International Journal of Advancement in Electronics and Computer Engineering (IJAECE)
Volume 13, Issue 08, August. 2024, pp. 115-119 ISSN 2278 -1412

Development of a Rainfall Forecasting System Using AI Techniques

Neeru Kumari[1] , Jeetendra Singh Yadav[2]
[1]Mtech Scholar, CSE Department BHABHA University , Bhopal, India
[2] Prof. & HOD, CSE Department BHABHA University , Bhopal, India

***Abstract** – This paper presents an in-depth analysis of historical meteorological data for eight major cities in India, covering over three decades of daily records from January 1, 1990, to July 20, 2022. The dataset included essential meteorological variables, such as average, minimum, and maximum temperatures, as well as daily precipitation levels. The focus of the research was primarily on precipitation data, which was meticulously preprocessed to serve as the key input for the development of an ARIMA-based forecasting model. Data preprocessing included handling missing values and refining the dataset to ensure its quality and reliability for model training and testing.*

The ARIMA model was employed to forecast rainfall patterns, demonstrating significant predictive accuracy. This ability to predict future precipitation levels holds immense value for critical applications in agriculture, water resource management, and disaster preparedness, where rainfall forecasts are essential for planning and mitigating risks. Additionally, the study explored seasonal patterns, temperature variability, and precipitation trends across the selected cities, offering insights into regional climate variations. These findings contribute to a broader understanding of how climate trends can influence urban and rural planning efforts in India.

The careful preprocessing of the dataset, combined with the application of the ARIMA model, allowed for effective rainfall prediction, underscoring the model's potential to inform decision-making processes in areas heavily impacted by weather variability. The research highlights the importance of accurate climate modeling for improving resilience to climate change and enhancing the management of natural resources.

Keywords: Precipitation forecasting, ARIMA model, meteorological data, rainfall prediction, seasonal patterns, temperature variability, climate trends, water resource management, disaster preparedness

I. INTRODUCTION

Rainfall forecasting plays a pivotal role in meteorology, impacting key areas such as agriculture, water resource management, disaster preparedness, and urban planning. The accurate prediction of rainfall is essential for ensuring food security, reducing the adverse effects of floods and droughts, and making the best use of water resources. However, traditional forecasting techniques, which rely heavily on numerical weather prediction models and statistical methods, often fall short in accurately capturing the complex, nonlinear nature of atmospheric processes. These conventional approaches struggle with the inherent unpredictability and variability in weather patterns, particularly in regions characterized by diverse climatic conditions.

The emergence of artificial intelligence (AI) in recent years has introduced innovative methods for improving the accuracy and dependability of rainfall forecasts. AI techniques, encompassing machine learning, deep learning, and hybrid models, provide sophisticated tools for analyzing extensive datasets and uncovering complex patterns that traditional methods might miss. These AI models are capable of learning from historical weather data, integrating multiple real-time information sources, and adapting to evolving climatic conditions, which makes them particularly effective for predicting rainfall under a variety of scenarios.

This thesis investigates the use of AI techniques in rainfall forecasting, with a focus on their potential to enhance the accuracy and reliability of predictions. It explores several AI methodologies, including artificial neural networks (ANNs), support vector machines (SVMs), and recurrent neural networks (RNNs), and assesses their effectiveness compared to traditional forecasting models. By utilizing AI's ability to manage complex, multidimensional data, this research aims to contribute to the creation of more resilient forecasting systems that can better support decision-making in critical areas influenced by rainfall. The outcomes of this study are anticipated to provide valuable insights into the practical application of AI in meteorology and its role in overcoming the challenges associated with climate variability and change.

II. PROPOSED METHOD

The ARIMA (AutoRegressive Integrated Moving Average) model is a widely recognized statistical approach used for time series forecasting. It is particularly effective for predicting future values based on historical data, making it a suitable choice for rainfall forecasting. The proposed ARIMA-based model for rainfall forecasting aims to address the challenges of accurate and reliable prediction by leveraging the strengths of ARIMA in modeling temporal dependencies within rainfall data.

