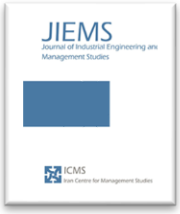
Vol. 13, No. 1, 2026, Pages 125-133

JIEMS
Journal of Industrial Engineering and Management Studies

journal homepage: www.jiems.icms.ac.ir



Monthly rainfall forecasting using artificial neural network model and regression method

Paria Soleimani *[1], Shahnaz Piroozfar [1]

[1] Department of Industrial Engineering, S.T.C., Islamic Azad University, Tehran, Iran.

Received: Dec 2025-25 / Revised: Jun 2026-20 / Accepted: July 2026-30

Abstract

Rainfall plays a vital role in water resource management and is a critical variable for effective decision making and planning. Rainfall strongly controls the dynamic of the water resource. It supports ecosystems, regulates temperature, cleans the atmosphere by removing pollutants, and creates landscapes. Accurate estimation of monthly rainfall is essential for various applications such as flood forecasting, drought management, irrigation planning, and watershed management. In this study, monthly rainfall in Tehran city is predicted using an artificial neural network (ANN) model. ANNs allow to solve complex problems that might have difficulty been solved through the traditional algorithms. The optimal structure is selected based on the valid statistics after testing different numbers of parameters. The results are also compared with regression-based approach. Support vector regression (SVR) aims to find a function that can accurately predict the monthly rainfall. The dataset includes 60 years of meteorological records such as rainfall, humidity, temperature, total monthly sunshine hours, dew point temperature, and wind speed. The results show the artificial neural network method is more proper for predicting monthly rainfall in Tehran due to its lower error rate compared to the support vector regression method. The results of paper can help and show a guidance for the managers to choose the proper method for forecasting of rainfall leading to the management of primary source of fresh water.

Keywords: Rainfall Forecasting; Artificial Neural Networks; Regression; Meteorological Parameters

Paper Type: Original Research

1. Introduction

Rainfall is considered one of the most important atmospheric phenomena that is playing a direct role in the human life cycle and serving as a main source of rivers. Nowadays, rainfall forecasting is a challengeable subject because of its portion as a vital tool for water resource management in many sectors (Chuasuka et al.,2025). Both excessive and insufficient rainfall can lead to undesirable climatic events such as floods and droughts, resulting in significant human and economic sufferers. Artificial neural networks (ANNs) have been identified as one of the most suitable and reliable methods for forecasting. Numerous variables such as pressure, temperature, and humidity affect rainfall patterns, and the relationship among these parameters is often complex and nonlinear. This study explores methods for accurately rainfall predicting. Since, that is a stochastic phenomenon, various black-box time series models can be employed for its estimation. Traditional methods, such as regression, can also be compared with the intelligent approaches (Nieto et al.,2025). This study adopts an ANN-based methodology and compares the results with those derived from the regression technique. Real-world meteorological data are used in this research. We used different indices to evaluate and compare the performances of proposed methods. We show the methods work properly for a real data set from Tehran city. Hence, the paper can illustrate a scientific way to predict rainfall for the city that has strong problems by lack of sufficient water.

1. Literature review

1.1. Neural Networks

Artificial Neural Networks (ANNs) have properties that make them effective in many applications, such as pattern recognition and control. Generally, ANNs are appropriate both for linear and nonlinear mappings. ANNs can be trained using three primary methods: supervised, unsupervised and reinforcement learning (Nieto et al.,2025). Neural networks can be classified in several ways, depending on the parameters such as number of layers, cell structure or arrangement of input and output neurons for data processing. In addition, an essential aspect of any model design is the evaluation of validity and performance of the model. There are various methods to assess the

*Corresponding Author: p_soleimani@iau.ac.ir

performance of a trained network. The simplest approach is to evaluate it using the same training data, which is trained, while a separate test dataset can be used to estimate the network ability (Gnanasankaran & Ramara ,2024).

