Rainfall Analysis and Forecasting Using Deep Learning Technique

Pragati Kanchan¹, Nikhil Kumar Shardoor²

¹²Department of Computer Science and Engineering, School of Engineering, MIT ADT University, Pune, India
¹pragatikanchan04@gmail.com

Abstract
Rainfall forecasting is very challenging due to its uncertain nature and dynamic climate change. It’s always been a challenging task for meteorologists. In various papers for rainfall prediction, different Data Mining and Machine Learning (ML) techniques have been used. These techniques show better predictive accuracy. A deep learning approach has been used in this study to analyze the rainfall data of the Karnataka Subdivision. Three deep learning methods have been used for prediction such as Artificial Neural Network (ANN) - Feed Forward Neural Network, Simple Recurrent Neural Network (RNN), and the Long Short-Term Memory (LSTM) optimized RNN Technique. In this paper, a comparative study of these three techniques for monthly rainfall prediction has been given and the prediction performance of these three techniques has been evaluated using the Mean Absolute Percentage Error (MAPE%) and a Root Mean Squared Error (RMSE%). The results show that the LSTM Model shows better performance as compared to ANN and RNN for Prediction. The LSTM model shows better performance with minimum Mean Absolute Percentage Error (MAPE%) and Root Mean Squared Error (RMSE%).

Keywords
Rainfall Prediction, Deep Learning, ANN, RNN, Long Short-Term Memory

1. Introduction
Rainfall Prediction will always help to make decisions on agriculture, fisheries, forestry, tourism, etc. Monsoon plays a significant role in agriculture production. For countries like India, where agricultural production has been one of the main factors affecting the economy of India, a decent amount of rainfall gives the entire country an economic outlook and boosts the economy. A decent amount of rain enhances crop productivity and also increases water resources. Where an excess amount of rainfall brings a flood, which destroys crops, causes structural damage, threatens human life. In India, floods occurred in 2019 due to excessive rainfall in July and August, which had affected 13 states, Karnataka and Maharashtra were the most
severely affected states [17]. The early prediction of rainfall is therefore essential. Rainfall prediction will help farmers to make decisions on crop production and harvesting, as well as help prevent flooding, protect human lives and resources. Rainfall forecasting is very challenging due to its uncertain nature and dynamically changing climate. It is an application of science and technology to predict precipitation in advance. It’s always been a challenging task for meteorologists. Prediction of precipitation is categorized into short-range prediction and long-range prediction [5]. Forecasting is done through the collection and analysis of weather and climate data. Rainfall is computed based on various attributes, like Temperature, Humidity, Atmospheric Pressure, Evaporation, Sunshine, and Rainfall Amount(mm)–Hourly, Monthly, Annual, etc. Nowadays, artificial intelligence techniques are booming in the market, are being used for data analysis and prediction purposes in different sectors. In various papers for rainfall prediction different Data Mining and Machine Learning (ML) techniques have been used. These techniques show better predictive accuracy. A deep learning approach is used in this study to analyze the rainfall data of the Karnataka Subdivision. Deep learning is capable of handling a vast amount of data and is capable of handling complex problems. In this study, a Long Short-Term Memory (LSTM) technique has been used for monthly rainfall prediction of the Karnataka subdivision [20-26]. LSTM is evolved version of RNN. The results of the ANN-FFNN (Feed-Forward Neural Network) and RNN model were compared with the performance of LSTM.

2. Literature Survey

Dash, Y. et al. [1] has used three artificial intelligence approaches like K-Nearest Neighbor (KNN), Extreme Learning Machine (ELM), and Artificial Neural Network (ANN), for seasonal forecasting of the monsoon. These three techniques were used for predicting rainfall for the Kerala subdivision. The author found that the Extreme Learning Machine shows better performances as compared to KNN and ANN. ELM structure (8-25-1) gives better predictive accuracy with minimal Mean Absolute Percentage Error for both summer monsoon (June-September) and post-monsoon (October-December).

Kashiwao, T., et al. [2] proposed a model to predict local rainfall in the region of Japan. Data were collected from the Japan Meteorological Agency(JMA). The proposed model automatically collected meteorological data of temperature, atmospheric pressure, vapor pressure, amount of precipitation, wind velocity, and humidity. Two methods, such as Multi-layer Perceptron (MLP) and Radial Basis Function Network(RBFN), were used for rainfall prediction. The result of this study showed that the MLP model was superior to that of the RBFN model for rainfall prediction.

Nurcahyo, S. et al. [3] conducted research on the weather forecast of rainfall over Kemayaran Jakarta. The system was built using a combined hybrid Genetic Algorithm(GA) and Partially Connected Feedforward Neural Network(PCFNN) to predict rainfall for 7 days ahead in Kemayaran Jakarta. Rainfall was predicted with 81.52 % accuracy.

