**Leveraging Artificial Neural Networks for Rainfall Prediction in Medak District, Central Telangana, India**

**Abstract**

This study explores the application of artificial neural networks (ANNs) for monthly rainfall prediction at the Medak station in Central Telangana, India. A total of 113 years of rainfall data was utilized, with 85 years (January 1901 to December 1985) used for model training and 28 years (January 1986 to December 2014) for testing. Input variable selection was carried out using the Gamma test, autocorrelation function, and cross-correlation function. The models were developed using a multilayer perceptron (MLP) trained with two learning algorithms—Levenberg-Marquardt and Delta-bar-delta—and employed two transfer functions: Sigmoid and Tanh. Model performance was evaluated both visually and through quantitative indices, including Root Mean Square Error (RMSE), Correlation Coefficient (R), Coefficient of Efficiency (CE), Percent Bias (PBIAS), and Integral Square Error (ISE). The results demonstrated that models using rainfall data from adjoining stations as inputs outperformed those using lagged data from the same station. Among all models, the M-8 model showed superior performance with higher R and CE values, and lower RMSE, PBIAS, and ISE values. These results indicate that the M-8 model is a promising tool for reliable monthly rainfall prediction in the Central Telangana region.

**Keywords:** ANN Model, Artificial Neural Network, Gamma test, Rainfall prediction.

**Introduction**

India is dealing with a serious water issue that affects both an essential needs and the demands of agriculture. In the upcoming decades, it is anticipated that the situation would deteriorate as water demand rises across all industries. Localized water demands get worse by increased urbanization and the development of sanitary infrastructure. Since the agricultural sector uses 70% of the nation's water **(NITI Aayog (2018)** Composite Water Management Index. Discusses the growing stress on India's water resources due to urbanization, industrialization, and population growth., it is likely to continue to dominate water usage. Nonetheless, the industrial sector, energy generation, and municipal supply system are all seeing significant increases in water demand. “A significant contributing aspect to this problem is the expanding population, which is expected to reach 1.66 billion by 2050” **(NITI Aayog (2018). “**Annual food consumption needs projected to be surpassed 250 million tons by 2050, as well as the entire demand for grain, including livestock feed, will exceed 375 million tons (Planning Commission, Government of India.). However, per capita cereal consumption is projected to decline by 9%, 47%, and 60% for rice, coarse cereals, and maize, respectively. In contrast, significant increases are expected in the consumption of sugar (32%), fruits (65%), and vegetables (78%) between 2000 and 2050. This dietary shift will escalate water demand, with livestock water use rising from 2.3 BCM in 2000 to 2.8 BCM by 2025 and 3.2 BCM by 2050” (The Future of Food and Farming: Challenges and Choices for Global Sustainability 2011.). Hence, a comprehensive approach is needed for addressing these issues in order to maximize water supplies and meet the requirements of a growing economy under diverse Indian rainfall pattern.

Rainfall pattern in India is influenced by the southwest and northeast monsoons, cyclonic depressions, disturbances, and localized violent storms. These storms typically form in areas where cool, humid sea winds meet hot, dry land winds, sometimes escalating to cyclonic dimensions. “The rainfall occurs in India mostly during the South-West monsoon season from June to September, except in Tamil Nadu” (Goswami, B. N., et al)., where the North-East monsoon during October and November is the primary source. Accurate and timely prediction of rainfall is crucial for effective water resource management. This has made rainfall forecasting a critical area of research, with methods largely relying on statistical analysis of historical data. Rainfall, as a dynamic weather parameter, exhibits nonlinear and multi-layered patterns that vary across time and space. These variations, combined with climate fluctuations and irregular spatial and temporal rainfall distribution, often lead to challenges such as floods and droughts (Danladi et al., 2018).

Rainfall prediction models are essential for water resource management, agricultural planning, and disaster prediction. Model assist to optimize crop scheduling, mitigate flood risks, and ensure efficient water allocation in different sectors. Rainfall prediction models can be empirical, black-box, or physically based, and each has many benefits and drawbacks of its own. While conceptual and physically based models help in understanding hydrological variables and physical processes, they often require extensive data and are computationally demanding. On the other hand, empirical models are advantageous in scenarios with limited data and prioritize accuracy over physical representation.

Considering the circumstances, artificial neural networks (ANNs) have emerged as one of the most efficient methods for rainfall forecasting, contributing significantly to optimizing agricultural production from limited land and water resources. Accurate rainfall prediction remains a formidable challenge in hydrology. While various mathematical models have been proposed, contemporary research highlights ANNs as a preferred approach. Over the past decade, ANNs have gained popularity as robust tools in hydrology and water resource management, often serving as alternatives to traditional modelling techniques (Hsu et al., 1995; Hu et al., 2005; Keskin et al., 2006).

Soft computing techniques, an emerging field, have provided innovative tools for rainfall prediction. These include fuzzy logic, neural computing, evolutionary computation, machine learning, and probabilistic reasoning. Among these, artificial neural networks (ANNs) have gained prominence due to their ability to model dynamic, nonlinear systems effectively **(Sudheer, K. P., & Jain, S. K. (2004.)** Moreover, ANNs, considered black-box models, are particularly suited for rainfall forecasting due to their capacity to handle complex, nonlinear relationships without requiring a detailed understanding of underlying physical processes **(Sudheer, K. P., & Jain, S. K. (2004).** Recent applications of ANN techniques have shown promising results, offering a robust and efficient tool for tackling the challenges of rainfall prediction (Sammen *et al.,*2023).

Despite their flexibility and adaptability, it is not conclusively proven that neural network models consistently outperform traditional models. A notable limitation of ANN models is their "black-box" nature, which can lead to issues like overfitting and equifinality (Beven and Binley, 1992). However, ANNs offer distinct advantages, including the ability to model complex, nonlinear relationships between inputs and outputs, much like the neural processes of the human brain. These models have been extensively utilized in hydrological studies, particularly in rainfall-runoff modelling, and are powerful tools for addressing complex river flow forecasting challenges, especially when rapid prediction is required. Mathematically, ANNs are universal approximators capable of learning from data without requiring explicit physical parameters (ASCE, 2000 a, b).

The present study explores the application of ANN in predicting rainfall, contributing to sustainable resource planning. Background Rainfall prediction is a critical aspect of agricultural planning, water resource management, and disaster preparedness, particularly in regions like Medak District in Central Telangana, where rainfall significantly influences livelihoods and ecosystems.

In Telangana, where approximately 55.49% of the population depends on agriculture and allied activities for their livelihood, rainfall is a crucial factor . Telangana state, India's 29th state, was separated from the northwest region of Andhra Pradesh on June 2nd, 2014. The state receives an average annual rainfall of 906 mm, with 80% occurring during the South-West monsoon (Agricultural statistics at a glance Telangana 2013-14). Agriculture in Telangana is highly vulnerable to climate change, particularly rainfall variability, which poses significant challenges. Variations in monsoon patterns directly impact the yield and profitability of rainfed crops. Prolonged dry spells or heatwave conditions during critical crop stages exacerbate moisture and thermal stress, reducing productivity.

Additionally, Telangana is prone to cyclones, heavy rainfall, floods, and droughts due to its geographical peculiarities. Droughts, in particular, are a recurring challenge. In light of these factors, a focused effort has been undertaken to estimate monthly rainfall in Central Telangana, utilizing advanced methodologies to address the inherent variability and challenges associated with climate and rainfall patterns. Accurate forecasting models, such as Artificial Neural Networks (ANNs), have emerged as powerful tools for analysing complex weather patterns and predicting rainfall with improved precision.

Some studies have been performed where researchers used the ANN model to predict the versatile factors pertaining to watershed management, surface runoff, etc., while ANN models were performed along with different algorithms to find the adequacy of the model. Anmala et al. (2000) used “ANN models for estimating runoff over three different medium sized watersheds which were found in Kansas. mitigating climate-related They also explained feed-forward neural networks without time delayed input did not provide significant improvement over other regression approaches”. “However, inclusion of feed forward with RNN resulted in better performance. ASCE (2000) examined the role of ANN in various branches of hydrology and found that ANNs were robust tool for modelling many of nonlinear hydrologic processes such as rainfall-runoff, stream flow, ground water management, water quality simulation and precipitation” (Spandana K. et al., 2021). Sahai et al. (2000) described “the artificial neural network (ANN) technique with error-back propagation algorithm to provide prediction of Indian Summer Monsoon Rainfall on monthly and seasonal time scales. It is observed by various researchers that with the passage of time the relationships between various predictors and Indian monsoon were changing, leading to changes in monsoon predictability”. Toth et al. (2000) studied and compared “the accuracy of the short-term rainfall forecasts obtained with time-series analysis techniques using past rainfall depths as the only input information”. “The results also indicated how the considered time-series analysis techniques, and especially those based on the use of ANN, provide a significant improvement in the flood forecasting accuracy in comparison to the use of simple rainfall prediction approaches of heuristic type, which were often applied in hydrological practice” (Spandana K. et al., 2021). Luk et al. (2001) and (Spandana K. et al., 2021). “developed three different types of ANN viz. multilayer feed forward neural networks, partial recurrent neural networks and time delay neural networks and found to provide reasonable predictions of the rainfall depth one time-step in advance”. Rajurkar et al. (2002) studied “the application of artificial neural network methodology for modelling daily flows during monsoon flood events for a large size catchment of the Narmada River in Madhya Pradesh, India”. The model provides a systematic approach for runoff estimation and represents improvement in prediction accuracy over the other models. Wilby et al. (2003) considered “precipitation, evaporation and discharge data for developing conceptual and neural network rainfall-runoff model”.