1.2. Regression

Regression is one of the most powerful tools for explaining the relationship between variables (Gnanasankaran & Ramara ,2024). It formulates correlation between independent and dependent variables(Jiryaei, et al.,2022). In this study, using data on rainfall, humidity, air temperature, wind speed, dew point temperature and total sunshine hours in Tehran city, regression equation is determined. Support vector regression (SVR) aims to find a function that can accurately predict the output value based on the input values. It is used both for grouping and for estimating the data fitting function in regression problems (Krishna Das, 2024). The SVR method is the support vector machine (SVM) regression model. Based on the related research, the advantage of SVM over the multilayer neural networks is the elimination of local optimization. Hence, the maximization can be generalized. SVM provides the best answer by minimizing the structural risk, and is a powerful algorithm to classify data (Raei, Shirkavand & Jamali,2024);but one of the disadvantages of it in regression modelling is the long training time and the complexity for multi-class problems (Chuasuka et al.,2025).

1.3. Research Background

Nowadays, climate change has become a serious threat to humanity (Alotaibi et al.,2025). Due to uncertainty in climate conditions, proper planning seems essential. Climate change and variability are two important factors to understand climate behavior. In addition, climate studies can provide actionable managerial insights to support supply chain resilience initiatives (Guitouni et al.,2025).So, accuracy in forecasting is vital (Sedighi &Madanchi Zj,2023). studied artificial neural network models to predict monthly rainfall in arid regions (Aksoy& Dahamsheh, 2018). (Dariane et al., 2019) used neural networks and large-scale weather signals to predict rainfall at six stations in Iran. (Benevides et al.,2019) predicted and classified heavy rainfall events using global navigation satellite systems. (Kang, et al., 2025) applied artificial neural networks to predict precipitation in Jingdezhen, Jiangxi province, China. They enhanced the accuracy of precipitation predictions by removing the input variables with poor correlation. Support vector regression is studied for its effectiveness in solving linear and nonlinear problems (Wang et al., 2024). In this study, the multilayer perceptron method of ANN is employed to predict monthly rainfall in Tehran city, while a regression model is also applied to forecast the monthly precipitation. Finally, the results are compared by using of statistic indexes.

2. Research methodology

2.1. Modeling Using Neural Networks

In this study, the Artificial Neural Network (ANN) perceptron model is employed, using meteorological data from Tehran city. Initially, the data were normalized and divided in to three subsets: training (70%), validation (15%) and testing (15%) sets. A neural network model is created, consisting of an input layer, hidden layers, and an output layer. Input variables include average relative humidity, average temperature, average wind speed, average dew point temperature, and total sunshine hours, while the output variable is the total monthly rainfall. Sensitivity analysis is conducted on the network parameters to minimize the prediction error. The training method was the Levenberg-Marquardt algorithm, and the number of iterations (epochs) was set to 1000. The steps for constructing the neural network are as follows:

3.1.2. Model Building and Validation

The number of neurons in the hidden layers is determined experimentally. Simulations are conducted using sigmoid and tangent-linear activation function for hidden layers, with two to ten neurons in each hidden layers. Finally, in this study, a three-hidden layer perceptron network is used with a tansig transfer function in the hidden layers and a linear (Purelin) function in the output layer. This structure, employing a sigmoid derivative and nonlinear function in the hidden layers, along with a linear function in the output layer, allows the network to learn both linear and nonlinear relationships between the inputs and output variables. Additionally, a gradient-based training algorithm is studied for network training. To assess the accuracy of the results, various statistical criteria are used, including the correlation coefficient (CC), root mean square error (RMSE) and mean absolute error (MAE).