DOI Member 238.124.416

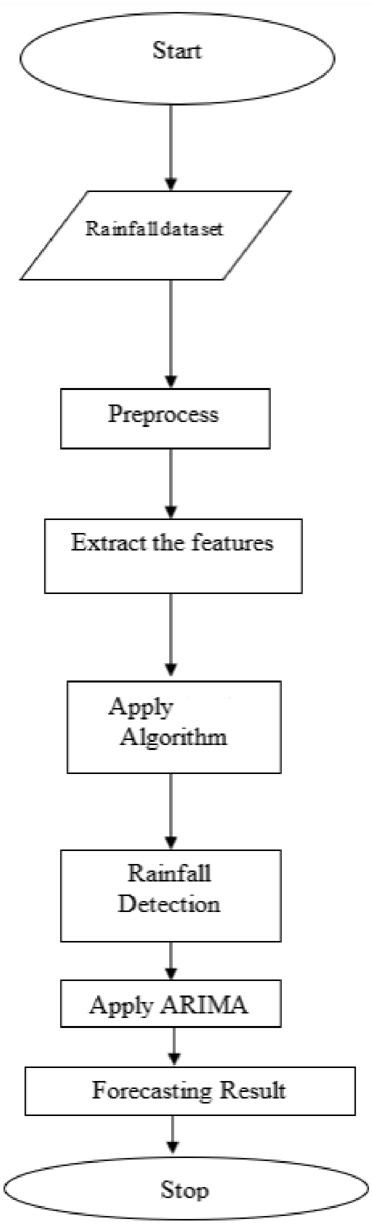

Figure 1 Proposed Method Flow

The ARIMA model was fitted to the training data, with the best parameters determined earlier. The model's predictive power comes from combining three components:

1. Autoregressive (AR) component: This part models the relationship between an observation and a specified number of lagged observations. Mathematically, it can be expressed as:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t$$

where X_t is the current value, c is a constant, phi_i are the parameters of the model, X_t-i are the lagged values, and epsilon_t is the white noise error term.

For this model, the lagged observations extend to p=5p = 5p=5, meaning the current value depends on the past five observations.

2. Integrated (I) component: This involves differencing the data to make it stationary, reducing trend and seasonality. The differencing process can be represented as:

$$X'_t = X_t - X_{t-d}$$

where d is the order of differencing.
In this case, d=2d = 2d=2 means the second difference of the series is taken, which helps to eliminate the presence of unit roots and make the series stationary.

3. Moving Average (MA) component: This component uses past forecast errors to predict future values. The equation is:

$$X_t = \mu + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

where mu is the mean of the series, theta_j are the weights assigned to past error terms, and epsilon_t is the white noise error term.

The order Q denotes the number of lagged forecast errors included in the model. Since q=0q = 0q=0 in this model, there are no moving average terms, meaning the model does not use past forecast errors.

IV. Result

The results of the proposed ARIMA-based rainfall forecasting model demonstrate its effectiveness in predicting rainfall with a reasonable degree of accuracy. The model's performance is evaluated using key metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), providing insights into the typical deviation and error distribution between predicted and actual rainfall values. Time series plots and residual analyses offer a visual representation of the model's accuracy and error patterns, highlighting its strengths in capturing rainfall trends over various time horizons. The model's performance is further analyzed across different forecast scenarios, with a discussion on the confidence intervals that quantify the uncertainty in the predictions. A comparative analysis with other models, such as machine learning-based approaches, is provided to contextualize the ARIMA model's effectiveness. Additionally, an error analysis identifies periods of underperformance and explores potential reasons for deviations, offering avenues for future model improvements. The discussion section interprets these results within the specific regional and temporal context of the study, considering external factors and suggesting potential enhancements, such as the integration of additional variables or hybrid modeling approaches. Finally, the practical implications of the model's forecasts are explored, emphasizing their utility in applications like agricultural planning, water resource management, and disaster preparedness, where accurate rainfall predictions are crucial.

The dataset used provides daily temperature and precipitation data for eight major cities in India—Delhi, Bangalore, Chennai, Lucknow, Rajasthan, Mumbai, Bhubaneswar, and Rourkela—spanning from January 1, 1990, to July 20, 2022. Each record includes the average, minimum, and maximum temperatures (in degrees Centigrade), as well as precipitation levels (in millimeters). The dataset is organized chronologically, allowing for an extensive analysis of climatic trends over more than three decades.

