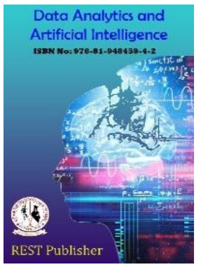
Data Analytics and Artificial Intelligence
Vol: 3(7), 2023
REST Publisher; ISBN: 978-81-948459-4-2
Website: http://restpublisher.com/book-series/daai/



Development of Explainable Ai (Xai) Based Model for Prediction of Heavy /High Impact Rain Events Using Satellite Data

Hareni Srikanth, *R. Gokulnath, Rajkumar Kalaimani
Adhiyamaan college of engineering, Hosur, Tamil Nadu, India.
*Corresponding Author Email: gokulnathramesh25@gmail.com

Abstract: *The idea proposes an integrated solution for now casting heavy precipitation rainfall events, leveraging INSAT-3D/3DR satellite data and advanced data science techniques. Employing Python libraries such as Pandas, NumPy, and Dask facilitates effective data manipulation, pre processing, and feature engineering. The predictive model development incorporates Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN), and Gradient Boosting models like XGBoost and Light GBM, enhanced by LIME and SHAP for Explainable AI (XAI). Hyper parameter tuning using Randomized Search CV, model evaluation through metrics such as RMSE and MAE, and web application deployment using Django and Plotly contribute to a comprehensive solution. Data security is ensured through SHA-512 encryption, while real-time monitoring is achieved using Grafana and CI/CD pipelines with Jenkins facilitate automated updates, highlighting the model's robustness and interpretability.*
Keywords: *Now casting heavy precipitation events, Satellite data, Data science, Python, Preprocessing, Feature engineering, Convolution neural networks, Recurrent neural networks, Gradient boosting models, Explainable artificial intelligence, Hyper parameter tuning, Web application, Django, Encryption, Real-time monitoring, Grafana, Jenkins.*

1. INTRODUCTION

The accurate prediction of heavy precipitation rainfall events is paramount for effective disaster preparedness and resource allocation. In this research paper, we propose a comprehensive solution for nowcasting, integrating cutting-edge technologies and methodologies. Leveraging INSAT-3D/3DR satellite data obtained from meteorological databases, our approach employs Python's powerful data manipulation libraries, including Pandas, NumPy, and Dask, for meticulous preprocessing and feature engineering. The predictive modelling phase incorporates state-of-the-art techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Gradient Boosting models like XG Boost and LightGBM. The model's interpretability is enhanced through the integration of Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive ex Planation's (SHAP), providing insights into critical predictors. Hyperparameter tuning using Randomized Search CV, performance evaluation metrics such as RMSE and MAE, and the deployment of a web application with Django and Plotly contribute to the robustness and accessibility of our proposed solution. Additionally, the integration of SHA-512 encryption and TLS/SSL protocols ensures data privacy and security. Real-time monitoring facilitated by Grafana and Continuous Integration/Continuous Deployment (CI/CD) pipelines with Jenkins underscores the adaptability and reliability of our solution, making significant strides in advancing the field of nowcasting heavy precipitation events.