3.1.3. Confusion Matrix and AUC Diagram

The accuracy index is one of the most common and fundamental criteria for measuring the quality of classification. It is calculated as a percentage, as follows in Equation (1):

$$\text{Accuracy} = (TP+TN) / (TP+FN+FP+TN) \qquad (1)$$

where TP, TN, FN, and FP represent the True Positives, True Negatives, False Negatives and False Positives, respectively. Another evaluation criterion is commonly used as sensitivity, also known as the True Positive Rate (TPR). This index is calculated in Equation (2):

$$\text{Sensitivity (TPR)} = TP / (TP+FN) \qquad (2)$$

This criterion indicates how much the classification method is successful for detecting the all rainfall months. The specificity (True Negative Rate, TNR) is another important index, which is calculated in Equation (3):

$$\text{Specificity (TNR)} = TN / (TN+FP) \qquad (3)$$

An additional tool for evaluating the quality of classifications is the Receiver Operating Characteristic (ROC) curve, which illustrates the relationship between sensitivity and specificity. The area under this curve (AUC) is typically close to one. By examining this area, the accuracy of predictions can be assessed. The confusion matrix and the ROC diagram are useful evaluation tools for binary classification prediction, such as predicting precipitation or no precipitation. These methods are employed in this research.

3.2. Support Vector Regression

In SVR method, it is assumed that the relationship between independent and dependent variables is determined by an algebraic function, f(x), as follows in Equations (4) and (5).

$$f(x)=WT.\emptyset(x)+b \qquad (4)$$

$$y=f(x)+noise \qquad (5)$$

where W is the coefficient vector, b is a constant in the regression function, and $\emptyset(x)$ is a kernel function. The goal is to find a functional form for f(x), which is achieved by training the SVM model using the training data set. In SVR, the output layer contains P as the precipitation amount (mm) and the input layer consist of the following variables: V for wind speed (m/s), M for humidity (percentage), S for sunshine hours(h), T for temperature(c), and D for dew point temperature(c). 60-years delay is used to predict the next 12 months. For modelling with SVM, the fit rsvm command in MATLAB software is employed with a kernel function (Leong& Kelani,2020) as follows in Equation (6).

$$\emptyset(u,v)=u.v \qquad (6)$$

The kernel function $\emptyset(u,v)$ generates inner products that allows for nonlinear decision boundaries within the data environment. In the next section, the predicted values from the neural network model and the support vector regression method are compared and analysed.

4. Data Analysis and Case Study

According to the average annual rainfall map of Iran, Tehran province is one of the regions with low rainfall. Due to the large population of the city, water shortage or flood can lead to significant problems for this metropolis. Therefore, accurate rainfall prediction can help implement appropriate strategies for managing water shortage or preparing for heavy rainfall. In this study, data from a meteorological station in Tehran is used as a case study to analyse rainfall predictions for the city. The monthly statistical data for variables such as rainfall, humidity, air temperature, wind speed, dew point temperature and sunny hours are presented for 60 years period. The statistical characteristics, including the minimum and maximum monthly rainfall, mean, skewness and standard deviation are shown in Table 1.

Table 1, Monthly meteorological data

Statistical Parameter / Meteorological parameter	Average	Standard deviation	Skewness	Min	Max
Rainfall (mm)	19.4	23.3	1.7	0	141.7
Temperature(c)	17.4	9.7	-0.07	-4.7	33.9
Humidity(%)	39.6	15.5	0.5	14.2	84.3
Wind speed(m/s)	2.7	0.7	0.2	0.8	5.1
Sunny hours(h)	250.2	72.9	0	73.5	387.4
Dew temperature(c)	1.2	4.3	0.3	-10.3	18.7

4.1. Correlation Coefficient of Independent and Dependent Variables

The correlation coefficients between monthly rainfall and independent variables are analysed to determine the effective variables for predicting monthly rainfall in Tehran. These coefficients are presented in Table 2.

Table 2. Correlation coefficient between independent and dependent variables

Independent Variables / Dependent Variable	Wind Speed	Humidity	Sunny Hours	Temperature	Dew point
Rainfall	-0.12	-0.69	-0.65	-0.60	-0.31

In Table 2, humidity, sunny hours and temperature have the strongest correlation with rainfall, respectively. Table 3 displays the correlations between the predictor variables.

