**IAES International Journal of Artificial Intelligence (IJ-AI)**
Vol. 14, No. 1, February 2025, pp. 182∼192
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i1.pp182-192


# Levenberg-Marquardt-optimized neural network for rainfall forecasting

**Gujanatti Rudrappa[1,2], Nataraj Vijapur[3], Sateesh Hosamane[4]**
[1]Department of Electronics and Communication Engineering, KLE Technological University Dr M. S. Sheshgiri Campus, Belagavi, India
[2]Research Centre, Department of Electronics and Communication, RV Institute of Technology and Management, Visvesvaraya Technological University, Belagavi, India
[3]Department of Electronics and Communication Engineering, RV Institute of Technology and Management, Visvesvaraya Technological University, Bangalore, India
[4]Department of Chemical Engineering, KLE Dr. M. S. Sheshgiri College of Engineering and Technology, KLE Technological University, Belagavi, India

**Article Info**

***Article history:***

Received Jan 5, 2024
Revised Jul 10, 2024
Accepted Jul 26, 2024

***Keywords:***

Artificial neural network
Levenberg-Marquardt-optimized neural network
Principal component analysis
Principal component regression
Rainfall forecasting

**ABSTRACT**

Rainfall is a crucial meteorological indicator with applications in agriculture, aviation, and military. Forecasting is essential due to unpredictable environmental changes. Current methods use complex statistical models, which are time-consuming. The present study is targeted for forecasting rainfall with the help of meteorological parameters, viz., temperature, humidity, wind speed, wind direction, and rain, using a specialized artificial intelligence (AI) method and real-time data captured over the study area. The weather station installed at KLE Dr. M. S. Sheshgiri College of Engineering and Technology in Karnataka, India, collects meteorological data. The models used were principal component regression (PCR) and Levenberg-Marquardt -optimized neural network (LMAONN). The Levenberg-Marquardt (LMA) backpropagation (BP) algorithm performed better than other BP algorithms. The coefficient of determination ($R^2$) observed for the PCR and LMAONN models were 0.57 and 0.87, respectively. The LMAONN model provided a better fit for rainfall forecasting than the PCR model, with an index of agreement (IoA) of 0.96, indicating good forecasting.



***Corresponding Author:***

Gujanatti Rudrappa
Department of Electronics and Communication Engineering
KLE Technological University Dr M. S. Sheshgiri Campus
Belagavi, Karnataka 590008, India
Email: rudraguj@gmail.com

## 1. INTRODUCTION

Rainfall is a natural phenomenon and affects the meteorology and hydrology domains. Rainfall affects many fields, such as water resource management [1], aviation industries [2], agriculture [3], and many other areas. Rainfall is affected by many factors, such as minimum and maximum temperature, humidity, atmospheric pressure, altitude, and other meteorological parameters [4].

Predicting rainfall is crucial for agriculture, aviation, and other industries. Farmers rely on rainfall records and improved forecasting skills to manage crops and boost the economy [5]–[7]. Aviation is heavily affected by meteorological parameters, and rainfall can disrupt operations [2]. Flooding is another area affected by rainfall, affecting people's lives and property [8], [9]. Forecasting rainfall helps prevent flooding, saves property, and manages water resources, helping overcome drought conditions. Numerous researchers have proposed ways for rainfall forecasting, delivering accurate and timely forecasts to enhance human mobility activities and agricultural/industrial development.

Forecasting the rain will be challenging for meteorologists due to the unpredictability of both the time and the amount of precipitation. For a decade, statisticians have used geographical coordinates (such as latitude and longitude) and other meteorological parameters (such as pressure, temperature, wind speed, and humidity) to examine the relationship between precipitation and rainfall forecasts [10]. Rainfall forecasting traditionally relies on numerical weather prediction (NWP) and statistical models [10], [11], which face challenges due to nonlinear changes in rainfall dynamics [12]. Linear regression (LR) and its variations are used, but capturing the nonlinearity of rain is challenging due to various meteorological elements [13]. A rainfall forecasting model using multiple linear regression (MLR) and multi-resolution analysis (MRA) was presented for South Australia [14], showing more accurate monthly rainfall estimations compared to traditional regression models.

Principal component analysis (PCA) has been used to overcome collinearity problems and reduce the number of independent variables in LR, often called principal component regression (PCR) [15]. However, PCR has been found to result in heterogeneous errors due to multicollinearity problems, leading to the incorporation of Dummy variables generated by partial least squares regression (PLSR). Meta-regressor XGBoost regressor was employed over the Karnataka region which outperformed the LR and random forest (RF) regression [16]. Artificial neural networks (ANNs) are more popular for forecasting nonlinear characteristics like rainfall due to their capacity for pattern learning [17], [18]. ANNs outperform traditional statistical methods in terms of accuracy and prediction performance, making them a preferred method for rainfall prediction with high accuracy. ANNs can be trained using optimization algorithms such as backpropagation (BP) and genetic algorithms, allowing for efficient tuning of model parameters and improved prediction accuracy [19].

Traditional approaches to weather prediction are greatly hindered by the nonlinear characteristics of precipitation dynamics. Although LR and statistical models have been employed, they frequently fail to grasp the intricate, nonlinear correlations present in weather data. We need more precise and dependable forecasting methods since the timing and amount of precipitation are so unpredictable. Many different types of businesses rely on accurate weather forecasts to help them prepare for and respond to unpredictable rains. There is a lack of knowledge and forecasting methodologies for other areas because most of the available research only covers one or two nations. Because of its distinct climate and the significance of precipitation for many industries, the Indian state of Karnataka requires specialized methods for predicting when and how much rain will fall.