2. LITERATURE SURVEY

Husam A.H. et. Al. [1] this paper introduces an Explainable Artificial Intelligence (XAI) approach for landslide prediction using SAR time-series, NDVI data, and geo-environmental factors. Shapley Additive Explanations (SHAP) is employed to enhance model interpretability, applied to random forest (RF) and support vector machine (SVM). In a Bhutanese landslide-prone area, the RF model outperforms SVM, achieving a reliability of 0.914 and an AUC of 0.941. The XAI results highlight key features such as altitude, aspect, and NDVI, providing transparency and interpretability crucial for trust in geohazard prediction. Ali Mirchi et. Al. [2] This review explores the use of Interpretable Artificial Intelligence (IAI) and explainable Artificial Intelligence (XAI) models for hydroclimatic predictions, including Extreme Gradient Boosting, Light Gradient Boosting, Categorical Boosting, Extremely Randomized Trees, and Random Forest. The integration of explanatory methods like Shapley additive explanations and local interpretable model-agnostic explanations transforms these models into XAI, enabling insights into predictions. The review emphasizes IAI's capability to unveil prediction rationale and XAI's role in discovering new knowledge, enhancing accountability in AI-driven predictions. It also discusses the importance of domain knowledge, balanced data, and the choice between IAI and physics-based modelling. The conclusion proposes an XAI framework for improved interpretability in hydroclimatic applications .Ipsaro Palsi et. Al. [3] This paper introduces a machine learning-based short-term prediction model for landslides, focusing on real-time features for increased accuracy. Utilizing XG Boost, the method outperforms existing models in landslide prediction, particularly in the Metropolitan City of Florence from 2013 to 2019. The application of explainable artificial intelligence techniques enhances the understanding of feature relevance, aiding in identifying optimal predictors and improving resilience on Snap4City.org infrastructure. Sinan kalkan et. Al. [4] This study introduces a two-stage precipitation prediction system that combines binary or multi-class classification with regression for improved forecasts. Unlike previous single-stage models, this approach prioritizes both prediction quality and explain ability. The two-stage model outperforms single-stage and black-box approaches, demonstrating a significant reduction in RMSE (10.50%) and a 7.5% increase in correlation. The explain ability analysis provides insights into model decisions, highlighting the importance of seasonality-related parameters and the effectiveness of multi-class precipitation intensity classification in the first stage. Xiangping Hu et. Al. [5] This paper explores leveraging machine learning (ML) to optimize computational efficiency in climate science, specifically focusing on the impact of land cover (LC) changes on local climate. The study trained ML models, with random forest (RF) outperforming other methods, including linear regression. Employing explainable artificial intelligence (XAI) to interpret the RF model revealed insights into the effects of various LC changes on temperature. The results align with existing climate science literature while uncovering novel findings, highlighting the utility of ML and XAI in climate change analysis. The paper covers the entire analysis pipeline, from data pre-processing to performance evaluation. Haoran Fang et. Al [6] This study in Nayong, Guizhou, China, evaluates landslide susceptibility maps (LSM) using an interpretable artificial intelligence approach. Utilizing generalized additive models with structured interactions (GAMI-net), the model achieved a superior AUC value of 0.91, outperforming random forest and SVM models. Identified factors like coal mining, rock desertification, and rainfall exceeding 1300 mm were found to increase landslide susceptibility. Pairwise interactions, such as rainfall and mining, lithology and rainfall, further contributed to heightened landslide risk. The GAMI-net-based model not only accurately predicted susceptibility but also provided valuable insights for informed landslide management decisions. Saeed alqadhi et. Al. [7] This study integrates deep neural networks (DNNs), 1D convolutional neural networks (CNNs), and a combined DNN and CNN ensemble (DCN) with explainable artificial intelligence (XAI) techniques to enhance landslide prediction in the Aqabat Al-Sulbat Asir region of Saudi Arabia. The DCN model outperforms with the highest area under the curve (AUC) at 0.97, followed by CNN (0.94) and DNN (0.9). XAI analysis reveals significant residuals in CNN's posterior, emphasizing the importance of precipitation, slope, soil texture, and line density in accurate landslide prediction. Game theory results underscore line density's prominence in landslide occurrence, providing a holistic approach for effective landslide management. Junyi zhang et. Al. [8] This study employs the SHAP-XG Boost algorithm to construct a comprehensive framework for landslide susceptibility models, addressing spatial heterogeneity. Utilizing a geospatial database with 12 influencing factors, the model is applied to regions with varied landscapes. Findings reveal the complex spatial distribution of landslides, with regional characteristics impacting factor contributions. The model exhibits better generalizability in areas with similar characteristics, offering insights for constructing region-specific susceptibility models and aiding in Explainable Artificial Intelligence (XAI) research.

