

Novel Flood Prediction System based on Big Data and Machine Learning

Ajay N¹, Mohan H S², Shwetha B V³ and Shrihari M R⁴

¹S J C Institute of Technology, VTU University, India
Email: n.ajay2@gmail.com

²New Horizon College of Engineering, VTU University, India

³Nagarjuna College of Engineering and Technology, VTU University, India

⁴S J C Institute of Technology, VTU University, India
Email:shrihari.mr@gmail.com

Abstract—The One of nature's most destructive hazards, floods cause both human fatalities and property destruction. Numerous cities are influenced by the monsoon, therefore they frequently experience calamity. Early warning of a flood incident could help the public and the authorities plan both short-term and long-term preventive actions, prepare for evacuation and rescue efforts, and provide relief for flood victims. For example, one of the main factors in most flood management is the geographic locations of the affected areas and their respective severity. There is still no reliable method for predicting floods in advance. Existing technologies typically relied on prepared and manually entered data. Since the procedures were time-consuming, making real-time and early forecasts impossible. Additionally, these systems did not fully utilize the more thorough data that was accessible in the most recent big data platforms. As a result, this article suggests a unique flood forecasting method based on combining large data from crowd sources, geospatial systems, and meteorological systems. Modern learning strategies were at the heart of data intelligence. Both subjective and objective assessments showed that the proposed method could predict flood episodes that would occur in particular locations and times. Later benchmarking studies showed that the MLP ANN-equipped system provided the best accurate prediction, with correct percentage, Kappa, MAE, and RMSE values of 97.93, 0.89, 0.01, and 0.10, respectively.

Index Terms— Machine Learning, Big Data, Flood Forecasting.

I. INTRODUCTION

One of the most frequent natural calamities is flooding. Major flood disasters always result in significant property damage and, more often than not, fatalities. This contrasts with the infrequently seen stagnant water flow in poorly built communities. Due to the fact that many Asian nations, including Thailand, are vulnerable to both the southwest and northeast monsoons, the majority of the countries experience seasonal flooding virtually every year.

The most important of the prominent reasons in Thailand is the unexpected and persistent heavy rain. Furthermore, damage can be greatly dispersed by overflow from major rivers along shore sides to neighboring basins. Unsuitable land uses prevent an area from effectively discharging stored rainwater, while being placed further from a river. As a result, it is unavoidably more prone to flooding, which occurs more frequently.

No of the reason, a flood is typically unexpected and makes it nearly impossible for the general people and any pertinent organizations to be effectively prepared for the occurrence. This is primarily caused by the absence of a reliable method for foreseeing the tragedy well in advance.

The majority of computerized flood predicting systems continue to be based on current precipitation, as measured by rain stations or rain gauges, despite recent substantial development in this area. Normally, meteorology departments or other comparable entities are in charge of these installations. Moreover, because their installation and upkeep are expensive, they are only sparsely distributed in a few places. As a result, it might be challenging to precisely estimate precipitation or flood, especially in locations without such a facility. To address this problem, precipitation in these locations was often calculated by extrapolating or interpolating from regions where rain stations were present.

Because there aren't enough of these stations, the data from one location might not be a useful indicator for others. As a result, the anticipated precipitation was too inaccurate to use in a forecast. The measurement, processing, recording, and transfer of traditional meteorological readings—such as precipitation, temperature, and humidity—to pertinent entities takes a very lengthy period. It was known that assessments based on prior precipitation had a number of flaws.

For example, they help make flood predictions that are frequently wrong and out of date. The true potential of this strategy was being undermined by small sample sizes, poor processing power, and ineffective prediction techniques. Nevertheless, computerized flood forecasting[2][3] based on theme elements has been extensively studied in light of current developments in distributed computing, particularly modern machine learning (ML), which resembles human intelligence. Additionally, Due to the rise in the volume of both open and private data sources, big data has now established itself as a crucial information source in these endeavors.

