International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Flood Prediction using Machine Learning

Prof. Priyanka Pujari, Nidhi Kulkarni, Vinay C Hiremath, Vinay Bhushi

Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India

A B S T R A C T

India is subject to frequent natural disasters in the form of floods, resulting in significant loss of life and property. Accurate prediction of flood onset and progression in real time is critical to minimizing flood impacts. This research paper focuses on a comparative study of different machine learning models for flood forecasting in India. Models analyzed include K-nearest neighbor (KNN), support vector classification (SVC), decision tree classification, binary logistic regression, and stacked generalization (stacking). Accurately predicting the onset and progression of floods in real time is difficult. To estimate water levels and velocities over a large area, it is necessary to combine the data with computationally demanding flood propagation models. This paper aims to reduce the extreme risks of this natural disaster and contribute to policy proposals through flood forecasts using various machine learning models.

Keywords: Decision tree, Support Vector Machine (SVM), Logistic Regression, Stacked Generalization.

1. Introduction

Floods caused by overflowing lakes, rivers or oceans constantly affect millions of people around the world. In India, Bangladesh and China alone, 4.84 million, 3.84 million and 3.28 million people respectively struggle against annual floods [1]. Many countries are prone to flooding, including the Netherlands, Monaco, Bahrain and low-lying areas. Between 1997 and 2008, Australia documented 73 people lost to floods [2]. In the United States, floods claim about 100 lives each year and cause a total of $7.5 billion in damage [3]. The World Resources Institute estimates that by 2030, floods will affect 147 million people and cause $174 billion to $712 billion in property damage. Several studies have used forecasting models such as NARX and BPNN to predict floods with varying degrees of success. This study focuses on improving the accuracy of flood forecasting using binary logistic regression, a method that classifies flood data sections using a binary classification. Traditional machine learning techniques such as SVC and KNN have shown superior accuracy. Despite the recurring nature of natural disasters, especially floods, recovery efforts should not be taken lightly. The accuracy of predicting river levels after heavy rain is critical for public safety, environmental considerations, and effective management of water resources. Learning from historical data, machine learning independently creates predictive models that are key to predicting proactive actions and recommendations for future events. The once-a-year catastrophic flood in Kerala, India, highlights the urgent need to monitor rainfall patterns, forcing a thorough investigation of flood forecasting methods..

2. Methodology

In undertaking a flood prediction project utilizing machine learning (ML), the initial step involves precisely defining the problem scope, specifying the types of floods targeted for prediction and the geographical areas of interest. Subsequently, a comprehensive collection of relevant datasets is performed, encompassing historical weather patterns, river/streamflow data, topography, and land-use variables. Rigorous data preprocessing ensues, addressing issues such as missing values and outliers, accompanied by feature engineering to extract pertinent predictors. The exploration of data patterns and relationships follows through exploratory data analysis (EDA). Feature selection is then executed, deploying statistical methods or ML techniques to identify the most impactful variables for flood prediction. The selection of ML models, including decision trees, random forests, and neural networks, is undertaken, with subsequent training on a partitioned dataset and optimization of hyperparameters for enhanced performance. Evaluation metrics such as accuracy, precision, and recall guide the assessment of model performance. Iterative tuning and cross-validation ensure model robustness, culminating in deployment for real-time flood prediction. The ongoing monitoring, maintenance, documentation, and transparent communication of findings and limitations to stakeholders form integral components of this methodology.

Once trained and fine-tuned, the models are deployed for real-time flood prediction, a moment where the theoretical work transforms into practical utility. Continuous monitoring and maintenance ensure the model's adaptability to evolving conditions, while thorough documentation and transparent communication with stakeholders provide a comprehensive understanding of the model's capabilities and limitations, fostering trust in its application within flood-prone areas.

