iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51528

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Flood Prediction using Machine Learning

Asst. Prof. Anil Kumar Ambore[1], T. Sri Sai Charan[2], U. Rohit Reddy[3], T. Samara Simha Reddy[4], Tarun.G[5]
[1, 2, 3, 4, 5]*School of Computer science &Engineering REVA University Bengaluru, India*

Abstract: ***The most frequent natural disaster in the world, flooding affects hundreds of millions of people and kills between6,000 and 18,000 people annually, with 20% of those deaths occurring in India. Several people lack access to reliable early warning systems, despite the fact that those systems already exists demonstrated may avoid an large portion of economy anddeath loss. Improved performance and cost-effective solutions are offered by this prediction system's development. In order toforecast the occurrence of floods brought on by rainfall, a prediction model is created in this article. Based on the rainfall range for certain places, the model forecasts if "flood mayhappen or not". information about rainfall in Indian districts.***
Keywords: Machine Learning, Flood Prediction, Classification,Disaster Management

I. INTRODUCTION

India is the country with the highest annual risk of flooding in the entire world. In large cities, low-lying areas are where water logging typically happens. Forecasting floods is so crucial in these areas. There were numerous flood-proneregions in recent years, including Assam, Bihar, Goa, Orissa,Pune, Maharashtra, Tamil Nadu, Karnataka, Kerala, and Gujarat.

In November 2015, Chennai saw rainfall of 1049 millimeters (mm). The best November precipitation total since 1918 was 1088 mm. In the Kanchipuram district, 64 cmof rain falls on average between October and December. It received the most precipitation, 181.5 cm, which is 183% more than average. The average rainfall in the Tiruvallur district is 59 cm, but 146 cm was recorded.

There has been a lot of research into flood prediction, but notmany methods provide an accurate estimate. Machine Learning is heavily used in flood prediction analysis (ML). Machine learning provides a wide range of approaches for more precise problem prediction. In this paper, we recommended estimating the flash flood to prevent flood- prone areas. The strategy is to establish the ML algorithm model. It incorporates the flood factor to provide more accurate short-term predictions in urban areas. Depending on the method of data transmission, retrieval of information could take hours. Still, images and video streams can provide useful information for a variety of applications. One of the applications for the technique of visual sensing is an early alert system for controlling and preventing flooding. Image processing is the process of extracting useful information from digital images using computer algorithms, which is a critical procedure in visual sensing systems.

Image segmentation is commonly used to partition an imageinto several regions to understand its content, which frequently depend on its pixel's characteristics. Many applications have made use of image classification., incorporating flood control, autonomous driving, and medical image processing. It can involve separating the subject from the backdrop in flood disaster scenarios. Image segmentation techniques used by researchers and industry today include thresholding, boundary-based, region-based, and alternative methods. Specific visual classification methods created for flood disaster scenarios were covered incertain articles.

II. LITERATURE REVIEW

Recent floods in several parts of southern India caused significant harm to both persons and property. Flooding is one of the most severe natural disasters, and it takes time to resume normal life. One of the various technologies used after a disaster to expedite rescue efforts and lessen damage is drones. Many algorithms are required for the automatic studying of aerial and remote sensing pictures. With a high degree of accuracy, Support Vector Machine (SVM) [1] andk-means clustering correctly classified roughly 92% of the pictures to identify flooded areas. Several kernel functions are changed in order to assess the performance of SVM.

Flood forecasting (FF) [2] is one of hydrology's most important and difficult problems. Flood forecasting and warning are widely acknowledged as the most important non- structural terms for reducing flood damage. A flood forecast system must give communities enough time to prepare. The goal of forecast reliability is to provide authorities and the general public with as much advance warning of an impending flood as possible. This paper investigates various aspects of flood forecasting, such as the models used, input collection and display techniques, and warnings.

In Malaysia, particularly on north shore, the end of the year is typically marked by devastating floods brought on by rainy season. Numerous people experienced revenue damage and property destruction.

©IJRASET: All Rights are Reserved

The prediction performances of models for flood forecasting developed using the Multiple-Input Single-Output (MISO) [3] Auto regressive with flowed and MISO Auto regressive Moving Average with flowed structures were compared in this article. The Matlab System Identification toolkit was used to generate the prototypes.