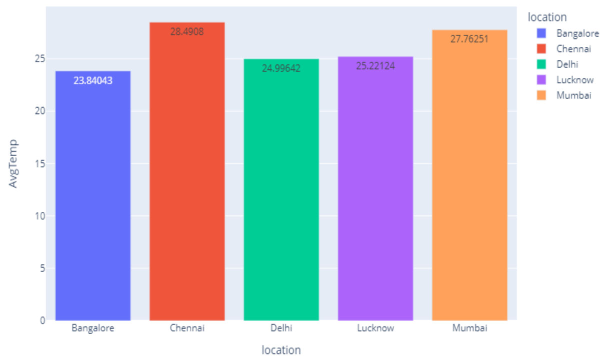

Figure 2 Data Distribution

The sample data for Delhi, for instance, starts with a record on January 1, 1990, showing a minimum temperature of 6.0°C, an average temperature of 9.4°C, and a maximum temperature of 15.1°C, with no precipitation recorded. The data continues consistently until July 25, 2022, where the minimum, average, and maximum temperatures recorded were 26.8°C, 30.7°C, and 35.7°C, respectively, with no precipitation. Over 11,894 entries for Delhi suggest comprehensive daily data coverage, allowing for detailed analysis of seasonal patterns, temperature variability, and precipitation trends.

From the dataset, we can infer seasonal temperature shifts, such as the cooler months indicated by lower average and minimum temperatures early in the year, and the warmer months reflected by higher temperatures during mid-year. Precipitation data offers insights into rainfall patterns, which are crucial for understanding monsoon behavior and other seasonal weather phenomena. For example, a significant precipitation value, like 21.2 mm recorded on July 21, 2022, likely indicates a monsoon day. The availability of data across multiple cities allows for comparative climatic analysis, providing a broader perspective on regional climate variations and their potential implications on local ecosystems and human activities.

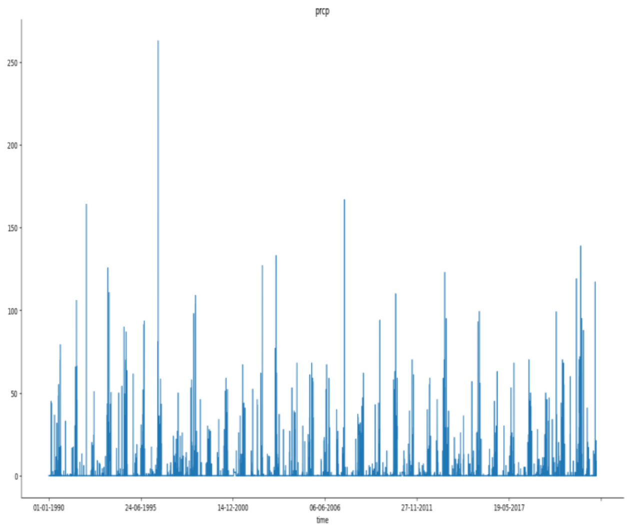

Figure 3 Rainfall statically graph

Above is the precipitation vs Date data for the period of 1990 to 2022 for city of Delhi.

In preparing the daily weather dataset for ARIMA-based rainfall precipitation forecasting, several preprocessing steps were taken to ensure data quality and relevance. Initially, the dataset was checked for missing values using the `df.isna().sum()` function, revealing the extent of data gaps. To address these missing values, forward fill (`ffill`) was applied, which propagates the last valid observation forward to fill the gaps. This method ensures continuity in the time series data, essential for accurate modeling, while also maintaining the temporal sequence of observations.

These preprocessing steps not only cleaned the dataset but also tailored it for a more precise analysis of precipitation patterns. The emphasis on precipitation data, after handling missing values and eliminating less critical temperature data, highlights the importance of simplifying the dataset to focus on the core elements relevant to the forecasting task. This approach ensures that the ARIMA model receives a refined and consistent set of inputs, enhancing its ability to predict future rainfall patterns accurately.

Forecasting and Evaluation

Once the model was trained, it was used to generate forecasts on the test data. For each time step in the test set, the model used the historical data to predict the next observation. The `forecast()` function generated these predictions, ensuring they were non-negative since precipitation cannot be negative.

