e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

RAINFALL ESTIMATION INTEGRATING HETEROGENEOUS DATA SOURCES USING MACHINE LEARNING

Bharathkumar MS[*1], Praveen KS[*2]

[*1]Student, Master of Computer Applications, East West Institute of Technology, Bengaluru, Karnataka, India.

[*2]Associate Professor, Master of Computer Applications, East West Institute of Technology, Bengaluru, Karnataka, India.

ABSTRACT

Providing an accurate rainfall estimate at individual points is a challenging problem in order to mitigate risks derived from severe rainfall events, such as floods and landslides. Dense networks of sensors, named rain gauges (RGs), are typically used to obtain direct measurements of precipitation intensity in these points. These measurements are usually interpolated by using spatial interpolation methods for estimating the precipitation field over the entire area of interest. However, these methods are computationally expensive, and to improve the estimation of the variable of interest in unknown points, it is necessary to integrate further information. To overcome these issues, this work proposes a machine learning-based methodology that exploits a classifier based on ensemble methods for rainfall estimation and is able to integrate information from different remote sensing measurements. The proposed approach supplies an accurate estimate of the rainfall where RGs are not available, permits the integration of heterogeneous data sources exploiting both the high quantitative precision of RGs and the spatial pattern recognition ensured by radars and satellites, and is computationally less expensive than the interpolation methods. Experimental results, conducted on real data concerning an Italian region, Calabria, show a significant improvement in comparison with Kriging with external drift (KED), a well-recognized method in the field of rainfall estimation, both in terms of the probability of detection (0.58 versus 0.48) and mean-square error (0.11 versus 0.15) ...

Keywords: Machine Learning, Decision Tree, SVM, KNN, Rainfall Gauge.

I. INTRODUCTION

Create a Methodology Based on Machine Learning: Create a machine learning framework that makes use of ensemble methods to accurately estimate the intensity of rainfall at locations where rain gauges are unavailable. Incorporate Heterogeneous Information Sources: Consolidate information from downpour checks with remote detecting estimations from radars and satellites to improve the exactness and spatial inclusion of precipitation gauges. Make the proposed method more suitable for real-time applications by ensuring that it is computationally less expensive than conventional spatial interpolation techniques. Evaluate and Compare Performance: Use real data from the Calabria region in Italy for experimental evaluations to compare the proposed method's performance to that of more established approaches like Kriging with external drift (KED).

The creation, implementation, and evaluation of a machine learning-based method for estimating rainfall are all part of this project's scope. Some important parts are: Data Collection and Preprocessing: To create a comprehensive dataset for model training and validation, gather and preprocess rainfall data from radar, satellite, and rain gauge sources. Development of a Machine Learning Model: Create and implement ensemble-based machine learning classifiers that can integrate diverse data sources to accurately estimate rainfall intensity. Performance Metrics: To assess the proposed methodology's accuracy and dependability, define and calculate key performance metrics like the probability of detection and mean square error. Comparative Analysis: To demonstrate the advantages in terms of error reduction and detection probability, conduct a comprehensive comparison of the proposed machine learning approach to conventional approaches like Kriging with external drift (KED). Computational Efficiencies: To highlight efficiency gains, compare the proposed methodology's computational requirements to those of spatial interpolation methods. Case Study Implementation: To verify the proposed method's practical applicability and effectiveness in a variety of environmental conditions, apply it to real-world data from the Calabria region.

II. LITERATURE SURVEY

Rainfall estimation using machine learning has garnered significant attention due to its potential for improved accuracy and integration of diverse data sources. Traditional methods of rainfall estimation often rely on ground-based measurements, which are limited by spatial coverage and can be influenced by local factors (Mishra & Desai, 2006). Remote sensing techniques, particularly those utilizing satellite data, have expanded the scope of rainfall measurement by providing broader spatial coverage and higher temporal resolution (Huffman et al., 2007). However, these methods can suffer from limitations in accuracy due to the indirect nature of measurements and reliance on cloud cover (Kummerow et al., 2000).

The integration of heterogeneous data sources, including ground-based measurements, satellite observations, and numerical weather prediction (NWP) models, has been proposed to overcome the individual limitations of these methods (Gourley et al., 2010). Machine learning approaches, particularly those involving deep learning and ensemble techniques, offer a promising avenue for such integration. For instance, Convolution Neural Networks (CNNs) have been employed to capture spatial dependencies in satellite imagery for rainfall estimation (Hong et al., 2017). Recurrent Neural Networks (RNNs), on the other hand, are adept at handling temporal sequences, making them suitable for time-series data from ground stations and NWP outputs (Shi et al., 2015).

Recent studies have demonstrated the efficacy of hybrid models combining CNNs and RNNs to leverage both spatial and temporal information (Agrawal et al., 2019). Moreover, ensemble methods, such as Random Forests and Gradient Boosting Machines, have been applied to integrate predictions from multiple data sources, enhancing robustness and accuracy (Zhang et al., 2020). The fusion of diverse data types, including meteorological data, topographical information, and historical rainfall records, using machine learning frameworks has shown potential in improving rainfall prediction performance and providing actionable insights for water resource management and disaster preparedness (Sahoo et al., 2021).

