Rainfall Prediction Using Deep Learning Techniques

Vighnesh Cilamkoti, Govreddy Jaya Prakash Reddy, Akula Tarun Sai, Gaddam Bharath Reddy
Department of Information Technology, Guru Nanak Institutions Technical Campus, Hyderabad, India.

Abstract: Predicting rainfall accurately is crucial for various sectors such as agriculture, water resource management, disaster preparedness, and climate research. Traditional methods of rainfall prediction often rely on numerical weather models, which may have limitations in accuracy and efficiency. In recent years, deep learning techniques have emerged as promising tools for improving the accuracy of rainfall prediction. This paper provides an overview of the application of deep learning techniques in rainfall prediction, discussing various methodologies, challenges, and future directions in this field. Predicting rainfall is the most challenging assignment in meteorology. We have developed a rainfall prediction model in our work that may be readily estimated via the use of LSTM and artificial intelligence approaches. This rainfall calculator is an advanced approach. For the application of this kind of strategy and its accuracy findings, the deep learning approach is most beneficial. The memory sequence data measurement process uses a long short-term memory method, which computes historical data quickly and produces the best forecast. Since agriculture is the main source of income for the majority of the population in this nation, this prediction method is essential. Crop yields will rise and agricultural expenses will decrease with prompt rainfall assessment. Our model, which will assist us in estimating the quantity of rainfall, was developed taking into account all of these variables. To achieve this, we have gathered information from six areas. We have used six factors in our prediction: temperature, dew point, humidity, wind pressure, wind direction, and wind speed. Our task was completed with 76% accuracy after all of our data was analysed. For improved results, we also concentrate on a large dataset on long-term weather.

Keywords: long short-term memory; predictive analytics; rainfall prediction; recurrent neural network

I.INTRODUCTION

Rainfall prediction plays a significant role in various aspects of human life, including agriculture, hydrology, and disaster management. Accurate rainfall prediction allows for better planning and decision-making in these sectors, leading to improved resource allocation and risk mitigation strategies. Traditional methods of rainfall prediction involve statistical models and numerical weather prediction models, which have limitations in accuracy, especially for short-term and localized predictions. Deep learning techniques offer a promising alternative for enhancing the accuracy of rainfall prediction by leveraging the power of neural networks to extract complex patterns from large datasets.

Rainfall prediction is vital for various sectors such as agriculture, water resource management, and disaster preparedness. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), have shown promise in capturing temporal dependencies and patterns in sequential data, making them suitable for rainfall prediction tasks. This paper explores the application of LSTM networks in rainfall prediction, discussing data preprocessing, model architecture, training strategies, and evaluation metrics. Additionally, it examines challenges and future directions in leveraging LSTM networks for accurate and reliable rainfall prediction.

Rainfall prediction is crucial for numerous applications, including agriculture, hydrology, and disaster management. Accurate rainfall forecasting enables proactive planning and decision-making, leading to improved resource allocation and risk mitigation strategies. LSTM networks, with their ability to capture long-term dependencies in sequential data, have emerged as effective tools for rainfall prediction tasks. LSTM networks consist of memory cells and gating mechanisms that control the flow of information over time. This architecture enables LSTMs to capture long-term dependencies in sequential data while mitigating the vanishing gradient problem encountered in traditional RNNs.

Accurate rainfall prediction is crucial for various sectors such as agriculture, water resource management, and disaster preparedness. Neural networks (NNs) have emerged as powerful tools for modeling complex relationships in data, making them promising for rainfall prediction tasks. This paper provides an overview of the application of neural networks in rainfall prediction, discussing different types of NNs, data preprocessing techniques, model architecture, training strategies, evaluation metrics, and challenges. Additionally, it explores potential applications and future directions in leveraging NNs for improved rainfall prediction. Rainfall prediction plays a vital role in decision-making processes across multiple domains. Neural networks offer a flexible framework for capturing nonlinear relationships in rainfall data, enabling accurate prediction at various spatial and temporal scales. This paper explores the application of neural networks in rainfall prediction and discusses methodologies to enhance prediction accuracy and reliability.

