International Academic Institute for Science and Technology

International Academic Journal of Innovative Research
Vol. 7, No. 2, 2020, pp. 31-36.

ISSN 2454-390X

International Academic Journal of Innovative Research

www.iaiest.com

Harnessing Machine Learning for Flood Prediction and Future Directions

Amit Khare [a], Isha Ghosh [b]

[a] *Department of Instrumentation Engineering, Triveni University of Science, Patna, India.*
[b] *Department of Instrumentation Engineering, Triveni University of Science, Patna, India.*

Abstract

Natural disasters have become more common in recent decades due to a variety of factors, including changes in the climate and environmental circumstances. Floods, earthquakes, tsunamis, and storms in particular have been more common in recent years and have the potential to cause property damage and human casualties. All living things are impacted by these natural disasters, including marine and wildlife species. Floods are stated to have a direct impact on people's life by destroying their means of subsistence as well as their private and public property. Increased water flow in typical streams is somewhat regulated, but when it beyond a typical threshold, it can have catastrophic consequences. The main cause of floods is the excessive use of wetlands and deforestation for human habitation. Development operations often alter the natural flow of rivers, ponds, and lakes, resulting in negative impacts or alterations to the water's real flow.

Keywords: Machine Learning, Natural Disaster, Flood.

Received: 12 September 2020/Revised: 24 October 2020/Accepted: 30 November 2020/Published: 31 December 2020

Introduction

Floods can be divided into river floods and flash floods based on the location and environmental factors. Flash floods are caused by an abrupt rise in water levels that rushes out on a dry stream when there is an excessive amount of rainfall in an area that normally does not receive consistent rainfall throughout the year (Oprea et al., 2020). These floods are frequent in arid climates and rocky terrains, and they are fatal. On the other side, river floods occur when too much rain overwhelms rivers' natural flow, raising their water levels. In areas where people have settled, river floods occur frequently and wreak severe damage to both people's lives and property. Floods can be forecast, unlike earthquakes, forest fires, and tsunamis, by properly tracking and monitoring rainfall-related events and by understanding past occurrence patterns (Bhuyan et al., 2014).

It has been demonstrated that machine learning algorithms are a successful choice for intricate numerical operations that may be carried out using statistics gathered over many years with straightforward structures (Jain, 2019). A kind of artificial intelligence called machine learning allows models to predict the necessary circumstances by using patterns found in historical data. Better outcomes are promised by machine learning algorithms, which get their forecasts from any event-based study. In order to improve understanding of floods, data-driven models are known to take historical data and other hydro-meteorological indices into account while calculating climatic parameters. The main focus of the statistical models is to anticipate floods by analyzing their frequency. Autoregressive moving average, linear regression, multiple linear regression, autoregressive integrated moving average, and other well-known models have been created for flood forecasting (Baryannis et al., 2019).

Related Works

Based on the information already available, there are a number of intermediate steps and measures to anticipate the likelihood of flooding. Figure 1 depicts the generalized flow of actions or processes of a flood prediction model based on a summary of current state-of-the-art methodologies.

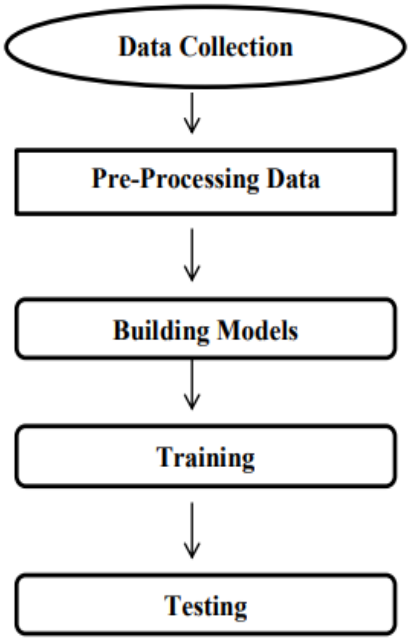


Fig. 1: Flood Prediction Model

Numerous factors, including water level, rainfall discharge, seasons, precipitation, soil level, soil moisture content, river inflow and outflow, seasonal flow, peak discharge, frequency, storm surge, surge level, plain flood, urban flood, typhoons, etc., are dependent on all flood prediction applications (Bilal et al., 2016). A deeper understanding of the qualitative and quantitative forecasting of rainfall, avalanches, snowmelt rate, etc., is provided by the spatial analysis of hydro-geomorphic characteristics. Experts from many fields gather and record all of these factors in order to filter them into prediction models. A flood prediction system, in contrast to other prediction models, incorporates numerous modalities and characteristics to provide an infallible mechanism (Shen, 2018).