This study emphasizes the implementation of ANNs, with a focus to model architecture, feature selection, and performance assessment. The objective of the study is to validate the developed ANN models for the Central Telangana region, and assess its performance in order to identify the best-suited model.

**Materials and Methods**

This study aimed to developed the effective artificial neural network models with two different activation functions for forecasting of monsoon rainfall, whereas the gamma test used for the selection of the inputs data. The methods used for model validation and calibration was presented as well as several standards for assessing the models' performance.

***Study Area Profile***

The Godavari and Krishna are two major rivers basin spread in Telangana state. Based on agro-climatic zone, Telangana state is divided into three zones namely northern Telangana zone, central Telangana zone and southern Telangana zone. The study area in this research was focused in central Telangana zone and consists of Medak. The rain gauge site's specifics and the stations' data length are provided in Table 1. In this study, neural network models for monsoon rainfall forecasting are developed, comprising data collecting, model building, and evaluation, with a focus on rain gauge site details and data length.

**Table 1** Detail of rain gauge site and data length

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **District** | **Latitude** | **Longitude,** | **Altitude, (m)** | **Data** |
| Medak | 18.03°N | 78.27° E | 442 | 1901-2014 |

***Climate***

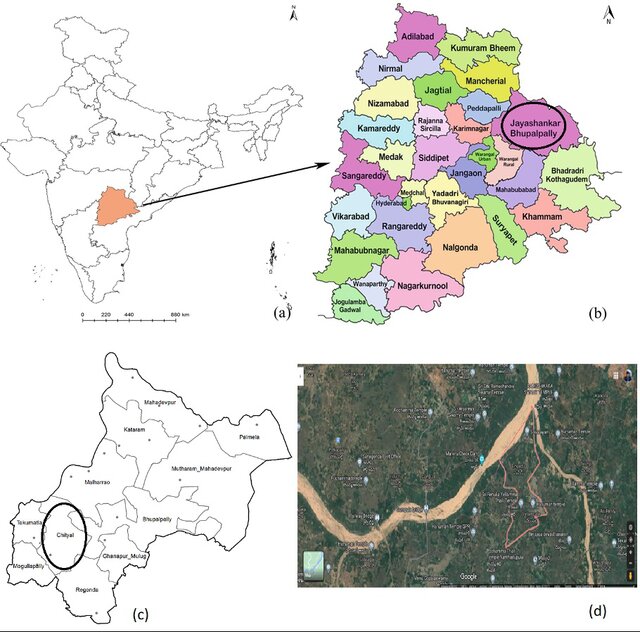
The area comes under semi-arid area and has a mainly dry and warm climate. The summer is beginning in March and has high temperature reach in the May month with normal temperatures around the 42ºC. The season for the study area as per meteorological study are divided into four distinct parts, i.e. monsoon (June to September), post monsoon (October to November), winter (December to February) and summer (March to May). The monsoon arrives in month of June and continues up to September with around 755 mm of average annual rainfall.

***Soil Characteristics***

The characteristics of soil provides adequate information about natural vegetation, infiltration of soil, types of soil, land forms, as well as nature of soils which can be used for land and development. The central Telangana zone have different types of soils such as red soils (48%), black soils (25%), laterite soils (7%) which is mostly present in Medak, districts of Telangana region. The water holding capacity varies from moderate to high. Soils have drainage behaviour ranging from well drained to poorly drained state under low land situation. There is a wide range of variation in nitrogen and phosphorus status of the soils. Most of the soils belong to land use capability classes I-III of which some pockets are problematic due to erosion and flood.

***Selected region of Telangana and data acquisition***

According to the government report (Government of India 2023). The central Telangana region is divided into areas having agriculture (43%), forest (24%), current fallow lands (8%), non-agricultural uses (7.80%), barren and uncultivable land (5%), other fallows (6%), and cultivable waste/permanent pastures (5%) land are covered. In central Telangana, 68% of the total gross cropped area occupied by Rice, Cotton, Maize, Soybean, Bengal gram, Maize, Green gram, Red gram, Black gram, Groundnut, Sunflower and Tobacco.



**Fig. 1** Selected region in central Telangana under agro-climatic zon

The Telangana region, with a specific focus on Medak district (Figure 1), was selected for this study based on historical rainfall data sourced from the Indian Water Portal for the period 1901 to 2002. Monthly rainfall data was obtained from the India Meteorological Department (IMD), covering the period from January 1901 to December 2014. Additional district-level rainfall records from IMD for the years 2003 to 2014 were also incorporated to ensure dataset continuity.

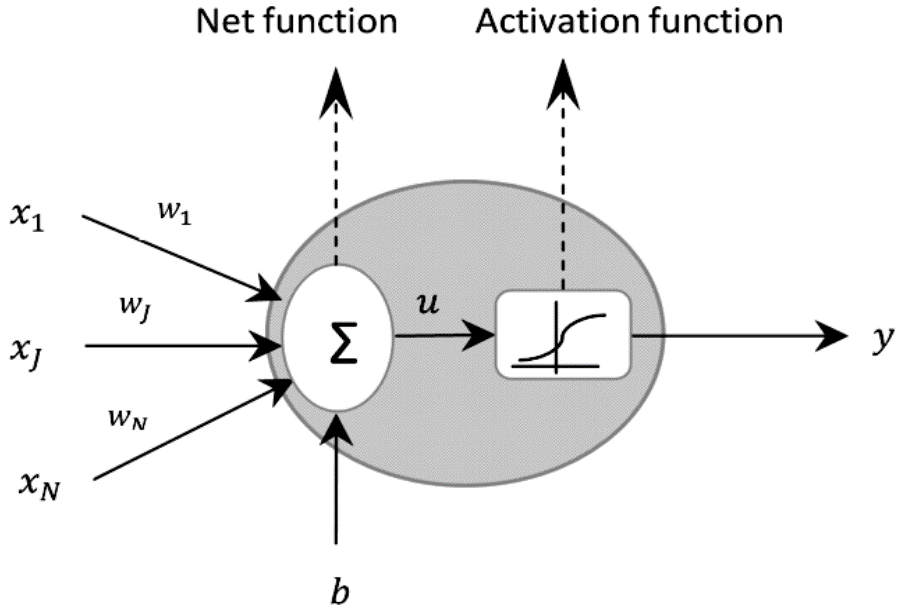
The complete dataset spans 113 years, from January 1901 to December 2014, and was divided into two phases for model development. The first phase, comprising 85 years of data (January 1901 to December 1985), was utilized for training the model. The second phase, encompassing the remaining 28 years (January 1986 to December 2014), was employed for validating the model's performance. This phased approach enables robust model training and assessment under diverse climatic conditions.

**Artificial Neural Networks**

Artificial neural networks (ANNs), are inspired by biological neural systems. They consist of interconnected units or nodes, referred to as artificial neurons. These connections allow these neurons to transfer signals with one another. When a neuron receives a signal, it processes the information and may pass the signal along to other connected neurons. Significant advancements occurred in the 1980s when principles, applications, and algorithms for artificial neural networks were refined. Rumelhart et al. (1986) introduced the multilayer perceptron and the backpropagation algorithm, which became a cornerstone for training neural networks and facilitated their application in a variety of fields.

**Concept of artificial neural networks (ANNs)**

Neurons, also known as nodes, are the essential building blocks of artificial neural networks. Neuron, is able to acknowledge and pass on signals starting with one neuron then onto the next neuron. The most widely recognized structure of a neuron is shown in figure 2.

****

**Fig. 2** Basic structure of neural network (Michael Nielsen, *Neural Networks and Deep Learning* (Online, 2015)

In neural system each neuron has a various information sources A neuron calculates an output by applying net and activation function on inputs.

Where xi is an input vector, w is the association weight from the ith neuron in the input layer and b is the threshold value or the inclination of the neuron. The nodes can take input data information can perform basic activity on information on data.

An artificial neural network typically consists of three distinct layers of neurons: an input layer, one or more hidden layers, and an output layer. Usually, a single concealed layer with enough neurons, usually leads to an accurate approximation needed. Having a greater number of hidden neurons, gives the network flexibility to solve more complex problems, while having excessive numerous neurons. may cause over fitting problem. When too many neurons are used, overfitting may result. To address this, many strategies have been proposed, where with the trial-and-error method being one of the most effective.

