ML) and deep learning (DL) have introduced powerful new tools for enhancing the accuracy and reliability of IDF curves. These approaches, which include techniques such as Support Vector Regression (SVR), Artificial Neural Networks (ANNs), and particularly Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units, have shown significant promise in modeling complex, non-linear hydrological processes [8]. For instance, studies have demonstrated that LSTM networks, which are adept at handling time-series data, can significantly outperform traditional statistical methods in predicting rainfall intensities over various durations and frequencies [9]. Machine learning methods such as SVR have been effectively employed to model rainfall data, offering improvements in capturing the non-linear relationships between rainfall intensity and other climatic variables [10,11]. Similarly, ANNs have been used to develop regional IDF curves with enhanced accuracy, adapting to the unique climatic conditions of different regions [2]. These methods not only improve the precision of IDF curve predictions, but also allow for the integration of additional data sources, such as satellite observations and climate model outputs, providing a more comprehensive understanding of rainfall patterns [10].

Moreover, recent advancements in deep learning have significantly improved predictive capabilities in various domains. For instance, studies like [12]. Demonstrate the effectiveness of LSTM models in forecasting temporal patterns in energy systems, achieving superior accuracy through their ability to capture non-linear dependencies in time-series data. Similarly [13], highlights the application of LSTM networks in fault classification, leveraging historical data to predict potential failures with high precision. Additionally [14], explores the integration of deep learning in renewable energy, showcasing how LSTM can address short-term prediction challenges in dynamic environments. These applications emphasize the robustness of deep learning methods in handling sequential data, reinforcing their relevance to rainfall intensity-duration-frequency (IDF) curve modeling. Incorporating such techniques into hydrological studies not only enhances prediction accuracy but also adapts to the increasing complexity posed by climate variability.

The importance of accurate IDF curves is particularly evident in regions with complex climatic conditions, such as the Middle East and North Africa region. In Iraq, for example, accurate IDF curve estimation is vital for managing water resources and mitigating flood risks, especially in urban areas where infrastructure is vulnerable to extreme weather events [15−17] Recent studies in the Kurdistan region of Iraq have shown that deep learning models, particularly LSTM networks, can significantly enhance the accuracy of IDF curves by better capturing the temporal dependencies in rainfall data [2]. Moreover, the integration of ML and DL techniques into hydrological modeling is not only a response to the challenges posed by climate change but also a proactive approach to improving the resilience of urban infrastructure. By continuously updating IDF curves with new data and refined models, these techniques help ensure that flood management strategies remain effective in the face of evolving climatic conditions [18,19].

This current research aims to explore and compare the efficacy of various methods, including traditional statistical approaches, machine learning techniques, and advanced deep learning models, in estimating IDF curves for Koya City, Iraq. By analyzing rainfall data from 2005 to 2022 and applying these diverse methodologies, this research seeks to identify the most reliable techniques for predicting future rainfall patterns and supporting flood risk management in the region. To date, this kind of the study has not been carried out in the selected area.

2. Materials and methods

2.1. Dataset description

The study focuses on Koya City, as illustrated in Fig. 1, a region known for its distinctive climatic conditions that influence rainfall patterns. This city was selected as a case study due to the availability of comprehensive rainfall data over an extended period and its vulnerability to flooding, necessitating precise hydrological modeling. Rainfall data spanning from 2005 to 2022 was sourced from local meteorological stations and government agencies responsible for monitoring hydrological data [20].

The dataset includes daily rainfall amounts recorded over ten distinct durations: 5 min, 10 min, 20 min, 30 min, 60 min, 120 min, 180 min, 360 min, 720 min, and 1440 min. These durations, standard in hydrological studies, provide a detailed view of rainfall intensity over both short and extended periods. The data is well-suited for generating IDF-curves and analyzing future rainfall patterns.

2.2. Preprocessing and inputs

The rainfall data was preprocessed to ensure quality and compatibility for predictive modeling. Z-score normalization was applied to standardize the data, ensuring consistency across features. Missing values were addressed using mean imputation, while outliers were replaced with median values to minimize distortion. The data was codified into sequences based on standard durations (5−1440 min) and split into training (2005−2015) and validation (2016−2022) subsets. These preprocessing steps ensured the dataset was clean, consistent, and ready for analysis.

2.3. Methods

To predict future IDF curves, four different modeling approaches were employed: Gumbel Distribution (GD), Linear Regression (LR), SVR, and RNN with LSTM. The dataset was divided into two subsets: the period from 2005 to 2015 was used for training the models, while the period from 2016 to 2022 was reserved for validation.

The prediction workflow for each method is depicted in Fig. 2. Each modeling approach was applied to simulate IDF curves, and their performances were assessed using three metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). These metrics provide a comprehensive evaluation of each method's predictive accuracy, with RMSE quantifying the average magnitude of prediction