Methodology

Dataset

To guarantee that the gathered data is accurate and comprehensive, it must be cleansed and pre-processed. Because machine learning algorithms rely on entire data sets to predict outcomes, there is a significant risk of false positives when there are missing values. Every machine learning algorithm will have various values, suggesting that depending on the application and prediction type, different data sets are required for a machine learning approach to flood prediction. Models developed for long-term stream flow predictions are entirely different from machine learning methods utilized for short-term flood prediction. In certain data sets, photographs are also included to help visualize the consequences of floods, and information may not always be in a single format. In contrast to textual and numerical data, photographs require a thorough examination of time series, background separation, thresholding, morphological procedures, and noise level removal (Sun & Scanlon, 2019).

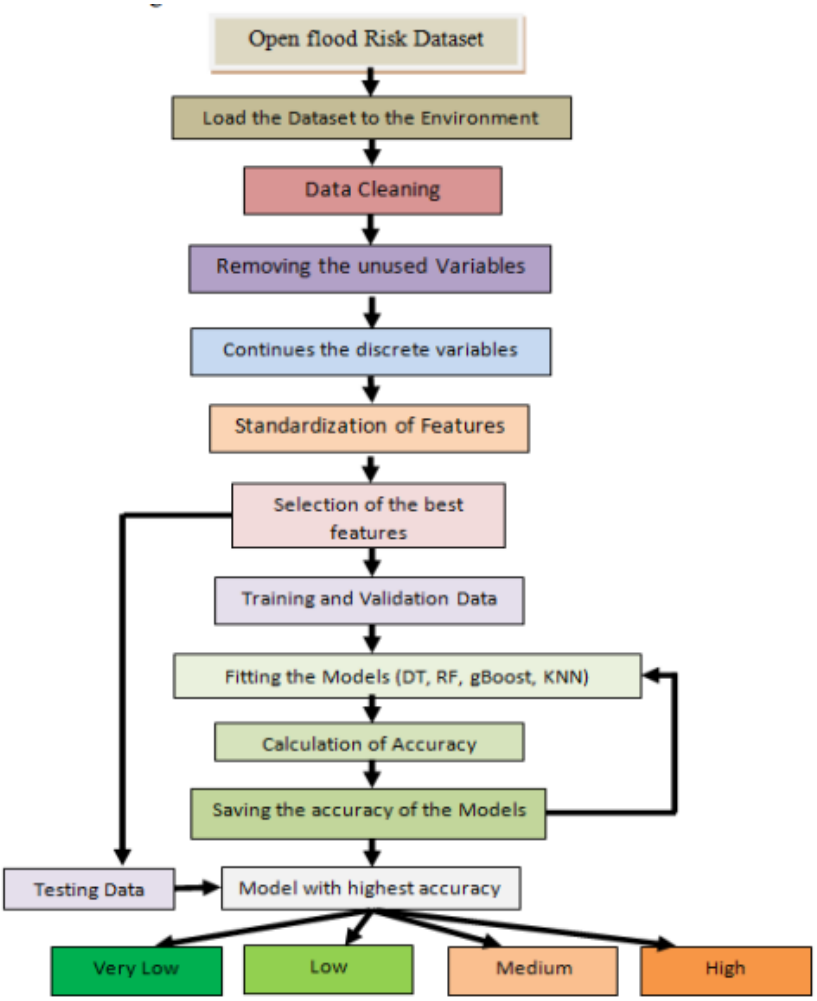


Fig. 2: Design Framework

Methods

We used k-means clustering, an unsupervised machine learning technique, to efficiently group these observations according to comparable levels of total risk. Cluster analysis is essentially a data organization technique that groups observations according to user-selected features (Karpatne et al., 2018).