In terms of the pattern of connections between the layers, ANN can be designed in feed- forward or recurrent form. Recurrent neural networks operate well with temporal input, whereas feed-forward neural networks is more often referred to as "neural networks" perform best with static patterns in data, so that some users identify the phrase “neural networks”, only feed-forward network. There is different type of feed-forward neural networks such as multilayer perceptron (MLP) and the radial basis function (RBF). The most popular neural network paradigm in hydrology is the multilayer feed-forward neural networks (Fernando and Jayawardena, 1998; ASCE task committee, 2000a and Dawson and Wily, 2001), which is used in this study.

*Learning algorithm*

Fundamentally, neural networks depend on a learning or training algorithm that, given a set of training data, modifies the network's parameters to produce the intended model performance. In response to the input data supplied at the input layer and, based on the learning rule, at the output layer, the weights are modified. The learning process enables the network to adjust its responses over time to produce the desired output and produced model based on training data set. There are three main classifications for ANN learning, supervised, unsupervised and reinforcement. Supervised learning, the most widely used model, requires both input and corresponding output for training the network. It is well-suited for handling time series forecasting problems.

Under the supervised learning, back-propagation algorithm is more often adopted to train the neural networks. Back-propagation algorithm (BP) is administered algorithm which adjusts the association weights and biases in the backward direction. It is an optimization procedure based on gradient descent to minimize the total error between the desired and actual outputs. The information data are multiplied by the initial weights, then the weighted information are added by simple summation to yield the net input to each neuron.

…3

…4

Where, Xi is the input to any neuron, wji is the connection weighted between jth layer to ith layer, N is the number of inputs and Net is the net for jth neuron. The output of kth node of the hidden layer bk is given as:

…5

Where is the activation function, example a tanh activation function. This can be represented as:

…6

The error calculated at the output layer is propagated back to the hidden layers and then to the input layer, in order to determine the updates for the weights. The mean sum of square error E for a single input-output pair data set is given as:

…7

Where, E is the Total error, ci is the observed or calculated output at ith node and di is the target or desired output at ith node.

A set of pattern examples is used under training process. Each example consisting of a pair with the input and corresponding target output. The patterns are introduced to the network sequentially in an iterative manner, the appropriate weight corrections being performed during the process to adapt the network to the desired behaviour. This repeating continues until the connection weight values allow the network to perform the required mapping. Each presentation of the whole pattern set is named an epoch. After this the term repetition will refer either to a pattern presentation or to a complete epoch depending on the situation. The generalized delta rule is used to calculate the values of the local gradients. Each weight update is defined as:

…8

The generalized delta rule used to calculate the values are as follow:

…9

…10

The weight update for the output units can be calculated using directly available values since the error measure is based on the difference between the desired tj and actual aj values. However, that measure is not available for the hidden neurons. The solution is to back-propagate the j values layer by layer through the network. The weight update for the output units can be calculated using directly available values since the error measure is based on the difference between the desired tj and actual aj values. However, that measure is not available for the hidden neurons. The solution is to back-propagate the j values layer by layer through the network. The learning algorithm also called gradient search is used to calculate the adjusted weights and biases of the network to minimize the error between computed and observed output. In searching with the momentum component there are two parameters to be selected, the step size and the momentum. The Neural Builder provides a default value for the learning rates. In this study, Delta-Bar-Delta learning and Levenberg–Marquardt algorithms are used.

1. *Delta Bar Delta*

The Delta-Bar-Delta (DBD) algorithm is a meta- processing algorithm in the sense that it learns the learning-rate parameters of an underlying base learning system. The Delta-Bar-Delta (DBD) network used the same architecture as a back-propagation network used to get the information from previous weights.

The Delta Bar Delta is an adaptive step-size system for searching a performance surface. Step size and momentum are modified according to the previous values of the error at the neurons. If the current and past weight update is both of the same symbols, it increases the learning rate linearly. The reasoning is that if the weight is individual moved in the same direction to decrease the error, and then it will get there faster with a larger step size. If the updates have different symbols, this is a signal that the weight has been moved too far. When this happens, the learning rate decrease geometrically to avoid difference. It was urbanized for quadratic error functions;

∆wji = α (t3-Y3) g (h3) xi  …11

Where, α = Constant learning rate, g (x) = neuron's activation function, t3 = targeted output, h3 = weighted sum of the neuron's inputs, y3 = actual output, Xi = ith input.

1. *Levenberg–Marquardt*

The Levenberg–Marquardt algorithm (LM) was most proper higher-order adaptive algorithms known for minimizing the mean square error (MSE) of a neural network. The Standard gradient descent algorithms use only the local estimate of the slope of the performance surface (error vs. weights) to establish the best direction to travel the weights for lowering the error. A key advantage of the LM algorithm is its ability to default to gradient search when the presentation surface deviates from a parabola, a common occurrence in neural computing. The challenge of training the MLP must be expressed as a nonlinear optimisation in order to use LM.

Wk+1 = Wk+1 Wk- ( JTkJTk +µl)-1 JTk e …12

Where, µ = Parameter changed during the training process. If µ = 0 Algorithm works as LM method and for large values of μ. algorithm works as steepest decent method.

*Connection weights*

The function of increasing decreasing, or altering the sign of the input signal determines the connection weights. An inhibitory association between two nodes is shown by a negative weight, whereas the absence of a connection is represented by a zero weight. In general, the input of from node xi is multiplied by the weight of the connection 𝑤𝑖 to produce the input signal to node. Hence connection weights represent the strength of the connection between two nodes. Weights are stored in the local memory of nodes and also hold the long-term memory of the network. Connection weights signify the strength of the connection between two nodes, and weights are stored in the local memory of the nodes, which serve as the network's long-term memory.

*Threshold*

Threshold is calculated by a set value based upon the final output of the network and used in the activation function. To determine the network output, the computed net input and the threshold are compared. For each and every application, there is a threshold limit. The activation function using threshold can be defined as:

…13

*Activation function*

A computational network that determines a processing element's output based on an input signal is called an activation function. It uses the processing element by using trial and error method. Different kinds of activation functions are frequently employed in hydrological models. Although, the most common activation function is Sigmoid and hyperbolic tangent.

**Gamma test**

GT from input variables can be researched for modeling output data and in addition building up a model smooth way. GT calculates the minimum square errors which are obtainable in continuous non-linear models and unseen data. Assume that Xa and Xb are nearest to each other; then, Ya and Yb ought to likewise be near every other. In GT, it is attempted to make this view qualitative through mean difference between the closest neighbor limited arrangement of Xa and Xb and mean length between the comparing output purposes of Ya and Yb and accomplish estimation for error value. Let assume that there is a data set as follows:

…14

Where, is the input vector in the output vector’s areas of *y* and *m* is total number of input variables. If the relationship is established between the set members:

…15

Where *f* is smooth function and *r* is a random variable. GT is an estimated output variance of a non-smooth model. GT is based on the *kth* (*1 ≤ k ≤ p*) nearest neighbors for each vector *Xi (1 ≤ i ≤ M)*. Delta function calculates the mean square distance of the *kth* neighbor:

…16

In which *| |* indicates Euclidean distance, corresponding gamma function is as follows:

…17

Where, *YN [i,k]* is value of *y* corresponding to the *kth* neighbor of *Xi* in Eq. 3. In order to calculate gamma, the linear regression is fitted from *P* spot to values of and .

…18

The intercept of this line =0 indicates the gamma value and is equal to the errors variance and it is provided that the combination *2n - 1* would be among them. Reviewing all the alternative combinations takes much time. Gamma test can determine the most efficient variable in modeling and the best input combination (Lafdani et al., 2013).

**RESULTS AND DISCUSSION**

This chapter focuses on selecting the best input variables based on the autocorrelation function, cross-correlation function, and Gamma test. These variables were used for the development and application of an artificial neural network model to predict monthly rainfall for the Central Telangana district of Medak, utilizing monthly rainfall data spanning 113 years (from January 1901 to December 2014). The data was divided into two sets viz. training data set (January, 1901 to December, 1985) and testing data set (January 1986, to December 2014) and were used for training and testing of developed models. Performance of the developed models was evaluated qualitatively and quantitatively by visual observations and quantitatively employing various statistical and hydrological indices viz. Root Mean Square Error, Integral Square Error, Percent Bias, Correlation Coefficient and Coefficient of Efficiency. The best model is selected based on lower value of Root Mean Square Error, Integral Square Error, Percent Bias and higher values of Correlation Coefficient and Coefficient of Efficiency.

