Journal of Hydroinformatics

Journal of Hydroinformatics Vol 26 No 4, 904 doi: 10.2166/hydro.2024.014

Developing an innovative machine learning model for rainfall prediction in a semi-arid region

Sarmad Dashti Latif *, Dyar Othman Mohammed and Alhassan Jaafar

Civil Engineering Department, College of Engineering, Komar University of Science and Technology, Sulaymaniyah, Kurdistan Region 46001, Iraq
*Corresponding author. E-mail: sarmad.latif@komar.edu.iq

SDL, 0000-0002-0417-3545

ABSTRACT

Due to global climate change, managing water resources is one of the most critical challenges for most countries in the world, especially in the Middle East. In the Kurdistan Region of Iraq (KRI), there is a good amount of precipitation, surface water, and groundwater, but the main issue is mismanagement of those sources. Rainfall is one of the major sources of water resources in KRI. In order to manage the available water resources and prevent natural disasters such as floods and droughts, there is a need for reliable models for forecasting rainfall. The current study focuses on developing a hybrid model, namely seasonal autoregressive integrated moving average combined with an artificial neural network (SARIMA-ANN) for forecasting monthly rainfall at Sulaymaniyah City for the duration of 1938–2012. For comparison purposes, a conventional machine learning model, namely artificial neural networks (ANN) has been applied on the same data. Two different statistical measurements, namely, root mean square error (RMSE) and coefficient of determination (R^2), have been used to check the accuracy of the proposed models. According to the findings, SARIMA-ANN outperformed ANN with RMSE = 11.5, RMSE = 51.002, $R^2 = 0.98$, $R^2 = 0.43$, respectively. The findings of the current study could contribute to Sustainable Development Goal (SDG) 6.

Key words: hybrid model, rainfall prediction, SARIMA-ANN, Sulaymaniyah City, Sustainable Development Goal (SDG) 6, water resources management

HIGHLIGHTS

- An innovative hybrid model has been developed for rainfall prediction.
- Accurate rainfall prediction will lead to better management of water resources.
- Rainfall prediction is an important hydrological tool due to global climate change.

ABBREVIATIONS

SARIMA-ANN	Seasonal autoregressive integrated moving average combined with artificial neural network
ANN	Artificial neural network
RMSE	Root mean square error
R2	Coefficient of determination
SDG	Sustainable development goal
ML	Machine learning
SVM	Support vector machine

1. INTRODUCTION

Water is one of the basic needs of humans as 60% of the human body is made of water. Water is utilized for different categories such as washing, cleaning, cooking, and so on. Just as life does not exist without oxygen, so it does not exist without water, and it is the only source that supports the Earth to distinguish itself from other planets and provide opportunities for life. The Earth's surface consists of three parts water and one part land. Finding water anywhere else in the universe is nearly impossible. It is important for all living things; therefore, water must be preserved for future generations. Water resource management is the process of planning, protecting, and improving water resources in terms of quality and quantity. Managing water resources is important in order to be able to supply water for domestic, irrigation, hydropower production, etc. Water resources management also helps to reduce the risks of extreme events such as floods (Akram & Hamid 2013; Vieira *et al.* 2020; Yang & Liu 2020; Fereshtehpour *et al.* 2021).

Nowadays, machine learning (ML) models have been used in the field of water management, especially to predict rainfall (Latif *et al.* 2023). ML is able to predict rainfall based on historical rainfall data. Nowadays, accurate prediction of rainfall is a big challenge due to climate change (Basha *et al.* 2020). ML models are widely used to predict future rainfall for the short and long term. The ML model relies on hydrological variables and data on previous rainfall and accurately processes it to predict future rainfall. Researchers around the world rely on ML due to the in-depth and accurate predictions it makes for the data (Parmar *et al.* 2017; Ahmed *et al.* 2020).

Nowadays, the Kurdistan Region of Iraq (KRI) faces water scarcity. The water level has decreased in both Dokan and Darbandikhan dams which are the main sources of water for Sulaymaniyah City in terms of supplying water for society. This water scarcity will lead to other issues in the future if the authority is not able to resolve it. The biggest problem in the KRI is the lack of a good water management plan. Developing a reliable model for predicting rainfall is considered one of the most important tools for managing water resources (Tinti 2017; Mohammed *et al.* 2018).

Having a wider population would increase water demand. To ensure there is enough water for humans, agriculture, and industry, effective water management is needed. Water resources are significantly affected by climate change, including shifts in precipitation patterns, more severe droughts, and more intense rainfall events (Boretti & Rosa 2019; Zubaidi *et al.* 2020; Barker *et al.* 2021). In today's world, starting to predict water resources with ML techniques is increasingly considered an important approach due to their many advantages over conventional modeling techniques (Latif & Ahmed 2023). ML algorithms can process a lot of data and find patterns that manual analysis might miss. Predicting the availability, demand, and quality of water can be made more accurate with ML to recognize complex relationships and patterns in data. The speed with which ML algorithms can process a large amount of data makes it ideal for real-time applications like flood forecasting and water management. ML algorithms are useful for predicting water resources under changing climate conditions and other environmental factors since they can adapt to changing conditions and learn from the new data. By determining the best methods for allocating and managing water to satisfy the requirements of various stakeholders, ML algorithms can enhance water management strategies. By automating data processing and analysis, ML methods can lower the cost of water resource management (Shen 2018).