Table 3. Correlation between rainfall predictor variables

Variables	Dew Point	Temperature	Sunny Hours	Humidity	Wind Speed
Dew Point	1	0.80	0.62	-0.47	0.12
Temperature	0.80	1	0.89	-0.89	0.33
Sunny Hours	0.62	0.89	1	-0.85	0.31
Humidity	-0.47	-0.89	-0.85	1	-0.43
Wind Speed	0.12	0.33	0.31	-0.43	1

4.2. Monthly Time Series Data for Meteorological Data

At first, data is normalized and pre-processing done, then the backpropagation network is trained. Delays of 50 and 60 years are also applied to predict the following 12 months. The number of neurons in the hidden layers was determined through trial and error. After testing various options (ranging from 2 to 10 neurons), it was concluded that fewer neurons in the hidden layers produced more accurate results. A larger number of neurons did not improve the results and increased training time. For example, in Model 1(net1), the hidden layer consists 3 layers, with 2 neurons per layer and the model is trained for 10,000 iterations (Appendix A). The network with the lowest error (RMSE) is considered as the best model.

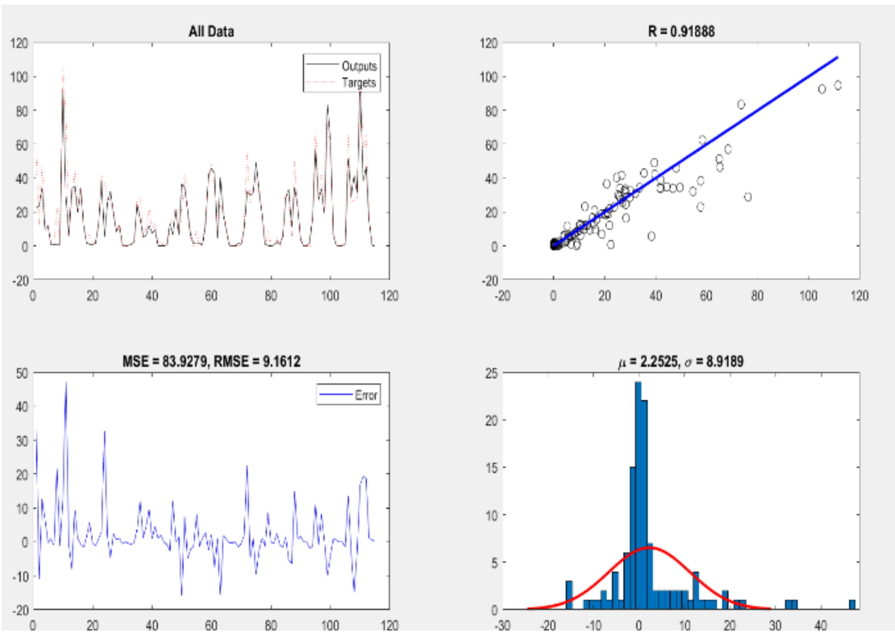


Figure 1. Graphs of Model 1 neural network

The red line in Figure 1 represents the actual rainfall values, while the black line shows the predicted values (output) for net1. It is evident that the predicted values closely follow the actual values. This indicates that the model accurately reflects the variations in rainfall. Additionally, the high determination coefficient (0.91) demonstrates the strong performance of Model 1(net1) for predicting rainfall.

4.3. Calculation and Evaluation of the Confusion Matrix and AUC Diagram

To construct the confusion matrix, the threshold values equal to 0.1, 2, 10, 15 and 25 are used. Values greater than the threshold were set to 1, while values below that were assigned as 0. The True Positive Rate (TPR) and False Positive Rate (FPR) values are calculated. An ROC diagram is generated for each class. Based on the calculated AUC value (which exceeds 0.5), the model accuracy can be confirmed. The confusion matrix and ROC diagram for the different threshold values are provided in Appendix B. Finally, the predicted rainfall values for Tehran city over the next 12 months are calculated using the artificial neural network model and presented in Table 4.