**Development for rainfall prediction models**

The prediction of rainfall is an enormously complex, vibrant, dynamic, and nonlinear process, which is affected by many factors that are frequently interrelated. The methodologies used for rainfall prediction cover a wide range of methods from completely black-box models to very detailed conceptual models. “Previous researches were mainly emphasis on two models: (a) deterministic/conceptual models that consider the dynamics of the principal process, and (b) systems theoretical/black-box models that do not consider the principal dynamics of the process. A black box model is an input-output pattern of which there is no erstwhile information available and these models define the causal link between input-output patterns by alteration. One of the approaches for system theoretical modeling based on artificial neural networks has recently become very popular in hydrological modeling and engineering due to their simplicity and adaptability to mug up and gather information from situations. In the present study, artificial neural network models with different activation functions have been developed to predict rainfall monthly. Table 2 presented the pattern of all activation function having input and output pairs for training and testing data pattern for rainfall prediction. The rainfall of a station has also predicted using rainfall of adjoining two stations and different combinations of lag rainfall (Spandana, K. et al., 2021).

**Table 2** Input-output pairs in training and testing for rainfall prediction

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Year (i)** | **Month**  **(j)** | **Input for jth months** | | | | **Output** |
| X1 | X2 | X3 | X4 | Y |
| Pi,j-1 | Pi,j-2 | Pi,j-3 | Pi,j-4 | Pi,j |
| 1 | 1 | P1,2 | P1,3 | P1,4 | P1,5 | P1,1 |
| 2 | P1,3 | P1,4 | P1,5 | P1,6 | P1,2 |
| 3 | P1,4 | P1,5 | P1,6 | P1,7 | P1,3 |
| … | … | … | … | … | … |
| 12 | P1,11 | P1,10 | P1,9 | P1,8 | P1,12 |
| 2 | 1 | P2,2 | P1,3 | P1,2 | P1,4 | P2,1 |
| 2 | P2,3 | P2,4 | P2,5 | P2,6 | P2,2 |
| 3 | P2,4 | P2,5 | P2,6 | P2,7 | P2,3 |
| … | … | … | … | … | … |
| 12 | P2,11 | P2,10 | P2,9 | P2,8 | P2,12 |
| … | … | … | … | … | … | … |
| 114 | 1 | P114,1 | P114,2 | P114,3 | P114,4 | P114,1 |
| 2 | P114,2 | P114,3 | P114,4 | P114,5 | P114,2 |
| 3 | P114,3 | P114,4 | P114,5 | P114,6 | P114,3 |
| … | … | … | … | … | … |
| 12 | P114,11 | P114,10 | P114,9 | P114,8 | P114,12 |

“The gamma test was used for selection of input data. In this study, Gamma test is used for identification of input parameter for monthly rainfall prediction and remove those input parameters which has insignificant contribution to output. The large number of inputs and insignificant inputs variable has increased the complexity of model and it is main cause of the overfitting of model. The gamma test helps to take decisions about selection of input data or input which are actually affecting the result of developed models. Number of inputs was selected on the basis of gamma value (┌), standard error and V-ratio” (Spandana K. et al., 2024). “The maximum variation in values of gamma and standard error were considered as the most influence model” (Lafdani et al., 2013). “The best input selection procedure, different combinations of input data were explored to assess their influence on the monthly rainfall prediction, from which meaningful. combinations are given in Tables 2 through Table 3. Also, input parameters were selected using autocorrelation function and cross correlation. Based on comparison of results of gamma test and autocorrelation function and cross correlation common input parameter was taken as appropriate input variables. After best input combinations are chosen software of Neuo- Solutions was used for development of best ANN models” (Spandana K. et al., 2021).

It is observed from results of gamma test and autocorrelation function that the present month rainfall (R M ) of Medak station depends on the two lag months of rainfall as given in Table 2. Finally, different inputs were selected for three different stations are i.e.

Table 3. Gamma test and autocorrelation function analysis for Month rainfall of Medak station

(R M = R M-1 , R M-2 ) for Medak station.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Output** | **Input** | **Mask** | **Gamma** | **SE** |
| R M | R M-1, R M-2, R M-3, R M-4, R M-5 | 11111 | 0.08710 | 0.00636 |
| R M | R M-1, R M-2, R M-3, R M-4 | 11110 | 0.09123 | 0.00513 |
| R M | R M-1, R M-2, R M-3, R M-5 | 11101 | 0.08800 | 0.00682 |
| R M | R M-1, R M-2, R M-4, R M-5 | 11011 | 0.09195 | 0.00687 |
| R M | R M-1, R M-3, R M-4, R M-5 | **10111** | **0.09890** | **0.00382** |
| R M | R M-2, R M-3, R M-4, R M-5 | **01111** | **0.09408** | **0.00387** |

**Table 4 : ANN models of** **Medak, Khammam and Warangal district for monthly rainfall prediction**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Input**  **variables** | **Learning**  **algorithm** | **Transfer**  **function** | **Learning**  **rule** | **Hidden**  **layer** |
| M1 | R M-1, R M-2 | MLP | Sigmoid axon | LM | Single |
| M2 | R M-1, R M-2 | MLP | Sigmoid axon | DBD | Single |
| M3 | R M-1, R M-2 | MLP | Tanh axon | LM | Single |
| M4 | R M-1, R M-2 | MLP | Tanh axon | DBD | Single |
| M5 | R W , R W-1, R K | MLP | Sigmoid axon | LM | Single |
| M6 | R W , R W-1, R K | MLP | Sigmoid axon | DBD | Single |
| M7 | R W , R W-1, R K | MLP | Tanh axon | LM | Single |
| M8 | R W , R W-1, R K | MLP | Tanh axon | DBD | Single |

**Performance Evaluation of Models**

*Percentage bias*

It calculates how likely simulated values are on average to be greater or less than the actual values. Zero is the ideal PBIAS value, while low magnitude values signify accurate model simulations.

…18

Where, Qp = predicted monthly rainfall, and Qo = observed monthly rainfall

*Root mean square error*

This method is employed to measure the prediction accuracy of a model, compare the difference between predicted and observed values, and get the information on short term performance. The RMSE is zero for perfect fit and increased values indicate higher deviation between predicted and observed values. The root mean square error (RMSE) is determined by following relationship:

…19

Where, Qp =predicted monthly rainfall, and Qo =observed monthly rainfall

*Integral Square error*

Integral squared error is a measure of system performance formed by integrating the square of system error over a fixed interval of time; and is calculated by following equation:

…20

Where, Qp =predicted monthly rainfall, and Qo =observed monthly rainfall

*Correlation coefficient*

It is an indicator of degree of range between observed and predicted values and provides the level of variance explained between observed and predicted. If observed and predicted values are completely independent, the r will be zero (Smith et al. (2023). The correlation coefficient is determined using the following equation:

…21

Where, Qp =predicted monthly rainfall, Qo =observed monthly rainfall, = average of the observed monthly rainfall and = average of predicted monthly rainfall

*Coefficient of efficiency*

The Coefficient of efficiency provides the proportions of variance of the observation for model and often used in hydrological analysis. Nash-Sutcliffe coefficient of efficiencies range between to 1. It’s value signify 1 for a perfect match and 0 when predictions equal the mean for observed data series. The Coefficient of efficiency is determined by using the following equation:

…22

Where, Qp =predicted monthly rainfall, Qo =observed monthly rainfall and = average of the observed monthly rainfall

**Comprehensive Insights into ANN Model Performance**

The model's qualitative performance was assessed by visually comparing predicted and observed monthly rainfall, evaluating over prediction and underproduction during training and testing across the station. Monthly rainfall predictions for Medak station using ANN models (M-1 to M-8) with two learning rules (Levenberg-Marquardt, Delta-bar-delta) and transfer functions (Sigmoid, Tanh) were evaluated qualitatively using prepared plot between observed and predicted plots and scatter plots for training (1901-1985) and testing (1986-2014) periods. Scatter plots show that models M-1 to M-4 under predicted rainfall, while M-5 to M-8 closely matched observed values, with overall predictions aligning well with actual rainfall.

