Multilayer perception and radial basis function models for predicting trends of rainfall in Asian megacity Dhaka, Bangladesh

Abu Reza Md. Towfiqul Islam (towfiq_dm@brur.ac.bd)
Begum Rokeya University
Subodh Chandra Pal
The University of Burdwan
Rabin Chakrabortty
The University of Burdwan
Javed Mallick
King Khalid University
Showmitra Kumar Sarkar
Khulna University of Engineering & Technology
Md. Abdul Fattah
Khulna University of Engineering & Technology
M. Safiur Rahman
Bangladesh Atomic Energy Commission

Research Article

Keywords: Multilayer perception neural network, Radial bias function neural network, multi-collinearity assessment, Dhaka

Posted Date: June 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1720286/v1

Multilayer perception and radial basis function models for predicting trends of rainfall in Asian megacity Dhaka, Bangladesh

Abu Reza Md. Towfiqul Islam[1*], Subodh Chandra Pal[2*], Rabin Chakrabortty[2], Javed Mallick[3*], Showmitra Kumar Sarkar[4], Md. Abdul Fattah[4], M. Safiur Rahman[5],

[1]Department of Disaster Management, Begum Rokeya University, Rangpur 5400, Bangladesh

[2]Department of Geography, The University of Burdwan, Bardhaman, West Bengal, 713104, India; Email: geo.subodh@gmail.com; rabingeo8@gmail.com

[3]Department of Civil Engineering, King Khalid University, Abha, 62529, Saudi Arabia; Email: jmallick@kku.edu.sa

[4]Department of Urban and Regional Planning, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh; email:mail4dhrubo@gmail.com; mafattah.kuet@gmail.com

[5]Water Quality Research Laboratory, Division of Chemistry, Atomic Energy Centre, Bangladesh Atomic Energy Commission, Dhaka-1000, Bangladesh; Email: safiur.rahman@dal.ca

[*]Corresponding author: towfiq_dm@brur.ac.bd; geo.subodh@gmail.com; jmallick@kku.edu.sa

Abu Reza Md. Towfiqul Islam, PhD

Subodh Chandra Pal, PhD

Javed Mallick, PhD

ORCID: 0000-0001-5779-1382

Tel: +880-2-58616687

Fax: +880-2-58617946

Abstract

Rainfall prediction is a fascinating topic, particularly in an urban city experiencing climate change; it is also required for hydrologic system analysis and design. Most real-time rainfall prediction algorithms use conceptual models that simulate the hydrological cycle in a changing climate. However, calibration of "conceptual" or "physically based models" is typically challenging and time-consuming due to the large number of variables and factors. Simpler "artificial neural network (ANN)" predictions may thus seem promising. To this end, this study aimed to evaluate the performance of two of the most commonly used ANN models, namely "Multilayer Perception (MLP)" and "Radial Bias Function (RBF)", by predicting the rainfall trend patterns in the mega city of Dhaka, Bangladesh. In this perspective, rainfall is considered as a dependent variable and the rest of the parameters are considered as independent variables for predicting the trends of rainfall in this region. In the prediction models, fifteen conditioned atmospheric and meteorological parameters were used, and the multi-collinearity of these parameters was checked by the "Variance Inflation Factor" and "Tolerance" methods. The performance of the ANN models was evaluated by comparing the predicted and residuals and also by using AOC and ROC curves. The importance study from the MLP model revealed that PM_{10}, O_3, $PM_{2.5}$, NO_x, and wind speed are the highest causal factors influencing the rainfall changes in Dhaka, Bangladesh. Though both ANN models produced similar robustness in rainfall prediction, results showed that MLP performed well with an AUC value of 0.941, compared to the RBF model with an AUC value of 0.915. Therefore, the application of the MLP model can be suggested as an alternative to predict the pattern of rainfall as well as meteorological and atmospheric variables based on historical recorded datasets.

Keywords: Multilayer perception neural network, Radial bias function neural network, multi-collinearity assessment, Dhaka

1. Introduction

Changing patterns of rainfall have a direct impact on water supplies since they are an essential aspect of the hydrological cycle (Islam et al., 2012). Water resource managers and hydrologists are increasingly concerned about the impact of climate change on rainfall patterns (Jhajharia et al. 2014; Gajbhiye et al., 2015). The variations in rainfall quantity and frequency directly affect the stream flow pattern, the allocation of run-off, ground water reserves, and soil moisture (Srivastava et al., 2014; Islam et al., 2022). Because of the significant variations in the rainfall pattern, drought and flood-like hazardous occurrences might occur on a regular basis (Srivastava et al., 2015). As a result, the most important approach to long-term water resource management is studying how rainfall patterns and distribution are changing as a result of climate change. Precipitation patterns are crucial in an ever-changing environment for making better decisions and improving communities' capacity to survive significant weather events. If the air humidity is high and profound and significant convection develops, surface rainfall is due to intensify even more at increased aerosol concentrations. Greater urban surface area is unlikely to have a significant influence on urban-induced rainfall. Larger urban areas, on the other hand, can interrupt or split precipitation convection structures that originate outside cities and pass over them. Rainfall over and/or leeward of cities can be increased or decreased by these urban-modified rainfall patterns. Due to human activity, aerosols have a variety of sizes and chemical compounds that are discharged into the urban atmosphere. Anthropogenic aerosols, in combination with toxic gases, degrade urban air quality and are detrimental to human health. In hydrological and climatological research and practice, the pattern of rainfall is a crucial input and a critical concern (Sangati and Borga 2009; Islam et al. 2012). Hydrological modeling and drainage system design need this data, as does flood forecasting (Lagouvardos et al. 2013). As a result, examining the changes in trend and presence of a trend in rainfall is one of the most important areas of hydrology, climatology, and meteorology research throughout the world (Islam et al., 2012; Chatterjee et al., 2016; Tian et al., 2017; Talaee, 2014). Statistical tests such as "regression" (Piao et al., 2010), the "Mann–Kendall" test (Mann, 1945), the "Kendall rank correlation" test (Kendall, 1995), "Sen's slope estimation" (Pingale, 2014), and the "Spearman rank correlation" test (McGhee, 1985) have all been used in the majority of investigations. The Mann-Kendall test was used in this investigation since it is one of the most frequently used worldwide techniques for trend identification in hydrology, climatology, and meteorology (Batisani and Yarnal, 2010; Du and Shi, 2012; Singh, 2008; Islam et al. 2019).

For planners, predicting the future is more helpful than trend research because it allows them to prepare for future climatic changes, which are more likely to occur. In addition to the complex global climate model, statistical approaches and machine learning techniques (Darji et al. 2015) might be used for regional predicting of climatic variables in the future. A significant number of databases, a high-configuration system, modern technology, and a technical specialist are needed to run the physical models. There is no substitute for physical models in terms of efficiency or cost. A number of statistical approaches have restrictions, such as the "auto regressive (AR)" model regresses past values, while the "moving average (MA)" model employs past error as the explanatory factor, and the "auto regressive moving average (ARMA)" model can only be performed for stationary time series data. Recent work has focused on the implementation of AI models such as machine learning techniques (Pal and Talukdar 2020). Due to the fact that it doesn't need a lot of information yet can handle complex and large data sets if given, AI models perform quite well.

Artificial neural network (ANN) has been extensively applied in various fields of water resource planning and management as well as environmental research. It is possible to construct an ANN by using several linked or weighted processing elements (PEs) or neurons, which are then assembled into layers and interconnected. Data is fed into the input layer and processed by the network until an output is generated at the end of its path. It is via these weighted connections that each neuron receives several inputs from its neighboring neurons. The total of these weighted inputs plus the addition of a standard cutoff generates the argument for a transfer function (often "linear," "logistic," or "hyperbolic tangent"), which in turn provides the final output of the neuron. According to Darji et al. (2015), a thorough study on the use of neural networks for rainfall forecasting, back propagation based neural networks perform well. Rather than relying on a single technique, we built a neural network around the "multilayer perception algorithm (MLP)" and "Radial Basis Function (RBF)". The MLP and RBF-based ANNs use a back propagation technique for their computations. The use of MLP-RBF-based ANN is becoming more popular because of the high reliability of their findings in predicting a wide range of hydrological and climatic events (Memarian & Balasundram, 2012; Samantarayet al., 2020; Liu, 2019). In the current study, MLP and RBF models were applied to predict the trends of rainfall in the Asian mega city of Dhaka, Bangladesh.

2. Materials and methods

2.1. Study area

Bangladesh is a low-lying, riverine country that is one of the most densely populated in the South Asian region. It has a 580 km long coastline along the Bay of Bengal. The climate of this country is a subtropical monsoon, with moderately warm temperatures, seasonal rainfall variations, and increased relative humidity. Bangladesh's summer (March to June), monsoon (June to October), and winter (October to March) are generally recognized as distinct seasons of Bangladesh (Mallick et al. 2022). Heavy rainfall is one of the main climatic characteristics of Bangladesh because of its geographical location in the southern corner of the foothills of the Himalayas. The monsoon season brings 80% of the country’s total rainfall. The monsoons are caused by differences in air pressure caused by the differential heating of land and sea in Bangladesh. Annual rainfall has ranged between 32,800 mm and 47,800 mm per year over the last 50 years (Fattah and Morshed, 2022).

Dhaka, the capital city of Bangladesh (Figure 1), is one of the fastest-growing megacities in the world and has been chosen as a study area in this research. Dhaka district is located between north latitudes 23°53' and 24°06' and the east longitudes of 90°01' and 90°37'. It is the world's eighth-largest and fourth-most densely populated city, with 8.9 million residents in the main city (Dhaka Metropolitan Area) and around 21.7 million in the entire Dhaka (Faisal et al. 2021). According to the Köppen climatic classification, the city has a tropical savanna climate and a distinct monsoonal season. With a distinct monsoonal season, temperatures range from 19°C in January to 29°C in May. Between May and October, nearly 87 percent of the average annual rainfall (2,123 mm) falls. Rapid population growth, unplanned urbanization, and industrialization have led to an increase in urban activities in Dhaka City, which has been affecting the climate as well as the environment (Chowdhury et al. 2021). Moreover, the geographical location of the city (lower reaches of the Ganges Delta) and the flat land close to sea level leave the city susceptible to flooding, especially in the rainy season due to heavy rainfall. Though the city's residents have faced water logging problems in recent years, the ground water level of Dhaka City is declining at a rate of 2-3 m/year due to less rainfall in other seasons and excessive settlement and population growth (Mamoon et al. 2020).