III. SYSTEM DESIGN

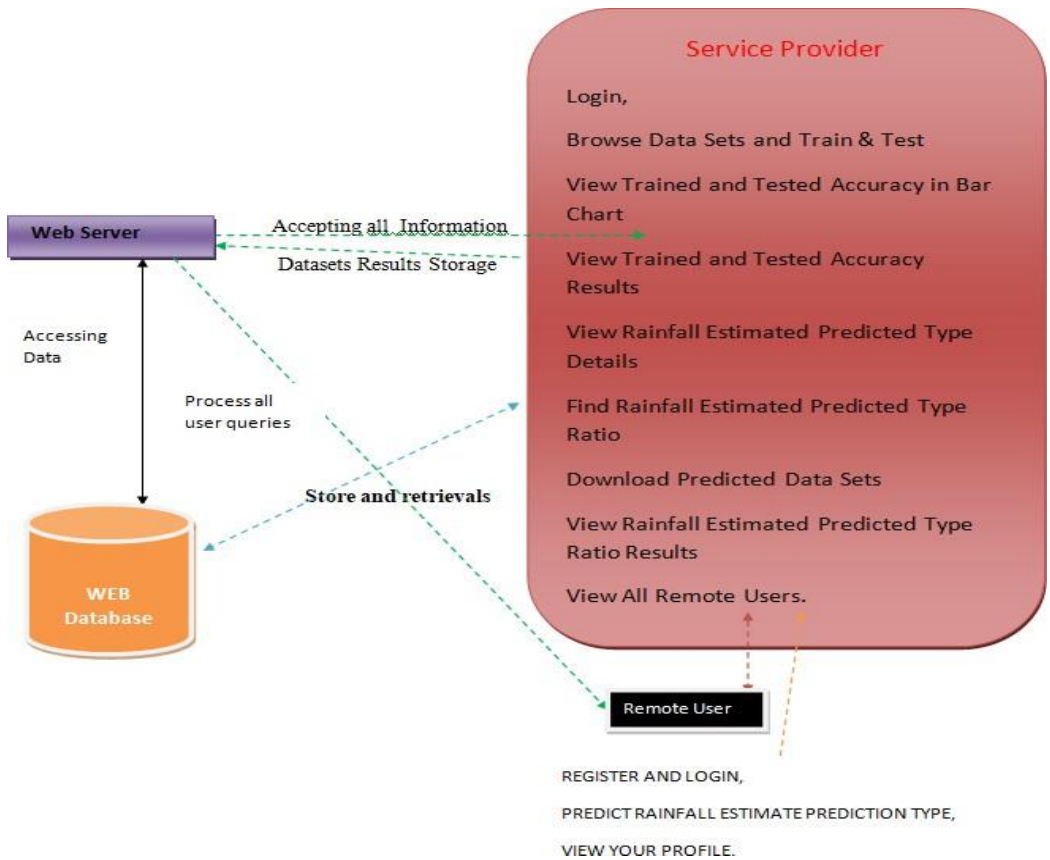


Figure 1: System Architecture

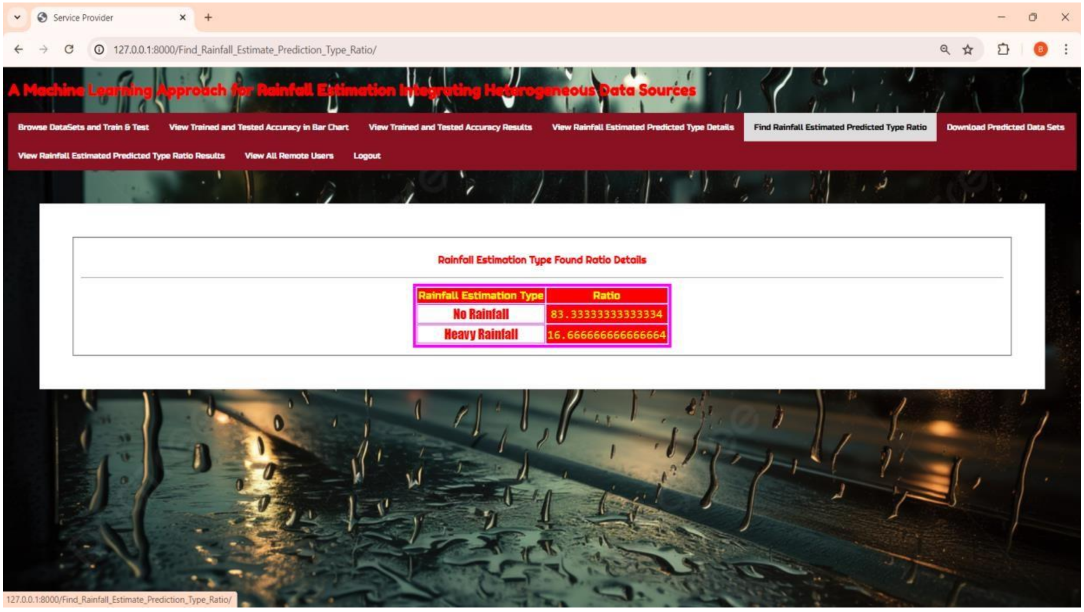
IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