Figure 3 presented the time series graph of training and testing data between observed and predicted rainfall of M1 model in Medak station. This graph compared actual and predicted rainfall over time. The blue line shows the real rainfall, and the red line shows the model's predictions. The predictions often underestimate the actual rainfall, but they follow the overall trend

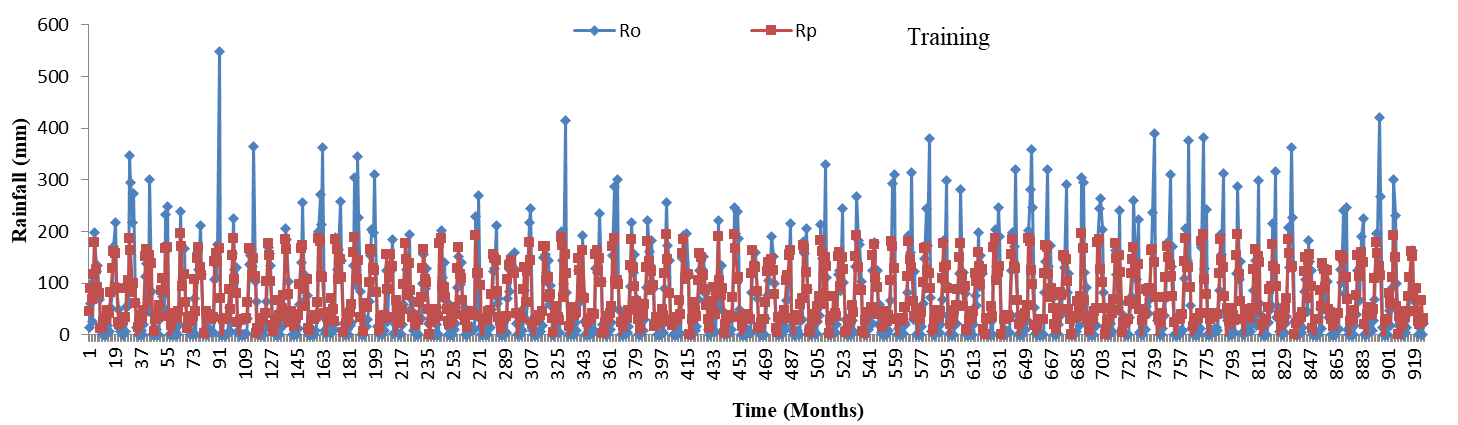
**Fig. 3 Monthly Observed and Predicted Rainfall data for M-1 Model at Medak Station**

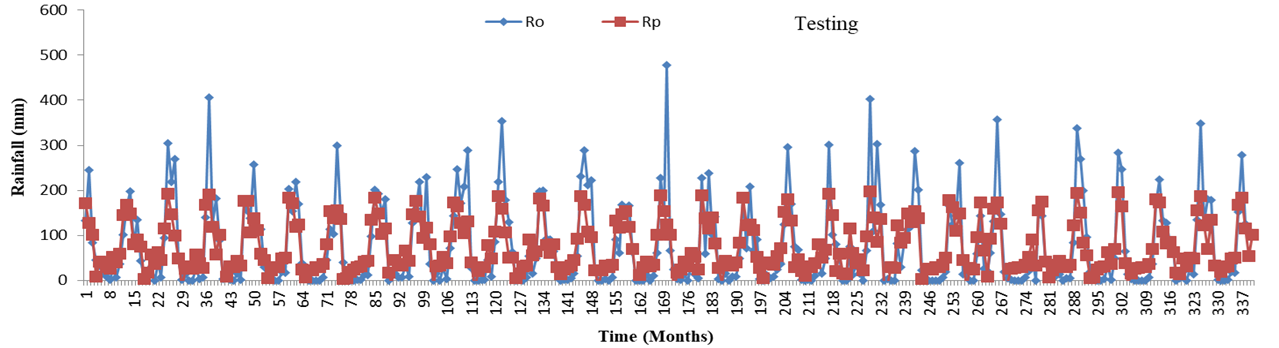
Scattered plot for M1 model was presented in figure 4, where it presented that how well the model's predicted rainfall matches the actual rainfall. Each red dot is a data point, and the black line shows the general trend of the predictions. The equation y=0.5378x+32.971y = 0.5378x + 32.971, y=0.5378x+32.971 means the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.694, 0.679 shows the predictions are fairly accurate but not perfect.

|  |  |
| --- | --- |
| **Training data** | **Testing data** |
| **M1 Model** | |
| **Training data** | **Testing data** |
| **M2 Model** | |

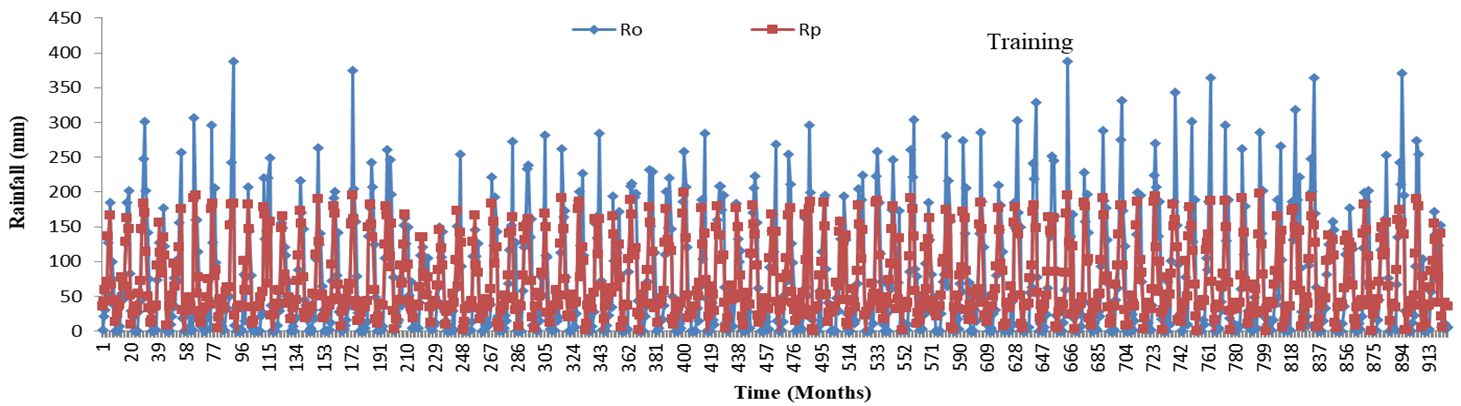
**Fig. 4 Scatter diagram of M 1 and M 2 model of Medak station**

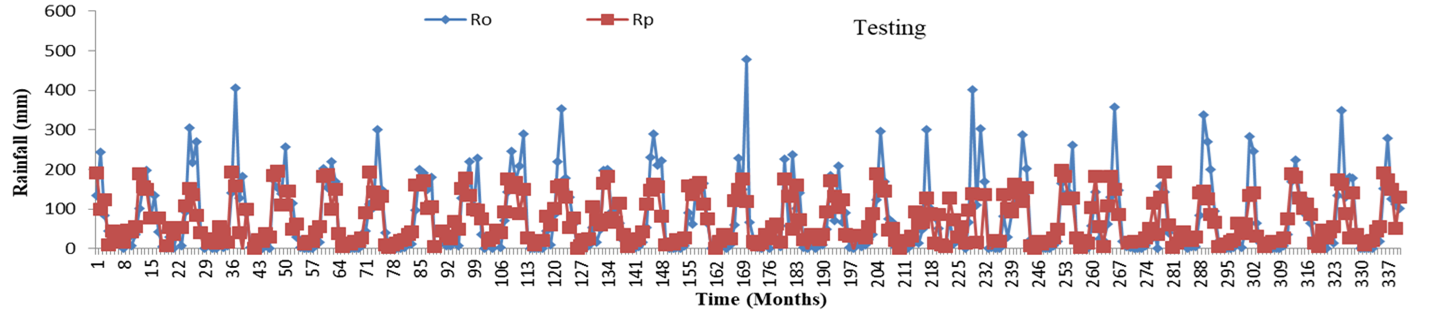
Similarly, for the M2 Model, the scatter diagram expressed the degree of alignment between the predicted rainfall and the actual rainfall. Each red dot represents an individual data point, while the black line indicates the overall trend of the predictions. The equation y=0.4853+31.961, Y= 0.5213+35.17, means the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.751, 0.691, shows the predictions are fairly accurate but not perfect.