Recent assessments indicate that up until now, the majority of flood forecasting systems have mostly relied on monitored precipitation data or data gathered from a single source[15]. In addition to the aforementioned drawbacks, current systems still have many other shortcomings. For instance, there wouldn't be any precipitation data available for critical studies in the event that monitoring facilities or a communication network malfunctioned. Additionally, to the finest of our knowledge, no programme or application existed that could provide thorough forecasting for a specific area. Furthermore, the use of the available tools was heavily reliant on the labor-intensive collection and preparation of data from a variety of sources, including Big Data. As a result, it has remained extremely difficult to automatically and spontaneously alert the public and authorities about flood crises or to realistically anticipate them[7].

Additionally, ML-based flood forecasting systems have recently undergone development [8]. Both characteristics and crowd sourcing data were incorporated into these systems' ML frameworks. The majority of current systems, however, function by first processing these data offline and onsite before displaying their prediction results on multiple platforms. A typical method was put out in where an ML was trained using real-time data from stream flow, rainfall, and other sources. However, it was not clear how a forecast result that was obtained from crowd sourcing was compared to an actual event.[14]

The primary issue claims are one of nature's most destructive hazards, floods cause both human fatalities and property destruction. Numerous cities are influenced by the monsoon, therefore they frequently experience calamity. The public and government may benefit from early reporting of flood incidents. Preparing for both short-term and long-term preventative measures, rescue missions, and providing relief for flood victims. Existing technologies typically relied on prepared and manually entered data. Since the procedures were time-consuming, making real-time and early forecasts impossible. More extensive information that was available in current big data platforms was not completely utilized by existing technologies. In this article, an unique flood forecasting method based on combining meteorological, hydrological, geospatial, and crowd sourced data[1] is proposed.

II. RELATED WORK

Experiments were conducted in coastal areas of Chennai in order to demonstrate the benefits of the suggested plan. When creating the dataset, we will take geographic information into consideration. Both the southeast and northwest monsoons have an impact on these regions. They are both hence vulnerable to significant flooding. This strategy would demonstrate that there were no data or core process restrictions placed on these particular provinces. Consequently, the suggested plan might be adapted and used successfully in various contexts[5].

To gather crowdsourced (or volunteer) data for flood forecasting, such as geospatial, meteorological, and hydrological information from GLOFAS. Data needs to be normalised, and a common coordinate frame needs to be created. data preprocessing for data integration. MLP ANN, RF, and SVM are used to predict and evaluate the outcomes[4].

A novel distributed flood forecasting system based on the fusion of atmospheric, hydrologic, geospatial, and crowd source data was presented in this research and is depicted in Figure.1. Big data from well-known agencies was obtained using a variety of cross-platform APIs. On the basis of these data, which cutting-edge ML techniques learnt, forecasting was done[9]. They included fuzzy logics, RF, Naive Bayes, MLP and RBF ANN, SVM, and decision trees. It was demonstrated empirically that the created system may be used to warn the public and government of future floods in addition to the current one. Additionally, this system improved user experience with responsive graphical user interfaces that were compatible with a variety of computer devices, including mobile ones[6]. This benefit effectively encouraged more public participation in crowdsourcing data, enriching data aggregation, and ultimately improving system correctness and dependability.

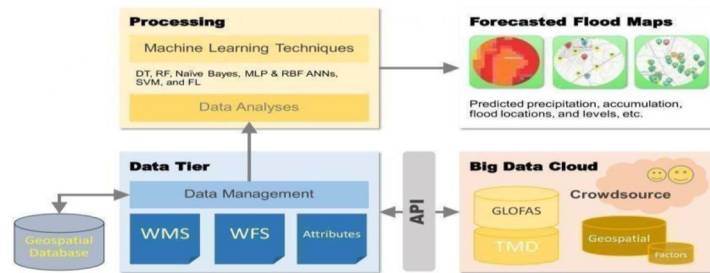


Figure 1 Process framework of the suggested plan

Similar to game planning, the information stream outline shows the graphical portrayal. It is used to communicate with the information through the information sources, different information types are examined, and the desired yield is created. These components will be used to illustrate the framework and show how to quickly think about the information. The Figure.2 Shows DFD will show the stream of complete parts in the framework outline. Using this framework, the data stream will alter in its layout.