The study describes the use of a hybrid artificial neural networks (HANN) called Levenberg-Marquardt -optimized neural network (LMAONN) for rainfall forecasting in Belagavi City, Karnataka, India. This hybrid technique improves predicting accuracy by combining the strengths of ANNs with real-time meteorological data. The study not only describes the LMAONN but also compares its performance to that of the PCR model. This comparison analysis sheds light on the efficacy of the suggested LMAONN in enhancing rainfall forecasts when compared to existing statistical approaches. By concentrating on the inefficiencies of present methodologies, the study presents to help design a more efficient and accurate rainfall forecasting model. This fills a substantial vacuum in the literature and has the potential to benefit a variety of sectors that rely on precise rainfall predictions.

The rest of the paper is organized as mentioned – section 2 provides brief details of the study area and data acquisition and pre-processing. Section 3 gives details of PCR. Section 4 gives an insight into the LMOANN model. Section 5 provides the details of the results and discussion and also gives the comparison with different models, and section 6 gives the conclusion for the proposed work.

## 2. METHOD

This section provides the details of the study area and also details on the experimental setup used for data acquisition, data pre-processing and data transformation.

### 2.1. Study area

Belagavi, formerly known as Belgaum, is a district of the State Karnataka and Country India. At a mean altitude of 779 meters, Belagavi is located in the Northwestern parts of Karnataka and is situated at the foothills of the Sahyadri Mountains. It lies close to the border of two states of India viz., Maharastra and Goa. The study area is located within Belagavi at 15°49'09.3"N 74°29'54.4"E as shown in Figure 1. Belagavi is known for its mild climate throughout the year. Winter is the coldest season (the lowest temperature in Karnataka is typically recorded in Belgaum), and the monsoon rainfall is nearly continuous from June to

September. April is sometimes marked by hailstorms in Belgaum. The minimum and maximum rain was 0.25 to 109.50 mm during the study period.

Figure 1. Satellite image of study area and weather monitoring station

### 2.2. Data acquisition

Data used in the study is acquired from May 2022 to October 2022. The station collects meteorological data every 10 minutes and the data acquired is stored in the cloud storage [20]. In total, five meteorological parameters are collected, which include temperature (T), humidity (H), wind speed (WS), wind direction (WD), and rainfall (R). The units for T, H, WS, WD, and R are degrees celsius ($^o$C), percentage (%), meters per second (m/s), degrees (deg), and millimetres (mm), respectively. All the five meteorological parameters are numerical values. Figure 2 shows the installed weather station at the location mentioned in the study area.


Figure 2. Weather station installed at study area

### 2.3. Data pre-processing

The dataset consists of 15906 data points, but due to internet, device, or sensor issues, some data points were missing or had inconsistent values. To address these issues, data cleaning was performed, replacing missing or noisy values with previous points in time. Outliers were also removed, resulting in a total of 7600 data points. The statistical values for meteorological parameters are presented in Table 1, including minimum (Min), maximum (Max), mean ($\mu$), and standard deviation (SD).

Table 1. Basic statistical parameters of the meteorological parameters

| Weather parameter (Unit) | Min | Max | $\mu$ | SD |
|---|---|---|---|---|
| Temperature ($^o$C) | 19.00 | 30.00 | 22.26 | 1.47 |
| Humidity (%) | 50.90 | 89.60 | 78.67 | 5.52 |
| Wind speed (m/s) | 0.20 | 14.10 | 3.19 | 2.052 |
| Wind direction (deg) | 0.00 | 359.00 | 190.77 | 62.38 |
| Rainfall (mm) | 0.25 | 109.50 | 9.96 | 16.84 |

### 2.4. Data transformation

Data transformation is crucial in decision-making systems, impacting model performance and interpretability. PCA is used to transform 7600 data points, reducing the number of dimensions in large datasets with numerous dimensions/features per observation [21]. This process improves data interpretability while maintaining information and facilitating multidimensional data display. PCA reduces the number of dimensions in a dataset, transforming information into a new set of principle components (PCs) values. This reduces data variation, mitigates overfitting, and removes multicollinearity using PCs generated from input data [22]. The number of PCs is the same as the number of original features. In (1) is used to estimate the standardized matrix's Eigenvalues.

$$|\mathrm{Cov(X, Y)} - \lambda \mathrm{I}| = 0 \tag{1}$$

Where $\lambda$ is Eigenvalue, and I is Indentiy matrix.

Now, if v represents the Eigenvector, then solving in (2) gives the Eigenvectors corresponding to the $\lambda$ Eigenvalues.

$$\mathrm{Cov\,(X, Y)} * \mathrm{v} = \lambda * \mathrm{v} \tag{2}$$

From the correlation matrix, the Eigenvectors are calculated. The sum of the variances explained by each Eigenvector is represented by its corresponding Eigenvalue. PCs with high orders contributing little to the overall variation are essentially noise. In (3) presents the $\mathrm{i^{th}}$ variance of the $\mathrm{i^{th}}$ PC.

$$\mathrm{Variance} = \frac{\lambda_\mathrm{i}}{\sum_\mathrm{n} \lambda_\mathrm{n}} \tag{3}$$

After obtaining all PCs, the original data set is turned into the orthogonal set by Eigenvector multiplication. This accounts for the data transformation step in the data pre-processing.

## 3. PRINCIPAL COMPONENT REGRESSION

A statistical technique called regression analysis finds the link between a dependent variable and a collection of independent (explanatory) factors [23]. The regression model can reveal whether or not changes in the dependent variable are related to any of the explanatory variables. The regression model generates the response to determine the model's unknown coefficients.

