

A Hybrid Rainfall Prediction Model Using a Long Short Term Memory Neural Network and Genetic Algorithm

Dr. Vimal Patel¹, Dr. Tejash Chaudhari², Dr. Chetan Rathod³, Dr. Bhumika Charnanand⁴

¹ College of Agricultural Engineering and Technology, NAU, Dediapada, Gujarat, India

² College of Agriculture, NAU, Waghai, Dangs, Gujarat, India

³ Vimal Tormal Poddar BCA College, Pandesara, Surat, Gujarat, India

⁴ Department of Computer Science, School of Science & Technology, Vanita Vishram Women's University, Surat, Gujarat, India

Abstract: Rain is a natural source of water and it is not only important for the human beings but also for the survival of plants and animals. It plays a significant role in industry, agriculture, electrical energy and others. Due to the uncertainty and irregularity of rain can lead to occurrence of natural disasters like floods and droughts and also make harm to the infrastructure and life of the living being. Therefore, it is essential to predict rainfall for the effective management in agricultural sector, drought assessment, water management and other areas. Now days, many more neural networks are used for rainfall prediction. One of these is the long short-term memory (LSTM), which has good prediction results and can be trained for sequence synthesis by processing real data sequences step by step. In this study, a hybrid rainfall prediction model is developed using Long Short Term Memory neural network and Genetic Algorithm using historical data of 20 years of rainfall. To assess the model's performance, a categorical cross-entropy loss function is employed. The outcomes of the experiment show that a hybrid rainfall prediction model is suitable for rainfall prediction and has exceptional prediction accuracy.

Keywords: LSTM, Genetic algorithm, rainfall prediction, neural network

Introduction: Rain is a natural source of freshwater that comes down to earth directly from the sky. Rain generally has a significant impact on the productivity of agricultural products and agriculture-dependent industrial development. In India, rainfall is one of the factors that affect crop cultivation. The unpredictable rain and variations in rainfall patterns can harm agricultural crops, infrastructure, and the lives of living beings. Thus, a significant effort is required to safeguard the severely affected agriculture and natural disaster due to any variation in rainfall pattern [1] [2]. Rainfall prediction is the process of using science and technology to predict the amount of rain for a specific area and time frame [3]. Prior information about the rain can help to save the infrastructure and life of the living being from the natural calamities like floods, droughts, tsunamis and others. Rainfall prediction is a persistent, changing and challenging phenomenon in the modern world. Thus, a significant effort is required to accurately predict the rain, which helps to take crucial decision in various sectors [4].

The term "Artificial Neural Network" is derived from biological neural networks, which aid in the formation of the human brain's structure. The multiple layers of ANNs' neurons are interconnected with one another in a manner similar to the way neurons in the human brain are linked to each other. The ANN is a set of interconnected artificial neurons that can store and disseminate experiential knowledge to practitioners and researchers as an alternative tool. Nonlinear data-driven ANNs

are able to perform nonlinear modeling without having prior expertise in the connections between the input and output variables [5]. It can predict problems that occur in real-life situations and generate almost flawless solutions with the help of a suitable training and testing process [6]. Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory Networks (LSTMs) are the different types of ANN. The RNNs can be trained to learn sequential or time-varying patterns by facilitate time delay units through feedback connections [7]. LSTM is the most common variant of RNN. A traditional RNN only has one hidden state, so it could be difficult for the network to learn long-term dependencies. LSTM can automatically store data for a long time in a memory cell and solved the problem of RNNs' long-term dependency. The ability of LSTM networks to learn long-term dependencies in sequential data makes them ideal for tasks like time series forecasting, speech recognition, and language translation [8]. Charles Darwin's "Survival of the Fittest" theory serves as the foundation for the Genetic algorithm, a search-based optimization method. In 1992, J.H. Holland proposed GA. [9]. GA draws inspiration from the way that species naturally evolve by selecting the fit individuals to create the strongest possible solution [10]. The population's best individual will use evolutionary operators including selection, crossover, and mutation to find the best solution. This process starts with an initial population that was generated at random. The selection operator selects parents for mating based on their fitness and applies the crossover and mutation operators to produce new children in each generation. Finally, the fittest children replace the population's existing parents and the process repeats itself [11]. The GA flow chart is shown in figure 1 below. The fundamental components of GA are chromosome representation, computing fitness functions, selection, crossover and mutation.

