University of Baghdad
College of Engineering
JOURNAL OF ENGINEERING

Journal of Engineering
journal homepage: www.jcoeng.edu.iq

Volume 31 Number 5 May 2025

Evaluation of Data Mining and Artificial Intelligence Methods to Predict Daily Precipitation

Yaseen Ahmed Hamaamin

Department of Civil Engineering, College of Engineering, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

Precipitation is the most important weather parameter, which has direct and indirect effects on life on our planet. One of the most relevant climate change consequences is extreme weather conditions, such as floods and droughts. Up to now, climate model projections have remained uncertain, therefore, more accurate precipitation modelling techniques are necessary. Due to the complexity of relationships among metrological parameters, traditional statistical modelling tools are ineffective in forecasting and predicting weather conditions. In this study, data mining techniques were applied to metrological data to predict daily precipitation using Multilinear regression (MLR) along with two artificial intelligence (AI) techniques, specifically Artificial Neural Networks (ANN) and Neuro-Fuzzy Inference System (ANFIS). A total of 10 daily metrological variables, for 13 years, namely Maximum temperature (Tmax), minimum temperature (Tmin), maximum humidity (Hmax), minimum humidity (Hmin), wind speed (Ws), wind direction (Wd), cloud cover (Cv), sea pressure (SEAp), station pressure (STAp) and relative humidity (RH) are used to predict precipitation (P). Both AI systems showed acceptable results predicting daily precipitation from observed meteorological parameters with a coefficient of determination (R^2) of 0.75 and 0.72 for model calibration of ANN and ANFIS methods, respectively. The results of the ANN and ANFIS testing methods were 0.55 and 0.62, respectively. Outcomes of the study showed that the ANN model may have overfitted the results in the calibration section of the process compared to the ANFIS method, which performed better in the testing section of the evaluation process. In ANFIS modelling, for several input variables up to 6 variables, this study recommends using the grid partition method to divide variable ranges into membership functions. For input variables of more than 6 variables, sub-clustering method is recommended.

Keywords: Data mining, AI, ANN, ANFIS, Precipitation, Model.

---

*Corresponding author
Peer review under the responsibility of University of Baghdad.
https://doi.org/10.31026/j.eng.2025.05.06

Article received: 16/09/2024
Article revised: 07/02/2025
Article accepted: 09/03/2025
Article published: 01/05/2025

1. INTRODUCTION

Water is essential for life; without a sufficient amount of clean freshwater, the lives of the majority of organisms will be impossible. Precipitation is a significant aspect of the water cycle, which recharges freshwater resources. Climate change substantially impacted on water cycle, altering evaporation rate and precipitation location, amount, and patterns. Consequently, vigorous management of rainfall is the key factor to adapt to those changes **(Lind, 2023)**. Climate change has made severe weather events such as floods and droughts more frequent, which have profound impacts on agriculture, water resource management, and disaster mitigation **(Hamdan, 2024)**. As climate conditions change, traditional modelling methods are not able to accurately predict weather conditions; therefore, for sustainable stormwater management, robust precipitation forecasting techniques are required **(Jayasingh and Mantri, 2019)**. Data mining applications have been proven to be powerful in this task, which extracts the hidden patterns between metrological measured parameters. In the last decade, much research has been performed to increase the accuracy of rainfall prediction using data mining techniques **(Aftab et al., 2018)**. Data mining tools apply a mixture of machine learning (ML) methods, statistical approaches, and AI systems on recorded weather data collected in weather stations to recognize trends and then predict upcoming precipitation events **(Liyew and Melese, 2021)**.

(Kisi et al., 2013) obtained improved results in modeling rainfall-runoff from different artificial intelligence (AI) techniques, ANN, Adaptive Neuro-Fuzzy Inference System (ANFIS), and Gene Expression Programming (GEP) approaches compared to multiple linear regression (MLR). ANN models were successfully used to predict monthly rainfall from eight weather stations in different locations of the Sulaymaniyah governorate **(Suhaili and Karim, 2014)**. Linear, polynomial, and radial basis functions are applied to a support vector regression (SVR) to predict the monthly amount of precipitation. The results of the research confirmed that the ANFIS model was superior to the SVR model **(Shamshirb et al., 2014)**. A fuzzy Logic (FL) rule-based system is used to map the nonlinear relationship between two input variables of temperature and wind speed to predict rainfall to improve agricultural management **(Rahman, 2020).** The applicability of four different artificial neural network (ANN) models was tested to predict rainfall-runoff. ANN models exhibit acceptable ability in predicting and modelling non-linear hydrological correlations. In addition to more predictors in the models, this results in more accurate results **(Aoulmi et al., 2021)**. Modelling of rainfall is a difficult task due to the dynamic nature of climatic events. An ML-based feature selection technique was used successfully to find the key meteorological factors that contribute to accurate rainfall predictions **(Pathan et al., 2021)**. ANN and multi-linear regression (MLR) models were used to estimate rainfall-runoff in Turkey. Optimum ANN models were obtained by using three different transformation functions, which were found to be superior in predicting rainfall runoff compared to the other methods **(Turhan, 2021)**. Autocorrelation, along with several ML techniques used to predict weather patterns, namely, Linear, Exponential, Quadratic, Additive Seasonality, Additive Seasonality Quadratic Trend, Multiplicative Seasonality, and Multiplicative Seasonality Linear Trend, were tested to model weather parameters **(Gore and Gawali, 2023)**. A hybrid framework named composite of the least absolute shrinkage and selection operator (LASSO), ANN, and support vector machine (SVM) was used to predict daily runoff in Pakistan. The hybrid system is efficient in modeling runoff perfectly, which is important in long-term water management and protection **(Shabbir and Chand, 2023)**. Proper water resources planning and management demand effective precipitation forecasting. Soft computing tools, which are inspired by human reasoning and intuition, offer a more efficient and time-saving approach compared to traditional mathematical techniques. Soft computing techniques, specifically ANN, Genetic Algorithms (GA), Support Vector Regression (SVR), and FL used to model rainfall and runoff successfully **(Sammen et al., 2023)**. Data mining algorithms can analyze historical precipitation data to predict future rainfall patterns in agricultural planning and water resource management **(Li et al., 2024)**.