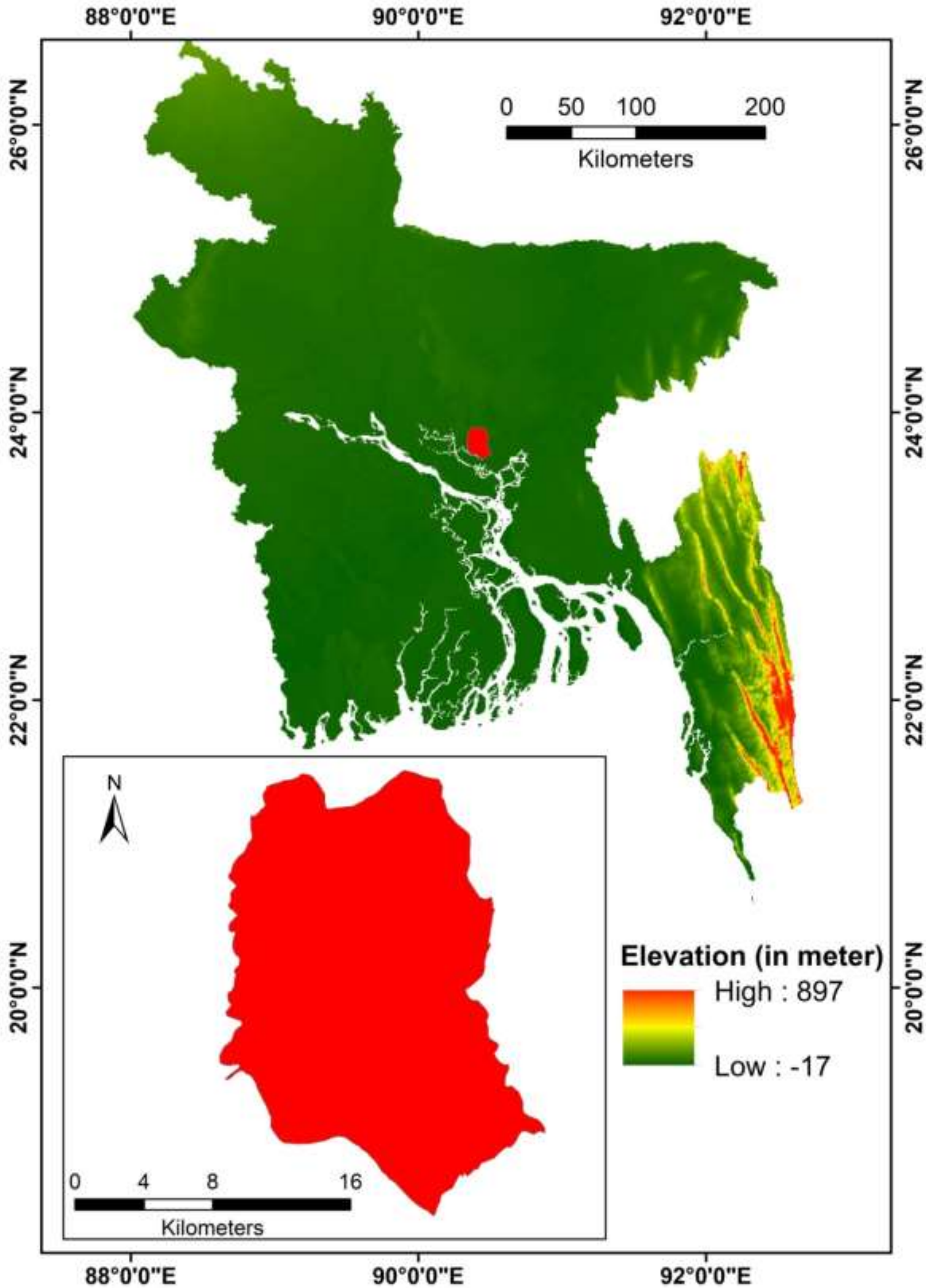


Figure 1 Location map showing the study area

2.2. Database and quality check

The amount of rainfall in any geographical location is affected by many atmospheric conditions, including air pollutants (SO_2, $PM_{2.5}$, PM_{10}, O_3, NOx, NO, NO_2, CO) and meteorological factors such as "air temperature, solar radiation, relative humidity, wind speed, wind direction". To predict the rainfall, change pattern in our study area, we have collected the atmospheric data from the "Ministry of Environment", and the meteorological data from the "Bangladesh Meteorological Department (BMD)". The quality control of the datasets was checked by BMD staff.

2.3. Methodology

As given Fig. 2, the process of predicting rainfall trend pattern consists of the following steps:

1. Input of air quality and meteorological parameters
2. Pre-processing of the input dataset
3. Building ANNs model with training (70%)
4. Validating ANNs model with testing (30%)
5. Importance of variable analysis by ANNs model
6. Evaluating the model by the taking difference between predicted and residual outputs

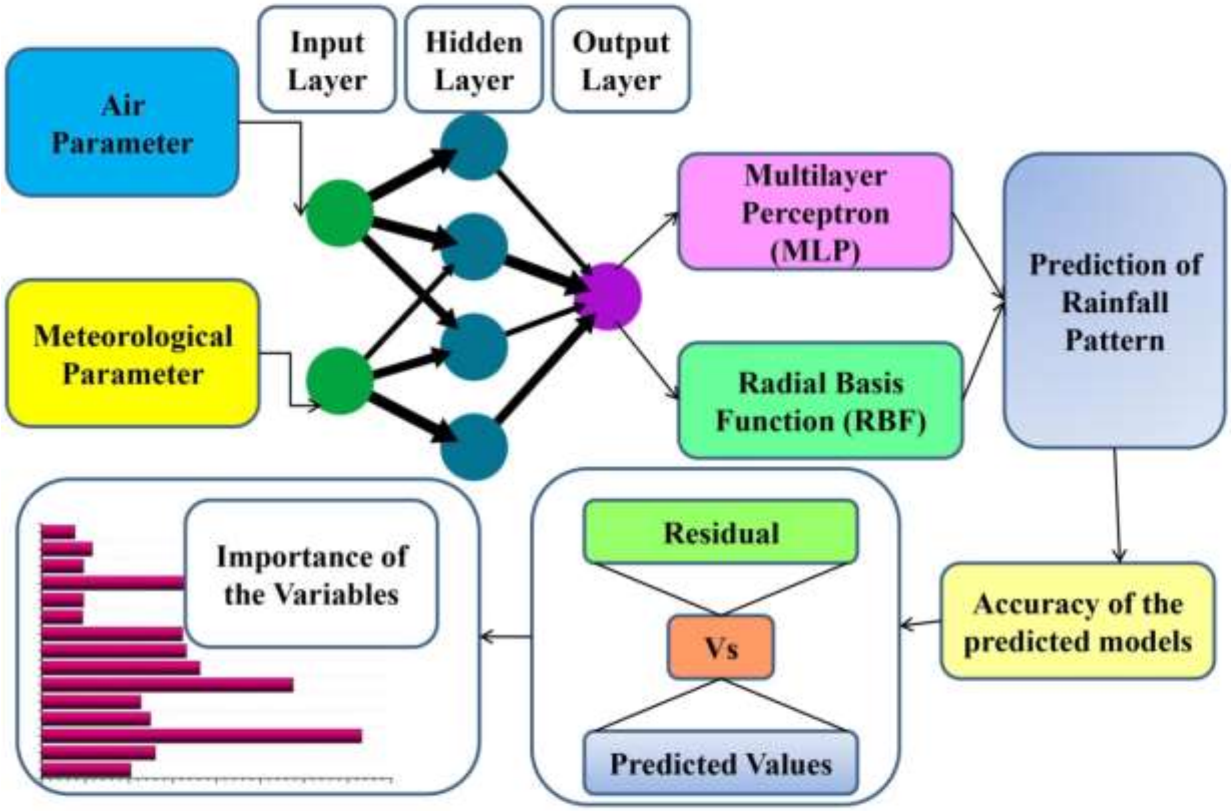


Figure 2 Flow Chart adopted for the current study

2.4. Trends of rainfall prediction

Multilayer Perception

The ANN is a machine learning-based computing method of the biological brain. The ANN-based Multilayer Perception (MLP) model is considered the most popular model in feedforward networks (Kuo et al. 2007, Hanoon et al. 2021). The MLP model is a set of organized, interconnected layers or nodes, namely "input", "hidden", and "output layers". The data is delivered in the input layer, and the neurons in the input layer transfer the weighted data to the hidden layers with the randomly set bias. A transfer function is used to generate an output response at the node after determining the net sum at the hidden node (Hanoon et al. 2021; Memarian and Balasundram, 2012).

In this study we have used two activation function namely hyperbolic tangent function (HTF) and the logistic function (LF). The HTF ranges between –1 and +1 while the LF ranges between 0 and +1. The two equations are expressed as follows:

$$\phi(y_i) = \tanh(\boldsymbol{w_i}) \qquad (1)$$

$$\phi(y_i) = \ (1 + e^{-w_i})^{-1} \qquad (2)$$

Here y_i denotes "the outcome of the i^{th} neuron, $\boldsymbol{w_i}$denotes the total weight of the input neurons. The equation 1 represents the HTF and equation 2 represents the LF".

The MLP network was trained using an error corrective learning approach. The desired output must be known in this technique. The instantaneous error $(\varepsilon_i(n))$is defined as "the difference between the desired response $\delta_i(n)$, and the system response at PEi at iteration n, $y_i(n)$".

$$\varepsilon_i(n) = \ \delta_i(n) - \ y_i(n) \qquad (3)$$

Each weight in the network can be updated using gradient descent learning theory. The weights update processing could take place after each training pattern has been presented, or after full sets of training patterns have been presented. A training epoch is deemed complete in both circumstances when each training pattern has been presented to the MLP once. Even in simple things, an accurate MLP needs to be well-trained for many epochs. The change of weight can be calculated by the "back propagation algorithm (equation 4) of momentum learning, which helps to calculate a cost function's sensitivity with respect to each weight in the MLP network and then updates each weight according to the sensitivity" (Behrang et al. 2010).

$$w_{ij}(n+1) = \ \boldsymbol{w}_{ij}(n) + \ \partial\left(\boldsymbol{w}_{ij}(n) - \boldsymbol{w}_{ij}(n-1)\right) + \ \eta\varepsilon_i(n)\boldsymbol{x}_j \qquad (4)$$

Here, “∂ = Momentum, η = step size, $\varepsilon_i(n)$= local error/ instantaneous error”.

Radial Basis Function

The “Radial bias function neural network (RBFNN)” is a non-normalized method of the Gaussian distribution nonlinear-function with a lot of excellent features for better learning. Nonlinear dynamic systems can also be learned, identified, synchronized, and controlled using Gaussian neural networks (Gholamreza et al. 2016).To solve the problems more intelligently the RBFNN model is used and it is expected to execute more precise simulations even in complex situation (Gholami et al. 2019). The architecture of the adopted RBFNN in this study is presented in Figure 6. This model is made up of three layers: input, hidden/radial basis, and output, all of which are made up of neurons or nodes. The input layer work as the container of input variables, radial layer contains the function of the radial basis. The output layer is linked with the radial basis nodes by linear weights and shows the output of the problem (Mustafa et al. 2012).

In Figure 6, the notation SO_2, $PM_{2.5}$, PM_{10}, O_3, NOx, NO, NO_2, CO, …... V_wind_speed represents the input layers, H(1), H(2), H(3),….. H(10) represents the hidden layers contains the radial basis function, and the Rain is the output layer. If we represent the output as R, the relation between the input variables and output for the i^{th} radial basis layer of the RBFNN model is as follows:

$$R(\boldsymbol{x}) = w_o + \sum_{\boldsymbol{i}=1}^{N} W_i \times H_i(\boldsymbol{x}) \qquad (5)(5)$$

Here, $R(\boldsymbol{x})$denotes “the output of the input vector $\boldsymbol{x}$, H_i denotes the basis function,$w_o$denotes the bias, W_i denotes the linear weight between the output layer and the hidden layer, N denotes the number of neurons in the radial basis node”.

The basis function $H_i(\boldsymbol{x})$ is the normalized-Gaussian function which is shown in following:

$$H_i(\boldsymbol{x}) = \exp\left(-\frac{\|x - \emptyset_i\|^2}{2\partial_i^2}\right) \qquad (6) \quad (6)$$

Here,||. ||represents "the Euclidean distance, ∂i are width parameter and $\emptyset_i$ is the center parameter of the basis function which was obtained by using random sampling method (Broomhead and Lowe, 1988) of the input data.$H_i(\boldsymbol{x})$**".** The matrix form of the equation 6 can be expressed by equation 7.

$$R(\mathbf{x}) = \boldsymbol{w}H \qquad (7)$$

Here, $\boldsymbol{w}$ = Weight vector.

For training the RBFNN model, two steps are followed. Firstly, established the width/ spread (∂) and the centre $\emptyset$ of the basis function from the input layer. After that adjusted the weights $\boldsymbol{w}$in order to reduce the error function (ε) by using equation 8.

$$\varepsilon = 0.5 \sum_{j=1}^{M} \sum_{i=1}^{N} \left[R_i(\boldsymbol{x_j}) - \boldsymbol{T_{ij}}\right]^2 \qquad (8)$$

Here, "$R_i(\boldsymbol{x_j})$ denotes the network output and $\boldsymbol{T_{ij}}$ denotes the target".