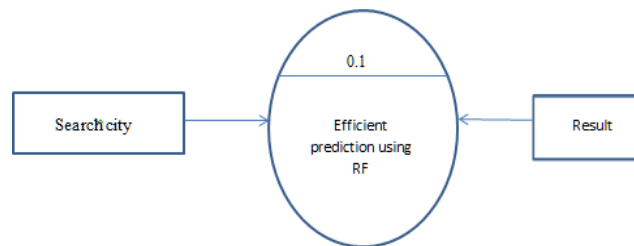


Figure 2 Data Flow Diagram Level 0

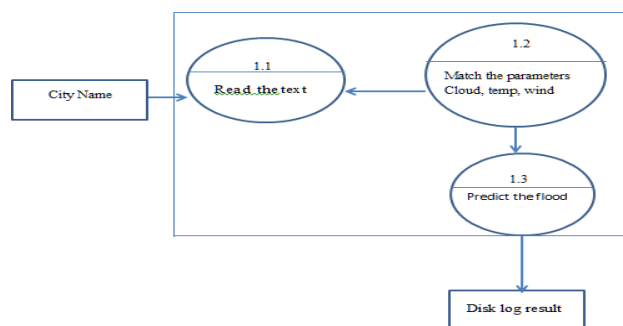


Figure 3 Data Flow Diagram Level 1

Figure 3 depicts how entering a city name and factors like cloud, temperature, and wind cause the system to interpret the data and forecast a flood.

The viewer may examine the map and accompanying statistics, including the weather prediction and the likelihood of flooding, as shown in Figure 4. By taking into account factors like rainfall, GLOFAS, and weather forecasts, the data is associated, and by doing so, floods are forecasted and the pertinent data and flood predictions are updated.

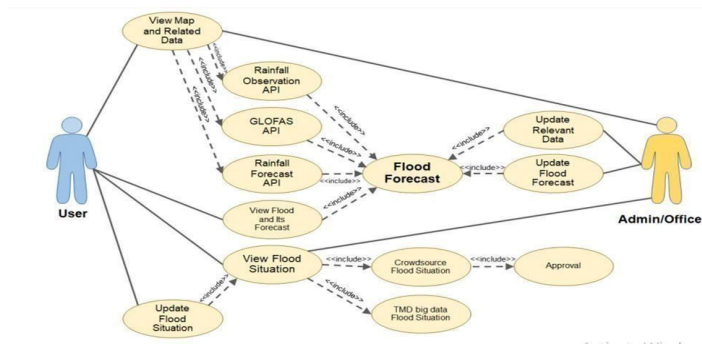


Figure 4 Use Case Diagram

III. METHODS

A. Predictionmodule

Step1:Startpredictionmodel

Step2:Usersshouldselect astaterequiredtoenterthestatenam

Step3:Itcheckstheall-weatherandparameters

Step4:Itdisplaysafe/unsafeforthefloodandalsodisplaystemperature, humidity,cloud, parameters

Step5:Stop

B. Usermodule

Step1:Startusermodel

Step2:viewmaprelateddata

Step3:Ifcity==enteredcityname

If true

Viewfloodstatues

Step4:floodforecastinginformation

Step5:endifcondition

Step6:Stop

C. Data Integration Module

Input:{Cityname}

Output:{Viewmaprelateddata}

Begin

Step1:Addcityname

Step2:connectedwith mapdata

Step3:checkstheweatherdata

Step4:Showstheweatherrelateddata

End

D. Data Processing Module

Input:{City name} **Output:**floodresult

Begin

Step1:Readcityname

Step2:Preprocessquery

Step3:Checksthecitynameinmap

Step4:ApplyRF

Step5:Findfloodrelateddata

Step6:returnsresult

END

One of the worst and most common natural catastrophes in the world is flooding. Millions of people lose their homes and means of support each year, and thousands of others perish in floods. Some of the world's deadliest floods have killed thousands, if not millions, of people in many states surrounding the Yellow River Valley in India and China.