In this study, data mining practice is applied to daily meteorological measured data to find an efficient modeling technique to estimate the non-linear relationship to predict daily precipitation effectively. A set of 10 metrological records for 13 years was used to estimate the daily amount of precipitation using MLR, besides two AI techniques, specifically ANN and ANFIS. The daily meteorological data was analyzed for the hidden relationships between predictors and the output variables to find a robust modeling method to forecast the precipitation.

2. MATERIALS AND METHODS

2.1 Data Collection

Metrological weather data were used from the Sulaymaniyah weather station located at coordinates of Latitude 35.55580 N and Longitude 45.45320^{o} E with an elevation of 884.8m. The average annual rainfall of the station is about 700mm. The area climate is the Mediterranean climate, which is hot and dry in summer and wet and cold in winter, with a yearly average temperature of 19.59 ^{o}C **(Hamaamin, 2017; Abdullah and Hamaamin, 2020; Hamaamin and Faraj, 2023; Rashid, 2024)**. From historical metrological records of the station, 10 weather parameters were used to predict daily precipitation. Maximum temperature (Tmax)in (^{o}C) units, minimum temperature (Tmin) in (^{o}C) units, maximum humidity (Hmax) (%), minimum humidity (Hmin) (%), wind speed (Ws) in (m/s) units, wind direction (Wd) (with north), cloud cover (Cv) in (oktas) units, sea pressure (SEAp) in (millibar) units, station pressure (STAp) in (millibar) units and relative humidity (RH) in (%) used to predict precipitation (P) in (mm) units.

2.2 Data Preparation

In this study, 13 years (2010 to 2023) of daily weather records from the Sulaymaniyah weather station were used. The 10 independent metrological parameters are used to predict the output parameter of precipitation. The data mining phases were shortened into the following main steps **(Mucherino et al., 2009; Judd et al., 2017; Gupta, 2024)**:

2.2.1. Cleaning and Integration of the Data Set

At the very beginning of the process, all incomplete data points were removed, and any data point or daily records lacking one of the selected 11 parameters were removed. The selected remaining data points were collected for the next step of the data mining process.

2.2.2. Reduction of the Data Set

In the second stage of the data preparation, data points that are expected to make noise and do not add any useful information to the data-driven process are removed. Consequently, for each year, dry months are excluded from data set, specifically June, July, August, and September.

2.2.3. Data Processing and Assumptions

In this step, many tasks were done in data mining, including data transformation, pattern analysis, correlation, modeling, and reporting the performance of the methods. Unfortunately, the output of these data processing steps was not different from the regular regression shown in the results section. Afterwards, nonlinear modeling (ANN and ANFIS) techniques were used to find these methods' ability to model precipitation. In this research, the performance of different models was evaluated using a coefficient of determination (R^2) and root mean square errors (RMSE).
The collected 13 years of data points were divided into two main sets, training (calibration) and testing (validation) sets. Data points from 10 years were used as a training set, and 3 years' data points were used as a testing set.

2.3 ANN Method

Artificial Neural Network (ANN) is a computational modeling process loosely inspired by biological neural networks. This modeling system consists of three main layers: input layer, hidden layers, and output layer, as shown in **Fig. 1**. The hidden layer can be a single layer or multiple layers. In the machine learning process, usually, historical input and output data are available, and then the system tries to extract an efficient weight to map the input to the output, which is known as the data-driven method. The resulting mapping function can be saved for future predictions using new input data to estimate the output data. In an artificial neural network, the neurons in the input layer transfer data via synapses to the next layer (hidden layer), and similarly, the hidden layer transfers this data to the last layer (output layer). This process in the ANN system is performed in the hidden layers by applying a weighting function during the training process, in which the relationship between the input to obtain the output values is extracted. The training process could be forward, mapping from the input to the output, or backward, in which output values can be used to update or train the weighting values in a backward process or training. Hidden layers process the signals from input to output implicitly, which is why Neural Network Modeling is called a black box **(Beale et al., 2017; Han et al., 2018; Astutik et al., 2021)**.

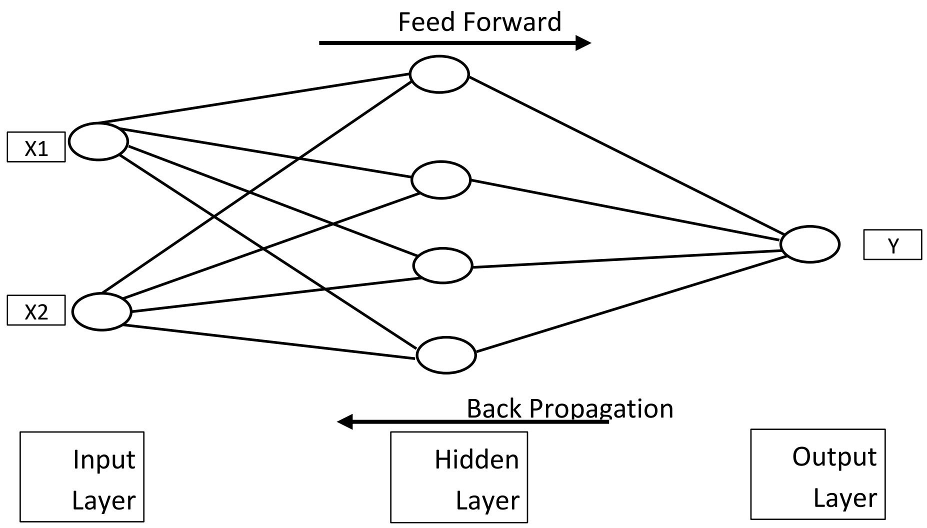

Figure 1. ANN model architecture **(Han et al., 2018)**.