Using the normalization technique of spread/ width (∂) determination, the spread of the basis function (∂) was calculated, after the establishment of the centre of the basis function ($\emptyset$). The equation for ∂i is expressed as:

$$\partial i = 2 \sum_{j=1}^{N_p} \frac{|\delta_{j=1} - \delta_j|}{N_p} \qquad (9)$$

Here, "N_p = Number of centers of basis function, $\delta_{j=1}$and δ_j denotes the successive centers of the basis functions. The network weights ($\boldsymbol{w}$)were estimated using the pseudo-inverse equation once the basis function's center and spread were determined" (Bishop, 1995).

$$\boldsymbol{w}^t = [\boldsymbol{\phi}^t\boldsymbol{\phi}]^{-1} \times \boldsymbol{\phi}^t \times \boldsymbol{T} \qquad (10)$$

Here, $\boldsymbol{T} = [\boldsymbol{T}_{ij}]$, $\boldsymbol{\phi} = H_i(\boldsymbol{x})$

Multi-collinearity assessment

In regression related research, collinearity means the non-independence of predictor variables. To identify the multi-collinearity among the input variables of the prediction model of this study, we used two collinearity-diagnostic factors, known as Variance Inflation Factor (VIF) and Tolerance. They are calculated as:

$$Tolerance = 1 - R^2 \qquad (11)$$

$$VIF = (1 - R^2)^{-1} = Tolerance^{-1} \qquad (12)$$

The VIF is the inverse of tolerance (1/Tolerance), and the value of VIF is always ≥ 1. VIF values more than 10 are frequently interpreted as showing multi-collinearity, 1 < VIF <5 indicates moderate correlation among variables, and VIF ≥ 5 indicates multi-collinearity among the variables. VIF > 10 means that regression coefficients aren't very well estimated when there is a lot of multi-collinearities in the data (Talukdar et al. 2021; Shrestha, 2020).

Model assessment

In machine learning, the evaluation of the performance of the models is one of the most essential tasks. The performance of the MLP and RBFNN models was evaluated by using the "receiver operating characteristic (ROC)" curve and the area under the ROC curve (AUC). The ROC and AUC are popular tools for evaluating the performance of binary classifiers. The value of AOC is near to 1, which indicates excellent measures, while close to 0 indicates the worst measures. When the AUC is 0.5, the model has no capacity for class separation (Namdar et al. 2021).

3. Results

3.1. Multi-collinearity assessment

VIF and tolerances were utilized in this research to identify multi-collinearity among independent variables. Talukdar et al. (2021) report that if the VIF score is more than 10 and the tolerance value is less than 0.2, it's a problem. As found by the findings of the multi-collinearity test, most of the components have VIF and tolerance values of less than 10 and more than 0.2, respectively (Table 1). As a result, multi-collinearity between independent variables is not an issue. In the present research, all fifteen conditioning factors were employed to predict rainfall trends.

Table 1 Results of multi-collinearity test

Parameters	Collinearity Statistics	
	Tolerance	VIF
SO2	0.36	2.79
NO	0.77	1.28
NO2	0.84	1.18
NOX	0.20	4.93
CO	0.55	1.83
O3	0.55	1.81
PM2.5	0.66	1.49
PM10	0.20	4.93
Wind Speed	0.43	2.33
Wind Direction	0.31	3.24
Temperature	0.66	1.52
RH	0.71	1.39
Solar Radiation	0.57	1.76
BP	0.78	1.28
V_Wind_Speed	0.65	1.53

3.2. Trends of rainfall pattern

The predicted and actual rainfall in the MLP classifier were plotted and shown in Figure 3. The plot illustrates that the expected values are proportional to observed values, and the majority of the observed and predicted values show positive trends and good agreement with the perfect line of agreement. The MLP model behaved well and effectively decreased network errors to provide predicted values that were very close to the observed data.

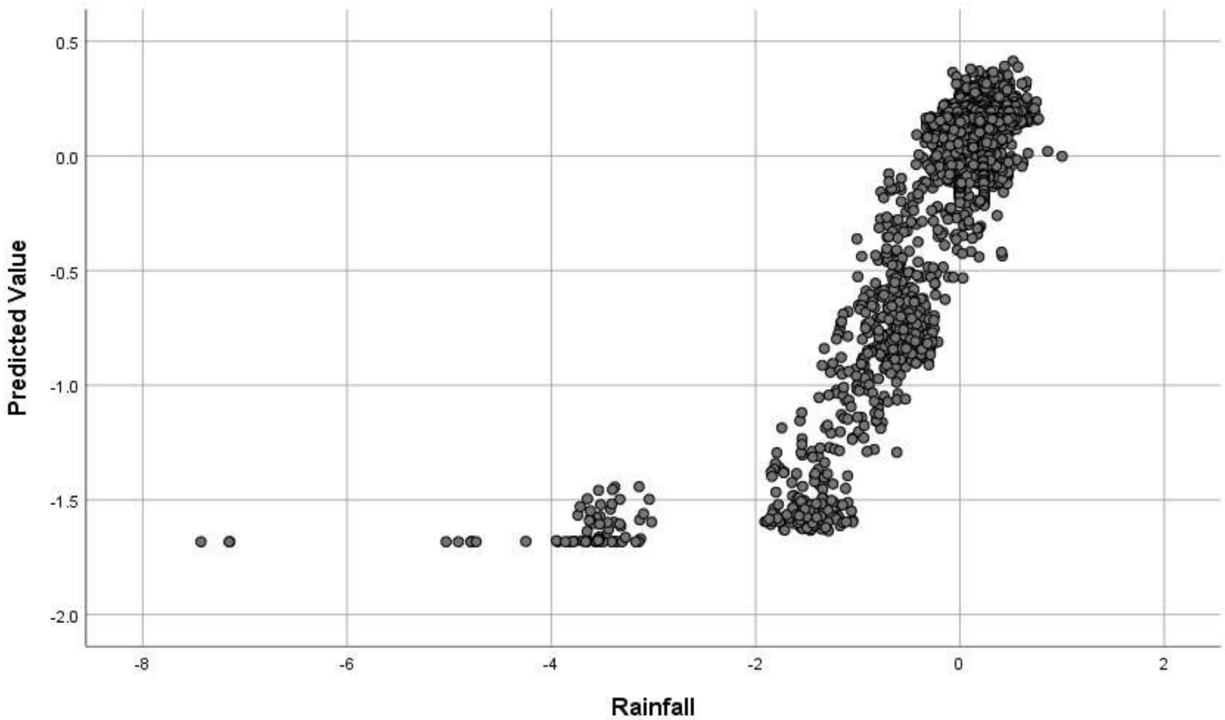


Figure 3 Comparison between predicted and actual rainfall in MLP classifier.

Figure 4 illustrates the structure of the neural network in the MLP classifier. Conditioning variables were used as inputs in the model, and fifteen units were used in the input layers. The number of hidden levels in the MLP model was determined to be one, with five units in each hidden layer. The activation function in the hidden layer was a hyperbolic tangent for the model. The model's output was rainfall, and the output layer's activation function was identity. For training data, the Sum of Squares Error and Relative Error were found to be 6741.319 and 0.341, respectively. On the other hand, Sum of Squares Error and Relative Error were both determined to be 2032.520 and 0.301 for the testing data.

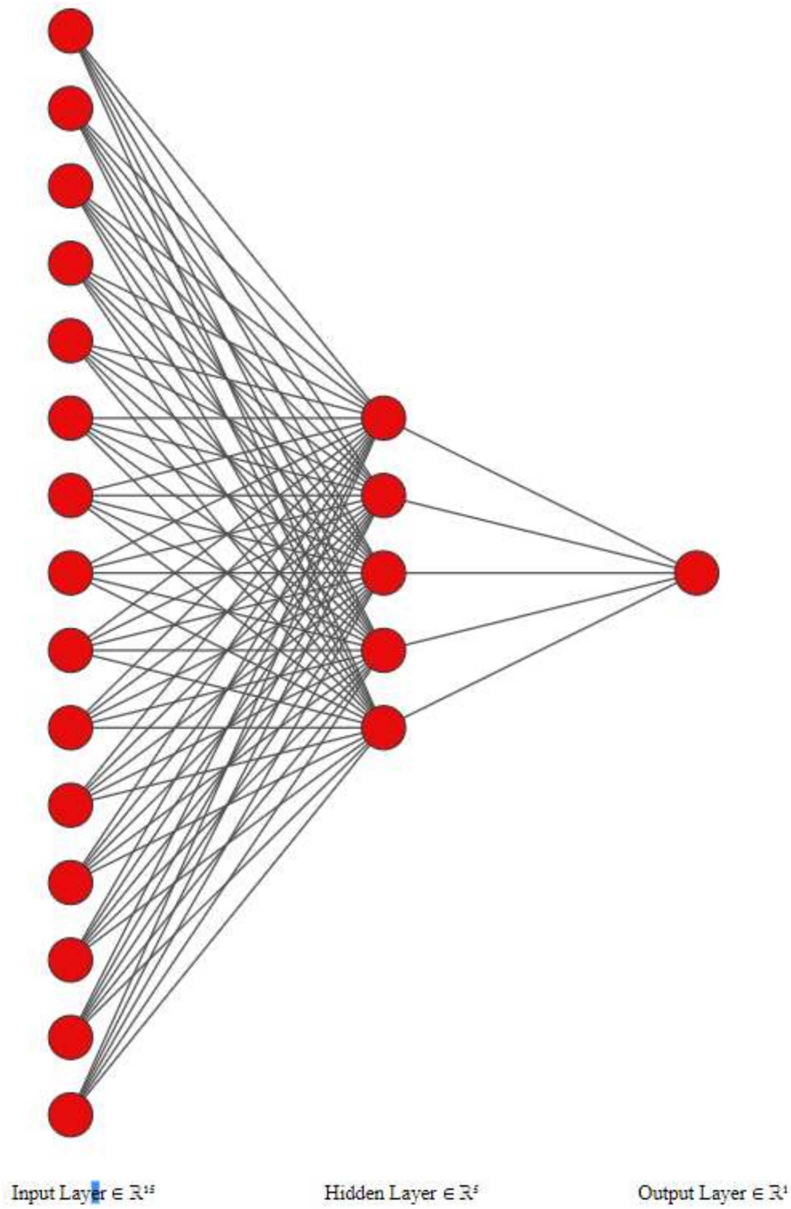


Figure 4 Structure of the neural network in MLP classifier.

Figure 5 depicts the comparison between the projected and observed rainfall based on the RBF classifier. Between all of the observed and projected data, a perfect line of agreement could be established. The concordance between predicted and observed data during testing may demonstrate the RBF model's ability to generalise successfully. The RBF model performed well, with network defects reduced to levels that were very near to those seen in the actual data.