Keras is an excellent high-level open-source deep learning API framework written in Python [6]. It is a library that provides highly powerful and abstract building blocks to build deep learning networks using other DL tools such as TensorFlow, CNTK (Cognitive Toolkit) and Theano. To create and maintain Keras, Francois Chollet used four guiding principles: modularity, simplicity and ease of use, extensibility, and adherence to the Python model [12].

The loss function is used to determine the loss or error of the model by comparing the result data with the original data [13]. The associated loss or error is the difference between the predicted values derived from the output layer and the actual values. The aim of the loss functions is to reduce the loss as much as close to zero. A model's performance improves with decreasing loss. Loss functions assist the model to determine if it is progressing correctly or not during training. Different types

of loss functions are available in Keras, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Hinge, Binary Cross-Entropy, Categorical Cross-Entropy and many more [14].

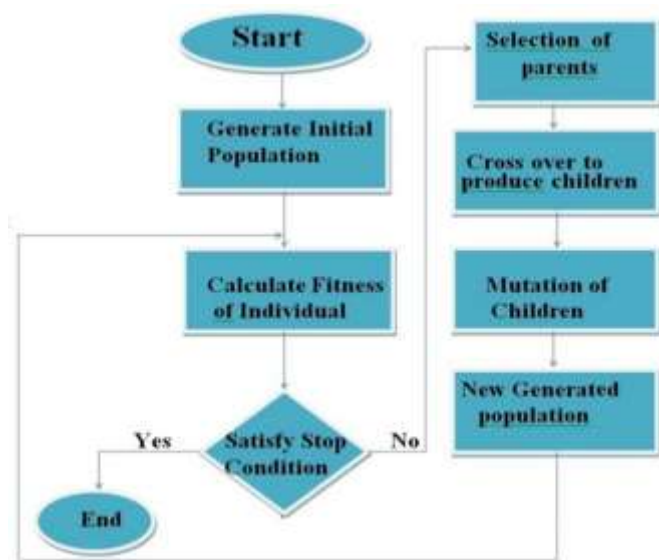


Fig - 1: Flow chart of Genetic Algorithm

Literature Review: The authors of the publication [15] conducted a thorough survey and analysed different NN designs utilised by various studies for rainfall prediction, including Feed Forward Neural Network (FFNN), Recurrent Neural Network (RNN), and Time Delay Neural Network (TDNN).

A variety of metrics, including data size, delay, required connections, training memory, learning speed, and others, are used to compare the performance of this NN. In order to predict daily, monthly, and annual rainfall using different NNs, they identified and examined a number of issues, such as the number of hidden layers and nodes to be used, the separation of training and test data, overfitting, the choice of performance measures, the training algorithm, and the activation function. According to the survey results, the author found that FFNN is the best method for predicting monthly and daily rainfall, while TDNN is the best method for predicting yearly rainfall. The survey shows that when compared to other rainfall prediction methods including statistical and numerical approaches, RNN, FFNN, and TDNN yield significant results.

The authors of the paper [16] used the long short-term memory (LSTM), which can be trained for sequence generation by processing real data sequences one step at a time and has good prediction results to investigate the changes of rainfall-runoff values and make a prediction. They used the cross-validation method to ensure the accuracy of the trained model by dividing the training data into 12 parts. A method of modeling analysis and parameter estimation of hydrological runoff and discharge relationship based on machine learning is designed. From the experimental results, the average risk of this method is 61.23%, which is 15.104% and 13.397% less than that of the other two existing methods, respectively. It proves that the method of hydrological runoff relationship modeling and parameter estimation integrated with machine learning has better practical application effect. Thus, the prediction result shows high accuracy and good consistency with the real results.