It can be concluded that ML methods are important for predicting water resource parameters since they improve accuracy, speed, adaptability, optimization, and cost-effectiveness for effective management of water resources in today's rapidly changing world (Li & Sansalone 2022). Many studies have been proposed for forecasting monthly rainfall in different regions. For example, a study conducted by Ali & Shahbaz (2020) proposes an effective method for runoff forecasting by modeling the relationship between precipitation and runoff to identify the optimal rainfall pattern for forecasting daytime streamflow. They have identified different sets of rainfall antecedent components and developed an artificial neural network (ANN) model. Their findings show the potential of ANN based approaches as an effective alternative for solving hydrological problems. Furthermore, another study proposed by Adede *et al.* (2015) aims to describe an ML experiment that uses ANN to predict rainfall. The data set was divided into three subsets of training, validation, and testing. The test was run 100 times with different random distribution of data. For each fold, ANN was trained 100 times resulting in 10,000 prediction ensembles. The goal of repeatedly training an ANN was to build a more robust model that could handle different data variations. Moreover, another study conducted by Pham *et al.* (2020) uses meteorological variables such as maximum and minimum temperature, wind speed, relative humidity, and solar radiation at different altitudes as input parameters to predict daily rainfall in Hoa Province in Vietnam. The utilized models were adaptive network-based fuzzy inference systems (ANFIS) optimized with particle swarm optimization (PSO), ANN, and support vector machine (SVM). The results showed that all AI models provided adequate forecasts, but SVM proved to be the best method for forecasting rainfall. In another study, Malki *et al.* (2020) aimed to investigate the relationship between weather variables, particularly temperature and humidity, and the spread of COVID-19. Various ML models were proposed and employed to extract this relationship using data on the number of confirmed cases and weather variables in certain regions. The study found that weather variables, particularly temperature, are more relevant in predicting the mortality rate of COVID-19 compared to census variables. This suggests that temperature and humidity are important features for predicting the mortality rate of COVID-19. Furthermore, their study indicates that higher temperatures are associated with lower numbers of infection cases, suggesting a potential relationship between temperature and the spread of COVID-19.

The aim of the current study is to develop a hybrid ML model, namely SARIMA-ANN, for forecasting monthly rainfall in Sulaymaniyah City, located in the north of Iraq. In order to compare the accuracy of the proposed hybrid SARIMA-ANN model, ANN has been also applied to the same dataset.

2. MATERIALS AND METHODS

2.1. Study area

Sulaymaniyah is a city in the KRI located approximately 370 km northeast of Iraq's capital, Baghdad. The city is located in a low-lying territory that is around 2,400 km^2 in size, with Goizha Mountain to the north and Qaradagh Mountain to the south. The study area is characterized by a distinct continental interior climate of Mediterranean type, with an average annual precipitation ranging from 500 to 700 mm. There were around 856,990 people living in the Sulaymaniyah governorate in 2016 (Alkaradaghi *et al.* 2019). The reason for selecting Sulaymaniyah as the study area is due to several important points. For instance, Sulaymaniyah is an important center in the region due to its fertile soil for agriculture and its geographical location between Iraq and Iran. Also, it has two important dams, namely Darbandikhan and Dokan, along with several other small dams that work to store water and reduce the risk of flooding. Furthermore, Sulaymaniyah is also rich in natural resources such as gas and oil, and is considered an important center for energy. Its mountains and valleys exhibit a beautiful landscape and a geological structure. The climate of Sulaymaniyah is relatively dry, cold in winter, and hot in summer. Finally, the position of Sulaymaniyah is important because of its geographical and strategic location. This has made it a valuable city in the region. Figure 1 represents the location of the study area.

2.2. Data

The weather in Sulaymaniyah is constantly changing throughout the year, the weather in summer is generally dry and hot, the highest temperature in some areas reaches 50 °C, winter is generally cold and rainy, and the temperature reaches 5 °C or even less. In spring and autumn, the weather is generally cool and occasionally rainy. The climate of Sulaymaniyah is generally Mediterranean with moderate rainfall. In this research, rainfall data of Sulaymaniyah City has been collected in a monthly manner recorded in the rainfall station of Sulaymaniyah in the years 1938–2012. The statistics showed that the amount of rainfall in January is the highest compared to all other months in Sulaymaniyah since the sky is cloudy and the temperature is cool. The statistical data were determined in order to understand the environmental condition. Figure 2 represents the monthly rainfall from 1938 to 2012. Figure 3 shows the flowchart of the proposed SARIMA-ANN and single ANN ML models.

2.3. Artificial neural network

ANN is one of the most commonly used ML techniques for prediction purposes. Layers of interconnected artificial neurons, or nodes, make up ANN models, which use a number of mathematical operations to analyze data. Each neuron in the network takes inputs from other neurons or external data sources, processes those inputs to produce an output signal and then transmits that signal to other neurons. Once an ANN model has been trained on a substantial amount of data, it can be applied to forecast the data (Latif *et al.* 2020). ANN models can be used in conjunction with other models and data sources to increase the accuracy of rainfall forecasts. The structure of the current study's ANN model is shown in Figure 4.