II.RELATED WORKS

In the modern world, rainfall is thought to be one of the few causes accountable for the majority of noteworthy events worldwide. Since agriculture in India is entirely reliant on rainfall, it is said to be one of the key determinants of the nation's economy.

Separately Therefore, it is important to ascertain the quantity of rainfall in coastal locations worldwide. Rainfall forecasts should be made in advance in order to build rainwater harvesters in certain water-scarce places. The goal of this research is to use neural networks and machine learning to forecast rainfall. The research does a comparison analysis of neural network and machine learning techniques, and then presents the most effective method for rainfall prediction. Initially, preprocessing is carried out. Preprocessing involves displaying the information as a variety of graphs, including bar graphs and histograms. LSTM regression is utilised in machine learning, whereas the ANN (Artificial Neural Network) technique is employed in neural networks. Following computation, the accuracy of LSTM and ANN was compared, and conclusions were drawn based on the sorts of mistakes. Using the more accurate technique, the forecast has been made in order to simplify the system. The dataset, which includes rainfall data for various locations of the nation from 1901 to 2015, was used to make the projection. It includes monthly rainfall data for the same as well as yearly rainfall data.

Predicting the amount of rainfall is now a crucial component of most water conservation schemes in the nation. The intricacy of rainfall data is one of the main obstacles. These days, the majority of rainfall prediction systems are unable to identify any non-linear patterns or hidden layers inside the system. This effort will help identify all of the non-linear patterns and hidden layers, which is helpful for making accurate rainfall predictions [1]. The tool used to forecast rainfall in a certain area is called rainfall prediction. There are two ways to do it. The first step is to examine the physical rule that governs rainfall; the second is to create a system that will identify oblique patterns or characteristics that influence the physical elements and the method by which it is accomplished. The second approach is superior as it can be used to complicated and non-linear data and does not need any kind of mathematical computations [2]. Owing to the existence of a system that is inaccurate in identifying nonlinear patterns and hidden layers, predictions are often incorrect, which may result in significant losses. Thus, the primary goal of this study is to design a system that can identify complexity and hidden layers in order to provide accurate and precise predictions that will aid in the development of the nation's economy and agricultural sector [3].

The first is known as hybridization, which denotes the use of many machine learning techniques in tandem with prediction. The second one focuses on raising the quality of the dataset that is being used.

The term "decomposition technique" refers to a method used to improve the quality of the dataset. The last strategy involves utilising an add-on optimizer to increase the algorithm's accuracy, while the third strategy involves employing a combination of methods to increase the algorithm's ability.

This system's capacity to improve algorithm and dataset quality is one of its main benefits. The forecast will be more accurate the more efficiently the data were utilised to make it. The algorithm's accuracy is no different. The error detection method of this system is among its main drawbacks. There are several other faults that are not computed in this method but may have a detrimental impact on the correctness of an algorithm, such as mean squared error (MSME), mean absolute error (MAE), etc. Merely R2 (R-Squared) and RMSE (Root Mean Square Error) are assessed in order to determine the algorithm's accuracy. Therefore, if the number of faults influencing an algorithm is increased, the method's correctness may be examined more thoroughly.

It is claimed that the neural network technique is superior to all other machine learning approaches and more successful due to its ability to identify any nonlinear pattern existing in a system by evaluating a small number of accuracy levels [4].

The use of data mining techniques aids in the discovery of hidden patterns that aid in accurate rainfall prediction. This method predicts future rainfall by taking into account all the variables that impact it, such wind speed and climate. Rainfall is predicted using a customised, integrated, and updated data mining approach. Rainfall predictions are based on a wide range of climatic factors. Polarity, weather, wind, highest and lowest temperatures, and other variables are being recorded. It states that there is no guarantee that the forecast will be more accurate when the maximum parameters are used. For prediction, both supervised and unsupervised methods are used. This report states that other nations, including India, Australia, Columbia, Indonesia, Malaysia, and others, attempted the forecast. The area where the forecast is to be made, the climatic parameters that were used as predictors, and the historical weather data that was used to train the algorithm were the main aspects that impacted the outcome.

