DOI:10.26524/krj.2024.2

Kongunadu Research Journal 11 (1): 9-14, 2024
Publisher: Kongunadu Arts and Science College, Coimbatore.

https://krjournal.com/index.php/krj/article/view/641

REVIEW ARTICLE

Exploring rainfall prediction through regression models: A systematic literature review

Velmurugan, S.[1,*], Saravana Moorthy, R.[1], Subramanian, K.[2] and Angel, S.[3]

[1]Department of Computer Science, Kongunadu Arts and Science College, Coimbatore -641029, Tamil Nadu, India
[2]Department of IT and Analytics, Xavier Institute of Management and Entrepreneurship (XIME), Bangalore, India
[3]Department of Computer Science (SF), Avanishilingam University, Coimbatore, Tamil Nadu, India

ABSTRACT

This study aims to develop a comprehensive rainfall forecasting system by employing advanced regression models such as CatBoost, XGBoost, Random Forest, SVM, Decision Tree, among others. The primary objectives include identifying, gathering, and preprocessing meteorological and environmental data that influence rainfall patterns. Integration of state-of-the-art regression models is intended to enhance the accuracy of rainfall predictions. The evaluation of these models involves rigorous assessment using performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared. Additionally, a user-friendly interface is designed to facilitate the input of meteorological data, benefiting a wide range of users, including meteorological experts and the general public. The applicability of the system spans across various domains, including agriculture, water resource management, and disaster preparedness.

Keywords: Provide Rainfall prediction, Regression model, Meteorological data, Hyperparameter tuning, Regression analysis

1. INTRODUCTION

Accurate rainfall prediction is indispensable across various sectors, including agriculture, disaster management, and water resource planning. This research paper is dedicated to enhancing the precision of rainfall forecasts through the utilization of robust machine learning techniques, including CatBoost, XGBoost, Random Forest, Logistic Regression, and Support Vector Classifier. These models will be refined and integrated into a user-friendly interface, empowering stakeholders to leverage predictions for informed decision-making.

This initiative responds to the increasing demand for precise meteorological forecasts, particularly in regions where agriculture is economically pivotal, and rainfall variations carry significant socio-economic consequences. By harnessing established and interpretable regression models, our aim is to advance the accuracy of rainfall pattern forecasts, ultimately contributing to the mitigation of uncertainties in weather conditions.

Rainfall prediction presents a formidable challenge in climate forecasting. Machine learning techniques offer a means of unveiling hidden patterns within recorded weather data to forecast rainfall. Such predictions rely on weather-related parameters like temperature, pressure, humidity, and wind speed. Temperature, notably, holds immense importance across various applications, including environmental considerations, manufacturing processes, agriculture, and energy management.

Given that precipitation, wind speed, temperature, and humidity are numerical data, regression analysis will facilitate rainfall prediction. Regression analysis, a statistical method used to establish relationships between independent and dependent variables, is well-suited for forecasting and prediction tasks.

While traditional methods in rainfall prediction have predominantly relied on statistical techniques to examine relationships among geographical coordinates, precipitation, and atmospheric parameters such as humidity, pressure, wind speed, and temperature, the intricate and non-

*Correspondence: Velmurugan, S., Department of Computer Science, Kongunadu Arts and Science College (Autonomous), Coimbatore - 641029, Tamil Nadu, India. E.mail: svelmurugan_cs@kongunaducollege.ac.in

linear nature of rainfall patterns has posed formidable obstacles to achieving precise forecasts. Consequently, efforts have been made to address and mitigate this non-linearity by employing methodologies like Empirical Mode Decomposition, Wavelet analysis, and Singular Spectrum Analysis.

Accurate rainfall predictions can raise public awareness about potential risks associated with heavy rainfall, enabling communities to better prepare with emergency kits, evacuation plans, and staying updated on weather forecasts. Water resource management authorities can optimize water allocation for irrigation and domestic use, enhancing the efficient utilization of water resources, particularly in areas with water scarcity.