PCR is a regression method that uses PCs as input parameters. The success of PCR is contingent on the selection of the essential factors employed for regression; PCR can accurately forecast the outcome based on model assumptions by using the decided number of PCs, which will explain about 60% of the variance. In (3) represents the PCR equation.

$$\mathrm{Y} = \alpha_0 + \alpha_1 * \mathrm{PC_1} + \ldots + \alpha_\mathrm{N} * \mathrm{PC_N} + \mathrm{e} \tag{4}$$

Where, $\alpha_0 \ldots \alpha_N$ are the coefficients of the regression model estimated from the least-squares method, $PC_1 \ldots PC_N$ are the N principal components, and $e$ denotes the random error.

## 4. LEVENBERG-MARQUARDT - OPTIMIZED NEURAL NETWORK MODEL

Multilayer perceptron (MLP) ANN models are used to solve nonlinear modelling problems by transforming input data to approximate output data. These models learn from monitored data examples and have the ability to generalize. The LMAONN model, which consists of three layers, is used in this work. The input is PCs obtained by applying PCA on input parameters, while the output layer outputs rainfall forecasting results. The hidden layer is enhanced through multiple trials using various architectures, resulting in low mean square error (MSE). The LMAONN model's architecture is illustrated in Figure 3.

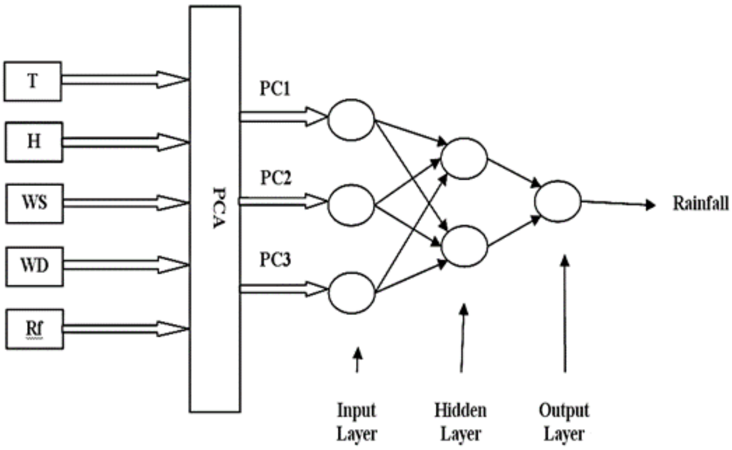


Figure 3. The architecture of the LMAONN model

## 5. RESULTS AND DISCUSSIONS

The correlation coefficient estimates the strength of the association between meteorological parameters [24]. The values are obtained by using (5).

$$\rho_{\mathrm{XY}} = \frac{\mathrm{Cov}\,(\mathrm{X}, \mathrm{Y})}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}} \tag{5}$$

Where $\rho_{XY}$ is the correlation coefficient, $\mathrm{Cov}(X, Y)$ is the covariance of variables $X$ and $Y$, and $\sigma_X$ and $\sigma_Y$ are the standard deviations of variables $X$ and $Y$, respectively.

Figure 4 shows the correlation of meteorological parameters for forecasting rainfall. It shows a negative correlation between rainfall and temperature, while humidity, wind speed, and wind direction are positively correlated to rainfall. High collinearity is observed between temperature and humidity, indicating the presence of collinearity between the independent variables. Measures such as the index of agreement (IoA), correlation coefficient (R), root mean square error (RMSE), normalised root mean square error (NRMSE), mean absolute error (MAE), R-squared ($R^2$), and mean square error (MSE) are employed to assess the forecasting model's performance [25].

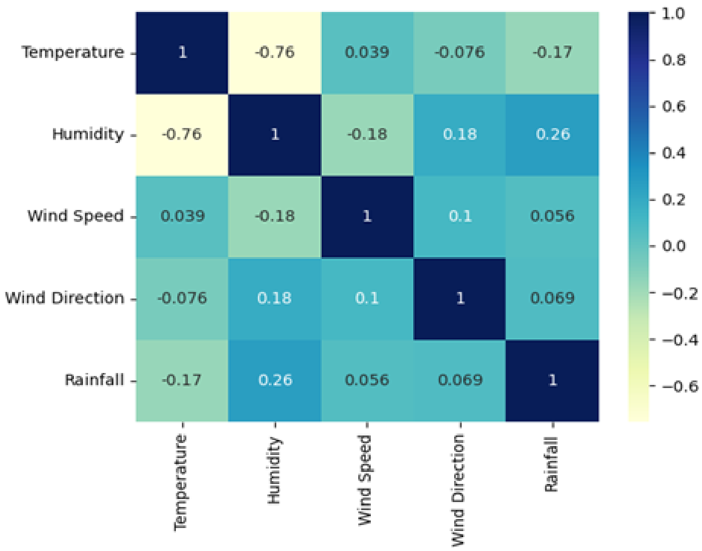


Figure 4. Correlation heatmap

The PCA analysis was conducted using Minitab17 statistical software to extract critical information (PCs) from data and rearrange the dataset description. PCA is advantageous in minimizing the dimension of input vectors, as it minimizes the redundancy of the vector components. Table 2 provides the PCA summary along with each PC's explained variance. The first and second PCs explain 38% and 22% of the total variation of parameters, respectively, making up 60% of the total variation used in forecasting models. Temperature and humidity explain 63% and 67% of rain variability, respectively, from the first PC, while wind speed and direction explain 75% and 57% of rain variability from the second PC.