While some researchers in this field are more interested in determining the association between weather characteristics, other researchers place more emphasis on the data that is used to train the algorithm. Data mining techniques can identify all of the hidden patterns that are there and have the potential to impact rainfall predictions. It is necessary to combine and optimise this approach so that all predictions are more error-free. Therefore, the remaining work will be to improve, optimise, and integrate this data mining technique in a way that will solve all current issues related to uncovering hidden patterns and, concurrently, determine an appropriate correlation between weather factors [5]. Three steps are included in the deep learning process.

The first step is to identify the most appropriate algorithm. The method that performs best with the provided dataset is chosen based on the dataset that is currently available. The task of determining which model best fits the algorithm is

handled in the second step. A few metrics items are being used for threat score, false alarm ratio, and quantitative precipitation forecasting in the final stage.
The first two phases of the computation had negative numbers since it was based on the original expected time series. This method has the advantage of accurately identifying all nonlinear patterns and calculating correlation coefficients because it uses original predicted data for calculation; however, this method occasionally takes a while to compute and produces inaccurate results [6]. ConvLSTM, a brand-new LSTM (long short term memory) extension, is being considered.

One kind of RNN is the LSTM (Recurrent neural network). One of RNN's main benefits is that it can accurately anticipate data stored in short-term memory; but, if the data is stored in long-term memory, it approaches long short-term memory (LSTM). RNN's efficiency decreases with increasing gap length. This study work's main benefit is its capacity to retain and retrieve data that has been held for extended periods of time, but it also demonstrates the neural network's incapacity to make predictions based on that data. To recover data that has been kept for a long time, RNN requires LSTM or ConvLSTM. This procedure addresses both obtaining the data from the data source and putting it to good use [7].

III. PROPSED WORK

The more accurate method is predicted by the proposed system to include rainfall. The collection of data is gathered. There are two methods for forecasting precipitation. LSTM regression is a component of the machine learning technique, which is the first. The neural network method is the second. This system uses the best algorithm to provide results after first comparing the two processes. The steps involved in the suggested system include data input, preprocessing, data splitting, training, testing, dataset comparison, identifying the best algorithm, prediction using the more accurate algorithm, and final output.

The primary motivation for not using both algorithms for prediction is to simplify the system as a whole. To this end, the system first determines which of the machine learning and neural network algorithms is the most accurate, then makes predictions using that method. You will get the results in the form of Excel sheets and graphs. All of the results for the preprocess will be provided as various graphs, and the accuracy for the machine learning and neural network will be provided as metrics and an excel sheet, with the predicted value correspondingly being provided as a sheet with two columns: ID and predicted value. The IDs will match those found in the datasheet. IDs need to match the IDs in the dataset in order to determine which area the forecast is for.

The User Interface (UI) module is the first one. Python is used in UI coding. There are five buttons on it: Quit, LSTM, Preprocess, Clear, Browse, and Neural Network.To choose a dataset from the system, utilise the Browse button.To remove the chosen dataset, click the Clear button. Preprocessing is used for visualisation, LSTM and neural networks are used for prediction, and stop is used to end an application.

Any procedure that ends will show a dialogue window with the phrase "Process successfully completed." Python is used to write all of the UI code. It contains several code snippets that link to every other Python coded file. Every other module in the system is linked to this one. The visualisation module is linked upon selecting the preprocess button; the