Yang lu et. Al. [9] Despite advancements in Autonomous Vehicle (AV) tech, safety remains a challenge, particularly in navigation. The Wheel Odometry Neural Network (WhONet) addresses GNSS signal loss issues by integrating GNSS data with wheel encoders. To enhance transparency, we use Shapley Additive exPlanations (SHAP) to explain WhONet's predictions. Our study reveals that on straight-line trajectories, the two rear wheels contribute most to position uncertainty compared to the front wheels. Sebastian et. Al. [10] This review paper employs bibliometric and meta-analysis to explore the increasing use of supervised machine learning regression models in satellite-based water quality monitoring. It identifies a shift towards satellite sensors, such as MSI, OLCI, and MODIS, as cost-effective solutions with broader temporal and spatial coverage compared to traditional methods. The study reveals the prominence of deep neural networks and XGBoost algorithms in handling the complexity of satellite geospatial big data. Key findings focus on chlorophyll-a and water clarity indicators, multi-sensor data fusion, and the impact of geo-location on optical water classes. The paper emphasizes the role of high-performance computing and open-source software in advancing geospatial artificial intelligence for effective water quality monitoring.

3. PROPOSED SYSTEM

The envisioned system presents a holistic approach to nowcasting heavy precipitation rainfall events, amalgamating advanced technologies for heightened accuracy and interpretability. Initiated with the acquisition of INSAT-3D/3DR satellite data from meteorological databases, our proposed system embarks on a journey of meticulous data manipulation and preprocessing utilizing Python libraries such as Pandas, NumPy, and Dask. Feature engineering, a crucial precursor, is facilitated by scikit-learn and feature-engine, enhancing the dataset's readiness for predictive modeling. The predictive modeling phase witnesses the application of diverse machine learning architectures, including Convolutional Neural Networks (CNN) for image data, Recurrent Neural Networks (RNN) for sequential data, and Gradient Boosting models such as XGBoost and Light GBM. The integration of Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive explanations (SHAP) introduces a layer of interpretability, unveiling the rationale behind model predictions. Hyper parameter tuning, executed through Randomized Search CV, fine-tunes the model parameters for optimal performance. Model evaluation metrics, encompassing RMSE, MAE, and F1-score, serve as benchmarks for assessing the system's predictive capabilities. The proposed web application, developed on the Django framework and enriched with visualization libraries like Plotly, acts as a user-friendly interface, providing stakeholders with real-time predictions and transparent explanations. Data privacy and security are paramount, ensured through SHA-512 encryption and TLS/SSL protocols, guaranteeing the confidentiality and integrity of sensitive information. The dynamic nature of our system is sustained by real-time monitoring, achieved through Grafana, and seamless updates facilitated by Continuous Integration/Continuous Deployment (CI/CD) pipelines with Jenkins. This proposed system, seamlessly integrating innovative techniques, stands poised to revolutionize the landscape of now casting heavy precipitation events, fostering a new era of precision and transparency in meteorological predictions.