2.4 ANFIS Method

Adaptive Neuro-Fuzzy Inference System is a hybrid of fuzzy logic with an Artificial Neural Network, which combines the goodness of both techniques in a single modelling system. ANFIS contains all steps of Sugeno fuzzy logic from fuzzification, inference, and defuzzification of the data, enhanced with the ANN's tuning ability of the inference system parameters **(Hamaamin et al., 2022)**. ANFIS automatically selects the type and shape of the membership function, which is useful in situations when we do not have any previous information about the data set distribution **(MathWorks, 2018)**. The parameters tuning process is in an iterative manner, which consists of two parts, forward and backward. During each iteration, the previous estimated parameters will be updated through backpropagated errors to update the parameters for the next step until the desired error is reached, the ANFIS model architecture in **Fig. 2**.
The input data is divided into at least two subsets, which are used to generate fuzzy membership functions. In MATLAB, there are two methods to obtain subsets: the grid partitioning method and the clustering method, using MATLAB's functions of genfis1 and genfis2, respectively **(Hamaamin et al., 2018)**. In this study, due to a high number of data points clustering method was used to create the subsets using the genfis2 function.

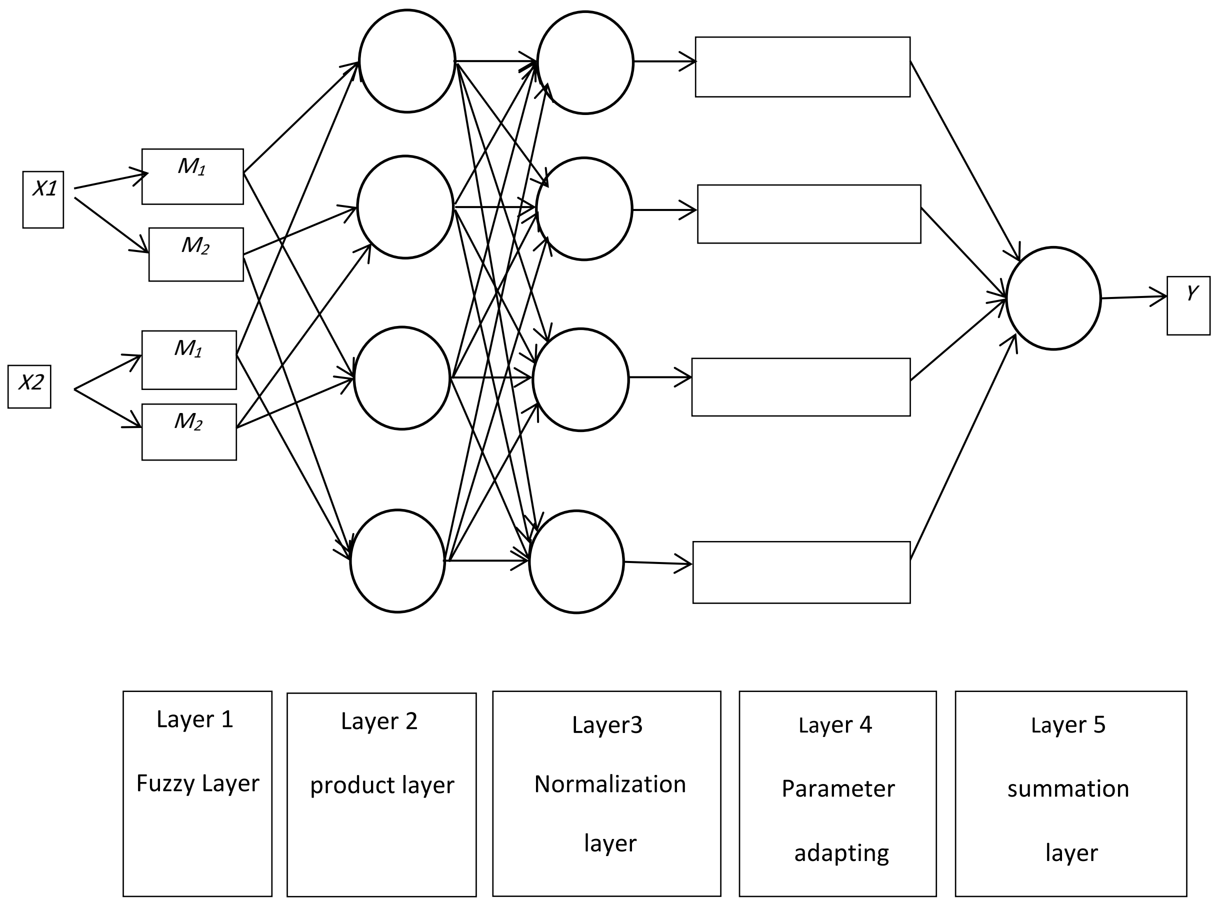

Figure 2. ANFIS model architecture **(Hamaamin et al., 2022)**.

3. RESULTS AND DISCUSSIONS

In this paper, 10 daily meteorological measures are used to predict daily precipitation. A total of 13 years of daily data from the Sulaymaniyah weather station was used in the modeling process. For the modelling process, the 13 years' data points were split into two sets, 10 years as the calibration set and 3 years as a validation set.

3.1 Data Mining

The data mining process started with removing dry months of June, July, August, and September and data cleaning, in which complete records were removed, as a result, a total of 4827 daily records remained. For modeling purposes, the data was split into two sets, 85% for training and 15% for testing. **Table 1** shows descriptive statistics of all variables along with the normality test of the variables data points, the table confirms that all variables have non-normal distribution from the Anderson Darling (AD) normality test p-value (The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05) **(Ott and Longnecker, 2015)**. In **Table 1**, StDev is the standard deviation of the data (with the same units of each variable), Q1 is the first quartile, and Q3 is the third quartile.

Table 1. Descriptive statistics and normality test of all variables.

Variable	Mean	StDev	Min	Q1	Median	Q3	Max	AD (p-value)
Tmax(°C)	21.586	9.245	-0.2	14.5	20	28.3	47	<0.005
Tmin(°C)	10.501	7.305	-7.5	4.9	9.5	15.8	31.6	<0.005
Hmax(%)	72.695	18.821	22	57	76	89	100	<0.005
Hmin(%)	39.386	18.302	8.1	25	36	51	93	<0.005
Ws(m/s)	1.2432	1.1509	0	0.5	1	1.6	11.8	<0.005
Wd(with north)	179.64	104.28	0	100	200	270	360	<0.005
Cv	3.0274	2.4376	0	1	3	5	8	<0.005
SEAp(millibar)	1017.6	6.76	996.2	1013.1	1018	1022.7	1055.9	<0.005
STAp(millibar)	919.48	6.61	903.5	914.9	918.6	923.8	938.9	<0.005
RH(%)	56.456	19.098	6.7	40	56	72	98	<0.005
P(mm)	2.745	8.26	0	0	0	0.6	131.8	<0.005

Fig. 3 shows a box plot of all variables to check the linearity and normality of the parameters; none of the variables has a normal distribution. This means that linear modelling, such as multiple linear regression, would not be expected to give reasonable results. The non-normality of all variables and the existence of outliers for some variables **Fig. 3** would worsen modelling outcomes and the linear modelling performance.