IV. OPERATION PERFORMED

There are three modules in the Novel Approached Flood Forecasting System Using ML and BD. They are:-

- Plots
- InteractiveHeat maps
- FloodPredictions

A. Plots

With consideration for variables including precipitation, wind speed, humidity, temperature, cloud cover, and past data history, the visualisations depict flood forecasts, damage projections, and heavy rainfall predictions over India. Plots of flood forecast, precipitation analysis, and damage analysis are available for viewing.

B. Interactive Heat maps

Precipitation, wind speed, humidity, temperature, cloud cover, and past data history are all taken into account when predicting floods, damage, and severe rains across India using heat maps.

C. Flood prediction

The location of the impending flood is indicated by red dots in the plot below, which was generated using machine learning.

Precipitation

The nation's current precipitation data are displayed in the graphic below. More precipitation is indicated by the bigger bubbles as shown in Figure 5.

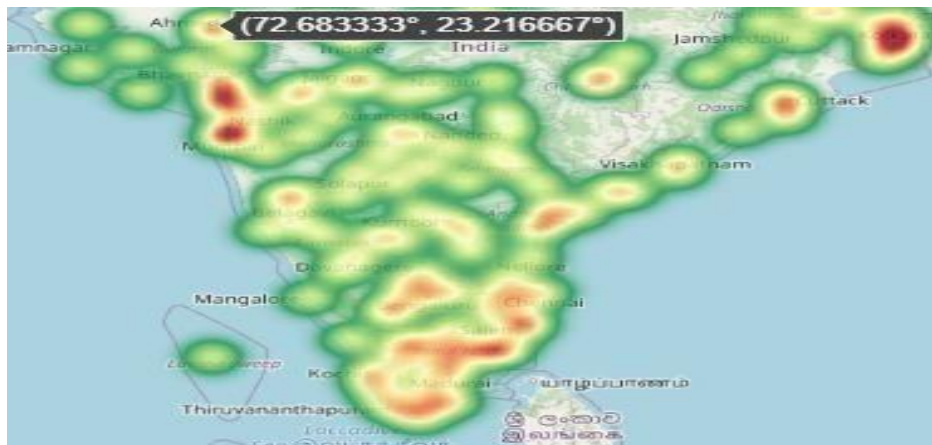


Figure 5 Precipitation

Damage Analysis

Based on the forecasted flood risk, the cost and damage calculations shown below. The heatmap's colour scale shows the amount of anticipated financial harm, calculated in INRas shown in Figure 6.

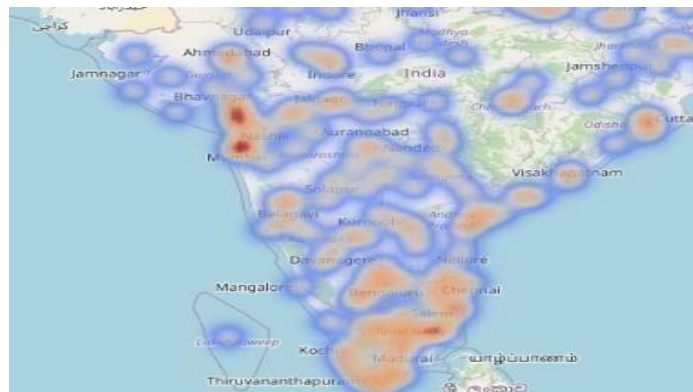


Figure 6 Damage analysis

Flood Prediction

The darker areas in the map below represent the areas where a flood is most likely to occur given the current environmental conditions red spots as shown in Figure 7.