Table 2. PC values, Eigenvalues, and the variance of various parameters (PCA)

| Variable | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Temperature | -0.631 | 0.086 | -0.021 | 0.408 | 0.654 |
| Humidity | 0.671 | -0.101 | 0.071 | -0.097 | 0.724 |
| Wind speed | -0.116 | 0.750 | -0.217 | -0.594 | 0.154 |
| Wind direction | 0.196 | 0.569 | 0.681 | 0.401 | -0.114 |
| Rainfall | 0.315 | 0.310 | -0.695 | 0.557 | -0.106 |
| Eigenvalue | 1.918 | 1.129 | 0.917 | 0.821 | 0.211 |
| Proportion of variance | 0.384 | 0.226 | 0.184 | 0.164 | 0.042 |
| Cumulative | 0.384 | 0.610 | 0.793 | 0.958 | 1.000 |

LR model with the help of two PCs as input variables are used to forecast rain, which is the PCR model. The regression equation used for rainfall forecasting using the PCR model is given by (6).

$$\text{Rainfall}{=}-83.641 + 2.1324{*}\text{vPC1} - 0.71003{*}\text{vPC2} \quad (6)$$

The degree of determination of the regression lines with the PCR model was 0.76 which is shown in Figure 5.

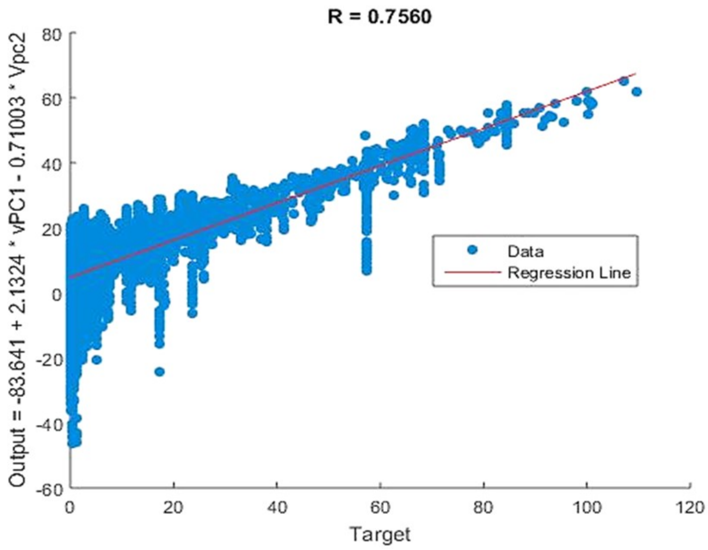


Figure 5. Validation for rainfall forecasting using PCR model

### 5.1. Levenberg-Marquardt - optimized neural network model

The three-layered LMAONN with BP is utilized for forecasting rainfall. The first two PCs, which account for 60% of the total parameter variation in PCA, are used for forecasting. The ANN's performance is enhanced by applying PCA to the input parameters [26]. The original data is divided into three subsets: 70% training, 15% testing, and 15% validation. The network's weights are initialized, and adjusted using the BP algorithm to obtain the lower MSE. The significant component transformation matrix transforms future inputs. The development of BP networks has significantly improved the performance of the selected architecture. BP-type networks use a gradient approach to MSE, providing a gradient method with the gradient of MSE [27].

A three-layer network was created and trained considering 8 BP algorithms. Table 3 summarizes the MLP-BP designs that produced the best LMAONN models. The top-performing LMAONN models involved the usage of a logistic activation function in the hidden layer. It was found that the Levenberg-Marquardt (LMA) BP algorithm is very efficient for the present problem when compared to other BP algorithms as the robust BP algorithm and the conjugate gradient algorithm. LMA was proven to provide a lesser MSE of 40.39. The LMA works best for these situations, and the approximation was precise.

Table 3. Performance indices for the 8 BP algorithms

| Index | BFG* | CGB* | CGF* | CGP* | GD* | GDM* | RP* | LMA* |
|---|---|---|---|---|---|---|---|---|
| MAE | 11.09 | 4.46 | 4.21 | 4.29 | 11.09 | 7.73 | 3.93 | 3.04 |
| MSE | 438.34 | 46.08 | 44.19 | 44.50 | 438.34 | 232.48 | 42.06 | 40.39 |
| RMSE | 20.93 | 6.78 | 6.64 | 6.67 | 20.93 | 15.24 | 6.48 | 6.35 |
| NRMSE | 0.19 | 0.06 | 0.06 | 0.060 | 0.19 | 0.13 | 0.06 | 0.05 |
| IoA | 0.38 | 0.95 | 0.96 | 0.96 | 0.38 | 0.84 | 0.96 | 0.96 |
| R | - | 0.92 | 0.92 | 0.92 | - | 0.75 | 0.93 | 0.93 |
| $R^2$ | 0.39 | 0.85 | 0.85 | 0.85 | 0.39 | 0.26 | 0.86 | 0.87 |

* BFG - BFGS-Quasi-Newton Backpropagation; CGB - Powell-Beale conjugate gradient Backpropagation; CGF - Fletcher-Reeves conjugate gradient Backpropagation; CGP - Polak-Ribiere conjugate gradient Backpropagation; GD - Batch Gradient Descent; GDM - Batch Gradient Descent with Momentum; RP - Resilient Backpropagation; LMA - Levenberg-Marquardt Backpropagation

The study aimed to optimize a network for rainfall forecasting by varying the number of neurons in the hidden layer. The number of neurons and the MSE were optimized, with twelve neurons making great predictions. The weight update with rotation pattern improved the performance of most trained networks. The network performed exceptionally well with changing neurons at the hidden layers, with $R^2$ values of 0.87, 0.86, 0.88, and 0.86, respectively for training, testing, and validation. The method by which rain is estimated most effectively explains its superior performance. The logistic activation function was used in a single hidden layer of 12 neurons in LMAONN models designed to forecast rainfall. The results showed that twelve neurons provided a lower MSE of 40.39 for the LMAONN model for each subset. Reducing the number of neurons below or above twelve increased the MSE values. Optimization of the number of neurons is crucial for the efficiency of the developed network. Therefore, the number of neurons and the MSE were optimized. The network optimization was carried out by guessing neurons in each hidden layer. The MSE values for varying numbers of neurons are presented in Table 4.