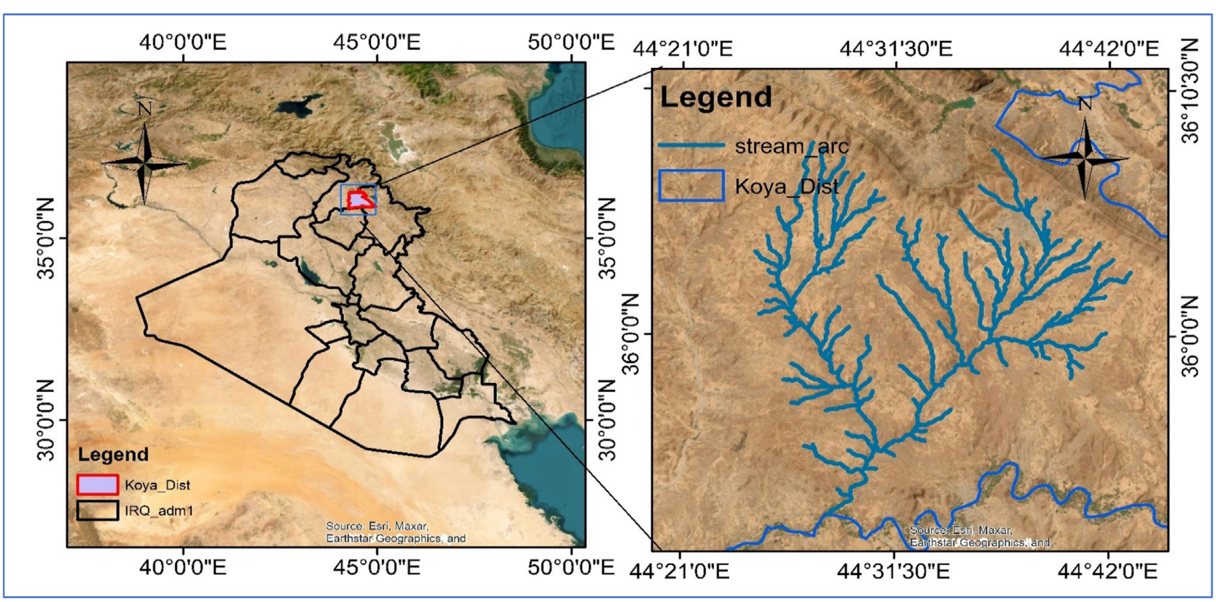

Fig. 1. Map of study area Koya City.

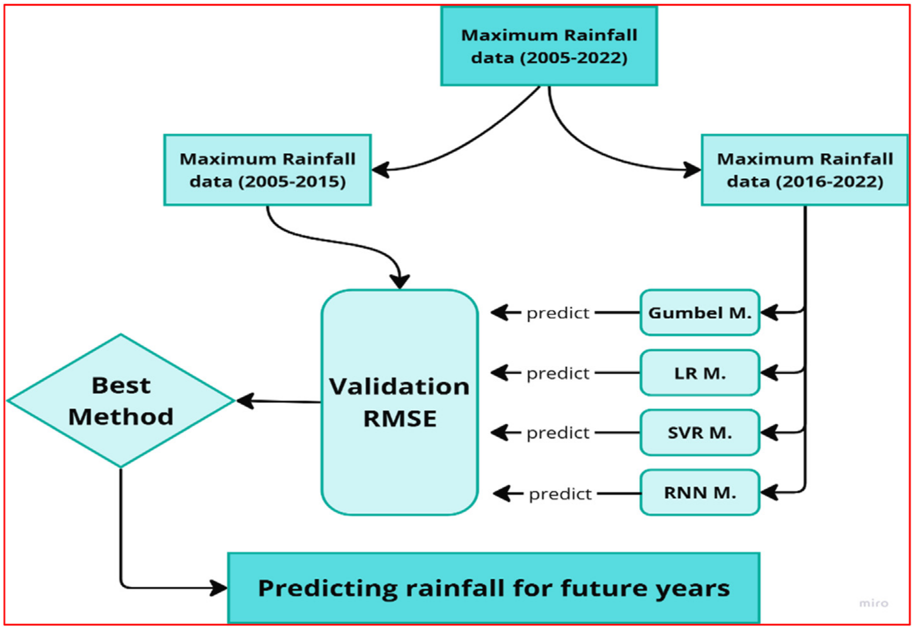

Fig. 2. Prediction of rainfall.

errors, MAE reflecting the average absolute deviation of predictions from actual values, and R^2 indicating the proportion of variance explained by the model. By incorporating multiple metrics, we ensured a more robust and transparent assessment of the predictive capabilities of each approach.

2.3.1. Gumbel Distribution method

The Gumbel Distribution is a well-recognized statistical method commonly employed to model extreme rainfall intensities over various time durations. Valued for its simplicity and reliability, it is frequently used in hydrology to estimate the likelihood of rare, high-intensity rainfall events, making it particularly useful when working with limited datasets (Gumbel, 1958). In this study, the Gumbel Distribution served as a baseline model for forecasting rainfall intensities and was instrumental in generating IDF curves, which are essential for assessing flood risks and evaluating the resilience of infrastructure against extreme weather events.

Key Steps

1. **Calculation of Statistical Parameters:** The mean (μ) and standard deviation (σ) were calculated for each of the 10 rainfall durations (5−1440 min), establishing the parameters needed to fit the Gumbel Distribution to the data.
2. **Gumbel Distribution Function:** The cumulative distribution function (CDF) of the Gumbel Distribution, defined as $F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$ …(1) was used to estimate the probability of extreme rainfall events, x is the rainfall intensity, μ is the location parameter (mean), $\boldsymbol{\sigma}$ is the scale parameter (standard deviation).
3. **Calculation of Intensity for Return Periods:** For a given return period **T**, the rainfall intensity **I(T)** was calculated as:

$$I(T) = \mu + \boldsymbol{\sigma} K_t \quad (2)$$

where: P = 1/T is the probability of exceedance associated with the return period T.

4. **Prediction of Rainfall Intensities and Generation of IDF Curves:** With the calculated parameters, IDF curves were created to illustrate the relationship between rainfall intensity, duration, and frequency, as displayed in Fig. 3. These curves are a crucial tool for flood risk management and infrastructure planning, providing essential benchmarks for designing systems capable of withstanding extreme weather events.
5. **Limitations of the Gumbel Distribution and the Role of Machine Learning:** Although the Gumbel Distribution effectively models extreme rainfall events, its assumption of independent, identically distributed data restricts its ability to adapt to the complex, non-linear patterns found in Koya City's rainfall data. These limitations underscore the benefits of machine learning approaches, which offer greater flexibility and