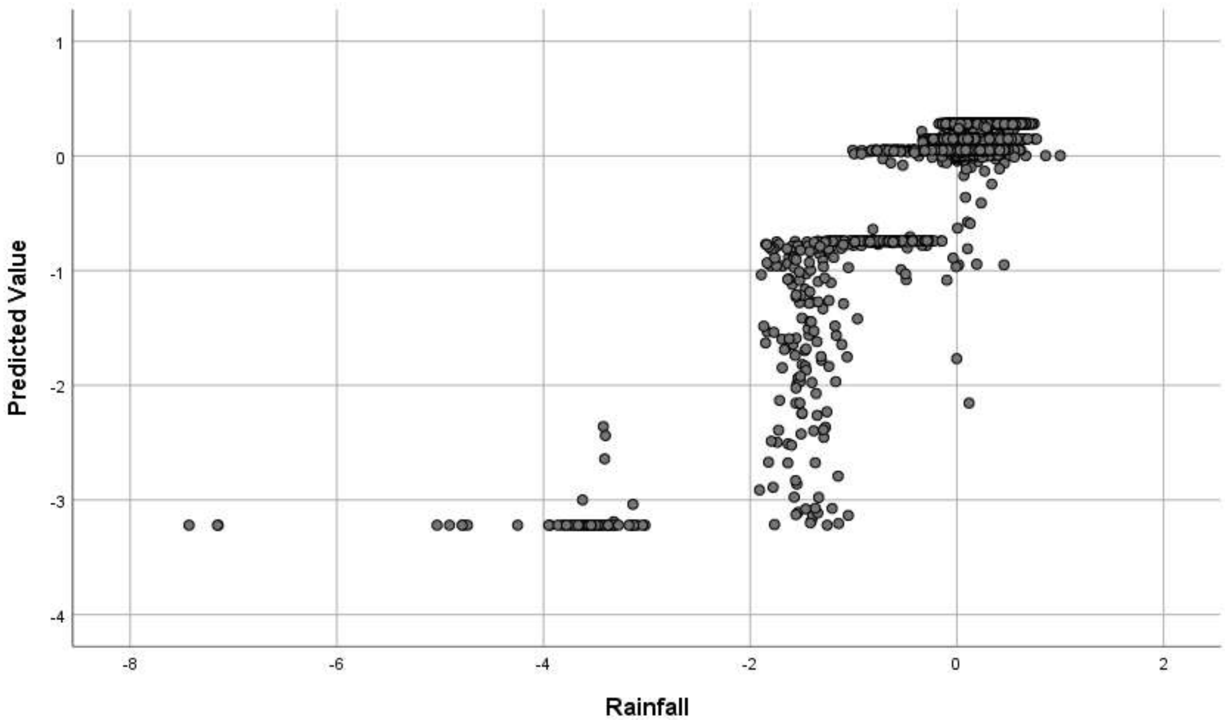


Figure 5 Comparison between predicted and actual rainfall in RBF classifier.

The architecture of the neural network used in the RBF classifier is shown in Figure 6. Model inputs were provided through conditioning variables, and fifteen units were employed in the model's initialization and initialization layers. The number of hidden levels in the RBF model was discovered to be ten, and the number of units in each hidden layer was determined to be one. The number of hidden units is the one that yields the smallest error in the testing data. The activation function in the hidden layer was a Softmax for the model, which was used to create the model. Rainfall was the model's output, and the activation function in the output layer was the same as the model's activation function. The Sum of Squares Error and the Relative Error for training data were determined to be 4092.408 and 0.207, respectively, for the training data. The Sum of Squares Error and the Relative Error, on the other hand, were both judged to be 1457.757 and 0.207 for the testing data, respectively.

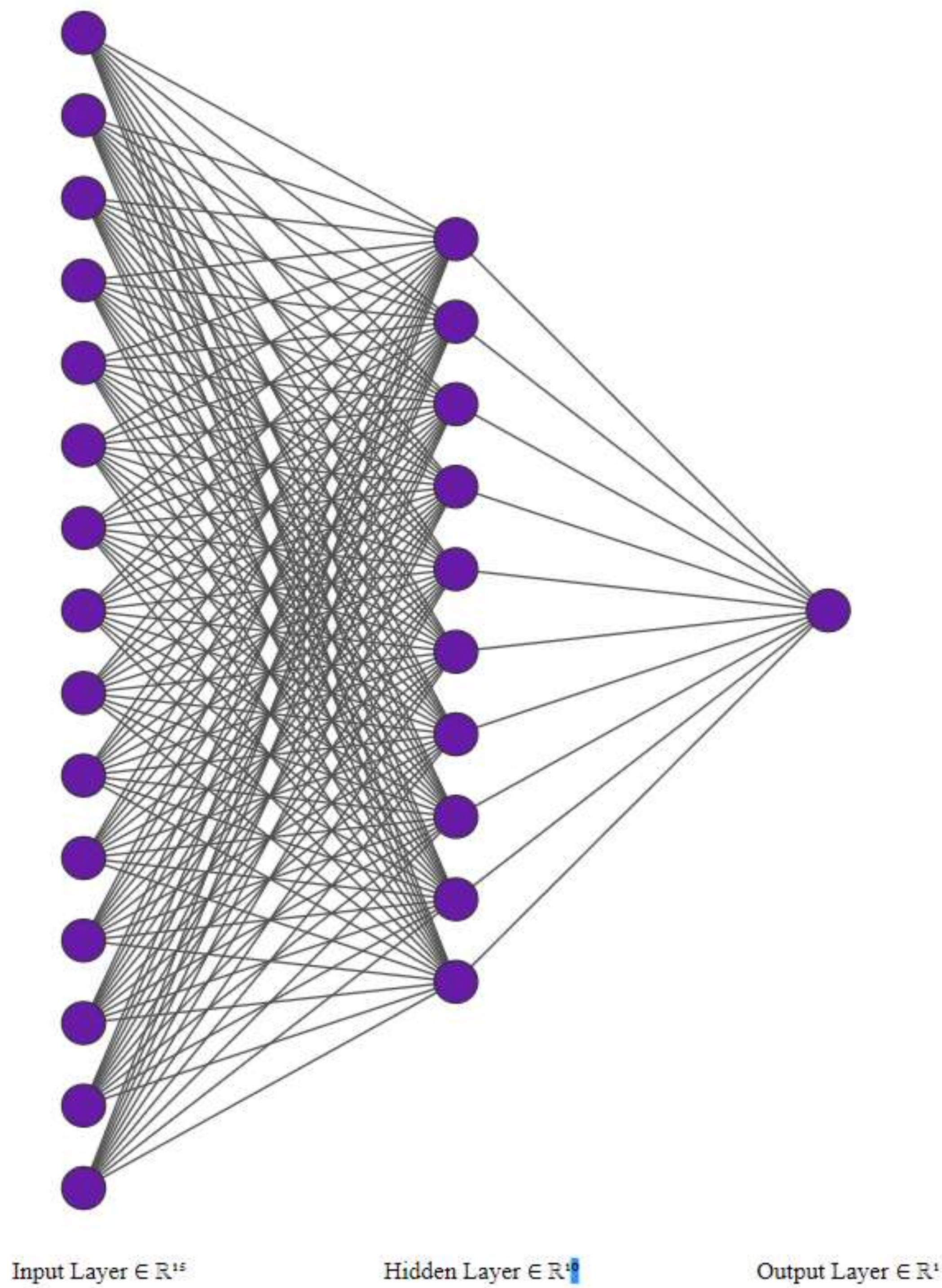


Figure 6 Structure of the neural network in RBF classifier.

3.3. Importance of the variables

The normalized importance of the conditioning variables in the MLP classifier is shown in Figure 7. The highest order of importance was found for PM_{10} (58.39), followed by O_3 (51.26), $PM_{2.5}$ (49.33), NO_X (42.37) and wind speed (41.1). Temperature and CO, on the other hand, were shown to have normalised importance of 8.26 and 8.3, respectively. According to the classification, 2, 4, 3, 1, 3, 2 variables were found for 0-10, 10-20, 20-30, 30-40, 40-50 and 50-60 normalized importance classes, respectively.

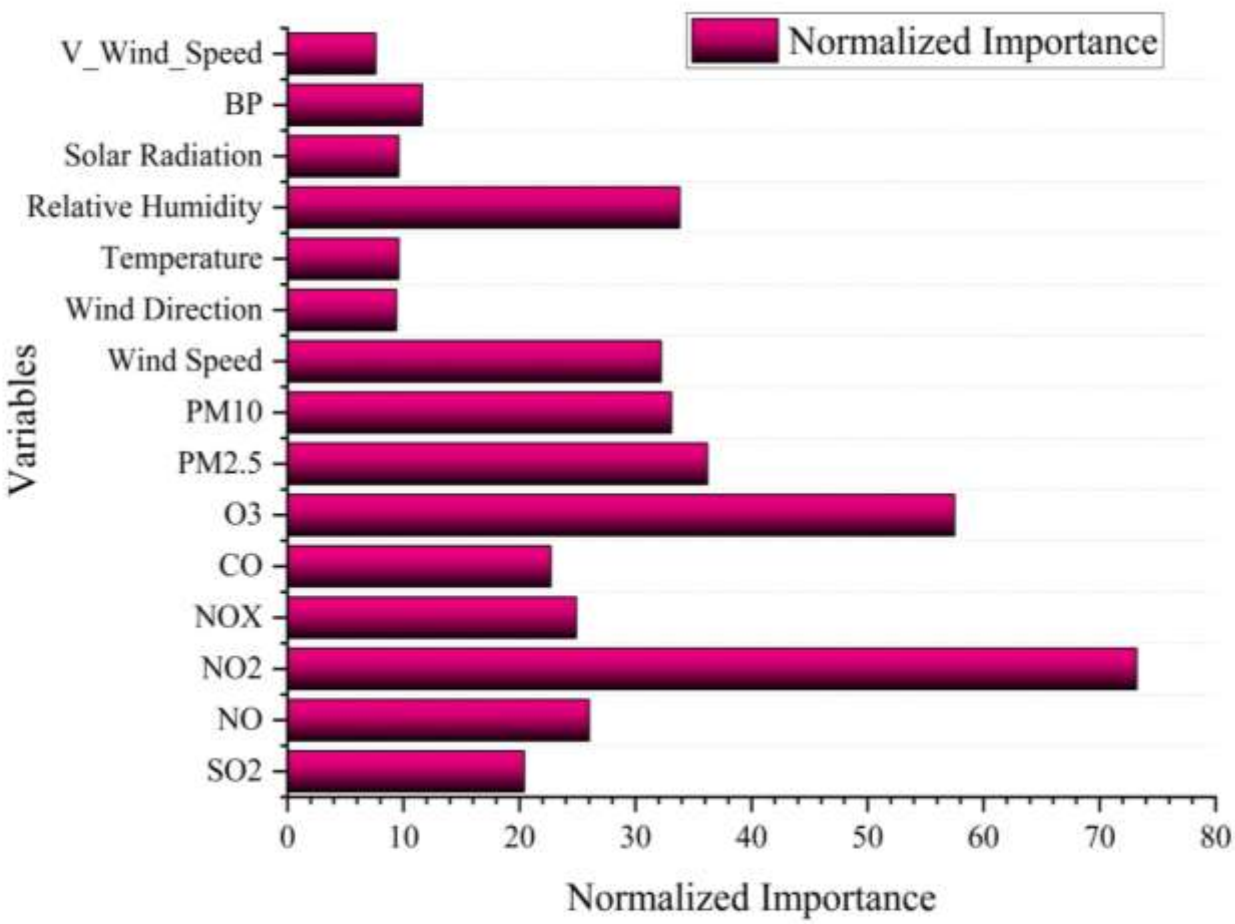


Figure 7 Importance of the variables in MLP classifier.

Figure 8 depicts the normalized importance of the conditioning variables in the RBF classifier with respect to each other. PM_{10} was shown to be the most significant component for predicting the trends of rainfall pattern. O_3, $PM_{2.5}$, NO_X, and wind speed were also influential component for the prediction under RBL model. Temperature, CO, solar radiation and wind direction were found the less important variables.

