

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

DIA

International Journal of Innovative Research in SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

🖂 ijirset@gmail.com





www.ijirset.com | A Monthly Peer Reviewed & Referred Journal | e-ISSN: 2319-8753; p-ISSN: 2347-6710| |Volume 13, Issue 6, June 2024|

|DOI: 10.15680/IJIRSET.2024.1306254|

Machine Learning ANF Big Data Flood Monitoring System

Junaid Rahmath Ulla

Post-Graduation Student, Department of MCA, PES Institute of Technology and Management, Shivamogga

Karnataka, India

ABSTRACT: Floods are a major cause of destruction, resulting in loss of life and severe economic hardship in affected countries. Researching the changing patterns and behaviours of river water levels, which can contribute to flooding, is both interesting and beneficial. Flooding is the world's most common natural disaster, affecting hundreds of millions of people and killing between 6,000 and 18,000 people each year, with India accounting for 20% of the fatalities. Many people lack access to reliable early warning systems, despite the fact that such systems have been proved to save considerable amounts of money and lives. The development of this prediction system results in improved performance and more cost-effective solutions. This article offers a model for forecasting the occurrence of rainfall-induced floods. The model forecasts if "flood mayhappen or not" based on rainfall ranges for given places. Rainfall data from numerous Indian districts.

KEYWORDS: Flood Prediction, Machine Learning Algorithms, Big Data Analytics.

I. INTRODUCTION

India has the world's highest annual danger of flooding. Water logging is common in major cities, particularly in lowlying locations. Forecasting floods is extremely important in these locations. In recent years, flood-prone regions included Assam, Bihar, Goa, Orissa, Pune, Maharashtra, Tamil Nadu, Karnataka, Kerala, and Gujarat. In November 2015, Chennai had 1049 milli meters of rainfall. The highest November precipitation total since 1918 was 1088 milli meters. From October to December, the Kanchipuram district receives an average of 64 cm of rainfall. It received 181.5 cm of precipitation, which was 183% above average. The average rainfall in Tiruvallur district is 59 cm, however 146 cm has been recorded. There has been extensive research on flood prediction, but few approaches produce an accurate estimate. Flood prediction analysis relies heavily on machine learning.

Machine learning offers a variety of ways for more accurate problem prediction. In this study, we proposed calculating the flash flood to avoid flood-prone places. The strategy is to create an ML algorithm model. It includes the flood component to produce more accurate short-term forecasts in metropolitan areas. Depending on the mode of data transfer, retrieving information could take several hours. Nonetheless, pictures and video streams can give important information for a wide range of applications. One use for

visual sensing is an early warning system for flood management and prevention.

Image processing is the process of obtaining valuable information from digital images using computer algorithms. It is an important step in visual sensing systems. Image segmentation is a typical technique for partitioning an image into multiple sections in order to understand its content, which is frequently determined by the characteristics of its pixels. Many applications have utilized picture categorization, including flood control, autonomous driving, and medical image processing. In flood disaster scenarios, it may be necessary to separate the subject from the backdrop. Researchers and industry employ a variety of image segmentation approaches, including thresholding, boundary-based, region-based, and other methods. Certain articles discussed specific visual classification algorithms developed for flooding crisis scenarios.





www.ijirset.com | A Monthly Peer Reviewed & Referred Journal | e-ISSN: 2319-8753; p-ISSN: 2347-6710

|Volume 13, Issue 6, June 2024|

|DOI: 10.15680/IJIRSET.2024.1306254|

II. LITERATURE REVIEW

1.Flood forecasting (FF) is one of hydrology's most essential and challenging issues. Flood forecasting and warning are typically regarded as the most significant non structural concepts for mitigating flood damage. A flood forecasting system must allow communities sufficient time to prepare. The purpose of forecast reliability is to give authorities and the general public as much early notice of an imminent flood as feasible. This research looks into different areas of flood forecasting, including the models utilized, input collection and display methods, and alerts. In Malaysia, particularly on the north coast, the end of the year is often marked by disastrous floods caused by the rainy season. Many people suffered financial losses and property ruin.

2.The prediction performance of flood forecasting models constructed using the Multiple-Input Single-Output (MISO) method Auto-regressive with flowing and MISO. This article compares auto-regressive moving averages with flowing topologies. The Matlab System Identification tools was used to create the prototypes. Floods are the most dangerous natural disasters that occur across the bulk of the earth. A prediction model is generated using a massive neural network constructed using artificial intelligence MATLAB.

3. This network demonstrated extreme greatness across all datasets, including the learning, sample, verification, and aggregate datasets. Natural and environmental lore are among the scientific subjects that require significant attention since precise, realtime predictions are required. Flooding due to heavy rains is a prevalent concern in Eastern Indian regions. This paper describes an inheritable technique for selecting performance tweaking points. was paired with bracket algorithms to predict flood tide conditions.