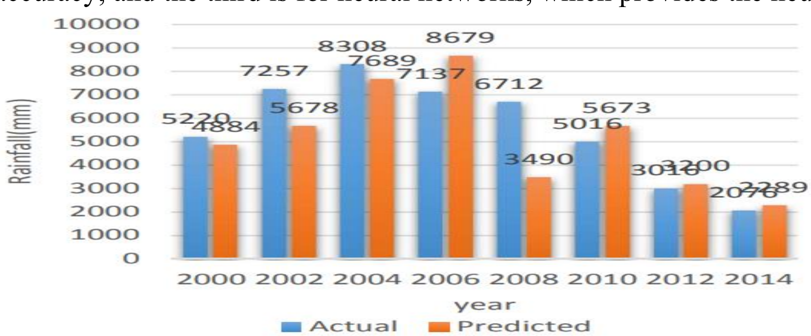
LSTM module is connected upon selecting the LSTM button; and the neural network module is connected upon selecting the Neural Network button. Thus, this module is related to all the other modules in this manner.

Visualisation is the second module. This procedure may be completed prior to making the forecast. It is a graph-based representation of the dataset. It makes the comparing process easier. Following the preprocess button click, this process is carried out, and many graphs are produced.

Even before the neural network and machine learning algorithms are executed, they first pass through this module before moving on to their own. In such scenario, no graphs are constructed since doing so would introduce duplication, which would complicate the process more. Correlation metrics, scatter metrics, maximum value for month-by-month, subdivision mean value for month-by-month, sum of every quarter of subdivision, sum of every quarterly, sum of year by year, and sum of month year by year are the graphs that appear after selecting the preprocess button. LSTM is the third module.

IV.RESULTS AND DISCUSSION

The dataset is entered into the system by the user. There will also be three more buttons available. The first one is for preprocessing, which uses graphs to describe the dataset; the second is for LSTM, which provides the LSTM regression accuracy; and the third is for neural networks, which provides the neural network accuracy.



Preprocessing should be carried out first in order to improve comparability and get a deeper grasp of the dataset. Preprocessing may be carried out before to making the forecast. It is a graph-based representation of the dataset. It facilitates the comparison process and improves comprehension of the available dataset at the same time. The dataset need to be divided into two sections: the first should be used to train the algorithm, and the second should be used to forecast the quantity of rainfall. Only the algorithm that predicts rainfall with greater accuracy is used. The utilised algorithm has to be trained before it can make predictions. Thus, the training has been completed in this section of the system. This is carried out in both LSTM and ANN methods. This stage provides a clear understanding of which of the two algorithms is more accurate. Then, rainfall prediction is performed using the leftover dataset (which was not utilised for training).

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

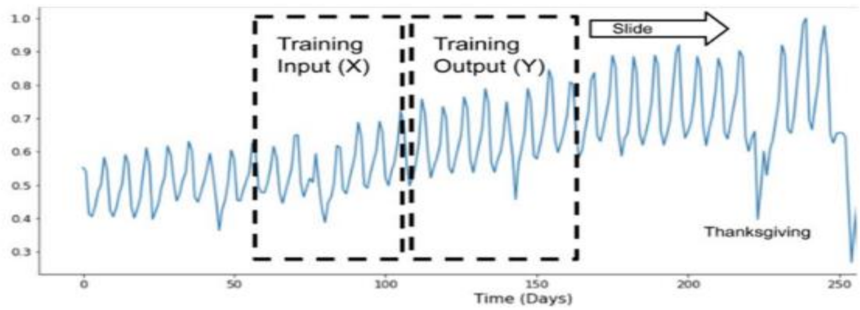
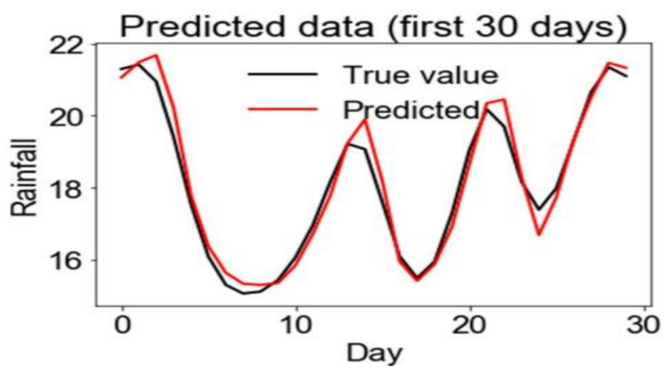


Figure 1. Time Frequency Architecture for Rainfall.