Dash, Y. et al. [4] in this study, artificial Intelligent (AI) Methods such as Extreme Learning Machine (ELM) and Single Layer Feed-Forward Network (SLFM) were used to predict Summer Monsoon in Kerala. Results of this study showed that ELM shows better results as compared to SLFM. The performance of these techniques was evaluated based on Mean Absolute Error and Root Mean Squared Error.

Dutta, P. S. et al. [5] proposed a model using data mining techniques for monthly rainfall prediction over Assam. Statistical technique - Multiple Linear Regression was used for prediction. The performance of the proposed model was measured in adjusted R-squared. 63% accuracy obtained using the given model.

Haidar, A. et al. [6] developed a monthly rainfall prediction model. A deep convolution neural network (CNN) was used for prediction. Performance of the proposed model compared with the first version of the Australian Community Climate and Earth-System Simulator (ACCESS-SI) and Multi-Layer Perceptron (MLP). The proposed model CNN gives better performance for rainfall prediction.

Thirumalai, C. et al. [7] presented machine learning techniques for heuristic prediction of rainfall. In this study rainfall da-
ta in previous years according to crop season like Rabi, Kharif, Zaid was considered for future prediction of rainfall. Linear Regression model was used for the early prediction of Rainfall.

A. kala et al. [8] in this study model built using Artificial Neural Network (ANN) such as Feed Forward Neural Network (FFNN) for predicting rainfall. Four parameters like Temperature, Cloud Cover, Vapor Pressure, and Precipitation were taken for predicting the rainfall. Root Mean Squared Error (RMSE) and Confusion matrix were used to measured prediction accuracy. The proposed model based on ANO indicates acceptable accuracy.

Qiu, M. et al. [9] in this paper, proposed a Multi-Task Convolution Neural Network(MT-CNN) model for rainfall prediction. The proposed approach automatically extracts features from the time series measured at observation sites. Based on multisite features, predicted short term rainfall amount.

Rasel, R. et al. [10] presented the performance of machine learning and data mining techniques such as Support Vector Machine (SVR) and Artificial Neural Network (ANN) for weather forecasting. The results of this study showed that ANN produces a better result.

Chatterjee, S. et al. [11] proposed a model for rainfall prediction using Hybrid Neural Network (HNN) over west Bengal. Data were collected from Dumdum Meteorological Station. K-mean Clustering and Neural Networks were used to trained model. Performance of HNN in terms of F-measure, accuracy, precision, and recall compared with Multilayer Perceptron-Feedforward Neural Network (MLP-FFN). The proposed model predicted rainfall with 89.54% accuracy.

Sulaiman, J. et al. [12] in this study for precipitation prediction an Artificial Neural Network model was used. The Rainfall data was collected from the local meteorological department. The 80% data used for training and 20% of data were used for testing. Precipitation was predicted using Time Delay Neural Network and Auto-Regressive Integrated Moving Average model. The result of this study showed that TDNN outperformed the ARIMA model.

Kumar, R. et al. [13] in this research, the author presented different Data Mining Techniques for rainfall prediction. Performance and comparison of various data mining techniques like Decision Tree, Naive Bayes, K-Nearest Neighbour, Neural Network, and Fuzzy Logic were given.

Parmar, A. et al. [14] this paper reviewed different approaches and algorithm such as Artificial Neural Network (ANN) - Back-Propagation Neural Network, Cascade Forward Back Propagation Network, Support Vector Machine (SVR), Layer Recurrent Network, and Self Organizing Map (SOM) for rainfall prediction.

Poornima, S. et al. [15] proposed a model for rainfall prediction using Intensified LSTM based RNN. The rainfall dataset of the Hyderabad region was used for prediction. Minimum and Maximum Temperature, Wind Speed, Sunshine, Minimum and Maximum Relative Humidity, Evapotranspiration parameters were used for predicting rainfall. The performance of Intensified LSTM model compared with RNN, LSTM, ELM, Holt-Winters, ARIMA methods. The result of this study shows that Intensified LSTM gives better results as compared to other methods used.

Basha, C. Z. et al. [16] in this study, deep learning approach has represented for rainfall prediction. Deep learning techniques such as MLP and Auto-Encoder NN were used for predicting rainfall. In this study, the CNN technique was used for taking input from past data. Performance of these techniques evaluated using MSE and RMSE.

---

**Table 1. Different AI and ML Techniques For Rainfall Prediction**
<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>DATA SET TIME PERIOD AND REGION</th>
<th>TECHNIQUES</th>
<th>ATTRIBUTES</th>
<th>ACCURACY MEASURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rasel, R. et al. (2017) [10]</td>
<td>6-years (Chittagong Bangladesh)</td>
<td>SVR, ANN</td>
<td>Rainfall, Temperature</td>
<td>RMSE, MAE</td>
</tr>
</tbody>
</table>

3. Methodology
A rainfall prediction model illustrated in the Figure 1. A rainfall dataset has collected in the first step.