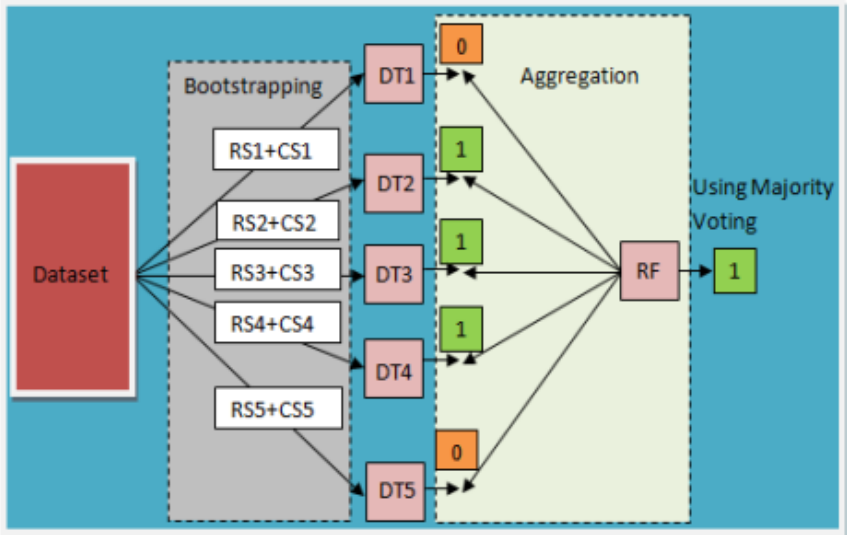


Fig. 3: Random Forest Classification

We can use this technique to group a collection of observations according to similar or dissimilar traits. One of the great things about cluster analysis is that it doesn't require prior knowledge of the classes, unlike supervised algorithms. Lastly, we employed the Random Forest algorithm to predict potential disasters.

Results and Discussion

Flood prediction is quite a complex issue that requires taking into account a variety of factors when identifying areas at risk of flooding. The accuracy of these predictions' hinges on numerous elements, including the specific geography of a location. Therefore, it's essential to develop a more comprehensive algorithm that incorporates a wider range of features to improve prediction accuracy. Explainable AI is an emerging field that offers opportunities to enhance transparency in flood risk mapping through the use of white box models. It's also crucial to address any biases that may arise during model training and result interpretation. For example, Fig. 4 shows the confusion matrix for the decision tree classifier, which aids in evaluating its accuracy.

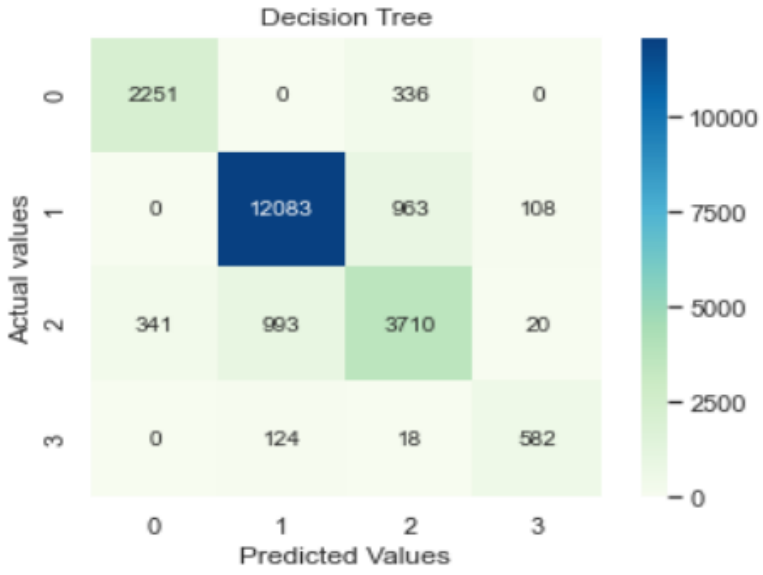


Fig. 4: Confusion Matrix

The decision tree, which represents the macro average, has a ROC accuracy score of 88.12%. Figure 5 illustrates the model's performance, scalability, and learning curve.