3. PROPOSED MODEL

In the proposed flood prediction model, machine learning algorithms will be leveraged to analyze historical weather data, river levels, and other relevant variables to develop a robust predictive model. The system will utilize advanced statistical and machine learning techniques to identify patterns and correlations that precede flooding events. Feature engineering will play a crucial role in selecting and transforming input variables to enhance the model's accuracy. Real-time data from various sources, including weather stations, satellite imagery, and river sensors, will be continuously fed into the model to ensure its responsiveness to changing conditions. The model will be trained on a diverse dataset encompassing different geographical regions and environmental conditions to enhance its generalization capabilities. Additionally, the incorporation of deep learning techniques, such as recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks, will enable the model to capture temporal dependencies within the data, improving its ability to forecast flood events with greater precision. The ultimate goal of this flood prediction model is to provide timely and accurate warnings, allowing for proactive disaster management and mitigation efforts to protect vulnerable communities and infrastructure.

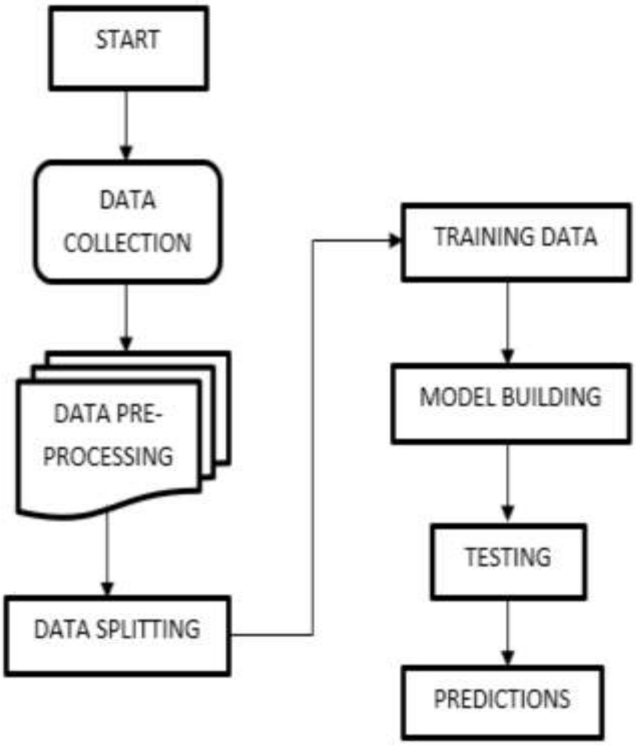

Fig 1: Proposed Model

1. **Data Collection:** Raw data collected from various sources (e.g. sensors, satellites or IoT devices) containing information about soil, weather, crops etc.
2. **Data Pre-processing:** Cleaning, organizing and preparing raw data for analysis. This step involves handling missing values, removing outliers, and converting the data to a usable format.
3. **Input dataset:** Processed data organized into a structured dataset ready for use by machine learning algorithms or analytical tools.
4. **Data Splitting data for training and testing:** Split the input data into two parts - one to train the model and one to test the trained model and performance. This ensures that the model not only remembers data, but also learns patterns.
5. **Model Building:** Adopting a model in precision agriculture means using machine learning or artificial intelligence models in an agricultural framework to analyze data and make informed decisions for better yields, soil health, pest control and overall agricultural efficiency.
6. **Prediction:** Applying the trained model to new or unseen data to generate predictions or insights. In precision farming, this could mean predicting crop yield, disease detection, or recommending optimal farming practices.