2. LITERATURE SURVEY

Rainfall prediction system based on machine learning techniques, specifically utilizing 'SoftMax' or multinomial logistic regression. The objective is to predict rainfall categories, wind speed, and temperature using historical weather data from eleven cities in Myanmar. The system categorizes rainfall into one of five classes: Light Rainfall (LR), Moderate Rainfall (MR), Heavy Rainfall (HR), Violet Rainfall (VR), and No Rain (NR). Weather forecasts are essential for society and sustainable development, especially for farmers to optimize agricultural activities.

The system's performance is evaluated using various metrics, including accuracy, precision, recall, F-measure, mean absolute error (MAE), and root mean squared error (RMSE). The system is tested on a dataset containing 1826 days (from 2018 to 2022) for eleven cities. The results show an average accuracy of 83%, with varying performance for different rainfall categories. MAE and RMSE values are also reported to assess the error in predictions. Rainfall prediction is crucial for various applications, and the paper proposes a machine learning-based system for this purpose. 'SoftMax' logistic regression is used to predict rainfall categories based on historical weather data. The system achieves an accuracy of 83% in the experimental results. Future work may involve further improving the system and comparing it with other machine learning techniques [1].

The Research focuses on the development of a rainfall forecasting model for the Jimma region in southwest Oromia, Ethiopia. Rainfall is a critical factor for managing reservoir water levels, especially in agriculture. The experimental results demonstrate the superiority of the proposed approach over various baseline models, including LSTM, bidirectional LSTM (BiLSTM), gated recurrent unit (GRU), and convolutional LSTM (ConvLSTM). The proposed approach achieves the best performance in terms of RMSE and R2, confirming its effectiveness in rainfall prediction. Statistical analyses, such as one-way analysis of variance (ANOVA) and the Wilcoxon signed-rank test, support the significance and stability of the proposed approach. The research contributes to deep learning approaches for rainfall forecasting, offering valuable insights for applications like smart farming. The study's dataset spans several decades and is drawn from Jimma, Ethiopia, with a focus on meteorological parameters [2].

The systematic literature review (SLR) on rainfall prediction, which is a challenging task in weather forecasting with significant implications for various sectors. Accurate and timely rainfall prediction is crucial for activities such as construction projects, transportation, agriculture, flight operations, and flood management. The study aims to provide a critical analysis and review of data mining techniques used for rainfall prediction, with a specific focus on papers published from 2013 to 2017. The research process is systematically outlined, starting with the identification of research questions. These questions serve as the basis for the SLR, aiming to find answers through a critical review. The questions span various aspects of rainfall prediction, including the techniques used, performance evaluation, data types, geographical locations, factors affecting results, and the latest trends in the domain. The study acknowledges its limitations, including the possibility of missing relevant work and the potential for authors' evaluations to introduce bias.

Despite these limitations, the systematic literature review provides a comprehensive overview of the state of rainfall prediction using data mining techniques from 2013 to 2017 and highlights the need for continued enhancements and optimizations in this domain [3]. The use of short-term rainfall data for rainwater harvesting system (RHS) modeling and the identification of representative time series lengths (RTSL) in 12 cities across various climatic zones. RHS has gained importance in mitigating urban water scarcity and flooding issues, making accurate modeling crucial. Short-term rainfall data with inadequate lengths can introduce significant errors in RHS modeling.

The study reveals that RTSL is not significantly correlated with mean annual rainfall or seasonality index but is significantly correlated with the variation coefficients of annual rainfall. The partial correlation coefficient between RTSL and the variation coefficients of annual rainfall is strong (0.878), while the partial correlation coefficients between RTSL and mean annual rainfall and seasonality index are negative (-0.569 and -0.522, respectively). These findings demonstrate the feasibility of using short-term rainfall data with sufficient length instead of long-term data for RHS modeling, offering valuable insights into the variability of RTSL in diverse climatic zones [4].

"Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques" focuses on providing insights into climate and its significance for various businesses, particularly agriculture. It highlights the importance of predicting meteorological parameters, with a specific focus on precipitation. The accuracy of various forecasting models is assessed by comparing the results with ground truth data. Rainfall prediction is crucial for agriculture in India, which heavily relies on precipitation as a water source. It references various studies related to rainfall prediction using machine learning techniques. K. Chowdari's work analyzes monthly climatic changes and seasonal rainfall variability using data from the Indian Meteorological Department and rain gauge stations. L. Ingsrisawang conducts a comparative study on rainfall prediction in the northeastern part of Thailand using different machine learning techniques, such as ANN, SVM, and KNN, with input parameters like temperature, humidity, pressure, and wind speed.

S.N. Kohail uses historical data for the Gaza city to predict and classify temperature, considering attributes like relative humidity, wind speed, and rainfall. Another study by Petre in 2008 uses decision tree algorithms for weather prediction using meteorological data from Hong Kong. In summary, it provides a comprehensive review of various studies related to rainfall prediction and proposes a fusion model for improved accuracy in rainfall forecasting. It emphasizes the importance of understanding the relationships between meteorological parameters and rainfall for more accurate predictions [5].

The study conducted an investigation into rainfall prediction in Lahore city using various data mining techniques. Rainfall prediction is vital for various applications, and given the challenges posed by climate variations, accurate forecasting is crucial. The research aimed to explore the effectiveness of different data mining techniques in predicting rainfall in Lahore city. The techniques employed in the study included Support Vector Machine (SVM), Naïve Bayes (NB), k Nearest Neighbor (kNN), Decision Tree (J48), and Multilayer Perceptron (MLP).

The findings presented in this study demonstrate that data mining techniques, when applied to the task of rainfall prediction in Lahore city, performed well for the "no-rain" class. However, their performance was notably lower when predicting the "rain" class. F-measure, which provides an average of precision and recall, is used as a key accuracy measure in this research. The limitations and potential reasons for lower performance in predicting the "rain" class were explored. This study underscores the significance of accounting for local climatic characteristics and data quality in improving the accuracy of rainfall prediction models, especially in areas with variable and the limited rainfall patterns [6].

3. MATERIALS AND METHODS

The system architecture for a Rainfall Prediction system using regression algorithms comprises several interconnected components that collaborate to deliver accurate predictions. This architecture is designed to integrate data collection, preprocessing, model training, user interaction, and prediction visualization, ensuring reliable and [10] timely rainfall forecasts for diverse applications. The system is engineered to be flexible, scalable, and maintainable to address real-world requirements effectively.

The system commences by aggregating extensive meteorological data from diverse sources, including weather stations, satellites, and remote [7] sensors. This data encompasses crucial variables such as temperature, humidity, wind speed, and atmospheric pressure. Prior to inputting the data into regression models, a preprocessing module undertakes data cleansing and standardization to enhance its quality and consistency. The system architecture is depicted in Fig 1. At the core of the architecture lies a collection of regression models, including Random Forest, XGBoost, CatBoost, and others. Each model is tasked with learning intricate patterns and relationships within the preprocessed [15] data. Leveraging historical [9] weather data, these models are trained to make predictions concerning future rainfall patterns. Ensemble [8] techniques may also be employed to amalgamate the outputs of multiple models, thereby bolstering prediction accuracy.