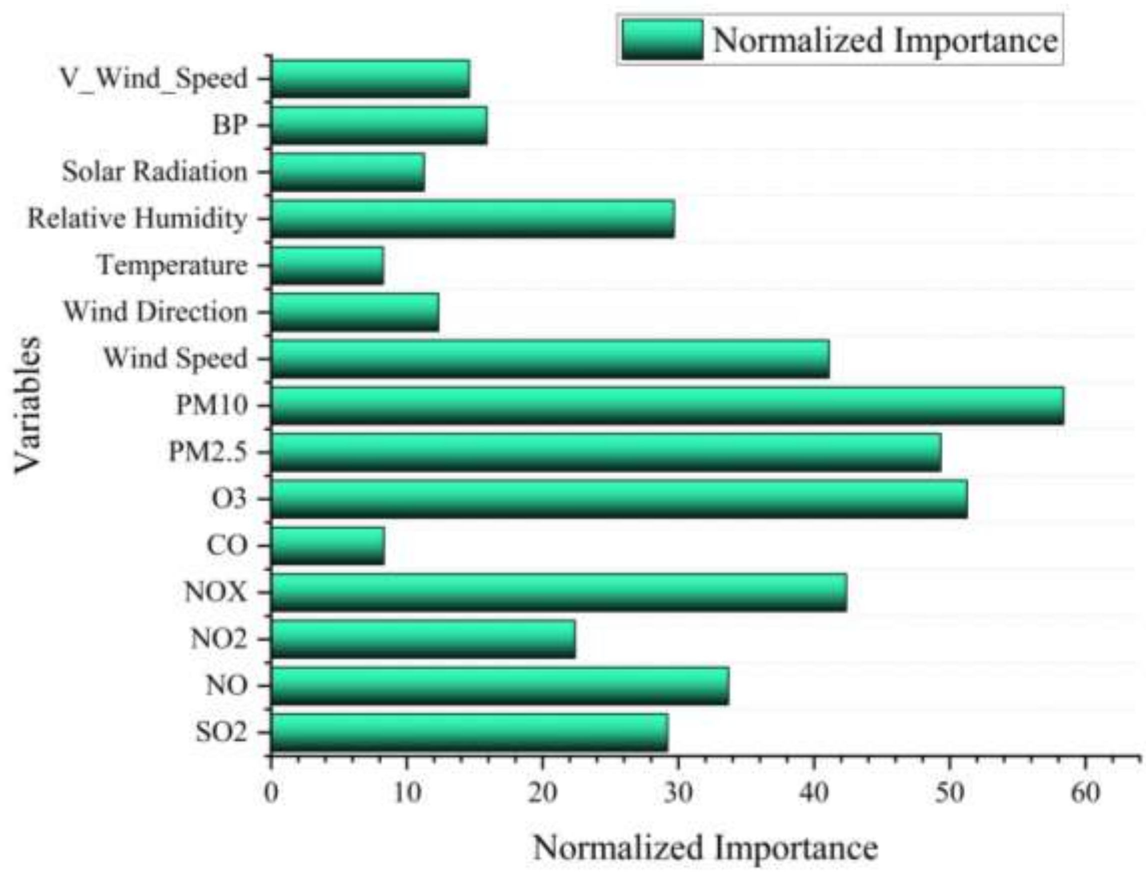


Figure 8 Importance of the variables in RBF classifier.

3.4. Accuracy of the predicted models

The predicted and residual values in the MLP classifier were plotted as shown in Figure 9. As seen in the plot, the majority of the values were close to the zero values. The MLP model showed higher accuracy in predicting the rainfall trend and reduced the network error to deliver results that were very close to the desired ones.

Figure 10 depicts a visualization of the predicted and residual values in the RBF classifier. The vast majority of the values were in the area of zero, as seen in the plot. While the MLP model also demonstrated more accuracy in predicting the rainfall pattern, it also showed lower network error, resulting in results that were very close to the anticipated ones.

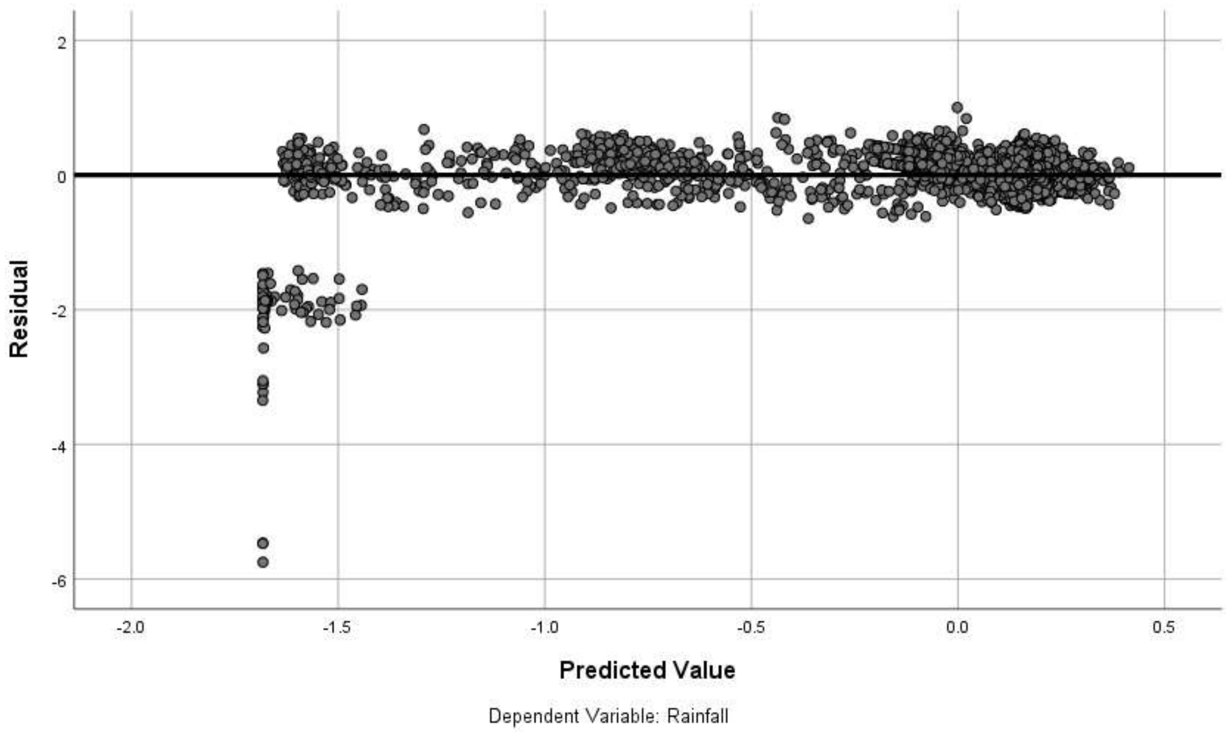


Figure 9 Accuracy of the MLP classifier.

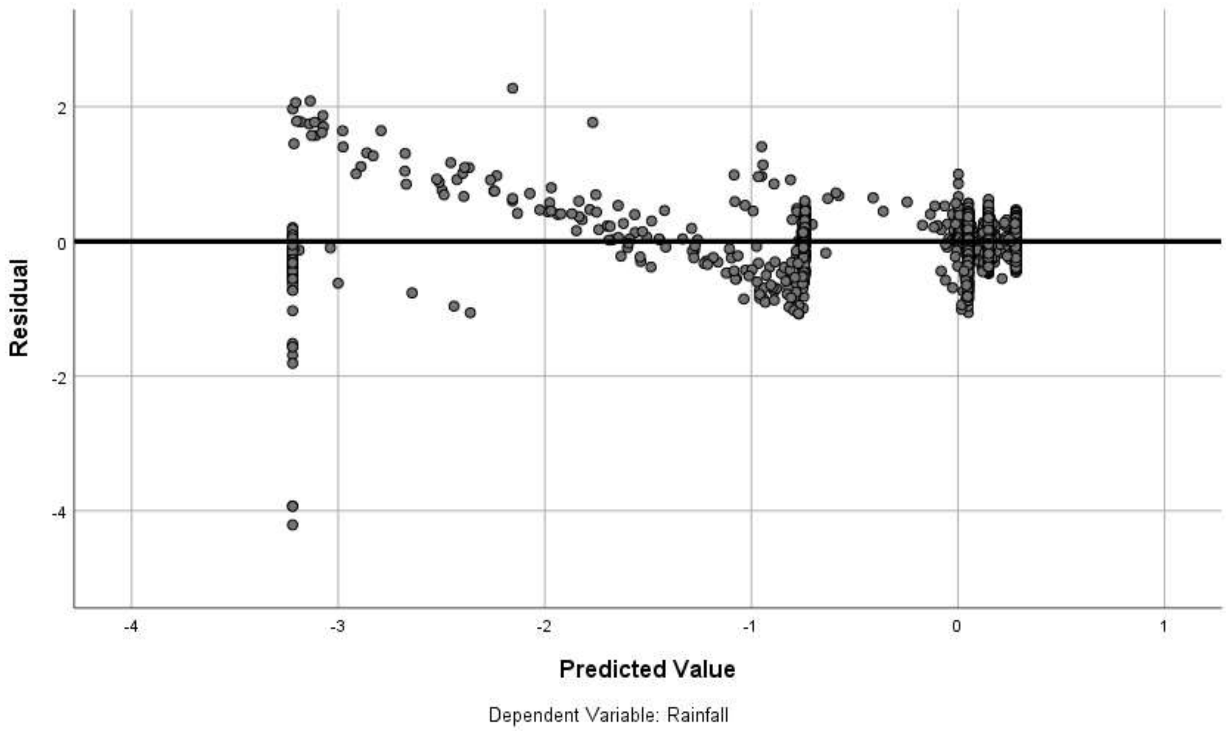
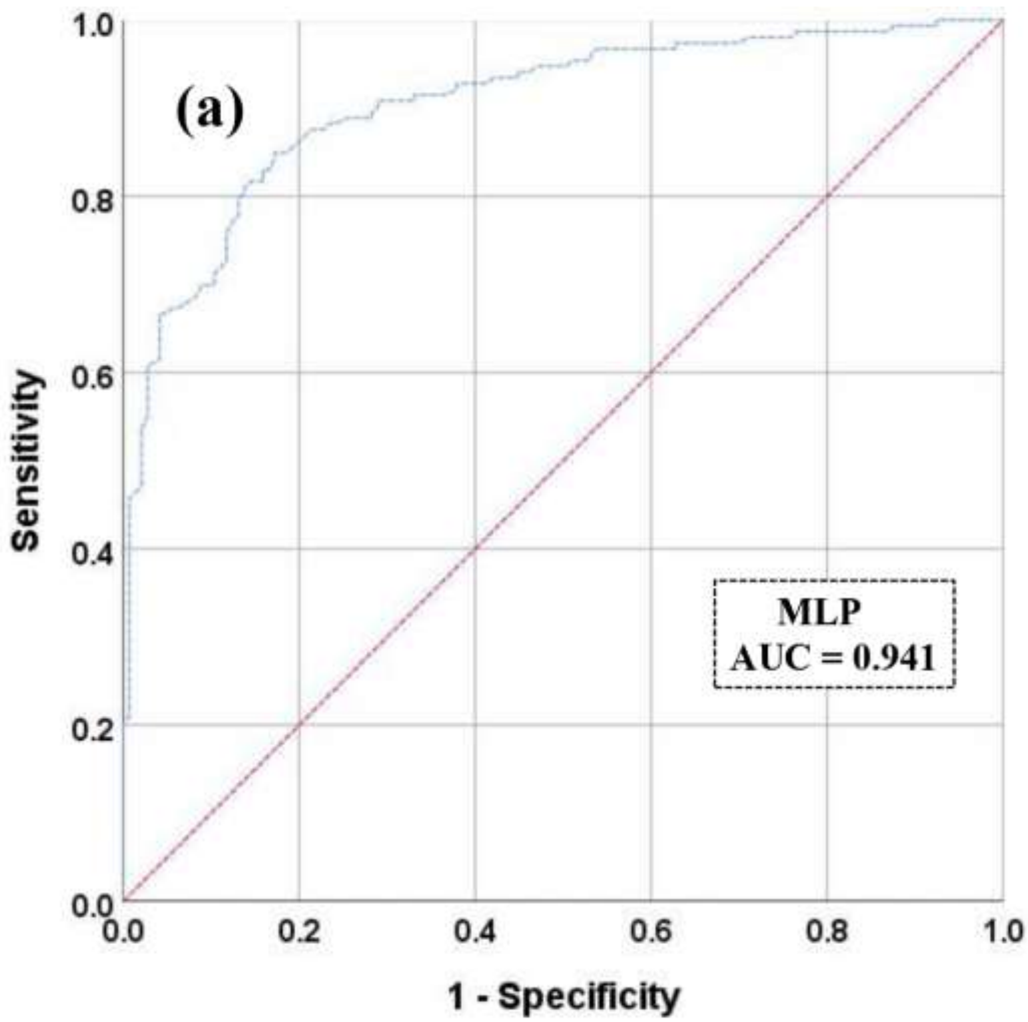


Figure 10 Accuracy of the RBF classifier.