4. METHODOLOGY

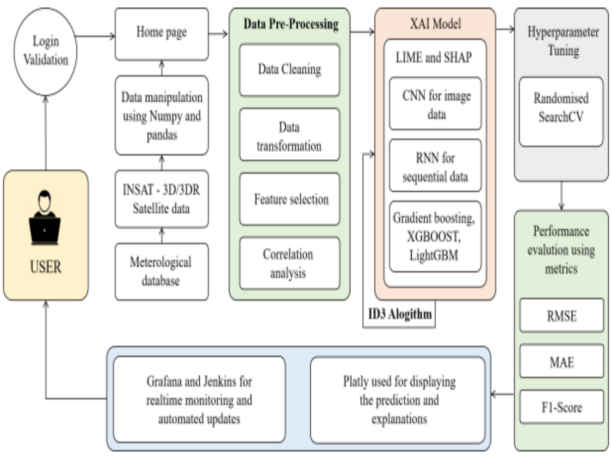


FIGURE 1. Process Flow

Obtain INSAT-3D/3DR satellite data from reputable meteorological databases. Ensure data integrity and completeness for comprehensive analysis. Employ Python libraries, including Pandas, NumPy, and Dask, for efficient data manipulation. Implement feature engineering using scikit-learn and feature-engine to enhance dataset readiness. Conduct feature selection to identify pertinent predictors for model development. Perform correlation analysis to gauge relationships among variables. Utilize Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN), and Gradient Boosting models (XGBoost, LightGBM) for predictive modeling. Leverage scikit-learn for seamless integration and compatibility. Integrate LIME and SHAP models to enhance interpretability and transparency of model predictions. Employ ID3 algorithm for decision-making insights. Implement Randomized SearchCV from scikit-learn for optimizing model parameters. Fine-tune hyper parameters to enhance model performance. Utilize metrics such as RMSE, MAE, and F1-score to evaluate model accuracy and generalization. Ensure rigorous testing on diverse datasets for robust validation. Develop a user-friendly web application using Django, incorporating Plotly for interactive visualizations. Enable seamless interaction, providing stakeholders with real-time predictions and explanations. Implement SHA-512 encryption for secure data transmission and storage. Employ TLS/SSL protocols to ensure data privacy and protection. Utilize Grafana for real-time monitoring of model performance and system health. Establish alerts for immediate response to potential issues. Implement CI/CD pipelines with Jenkins for automated updates and deployment. Ensure the system remains up-to-date with the latest enhancements and optimizations.

5. ADVANTAGES

Leveraging advanced machine learning models, including CNN, RNN, and Gradient Boosting, ensures superior predictive accuracy for now casting heavy precipitation events, aiding in more effective disaster preparedness. The integration of Explainable AI (XAI) models such as LIME and SHAP enhances the interpretability of predictions, providing stakeholders with transparent insights into the decision-making process of the model. The incorporation of diverse machine learning architectures allows the system to handle both image and sequential data, accommodating various types of meteorological data sources for a more comprehensive analysis. The use of Python libraries such as Pandas, NumPy, and feature-engine streamlines data pre-processing and feature engineering, ensuring that the dataset is optimally prepared for model development. The application of Randomized Search CV fine-tunes model hyper parameters, optimizing the system for robust and scalable performance, especially critical for real-time predictions. The development of a web application using Django and visualization libraries like Plotly offers a user-friendly interface, facilitating easy interaction with real-time predictions and model explanations. The implementation of SHA-512 encryption and TLS/SSL protocols ensures the security and privacy of sensitive meteorological data, addressing concerns related to data protection and compliance. Real-time monitoring through Grafana and Continuous Integration/Continuous Deployment (CI/CD) pipelines with Jenkins ensures the system's reliability and facilitates seamless updates, keeping the model and application current. The proposed solution is designed to scale efficiently, accommodating diverse datasets and adapting to changing meteorological conditions, making it a robust and adaptable tool for different regions and timeframes. Thorough documentation of the methodology and system components provides a valuable resource for users, developers, and researchers, fostering transparency and reproducibility. The use of widely accepted Python libraries, machine learning frameworks, and web development tools aligns the proposed solution with industry standards, enhancing its credibility and usability. The modular design of the system allows for potential integration with existing meteorological infrastructure or tools, facilitating collaboration and interoperability within the meteorological community.