Table 4. Predicted rainfall for the next 12 months

Month	Predicted Values
July	0
August	4.7
September	3.66
October	0.319
November	20.80
December	40.09
January	29.58
February	23.22
March	11.77
April	15.91
May	10.59
June	0

The results show that there are maximum rainfall for December and January months and minimum for July and June months.

4.4. Support Vector Regression (SVR) Method Results

The SVR model results are shown in Table 5.

Table 5. Support Vector Regression Results

MSE	RMSE	R2	R
154.3	12.4	0.62	0.78

The SVR output generated by MATLAB software is provided in Appendix C. Based on the results presented in the appendix, it is evident that the support vector regression model has not been able to predict the precipitation with high accuracy. Some sections of the model output even show inability to predict the rainfall effectively. As a result, this model is not suitable for accurate and efficient precipitation prediction. Additionally, the correlation coefficient between the observed and predicted values was calculated, as shown in Figure 2.

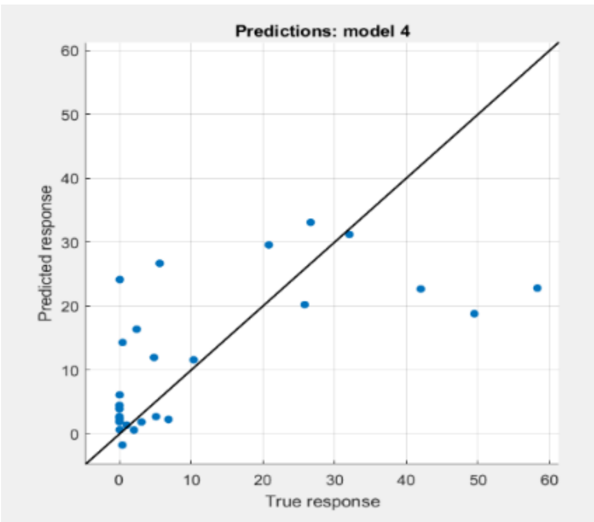


Figure 2. Diagram of R for support vector regression model

According to Table 5, the SVR model accounts for 62% of actual precipitation values, with an error margin of 12.4%. This indicates that the SVR model does not predict monthly rainfall in Tehran as accurately as the ANN model.

4.5. Comparison of Models Results

Four criteria are examined to compare the performance of two proposed methods, the coefficient of determination (R^2), correlation coefficient (R), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The results are presented in Table 6. It shows that the ANN model outperforms the SVR model in terms of rainfall prediction accuracy for the given dataset.

Table 6. Statistical criteria values of proposed models

Models	R2	R	MSE	RMSE
ANN	0.83	0.91	83.9	9.1
SVR	0.62	0.78	154.3	12.4

5. Discussion and Future Research

Climatic phenomena typically exhibit some trends, that they follow nonlinear patterns over time. Therefore, to predict and estimate such phenomena, artificial neural networks (ANNs) often outperform other models. Rainfall is a nonlinear phenomenon. It does not yield the acceptable results using linear or polynomial models, as semi-linear methods. The results of this study, after testing some networks with various structures, indicate that the use of a multilayer perceptron artificial neural network provides relatively accurate results. Generally, neural network models require substantial amounts of data to build a successful forecasting model. Furthermore, personal judgment and experience play an essential role in selecting the appropriate parameters. The selection of model parameters is typically done through trial and error. According to the most studies, perceptron networks with one or two hidden layers are preferred to reduce simulation time. However, based on the results of this study, it is advisable to select the optimal structure based on the valid statistics after testing different numbers of layers and neurons. Among the neural networks proposed for the case study, the MLP neural network with three hidden layers and two neurons per layer was found to be the best model. Among the training algorithms used, Levenberg-Marquardt (LM) algorithm was the most effective. Additionally, the results of the support vector regression (SVR) model with a kernel function demonstrated that SVR performed less effectively for rainfall forecasting compared to the neural network method .Future research could be considered to improve the rainfall forecasting accuracy, for example comparing autoregressive vector models (multivariate time series) with neural network models, using the other meteorological parameters (such as wet point temperature, air pressure or minimum and maximum temperature), exploring the seasonal modelling, and shorter time scales such as daily or hourly intervals. It is also suggested to combine the other artificial neural network training algorithms and compare the results.