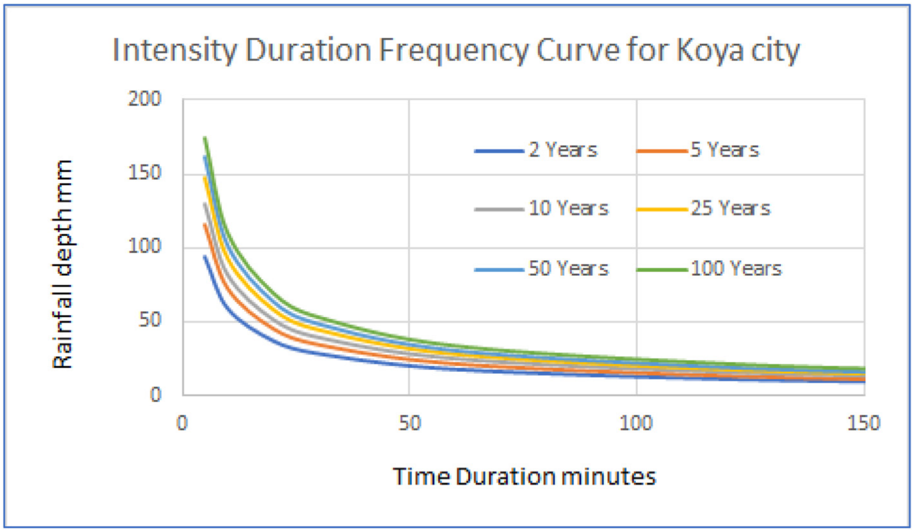

Fig. 3. Example of an IDF curve generated for Koya City.

accuracy in capturing such complex relationships, a key focus of this study.

2.3.2. *Machine learning*

IDFoffers robust tools for identifying complex, non-linear patterns in data, proving especially valuable when traditional statistical models like the Gumbel Distribution have limitations. Alongside the Gumbel approach, this study applied three ML techniques LR, SVR, and RNN-LSTM to estimate IDF curves and predict rainfall intensities. These ML methods enhance prediction accuracy, allowing for a deeper understanding of rainfall trends and improved infrastructure planning.

2.3.2.1. Linear Regression (LR). Linear Regression was used as a baseline model to establish a simple linear relationship between rainfall intensity and duration, represented by $I = \alpha D + \beta$, where I is the intensity, D the duration, α the slope, and β the intercept. Although LR assumes linearity, it provides a reference point for assessing the predictive performance of more complex models (Yu, Yang, and Lin, 2004). The LR model predicts future rainfall intensity based solely on historical rainfall data without explicitly incorporating “year” as an input feature. This approach assumes a direct relationship between rainfall intensities and durations, neglecting temporal trends. While this simplification facilitates implementation, it limits the model's ability to capture temporal variability.

2.3.2.2. Support Vector Regression. SVR extends Support Vector Machines to regression, using a Gaussian kernel to capture non-linear relationships in rainfall data. SVR identifies a function $f(x) = \omega \cdot \phi(x) + b$ that maintains predictions within an acceptable margin of error while minimizing deviations from observed values. This method offers an intermediate solution between simple linear models and deep learning (Ramaseshan, 1996).

2.3.2.3. RNN-LSTM model for rainfall prediction. The study employed an RNN-LSTM model to capture long-term dependencies within sequential rainfall data. LSTM networks are specifically designed to retain important historical information through memory cells, making them particularly effective for time-series forecasting. Here, the RNN-LSTM model was trained on historical rainfall data to predict future intensities, adeptly capturing the temporal patterns unique to Koya City's climate. As shown in Fig. 4, the RNN-LSTM's architecture leverages input, output, and forget gates to regulate data flow, enhancing its capacity to model complex rainfall sequences [9].

2.4. *Validation and comparison*

2.4.1. *Comparison metric*

To evaluate the accuracy and reliability of the rainfall intensity predictions, we used three comparison metrics: RMSE, MAE, and R^2. These metrics were chosen to provide a comprehensive assessment of the models' performance across various aspects. RMSE was used as the primary metric due to its sensitivity to large errors, MAE for assessing overall prediction accuracy without emphasizing larger errors, and R^2 for evaluating how well the predictions align with the observed data.

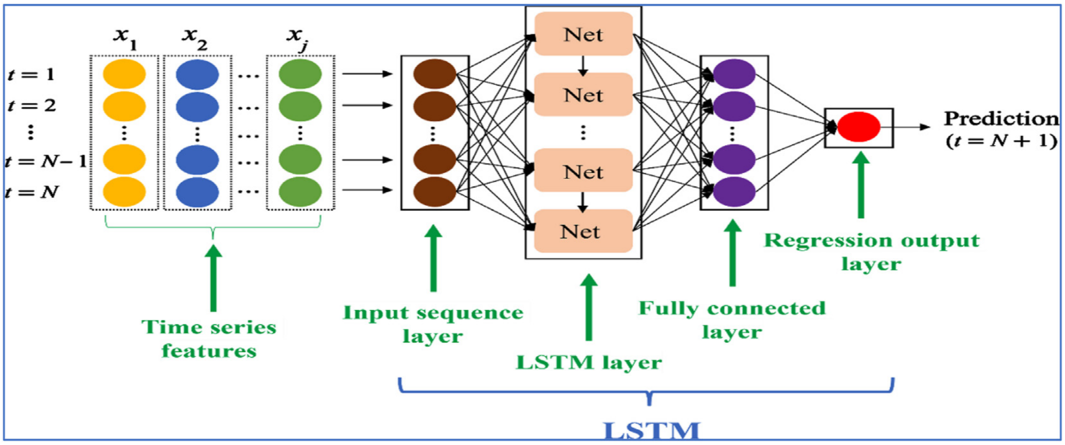

Fig. 4. Flow diagram of Recurrent Neural Networks (RNN), to predict rainfall intensities. LSTMs.

2.4.2. Calculation of RMSE

The Root Mean Square Error (RMSE) is a widely used metric to evaluate the accuracy of predictive models. In the context of this study, RMSE was used to quantify the difference between the predicted rainfall intensities (2016–2022) generated by the SVR models and the actual observed rainfall intensities from 2016 to 2022. The RMSE provides an indication of how closely the model's predictions align with the actual data, with lower values indicating better model performance.