3. Comparison Table :

SL No.	Title	Authors	Advantages	Disadvantages	Accuracy
1.	Flood Prediction Using Ensemble Machine Learning Model	Miah Mohammad Asif Syeed, Maisha Farzana, Ishadie Namir, Ipshita Ishrar, Meherin Hossain Nushra, Tanvir Rahman, Bhoktear Mahbub Khan	machine learning models can provide accurate and timely flood predictions, enabling disaster management authorities to take appropriate measures to minimize damage and save lives. Binary Logistic Regression is provide high accuracy.	The K-Nearest Neighbors (KNN) is provide low accuracy.	96.0%
2.	Flood Prediction Using Machine Learning Models	Miah Mohammad Asif Syeed, Maisha Farzana, Ishadie Namir, Ipshita Ishrar, Meherin Hossain Nushra, Tanvir Rahman	To reduce the extreme risks of this natural disaster and also contributes to policy suggestions by providing a prediction for floods using different machine learning models	The support vector classifier is provide low accuracy compare to other methods.	93.0%
3.	Flood Prediction using Deep Learning Models	Muhammad Hafizi Mohd Ali, Siti Azirah Asmai*, Z. Zainal Abidin, Zuraida Abal Abas, Nurul A. Emran	In terms of prediction accuracy, the experimental results also demonstrated that the deep recurrent neural network model with layer normalization and Leaky ReLU activation function performed better than other models.	long-term short memory (LSTM) is take more time to provide accuracy.	81.11%
4.	Machine learning approach for flood risks prediction	Nazim Razali, Shuhaida Ismail, Aida Mustapha	SMOTE method is found highly effective in dealing with imbalance dataset.	Decision Tree (DT), k-Nearest Neighbors (KNN) and Support Vector Machine (SVM) for flood risks prediction in this model.	99.0%
5.	Flood prediction forecasting using machine Learning Algorithms	Naveed Ahamed, S. Asha	There exist a lot of machine algorithms which generate models with more accuracy.	This model is not provide more accuracy than other models.	77.3%

Fig 2. Comparison Table

1. Flood Prediction Using Ensemble Machine Learning Model: This research paper focuses on a comparative study of various machine learning models for flood prediction in India. They used a dataset of rainfall to train and test the models. Results indicate that the stacked generalization model outperforms the other models, achieving an accuracy of 96.0% and Standard Deviation of 0.098.

2. Flood Prediction Using Machine Learning Models: This research will use Binary Logistic Regression, K-Nearest Neighbor (KNN), Support Vector Classifier (SVC) and Decision tree Classifier to provide an accurate prediction. And achieving an accuracy of 93.0%.

3. Flood Prediction using Deep Learning Models: This study proposed a time series model with layer normalization and Leaky ReLU activation function in multivariable long-term short memory (LSTM), bidirectional long-term short memory (BILSTM) and deep recurrent neural network (DRNN). And achieving an accuracy of 81.11%.

4. Machine learning approach for flood risks prediction: This study aims to develop a predictive modelling follow Cross Industry Standard Process for Data Mining (CRISP-DM) methodology by using Bayesian network (BN). The results showed that DT with SMOTE method performed the best compared to others by achieving 99.92% accuracy. Also, SMOTE method is found highly effective in dealing with imbalance dataset.

5. Flood prediction forecasting using machine Learning Algorithms: This study proposed a time series model with layer normalization and Leaky ReLU activation function in multivariable long-term short memory (LSTM), bidirectional long-term short memory (BILSTM) and deep recurrent neural network (DRNN). In terms of prediction accuracy, the experimental results also demonstrated that the deep recurrent neural network model with layer normalization and Leaky ReLU activation function performed better than other models. And achieving an accuracy of 77.3%.

4. Conclusion

In conclusion, the proposed flood forecasting model combines advanced machine learning techniques, interpretability and community participation to provide a comprehensive solution for proactive disaster management. With a focus on adaptability, transparency and real-time communication, the model aims to provide timely warnings to vulnerable areas. A user-friendly interface and a multi-channel alarm system improve its usability, while continuous collaboration with local communities ensures the effectiveness of the model and promotes sustainability. Combining technology with a community-based approach, the flood forecasting model is a promising tool to reduce the impact of floods and promote sustainable disaster preparedness.

References

[1] Miah Mohammad Asif Syeed, Maisha Farzana, Ishadie Namir, Ipshita Ishrar, Meherin Hossain Nushra, Tanvir Rahman, Bhoktear Mahbub Khan, A Review on Flood Prediction Using Ensemble Machine Learning Model, 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2023.

[2] Miah Mohammad Asif Syeed, Maisha Farzana, Ishadie Namir, Ipshita Ishrar, Meherin Hossain Nushra, Tanvir Rahman, Flood Prediction Using Machine Learning Models, 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2022.

[3] Muhammad Hafizi Mohd Ali, Siti Azirah Asmai, Z. Zainal Abidin, Zuraida Abal Abas, Nurul A. Emran, Flood Prediction using Deep Learning Models, International Journal of Advanced Computer Science and Applications (IJACSA), 2022.