The AUC of ROC curves was used to assess the accuracy of the predicted models. According to the results, the AUC under the ROC is 0.941 and 0.915 for the MLP and RBF classifiers, respectively (Figure 11). Despite the fact that both models produced AUC values more than 0.8, the findings might be regarded adequate. However, it can be stated that the MLP model outperformed the RBF model.


(a)
MLP
AUC = 0.941
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

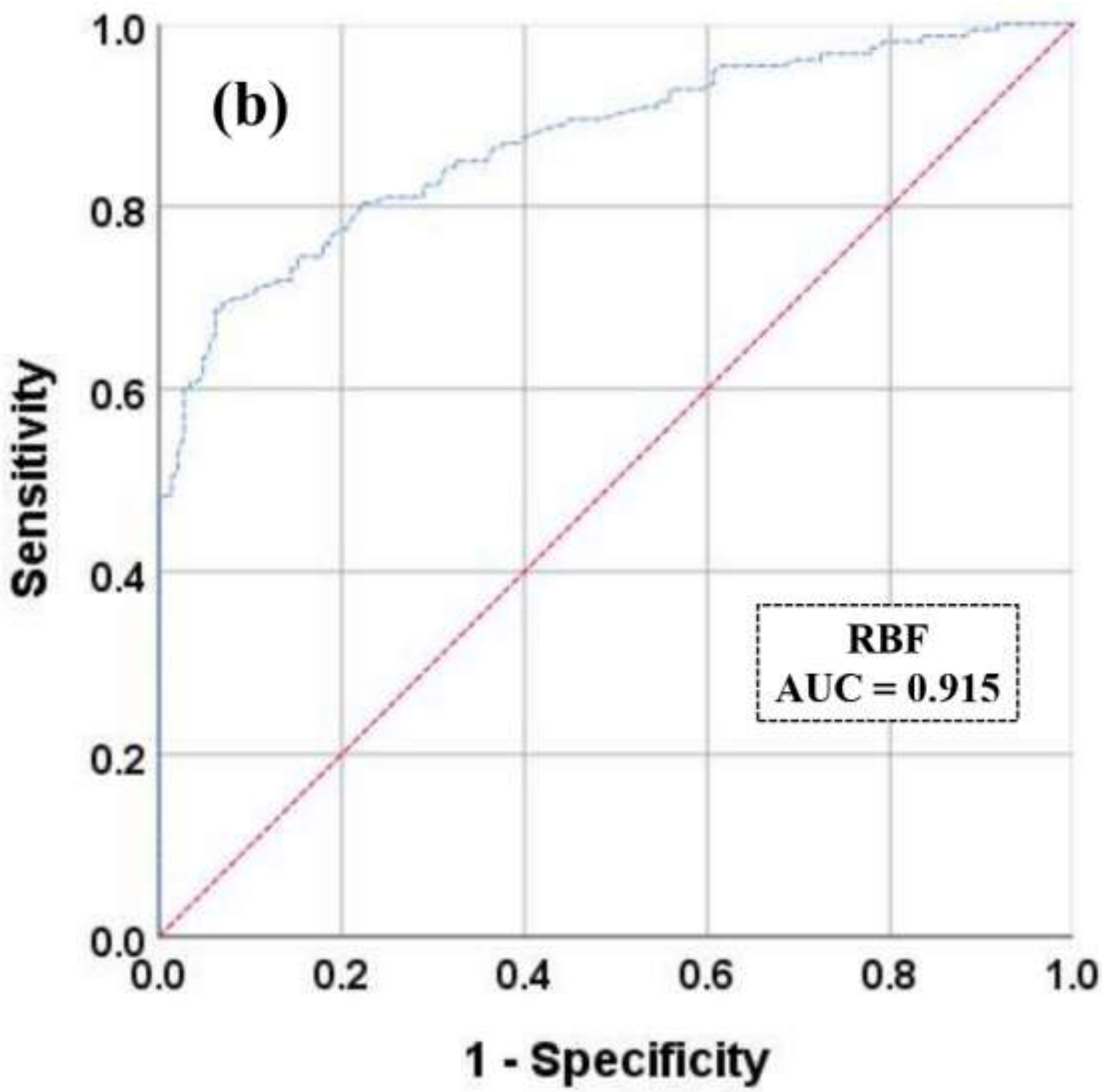


Figure 11 ROC curve of the MLP (a) and RBF (b) model.

3.5. Function of hidden layer in parameter estimates

The functions of different hidden layers in parameter estimates for the MLP and RBF classifiers are shown in Table 2 and Table 3 respectively. For the MLP classifier, fifteen input layers predict five units under a hidden layer, which were applied to predict the output rainfall. As found by the result, the fifth unit of the hidden layer has the highest weight in rainfall, followed by the fourth and third. For the RBF classifier, fifteen input layers predict ten units under ten hidden layers, which were applied to predict the output rainfall. The first hidden layer has the highest weight in rainfall, followed by the eighth and second for this classifier.

Table 2: Parameters estimates of MLP model

Predictor		Predicted: Hidden Layer 1 H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	Output Layer Rain
Input Layer	(Bias)	-.503	.502	-.556	-.338	.964	
	SO2	.014	-.224	.106	-.409	.047	
	NO	-.097	.098	.320	-.048	-.257	
	NO2	-.342	.186	.431	-.150	.246	
	NOX	.244	-.329	.469	.351	-.017	
	CO	.038	.184	-.177	.208	.008	
	O3	-.017	.257	-.064	.126	.333	
	PM2.5	-.265	-.125	-.261	.039	.192	
	PM10	-.179	.000	-.104	.148	-.023	
	Wind Speed	.520	-.464	-.700	-.286	-.043	
	Wind Direction	.236	-.300	-.105	-.222	.081	
	Temperature	.688	-.372	.604	.501	-.127	
	RH	-.170	-.506	.075	.264	-.284	
	Solar Radiation	-.097	.401	-.269	.287	.035	
	BP	.468	-.283	-.353	.144	-.077	
	Vertical Wind Speed	.103	-.126	.297	-.104	.012	
Hidden Layer 1	(Bias)						-1.430
	H(1:1)						-1.917
	H(1:2)						-1.282
	H(1:3)						.640
	H(1:4)						1.199
	H(1:5)						2.245

Parameter Estimates (MLP)

Table 2: Parameters estimates of RBF model

Parameter Estimates (RBF)												
		Predicted										
		Hidden Layer[a]										Output Layer
Predictor		H(1)	H(2)	H(3)	H(4)	H(5)	H(6)	H(7)	H(8)	H(9)	H(10)	Rain
Input Layer	SO2	.434	.416	-.080	-.138	.184	.094	.040	.025	-8.904	-11.161	
	NO	.353	-3.683	.056	-.017	.098	.117	.095	.166	-4.830	-15.732	
	NO2	.452	-2.493	.059	.076	.052	.097	.102	.388	-6.903	-15.355	
	NOX	.242	-5.258	.115	.101	.132	.146	.143	.050	-5.702	-14.000	
	CO	3.004	-1.017	1.411	.681	-.236	-.153	-.118	-.795	-2.787	.476	
	O3	.097	-1.069	-.140	-.151	2.394	-.100	.065	-.083	-5.889	-9.680	
	PM2.5	-.341	-3.678	.192	.189	.209	.216	.217	-2.267	-5.559	-13.686	
	PM10	-2.291	-2.321	-.995	-.302	.329	.337	.336	-2.319	-4.068	-10.741	
	Wind Speed	-3.221	-3.223	.264	.283	.313	.312	.315	-3.222	-3.247	-3.471	
	Wind Direction	-3.112	-3.229	.269	.289	.306	.310	.314	-3.111	-3.729	-4.810	
	Temperature	-.793	-1.060	-.719	-.595	-.456	-.498	1.720	-.806	-1.361	-2.397	
	RH	-2.840	-3.616	-.064	.200	.340	.320	.340	-2.960	-4.106	-4.344	
	Solar Radiation	-3.202	-3.277	.296	.301	.305	.305	.324	-3.215	-3.289	-3.308	
	BP	-2.382	-2.421	-2.379	-1.152	.361	.361	.544	-2.382	-2.574	-2.755	
	Vertical Wind Speed	.527	3.299	-.015	6.748	-.276	-.267	-.033	1.142	3.047	4.290	
Hidden Unit Width		.911	1.690	.735	1.708	.756	.370	.371	.787	1.849	5.331	
Hidden Layer	H(1)											1.437
	H(2)											.242
	H(3)											-.012
	H(4)											.027
	H(5)											-.018
	H(6)											-.011
	H(7)											.007
	H(8)											.748
	H(9)											-3.852
	H(10)											-16.762

a. Displays the center vector for each hidden unit.

Figure 12 illustrates the diversity of the samples under the study. For sample sizes less than 8500, the diversity increased with the increase in sample size. The diversity remained the same for the sample size of 8500–44000. For a sample size greater than or equal to 44000-55000, the diversity increased a little more and decreased again for a sample size greater than or equal to 55000. In this study, the sample size was 56649, and the diversity was close to 11. The variation of the samples from the mean value is shown in Figure 13. From which we can easily understand about the difference of the samples from the mean value. The probability of rainfall has been estimated with considering upper and lower percentile (Figure 14). From this estimation we found that there is a increasing probability of rainfall in future period. These scenarios suggest the extreme climatic condition and its associated water logging condition the study region.