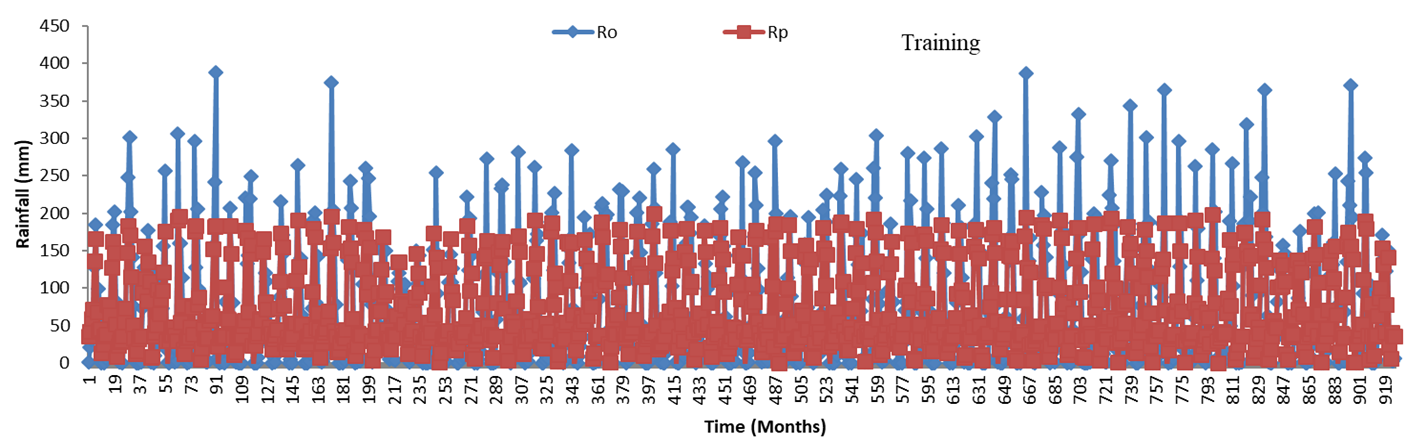


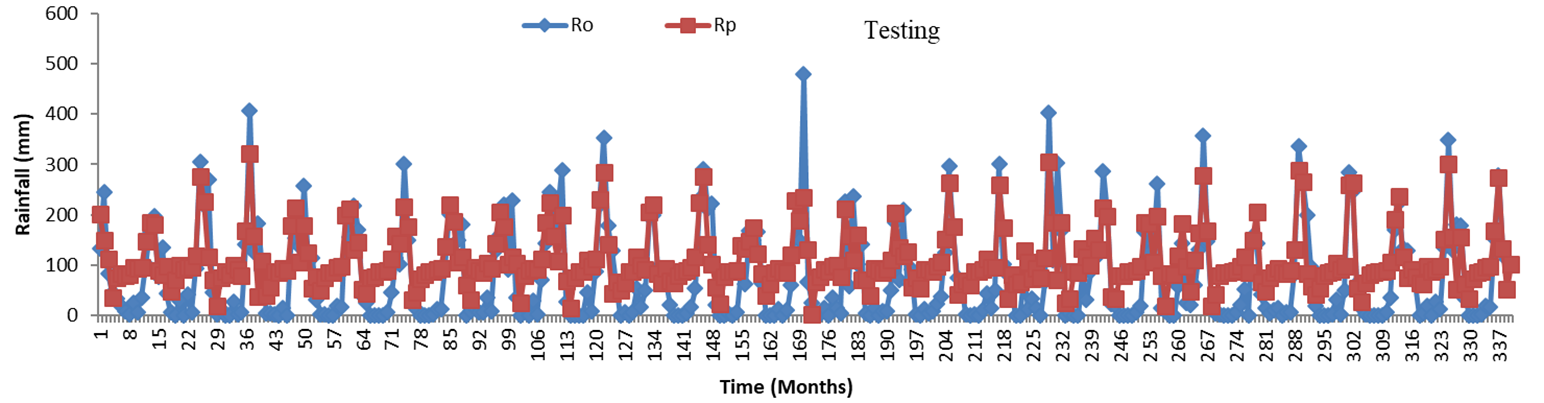
**Fig. 5 Monthly Observed and Predicted Rainfall data for M2 Model at Medak Station**





**Fig. 6 Monthly Observed and Predicted Rainfall data for M3 Model at Medak Station**



**Fig. 7 Monthly Observed and Predicted Rainfall data for M4 Model at Medak Station**

Figures 5, 6, 7, and 9 show the line graphs for the M2, M3, M4 and M5 models, which contrast the actual rainfall over time with the predicted rainfall values. The red line in these graphs highlighted the model's predictions, and the blue line shows the actual rainfall. The predictions often underestimate the actual rainfall, but they follow the overall trend.

|  |  |
| --- | --- |
| **Training data** | **Testing data** |
| **M3 Model** | |
| **Training data** | **Testing data** |
| **M4 Model** | |
|  |  |
| **M5 Model** | |

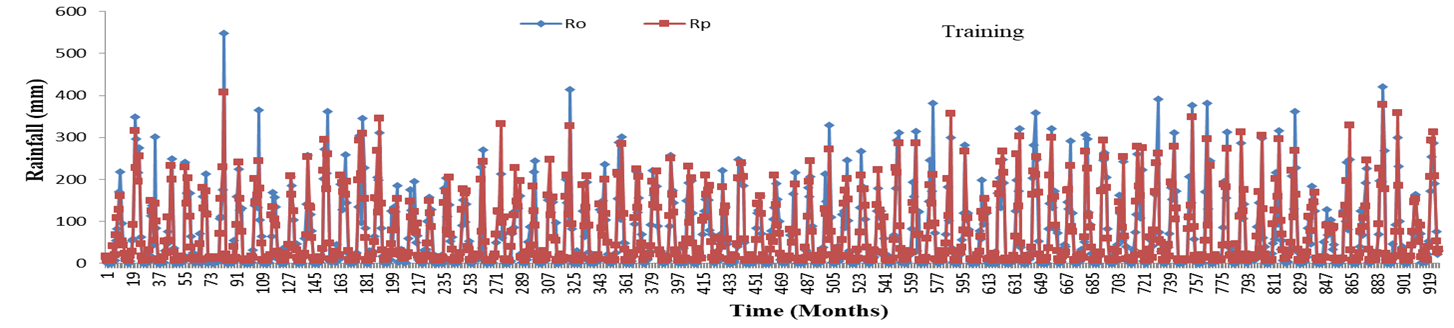
**Fig. 8 Scatter diagram of M 3, M4 and M5 model of Medak station**

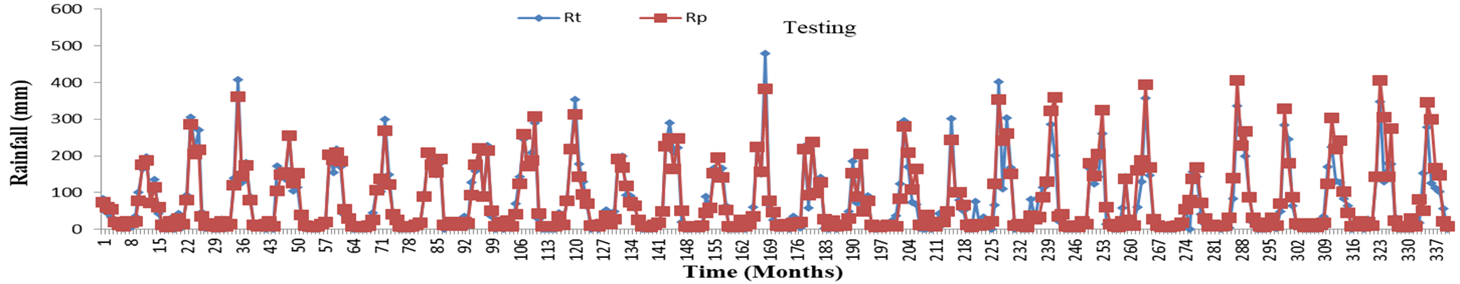
This scatter plot shows how well the model's predicted rainfall matches the actual rainfall. Each red dot is a data point, and the black line shows the general trend of the predictions. The equation y=0.5568X+29.79, Y= 0.4713X+33105, means the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.7757 slightly accurate, 0.5466, shows the predictions are very less accurate not perfect.

This scatter plot of M4 model shows how well the model's predicted rainfall matches the actual rainfall. Each red dot is a data point, and the black line shows the general trend of the predictions. The equation y=0.4587x+65.387 y=0.5123x+71.568means the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.692, 0.671 shows the predictions are fairly accurate not perfect.

This scatter plot made by M5 Model shows how well the model's predicted rainfall matches the actual rainfall. Each red dot is a data point, and the black line shows the general trend of the predictions. The equation y=0.8912x+7.6865, y=1.0338x+48753 means the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.89 good accurate, 0.849 shows the predictions are accurate& perfect.