2.4. SARIMA-ANN

For the purpose of forecasting time series, the SARIMA-ANN hybrid model combines the strengths of two separate models, the SARIMA and ANN. The hybrid model's SARIMA component focuses on identifying linear and seasonal trends in the data. SARIMA models, which include moving average, integrated, and autoregressive components, are frequently employed in time series analysis. By doing so, the historical dependencies and trends in the data are better captured. The SARIMA model could have trouble capturing subtle patterns and complex nonlinear interactions in the data. ANNs can capture nonlinearities and complex interactions and are very adaptable. The residuals (the discrepancies between the actual values and SARIMA predictions) are used to train an ANN, which allows the hybrid model to capture any patterns that the SARIMA model could have missed. By simulating the linear and seasonal components of the time series, the SARIMA component of the hybrid model offers a baseline forecast. Additional nonlinear patterns and dependencies are captured by the ANN

Figure 1 | Location of the study area (*Source*: Google Map, 2024).

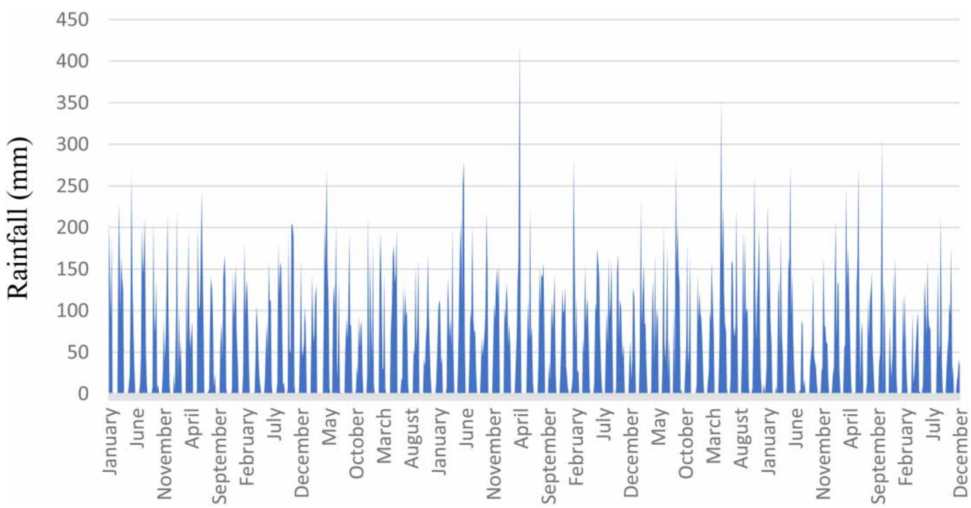

Figure 2 | Monthly rainfall in Sulaymaniyah City from January 1938 to October 2012.

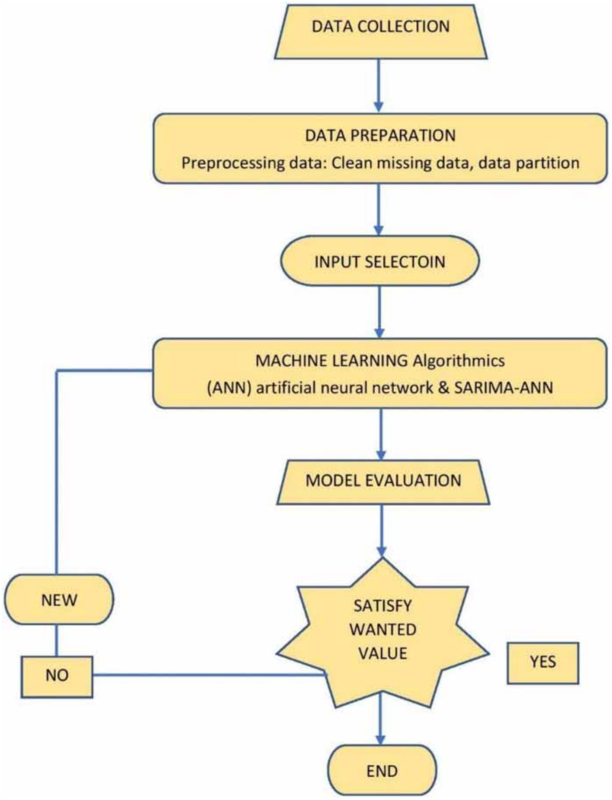

Figure 3 | The framework of the current study.

component. The final forecast is created by combining the results of the two models. The data must be divided into linear and nonlinear components before being used in a hybrid model:

$$R(t) = R_{\mathrm{L}}(t) + R_{\mathrm{N}}(t) \tag{1}$$

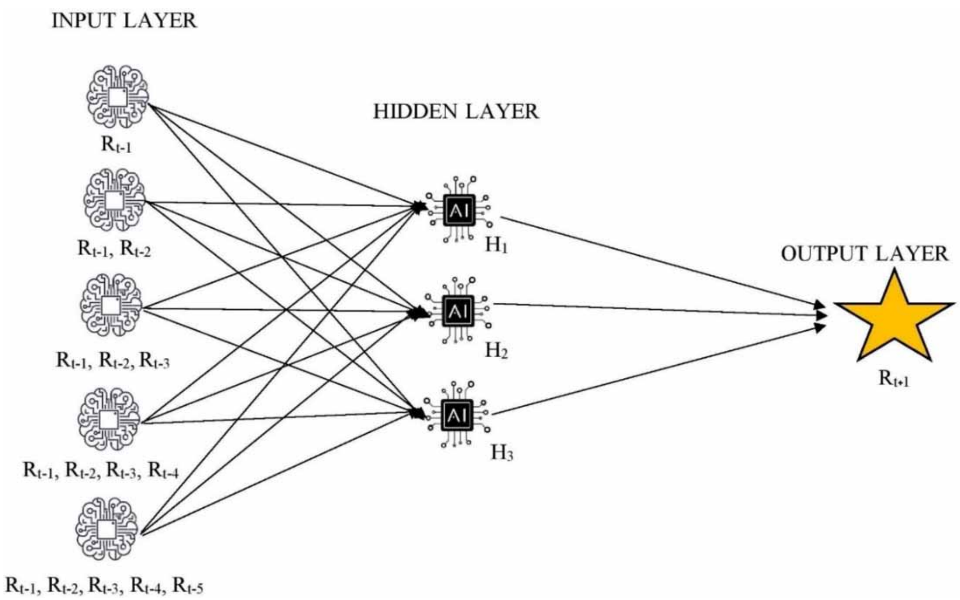

Figure 4 | The structure of the current study's ANN model.

where $R(t)$ is the actual rainfall, $R_L(t)$ is the linear component of actual rainfall, and $R_N(t)$ is the nonlinear component of actual rainfall (Moeeni & Bonakdari 2017). Iteratively fitting the SARIMA model to the data comes first, then an examination of the residuals in the SARIMA-ANN hybrid model. The ANN is then trained using the residuals as the variable, and the SARIMA model is then integrated for the outcome prediction. Using the right metrics, the hybrid model's performance is assessed, and changes can be made as needed to make the model better. Figure 5 shows the structure of the developed SARIMA-ANN model.

2.5. Input selections

Auto-correlation function (ACF) is a statistical technique used to determine how closely a time series of data and its lagged values correlate. In other words, it evaluates how closely a data point resembles its previous values. The correlation coefficient between time series data and a lagged version of that data at various time lags is used to calculate the ACF. The outcome is a collection of correlation coefficients, each of which is associated with a different time lag. The ideal lag value for a time series model can be chosen using ACF to spot patterns and trends in time series data. In this research, ACF was used to find the most correlated lagged data. All the training data in the ANN system were built on 4-lag-time as recommended by ACF.

2.6. Evaluation metrics

Metrics used to assess an ML model's performance are called evaluation metrics. With the support of these measures, it is possible to evaluate the model's effectiveness in terms of accuracy, precision, recall, and other crucial factors. This research depends on two types of evaluation metrics. The first type is RMSE, and the second type is R^2.

2.6.1. Root mean squared error

Root mean squared error (RMSE) is a widely used metric for assessing how well a ML model performs, particularly when doing regression tasks. It calculates the difference between the dataset's actual values and the predicted values.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_o - R_p)^2} \tag{2}$$

where R_o and R_p are observed and predicted rainfall values.

2.6.2. R^2

An evaluation of a regression model's ability to fit the data is done statistically using the R^2 metric. It displays the proportion of the dependent variable's variance that the model's independent variables are responsible for explaining.

$$R^2 = \left\{(1/N) * \sum [(xi - X) * (yi - Y)/(\sigma x * \sigma y\right\}^2 \tag{3}$$

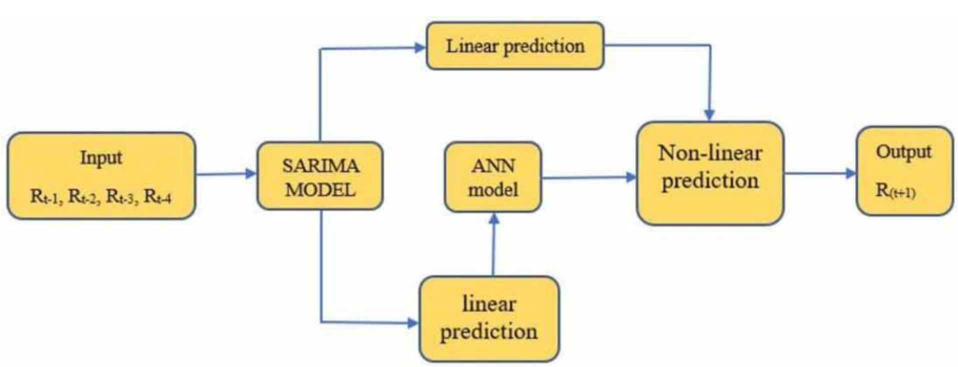

Figure 5 | The structure of the developed SARIMA-ANN model (Moeeni & Bonakdari 2017).

3. RESULT AND DISCUSSION

Predicting the amount of rainfall in Sulaymaniyah after testing the data through the ANN model could not achieve accurate results. In the months when the amount of rainfall in the city was high, the level of prediction was very poor in testing the data through the ANN model. In the next step, SARIMA-ANN was developed, and the best model obtained an accurate rainfall forecasting result.

3.1. ANN results

Initially, the ANN model was used to forecast the amount of rainfall in Sulaymaniyah based on data from previous years. As can be seen in the first testing set (Figure 6), in the first model, it is clear that the prediction is generally positive, and there is no negative prediction. However, the first model could not provide accurate and reliable predictions. The model failed to predict extreme events. The R^2 for the first model is 0.335, which is a high error rate, with RMSE of 55.05 which is also a high error rate. Obviously, the rainfall level in Sulaymaniyah is not stable, so this model could not predict the low and high levels of rainfall, especially in months such as January when usually heavy rains and sometimes floods occur in some parts of the city.