A multilinear regression model (MLR) was generated to predict P from the set of 10 predictors, but acceptable results were not obtained (R^2=0.29), which is very low in terms of model prediction. The regression equation is shown in Eq. (1), and the performance of the regression model is in **Table 2**, which shows R^2, adjusted R^2 (R^2(adj)), and prediction R^2 (R^2(pred)).

$$P=178-0.0814\,Tmax+0.1262\,Tmin-0.031\,Hmax+0.0827\,Hmin+0.576\,Ws$$
$$+0.00287\,Wd+0.863\,Cv-0.1699\,SEAp-0.014\,STAp+0.1019\,RH \qquad (1)$$

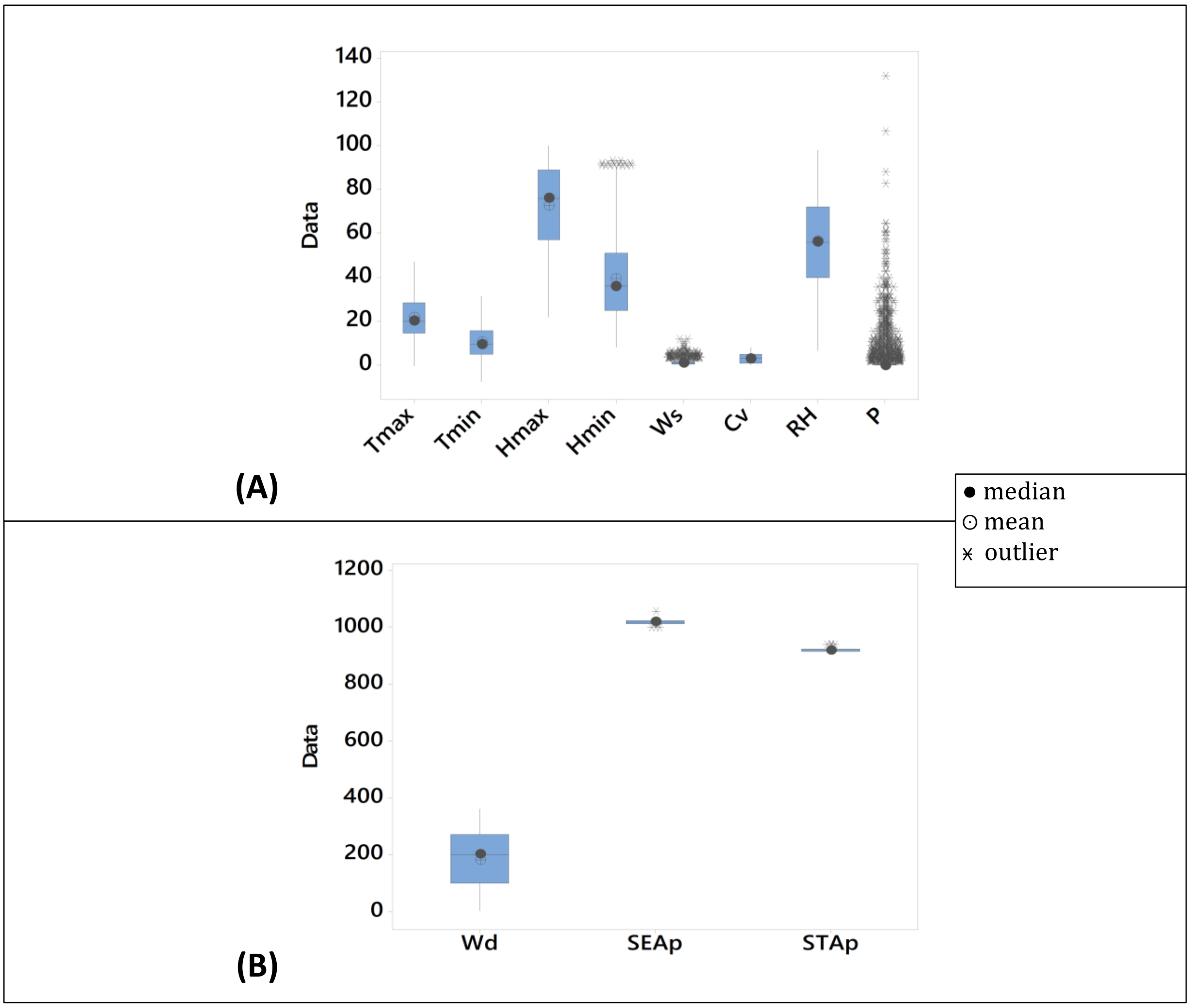

Figure 3. Box plots of all variables.

Table 2. Summary of the regression model.

Data	R^2	R^2(adj)	R^2(pred)	RMSE
Training	0.2936	0.2906	0.2853	353.7866
Testing	0.5600	0.3200	0.1110	105.4691

Correlation is the measure of how strongly two variables are related to each other; in other words, it shows how the two variables change together at a constant rate. Usually, correlation takes values from (-1 to 0) or (0 to 1), where (-1 or 1) is strongest and (0) is the weakest correlation. **Table 3** shows a correlation between all the variables, where Cloud cover (Cv) has the highest correlation value of 0.49 with the precipitation (P), while the lowest correlation value of 0.01) for wind direction (Wd) with the precipitation (P). Referring to regression Eq.1 and the correlation values in **Table 3**, a stepwise regression method was performed to obtain better results of an enhanced regression model.

Table 3. Correlation between all variables.