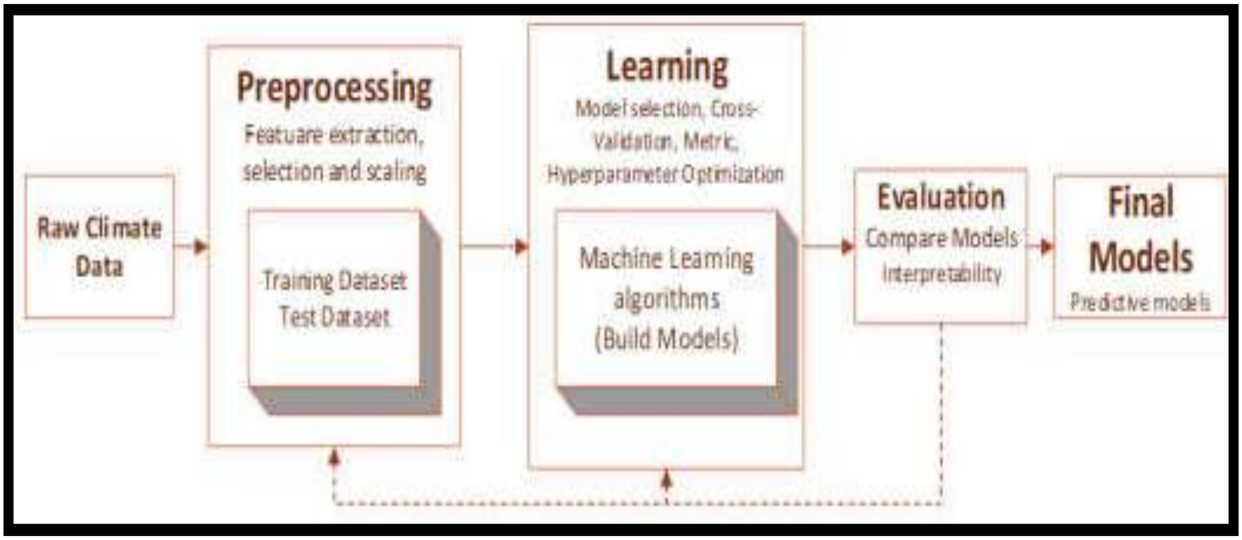

Figure 1. System Architecture

The system employs robust evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE) [14], and R-squared to gauge the efficacy [13] of the regression models. These metrics facilitate a comprehensive assessment of the models' performance, aiding in the refinement and optimization of prediction outcomes. The different kinds of input data used by the models to make predictions about rainfall, can be used to categorise the various kinds of rainfall forecasting studies. Input data used consideration are depicted in Fig. 2.

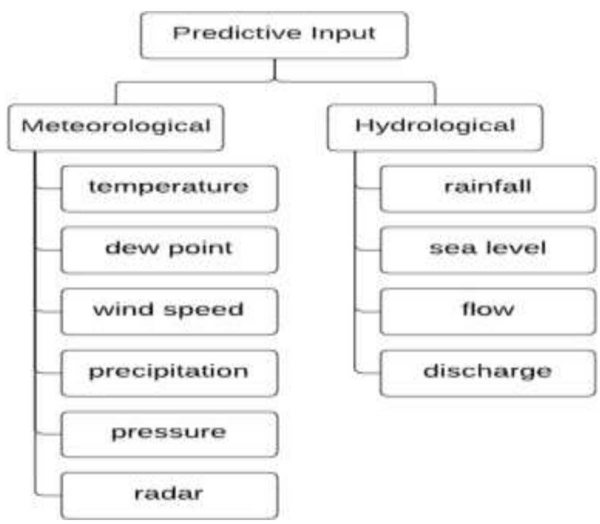

Figure 2. Types of predictive parameters.

Additionally, the system incorporates user interaction [12] features, enabling stakeholders to input specific parameters or preferences and visualize prediction results through intuitive interfaces. This interactive aspect enhances user engagement and facilitates informed decision-making across various sectors reliant on rainfall forecasts.

4. CONCLUSION

In conclusion, this research paper focuses on the development of a comprehensive rainfall prediction system that leverages advanced regression models, including CatBoost, XGBoost, Random Forest, SVM, Decision Tree, among others. The primary objectives of this project include the identification, acquisition, and preprocessing of meteorological and environmental data influencing rainfall patterns. By incorporating state-of-the-art regression models, the research aims to significantly enhance the accuracy of rainfall predictions.

The evaluation of these models employs well-established performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared. Moreover, the research emphasizes the importance of making this system easily accessible through a user-friendly interface, benefiting a diverse user base, including meteorological experts and the general public. The application of this system spans across various domains, such as agriculture, water resource management, and disaster preparedness. The literature survey within this research paper highlights the significance of accurate rainfall predictions for various sectors and demonstrates the value of machine learning and data mining techniques in addressing the challenges of forecasting rainfall. The integration of regression algorithms, coupled with the outlined system architecture, offers a promising approach to address the complexities of rainfall prediction and contribute to informed decision-making across a range of applications.