In the second experiment with the ANN model, which is model 2 as shown in Figure 7, it was found that the positive and negative ratios do not represent a good model for predicted rainfall and there is a significant error compared to the first model. In the ANN model, the R_{t-1}, and R_{t-2} were used for the second model as input combination selection. However, the model failed to provide an accurate rainfall forecast. The error rate of the second model is higher than the first model. The R^2 for the second model is 0.362, which is a high error rate and causes the instability of the prediction. The RMSE is 54.312 which is obviously a high error rate compared to the first model. The instability of rainfall and the complexity of the seasons in Sulaymaniyah resulted in high errors in prediction accuracy.

As shown in Figure 8, the error rate for the third model is higher than for the previous models. R_{t-1}, R_{t-2}, and R_{t-3} are used as input combinations for the third model. The ability to predict high levels of rainfall is weak, as shown in Figure 8. There are positive and negative parts. This shows the weaknesses and errors of the current model compared to the previous models. The R^2 for the third model is 0.386 which indicates a high error rate and causes the instability of the prediction. Also, RMSE is 53.292 for the first model which is a high error rate compared to the first and second models.

As shown in Figure 9, the lowest error rate has been obtained compared to the other previous models. The input combination selection for the fourth model was R_{t-1}, R_{t-2}, R_{t-3}, and R_{t-4}. The R^2 for the fourth model is 0.431, which shows a low error rate compared to the other five models. RMSE is 51.002, which shows a similarly low error rate. The availability of appropriate models with the least error will help predict low levels of rainfall in the future in conjunction with high levels of rainfall, which has concluded with the fourth model. The temperature difference in the city was an important factor in the error rate, especially in the first and second months of the year which had an unstable amount of rainfall. In the next

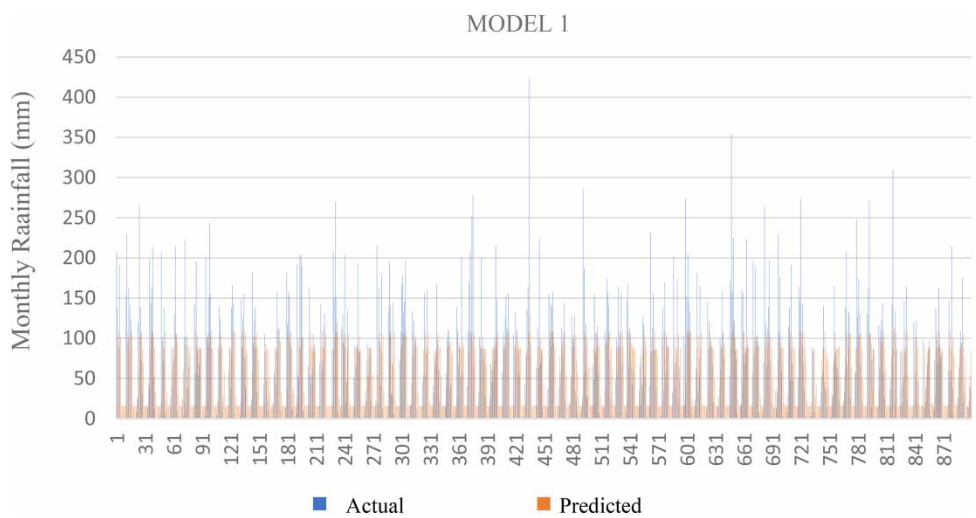

Figure 6 | Rainfall forecast results obtained for the training and testing periods in Model 1.

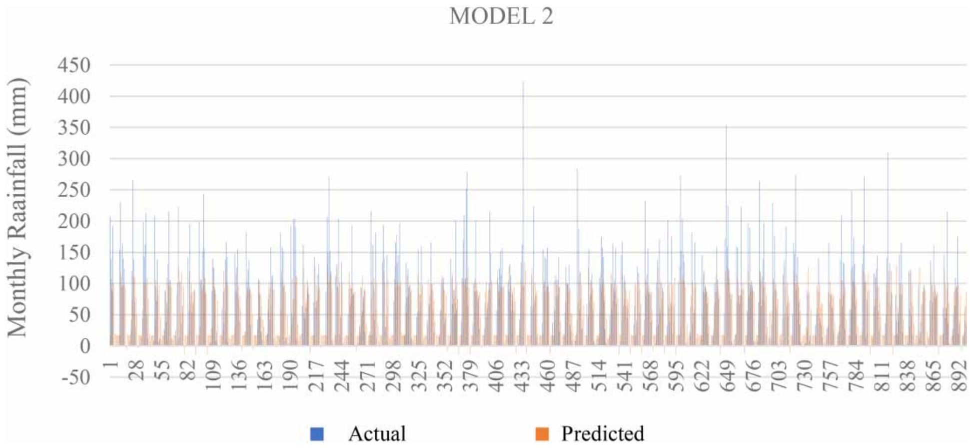

Figure 7 | Rainfall forecast results obtained for the training and testing periods in Model 2.