Variables	Tmax	Tmin	Hmax	Hmin	Ws	Wd	Cv	SEAp	STAp	RH	P
Tmax	-0.78	-0.64	0.94	0.94	-0.22	-0.08	0.60	0.51	0.24	1.00	0.40
Tmin	-0.30	-0.43	0.28	0.17	0.00	-0.19	-0.10	0.69	1.00	0.24	-0.09
Hmax	-0.73	-0.82	0.54	0.42	-0.16	-0.18	0.02	1.00	0.69	0.51	-0.02
Hmin	-0.43	-0.17	0.52	0.64	-0.03	-0.03	1.00	0.02	-0.10	0.60	0.49
Ws	0.15	0.14	-0.06	-0.09	-0.07	1.00	-0.03	-0.18	-0.19	-0.08	0.01
Wd	0.08	0.18	-0.24	-0.14	1.00	-0.07	-0.03	-0.16	0.00	-0.22	0.04
Cv	-0.74	-0.54	0.82	1.00	-0.14	-0.09	0.64	0.42	0.17	0.94	0.44
SEAp	-0.77	-0.69	1.00	0.82	-0.24	-0.06	0.52	0.54	0.28	0.94	0.33
STAp	0.90	1.00	-0.69	-0.54	0.18	0.14	-0.17	-0.82	-0.43	-0.64	-0.08
RH	1.00	0.90	-0.77	-0.74	0.08	0.15	-0.43	-0.73	-0.30	-0.78	-0.25
P	-0.25	-0.08	0.33	0.44	0.04	0.01	0.49	-0.02	-0.09	0.40	1.00

Also, several possible data transformation techniques were tried to obtain better results, but the results were still around what is presented in **Table 2** without any reasonable improvements. Afterwards, nonlinear artificial intelligence (AI) techniques of ANN and ANFIS were used to model P from the previously stated 10 predictors.

3.2 ANN Model

The proceeded 13 years' data set was randomly divided into 85% training and 15% testing data. The training 85% of the data feed to MATLAB's Neural Network Toolbox (nftool) to model precipitation (P) out of the 10 predictors (Tmax, Tmin, Hmax, Hmin, Ws, Wd, Cv, SEAp, STAp, and RH). After several trials of modeling the P by changing different modeling parameters and options, the best model was obtained using 14 hidden layers and the training algorithm of the Bayesian Regularization algorithm with 1000 iterations. Plots of training (calibration) and testing (validation) from the ANN model are shown in **Figs. 4 and 5**. The performance of the ANN method is shown in **Table 4**.

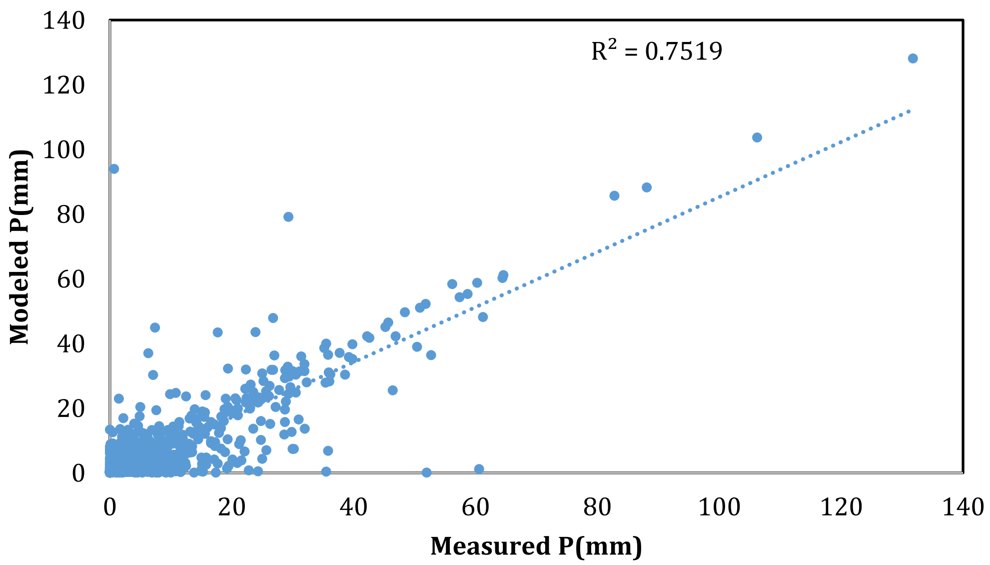

Figure 4. Training analysis of modeled and observed precipitation using an ANN model.

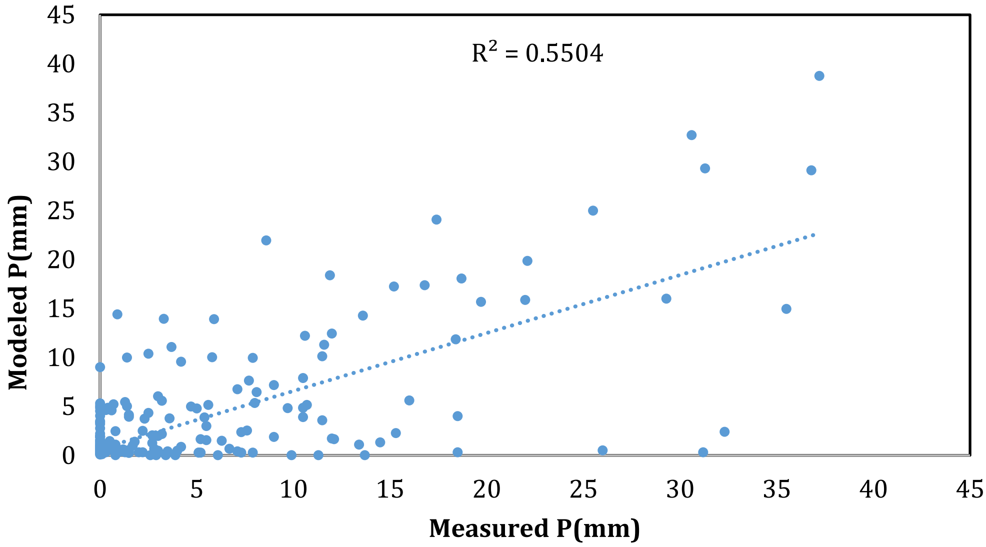

Figure 5. Testing analysis of modeled and observed precipitation using the ANN model.

Table 4. Performance of the ANN model.