6. Managerial Insights and Conclusion

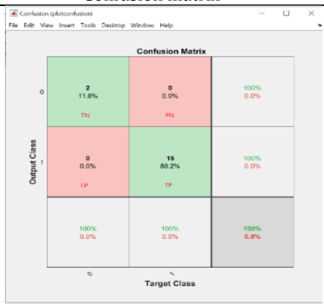
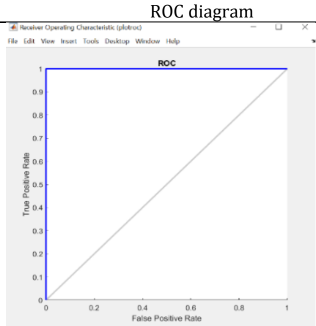
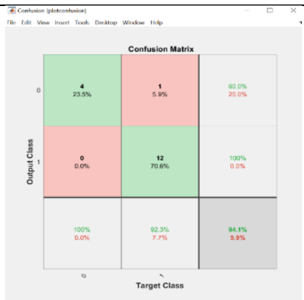
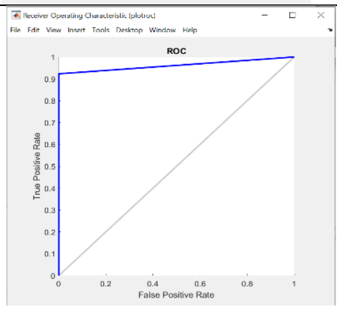
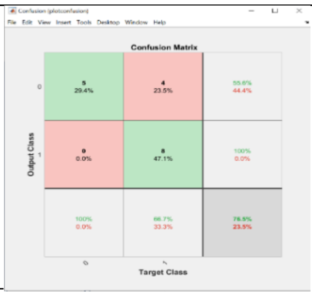
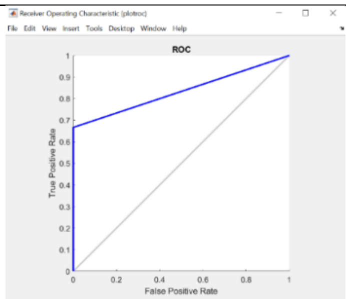
Todays, water shortage is an important subject for many countries. Therefore, it can threat human health and lead to diseases and dehydration. It has the important impacts on agriculture, which in turn affects food security. One of the main and basic factors to manage water resources is rainfall management. The accurate and precision prediction of rainfall is a key factor for programming and decision-making. Applying new scientific methods, especially artificial intelligence (AI) can help simplify, speed up, and increase the accuracy of calculations. This study shows that, based on the correlation coefficients, humidity, temperature, and sunny hours are the most important and effective variables for rainfall forecasting in Tehran city. A comparison of performance indices between the two applied methods reveals that the neural network model outperforms the regression method. Hence, the artificial neural network method is recommended for predicting monthly rainfall in Tehran due to its lower error rate compared to the support vector regression method .Ultimately, the presented findings here underscore the critical importance of applying the scientific and new methods for water resource management in today's world.