The authors of the paper [17] developed a rainfall prediction model using long short-term memory (LSTM) neural network approach. They have incorporated seasonal decomposition methods for rainfall prediction based on Australian preprocessed dataset for the year 2007 to 2016. The parameters and factors taken into consideration for improvement in accuracy and performance evaluation include learning rate, epochs, number of hidden layer neurons, loss and Root Mean Square Error (RMSE). They have compared the proposed LSTM based model results with ANN results and concluded that the LSTM network outperformed ANN in all the cases. They have also observed that the proposed model has predicting accurately for moderate cases, but there was an increment in the error rate in cases of outliers.

The authors of the paper [18] have adopted long short-term memory (LSTM) neural network deep learning approach to predict rainfall. They have used European Centre for Medium-Range Weather Forecasts data with seven weather parameters to predict hourly rainfall in the western region of Ghana. A correlation analysis was performed on the chosen parameters to get the best combination, which significantly influenced rainfall. Validations were carried out first on each parameter, Precipitation, Relative Humidity, MSLP, and Temperature; Secondly, on Precipitation (R) against Relative Humidity (H), Pressure (M), and temperature; Thirdly on (R) against (T)/(M), (R) against (T)/(H), (R) against (M)/(H). Finally, the testing on the (R) against (T)/(M)/(H) produced the best MSE result of 0.002 with a MAE result of 0.021.

A rainfall prediction model has been developed by Hetty Rohayani's team [19] utilizing the Feed Forward Neural Network (FFNN) and Genetic Algorithm (GA). Researchers used the backpropagation (BP) algorithm to train FFNNs to update their weights in a prior work. A limitation of the BP algorithm was discovered based on previous research. For basic training datasets, the BP algorithm produces good results; but, when used to large, complex datasets, the ANN-BP model's performance deteriorates and it is more prone to become trapped in a local minimum. According to some study, the BP technique should be replaced by evolutionary algorithms to update the NN's weights in order to get over this limitation. Thus in this study, researchers used an evolutionary method called the genetic algorithm (GA). In the proposed rainfall prediction model, they have substituted BP with GA as a learning method for FFNN to obtain the best weight to minimize errors and produce the desired output. The model's performance has been assessed using the Root Mean Square Error (RMSE).

Research Methodology: The proposed rainfall prediction model was developed using an LSTM neural network coupled with a genetic algorithm. The model's goal is to predict rainfall based on 20 years of historical rainfall data. The Keras API framework is used in the model's implementation. The performance of the model is evaluated using the Categorical Cross-entropy loss function. The proposed DLGA model approach entails gathering raw data and preprocessing it. A LSTM neural network is used for prediction by utilizing a genetic algorithm to update the weights. The loss function evaluates the model's performance.

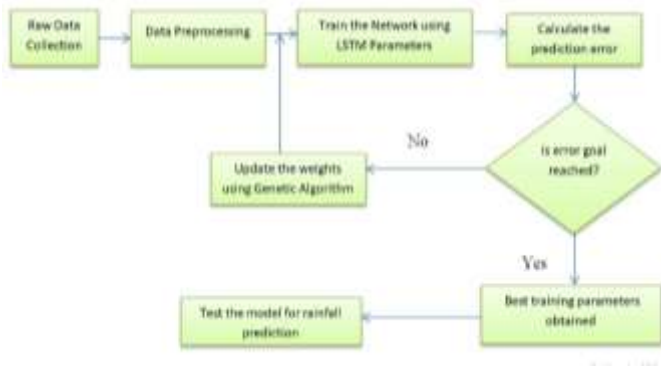


Fig – 2: Proposed hybrid rainfall prediction model

Navsari District's raw data (7306 records) spanning 20 years were used in the study. A training dataset and a testing dataset were created from the data, which included 70% (5114) records from 2001 to 2014 and 30% (2192 records) from 2015 to 2020. The eight input attributes that the proposed model employs are described in Table 1. Types of rainfall that come within the class [20] shown in Table 2 are the output attribute.