2.4.3. Mathematical definition of RMSE

The RMSE is defined as the square root of the mean of the squared differences between the predicted and actual values. Mathematically, for a given duration (d_i), the RMSE is calculated as follows:

$$RMSEd_i = \frac{1}{n}\sum_{k=1}^{n}\left(P_{k,di} - A_{k,di}\right)^2 \tag{5}$$

where:

n is the number of data points (years) in the prediction period (2016–2022), which is 7 in this case.

$P_{k,d}$ is the predicted rainfall for year k and duration d_i.

$A_{k,d}$ is the actual observed rainfall for year k and duration d_i.

The RMSE for each duration is computed by first calculating the squared differences between the predicted and actual values for each year, taking their average, and then finding the square root of this average.

2.4.4. Calculation of MAE

The Mean Absolute Error (MAE) measures the average magnitude of errors, treating all deviations equally without emphasizing larger errors. It is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n}\left|P_{k,d_i} - A_{k,d_i}\right|$$

MAE complements RMSE by offering insight into the overall accuracy of the model while being less sensitive to outliers.

2.4.5. Calculation of R^2

The **Coefficient of Determination (R^2)** evaluates how well the model explains the variance in the observed data. It is calculated as:

$$R^2 = 1 - \frac{\sum_{k=1}^{n}\left(A_k - P_k\right)^2}{\sum_{k=1}^{n}\left(A_k - \overline{A}\right)^2}$$

Where:

A_k Actual observed value for year.

P_k Predicted value for year k.

A¯ Mean of actual observed values.

An R^2 value closer to 1 indicates better model performance, while negative values suggest the model performs worse than predicting the mean of the observed data.

2.4.6. Interpretation of metrics

- RMSE and MAE together provide a detailed view of the model's accuracy, with RMSE emphasizing larger errors and MAE reflecting average error magnitudes.
- R^2 adds a measure of how well the model captures the variance in the observed data, further validating its predictive reliability.

By incorporating these metrics, this study ensures a robust and transparent evaluation of the models, highlighting the superior performance of RNN-LSTM compared to traditional methods like Gumbel Distribution, Linear Regression, and SVR.

2.5. Experimental study

In this study, we evaluated four approaches for predicting future IDF curves using rainfall data from Koya City, Iraq, spanning the period 2005–2022. The dataset was divided into two subsets: 2005–2015 for training and 2016–2022 for validation. The data encompassed various rainfall durations (5–1440 min) and return periods (2–100 years). The methods assessed were the GD, LR, SVR, and RNN-LSTM.

The Gumbel Distribution was applied to model extreme rainfall values by fitting the distribution to the annual maxima. Linear Regression established a baseline model, assuming a direct linear relationship between rainfall intensity, duration, and return period. SVR utilized machine learning to capture non-linear relationships in the data, optimizing its parameters to minimize prediction error. The RNN-LSTM model, leveraging deep learning, excelled at handling temporal dependencies and capturing complex rainfall patterns.

To evaluate model performance, three metrics were used: RMSE, MAE, and R^2. RMSE served as the primary metric, quantifying the difference between predicted and observed rainfall intensities. Based on the RMSE results, the RNN-LSTM model emerged as the best-performing approach, demonstrating superior accuracy. This model will be used to forecast future rainfall intensities for the period 2023–2043, utilizing the full dataset (2005–2022) for model training and calibration.

2.6. Experimental results

In this section, we analyze the performance of four different predictive models GD, LR, SVR, and RNN for estimating IDFcurves. Fig. 5 compares the performance of four prediction methods Gumbel Distribution (GD), Linear Regression (LR), SVR, and Recurrent Neural Network (RNN) in rainfall intensity prediction. Each method's performance is evaluated using three metrics: RMSE, MAE, and the R^2. These metrics provide insights into the accuracy, reliability, and consistency of the predictions, with lower RMSE and MAE values and higher R^2 indicating better performance. Among these methods, GD and LR represent more traditional statistical approaches, while SVR and RNN leverage machine learning for predictive modeling.

As seen in Fig. 5, the results highlight significant differences in the predictive performance of the four methods. The Recurrent Neural Network (RNN) achieves the best results with the lowest RMSE (1.44) and MAE (0.81041), alongside an impressive R^2 value of 0.99133, indicating high accuracy and reliability. In contrast, the traditional Gumbel Distribution (GD) and Linear Regression (LR) methods perform poorly, with negative R^2 values and significantly higher RMSE and MAE, reflecting limited predictive capability. SVR performs reasonably well with an RMSE of 6.19 and R^2 of −0.17, but it is outperformed by the RNN in all metrics, underscoring the superiority of deep learning in handling complex rainfall prediction tasks.

2.7. IDF curves: real vs. predicted

To evaluate the performance of each method, we compare the IDF curves generated by each model

Fig. 5. Histograms of error metrics for prediction methods.

with the actual IDF curve derived from historical data. Figs. 6–9 shows the Real vs. Predicted IDF Curves for GD, LR, SVR, and RNN.

In each of these figures, the x-axis represents the duration (in minutes), and the y-axis represents the intensity (in mm/hr). The real IDF curve (based on historical data) is plotted alongside the predicted IDF curve from each model. This visual comparison allows us to see where each model deviates from the actual data, providing insight into the strengths and weaknesses of each method.