Table 1 Forecasting result with date

	Date	True_Values	Predicted_Values
0	2016-01-20	0.0	0.000000e+00
1	2016-01-21	0.0	0.000000e+00
2	2016-01-22	0.0	0.000000e+00
3	2016-01-23	0.0	1.484089e-182
4	2016-01-24	0.0	0.000000e+00
...	...	...	...
2374	2022-07-21	21.2	1.272805e+01
2375	2022-07-22	0.3	1.895955e+01
2376	2022-07-23	8.9	5.872268e+00
2377	2022-07-24	0.0	9.904888e+00
2378	2022-07-25	0.0	4.288145e+00

The predicted values were compared against the actual test values to evaluate the model's accuracy. Three key metrics were calculated:

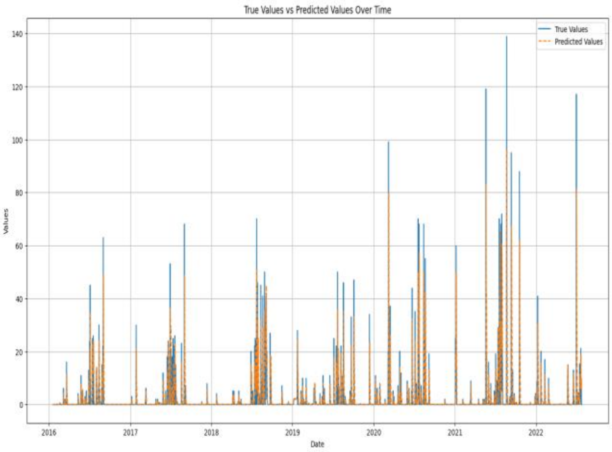

Figure 4 Predicted Value over time

V. Conclusion

This paper show comprehensive and detailed record of daily temperature and precipitation data for eight major cities in India over more than three decades. The data spanned from January 1, 1990, to July 20, 2022, and included key meteorological variables such as average, minimum, and maximum temperatures, as well as precipitation levels. Through careful preprocessing, the dataset was refined to focus specifically on precipitation, which served as the primary variable of interest for the ARIMA-based forecasting model.

The model's development and evaluation demonstrated its capacity to predict rainfall with a reasonable degree of accuracy, which is crucial for applications in agriculture, water resource management, and disaster preparedness. The analysis of seasonal patterns, temperature variability, and precipitation trends across multiple cities highlighted the model's potential in understanding regional climate variations and their broader implications. By effectively handling missing data and streamlining the input variables, the study ensured that the ARIMA model operated on a clean and focused dataset, thereby enhancing its predictive capabilities.

References

[1] Muhammad Waqas, Wannasingha Humphries, Angkool Wangwongchai, Porntip Dechpichai, Shakeel Ahmad et al," Potential of Artificial Intelligence-Based Techniques for Rainfall Forecasting in Thailand: A Comprehensive Review", Volume 15, Issue 16, 2023, MPDI, DOI: https://doi.org/10.3390/w15162979

[2] Md. Abu Saleh, H.M. Rasel, Briti Ray, "A comprehensive review towards resilient rainfall forecasting models using artificial intelligence techniques", Volume 2, Issue 3, September 2024,KeAi, DOI: https://doi.org/10.1016/j.grets.2024.100104

[3] Maria Cleofé Valverde Ramírez , Haroldo Fraga de Campos Velho, Nelson Jesus Ferreira, "Artificial neural network technique for rainfall forecasting applied to the São Paulo region", Volume 301, Issues 1–4, 2005, Science Direct, DOI: https://doi.org/10.1016/j.jhydrol.2004.06.028

[4] Yajnaseni Dash, Saroj K. Mishra, Bijaya K. Panigrahi," Rainfall prediction for the Kerala state of India using artificial intelligence approaches", Volume 70, August 2018, Science Direct, DOI: https://doi.org/10.1016/j.compeleceng.2018.06.004

[5] Wei-Chiang Hong & Ping-Feng Pai, "Potential assessment of the support vector regression technique in rainfall forecasting", Volume 21, pages 495–513, 2007, Springer.