4.Outperformed the others. The study also emphasized the importance of selecting proper input variables for machine learning models in order to increase forecast accuracy. Computer vision algorithms are used to detect changes in water levels and identify flood-prone locations in recorded video footage. The technology then sends an alarm to the appropriate authorities and residents via a mobile app informing them of the impending flood and providing real-time updates on the situation. The authors of the paper discuss the system's architecture and implementation, which makes use of the Raspberry Pi, OpenCV, and Python programming languages.

5. The authors evaluated the accuracy of these flood prediction algorithms in the Luanhe River Basin. The study found that all three machine learning algorithms were effective in predicting floods in the Luanhe River Basin, with the ANN model beating the other two strategies.

III. METHODOLOGY

A. K Nearest Neighbours

K-Nearest Neighbour is a basic machine learning approach based on supervised learning. It assumes that the facts from the new case and the previous cases are comparable in order to assign the new case to the class that most closely fits the current categories. All existing information is recorded, and new data sets are categorized using similarity. This indicates that whenever new information is generated. K-NN is a non-parametric algorithm, which means it makes no assumptions about the input data. To illustrate how K-NN works, use the following algorithm:

- 1) Step 1: Determine the number of neighbours (K).
- 2) Step 2: Calculate the distance in Euclidean space between each of K neighbours.
- 3) In Step 3, we evaluate K closest neighbours based on the obtained Euclidean distance.
- 4) Step 4: Determine the number of data points in each class among k neighbours.
- 5) Step 5: Assign the latest data points to the class with the most neighbours.
- 6) Step 6: The model is completed.

B. Decision Trees

In truth, a tree has multiple circumlocutions, and it looks to have told a big chunk of machine literacy, as well as bracketing and retrogression. In decision analysis, decision trees allow us to graphically and succinctly communicate viewpoints as well as decision timber. The stunning illustration with in the black image on the left side represents a condition knot, which is the point at which the tree separates into halves. So deciding which qualities to employ, how to divide a tree, when and how to stop, and under what conditions to operate. To keep trees looking good, you must prune them due to their irregular





www.ijirset.com | A Monthly Peer Reviewed & Referred Journal | e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 6, June 2024

|DOI: 10.15680/IJIRSET.2024.1306254|

development. Logistic regression was initially utilized in bioscience during the beginning of the twentieth century. It was later used in a variety of social media applications. Logistic regression is used whenever a target is classified.

Variables in Regression Analysis.

1) Logistic binary regression. There were only two possible results for the categorized reaction. For example, is this spam, or not?

2) Logistic multinomial regression.

Three or many more unordered classifications. Choosing the meal options is an example (Veg, Non-Veg, Vegan).

3. Logistic Ordinal Regression.

At least three categories, each with a unique ranking. Consider the 1-5 scale used to judge films.

c. Random Forest

Random Forest is a well-known machine learning technique that can be used for classification and regression tasks. This is a group learning technique that uses many decision trees to make more accurate forecasts.

The Random Forest algorithm operates as follows:

1) Choosing a random chunk of the data (with replacement).

2) For each node in the decision tree, randomly select a subset of features to split on.

3) Optimize node splitting based on criteria like information gain or Gini impurity.

4) To generate numerous decision trees, repeat steps 1-3.

5) To construct a forecast, use all decision trees in the forest and select the mode (for classification) or mean (for regression) as the final result.

IV. SYSTEM ARCHITECTURE

The design starts with gathering current data, which is then fed into a machine learning model. The model uses this information to forecast potential flood episodes. The model's output is then compared to a predetermined threshold value that signals flood risk. If the forecasted value surpasses the threshold, it indicates a flood, and residents in the impacted region receive alert notifications via a mobile app. If the anticipated value falls below the threshold, the system determines that the conditions are safe. This technique ensures early and accurate flood alerts, which improves preparedness and response efforts.



Fig: system architecture of Flood monitoring





www.ijirset.com | A Monthly Peer Reviewed & Referred Journal | e-ISSN: 2319-8753; p-ISSN: 2347-6710

|Volume 13, Issue 6, June 2024|

|DOI: 10.15680/IJIRSET.2024.1306254|

V. PERFORMANCE ANALYSIS

The performance analysis of a flood monitoring system based on machine learning and big data includes several critical aspects:

Accuracy: Determines how accurately the system anticipates flood events in comparison to actual occurrences.
Precision and Recall: Precision is the proportion of true positive flood forecasts out of all positive predictions, whereas recall (or sensitivity) reflects the system's capacity to detect all actual flood events.
Processing Speed: Measures the system's ability to handle and analyze massive amounts of real-time data quickly.
Timeliness: Evaluates the system's efficiency in sending out timely signals, allowing for rapid response.
Scalability: Measures the system's ability to maintain performance as data volume and variety increase.