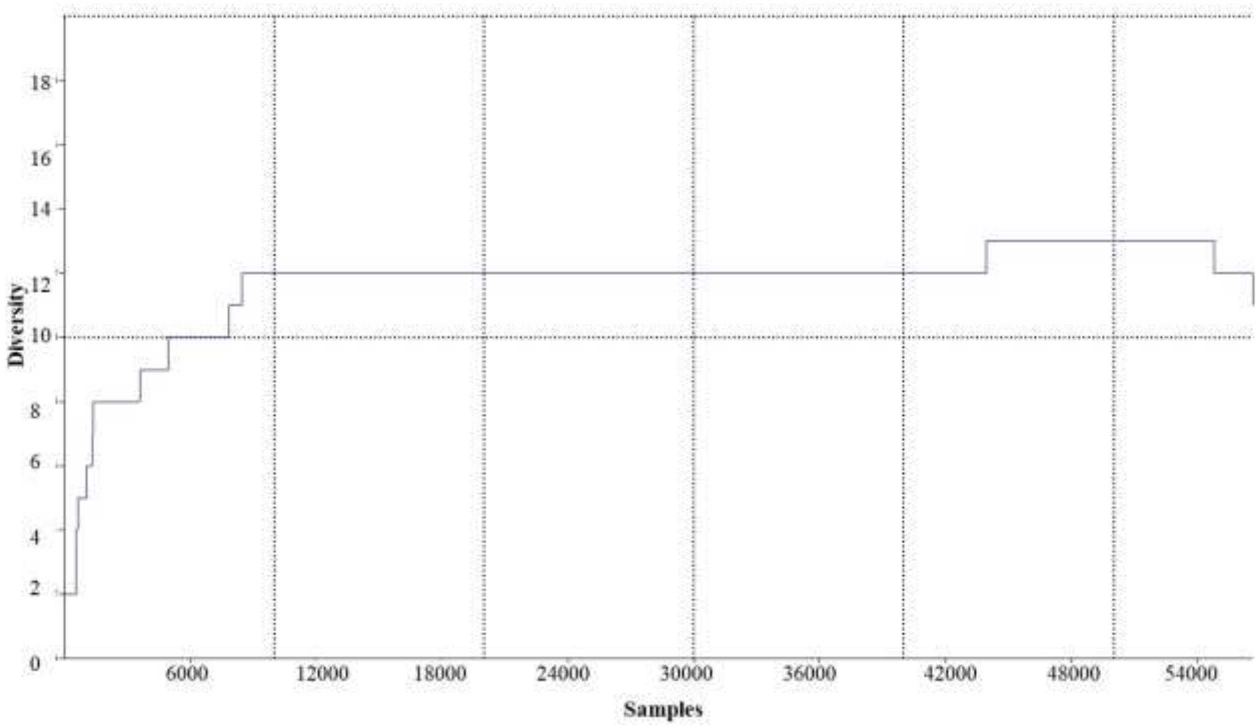
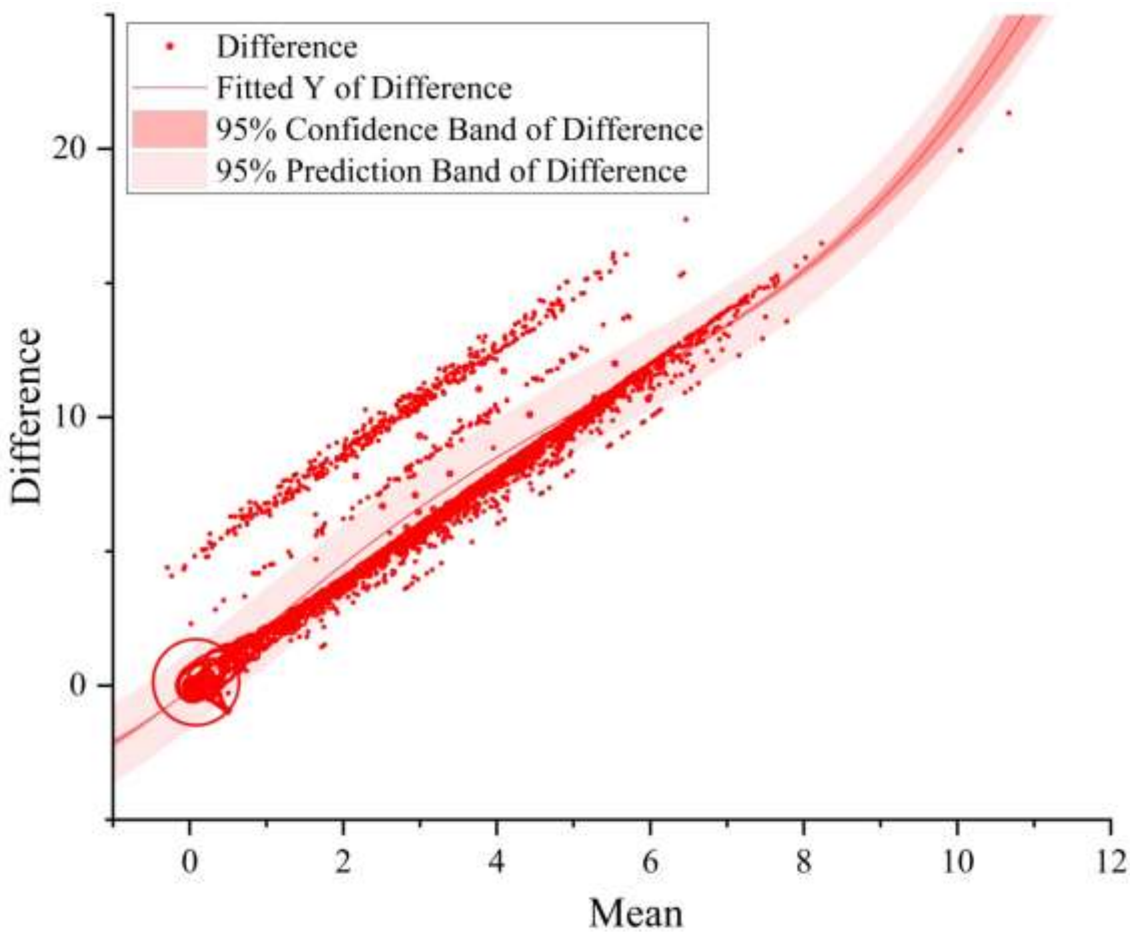


Figure 12 Diversity of samples in the current study



374

375 **Figure 13** Variations of the samples from mean value

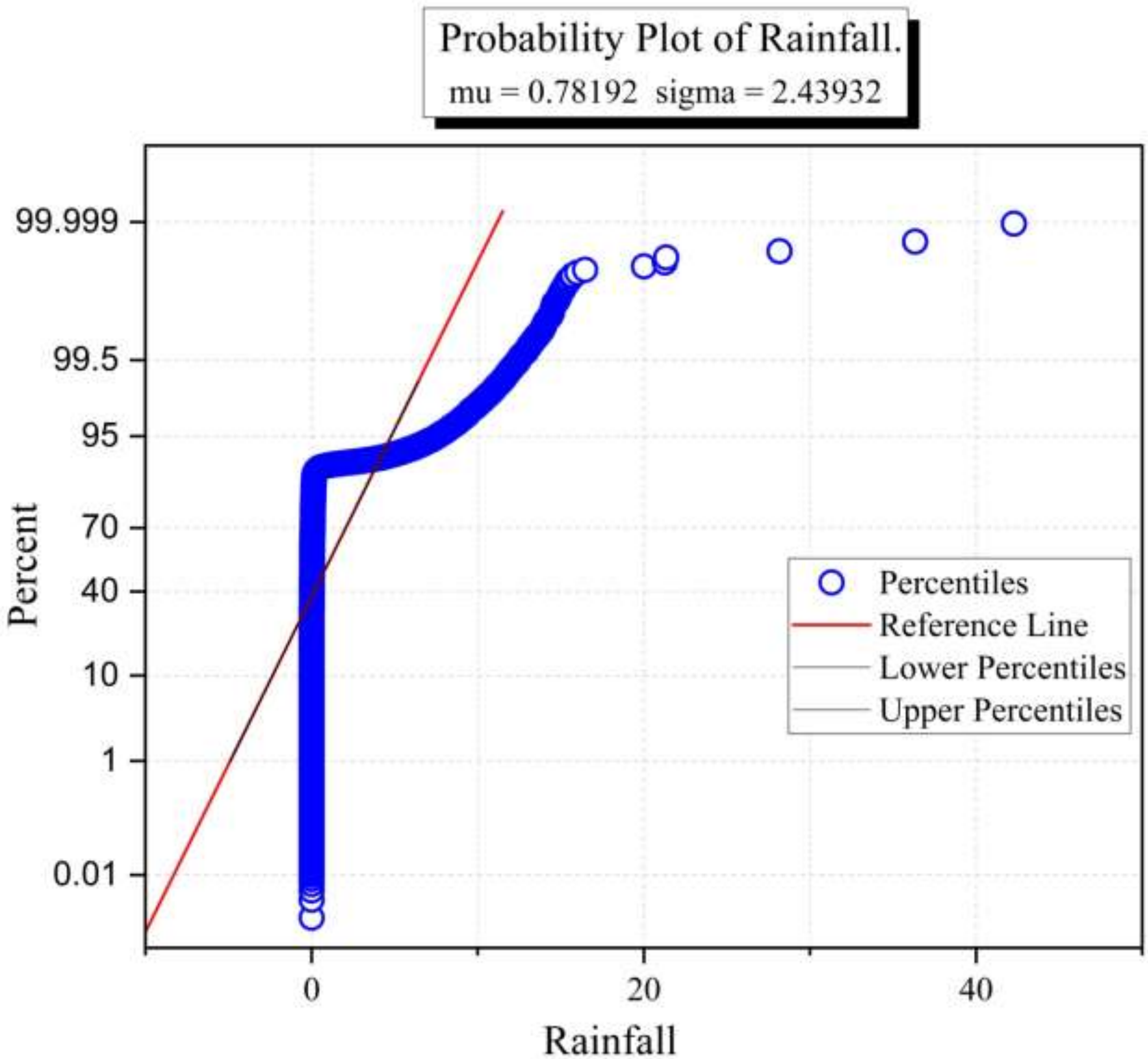


Figure 14 Probability of the rainfall in Dhaka megacity

4. Discussion

One of the most important aspects of managing climate-induced water-related problems in cities is forecasting rainfall patterns. The occurrences of rainfall in a specific region can be used to determine the overall climatic pattern. Rainfall and temperature are critical climatic inputs for economic productivity, particularly in light of climate change (Rahman and Islam 2019; Jhajharia et al. 2021; Talukdar et al. 2022). However, because of their probable interaction, precise study and modelling of the combined distribution of rainfall and temperature is challenging. Most scientists agree that there is a strong association between rainfall and temperature across tropical oceans and land (Jhajharia and Singh 2011). The most important fundamental physical elements in the climate are rainfall and temperatures, as these variables govern the environmental state of a place, which influences economic activity. Several methods are available and applied for predicting trends in rainfall patterns. This study aimed to evaluate and compare two different methods in order to select the best fit method for predicting the trends of rainfall patterns in Dhaka City. Fifteen air parameters and meteorology-related conditioning parameters were employed to predict rainfall trends and, according to VIF and tolerance values, most of the parameters have no multi-collinearity. Here, rainfall is considered as a dependent variable and the rest of the parameters have been considered as independent variables. Both the MLP and RBF models performed well, with network faults reduced to values very close to those observed in the real data. In both models, PM_{10} was the most significant component for predicting the trends of rainfall patterns.

In the last few years, the connection between rainfall and synoptic atmospheric parameters has led to the development of the weather-state model (Hughes et al. 1995; Shafie et al. 2012; Islam et al. 2020). The AUC under the ROC is 0.941 and 0.915 for the MLP and RBF classifiers, respectively. The MLP outperformed the RBF model in terms of performance. There is an increasing probability of rainfall in this region in upcoming times. So, it directly indicates the water logging condition of this study region. Citizens were reminded of the chaos that regularly surrounds metropolitan regions at this time, when roads turn into canals even if there is mild rain, as the monsoon pounded cities. During the monsoon, waterlogging, which is a precursor to urban floods, is a familiar sight in this region. Even as the weather system has changed, more high-intensity rain has fallen on fewer rain days, resulting in increased urban floods. Others have been caused by three to six-hour bouts of high-intensity rain that abruptly overwhelmed drainage infrastructure. The pattern of urban floods and waterlogging in urban areas of Bangladesh has persisted this year as well (Islam et al. 2021).

Compared to the applied MLP and the RBF models, the MLP performance is superior to the RBF, whereas the values of relative error and sum of square error for the MLP model are higher than those for the RBF model. Although the RBF model provides a lower error than the MLP model with a higher degree of efficiency in terms of error matrices, the relative error for the RBF model is consistently less than half of the relative error for the MLP model. Johnson and King (1988) reported that the MAPE of around 30% is regarded as a rational prediction as it is close to 10% and can be considered as precise, which is similar to our study.

Although the accuracy of the RBF model is less than that of the MLP model, other meteorological and environmental processes can be simulated by using both models with acceptable accuracy. The MLP model outperforms the RBF model in terms of stability. Although using artificial intelligence to predict was a viable method for generating a conservative estimate for the real monitoring dataset, the MLP model may have had a limit. The main drawback is the need for trial and error to find the best network structure. Future studies should combine the neural network model with a genetic algorithm to find the ideal neural network structure. PM_{10}, $PM_{2.5}$, and wind speed data might be used to assess the model's predicting ability.

5. Conclusion

Because of their ability to identify functional correlations between the model's inputs and outputs while explicitly considering the data generation process, machine learning-based ANN models are effective and powerful tools for simulating complex and poorly understood natural processes. In this study, we employed two ANN models of feed-forward networks, namely the MLP and RBF models, to predict the rainfall trend pattern of a South Asian megacity, Dhaka, Bangladesh. Fifteen atmospheric and meteorological factors were considered as input variables in both models. For predicting rainfall patterns in this region, rainfall is regarded as a dependent variable, whereas the remaining elements are considered independent variables. In the RBF model, ten basis radial functions were used. The multi-collinearity assessment among the input variables shows that most of the components have VIF and tolerance values of less than 10 and more than 0.2, i.e., no multi-collinearity. This type of outcome is helpful for projecting rainfall trends while considering various climatic and atmospheric parameters. The findings showed that the MLP is the leading machine learning-based ANN model compared to the RBF model in predicting rainfall trend patterns with an AUC value of 0.941.

Acknowledgement

The authors highly acknowledge the Bangladesh Meteorological Department (BMD) and Atomic Energy Commission (AEC) for providing the required datasets to conduct this study. The authors also thank the Department of Disaster Management, Begum Rokeya University, Rangpur, for providing the necessary supports to conduct this study.