Data	R^2	R^2(adj)	R^2(pred)	RMSE
Training	0.7519	0.7518	0.7495	4.376
Testing	0.5504	0.5493	0.5432	5.466

According to **(Moriasi et al., 2007)** coefficient of determination (R^2) describes the proportion of the variance in measured data explained by the model, which ranges between 0 and 1. The value of R^2 =1 describes the best performance of models. Typically, for hydrology and water quality models, R^2 values of more than 0.5 can be considered acceptable.

3.3 ANFIS Model

The 13 years' process split data was used to model P using ANFIS's toolbox of MATLAB. The 85% training data was uploaded to MATLAB's Fuzzy Logic Toolbox Neuro-Fuzzy Designer (Neuro-Fuzzy Designer). Due to the high number of data points, to generate the Fuzzy Inference System, a sub-clustering method is used to divide (partition) the data points into membership functions. The other method of dividing (grid partitioning) is useful for up to 6 variables with a relatively low number of data points. Sub-clustering method was used to divide the data points into membership functions with a range of influence of 0.3, squash factor of 1.25, acceptance ratio of 0.8, and reject ratio of 0.15. The best model for predicting was obtained after 200 epochs (iterations); the results of the ANFIS model are shown in **Table 5**. **Figs. 6 and 7** show the training and testing performance of the ANFIS model.

Table 5. Performance of the ANFIS model

Data	R^2	R^2(adj)	R^2(pred)	RMSE
Training	0.7269	0.7267	0.7234	4.479
Testing	0.6123	0.6114	0.6017	3.967

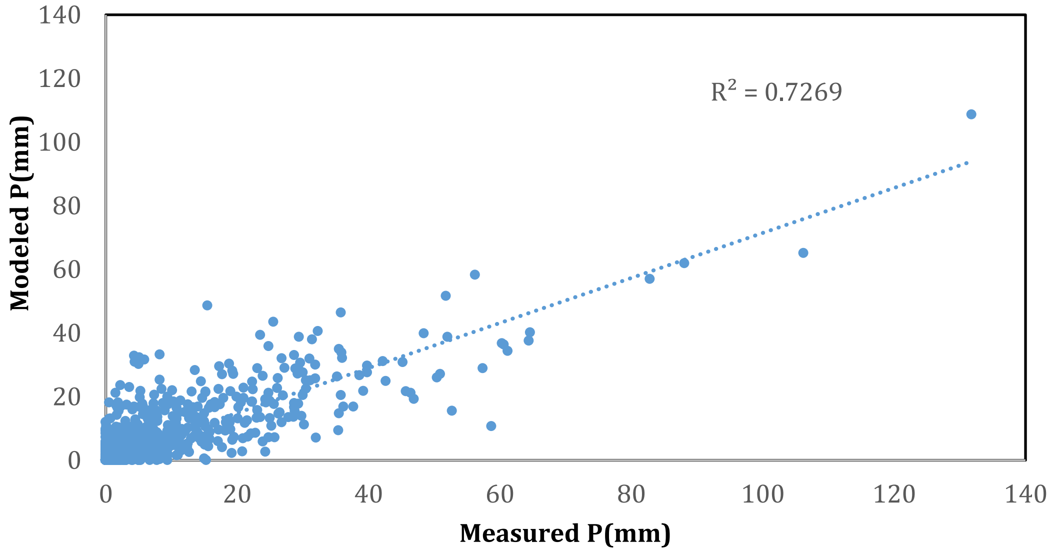

Figure 6. Training analysis of modeled and observed precipitation using ANFIS model.

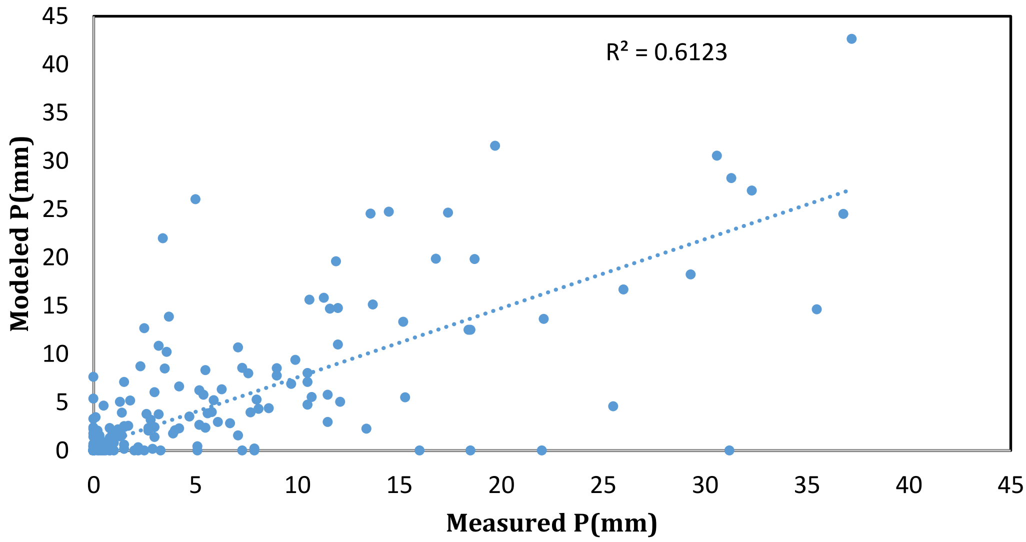

Figure 7. Testing analysis of modeled and observed precipitation using the ANFIS model.

In general performance of both AI methods is acceptable in predicting precipitation P. However, staring and comparing results from both ANN and ANFIS models, one can conclude that the performance of the ANFIS model is better, especially in the testing of the model, where the R^2=0.612 for the ANFIS using the testing part of the data compared to R^2=0.550 for the ANN model. Also, the values of RMSE of the ANFIS model are less than the ANN model, which confirms a better prediction of the ANFIS model. The drawback of the ANN model may be due to the overfitting of the results during training. the overfitting often occurs during the ANN training process, which results in high accuracy in correlating the training data but poor prediction performance **(Chu et al., 2021)**.