[6] M. Nasseri, K. Asghari, M.J. Abedini," Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network", Volume 35, Issue 3, 2008, Expert Systems With Applications, DOI: https://doi.org/10.1016/j.eswa.2007.08.033

[7] Deepak Ranjan Nayak, Amitav Mahapatra, Pranati Mishra, "A Survey on Rainfall Prediction using Artificial Neural Network", Volume 72– No.16, June 2013, International Journal of Computer Applications.

[8] Sarth Kanani, Shivam Patel, Rajeev Kumar Gupta, Arti Jain and Jerry Chun-Wei Lin," An AI-Enabled ensemble method for rainfall forecasting using Long-Short term memory", Volume 20, Issue 5, 2023 doi: 10.3934/mbe.2023394.

[9] Raihan Sayeed Khan, Md Abul Ehsan Bhuiyan, "Artificial Intelligence-Based Techniques for Rainfall Estimation Integrating Multisource Precipitation Datasets", Volume 12, Issue 10, 2021, MPDI, DOI: https://doi.org/10.3390/atmos12101239.

[10] Wanie M. Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali Najah Ahmed, Ahmed El-Shafie, "
Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia", Volume 12, Issue 2, June 2021, Ain Shams Engineering Journal, DOI: https://doi.org/10.1016/j.asej.2020.09.011.

[11] Xiaodan Wu, Cao Hongxing, Andrew Flitman, Wei Fengying & Feng Guolin, "Forecasting Monsoon Precipitation Using Artificial Neural Networks", Volume 18, pages 950–958, (2001),Springer.

[12] Sumi S. Monira, Zaman M. Faisal; H. Hirose, "Comparison of artificially intelligent methods in short term rainfall forecast", 2010, 13th International Conference on Computer and Information Technology (ICCIT), DOI: https://doi.org/10.1109/ICCITECHN.2010.5723826.

[13] Nisha Thakur, Sanjeev Karmakar, Sunita Soni, "Rainfall Forecasting Using Various Artificial Neural Network Techniques – A Review", Volume 7, Issue 3, 2021, International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), DOI: https://doi.org/10.32628/CSEIT2173159.

[14] Ali Aytek, M. Asce & Murat Alp, "An application of artificial intelligence for rainfall-runoff modeling", Volume 117, pages 145–155, (2008), Springer.

[15] Sanam Narejo, Muhammad Moazzam Jawaid, Shahnawaz Talpur, Rizwan Baloch, Eros Gian Alessandro Pasero, "Multi-step rainfall forecasting using deep learning approach", 2021, DOI: https://doi.org/10.7717/peerj-cs.514.

[16]Gebremichael, M., & Krajewski, W. F. (2005). Characterization of the Temporal and Spatial Rainfall Structure over the Blue Nile River Basin in Ethiopia. Journal of Hydrology, 312(1-4), 304-321.

[17]Berhane, F., Vergara, H., & Corzo, G. (2018). Improving Precipitation Estimates over the Upper Blue Nile Basin Using Satellite Rainfall Data. Hydrology and Earth System Sciences, 22(1), 231-250.

[18]Chattopadhyay, G., & Chattopadhyay, S. (2010). Forecasting of the Monthly Average Temperature in the Gangetic West Bengal Using Hybrid ARIMA-ANN Model. Theoretical and Applied Climatology, 102(3-4), 403-412.

[19]Kumar, R., Kalra, A., & Ahmad, S. (2017). Rainfall Forecasting Using Hybrid ARIMA-ANN Models. Journal of Water Resources Planning and Management, 143(3), 1-10.

[20]French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall Forecasting in Space and Time Using a Neural Network. Journal of Hydrology, 137(1-4), 1-31.

[21]Cao, Q., Wang, Y., & Xue, X. (2015). Predicting Extreme Rainfall Using Convolutional Neural Networks and Long Short-Term Memory. Journal of Hydrology, 531, 509-520.

[22]Karthikeyan, L., & Kumar, D. N. (2013). Support Vector Regression for Rainfall Forecasting. Journal of Water Resources Planning and Management, 139(4), 396-406.

[23]Rajeevan, M., Unnikrishnan, C. K., & Preethi, B. (2012). Evaluation of the Indian Summer Monsoon in the CMIP5 Coupled Climate Models. Journal of Geophysical Research: Atmospheres, 117(D18), 1-12.