Figure 7 MLpoweredprediction

V. RESULTS& DISCUSSION

The recommended technique used the responsive web design paradigm while creating the website in order to provide the greatest degree of adaptability and crowdsourced[10] coverage. Once implemented, the web application may work with a wide range of devices, including mobile phones, tablet computers, personal and portable computers, and PCs with varied screen sizes and resolutions. The Performance Comparison is shown in Figure 8 and flood prediction of a city is shown in Figure 9.

The Bootstrap platform made it possible for the CSS and JavaScript that were being shown to automatically align and change the layout of components and controls in order to maintain a uniform appearance and so give a positive user experience (UX). Additionally, the cost of development was kept to a minimum since attention was given to key features and editable content rather than platform-specific changes in frontend interfaces and layouts[11].

Even while some of the data used in this investigation, including GLOFAS, did not naturally have high spatial resolution, they were reliable. For example, the cumulative precipitation and its likelihood at various levels matched the actual occurrence. Their API system was dependable as well. The suggested system modified these attributes[12]. By adding crowdsourcing elements and other, more specific layers to the ML framework, their resolution flaws were fixed. Utilizing the Internet of Things (IoT) to measure accurate meteorological data with chosen coverage is one way to address this issue.



Figure 8 Performance Comparison

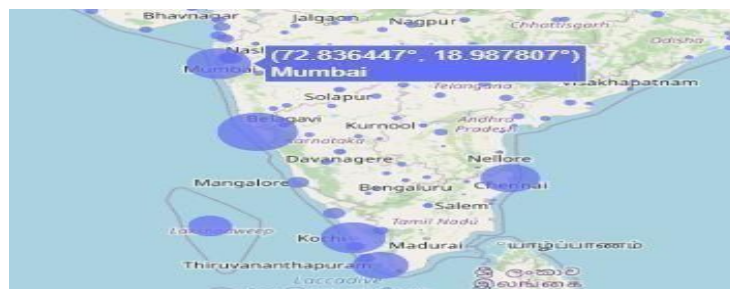


Figure 9Flood predictionofacity

TABLE I. ACCURACY OF FLOOD FORECASTING IS EVALUATED USING SEVERAL MACHINE LEARNING TECHNIQUES

S. No	ML	Correctly Forecasted	Kappa	MAE	RMSE
1	Decision Tree	95.33%	0.77	0.02	0.14
2	Random Forest	96.67%	0.84	0.02	0.10
3	Naïve Bayes	88.00%	0.59	0.61	0.24
4	Artificial Neural Network(MLP)	97.83%	0.89	0.01	0.10
5	Artificial Neural Network(RBF)	95.50%	0.80	0.02	0.14
6	Support Vector Machine	96.67%	0.84	0.01	0.12
7	Fuzzy Logic	95.67%	0.79	0.02	0.13

VI. CONCLUSIONS

Based on the suggested approach, a unique distributed flood forecasting system incorporates meteorological, hydrological, geospatial, and crowdsourcing data. Big data from well-known agencies was obtained using a variety of cross-platform APIs. On the basis of these data, which cutting-edge ML techniques learnt, forecasting was done [16]. They included fuzzy logics, decision trees, RF, Naive Bayes, MLP and RBF ANN, SVM, and naive Bayes. The technology could estimate flood occurrences very precisely, according to evaluation findings on the tested locations. Through interactive, responsive graphical user interfaces that work on a variety of computer platforms, including mobile ones, this technology also improved user experience. This benefit encouraged more in an effective way. In the future, the technology may be easily included into current flood management plans, such as those run by nonprofit or governmental groups. Additionally, the system's distributed architecture allows it to reach a larger audience, making it a useful tool for informing people about the disaster's state and progression (particularly the flood victims). In the future, [13] the technology may be easily included into current flood management plans, such as those run by nonprofit or governmental groups. Additionally, the system's distributed architecture allows it to reach a larger audience, making it a useful tool for informing people about the disaster's state and progression (particularly the flood victims).