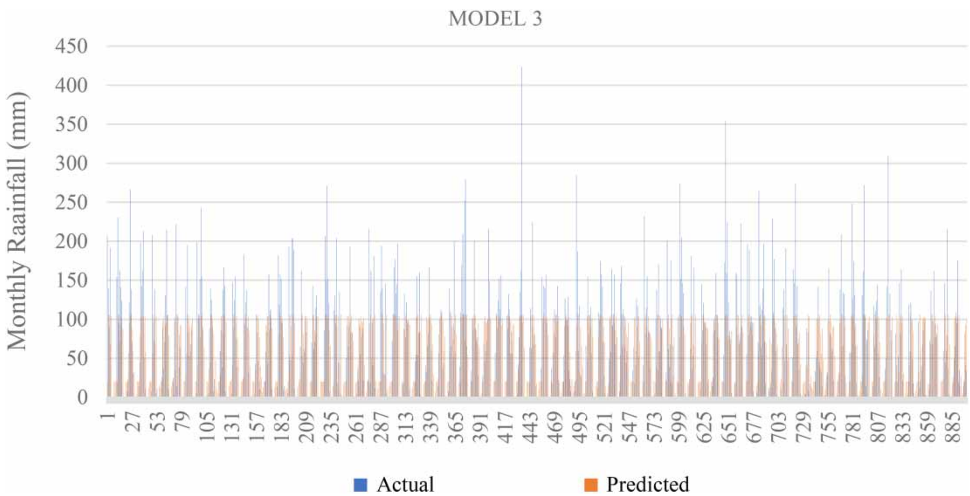

Figure 8 | Rainfall forecast results obtained for the training and testing periods in Model 3.

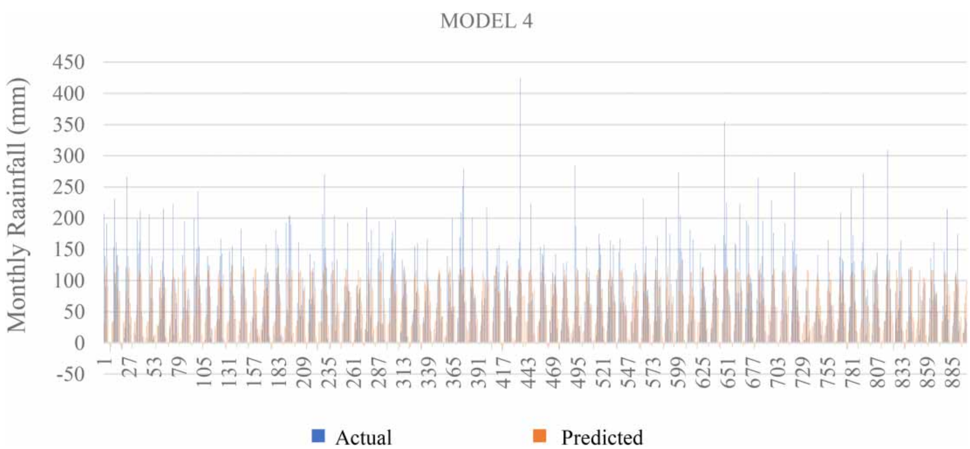

Figure 9 | Rainfall forecast results obtained for the training and testing periods in Model 4.

steps, Model 4 was developed which was an appropriate model compared to the others. The fourth model was successfully developed with the SARIMA-ANN.

It was found that Model 5 has a higher error rate than Model 4. Based on both RMSE and R^2, it showed that the fourth model is the most appropriate among all models. The R^2 for the fifth model is 0.372, which shows a high error rate compared to the other models, especially the fourth. The RMSE for Model 5 is 55.501, which showed another low coefficient compared to the fourth model as shown in Figure 10. According to the results given in Table 1, there are five different models with five different components, and Model 4 shows the better accuracy among the five. Finally, the input combination for Model 4 is selected to be developed utilizing the SARIMA-ANN model.

3.2. SARIMA-ANN results

Since Model 4 outperforms all models based on both RMSE and R^2 metrics, SARIMA-ANN has been developed based on the input combination scenario of Model 4. Table 2 shows the results of the developed SARIMA-ANN model for predicting rainfall. It shows that RMSE is 11.5 and R^2 is 0.98 which outperformed the most accurate model of ANN (Model 4) with a significant difference of RMSE = 51.002 and $R^2 = 0.431$. It shows the successful results of using the SARIMA-ANN model in this study. The SARIMA-ANN model successfully identified the data and accurately worked on it to forecast monthly

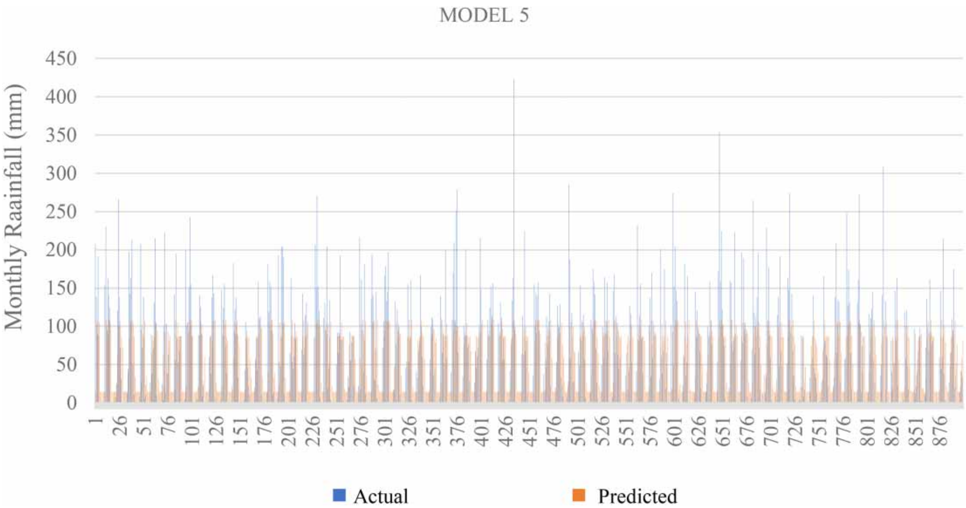

Figure 10 | Rainfall forecast results obtained for the training and testing periods in Model 5.