Table 4. Performance indices for varying numbers of neurons in the network for selected BP algorithm

| | Number of neurons | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| MAE | 4.24 | 4.08 | 4.09 | 4.10 | 4.09 | 4.01 | 4.02 | 4.01 | 3.99 | 3.97 | 3.04 | 4.02 |
| MSE | 45.68 | 41.96 | 41.95 | 41.88 | 42.11 | 41.07 | 40.85 | 40.69 | 41.00 | 40.98 | 40.38 | 41.10 |
| RMSE | 6.75 | 6.47 | 6.47 | 6.47 | 6.48 | 6.40 | 6.39 | 6.37 | 6.40 | 6.40 | 6.35 | 6.41 |
| NRMSE | 0.061 | 0.059 | 0.059 | 0.059 | 0.059 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 | 0.057 | 0.058 |
| R | 0.92 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 | 0.932 | 0.93 | 0.94 | 0.93 |
| $R^2$ | 0.85 | 0.86 | 0.86 | 0.86 | 0.86 | 0.86 | 0.87 | 0.87 | 0.86 | 0.87 | 0.88 | 0.86 |
| IoA | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.97 | 0.96 |

In Figure 6, the logarithms of the expected rain are plotted against the logarithms of the actual rain. The fact that the points tend to cluster along lines close to the 45-degree tangent line shows that the model is good. The model has the lowest $R^2$ (0.26) with the GDM BP algorithm and has the highest $R^2$ value of 0.88 for LMA. The PC application benefits the LMAONN model because it helps explain the change in the predictor. In this case, the hybrid method chose the meteorological variables temperature and humidity as the most critical predictor variables based on the first PC and wind speed and wind direction as the most vital predictor variables based on the second PC during model development.

The coefficient of correlation (R) for the model was 0.933, 0.931, 0.937, and 0.933, respectively, for training, testing, validation, and overall data. This is represented in Figure 7. Tests were carried out for 1000 epochs for the LMAONN model considering 12 neurons and the LMA algorithm. The best validation performance (MSE) of 39.96 was obtained at 52 epochs. The validation performance for the LMAONN model is shown in Figure 7.

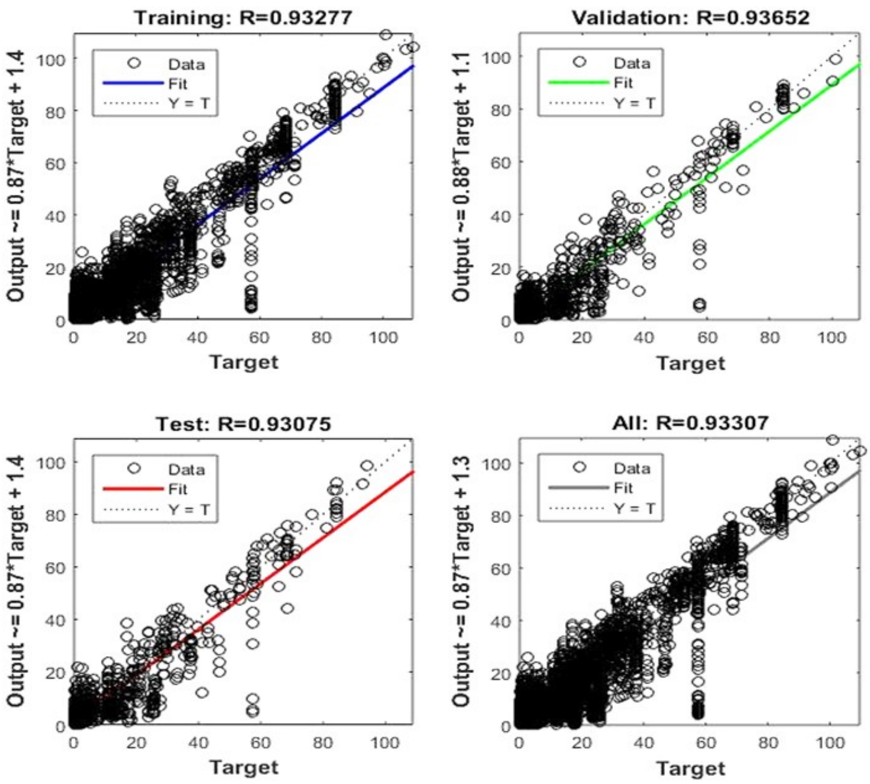


Figure 6. Training, testing, and validation for rainfall forecasting using the LMAONN model