2.8. Predictions

Fig. 10 presents a comparison of IDF curves for historical rainfall intensity from 2005 to 2022 with predicted intensities for 2023–2043. The analysis indicates that significant changes in rainfall patterns are expected in the coming decades. The predictive models consistently show an upward trend in both the frequency and intensity of extreme rainfall events, highlighting the increasing risk of flooding. These findings suggest a pressing

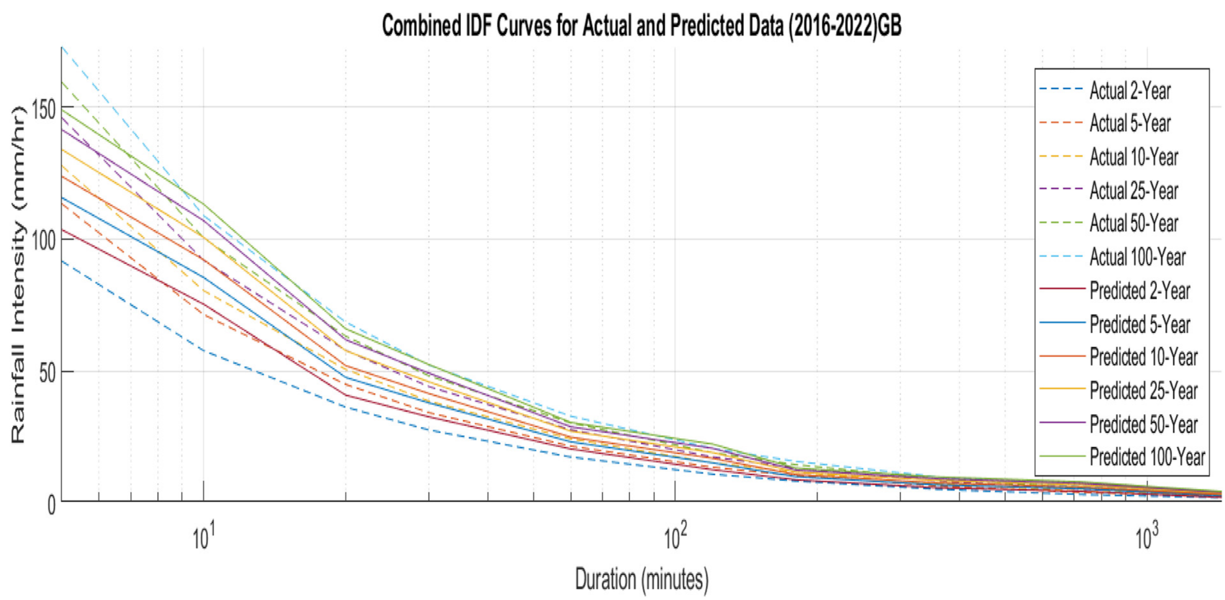

Fig. 6. IDF Curve – Actual and Gumbel Distribution (GD) (2016–2022).

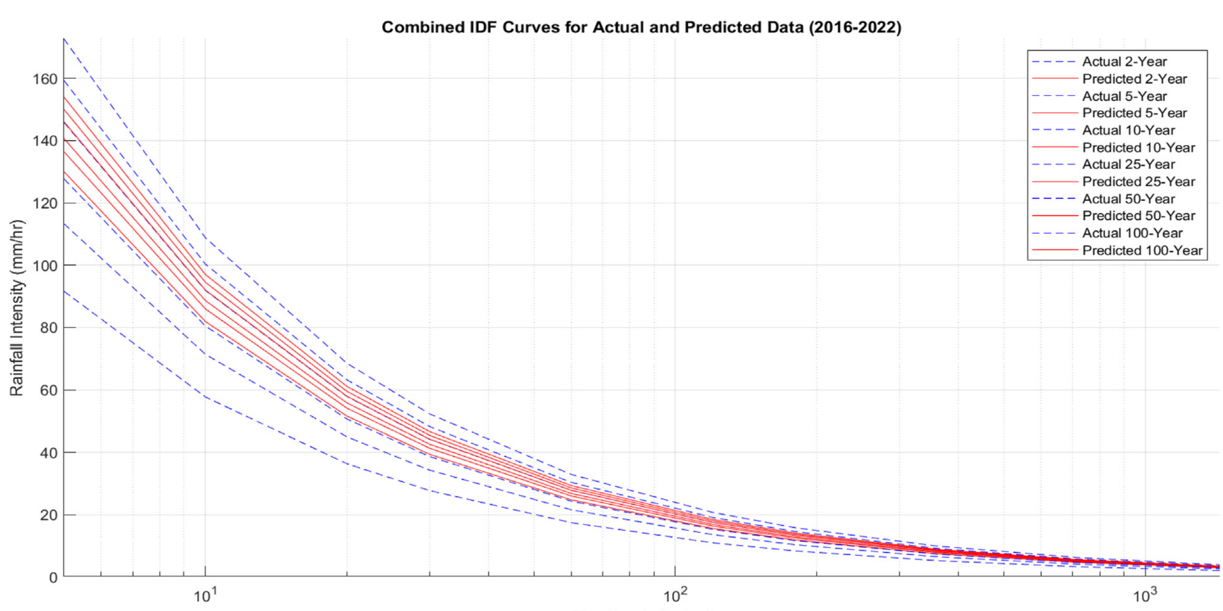

Fig. 7. IDF Curve – Actual and Linear Regression (LR).