**Fig. 9 Monthly Observed and Predicted Rainfall data for M5 Model at Medak Station**



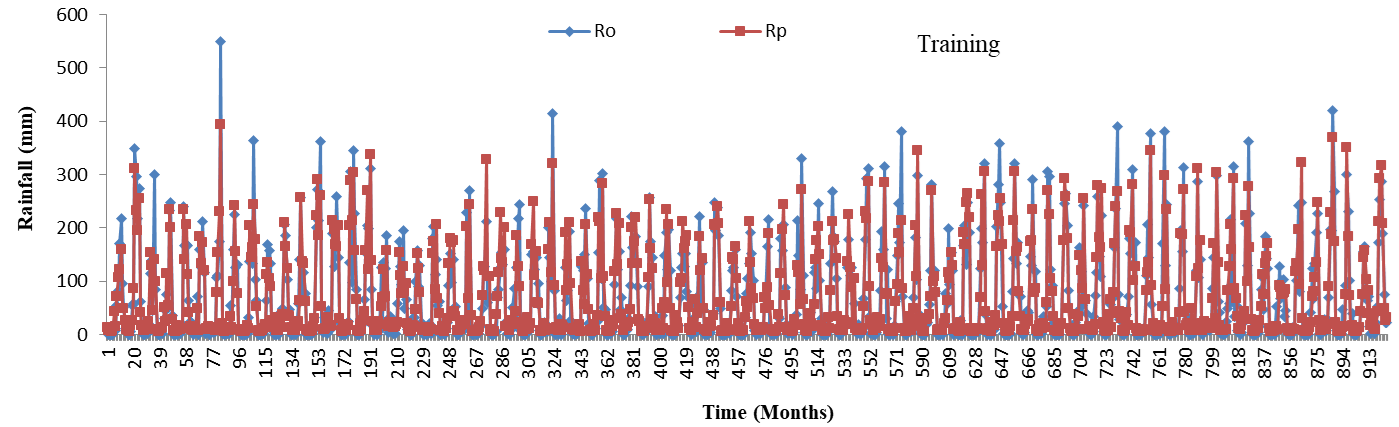


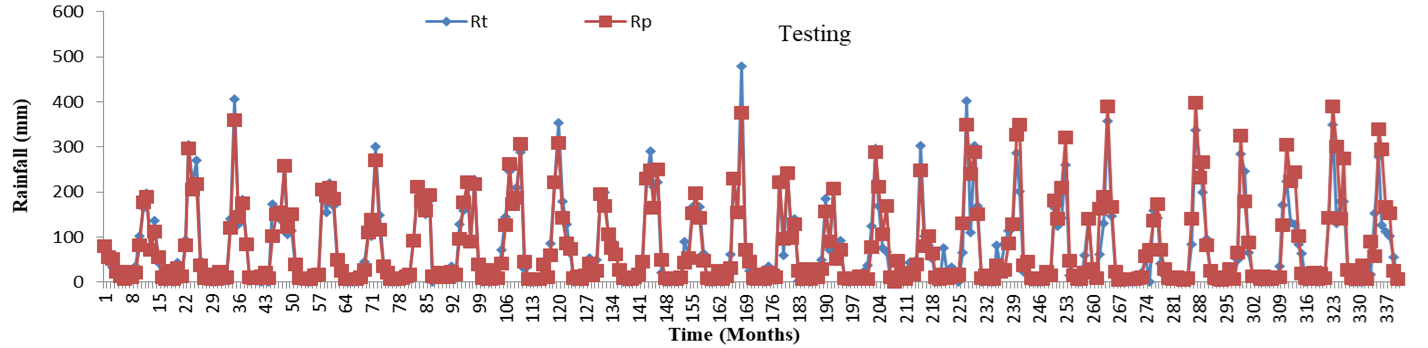
**Fig. 10 Monthly Observed and Predicted Rainfall data for M6 Model at Medak Station**

Figures 10, 12, and 13 display line graphs for the M6, M7, and M8 models, contrasting the actual rainfall over time with the predicted values. In these graphs, the red line represents the model's predictions, while the blue line indicates the observed rainfall. Although the predictions frequently underestimate the actual rainfall, they closely follow the overall trend.

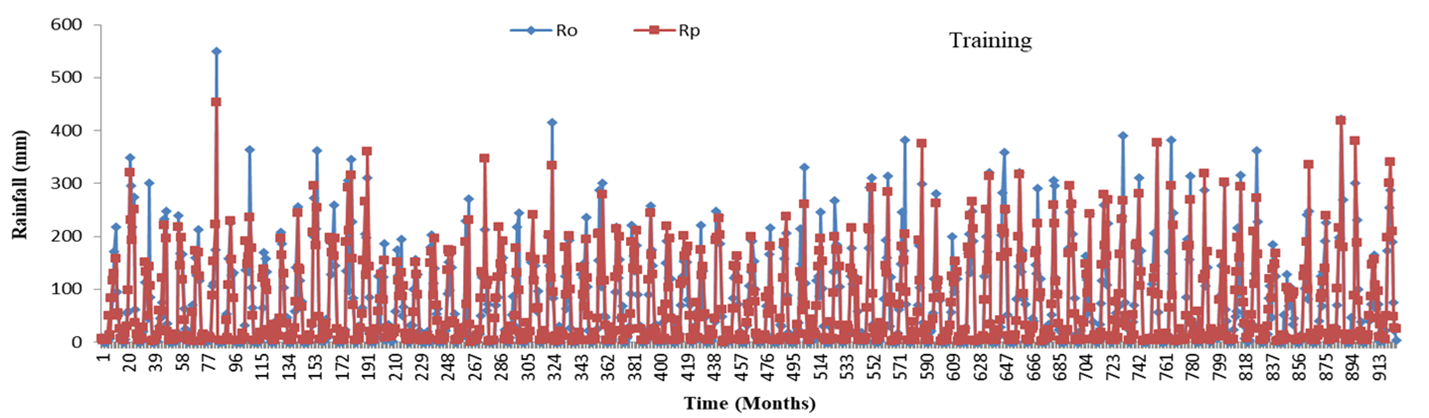
|  |  |
| --- | --- |
| **Training data** | **Testing data** |
| **M6 Model** | |
| **Training data** | **Testing data** |
| **M7 Model** | |
| **Training data** | **Testing data** |
| **M8 Model** | |

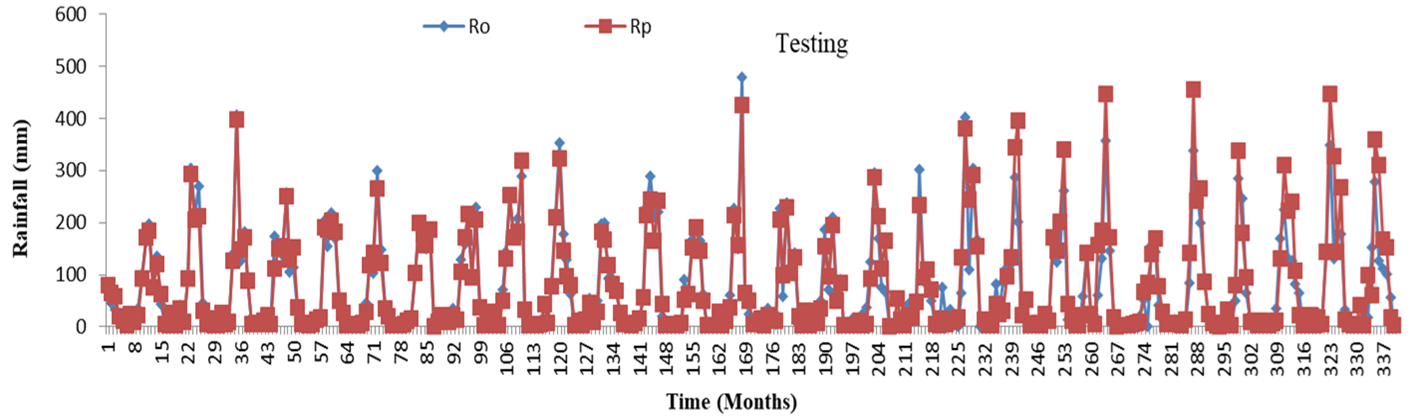
**Fig. 11 Scatter diagram of M6, M7 and M8 model of Medak station**





**Fig. 12 Monthly Observed and Predicted Rainfall data for M7 Model at Medak Station**

****

****

**Fig. 13 Monthly Observed and Predicted Rainfall data for M8 Model at Medak Station**

Figure 11 revealed the scattered diagram of training and testing data for M6, M7 and M8 model. This scatter plot of M6 model presented where each red dot is a data point, and the black line reflected the general trend of the predictions. The equation y=0.9643+8.3354, y=0.8652x+10.441 means the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.893 good accurate, 0.877 shows the predictions are accurate & perfect. Similarly, M7 and M8 Model scatter plot follow the same trend. The equation of M7 model was y=0.8707+9.5497, y=0.9695x+70612 signify that the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.877 good accurate, 0.895 shows the predictions are accurate & perfect. Moreover, M8 model’s equation y=0.8833x+8.394, y=1.011x+4.802 presents the model slightly underpredicts rainfall, and the R2R^2R2 value of 0.882 good accurate, 0.890 shows the predictions are accurate& perfect.

Based on a meticulous evaluation of the results of the ANN models, models M-1 to M-4 performed well for low monthly rainfall but struggled to accurately predict peak rainfall during both training and testing. Throughout the analysis, models M-5 to M-8 closely resemble the observed rainfall. Thus, the ANN models provide satisfactory predictions across all stations.

**Quantitative performance analysis**

*Correlation coefficient*

The correlation coefficients for ANN-based models were calculated, and the results for Medak station are shown in the table. For the best-performing model, M-8 (network 3-2-1), the correlation coefficients during the training period (1901-1985) and testing period (1986-2014) are 0.938 and 0.93, respectively. **Moriasi et al. (2007),** reported the NSE values classify model performance as unsatisfactory (>0.50), satisfactory (0.50–0.65), good (0.65–0.75), and very good (0.75–1.00). For the ANN-based model M-8 at Medak station, NSE values are 0.91 (training) and 0.93 (testing), both in the "very good" class, showing the model's strong performance.