Table 1 | Represents the results of ANN models in terms of RMSE and R^2

ANN models	Input selection	R^2	RMSE
Model 1	R_{t-1}	0.335	55.05
Model 2	R_{t-1}, R_{t-2}	0.362	54.312
Model 3	R_{t-1}, R_{t-2}, R_{t-3}	0.386	53.292
Model 4	R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}	0.431	51.002
Model 5	R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}	0.372	55.501

Table 2 | Represents SARIMA-ANN results with RMSE, R^2.

SARIMA-ANN	Input selection	R^2	RMSE
Developed model	R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}	0.98	11.5

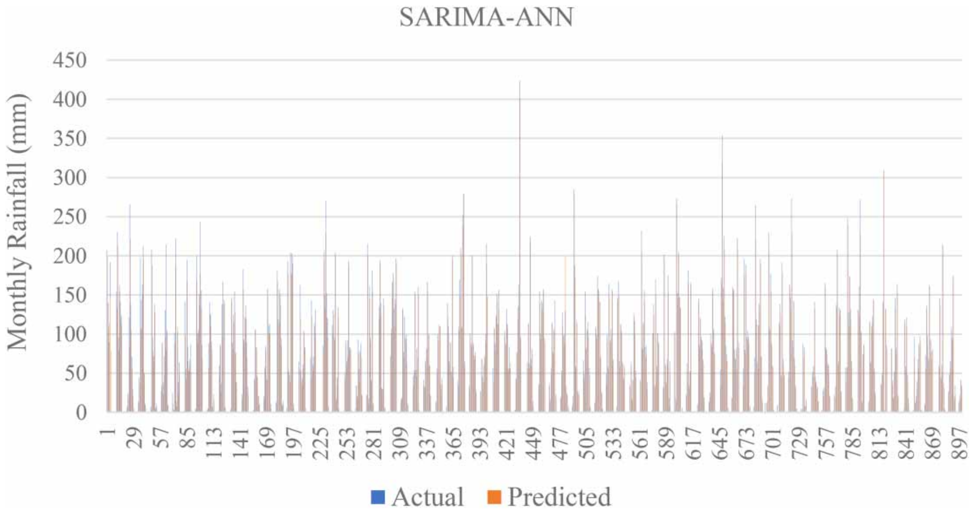

Figure 11 | Rainfall forecast results with SARIMA-ANN model.

rainfall in Sulaymaniyah. The model forecasted low and high levels of rainfall in Sulaymaniyah accurately, which helps sustainable water management.

The main problem in Sulaymaniyah City is that its summers are very hot, and its winters are very cold, and this vast change makes the prediction process harder and contains more errors. SARIMA-ANN can anticipate severe weather, such as heavy floods and drought. Therefore, through this model, weather disasters can be predicted before they occur. Also, this model was able to predict low levels of rain, unlike previous models. Finally, it can be mentioned that the developed SARIMA-ANN model is considered accurate and can be relied upon to predict the rainfall levels of Sulaymaniyah City accurately. Figure 11 represents the performance of the proposed hybrid SARIMA-ANN model.

Based on the findings of the current study, SARIMA-ANN could be very useful for Sulaymaniyah City since it could accurately predict the monthly rainfall data. On the other hand, studies for rainfall prediction in Sulaymaniyah using hybrid models are rarely found in the literature. Therefore, the findings of the current study could fill this gap in the literature.

4. CONCLUSION

The water level of Sulaymaniyah City rises and falls due to climate change. ANN failed to perform well in predicting monthly rainfall in Sulaymaniyah. The main obstacle to the low prediction results was the low quantity of data. The fourth model was selected as the highest accuracy in the ANN model. Later, the input selection scenario of the fourth model was developed using SARIMA-ANN and obtained a high accuracy of RMSE $= 12.209$ and $R^2 = 0.979$. This developed model of the current research could be considered as a suitable tool for managing rainfall in Sulaymaniyah City. It is recommended for future studies to use daily rainfall data since it will help the ML models train better and have more accurate predicted values.

AVAILABILITY OF DATA AND MATERIAL

Data will be made available on reasonable request.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Adede, C., Oboko, R., Wagacha, P. & Atzberger, C. 2015 A mixed model approach to drought prediction using artificial neural networks: Case of an operational drought monitoring environment. *MDPI Remote Sens* **11** (9), 1099.

Ahmed, A. N., Latif, S. D. & El-Shafie, A. 2020 *Research Series in Civil Engineering: Soft Computing Applications in Hydro-Environment*. UNITEN Press, Selangor.