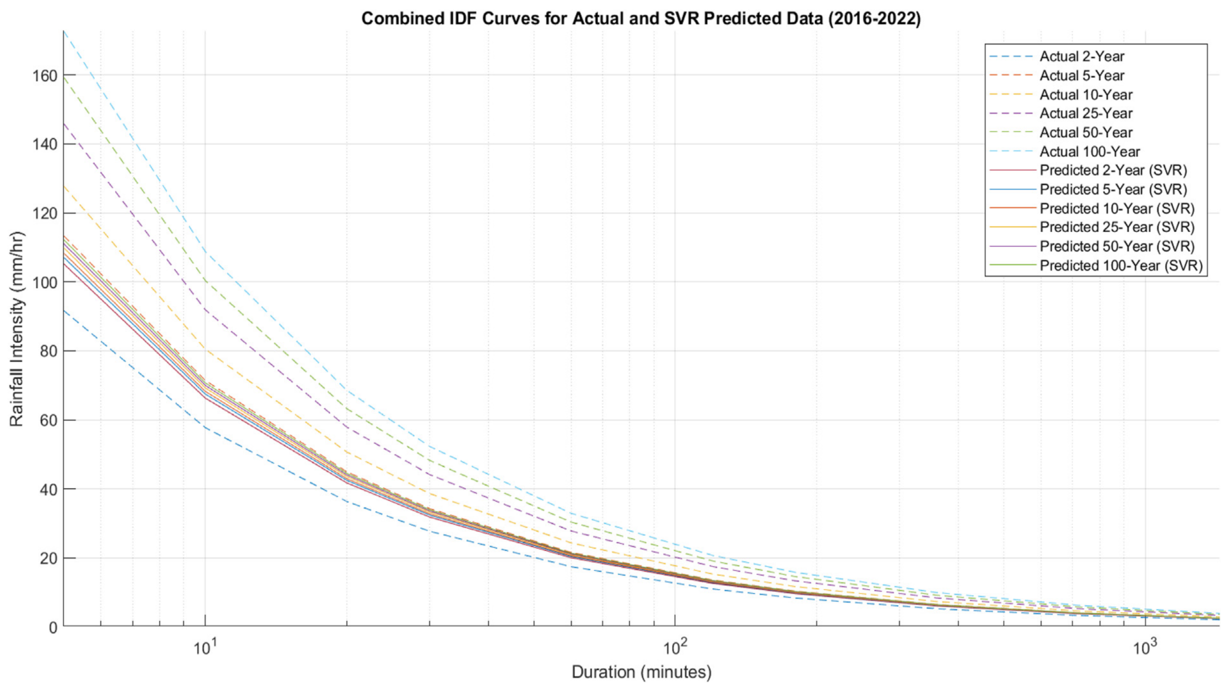

Fig. 8. IDF Curve — Actual and SVR.

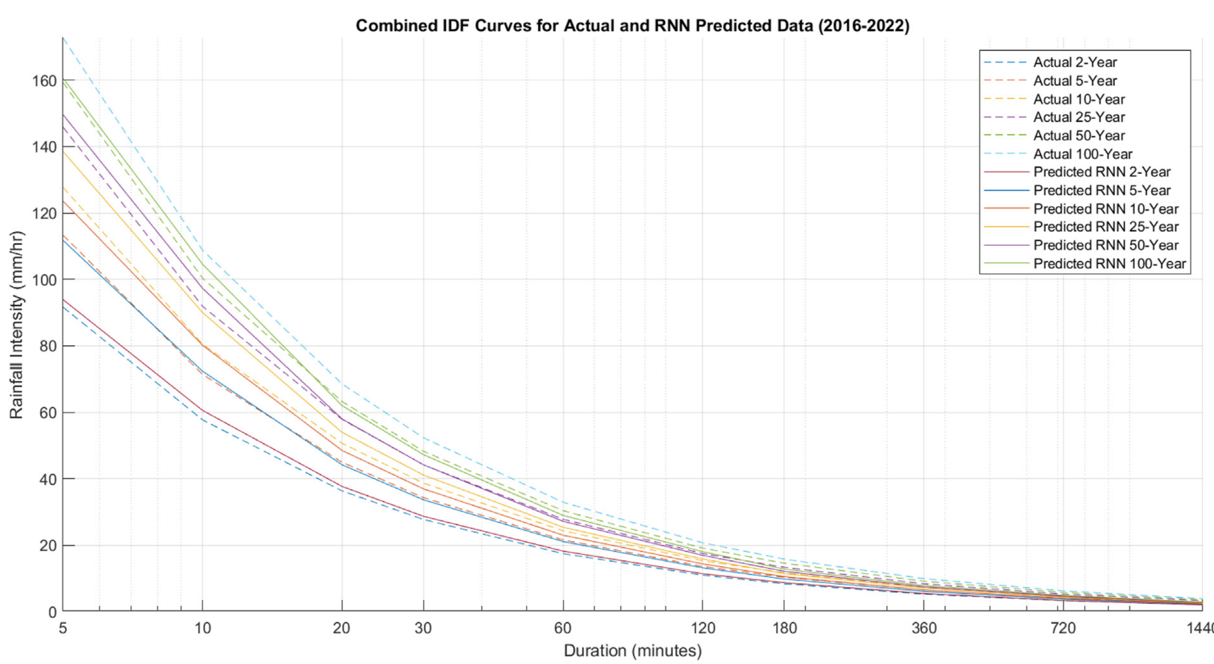

Fig. 9. IDF Curve - Actual and Recurrent Neural Network (RNN).

need for updated infrastructure planning and enhanced water resource management strategies. The predictions emphasize the critical importance of implementing adaptive measures to mitigate the impacts of climate change on both urban and rural environments.

2.9. Comparison of methods

The four models analyzed in this study GD, LR, SVR, RNN differ notably in their predictive capabilities for IDF curves. Among these, the RNN model demonstrates superior performance, with a