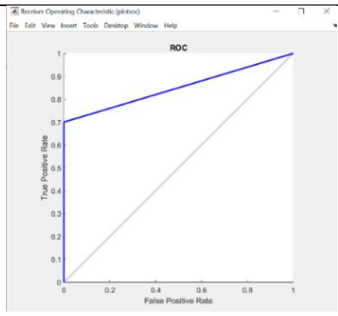
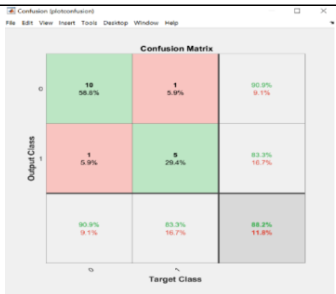
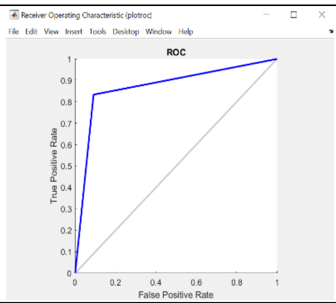
7. Appendix

Appendix A. Artificial neural network results for some models

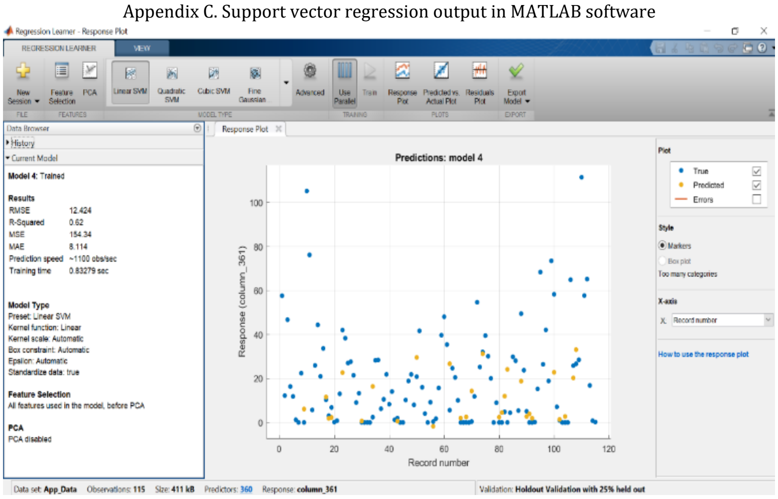
MSE Total	RMSE Total	R Test	R Validation	R Training	R Total	Delays	Network	Net
83.9	9.1	0.87	0.81	0.95	0.91	60 years old	2-2-2-1	net_1
88.6	9.4	0.66	0.80	0.92	0.89	60 years old	4-4-4-1	net_2
86.4	9.3	0.66	0.75	0.94	0.90	60 years old	5-5-5-1	net_3
96.04	9.8	0.71	0.65	0.95	0.90	60 years old	6-6-6-1	net_4
117.1	10.8	0.68	0.85	0.96	0.87	60 years old	8-8-8-1	net_5
136.9	11.7	0.66	0.61	0.93	0.86	60 years old	7-7-7-1	net_6
168.1	12.9	0.81	0.80	0.88	0.84	60 years old	2-2-2-1	net_7
187.5	13.6	0.81	0.80	0.80	0.80	50 years old	2-2-2-1	net_8
198.1	14	0.78	0.78	0.79	0.78	50 years old	2-2-2-1	net_9
198.8	14.1	0.81	0.78	0.78	0.78	50 years old	2-2-2-1	net_{10}
201.6	14.2	0.78	0.79	0.78	0.78	50 years old	2-2-2-1	net_{11}
206.2	14.3	0.77	0.77	0.81	0.78	50 years old	2-2-2-1	net_{12}
214.7	14.6	0.72	0.73	0.78	0.76	50 years old	2-2-2-1	net_{13}
216	14.6	0.71	0.81	0.77	0.76	50 years old	2-2-2-1	net_{14}
222.5	14.9	0.82	0.78	0.77	0.76	50 years old	2-2-2-1	net_{15}
247	15.7	0.80	0.73	0.75	0.74	50 years old	2-2-2-1	net_{16}
267.7	16.3	0.79	0.82	0.70	0.71	60 years old	4-4-4-4-1	net_{17}
275.4	16.5	0.71	0.78	0.70	0.71	50 years old	2-2-1	net_{18}
276.9	16.6	0.81	0.70	0.72	0.72	50 years old	2-2-2-1	net_{19}
284	16.8	0.64	0.72	0.65	0.66	60 years old	2-2-2-2-1	net_{20}
302.8	17.4	0.73	0.65	0.63	0.64	60 years old	10-10-10-1	net_{21}
309.4	17.5	0.66	0.60	0.61	0.62	60 years old	5-5-5-5-1	net_{22}