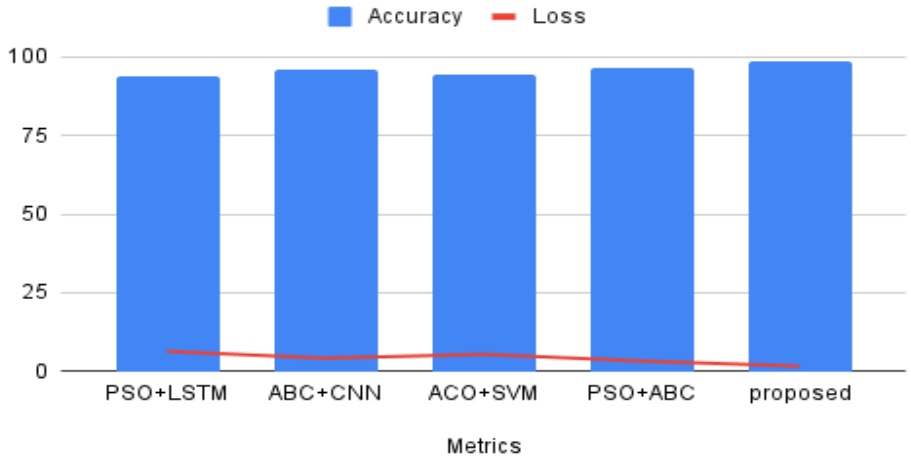


Fig. 5: Performance Metrics

The research so far has relied on manually inputting data into each machine learning model for predictions, which can be quite time-consuming. This proposed concept's main goal is to create a technique that automatically chooses the best flood prediction algorithm.

Conclusion

Every day, the severity of both man-made and natural disasters keeps growing. High population density, a high volume of traffic, and stringent building construction restrictions are the main factors that define smart cities. People feel safer when they are aware that government facilities and services, such as schools, hospitals, public transportation, communication, water, and energy, are easily accessible. By integrating information technology with physical systems and infrastructure, "smart cities" could promote improved mental clarity, closer links within the community, and more chances to test and apply novel concepts. The importance placed on these elements in the newly created notion of urban resilience is causing people's ideas of smart cities to change. In reality, resilience is dependent on the complex web of connections between individuals, groups, organizations, and systems. Natural disasters will therefore disproportionately affect cities, causing significantly more loss and destruction than smaller towns. The ability of "smart cities" to effectively handle crises is one of its most crucial characteristics. It has been demonstrated that machine learning algorithms are a successful alternative for intricate numerical operations that may be carried out using statistics gathered over many years using straightforward structures.

References

Oprea, S., Martinez-Gonzalez, P., Garcia-Garcia, A., Castro-Vargas, J. A., Orts-Escolano, S., Garcia-Rodriguez, J., & Argyros, A. (2020). A review on deep learning techniques for video prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(6), 2806-2826. https://doi.org/10.1109/TPAMI.2020.3045007

Bhuyan, M. H., Kashyap, H. J., Bhattacharyya, D. K., & Kalita, J. K. (2014). Detecting distributed denial of service attacks: methods, tools and future directions. *The Computer Journal, 57*(4), 537-556. https://doi.org/10.1093/comjnl/bxt031

JAIN, S. (2019). Beyond Traditional Algorithms: Harnessing Reinforcement Learning and Generative AI for Next-Generation Autonomous Systems.

Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: state of the art and future research directions. *International journal of production research, 57*(7), 2179-2202. https://doi.org/10.1080/00207543.2018.1530476

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., ... & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced engineering informatics, 30*(3), 500-521.https://doi.org/10.1016/j.aei.2016.07.001

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research, 54*(11), 8558-8593. https://doi.org/10.1029/2018WR022643

Sun, A. Y., & Scanlon, B. R. (2019). How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters, 14*(7), 073001. https://doi.org/10.1088/1748-9526/ab1b7d

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2018). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering, 31*(8), 1544-1554. https://doi.org/10.1109/TKDE.2018.2861006