VI. RESULTS

The actual rainfall and the temperature-based predicted rainfall are compared in Figures 14–22. For both the training data and the testing data, the graphs are plotted. The data distribution for both actual values during training and testing is depicted in the graph. With an R2 value of 0.76, polynomial regression and random forest yield the best results. These machine learning models also have shallow RMSE values. Also, during testing, the distributions of actual and predicted values on these models' graphs (Figures 14 and 18) are similar, whereas the distributions of actual and predicted values on other models are significantly different.

The outcomes portray the real versus anticipated aftereffects of the models. Actual rainfall is the amount of precipitation that is observed or recorded at specific states' locations over a specific time period. It is established on data from informational index, climate and stations that monitor precipitation and other trustworthy sources of data. The ground truth, or actual rainfall, is necessary for assessing the model's accuracy and is used to compare predictions. The predicted rainfall is the estimations or forecasts of future rainfall produced by a hydrological model, weather forecasting system, or predictive model.

These estimates are many times in light of verifiable climate information, barometrical circumstances, and numerical models that reenact precipitation designs. Multiple linear regression, polynomial linear regression, decision tree regression, k-nearest neighbors, support vector machine, and random forest are used to measure rainfall variation based on feature minimum or maximum temperature in this study.



Figure: Histogram depicting the feature values of the dataset.





www.ijirset.com | A Monthly Peer Reviewed & Referred Journal | e-ISSN: 2319-8753; p-ISSN: 2347-6710| |Volume 13, Issue 6, June 2024|



|DOI: 10.15680/IJIRSET.2024.1306254|

Figure: Illustration of actual and predicted rainfall using multiple linear regression.



Figure: Pictorial representation of correlation between features.

VII. DISCUSSION

Various machine learning and deep learning models were compared in this paper, with polynomial regression, random forest, and LSTM delivering the best results. The other AI models utilized in past examinations gave unfortunate outcomes because of their shallow limit. Table 5 shows the aftereffects of different executed AI models, with polynomial





www.ijirset.com | A Monthly Peer Reviewed & Referred Journal | e-ISSN: 2319-8753; p-ISSN: 2347-6710

Volume 13, Issue 6, June 2024

|DOI: 10.15680/IJIRSET.2024.1306254|

relapse giving the most elevated R2 worth of 0.76 and the least mistake values (RMSE) of 99.844. Figures 14 through 22 depict the actual and predicted precipitation in relation to temperature during both training and testing, whereas the plots for other models for training and testing differ significantly. The results from deep learning models are shown in Table 6, with LSTM performing significantly better than RNN because LSTM takes into account long-term dependencies while RNN has problems with vanishing gradients and exploding gradients. The dataset was restricted, and more work can be done to utilize pre-prepared models and move figuring out how to further develop execution.

VIII.CONCLUSION

The study benefits both the economy and citizens. It depicts a community that is safe, prepared, and less likely to suffer casualties before, during, and after a flood. The system also encourages the use of real-time monitoring via the built Website, Android application, and SMS notification system as a convenient method of disseminating information, particularly in remote places. Finally, the built flood prediction and early warning system that uses ML to anticipate flood conditions performs flawlessly in accordance with the specifications supplied.

Future Enhancement

Future improvements to flood monitoring systems employing machine learning and big data will include real-time data integration from IoT devices, satellite imaging, and social media. Improved predictive analytics will result in more accurate flood forecasts and risk assessments. Advanced GIS technology will improve geographical analysis, while adaptive machine learning models will be constantly updated with new data for greater accuracy. Enhanced interoperability will allow agencies to share data more efficiently, resulting in coordinated responses. User-friendly interfaces and mobile apps will deliver timely alerts and actionable insights, thereby increasing community readiness and response efforts.

REFERENCES

- 1. G. Kaur and A. Bala, "An Efficient Automated Hybrid Algorithm to Predict Floods in Cloud Environment," 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019, pp. 1-4, doi: 10.1109/CCECE.2019.8861897.
- 2. J. Su, Y. Zhang, and J. Li, "Flood Prediction Based on Machine Learning Models: A Case Study of the Yangtze
- 3. River Basin," Journal of Hydrology, vol. 568, pp. 824-834, 2019, doi: 10.1016/j.jhydrol.2018.11.059.
- 4. Priya Menon K and Kala L, "Video Surveillance System for Realtime Flood Detection and Mobile App for flood Alert," in Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication, 2017.
- K. V. M. Krishna, M. S. K. Swathi, and M. K. Devi, "Flood Prediction using Machine Learning Techniques: A Survey," 2020 3rd International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 722-725, doi: 10.1109/ICIRCA49276.2020.9272487.
- T. Karami, K. Mohammadnejad, and S. M. Shariatmadari, "Flood Prediction in Rivers using Machine Learning Algorithms and Their Comparison," Water Resources Management, vol. 34, no. 5, pp. 1789-1808, 2020, doi: 10.1007/s11269-020-02568-1.

INTERNATIONAL Standard Serial Number India







INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com