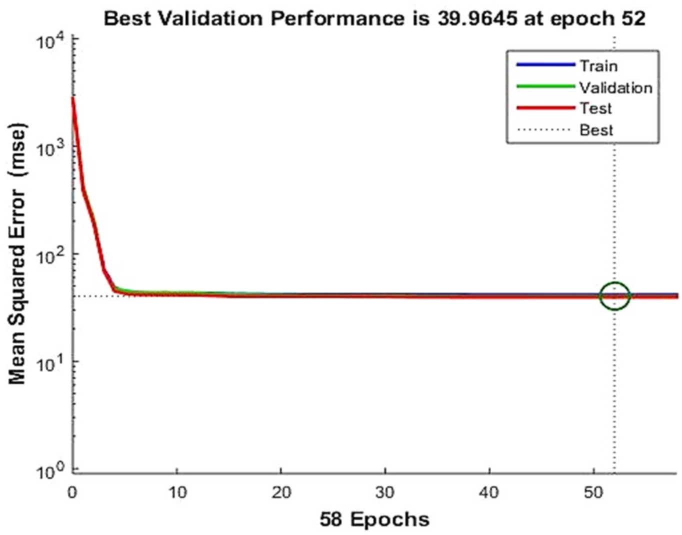


Figure 7. Training, testing, and validation MSE for rainfall forecasting using the LMAONN model

### 5.2. Comparison between principal component regression and LMAONN models

Various statistical evaluations were used for the PCR and LMAONN models to measure the goodness of fit. The performance metrics are obtained by comparing the actual and the forecasted rainfall estimates. Table 5 summarizes the statistical evaluation indices. The LMAONN model demonstrated a better fit for the problem than the PCR model, with $R^2$ values of 0.57 and 0.87, respectively. The fractional bias (FB) for both models was 0.0002 and 0.0037, indicating they greatly predicted observed rainfall. The LMAONN model's performance was evident in its MAE (3.04), MSE (40.38), RMSE (6.35), and $R^2$ (0.88), indicating its ability to accurately forecast rainfall in the Belagavi region

Table 5. Performance indices for comparing PCR and LMAONN models

| Index | PCR | LMAONN | Index | PCR | LMAONN | Index | PCR | LMAONN | Index | PCR | LMAONN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MAE | 8.9789 | 4.046 | RMSE | 11.614 | 6.355 | R | 0.756 | 0.934 | IoA | 0.8430 | 0.965 |
| MSE | 134.880 | 40.387 | NRMSE | 1.0470 | 0.0581 | $R^2$ | 0.572 | 0.8719 | FB | 0.0002 | 0.0037 |

The results match those of other studies done by a comparable group of researchers. Rainfall in Western Australia was forecast using LR [28]. Peak $\mathrm{R}$-values were between 0.47 and 0.53, and trough R-values were between 0.82 and 0.94. A reasonable range of $\mathrm{R}$ values (0.71-0.91), $\mathrm{RMSE}$ (10.68–23.08), $\mathrm{MAE}$ (9.00–20.90), and $\mathrm{d}$ (0.56-0.71) were produced by the use of non-linear regression (NLR) modelling for forecasting rainfall in the Australian Capital Territory [29]. For rainfall forecasting in the Indramayu district, PCR was employed, leading to an $\mathrm{R}$-value of 0.75 and an $\mathrm{RMSE}$ of 28.84 [15]. A better performance was obtained by applying meta-regressor XGBoost (XGB) over the Karnataka region [16]; the performance measures in terms of MAE, RMSE, R, and $R^2$ were 94.53, 158.27, 0.92 and 0.84 respectively. Rainfall prediction was obtained by the application of rainfall forecasting for the Semarang region by applying ANN [18]; the model's performance in terms of RMSE and MAE was 13.05 and 6.62 respectively. The Indus Basin in Pakistan used an ANN to predict when it will rain, and the results showed $\mathrm{R}^2$ values between 0.19 and 0.91 and prediction accuracies between 0.85 and 0.82 [30]. Rainfall forecasting in the research location (Belagavi) was well handled by the suggested LMAONN model, which utilized 12 neurons. The model's performance metrics is shown in Table 5. The comparison of LMAONN model with other models is shown in Table 6.

Table 6. Comparison of the various models discussed

| Performance measure | LR [28] | NLR [29] | PCR [15] | XGB [16] | ANN [18] | ANN [30] | LMAONN |
|---|---|---|---|---|---|---|---|
| $MAE$ | 13.06-23.52 | 9.00–20.90 | 15.06 | 94.53 | 6.62 | 7.03-24.75 | 4.05 |
| $RMSE$ | 21.21-36.45 | 10.68–23.08 | 28.84 | 158.27 | 13.05 | 16.46-26.01 | 6.36 |
| R | 0.47-0.92 | 0.71-0.91 | 0.75 | 0.92 | 0.89 | 0.44-0.89 | 0.93 |
| $R^2$ | 0.22-0.85 | 0.51-0.82 | 0.57 | 0.84 | 0.79 | 0.19-0.79 | 0.87 |

## 6. CONCLUSION

Two predictive models, PCR and LMAONN, were developed for rain forecasting at Belagavi, Karnataka and were compared with state-of-the-art models. The statistical analysis of meteorological data is evaluated for the significance of the collected data. The data was dimensionally reduced by applying PCA, and the PCs' results were used for the PCR model. The PCR model performance was relatively good, with an $R^2$ of 0.57. The PCR model performance was improved using a three-layered LMAONN model with eight BP algorithms. Based on the optimization and model performance, the LMA BP algorithm performed well with twelve neurons in hidden layers. The developed PCR and LMAONN performance is evaluated with the coefficient of determination ($R^2$) and was found to be 0.57 and 0.87, respectively. It was observed that PCA helped in capturing the variability of rainfall. MSE, RMSE, and MAE for the PCR model were 134.88, 11.61, and 8.98, respectively, whereas, for the LMAONN model, the values were 40.39, 6.355, and 4.046, respectively. The present study aims to enhance the precision of rainfall forecasting models by combining neural networks with other methodologies like fuzzy logic, evolutionary algorithms, and ensemble methods. By broadening the scope of meteorological factors and utilizing remote sensing technology like satellite imaging and radar data, the study aims to capture geographical and temporal fluctuations in precipitation patterns, particularly in areas with limited data or during severe weather occurrences. This approach can be applied to various domains such as farming, aviation, water resource management, and flood management.