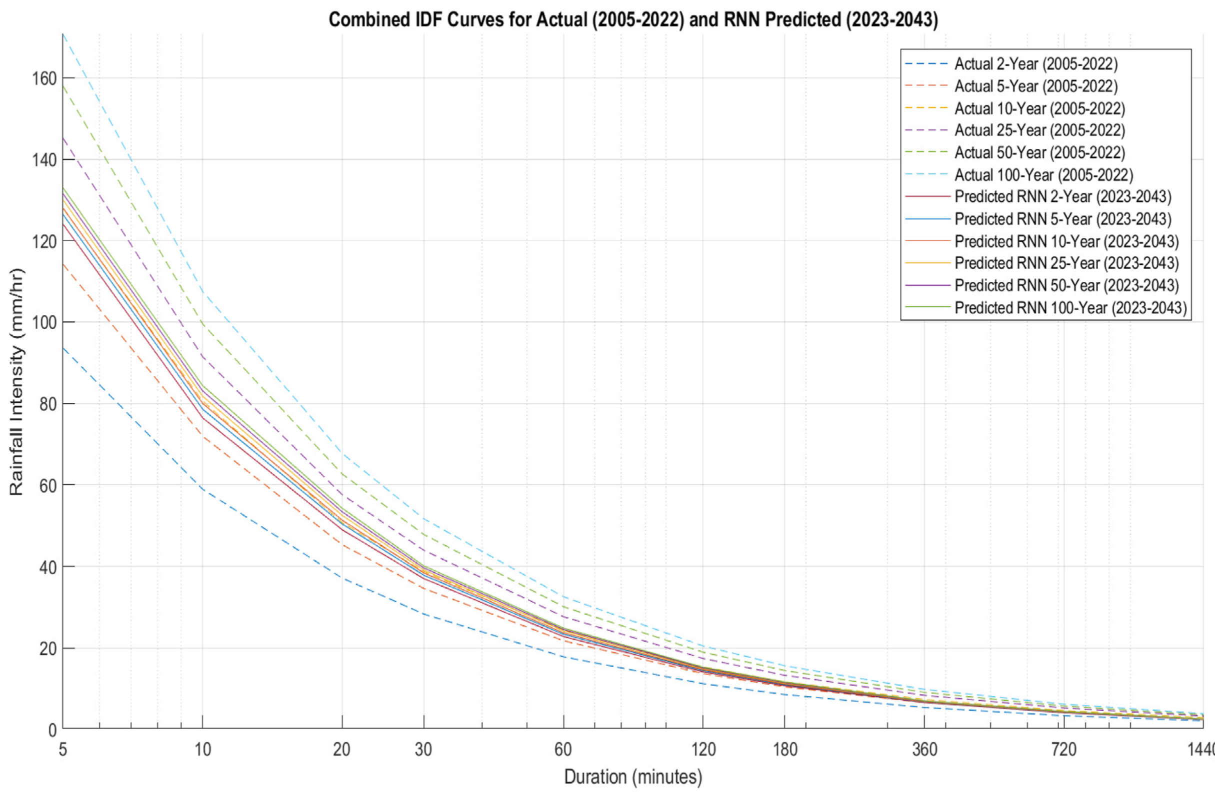

Fig. 10. IDF curves for historical rainfall intensity from 2005 to 2022 with predicted intensities for 2023−2043.

significantly lower RMSE and a predicted IDF curve that closely matches the actual data, as seen in Fig. 4. In contrast, the GD and LR models exhibit higher RMSE values and less accurate IDF curves (Figs. 1 and 2), suggesting they struggle to capture the non-linear complexities inherent in the data. The SVR model performs better than GD and LR but still does not match the accuracy of the RNN model, indicating that while SVR can handle some non-linearities, it may not fully capture the intricate dependencies present in the data (Fig. 3).

2.9.10. Focus on RMSE results

The RMSE is a key metric for assessing the accuracy of predictive models, reflecting the average error between observed and predicted values. The RNN model's exceptionally low RMSE (1.44) is corroborated by its IDF curve (Fig. 4), which aligns closely with the actual data. This strong alignment underscores the RNN model's ability to generalize effectively from training data, making it the most reliable choice for future predictions.

In contrast, the higher RMSE values for the GD and LR models indicate lower predictive accuracy, as evidenced by the discrepancies between their predicted and actual IDF curves (Figs. 1 and 2). These models are less adept at forecasting future rainfall intensities, particularly when faced with complex, non-linear data patterns.

3. Discussion and Conclusions

This study offers critical insights into projected changes in rainfall intensity and frequency over the next two decades, specifically tailored to the unique climate patterns and data constraints of Koya City. Through the application of an RNN-LSTM, the study underscores the model's robustness and effectiveness for predicting future IDF curves, especially important in the context of climate change, where complex and non-linear dependencies are expected to increase. The adaptability of the LSTM model to limited datasets highlights its suitability for regions like Koya City, where data scarcity and challenging mountainous terrain limit traditional modeling approaches. The analysis predicts an increase in short-duration, high-intensity rainfall events, pointing to a heightened risk of flash floods and prompting a need to reassess current flood prevention strategies and infrastructure resilience. Conversely, the expected decrease in long-duration rainfall events may impact water resource management, particularly in sustaining water supplies during prolonged dry spells. These region-specific insights emphasize the need for adaptive infrastructure to respond to evolving rainfall patterns effectively.

Comparing four predictive models GD, LR, SVR, and RNN-LSTM revealed notable differences in performance across three key metrics: RMSE,MAE, and R^2. Among these, the RNN-LSTM model consistently achieved the best results with the lowest RMSE (1.44 mm/hr) and MAE (0.81 mm/hr), and the highest R^2 (0.99), demonstrating superior alignment between predicted and historical IDF curves. In contrast, GD and LR performed poorly with significantly higher RMSE (9.13 and 10.76 mm/hr, respectively) and negative R^2 values, reflecting their limitations in capturing temporal trends. SVR provided moderate performance, with RMSE and MAE values of 6.19 mm/hr and 5.38 mm/hr, respectively, but it was outperformed by RNN-LSTM in all metrics.

The practical implications of these findings are significant. The RNN-LSTM model's ability to adapt to limited datasets and capture intricate rainfall dynamics underscores its utility for regions like Koya City, where data availability and variability pose challenges. By identifying future rainfall patterns with greater precision, the RNN-LSTM model supports informed decision-making in stormwater management and flood mitigation planning. Its superior accuracy in predicting short-duration, high-intensity events is particularly valuable for designing resilient urban infrastructure capable of mitigating flash flood risks. Furthermore, the model's ability to predict long-term trends supports sustainable water resource management, particularly in adapting to shifts in prolonged rainfall events.

This study's unique contribution lies in its application of RNN-LSTM modeling within a data-limited, region-specific context like Koya City. Through model optimization, including hyperparameter tuning, regularization, and validation adjustments, the research illustrates how deep learning can support flood risk management in challenging environments. Building on findings from [21], which suggest potential gains from adaptive learning rates or seasonal adjustments, future studies could consider these methods to improve model adaptability in regions with variable climates.