*Integral square error*

The integral square error (ISE) is a statistical measure used to evaluate the model's performance. For the best ANN model M-8 (3-2-1 network) with a tanh transfer function and delta bar delta activation function, the ISE values for Medak station are 1.405 during the training period and 2.31 during the testing period, as shown in the table 5.

*Percent bias*

According to Moriasi et al. (2007), PBIAS is classified into four categories: unsatisfactory (> ±25%), satisfactory (±15% to ±25%), good (±10% to ±15%), and very good (< ±10%). For the best ANN models: Medak station (Model M-8, 3-2-1 network, tanh transfer function, delta bar delta activation): PBIAS values are 1.42% (training) and 7.16% (testing), both "very good".

**Table 5. Comparison of different ANN models for rainfall prediction of Medak station**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | | M-1 | M-2 | M-3 | M-4 | M-5 | M-6 | M-7 | M-8 |
| Arch | | 2-1-1 | 2-4-1 | 2-4-1 | 2-20-1 | 3-5-1 | 3-8-1 | 3-1-1 | 3-2-1 |
| Training | r | 0.642 | 0.620 | 0.651 | 0.522 | 0.945 | 0.938 | 0.934 | 0.938 |
| RMSE | 0.013 | 0.013 | 0.049 | 0.062 | 0.002 | 0.003 | 0.011 | 0.014 |
| NSE | 0.67 | 0.66 | 0.59 | 0.68 | 0.92 | 0.91 | 0.91 | 0.91 |
| PBIAS | 0.04 | 1.56 | 0.08 | 38.05 | 0.24 | 1.4 | 0.48 | 1.42 |
| ISE | 2.31 | 2.35 | 2.62 | 2.73 | 1.31 | 1.39 | 1.38 | 1.405 |
| Testing | r | 0.821 | 0.882 | 0.861 | 0.851 | 0.942 | 0.945 | 0.944 | 0.946 |
| RMSE | 0.092 | 0.082 | 0.166 | 0.185 | 0.061 | 0.161 | 0.103 | 0.102 |
| NSE | 0.61 | 0.64 | 0.49 | 0.71 | 0.9 | 0.93 | 0.93 | 0.93 |
| PBIAS | 6.17 | 3.85 | 8.68 | 47.1 | 9.98 | 7.71 | 6.51 | 7.16 |
| ISE | 4.06 | 3.95 | 4.61 | 4.81 | 3.02 | 2.32 | 2.29 | 2.31 |

*Root mean square error*

The root mean square error (RMSE) of the M-8 model for Medak station is 0.0142 mm during training (1901-1985) and 0.102 mm during testing (1986-2014). Models M-5 to M-8, using rainfall data from adjoining stations, perform better than M-1 to M-4, which rely on same-station rainfall data. The M-8 model accurately predicts monthly rainfall for Medak station based on previous one-day and current-month rainfall from Warangal.

**Conclusion**

In conclusion, the analysis of the result provides the valuable insight into the ANN model-based Rainfall prediction in Medak District of Central Telangana. In order to anticipate monsoon rainfall, this study used ANN models, with two distinct activation function. Findings revealed that the M-5 model performed best with high accuracy, low error, and good consistency during training and testing. Although, model 6 and 7 also expressed strong performance. Models having 3 layers (M-5 to M-8) generally outperform simpler 2-layer models. On other side, the performance of model 4 (M-4) shown the poor result, having high error and biasness. However, some limitation was constrained the study like these results and findings are specific to the dataset and may not generalized well to others. Models lack interpretability, making it hard to explain predictions. Moreover, high bias was seen in some model like M-4, leading to systematic errors. By addressing the challenges in the study employing ANN models, this work opens new avenues for exploration to test the advanced models like RNNs or CNNs for better performance, because this study merely focused on testing the performance under feed-forward networks. Fine-tuned model parameters need to be used for acquiring better accuracy. Similarly, using techniques like SHAP or LIME enhances the interpretability of the model. Moreover, the use of combined strong models like M-5 and M-6 can create ensembles for improved rainfall predictions. Consequently, integrating the ANN prediction along with qualitative analysis bridges the gap between academia and practical application, assist to the government and local agencies in designing effective drought mitigation, crop insurance, and climate adaptation strategies. This research offers a standard for upcoming research and uses in regional climate analysis and meteorology.

**Disclaimer (Artificial intelligence)**

**Option 1:**

**Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.**

**Option 2:**

**Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology**

**Details of the AI usage are given below:**

**1.**

**2.**

**3.**

**References**

1. Agricultural Statistics at a Glance Telangana. (2013-14). Telangana state agriculture department.
2. Anmala, J., Hsu, K. L., & Sorooshian, S. (2000). Using artificial neural networks for estimating runoff at different time scales. *Water Resources Research, 36*(9), 2511-2519.
3. ASCE Task Committee. (2000a). Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering, 5*(2), 115-123.
4. ASCE Task Committee. (2000b). Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering, 5*(2), 124-137.
5. Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes, 6*(3), 279-298.
6. Danladi, A., Stephen, M., Aliyu, B. M., Gaya, G. K., Silikwa, N. W., & Machael, Y. (2018). Assessing the influence of weather parameters on rainfall to forecast river discharge based on short-term. Alexandria Engineering Journal, 57(2), 1157-1162.
7. Dawson, C. W., & Wilby, R. L. (2001). Hydrological modeling using artificial neural networks. *Progress in Physical Geography, 25*(1), 80-108.
8. Fernando, T. M. K. G., & Jayawardena, A. W. (1998). Runoff forecasting using RBF networks with OLS algorithm. *Journal of Hydrologic Engineering, 3*(3), 203-209.
9. Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., & Xavier, P. K. (2006). Increasing trend of extreme rain events over India in a warming environment. *Science, 314*(5804), 1442-1445.
10. Government of India. (2023). Land use and agriculture statistics. Ministry of Agriculture & Farmers Welfare.
11. Hsu, K. L., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research, 31*(10), 2517-2530.
12. Hu, T., Li, X., & Chen, X. (2005). Artificial neural networks and their applications in hydrology. *Advances in Water Science, 16*(2), 234-245.
13. Hung, N. Q., Babel, M. S., Weesakul, S., and Tripathi, N. K.(2009) An artificial neural network model for rainfall forecasting in Bangkok, Thailand, *Hydro. Earth Syst. Sci*., 13, 1413–1425.
14. Keskin, M. E., & Terzi, O. (2006). Artificial neural network models of lake water level fluctuations in Turkey. *Mathematical and Computational Applications, 11*(3), 193-200.
15. Lafdani, E. K., Bardossy, A., & He, Y. (2013). Selection of input variables for hydrological models. *Hydrology and Earth System Sciences, 17*(2), 689-700.
16. Luk, K. C., Ball, J. E., & Sharma, A. (2001). A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *Journal of Hydrology, 246*(1-4), 231-241.
17. Michael Nielsen. (2015). *Neural Networks and Deep Learning*. Online book.
18. Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE, 50*(3), 885-900.
19. NITI Aayog. (2018). Composite water management index: A tool for water management. Government of India.
20. Planning Commission, Government of India. (2011). *The Future of Food and Farming: Challenges and Choices for Global Sustainability.*
21. Rajurkar, M. P., Kothyari, U. C., & Chaube, U. C. (2002). Artificial neural networks for daily rainfall-runoff modelling. *Hydrological Sciences Journal, 47*(6), 865-877.
22. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536.
23. Sahai, A. K., & Mohanty, U. C. (2000). A study on artificial neural network (ANN) technique for the prediction of Indian summer monsoon rainfall. *Meteorology and Atmospheric Physics, 75*(1-2), 161-177.
24. Sammen, S. S., Kisi, O., Ehteram, M., El-Shafie, A., Al-Ansari, N., Ghorbani, M. A., ... & Shahid, S. (2023). Rainfall modeling using two different neural networks improved by metaheuristic algorithms. Environmental Sciences Europe, 35(1), 112.
25. Smith, J., Wang, R., & Li, X. (2023). Statistical evaluation methods for hydrological models. *Hydrology Research, 54*(2), 345-357.
26. Sudheer, K. P., & Jain, S. K. (2004). Rainfall-runoff modeling using artificial neural networks: Model performance comparison. *Hydrological Processes, 18*(4), 795-807.
27. Toth, E., Brath, A., & Montanari, A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology, 239*(1-4), 132-147.
28. Wilby, R. L., & Dawson, C. W. (2003). Using artificial neural networks for rainfall-runoff modeling in the UK. *Hydrology and Earth System Sciences, 7*(5), 629-639.
29. Spandana, K., Kumar, D., & Singh, V. K. (2021). Development of ANN models using monthly rainfall for central Telangana region. IJCS, 9(1), 1920-1933.