![Rainfall Prediction Model](image)

**Figure 1.** Rainfall Prediction Model.

### 3.1. Dataset

The Rainfall dataset has downloaded from the data.gov.in website. Dataset has subdivision wise rainfall data. Rainfall data of the Karnataka Subdivision has been used in this study. The dataset period is from 1901 to 2017. Rainfall data from the Jan to Dec has used for prediction purposes.

### 3.2. Data Preprocessing

Initially, the data set contains some missing values. Thus preprocessing data is very important for the accuracy of the model. The missing values are dropped and replaced with the mean value. Data scaling and normalization transform the data into a standardized form. Normalization helps to scale the data of an attribute so that it falls in a smaller range between 0 to 1 or -1 to 1. The min-max normalization method has been used for this study. Input data were normalized using the formula stated below in Equation 1.

$$Y_{\text{normalized}} = \frac{y - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}}$$  \hspace{1cm} (1)

Where $Y$ represents normalized data, $y$ is the actual value of rainfall data to be normalized, $y_{\text{min}}$ represents minimum value of rainfall data, $y_{\text{max}}$ represents max value of rainfall data respectively.

### 3.3. Training and Testing Data

The Rainfall dataset split into a training and test dataset. From 1940-2010 data has given to the training phase, and for testing, from 2011 to 2017 rainfall data used. The prediction model first trained using ANN and Simple RNN, then the model was trained with an Long Short Term Memory. The performance of the built model checked with the test dataset.

### 4. Techniques
4.1. Artificial Neural Network

An ANN is a type of neural network that is inspired by the biological neural network. It is a computational model that processes information like the way the human brain process information. ANN comprises multiple interconnected neurons that mostly operate in parallel. There are three types of neural network architecture. The first type is Single-layer feed-forward neural network architecture which contains only the Input and Output layers. The second type is a Multi-layer feed-forward neural network that comprises at least three layers of nodes: The input node, the Hidden node, and the output node. The third type of neural network architecture is Recurrent neural network architecture. Multilayer feed-forward neural network is also called Multilayer Perceptron (MLP). Multilayer Perceptron was used in this study to predict monthly precipitation. The neural network is mostly train using Back-Propagation.

Algorithm
Training dataset {\( \{x_1, t_1\}, \{x_2, t_2\}, \ldots, \{x_n, t_n\} \)} given as input.

Return Output with trained ANN.

1. The first step is to initialize the weight in the network.
2. Repeat
   For several epochs:
   - Process training data(Xn)
   - Calculate output(O)
   - Compare target output(T) with predicted output(O)
   - Calculate error(target output - calculated output) at the output layer.
   - BackPropagate error and update the weights in the network.
end
3. Until desired output obtained
4. return (Trained ANN)

4.2. Simple Recurrent Neural Network

In feed-forward neural networks, there is no feature of remembering the previous output as outputs in FFNN are independent of each other. Where RNN has the ability to remember previous information, and because of this RNN works better with time series prediction problems. A recurrent neural network is a deep learning technique, in which input for the current state (C_t) is the new input and output of the previous time step (C_{t-1}). For the next state (C_{t+1}), there are two inputs one is the new input and the other is the output of the previous time step (C_t). The RNN learns using backpropagation through time. The formula used to calculate the current state is stated in Equation 2.

\[ h_t = g(h_{t-1}, x_t) \]  

(2)

where current state is denoted by \( h_t \), \( h_{t-1} \) represent output from the previous state, \( x_t \) is a new input at time step \( t \) and \( g \) is a recursive function. For applying the activation function the formula is given below in Equation 3.

\[ h_t = \tanh(w_h h_{t-1} + w_x x_t) \]  

(3)

where \( w_h \) and \( w_x \) represent weight at recurrent neuron and weight at input neuron. The formula used to calculate output is stated in Equation 4.

\[ y_t = w_y h_t \]  

(4)

where output is denoted \( y_t \) and \( w_y \) represent weight at output neuron.
4.3. Long Short-Term Memory (LSTM)

With RNN, there is a problem of exploding and gradient vanishing. This problem can be solved with the help of an evolved version of Simple RNN, such as Gated Recurrent Units and Long Short-Term Memory (LSTM). In this study, the LSTM algorithm has used for rainfall prediction. This algorithm is best suited for prediction based on time series data. In RNN, after calculating loss weights are updated by multiplying gradient. If the gradient is too smaller then the updated weight would be negligible so the neural network wouldn’t learn at all. This problem is called a gradient vanishing problem. Long Short-Term Memory (LSTM) solves the problem of gradient and also gives better accuracy than RNN. The architecture of LSTM comprises of three gates: an input gate, an output gate, a forget gate, and one cell state. The architecture of LSTM is shown in Figure 2.