Akram, M. & Hamid, A. 2013 A comprehensive review on water balance. *Biomed. Prev. Nutr.* **3**, 193–195. https://doi.org/10.1016/j.bionut.2012.10.003.

Ali, S. & Shahbaz, M. 2020 Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. *Model. Earth Syst. Environ.* **6**, 1645–1656. https://doi.org/10.1007/s40808-020-00780-3.

Alkaradaghi, K., Ali, S. S., Al-ansari, N. & Laue, J. 2019 Landfill site selection using MCDM methods and GIS in the sulaymaniyahyah governorate, Iraq. *Sustainability*. https://doi.org/10.3390/su11174530.

Barker, A., Pitman, A., Evans, J. P., Spaninks, F. & Uthayakumaran, L. 2021 Drivers of future water demand in Sydney, Australia: Examining the contribution from population and climate change. *J. Water Clim. Chang* **12**, 1168–1183. https://doi.org/10.2166/wcc.2020.230.

Basha, C. Z., Bhavana, N., Bhavya, P. & Sowmya, V. 2020 Rainfall prediction using machine learning deep learning techniques. In: *Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020*. https://doi.org/10.1109/ICESC48915.2020.9155896.

Boretti, A. & Rosa, L. 2019 Reassessing the projections of the world water development report. *npj Clean Water* **2**. https://doi.org/10.1038/s41545-019-0039-9.

Fereshtehpour, M., Javidi Sabbaghian, R., Farrokhi, A., Jovein, E. B. & Ebrahimi Sarindizaj, E. 2021 Evaluation of factors governing the use of floating solar system: A study on Iran's important water infrastructures. *Renewable Energy* **171**, 1171–1187. https://doi.org/10.1016/j.renene.2020.12.005.

Latif, S. D. & Ahmed, A. N. 2023 A review of deep learning and machine learning techniques for hydrological inflow forecasting. *Environ. Dev. Sustainability*. https://doi.org/10.1007/s10668-023-03131-1.

Latif, S. D., Azmi, M. S. B. N., Ahmed, A. N., Fai, C. M. & El-Shafie, A. 2020 Application of artificial neural network for forecasting nitrate concentration as a water quality parameter: a case study of Feitsui Reservoir, Taiwan. *Int. J. Des. Nat. Ecodyn.*. https://doi.org/10.18280/ijdne.150505.

Latif, S. D., Alyaa Binti Hazrin, N., Hoon Koo, C., Lin Ng, J., Chaplot, B., Feng Huang, Y., El-Shafie, A. & Najah Ahmed, A. 2023 Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches. *Alexandria Eng. J.* **82**, 16–25. https://doi.org/10.1016/j.aej.2023.09.060.

Li, H. & Sansalone, J. 2022 Implementing machine learning to optimize the cost-benefit of urban water clarifier geometrics. *Water Res.* **220**, 118685. https://doi.org/10.1016/j.watres.2022.118685.

Malki, Z., Atlam, E. S., Hassanien, A. E., Dagnew, G., Elhosseini, M. A. & Gad, I. 2020 Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos, Solitons Fractals* **138**, 110137. https://doi.org/10.1016/j.chaos.2020.110137.

Moeeni, H. & Bonakdari, H. 2017 Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model. *Stochastic Environ. Res. Risk Assess.* **31**, 1997–2010. https://doi.org/10.1007/s00477-016-1273-z.

Mohammed, M. N., Daham, B. F. A. & Mohammed, K. S. 2018 Darbandikhan reservoir operation optimization using ant colony optimization algorithm. *Zanco J. Pure Appl. Sci.* **30**, 135–141. https://doi.org/10.21271/zjpas.30.s1.15.

Parmar, A., Mistree, K. & Sompura, M. 2017 Machine learning techniques for rainfall prediction: A review. *Int. Conf. Innov. Inf. Embed. Commun. Syst.*, Coimbatore, India, 17-18 March 2017. Curran Associates, Red Hook, NY.

Pham, B. T., Le, L. M., Le, T. T., Bui, K. T. T., Le, V. M., Ly, H. B. & Prakash, I. 2020 Development of advanced artificial intelligence models for daily rainfall prediction. *Atmos. Res.*. https://doi.org/10.1016/j.atmosres.2020.104845.

Shen, C. 2018 A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* **54**, 8558–8593. https://doi.org/10.1029/2018WR022643.

Tinti, A. 2017 *Water Resources Management in the Kurdistan Region of Iraq*, The Institute of Regional and International Studies.

Vieira, A. d. N., Kleinermanns, K., Martin, W. F. & Preiner, M. 2020 The ambivalent role of water at the origins of life. *FEBS Lett.* **594**, 2717–2733. https://doi.org/10.1002/1873-3468.13815.

Yang, T. H. & Liu, W. C. 2020 A general overview of the risk-reduction strategies for floods and droughts. *Sustainability* **12**, 1–20. https://doi.org/10.3390/su12072687.

Zubaidi, S. L., Ortega-Martorell, S., Al-Bugharbee, H., Olier, I., Hashim, K. S., Gharghan, S. K., Kot, P. & Al-Khaddar, R. 2020 Urban water demand prediction for a city that suffers from climate change and population growth: Gauteng province case study. *Water (Switzerland)* **12**, 1–17. https://doi.org/10.3390/W12071885.

First received 7 January 2024; accepted in revised form 10 March 2024. Available online 21 March 2024