Appendix B. The confusion matrix and ROC diagram for different threshold values

Threshold	Confusion matrix	ROC diagram
0.1		
2		
10		

15		
25		

Appendix C. Support vector regression output in MATLAB software



References

Aksoy, H.,& Dahamsheh, A.(2018). Artificial neural network models for forecasting monthly precipitation in Jordan. Stochastic Environmental Research and Risk Assessment, 23, 917-931.

Alotaibi et al.(2025). Role of risk perception and climate change beliefs in adoption of climate-resilient agricultural practices in Saudi Arabia. Climate Services, 38.

Benevides, P J., Catalao,A., Nico, G.(2019). Neural Network Approach to Forecast Hourly Intense Rainfall Using GNSS Precipitable Water Vapor and Meteorological Sensors. Remote Sensing. 11, 966.

Chuasuka, P., Bhatrasataponkul, T., & Akkarapongtrakul, A.(2025). Comparative analysis and enhancing rainfall prediction models for monthly rainfall prediction in the Eastern Thailand. Methods X, 103094.

Dariane, AB., M., Ashrafi, Gol M., Karami, F.(2019).Forecasting of rainfall using different input selection methods on climate signals for neural network inputs. Journal of Hydraulic Structures. 5, 42-59.

Gnanasankaran, N., & Ramara,j E. A.(2024). Multiple linear regression model to predict rainfall using indian meteorological data. International Journal of Advanced Science and Technology , 29(8).

Guitouni et al.(2025). Supply chain serviceability under climate change with application in the Arctic. Transportation Research,139.

Jiryaei Sharahi,Z., et al. Machine learningdecisiontree based on regression in data mining to extract more knowledge.(2022). Journal of Industrial Engineering and Management Studies,9(2),86-112.

Kang, J., et al.(2025). Prediction of precipitation based on recurrent neural networks in jingdezhen, Jiangxi Province, China. Atmosphere. 11, 246.

Krishna Das, P., Sahu, R L., Swain, PC. (2024). Comparative analysis of machine learning models for rainfall prediction. Journal of Atmospheric and Solar-Terrestrial Physics, 246.
Leong, WC. , Kelani, K.(2020). Prediction of air pollution index (API) using support vector machine (SVM), Environmental Chemical Engineering. 8,103208.
Raei,R.,Shirkavand,S.,& Jamali Neyshabour, Ali.(2024).Applying metaheuristics And SVMs to forecast stock price crashes in Tehran stock exchange. Journal of Industrial Engineering and Management Studies,11(2),45-52.
Sedighi,M., & Madanchi,M.(2023). An intelligent hybrid model for forecasting the stock price index volatility: the case of Tehran stock exchange. Journal of Industrial Engineering and Management Studies,10(2),116-130.
Nieto et al.(2025) Predicting the critical superconducting temperature using the random forest, MLP neural network, M5 model tree and multivariate linear regression. Alexandria Engineering Journal,86, 144–156.
Wang, W et al.(2024). Forecasting the market value of power battery industry chain: a novel RRMIDAS-SVR model. Journal of management science and engineering. 9, 474-489.