## ACKNOWLEDGEMENTS

The authors express gratitude to VGST for funding their ”Establishment of IoT and Artificial Intelligence Lab for Problem-Based Learning” laboratory (under KFIST-L1, Grant number 958), which enabled successful project implementation and real-time data access, crucial for achieving study objectives.

## REFERENCES


[1] Q. Y. Tarawneh and S. Chowdhury, "Trends of climate change in Saudi Arabia: Implications on water resources," *Climate*, vol. 6, no. 1, 2018, doi: 10.3390/cli6010008.

[2] I. Gultepe *et al.*, "A review of high impact weather for aviation meteorology," *Pure and Applied Geophysics*, vol. 176, no. 5, pp. 1869–1921, 2019, doi: 10.1007/s00024-019-02168-6.

[3] B. Praveen and P. Sharma, "A review of literature on climate change and its impacts on agriculture productivity," *Journal of Public Affairs*, vol. 19, no. 4, 2019, doi: 10.1002/pa.1960.

[4] A. Erdil and E. Arcaklioglu, "The prediction of meteorological variables using artificial neural network," *Neural Computing and Applications*, vol. 22, no. 7–8, pp. 1677–1683, 2013, doi: 10.1007/s00521-012-1210-0.

[5] R. Guntukula, "Assessing the impact of climate change on Indian agriculture: Evidence from major crop yields," *Journal of Public Affairs*, vol. 20, no. 1, 2020, doi: 10.1002/pa.2040.

[6] K. K. Kumar, K. R. Kumar, R. G. Ashrit, N. R. Deshpande, and J. W. Hansen, "Climate impacts on Indian agriculture," *International Journal of Climatology*, vol. 24, no. 11, pp. 1375–1393, 2004, doi: 10.1002/joc.1081.

[7] A. Azam and M. Shafique, "Agriculture in Pakistan and its impact on economy—a review," *International Journal of Advanced Science and Technology*, vol. 103, pp. 47–60, 2017, doi: 10.14257/ijast.2017.103.05.

[8] L. J. Speight, M. D. Cranston, C. J. White, and L. Kelly, "Operational and emerging capabilities for surface water flood forecasting," *Wiley Interdisciplinary Reviews: Water*, vol. 8, no. 3, 2021, doi: 10.1002/wat2.1517.

[9] Y. M. Choo, D. J. Jo, G. S. Yun, and E. H. Lee, "A study on the improvement of flood forecasting techniques in urban areas by considering rainfall intensity and duration," *Water*, vol. 11, no. 9, 2019, doi: 10.3390/w11091883.

[10] T. Zhang *et al.*, "Development and evaluation of a WRF-based mesoscale numerical weather prediction system in Northwestern China," *Atmosphere*, vol. 10, no. 6, 2019, doi: 10.3390/atmos10060344.

[11] P. Sun *et al.*, "Nonstationarities and at-site probabilistic forecasts of seasonal precipitation in the East River Basin, China," *International Journal of Disaster Risk Science*, vol. 9, no. 1, pp. 100–115, 2018, doi: 10.1007/s13753-018-0165-x.

[12] R. Das, J. Mishra, S. Mishra, and P. K. Pattnaik, "Design of mathematical model for the prediction of rainfall," *Journal of Interdisciplinary Mathematics*, vol. 25, no. 3, pp. 587–613, 2022, doi: 10.1080/09720502.2021.2016853.

[13] E. A. Hussein, M. Ghaziasgar, C. Thron, M. Vaccari, and Y. Jafta, "Rainfall prediction using machine learning models: literature survey," *Studies in Computational Intelligence*, vol. 1006, pp. 75–108, 2022, doi: 10.1007/978-3-030-92245-0_4.

[14] X. He, H. Guan, X. Zhang, and C. T. Simmons, "A wavelet-based multiple linear regression model for forecasting monthly rainfall," *International Journal of Climatology*, vol. 34, no. 6, pp. 1898–1912, 2014, doi: 10.1002/joc.3809.

[15] S. Sahriman, A. Djuraidah, and A. H. Wigena, "Application of principal component regression with dummy variable in statistical downscaling to forecast rainfall," *Open Journal of Statistics*, vol. 4, no. 9, pp. 678–686, 2014, doi: 10.4236/ojs.2014.49063.

[16] G. Meti, R. K. G. Krishnegowda, and G. S. Swamy, "Rainfall analysis and prediction using ensemble learning for Karnataka State," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, pp. 1187–1198, 2023, doi: 10.11591/ijeecs.v32.i2.pp1187-1198.

[17] J. Lee, C. G. Kim, J. E. Lee, N. W. Kim, and H. Kim, "Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea," *Water*, vol. 10, no. 10, 2018, doi: 10.3390/w10101448.

[18] C. D. Usman, A. P. Widodo, K. Adi, and R. Gernowo, "Rainfall prediction model in Semarang City using machine learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 2, pp. 1224–1231, 2023, doi: 10.11591/ijeecs.v30.i2.pp1224-1231.