4. CONCLUSIONS AND RECOMMENDATIONS

Whereas precipitation is the key factor in water resources management, better prediction of this important factor is vital for decision-makers and watershed managers. In this study, 10 daily meteorological parameters were used to predict daily precipitation. Due to the complexity and nonlinear relationship between variables, modeling precipitation is not an easy task. Many data mining approaches were tested in this study to model precipitation accurately, but the obtained linear model was not reasonable. After that, two AI modeling techniques were tested to predict precipitation. Results from both methods were acceptable, with R^2 =0.752 and 0.727 for calibration (training) of ANN and ANFIS models, respectively. Although R^2 results of the ANN model were better than the ANFIS model, it seems to be overfitting the results, which this confirmed in the validation (testing) part of the models, where new data was used to test them. The R^2 results of the testing of the two methods were 0.550 and 0.612 for ANN and ANFIS models, respectively. In conclusion, the ANFIS model can be selected as the best model to predict daily precipitation.

This study recommends that, in ANFIS modeling, for a high number of data points and several variables more than 6, to prevent a huge number of fuzzy inference rules, a sub-clustering method must be used. The other method of dividing the data into membership functions, namely grid partitioning, is useful for up to 6 input variables. Sub-clustering method was used to divide the data points into membership functions with a range of influence of 0.3, squash factor of 1.25, acceptance ratio of 0.8, and reject ratio of 0.15.

Acknowledgements

The author would like to extend sincere thanks to the Directorate of Meteorology and Seismology in Sulaymaniyah, KRG, Iraq, for providing the daily meteorological data used in this study.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abdullah, J. B., and Hamaamin, Y. A., 2020. Temporal variation of drinking water quality parameters for Sulaimani city, Kurdistan region, Iraq. *UHD Journal of Science and Technology*, 4(2), pp.99-106. https://doi.org/10.21928/uhdjst.v4n2y2020.pp99-106

Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., Nawaz, Z., 2018. Rainfall prediction using data mining techniques: A systematic literature review. *International journal of advanced computer science and applications*, 9(5). https://dx.doi.org/10.14569/IJACSA.2018.090518

AL-Suhaili, R. H., and Karim, R. A., 2014. Spatial prediction of monthly precipitation in Sulaimani governorate using artificial neural network models. *Journal of Engineering*, 20(03), pp.15-27. https://doi.org/10.31026/j.eng.2014.03.02

Aoulmi, Y., Marouf, N., Amireche, M., 2021. The assessment of artificial neural network rainfall-runoff models under different input meteorological parameters: Case study: Seybouse basin, Northeast Algeria. *Journal of Water and Land Development*, pp. 38-47. https://doi.org/10.24425/jwld.2021.138158

Beale, M. H., Hagan, M. T., and Demuth, H. B., 2017. Neural network toolbox. User's Guide. The MathWorks, Inc.

Chu, J., Liu, X., Zhang, Z., Zhang, Y., and He, M., 2021. A novel method overcoming overfitting of artificial neural networks for accurate prediction: application on thermophysical properties of natural gas. *Case Studies in Thermal Engineering*, 28, P. 101406. https://doi.org/10.1016/j.csite.2021.101406

Frost, J., 2019. Introduction to statistics: An intuitive guide for analyzing data and unlocking discoveries. Amazon: Bellevue, WA, USA, 2020.

Frost, J., 2020. Hypothesis Testing: An Intuitive Guide for Making Decisions. Statistics by Jim Publishing; Amazon: Bellevue, WA, USA, 2020.

Gore, R. D., and Gawali, B. W. 2023. Analysis of weather parameters using machine learning. *In the First International Conference on Advances in Computer Vision and Artificial Intelligence Technologies*. (ACVAIT 2022) (pp. 569-589). Atlantis Press. https://doi.org/10.2991/978-94-6463-196-8_44

Gupta, G. K., 2014. Introduction to data mining with case studies. PHI Learning Pvt. Ltd. ISBN-13 : 978 8120350021.

Hamaamin, Y. A., and Faraj, B.A., 2023. Optimal locations for BMPs stormwater management for Sulaymaniyah city industrial zone, KRG, Iraq. *Environment and Ecology Research* 11(3), pp.411-420. https://doi.org/10.13189/eer.2023.110301

Hamaamin, Y. A., 2017. Developing of rainfall intensity-duration-frequency model for Sulaimani city. Journal of Zankoy Sulaimani, 19(3-4), pp.10634. http://dx.doi.org/10.17656/jzs.10634

Hamaamin, Y. A., Rashed, K. A., Ali, Y. M., and Abdalla, T. A., 2022. Evaluation of ANFIS and regression techniques in estimating soil compression index for cohesive soils. *Journal of Engineering*, 28(10), pp.28-41. https://doi.org/10.31026/j.eng.2022.10.03

Hamdan, A., Ibekwe, K. I., Etukudoh, E. A., Umoh, A. A., and Ilojianya, V. I., 2024. AI and machine learning in climate change research: A review of predictive models and environmental impact. *World Journal of Advanced Research and Reviews*, 21(1), pp. 1999-2008. https://doi.org/10.30574/wjarr.2024.21.1.0257

Han, S. H., Kim, K. W., Kim, S., and Youn, Y. C., 2018. Artificial neural network: Understanding the basic concepts without mathematics. *Dementia and neurocognitive disorders*, 17(3), pp.83-89. https://doi.org/10.12779%2Fdnd.2018.17.3.83

Judd, C. M., McClelland, G. H., and Ryan, C. S., 2017. Data analysis: A model comparison approach to regression, ANOVA, and beyond. Routledge. ISBN 9781138819832.

Kisi, O., Shiri, J., and Tombul, M., 2013. Modeling rainfall-runoff process using soft computing techniques. *Computers & Geosciences*, 51, pp.108-117. https://doi.org/10.1016/j.cageo.2012.07.001

Lind, P., 2023. Kilometer-scale climate modeling of precipitation in the Nordic region, PhD dissertation, Department of Meteorology, Stockholm University.

Li, H., Li, S., and Ghorbani, H., 2024. Data-driven novel deep learning applications for the prediction of rainfall using meteorological data. *Frontiers in Environmental Science*, *12*, 1445967. https://doi.org/10.3389/fenvs.2024.1445967

Liyew, C. M., and Melese, H. A., 2021. Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, 8, pp.1-11. https://doi.org/10.1186/s40537-021-00545-4

Liu, H., Fung, J. C., Lau, A. K., and Li, Z., 2024. Enhancing quantitative precipitation estimation of NWP models with fundamental meteorological variables and a transformer-based deep learning model. *Earth and Space Science*, 11(4), e2023EA003234. https://doi.org/10.1029/2023EA003234

Mucherino, A., Papajorgji, P., and Pardalos, P. M., 2009. *Data mining in agriculture*. Springer Science & Business Media. eBook ISBN 978-0-387-88615-2

MathWorks, 2018. MATLAB Fuzzy Logic Toolbox User's Guide2, The MathWorks, Inc.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900.