Both methods also complete this section. Following the completion of all steps, a graph and table displaying the anticipated rainfall and algorithm accuracy are provided as the final results. Following preprocessing, the graphs that are obtained include those for correlation, scatter, and Maximum value for each month, Subdivision mean value for each month, Sum of each quarterly, Sum of each year, and Sum of month for each year are all examples of cumulative sums. Following that, metrics and an excel sheet including the accuracy are obtained for both the neural network and the LSTM.



Different kinds of mistakes are shown in metrics in addition to accuracy, and the same is exhibited in the excel sheet. Finally, the anticipated value is obtained and saved in an Excel file.

V.CONCLUSION

As one of the few factors contributing to India's greatest GDP, rainfall need to be the main worry for the majority of us. Since it cannot foresee the hidden layers that are there and are yet to be identified as being capable of completing the accurate prediction, the existing method for predicting rainfall fails in the majority of complicated scenarios. In an attempt to develop a reliable method for rainfall prediction, two approaches are contrasted. There are two approaches: the first uses machine learning, while the second uses artificial neural networks. The LSTM regression technique is used in the first. Effective comparison benefits from the visualisation of the dataset before the comparison is carried out. Test data is used to make predictions, while train data is used to train the algorithm. Along with accuracy, error categories including MSE, MAE, R-SQUARED, and RSME were taken into consideration while comparing these two procedures. The most accurate one was taken into account, and forecast was made using that method specifically. As part of the data being used to train the algorithm, the rainfall was forecasted from the testing data. The system concludes that the LSTM regression process is more accurate than the artificial neural network method after doing the comparison. We learned from the comparison that ANN has an accuracy of 77% and LSTM has an accuracy of around 94%. As a result, LSTM is the most effective analytical approach for forecasting the amount of rainfall in any given area. This project's next improvements may take the form of a strategy to lower the proportion of mistakes that are made. Reducing the ratio of train data to test data will also be one of the main improvements, helping to improve the degree of prediction within the constraints of time and complexity. Further testing of the algorithm's accuracy may be done as complexity rises. Any of the aforementioned methods' correctness may be tested by computing a wide range of additional error kinds. Therefore, an algorithm for assessing daily datasets rather than accumulated datasets may be crucial for future study.