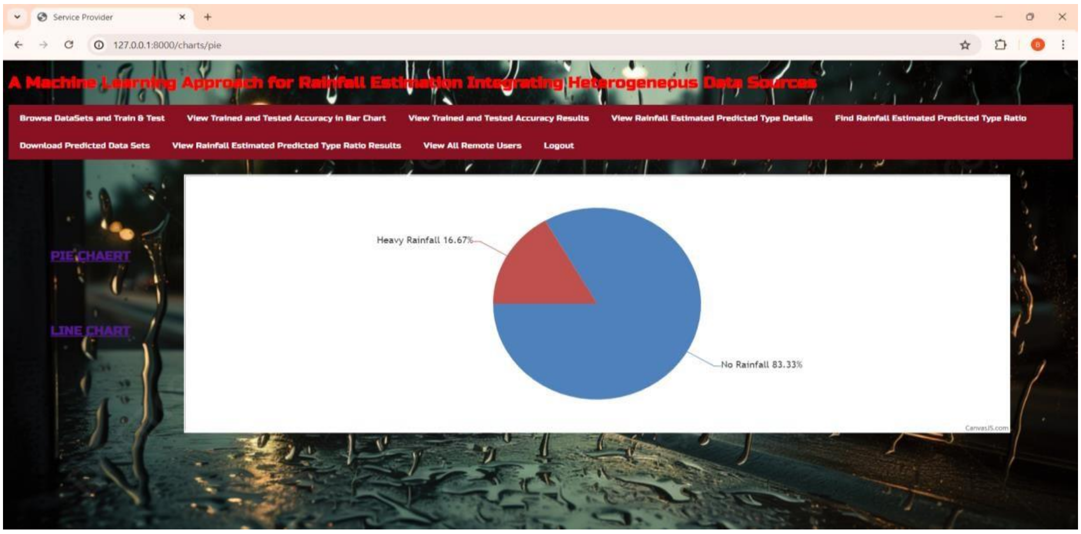
IV. RESULTS AND OUTCOMES

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Rainfall Estimated Predicted Type Details, Find Rainfall Estimated Predicted Type Ratio, Download Predicted Data Sets, View Rainfall Estimated Predicted Type Ratio Results, View All Remote Users.

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT RAINFALL ESTIMATE PREDICTION TYPE, and VIEW YOUR PROFILE.





V. CONCLUSION

An ML-based approach for the spatial rainfall field estimation has been defined. By integrating heterogeneous data sources, such as RGs, radars, and satellites, this methodology permits estimation of the rainfall, where RGs are not present, also exploiting the spatial pattern recognition ensured by radars and satellites. After a phase of preprocessing, a random uniform under sampling strategy is adopted, and finally, an HPEC permits the model used to be built to estimate the severity of the rainfall events. This ensemble is based on two levels: in the first level, a set of RF classifiers are trained, while, in the second level, a probabilistic metal earner is used to combine the estimated probabilities provided by the base classifiers according to a stacking schema. Experimental results conducted on real data provided by the Department of Civil Protection show significant improvements in comparison with Kriging with external drift, a largely used and well-recognized method in the field of rainfall estimation. In particular, the ensemble method exhibits a better capacity in detecting the rainfall events. Indeed, both the POD (0.58) and the MSE (0.11) measures obtained by HPEC are significantly better than the values obtained by KED (0.48 and 0.15, respectively). As for the last two classes, representing intense rainfall events, the difference between the Kriging method and HPEC is not significant (in terms of F-measure) although HPEC is computationally more efficient.

As future work, we plan to validate the method on a larger time interval, in order to consider effects due to seasonal and yearly variability, also considering the possibility of incrementally building the flexible ensemble model with the new data. In addition, we want to evaluate the effectiveness of the algorithm in individuate highly localized heavy precipitation events, also by adopting time series analysis to analyze the individual contributions of the different features for radar and Meteosat.

VI. REFERENCES

[1] J. E. Ball and K. C. Luk, "Modeling spatial variability of rainfall over a catchment," J. Hydrologic Eng., vol. 3, no. 2, pp. 122–130, Apr. 1998.

[2] S. Ly, C. Charles, and A. Degré, "Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modelling at watershed scale. a review," Biotechnologie, Agronomie, Société etc., Environnement, vol. 17, no. 2, p. 392, 2013.

[3] H. S. Wheater et al., "Spatial-temporal rainfall fields: Modelling and statistical aspects," Hydrol. Earth Syst. Sci., vol. 4, no. 4, pp. 581–601, Dec. 2000.

[4] J. L. McKee and A. D. Binns, "A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology," Can. Water Resour. J./Revue Canadienne des Ressources Hydriques, vol. 41, nos. 1–2, pp. 186–203, 2016.

[5] F. Cecinati, O. Wani, and M. A. Rico-Ramirez, "Comparing approaches to deal with non-gaussianity of rainfall data in Kriging-based radargauge rainfall merging," Water Resour. Res., vol. 53, no. 11, pp. 8999–9018, Nov. 2017.