ACKNOWLEDGEMENT

The authors would like to express our gratitude to acknowledge the Management and Faculties of Computer Science Department Kongunadu Arts and Science College (Autonomous), Coimbatore, Tamil Nadu, India who helped us during the writing of this manuscript for supporting all the facility to complete this study.

REFERENCES

1. Soe, M.M. (2023). Rainfall Prediction using Regression Model. In 2023 IEEE Conference on Computer Applications (ICCA) (pp. 113-117). IEEE.
2. El-kenawy, E.S.M., Abdelhamid, A.A., Alrowais, F., Abotaleb, M., Ibrahim, A. and Khafaga, D.S. (2023). l-Biruni Based Optimization of Rainfall Forecasting in Ethiopia. Computer Systems Science & Engineering, 46(1).
3. Rahman, A., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M.A. and Mosavi, A. (2022). Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. Sensors. https://doi.org/10.3390/s22093504.
4. Salaeh, N., Ditthakit, P., Pinthong, S., Hasan, M.A., Islam, S., Mohammadi, B. and Linh, N.T.T. (2022). Long-Short Term Memory Technique for Monthly Rainfall Prediction in Thale Sap Songkhla River Basin, Thailand. Symmetry. https://doi.org/10.3390/sym14081599.
5. Perera, A., Mudannayake, S.D., Azamathulla, H.M. and Rathnayake, U. (2020). Recent climatic trends in Trinidad and Tobago, West Indies. Asia-Pacific Journal of Science and Technology, 1-11. https://so01.tci-thaijo.org/index.php/APST/article/view/205343.
6. Mehta, D.J. and Yadav, S.M. (2022). Long-term trend analysis of climate variables for arid and semi-arid regions of an Indian State Rajasthan. International Journal of Hydrology Science and Technology. https://doi.org/10.1504/IJHST.2022.120639.
7. Mohammed, M., Kolapalli, R., Golla, N. and Maturi, S.S. (2020). Prediction of rainfall using machine learning techniques. International Journal of Science and Technology Research, 9, 3236-3240.
8. Chaplot, B. (2021). Prediction of rainfall time series using soft computing techniques. Environmental Monitoring and Assessment. https://doi.org/10.1007/s10661-021-09388-1.
9. An-Vo, D.A., Mushtaq, S., Reardon-Smith, K., Kouadio, L., Attard, S., Cobon, D. and Stone, R. (2019). Value of seasonal forecasting for sugarcane farm irrigation planning. European Journal of Agronomy, 104, 37-48. https://doi.org/10.1016/j.eja.2019.01.005
10. Miller, S., Mishra, V., Ellenburg, W.L., Adams, E., Roberts, J., Limaye, A. and Griffin, R. (2021). Analysis of a short-term and a seasonal precipitation forecast over Kenya. Atmosphere. https://doi.org/10.3390/atmos12111371.

11. Frame, J.M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L.M., Gupta, H.V. and Nearing, G.S. (2022). Deep learning rainfall-runoff predictions of extreme events. Hydrology and Earth System Sciences, 26, 3377-3392. https://doi.org/10.5194/hess-26-3377-2022.
12. Aftab, S., Ahmad, M., Hameed, N., Salman, M., Ali, I. and Nawaz, Z. (2018). Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review. International Journal of Advanced Computer Science Applications, 9, 143–150.
13. Yue, T., Zhang, S., Zhang, J., Zhang, B. and Li, R. (2020), "Variation of representative rainfall time series length for rainwater harvesting modelling in different climatic zones. Journal of Environmental Management, 269, p.110731.
14. Shah, U., Garg, S., Sisodiya, N., Dube, N. and Sharma, S. (2018), "December. Rainfall prediction: Accuracy enhancement using machine learning and forecasting techniques. In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 776-782). IEEE.
15. Aftab, S., Ahmad, M., Hameed, N., Bashir, S., Ali, I. and Nawaz, Z. (2018). Rainfall Prediction in Lahore City using Data Mining Techniques. International Journal of Advanced Computer Science and Applications, 9.