Code availability

Code will be available upon reasonable request on corresponding authors

Data availability

Data will be available upon reasonable request on corresponding authors

Author contribution

Islam, A.R.M.T., designed, conceptualized, drafted the original manuscript; Pal, S.C., Chakrabortty R,, Sarkar S.K, planned the documents; Fattah, M.A., and Mallick J., involved in the literature review, software, mapping, statistical analysis, interpretation of the analysis and discussion; Pal, S.C., and Chakrabortty R., contributed to instrumental setup, data analysis, validation; Rahman M.S., contributed to data collection and extraction; Islam, A.R.M.T., Mallick, J., Fattah, M.A., and Sarkar S.K., had done the internal review and proofreading during the manuscript drafting stage.

Ethical approval

Not applicable

Consent to participate

Not applicable

Consent to publish

Not applicable

Conflict of interest

None

Funding information

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through General program under grant number (RGP2/169/43)

References

1. Batisani, N., & Yarnal, B. (2010). Rainfall variability and trends in semi-arid Botswana: implications for climate change adaptation policy. Applied Geography, 30(4), 483-489.
2. Behrang, M. A., Assareh, E., Ghanbarzadeh, A. & Noghrehabadi, A. R. (2010) The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. Sol. Energy 84, 1468–1480.
3. Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, pages (165–171) p. 482
4. Broomhead, D.S., Lowe, D., 1988. Multivariable functional interpolation and adaptive networks. Complex Systems 2, 321–355

5. Chatterjee, S., Khan, A., Akbari, H., & Wang, Y. (2016). Monotonic trends in spatio-temporal distribution and concentration of monsoon precipitation (1901–2002), West Bengal, India. Atmospheric Research, 182, 54-75.
6. Chowdhury, M.A., Hasan, M.K., Islam, S.L.U. (2021) Climate Change Adaptation in Bangladesh: Current Practices, Challenges and Way Forward. The Journal of Climate Change and Health. 100108. https://doi.org/10.1016/j.joclim.2021.100108
7. Darji, M. P., Dabhi, V. K., & Prajapati, H. B. (2015, March). Rainfall forecasting using neural network: A survey. In 2015 international conference on advances in computer engineering and applications (pp. 706-713). IEEE.
8. Du, J., & Shi, C. X. (2012). Effects of climatic factors and human activities on runoff of the Weihe River in recent decades. Quaternary International, 282, 58-65.
9. Faisal, A-A., Kafy, A-A., Rakib, AA. et al. (2021) Assessing and predicting land use/land cover, land surface temperature and urban thermal field variance index using Landsat imagery for Dhaka Metropolitan area. Environmental Challenges. 4(100192). https://doi.org/10.1016/j.envc.2021.100192
10. Fattah, M., Morshed, S.R. Assessment of the responses of spatiotemporal vegetation changes to climatic variability in Bangladesh. Theor Appl Climatol (2022). https://doi.org/10.1007/s00704-022-03943-7
11. Gajbhiye, S., Meshram, C., Singh, S. K., Srivastava, P. K., & Islam, T. (2016). Precipitation trend analysis of Sindh River basin, India, from 102-year record (1901–2002). Atmospheric Science Letters, 17(1), 71-77.
12. Gholami, A., Bonakdari, H., Zaji, A.H. et al. An efficient classified radial basis neural network for prediction of flow variables in sharp open-channel bends. Appl Water Sci 9, 145 (2019). https://doi.org/10.1007/s13201-019-1020-y
13. Gholamreza, A., Afshin, M.D., Shiva, H.A., Nasrin, R. (2016) Application of artificial neural networks to predict total dissolved. Environ. Eng. Res, Vol. 21, No. 4, pp. 333-340.
14. Hanoon, M.S., Ahmed, A.N., Zaini, N. et al. Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu state in Malaysia. Sci Rep 11, 18935 (2021). https://doi.org/10.1038/s41598-021-96872-w
15. Hughes, J. P., Lettenmaier, D. P. & Guttorp, P. □□□□ A stochastic approach for assessing the effects of changes in regional circulation patterns on local precipitation. Water Resources Research 29, 3303–331.
16. Islam, T., Rico-Ramirez, M. A., Han, D., Srivastava, P. K., & Ishak, A. M. (2012). Performance evaluation of the TRMM precipitation estimation using ground-based radars from the GPM validation network. Journal of Atmospheric and Solar-Terrestrial Physics, 77, 194-208.
17. Islam ARMT, Pal SC et al. (2022) A coupled novel framework for assessing vulnerability of water resources using hydrochemical analysis and data-driven models, Elsevier, SCOPUS/ISI Index, Journal of Cleaner Production, IF:9.29, DOI: 10.1016/j.jclepro.2022.130407

18. Islam ARMT, Rahman MS, Khatun R, Hu Z (2020) Spatiotemporal trends in the frequency of daily rainfall in Bangladesh during 1975-2017, Theoretical and Applied climatology, 141(3-4), 869-887. DOI: 10.1007/s00704-02003244-x
19. Islam ARMT, Shen S, Yang SB, Hu Z, Chu R (2019) Assessing recent impacts of climate change on design water requirement of Boro rice season in Bangladesh, Theoretical and Applied Climatology, 138(1-2):97-113, https://doi.org/10.1007/s00704-019-02818-8
20. Islam ARMT, Islam HMT, Shahid S, Khatun MK, Ali MM, Rahman MS, Ibrahim SM, Almoajel AM (2021) Spatiotemporal nexus between vegetation change and extreme climatic indices and their possible causes of change, Journal of Environmental Management, 289: 112505, doi: https://doi.org/10.1016/j.jenvman.2021.112505
21. Kuo JT, Hsieh MH, Lung WS and N. She (2007) Using Artificial Neural Network for Reservoir Eutrophication Prediction," Ecological Modelling, Vol. 200, No. 1-2, pp. 171-177. doi.10.1016/j.ecolmodel.2006.06.018
22. Johnson, D. & King, M. (1988) Basic Forecasting Techniques. Butterworth, London, pp. 50–78.
23. Kendall, M. G. (1975). Rank correlation methods. Charles Griffin and Co. Ltd., London, U.K.
24. Jhajharia D, Dinpashoh Y, Kahya E, Choudhary RR, Singh VP (2014) Trends in temperature over Godavari river watershed in southern peninsular India. Inter J Climatol 34:1369–1384
25. Jhajharia D, Singh VP (2011) Trends in temperature, diurnal temperature range and sunshine duration in Northeast India. Int J Climatol 31(9):1353–1367
26. Jhajharia et al. (2021) Pan evaporative changes in transboundary Godavari river basin, India. Theoret Appl Climatol 145:1503–1520
27. Lagouvardos, K., Kotroni, V., Defer, E., & Bousquet, O. (2013). Study of a heavy precipitation event over southern France, in the frame of HYMEX project: Observational analysis and model results using assimilation of lightning. Atmospheric research, 134, 45-55.
28. Liu, Q., Zou, Y., Liu, X., & Linge, N. (2019). A survey on rainfall forecasting using artificial neural network. International Journal of Embedded Systems, 11(2), 240-249.
29. Mallick, J., Salam, R., Islam, H.M.T. et al. Recent changes in temperature extremes in subtropical climate region and the role of large-scale atmospheric oscillation patterns. Theor Appl Climatol (2022). https://doi.org/10.1007/s00704-021-03914-4
30. Mann, H. B. (1945). Nonparametric tests against trend. Econometrica: Journal of the econometric society, 245-259.
31. McGhee, J. W. (1985). Introductory statistics. West Publishing Co., New York.
32. Memarian, H., & Balasundram, S. K. (2012). Comparison between multi-layer perceptron and radial basis function networks for sediment load estimation in a tropical watershed. Journal of Water Resource and Protection, 4(10), 870.
33. Mondal, A., Khare, D., & Kundu, S. (2015). Spatial and temporal analysis of rainfall and temperature trend of India. Theoretical and applied climatology, 122(1), 143-158.

34. Mustafa, M.R., Rezaur, R.B., Rahardjo, H., Isa, M.H. (2012) Prediction of pore-water pressure using radial basis function neural network. Engineering Geology, Vol. 135-136. Pp. 40-47.
35. Namdar K, Haider MA and Khalvati F (2021) A Modified AUC for Training Convolutional Neural Networks: Taking Confidence Into Account. Front. Artif. Intell. 4:582928. doi: 10.3389/frai.2021.582928
36. Pal, S., Talukdar, S. (2020). Modelling seasonal flow regime and environmental flow in Punarbhaba river of India and Bangladesh. Journal of Cleaner Production, 252, 119724.
37. Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., ... & Fang, J. (2010). The impacts of climate change on water resources and agriculture in China. Nature, 467(7311), 43-51.
38. Pingale, S. M., Khare, D., Jat, M. K., & Adamowski, J. (2014). Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centers of the arid and semi-arid state of Rajasthan, India. Atmospheric Research, 138, 73-90.
39. Rahman, MS, Islam, ARMT (2019) Are precipitation concentration and intensity changing in Bangladesh overtimes? Analysis of the possible causes of changes in precipitation systems, Science of the Total Environment, 690:370-387, doi: 10.1016/j.scitotenv.2019.06.529
40. Samantaray, S., Tripathy, O., Sahoo, A., & Ghose, D. K. Rainfall Forecasting Through ANN and SVM in Bolangir Watershed, India. In Smart Intelligent Computing and Applications 767-774 (2020).
41. Sangati, M., & Borga, M. (2009). Influence of rainfall spatial resolution on flash flood modelling. Natural Hazards and Earth System Sciences, 9(2), 575-584.
42. Shafie, A.H.E., El-Shafie, A.,Almukhtar, A., Taha MR, El Mazoghi, H.G., Shehata A. (2012) Radial basis function neural networks for reliably forecasting rainfall, Journal of Water and Climate Change, 3(2), 125-138. doi: 10.2166/wcc.2012.017.
43. Shrestha, N. (2020) Detecting Multicollinearity in Regression Analysis. American Journal of Applied Mathematics and Statistics. 8(2), 39-42. DOI: 10.12691/ajams-8-2-1
44. Singh, P., Kumar, V., Thomas, T., & Arora, M. (2008). Changes in rainfall and relative humidity in river basins in northwest and central India. Hydrological Processes: An International Journal, 22(16), 2982-2992.
45. Srivastava, P. K., Han, D., Rico-Ramirez, M. A., & Islam, T. (2014). Sensitivity and uncertainty analysis of mesoscale model downscaled hydro-meteorological variables for discharge prediction. Hydrological Processes, 28(15), 4419-4432.
46. Srivastava, P. K., Mehta, A., Gupta, M., Singh, S. K., & Islam, T. (2015). Assessing impact of climate change on Mundra mangrove forest ecosystem, Gulf of Kutch, western coast of India: a synergistic evaluation using remote sensing. Theoretical and Applied Climatology, 120(3), 685-700.
47. Talukdar S, Naikoo, MW, Mallick J, Praveen B, Shahfahad, Sharma P, Islam ARMT, Pal S, Rahman A (2022) Coupling geographic information system integrated fuzzy logic-analytical hierarchy process with global and machine learning based sensitivity analysis for

agricultural suitability mapping, Agricultural Systems, 196, 103343; doi: https://doi.org/10.1016/j.agsy.2021.103343

48. Talukdar S, Eibek KU, Akhter S, Ziaul SK, Islam ARMT, Mallick J (2021) Modeling fragmentation probability of land-use and land-cover using the bagging, random forest and random subspace in the Teesta River Basin, Bangladesh, Ecological Indicators 126: 107612, doi: 10.1016/j.ecolind.2021.107612
49. Talaee, P. H. (2014). Iranian rainfall series analysis by means of nonparametric tests. Theoretical and applied climatology, 116(3), 597-607.
50. Tian, Y., Bai, X., Wang, S., Qin, L., & Li, Y. (2017). Spatial-temporal changes of vegetation cover in Guizhou Province, Southern China. Chinese Geographical Science, 27(1), 25-38.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementryMaterial06052022final.docx