Table -1: Input Attributes

Sr. no	Name	Type	Description	Masure. Unit
1.	Tmax	Numerical	Maximum Temperature	°C
2.	TMin	Numerical	Minimum Temperature	°C
3.	RH	Numerical	Relative Humidity	%
4.	WS	Numerical	Wind speed	km/h
5.	Evp	Numerical	Evaporation	mm
6.	VP	Numerical	Vapor Pressure	%
7.	BSSH	Numerical	Bright Sunshine hour	No. of hours
8.	Rf	Numerical	Rainfall	mm

Table -2: Output Attributes Classes

Sr. no.	Class Type	Rainfall Range (in mm)
1.	No Rain	0
2.	Light Rain	0.1 to 7.5
3.	Moderate Rain	7.6 to 64.4
4.	Heavy Rain	64.5 to 244.4
5.	Extremely Heavy Rain	Above 244.4

The data is cleaned during data preparation by substituting the mean value of the input parameters for the missing values [6]. To reduce errors, the data is scaled and normalized between values [0-1] after being cleaned [2]. The data was normalized using the min-max normalization method. The proposed rainfall prediction model is developed using the Keras API framework. Prediction models use LSTM neural networks, which are made up of input layers, output layers, completely connected hidden layers, and activation functions.

The input layer of the prediction model is formed by feeding all of the input variables specified in Table 1. The value of the output class attributes that were obtained from the output layer is displayed in Table 2. The output layer uses the Softmax activation function, whereas the input and hidden layers employ the ReLU activation function. The accuracy of the model is evaluated using the Categorical Cross-entropy loss function. A table 3 introduces the parameters of LSTM neural network and GA used to build the prediction model.

Table -3: LSTM neural network Parameters

Name	Description	Value
Input Parameter	The number of input parameters utilized by the model.	8
Iteration	The number of rotations or generation implied by the model to change the weight of the network while training.	300
Loss function	To determine the loss/error of the proposed prediction model.	Categorical cross entropy
Activation Function	It helps model to introduce and learned the data's non-linearity at each layer.	ReLU (input & hidden layer) Softmax (Output Layer)
Initial population	Randomly generate the initial population	Weights of the network layers
Fitness function	It defines how near our solution is to reaching the objective.	---
Selection of Parent	The method to select parents for mating to produce children for the new generation	Rank method
Crossover	The method used to merge the parental genetic characteristics for the new children	Single point
Mutation	The method add unique genetic features to the children that it different from its parents	Random resetting
Output Parameter	The number of output parameters utilized by the model.	1

Experimental Result

This section shows the experimental findings of the proposed rainfall prediction model. The model uses 7306 records, combining a training dataset of 5114 records and a testing dataset of 2192 records. Furthermore, the training dataset is once more divided into 20% validation records (1023 records) and 80% training records (4091 records). Table 3 displays the parameters that were utilized in the model calibrating process. Accuracy in training and validation is crucial for avoiding model overfitting. Figure 3 makes it very evident that with the each iteration, the difference between training and testing accuracy is smaller. Therefore, it is suggested that there be no chance of overfitting in the proposed model.