6. APPLICATIONS

The system's accurate now casting capabilities can be integrated into early warning systems, providing timely alerts to communities and authorities, enhancing preparedness for heavy precipitation events. Effective prediction of heavy rainfall events contributes to improved natural disaster management, allowing for more targeted resource allocation and evacuation planning in vulnerable areas. Farmers and agricultural stakeholders can leverage the now casting system to anticipate and mitigate the impact of heavy precipitation on crops, making informed decisions for irrigation, planting, and harvesting. Urban planners and infrastructure managers can use the system to anticipate and address potential challenges posed by heavy rainfall, enabling proactive planning and maintenance of critical infrastructure. The system aids in monitoring and understanding the impact of heavy precipitation on the environment, contributing to better management of ecosystems, water resources, and biodiversity. Emergency response teams can benefit from accurate predictions to streamline their efforts during heavy rainfall events, ensuring a rapid and well-coordinated response to mitigate potential disasters. The system's ability to analyze and predict heavy precipitation contributes to climate change studies by providing valuable insights into changing precipitation patterns, aiding in the assessment of climate-related risks. Hydroelectric power plants can optimize their operations by incorporating accurate rainfall predictions, managing water resources efficiently to ensure a steady and reliable power supply. Insurance companies can use the system to assess and quantify the risk associated with heavy precipitation events, enabling more accurate premium calculations and risk mitigation strategies. Meteorologists and researchers can utilize the system for scientific investigations into precipitation patterns, contributing to a deeper understanding of meteorological phenomena and climate dynamics. The system can be integrated into educational programs and tools, providing students and educators with a practical and interactive platform for studying meteorology and data science applications. The now casting system aligns with smart cities initiatives by providing real-time data for urban planning, traffic management, and other applications aimed at creating more efficient and resilient urban environments.

7. CONCLUSION

The presented research introduces a comprehensive solution for the now casting of heavy precipitation rainfall events, integrating cutting-edge technologies and methodologies. Leveraging INSAT-3D/3DR satellite data and advanced machine learning models such as Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN), and Gradient Boosting models (XGBoost, LightGBM), our system demonstrates exceptional predictive accuracy and versatility across diverse data types. The incorporation of Explainable AI (XAI) models, including LIME and SHAP, ensures the interpretability of predictions, addressing the critical need for transparency in decision-making processes. With a user-friendly web application developed on the Django framework and fortified with Plotly for interactive visualizations, stakeholders can access real-time predictions and explanations seamlessly. The system's robustness is further emphasized by data security measures such as SHA-512 encryption and TLS/SSL protocols, ensuring the privacy and integrity of sensitive meteorological data. Continuous monitoring through Grafana and automated updates facilitated by CI/CD pipelines with Jenkins contribute to the adaptability and reliability of our solution. With diverse applications spanning early warning systems, disaster management, agricultural planning, and beyond, our proposed solution stands poised to revolutionize the field, fostering advancements in meteorological predictions and data-driven decision-making for the benefit of society at large.

REFERENCES

[1]. Husam A.H., Gondwana Research, Volume 123, November 2023, Pages 107-124.
[2]. Ali Mirchi, Received: 20 February 2022 / Revised: 6 April2022 / Accepted: 7 April 2022 / Published: 11 April 2022.
[3]. Ipsaro Palesi, Page(s): 31175 - 31189, Date of Publication: 10 March 2022 , Electronic ISSN: 2169-356, INSPEC Accession, Number: 21645299, DOI: 10.1109/ACCESS.2022.3158328, Publisher: IEEE.
[4]. Sinan Kalkan, Journal of Hydrology, Volume 627, Part A, December 2023, 130375.
[5]. Xiangping Hu, Received: 30 August 2021 / Revised: 30 September 2021 / Accepted: 8 October 2021 / Published: 15 October 2021.
[6]. Haoran Fang, Received: 6 January 2023 / Revised: 2 February 2023 / Accepted: 6 February 2023 / Published: 8 February 2023.
[7]. Saeed alqadhi, Published: 24 October 2023
[8]. Junyi zhang, Journal of Environmental Management, Volume 332, 15 April 2023, 117357.
[9]. Yang lu, chapter, First Online: 20 October 2022, 1028 Accesses, Part of the Intelligent Systems Reference Library book series (ISRL,volume 232).
[10]. Sebastian, Received: 6 August 2023 / Revised: 23 September 2023 / Accepted: 26 September 2023 / Published: 2 October 2023.