Overall, the findings highlight the vital role of advanced machine learning models, particularly RNN-LSTM, in environmental forecasting and strategic planning. By integrating these predictive tools into urban and environmental planning, communities can construct resilient infrastructure, mitigate flood risks, and manage water resources sustainably in the face of increasingly unpredictable climate patterns.

Funding body

There is no funding.

Ethical statement

None.

Author contribution

The authors take full responsibility for the content and accuracy of the final version of the manuscript.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] Chow V Te, Maidment DR, Mays LW. Applied hydrology. New York City, New York, U.S.: McGraw Hill, Inc; 1988.
[2] Kareem AT, Ahmed AJ, Mohammed SH. Enhancing IDF curve prediction accuracy using deep learning techniques: A case study of the Kurdistan region. Water Resour Manag 2022;36(4):1231−47.
[3] Dawood AH, Mawlood DK. Derivation of the Rainfall Intensity- Duration- Frequency Curve for Erbil city, Using SCS II Method. Zanco J Pure Appl Sci 2023;35(SpA). https://doi.org/10.21271/zjpas.35.spa.1.
[4] Lionello P, Malanotte-rizzoli P, Boscolo R. MedCLIVAR Mediterranean CLImate VARiability. Earth Environ Sci 2006;(January).
[5] Taylor P. Generalized Gumbel distribution. J Appl Stat 2020; 37(January 2010):171−9. https://doi.org/10.1080/02664760802698995.
[6] Griffis VW, Asce M, Stedinger JR, Asce M. Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis I: Distribution Characteristics. J Hydrol Eng 2007;(October):482−91.
[7] Eregno FE, Xu C, Kitterod N. Modeling hydrological impacts of climate change in different climatic zones. Int J Clim Change Strat Manag 2013;5(July):344−65. https://doi.org/10.1108/IJCCSM-04-2012-0024.
[8] Ewea HA, Elfeki AM, Al-amri NS. Development of intensity − duration − frequency curves for the Kingdom of Saudi Arabia. 5705. Oxfordshire, United Kingdom: Taylor and Francis, Milton Park, Abingdon-on-Thames; 2017. https://doi.org/10.1080/19475705.2016.1250113.
[9] Yu Q, Tolson BA, Shen H, Han M, Mai J, Lin J. Enhancing LSTM-based streamflow prediction with a spatially distributed approach. Hydrol Earth Syst Sci 2023;1(November): 1−23.
[10] Sidek LM, Basri H, Sammen SS. Machine learning techniques for flood forecasting Uncorrected Proof Machine learning techniques for fl ood forecasting. J Hydroinf 2024; 1(March). https://doi.org/10.2166/hydro.2024.208.
[11] Elsebaie IH. Developing rainfall intensity − duration − frequency relationship for two regions in Saudi Arabia, vol. 1932. Riyadh, Saudi Arabia Kingdom: King Saud University − Engineering Sciences; 2012. p. 131−40. https://doi.org/10.1016/j.jksues.2011.06.001.

[12] Le Kim KN, Quoc TN, Thanh PN. Deep Learning Method For Load Forecast In Smart Solar System Based Long Short-Term Memory. In: 2024 8th international artificial intelligence and Data processing symposium (IDAP), malatya, Turkiye. New York, USA: IEEE, The Institute of Electrical and Electronics Engineers (IEEE); 2024. p. 1−5.

[13] Liu W-B, Bui T-M, Quoc TN, Thi VP, Cho M-Y, Da TN, et al. Failure Classified Method for Diesel Generators Based Long Short-Term Memory Approach. In: 2024 international conference on system science and engineering (ICSSE), hsinchu, Taiwan. New York, USA: IEEE, The Institute of Electrical and Electronics Engineers (IEEE); 2024. p. 1−6.

[14] Da TN, Yimin L, Peng CL, Cho M-Y, Le Kim KN, Thanh PN. Short-term Solar Power Prediction using Long Short-Term Memory in Solar Plant with Deep Learning Machine. In: 2022 6th international conference on green technology and sustainable development (GTSD), Nha Trang city, Vietnam. New York, USA: IEEE, The Institute of Electrical and Electronics Engineers (IEEE); 2022. p. 651−6.

[15] Azi SQ, Mizzouri NS, Ahmed KO, Hamaamin YA, Hawez DM. Intensity-Duration-Frequency (IDF) Modelling for Kurdistan Region Provinces- Intensity-Duration-Frequency (IDF) Modelling for Kurdistan Region Provinces-Iraq. Eur Chem Bull 2023;12(10):7177−85. https://doi.org/10.48047/ecb/2023.12.10.534.

[16] Aziz SQ, Saleh SM, Muhammad SH, Ismael SO, Ahmed BM. Flood disaster in Erbil city : problems and solutions flood disaster in Erbil city : problems and solutions. Environmental Protection Research, December. Singapore 238875: Universal Wiser Publisher; 2023. https://doi.org/10.37256/epr.3220232993.

[17] Dawood AH, Mawlood DK. Flood Risk Analysis in Potential Points, Zerin City in Erbil as a Case Study. Iraqi Geol J 2023;54:125−37. https://doi.org/10.46717/igj.56.1E.13ms-2023-5-23.

[18] Nhat LT, Tachikawa Y, Takara K. Estimation of rainfall intensity duration frequency distribution in the mountainous regions of Vietnam, vol. 49. Kyoto, 606-8501, Japan: Annuals of Disaster Prevention Research Institute, Kyoto University; 2006. p. 73−82.

[19] Eckersten H. Modeling hydrological impacts of climate change using statistical methods. J Hydrol 2016;542:213−26.

[20] Dawood AH, Mawlood DK, Al-ansari N. Flood Modeling on Koya Catchment Area Using Hyfran , Web Map Service , and HEC-RAS Software 2021:107−11. https://doi.org/10.14500/aro.10824.

[21] Du B, Li W, Tan X, Ye J, Chen W, Sun L. Development of load-temporal model to predict the further mechanical behaviors of tunnel structure under various boundary conditions. Tunn Undergr Space Technol 2021;116:104077.