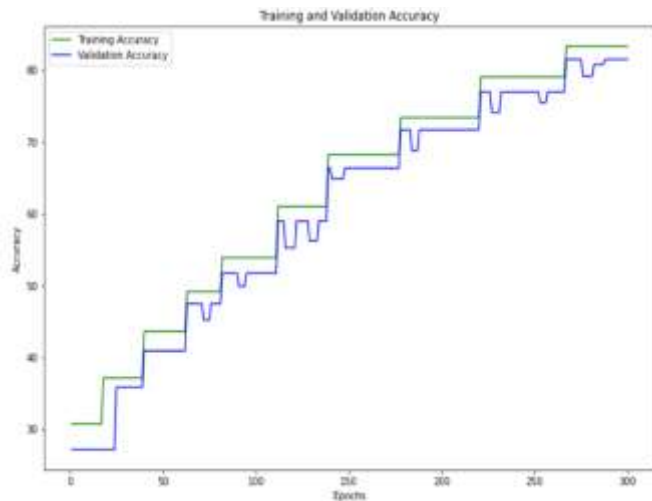


Fig – 3: Iteration wise training and validation accuracy

Table -1: True and false prediction records of the model

	Total Records	True Prediction	False Prediction	Accuracy
Training	4091	3409	682	83.33 %
Testing	2192	1795	397	81.89%

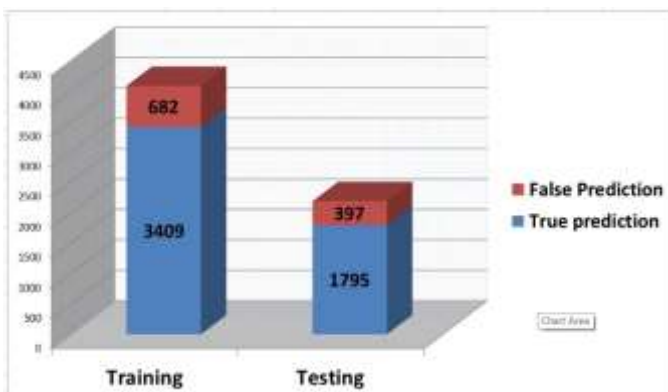


Fig - 4: Prediction records of the proposed model

After training and validation performed effectively, the model was tested using the testing dataset. The model's performance has been evaluated using the accuracy performance metrics. Table 4 above demonstrates that during the training phase, the proposed model generated 3409 accurate predictions and 682 incorrect predictions out of 4091 records, yielding an accuracy rate of 83.33%. While in the testing phase, the model generated 1795 accurate predictions and 397 wrong predictions from 2192 records, yielding an accuracy rating of 81.89%. A graphic representation of it is provided in Figure 4 above.

Conclusion: In this research, the proposed rainfall prediction model has developed based on 20 years of historical data of rainfall. The model is built by the Keras API framework using a genetic algorithm and an LSTM neural network. This

prediction aids in preventing unpredictable natural disasters like drought and floods and supports farmers in making important agricultural decisions. According to the experimental findings, models that are evaluated based on accuracy achieve a performance of about 83.33 percent. The study comes to the conclusion that the proposed model is ideally appropriate for rainfall prediction and has a respectable degree of accuracy.

References:

1. Nicholls, N, Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation, Springer Natural Hazards, Vol-23, 2001, pp. 137–155.
2. Sourabh Nema, M.K Awasthi and R K Nema, Trend Analysis of Annual and Seasonal Rainfall in Tawa command Area, International Journal of Environment, Agriculture and Biotechnology (IJEAB), Vol-1, Issue-4, Nov-Dec-2016, pp. 952-957.
3. Ganesh P. Gaikwad and Prof. V. B. Nikam, Different Rainfall Prediction Models And General Data Mining Rainfall, International Journal of Engineering Research & Technology (IJERT), Vol-2 Issue-7, July 2013, pp. 115-123.
4. H. M. Abd El Baki, H. Fujimaki, I. Tokumoto, and T. Saito, A new scheme to optimize irrigation depth using a numerical model of crop response to irrigation and quantitative weather forecasts, Computers and Electronics in Agriculture, Elsevier, Vol-150, July 2018, pp. 387-393.
5. S. Renuga Devi, P. Arulmozhiarman, C. Venkatesh and Pranay Agarwal, Performance Comparison of Artificial Neural Network Models for Daily Rainfall Prediction, International Journal of Automation and Computing, Springer, 2015, Pp. 417–427.
6. Vimal B. Patel and R. D. Morena, Use of Appropriate Loss Function in Rainfall Prediction using Deep Learning, International Journal of Engineering and Advanced Technology (IJEAT), Vol-9 Issue-6, August 2020, pp. 293-297.
7. Caihong Hu, Qiang Wu, Hui Li, Shengqi Jian, Nan Li and Zhengzheng Lou, Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation, Water, Vol-10 Issue-11, October 2018, pp. 1-16.
8. Thomas Fischer and Christopher Krauss, Deep learning with long short-term memory networks for financial market predictions, European Journal of Operational Research, Elsevier, Vol-270, Issue-2, October 2018, pp. 654-669.
9. Vimal B. Patel and R. D. Morena, Trend Analysis And Prediction Of Rainfall Using Genetic Algorithm, Shodh Shanchar Bulletin, Vol-10, Issue-40, October-December 2020, pp.130-139.