Ott, R. L., and Longnecker, M. 2015. An introduction to statistical methods and data analysis. *Cengage Learning Inc.* ISBN-13: 978-1305269477

Pathan, M. S., Wu, J., Lee, Y. H., Yan, J., and Dev, S., 2021. Analyzing the impact of meteorological parameters on rainfall prediction. *In 2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*, pp. 100-101. https://doi.org/10.23919/USNC-URSI51813.2021.9703664

Rahman, M. A., 2020. Improvement of rainfall prediction model by using fuzzy logic. *American Journal of Climate Change*, 9(4), pp.391-399. https://doi.org/10.4236/ajcc.2020.94024

Rashid, H. M., 2024. Drought assessment based on different meteorological drought indices in Sulaymaniyah governorate, KRG, Iraq. *Journal of Engineering*, 30(9), pp.190-215. https://doi.org/10.31026/j.eng.2024.09.10

Raval, M., Sivashanmugam, P., Pham, V., Gohel, H., Kaushik, A., and Wan, Y., 2021. Automated predictive analytics tool for rainfall forecasting. *Scientific Reports*, 11(1), p.17704. https://doi.org/10.1038/s41598-021-95735-8

Rojas-Campos, A., Langguth, M., Wittenbrink, M., and Pipa, G., 2023. Deep learning models for generation of precipitation maps based on numerical weather prediction. *Geoscientific Model Development*, 16(5), pp.1467-1480. https://doi.org/10.5194/gmd-16-1467-2023

Sammen, S. S., Kisi, O., Ehteram, M., El-Shafie, A., Al-Ansari, N., Ghorbani, M. A., and Shahid, S., 2023. Rainfall modeling using two different neural networks improved by metaheuristic algorithms. *Environmental Sciences Europe*, 35(1), p.112. https://doi.org/10.1186/s12302-023-00818-0

Shabbir, M., Chand, S., and Iqbal, F., 2023. A novel hybrid framework to model the relationship of daily river discharge with meteorological variables. *Meteorology, Hydrology, and Water Management. Research and Operational Applications*, *11*(2), pp.70-94. https://doi.org/10.26491/mhwm/187899

Shamshirband, S., Gocić, M., Petković, D., Saboohi, H., Herawan, T., Kiah, M. L. M., and Akib, S., 2014. Soft-computing methodologies for precipitation estimation: A case study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(3), pp.1353-1358. https://doi.org/10.1109/JSTARS.2014.2364075

Turhan, E., 2021. A comparative evaluation of the use of artificial neural networks for modeling the rainfall-runoff relationship in water resources management. *Journal of Ecological Engineering*, 22(5). https://doi.org/10.12911/22998993/135775

تقييم أساليب التعدين في البيانات والذكاء الاصطناعي للتنبؤ بهطول الأمطار اليومي

ياسين احمد حمه امين

قسم هندسة المدنية، كلية الهندسة، جامعة السليمانية، السليمانية، العراق

الخلاصة

يعتبر هطول الأمطار من عوامل المهمة للطقس الذي لها تأثيرات مباشرة وغير مباشرة على الحياة في كوكبنا. ومن أهم عواقب تغير المناخ هي الظروف الجوية القاسية مثل الفيضانات والجفاف. وحتى الآن ظلت توقعات نموذج المناخ غير مؤكدة، وبالتالي فإن تقنية نمذجة هطول الأمطار الأكثر دقة ضرورية. ونظرًا للتعقيد والعلاقة غير الخطية بين المعلمات المناخية، وعدم فعالية أدوات النمذجة الإحصائية التقليدية في التنبؤ بالظروف الجوية. في هذه الدراسة، تم تطبيق تقنيات استخراج البيانات على البيانات الانواء الجوية والمناخية للتنبؤ بهطول الأمطار اليومي جنبًا إلى جنب مع تقنيتين للذكاء الاصطناعي، الشبكات العصبية الاصطناعية (ANN) ونظام الاستدلال العصبي الضبابي (ANFIS). تم استخدام ما مجموعه 10 متغيرات الانواء الجوية اليومية، و لفترة 13 عاما، تحديدا الحد الأقصى لدرجة الحرارة (Tmax)، والحد الأدنى لدرجة الحرارة (Tmin)، والحد الأقصى للرطوبة (Hmax)، والحد الأدنى للرطوبة(Hmin)، وسرعة الرياح(Ws)، واتجاه الرياح(Wd)، والغطاء السحابي (Cv)، وضغط البحر (SEAp)، وضغط المحطة (STAp) والرطوبة النسبية (RH) للتنبؤ بهطول الأمطار اليومى(P). أظهر كلا نظامي الذكاء الاصطناعي نتائج مقبولة للتنبؤ بهطول الأمطار اليومي من المعلمات الانواء الجوية المرصودة مع معامل التحديد 0.75 (R^2) و0.72 لمعايرة النماذج ANN وANFIS على التوالي، بينما كانت نتائج اختبار طرق ANN و ANFIS و 0.55 و 0.62 على التوالي. وأظهرت نتائج الدراسة أن نموذج ANN قد يكون مبالغا في النتائج في قسم المعايرة من العملية مقارنة بطريقة ANFIS التي كان أداؤها أفضل في قسم الاختبار من عملية التقييم. في نمذجة ANFIS، لعدد من متغيرات الإدخال يصل إلى 6 متغيرات، توصي هذه الدراسة باستخدام طريقة تقسيم البيانات المعروف ب grid partition method . بالنسبة لمتغيرات الإدخال التي تزيد عن 6 متغيرات، يوصى باستخدام طريقة التجميع الفرعي المعروف ب sub clustering method .

الكلمات المفتاحية: استخراج البيانات، الذكاء الاصطناعي، ANN، ANFIS، هطول الأمطار، النمذجة.