REFERENCES

1. Hernández, E.; Sanchez-Anguix, V.; Julian, V.; Palanca, J.; Duque, N. Rainfall prediction: A deep learning approach. In International Conference on Hybrid Artificial Intelligence Systems; Springer: Cham, Switzerland, 2016; pp. 151–162.
2. Goswami, B.N. The challenge of weather prediction. Resonance 1996, 1, 8–17.
3. Nayak, D.R.; Mahapatra, A.; Mishra, P. A survey on rainfall prediction using artificial neural network. Int. J. Comput. Appl. 2013, 72, 16.
4. Ravindra Changala, "Brain Tumor Detection and Classification Using Deep Learning Models on MRI Scans", EAI Endorsed Transactions on Pervasive Health and Technology, Volume 10, March 2024.
5. Ravindra Changala, "Sentiment Analysis in Social Media Using Deep Learning Techniques", International Journal of Intelligent Systems and Applications in Engineering, 2024, 12(3), 1588–1597.
6. Ravindra Changala, "Integration of IoT and DNN Model to Support the Precision Crop", International Journal of Intelligent Systems and Applications in Engineering, Vol.12 No.16S (2024).
7. Ravindra Changala, "UI/UX Design for Online Learning Approach by Predictive Student Experience", 7th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2023 - Proceedings, 2023, pp. 794–799, IEEE Xplore.
8. Kashiwao, T.; Nakayama, K.; Ando, S.; Ikeda, K.; Lee, M.; Bahadori, A. A neural network-based local rainfall prediction system using meteorological data on the internet: A case study using data from the Japan meteorological agency. Appl. Soft Comput. 2017, 56, 317–330.
9. Mislan, H.; Hardwinarto, S.; Sumaryono, M.A. Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan, Indonesia. Procedia Comput. Sci. 2015, 59, 142–151.
10. Ravindra Changala, "Optimization of Irrigation and Herbicides Using Artificial Intelligence in Agriculture", International Journal of Intelligent Systems and Applications in Engineering, 2023, 11(3), pp. 503–518.
11. Ravindra Changala, Development of Predictive Model for Medical Domains to Predict Chronic Diseases (Diabetes) Using Machine Learning Algorithms And Classification Techniques, ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue 6, 2019.
12. Ravindra Changala, "Evaluation and Analysis of Discovered Patterns Using Pattern Classification Methods in Text Mining" in ARPN Journal of Engineering and Applied Sciences, Volume 13, Issue 11, Pages 3706-3717 with ISSN:1819-6608 in June 2018.
13. Muka, Z.; Maraj, E.; Kuka, S. Rainfall prediction using fuzzy logic. Int. J. Innov. Sci. Eng. Technol. 2017, 4, 1–5.
14. Jimoh, R.G.; Olagunju, M.; Folorunso, I.O.; Asiribo, M.A. Modeling rainfall prediction using fuzzy logic. Int. J. Innov. Res. Comput. Commun. Eng. 2013, 1, 929–936.

15. Wu, J.; Liu, H.; Wei, G.; Song, T.; Zhang, C.; Zhou, H. Flash flood forecasting using support vector regression model in a small mountainous catchment. Water 2019, 11, 1327.
16. Ravindra Changala "A Survey on Development of Pattern Evolving Model for Discovery of Patterns in Text Mining Using Data Mining Techniques" in Journal of Theoretical and Applied Information Technology, August 2017. Vol.95. No.16, ISSN: 1817-3195, pp.3974-3987.
17. Poornima, S.; Pushpalatha, M.; Sujit Shankar, J. Analysis of weather data using forecasting algorithms. In Computational Intelligence: Theories, Applications and Future Directions—Volume I. Advances in Intelligent Systems and Computing; Verma, N., Ghosh, A., Eds.; Springer: Singapore, 2019; Volume 798.
18. Ravindra Changala, Framework for Virtualized Network Functions (VNFs) in Cloud of Things Based on Network Traffic Services, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Volume 11, Issue 11s, August 2023.
19. Ravindra Changala, Block Chain and Machine Learning Models to Evaluate Faults in the Smart Manufacturing System, International Journal of Scientific Research in Science and Technology, Volume 10, Issue 5, ISSN: 2395-6011, Page Number 247-255, September-October-2023.
20. Ravindra Changala, AIML and Remote Sensing System Developing the Marketing Strategy of Organic Food by Choosing Healthy Food, International Journal of Scientific Research in Engineering and Management (IJSREM), Volume 07 Issue 09, ISSN: 2582-3930, September 2023.
21. Manideep, K.; Sekar, K.R. Rainfall prediction using different methods of Holt winters algorithm: A big data approach. Int. J. Pure Appl. Math. 2018, 119, 379–386.
22. Gundalia, M.J.; Dholakia, M.B. Prediction of maximum/minimum temperatures using Holt winters method with excel spread sheet for Junagadh region. Int. J. Eng. Res. Technol. 2012, 1, 1–8.

23. Ravindra Changala, Development of CNN Model to Avoid Food Spoiling Level, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN: 2456-3307, Volume 9, Issue 5, September-October-2023, Page Number 261-268.
24. Ravindra Changala, A Novel Prediction Model to Analyze Evolutionary Trends and Patterns in Forecasting of Crime Data Using Data Mining and Big Data Analytics, Mukt Shabd Journal, Volume XI, Issue X, October 2022, ISSN NO: 2347-3150.