10. Chayanika Sharma, Sangeeta Sabharwal and Ritu Sibal, A Survey on Software Testing Techniques using Genetic Algorithm, International Journal of Computer Science, Vol-10, Issue-1, January 2013, pp. 381-393.
11. Fateme Moslehi, Abdorrahman Haeri and Francisco Martinez-Alvarez, A novel hybrid GA-PSO framework for mining quantitative association rules, Soft Computing 24, Springer, July 2019, pp. 4655-4666.
12. Giang Nguyen, Stefan Dlugolinsky, Martin Bobak, Viet Tran, Alvaro Lopez Garcia, Ignacio Heredia, Peter Malik and Ladislav Hluchy, Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey, Springer, Artificial Intelligence Review, vol. 52, January 2019 pp. 77-124.
13. Hang Zhao, Orazio Gallo, Iuri Frosio and Jan Kautz, Loss Functions for Image Restoration with Neural Networks, IEEE Transactions on Computational Imaging, Vol-3, Issue-1, March 2017, pp. 47-57.
14. Yangfan Zhou, Xin Wang, Mingchuan Zhang, Junlong Zhu, Ruijuan Zheng And Qingtao W, MPCE: A Maximum Probability Based Cross Entropy Loss Function for Neural Network Classification, IEEE Access, Vol-7, October 2019, pp. 146331-146341.
15. Mohini P. Darji, Vipul K. Dabhi and H. B.Prajapati, Rainfall Forecasting Using Neural Network: A Survey, International Conference on Advances in Computer Engineering and Applications, IEEE, July 2015, pp. 706-713.
16. Longzhang Fang and Dongguo Shao, Application of Long Short-Term Memory (LSTM) on the Prediction of Rainfall-Runoff in Karst Area, Frontiers in Physics, Vol-9, January 2022, pp. 1-9.
17. Anjali Samad, Bhagyanidhi, Vaibhav Gautam, Piyush Jain, Sangeeta and Kanishka Sarkar, An Approach for Rainfall Prediction Using Long Short Term Memory Neural Network, 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), October 2020, pp. 190-195.
18. Charity Oseiwah Adjei, Wei Tian, Bernard-Marie Onzo, Emmanuel Adu Gyamfi Kedjanyi, Oscar Famous Darteh and Shiwei Chen, Rainfall Forecasting in Sub-Sahara Africa-Ghana using LSTM Deep Learning Approach, International Journal of Engineering Research & Technology, Vol. 10 Issue 03, March-2021, pp. 4664-470.
19. Hetty Rohayani, Tuga Mauritsius, Leslie H Spit Warnars Harco and Edi Abdurrachman, Evaluation Performance Neural Network Genetic Algorithm, Advances in Intelligent Systems Research, Vol-172, May 2020, pp. 426-431.
20. <https://www.imdpune.gov.in/Weather/Reports/glossary.pdf>, last accessed 30-08-2026.