[4] Nazim Razali, Shuhaida Ismail, Aida Mustapha, Machine learning approach for flood risks prediction, IAES International Journal of Artificial Intelligence (IJ-AI), 2020.

[5] Naveed Ahamed, S.Asha, Flood prediction forecasting using machine Learning Algorithms, International Journal of Scientific & Engineering Research, 2020.

[6] Akshay Kharche, Pratibha Bhagat, Syed Faiz Ibrahim, A Review on Flood Prediction Using Machine Learning based Apache SystemML Python Platform, 2019 JETIR January

[7] A. M. Kamal, M. Shamsudduha, B. Ahmed, S. K. Hassan, M. S. Islam, I. Kelman, and M. Fordham, "Resilience to flash floods in wetland communities of northeastern bangladesh", International journal of disaster risk reduction, vol. 31, pp. 478-488, 2018

[8] Dr. nasina krishna kuma, attla vinay kumar, Flood prediction using machine learning models and forecast in methods, Journal of Engineering Sciences, 2022.

[9] H. Shafizadeh-Moghadam, R. Valavi, H. Shahabi, K. Chapi, and A. Shirzadi, "Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping," Journal of environmental management, vol. 217, pp. 1-11, 2018.

[10] Amir Mosavi, Pinar Ozturk, and Kwok-wing Chau, Flood Prediction Using Machine Learning Models: Literature Review Department of Computer Science (IDI), Norwegian University of Science and Technology,2018.

[11] V. Yadav and K. Eliza, "A hybrid wavelet-support vector machine model for prediction of lake water level fluctuations using hydro-meteorological data," Journal of the International Measurement Confederation, vol. 103, pp. 2655–2675, 2017.

[12] A. D. A. Dali, N. A. Omar, and A. Mustapha, "Data mining approach to herbs classification," Indonesian Journal of Electrical Engineering and Computer Science, vol. 12, no. 2, pp. 570–576, 2018.

[13] S. F. A. Razak, C. P. Lee, K. M. Lim, and P. X. Tee, "Smart halal recognizer for muslim consumers," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 1, pp. 193–200, 2019.

[14] M. A. S. Anuar, R. Z. A. Rahman, S. B. Mohd, A. C. Soh, and Z. D. Zulkafli, "Early prediction system using neural network in kelantan river, malaysia", in Proceedings of the 15th IEEE Student Conference on Research and Development: Inspiring Technology for Humanity, SCOReD 2017, 2018, pp. 104–109.

[15] M. Abdullah, M. Othman, S. Kasim, and S. Mohamed, "Evolving spiking neural networks methods for classification problem: a case study in flood events risk assessment," Indonesian Journal of Electrical Engineering and Computer Science, vol. 16, no. 1, pp. 222–229, 2019.

[16] G. Zhao, B. Pang, Z. Xu, D. Peng, and L. Xu, "Assessment of urban flood susceptibility using semi-supervised machine learning model," Science of the Total Environment, vol. 659, no. 3, pp. 940–949, 2019.

[17] Z. Geng, X. Hu, Q. Zhu, Y. Han, Y. Xu, and Y. He, "Pattern recognition for water flooded layer based on ensemble classifier," in Proceedings of the 5th International Conference on Control, Decision and Information Technologies, CoDIT 2018, 2018, pp. 164–169.

[18] T. Ashizawa, N. Sudo, and H. Yamamoto, "How Do Floods Affect the Economy? An Empirical Analysis using Japanese Flood Data," 2022.

[19] M. S. M. Shaari, M. Z. Abd Karim, and B. Hasan-Basri, "Does flood disaster lessen GDP growth? Evidence from Malaysia's manufacturing and agricultural sectors," Malaysian J. Econ. Stud., vol. 54, no. 1, pp. 61–81, 2017, doi: 10.22452/mjes.vol54no1.4.

[20] K. A. Oladapo, S. A. Idowu, Y. Adekunle, and F. Ayankoya, "Categorization of Conditioning Variables for Pluvial Flood Risk Assessment," Int. J. Sci. Eng. Res., vol. 11, no. 8, pp. 355–368, 2020.