![Figure 2. Structure of LSTM Unit.](image)

Forget Gate determines which information is no longer required and will be thrown away from the block. Forget gate ($f_t$) takes the output of the old state and the input and multiply it with respective weights, and output is passed it through the sigmoid activation to cell state.

$$f_t = \sigma(w_f x_t + u_f h_{t-1} + b_f) \quad (5)$$

Input Gate determines which input values are to be written to the memory state. The input gate ($i_t$) takes input from the previous timestamps and the new input and passes it through sigmoid activation. The value of $i_t$ then multiply with $c'_t$ and result of this is then add to the cell state.

$$i_t = \sigma(w_i x_t + u_i h_{t-1} + b_i) \quad (6)$$

$$c'_t = \tanh(w_c x_t + u_c h_{t-1} + b_c) \quad (7)$$

In the next step the value of new cell state $c_t$ is obtained by first multiplying $c_{t-1}$ i.e. old cell state by forget gate ($f_t$), and then by adding $i_t * c'_t$.

$$c_t = f_t * c_{t-1} + i_t * c'_t \quad (8)$$

Output Gate($o_t$) determines which output is to be generated based on the current internal cell state.

$$o_t = \sigma(w_o x_t + u_o h_{t-1} + b_o) \quad (9)$$

After calculating the values of $o_t$ result of the output gate is multiplied with the cell state and passes through tanh activation.

$$h_t = o_t * \tanh(c_t) \quad (10)$$

5. Performance Measures

The Mean Absolute Percentage Error (MAPE%) and Root Mean Square Error (RMSE%) have used for evaluating the performance of the built model. The accuracy metrics used to measure the results of the techniques are presented in Equation 11

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$
and Equation 12.

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - P_i}{A_i} \right| \times 100
\]  \hspace{1cm} (11)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (A_i - P_i)^2}
\]  \hspace{1cm} (12)

Here, A, represents actual rainfall value, and P, represents the predicted rainfall value for the year i. n is the number of years to be predicted.

6. Result and Discussion

In this study, the performance of FFNN, Simple RNN, and LSTM for monthly rainfall prediction over Karnataka subdivision of India has evaluated. The performance of these three techniques in terms of MAPE, RMSE has given in this section. The data collected includes monthly rainfall measurements for 116 years. Keras neural network library has used for implementation. Keras is python based deep learning framework. It is a high-level API of TensorFlow, and it is run on top of TensorFlow. Adam optimizer has been used to train the deep neural network. Adam optimization is an extension of the stochastic gradient descent method to update network weights based on training data. Different combinations of input and hidden layers has examined for predicting rainfall. The performance of three techniques has evaluated with the help of accuracy matrices. As shown in Table 2. LSTM model (12-5-1) shows better performance with minimum Mean Absolute Percentage Error 79.0% and Root Mean Squared Error 135.4% as compared to MLP and Simple RNN for prediction.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Input, Hidden and output Node</th>
<th>MAPE%</th>
<th>RMSE%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN_MLP</td>
<td>12-5-1</td>
<td>109.6</td>
<td>162.5</td>
</tr>
<tr>
<td>RNN</td>
<td>12-5-1</td>
<td>99.5</td>
<td>153.8</td>
</tr>
<tr>
<td>LSTM</td>
<td>12-5-1</td>
<td>79</td>
<td>135.4</td>
</tr>
</tbody>
</table>

Prediction results for monthly rainfall(mm) using LSTM on train data from 1940 to 2010 and on test data from 2011 to 2017 is depicted in Figure 3 and Figure 4. Actual rainfall values has compared with the output of LSTM.

![Figure 3. Rainfall Prediction Using LSTM on Training Data.](image-url)
The average monthly rainfall for all years (From 1940-2017) of the Karnataka Subdivision is shown in Figure 5. In which Jun, July, August month have the highest rainfall, September, October, May month has moderate rainfall, and January, Feb, Mar, Apr, Nov, Dec month has the lowest rainfall.

As shown in figure 6. FNN gives 109.6% mean absolute percentage error, RNN gives 99.5% error and LSTM prediction model gives 79.0% error.
7. Conclusion

In this study deep learning techniques are used for predicting monthly rainfall over Karnataka subdivision. The results show that the LSTM optimized deep learning technique shows better predictive outcomes. The results of the ANN and RNN model were compared with the performance of LSTM. The performance of these three techniques has evaluated with the help of accuracy matrices. The LSTM model shows better performance with minimum Mean Absolute Percentage Error (0.79), and Root Mean Squared Error (1.35) for prediction.

References