[19] S. N. Hosamane and G. P. Desai, "Air pollution modelling from meteorological parameters using artificial neural network," *Computational Vision and Bio Inspired Computing*, vol. 28, pp. 466–475, 2018, doi: 10.1007/978-3-319-71767-8_39.

[20] "Atech data logger," *Data Loggers India*. [Online]. Available: https://dataloggersindia.org/logger/239

[21] B. M. S. Hasan and A. M. Abdulazeez, "A review of principal component analysis algorithm for dimensionality reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20–30, 2021, doi: 10.30880/jscdm.2021.02.01.003.

[22] K. Warne, G. Prasad, S. Rezvani, and L. Maguire, "Statistical and computational intelligence techniques for inferential model development: A comparative evaluation and a novel proposition for fusion," *Engineering Applications of Artificial Intelligence*, vol. 17, no. 8, pp. 871–885, 2004, doi: 10.1016/j.engappai.2004.08.020.

[23] M. Sarstedt and E. Mooi, "Regression analysis," in *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*, Berlin, Heidelberg: Springer, 2019, pp. 209–256, doi: 10.1007/978-3-662-56707-4.

[24] P. Schober and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia and Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018, doi: 10.1213/ANE.0000000000002864.

[25] I. R. Illaboya and O. E. Igbinedion, "Performance of multiple linear regression (MLR) and artificial neural network (ANN) for the prediction of monthly maximum rainfall in Benin City, Nigeria," *International Journal of Engineering Science and Application*, vol. 3, no. 1, pp. 21–37, 2019.

[26] K. Yetilmezsoy and A. Saral, "Stochastic modeling approaches based on neural network and linear-nonlinear regression techniques for the determination of single droplet collection efficiency of countercurrent spray towers," *Environmental Modeling and Assessment*, vol. 12, no. 1, pp. 13–26, 2007, doi: 10.1007/s10666-006-9048-4.

[27] S. N. Hosamane, "Prediction of PM10 pollution using principal component regression and hybrid artificial neural network model," *Songklanakarin Journal of Science and Technology*, vol. 44, no. 5, pp. 1256–1263, 2022.

[28] I. Hossain, H. M. Rasel, M. A. Imteaz, and F. Mekanik, "Long-term seasonal rainfall forecasting: efficiency of linear modelling technique," *Environmental Earth Sciences*, vol. 77, no. 7, 2018, doi: 10.1007/s12665-018-7444-0.

[29] I. Hossain, R. Esha, and M. A. Imteaz, "An attempt to use non-linear regression modelling technique in long-term seasonal rainfall forecasting for australian capital territory," *Geosciences*, vol. 8, no. 8, 2018, doi: 10.3390/geosciences8080282.

[30] M. Hammad, M. Shoaib, H. Salahudin, M. A. I. Baig, M. M. Khan, and M. K. Ullah, "Rainfall forecasting in upper Indus basin using various artificial intelligence techniques," *Stochastic Environmental Research and Risk Assessment*, vol. 35, no. 11, pp. 2213–2235, 2021, doi: 10.1007/s00477-021-02013-0.

**BIOGRAPHIES OF AUTHORS**


**Gujanatti Rudrappa** holds Masters of Technology Degree from Visvesvaraya Technological University in 2013. He received his B.E. degree from Visvesvaraya Technological University in 2008. He is currently an assistant professor in KLE Technological University, Belagavi, Karnataka, India since 2013. He is an accomplished academician and researcher with expertise in machine learning, image processing, and artificial intelligence. With over two years of experience in the field of software industry as a Programmer Analyst and ten years of experience in the field of academics as an Assistant Professor, he has consistently demonstrated a keen understanding of machine learning, image processing, artificial intelligence, and their related applications. His research and contributions have been published in prominent journals, and his work has significantly influenced modern approaches to machine learning, image processing, artificial intelligence, and their related applications. He can be contacted at email: rudraguj@gmail.com or rudrappa.gujanatti@klescet.ac.in.

**Dr. Nataraj Vijapur** holds a Ph.D. from Visvesvaraya Technological University and has since held positions at leading research institutions. He is currently Professor and Head for Electronics and Communication Department at RV Institute of Technology and Management, Bangalore, Karnataka, India. He is dedicated to advancing our understanding of precipitation forecasting and fostering sustainable practices for a better future. He is an accomplished academician and researcher with expertise in medical image processing, artificial intelligence, machine learning, VLSI, and embedded systems. With over twenty years of academic experience as an Professor, he has consistently demonstrated a keen understanding of medical image processing, artificial intelligence, machine learning, VLSI, and embedded systems. His research and contributions have been published in prominent journals, and his work has significantly influenced modern approaches to medical image processing, machine learning, artificial intelligence, and their related applications. He can be contacted at email: nvijapur2academic@gmail.com and natarajv.rvitm@rvei.edu.in.

**Dr. Sateesh Hosamane** holds a Ph.D. from Visvesvaraya Technological University and has since held positions at leading research institutions. He is currently Professor and Head for Chemical Department at KLE Technological University, Belagavi, Karantaka, India. He is dedicated to advancing our understanding of ari pollution forecasting and its effects, precipitation forecasting and fostering sustainable practices for a better future. He is an accomplished academician and researcher with expertise in artificial intelligence, machine learning, and chemical reaction analysis. His research and contributions have been published in prominent journals, and his work has significantly influenced modern approaches to medical image processing, machine learning, artificial intelligence, and their related applications. He can be contacted at email: satishosamane@gmail.com.