REFERENCES

- [1] Supattra Puttinaovarat, Paramate Horkaew, Flood Forecasting System Based on Integrated Big and Crowdsourced Data by Using Machine Learning Techniques Jan 2020,
- [2] Thorndahl, S.; Einfalt, T.; Willems, P.; Ellerbæk Nielsen, J.; Ten Veldhuis, M.C.; Arnbjerg-Nielsen, K.; Rasmussen, M.R.; Molnar, P. Weather radar rainfall data in urban hydrology. *Hydrol. Earth Syst. Sci.* 2017, 21, 1359–1380
- [3] Scott Miau and Wei-Hsiung River Flooding Forecasting and Anomaly Detection Based on Deep Learning 30 Oct 2020,
- [4] Sarawut Jamrusri, Yuji Toda. Simulating past severe flood events to evaluate the effectiveness of nonstructural flood countermeasures in the upper Chaophraya river basin, Thailand. *Journal of Hydrology Regional Studies*, 2017.
- [5] A. Mosavi, P. Ozturk, and K.-W. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, Oct. 2018.
- [6] S. Ghosh, S. Karmakar, A. Saha, M. P. Mohanty, S. Ali, S. K. Raju, and P. L. N. Murty, "Development of India's first integrated expert urban flood forecasting system for Chennai," *Current Sci.*, vol. 117, no. 5, pp. 741–745, 2019
- [7] L.-C. Chang, F.-J. Chang, S.-N. Yang, I.-F. Kao, Y.-Y. Ku, C.-L. Kuo, and I. Amin, "Building an intelligent hydroinformatics integration platform for regional flood inundation warning systems," *Water*, vol. 11, no. 1, p. 9, Dec. 2018.
- [8] Q. Hu, Z. Li, L. Wang, Y. Huang, Y. Wang, and L. Li, "Rainfall spatial estimations: A review from spatial interpolation to multi-source data merging," *Water*, vol. 11, no. 3, p. 579, Mar. 2019.
- [9] N. Belabid, F. Zhao, L. Brocca, Y. Huang, and Y. Tan, "Near-real-time flood forecasting based on satellite precipitation products," *Remote Sens.*, vol. 11, no. 3, p. 252, Jan. 2019.
- [10] S. K. Jain, P. Mani, S. K. Jain, P. Prakash, V. P. Singh, D. Tullos, S. Kumar, S. P. Agarwal, and A. P. Dimri, "A Brief review of flood forecasting techniques and their applications," *Int. J. River Basin Manage.*, vol. 16, no. 3, pp. 329–344, Jul. 2018.
- [11] Ajay N and T N Anitha, "supply of agriculture product by ensuring quality through block chain technology" in *IJRTER*, vol. 4, no. 6, pp. 2455–2461, 2018
- [12] A. Foehn, J. G. Hernández, B. Schaefer, and G. De Cesare, "Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments," *J. Hydrol.*, vol. 563, pp. 1092–1110, Aug. 2018.
- [13] L. Al-erfi, M. Berenguer, V. Knechtel, K. Liechti, D. Sempere-Torres, and M. Zappa, "Flash flood forecasting based on rainfall thresholds," in *Hand- book of Hydrometeorological Ensemble Forecasting*. Berlin, Germany: Springer-Verlag, 2019, pp. 1223–1260.

- [14] F. Dottori, M. Kalas, P. Salamon, A. Bianchi, L. Al_eri, and L. Feyen, "An operational procedure for rapid _ood risk assessment in Europe," *Nat. Hazards Earth Syst. Sci.*, vol. 17, no. 7, pp. 1111_1126, Jul. 2017.
- [15] J. A. Pollard, T. Spencer, and S. Jude, "Big data approaches for coastal flood risk assessment and emergency response," *WIREs Climate Change*, vol. 9, no. 5, p. e543, Sep. 2018.
- [16] Ajay N, H S Mohan, Shwetha B V, Shrihari M R, T N Anitha, " Access Control Framework in the Cloud based on Multiblockchain with light privacy protection" in IEEE, pp. 1-5, 2022.