The most dangerous natural events that occur on a majority of the planet is floods. A prediction model is developed usinga mega neural network constructed using artificial intelligence and MATLAB [4]. This network showed very strong awesomeness across all datasets, including the learning, sample, verification, and aggregate datasets. Natural and environmental lores are among the scientific disciplines that admit considerable focus because precise in real-time prognostications are needed. Floods from heavy rains is a common hazard in Eastern Indian countries. An performance tuning point selection inheritable technique is shown in this study. was combined with bracket algorithms to prognosticate flood tide circumstances. The experimental results show that in terms of delicacy and total prosecution time, the GA- SVM algorithm [5] outperforms other mongrelalgorithms. Eventually, the results are vindicated.

The study's findings revealed that all four machine learning models performed well in predicting flood events, with the RF model [6] outperforming the others. The study also emphasised the significance of choosing appropriate input variables for machine learning models in order to improve their predictive accuracy.

Computer vision algorithms are used to detect changes in water levels and identify flood-prone areas in the captured video footage. The system then sends an alert to the relevantauthorities and citizens via a mobile app [7], notifying them of the potential flood and providing them with real-timeupdates on the situation. The authors of the paper describe thesystem's design and implementation, which includes the use of the Raspberry Pi, OpenCV, and Python programminglanguages.

This article's authors present a study of different machine learning algorithms are utilized to detect floods. They discuss the importance of flood prediction, and the various factors that contribute to floods, such as rainfall, river flow, soil moisture, and topography. The authors then provide a thorough analysis of various machine learning algorithms such as decision tree algorithm, random forests, support vector machines, and neural networks with artificial intelligence,[8] that are used for flood prediction.

In the paper, the authors suggest using algorithms for machine learning. to predict floods in rivers. They compare the performance of Machine learning techniques that use neural networks include the random forest method (RF), artificially generated neural networks, artificial neural integrating the use of support vector machines, and k-nearestneighbour [9]. The authors used data from two rivers in Iran,the Karoon and Dez rivers, to create and evaluate the modelsfor machine learning. To predict flood events, they used a variety of input variables such as rainfall, temperature, and river flow rate.

The authors present a thorough examination of various Big data applications that can leverage machine learning methods. Neural networks that are artificial, supportinghidden markov model, decisions trees, regression trees, and other methods are among them. deep learning models. The paper also covers some of the latest developments in big dataanalytics [10], such as data stream mining and distributed machine learning.

The authors then describe the machine learning techniques used in the study, including [11]. Authors compared the accuracy of these flood prediction techniques in the Luanhe River Basin. The results of the study showed that all three machine learning techniques performed well in predicting floods in the Luanhe River Basin, with the ANN model outperforming the other two techniques.

III. METHODOLOGY/ALGORITHMS

A. XGBoost

XGBoost is an recently algorithm in Kaggle competitions fororganized data and used machine learning. Gradient-boosteddecision trees are created in this method with excellent performance. An gradient-boosting machine-learning technique built upondecision trees is termed as XGBoost. however, decision tree- based algorithms are currently considered best-in-class for small-to-medium structured/tabular data.

Bagging: Imagine an interview panel with each memberhaving a vote rather than a single interviewer. The techniqueof integrating the opinions of all interviewers for the ultimatejudgment through a democratic voting procedure is known asbagging, often known as bootstrap aggregation. Heterogeneous tree techniques include Gradient Boosting Machines and XGBoost that boost weak learners using the gradient descent architecture (CARTs in general). On the other hand, improves the base GBM framework via system optimization and algorithmic improvements.

B. K Nearest Neighbors

A basic machine learning approach that makes use ofsupervised learning is named as K-Nearest Neighbor.

Inorder to allocate the new case to the class that matches mostclosely to the existing categories, it makes the assumptionthat the new case's data and existing cases are comparable. All the existing information is recorded and fresh data setsare categorized using similarity. This implies that wheneverfresh information is produced.
K-NN is a non-parametric algorithm, meaning it makes noassumptions about the underlying data.
To explain how K-NN works, consider the followingalgorithm:

1) Step1: Count the number of neighbours K.
2) Step2: determine the distance in euclidean space amongeach of K neighbours.
3) Step3: From the calculated Euclidean distance we evaluateK closest neighbours.
4) Step4: Find no of data within every class among kneighbours.
5) Step5: Allocate the latest data points to the class that has themost neighbours.
6) Step6: Finally the model is completed.

C. *Decision Trees*

An tree is having numerous circumlocutions in reality, and it appears that it told a large section of machine literacy, as well as bracket & retrogression. With decision tree we can visually and clearly describe opinions as well as decision timber in decision analysis. The striking textbook with in the black image on the left side indicates a condition knot, the ground at which the tree divides to section borders.
So Making choices about which characteristics to use how to split a tree, when and how to stop, and what circumstances to utilise. You must prune trees because of their unpredictable growth if you want them to look attractive.

D. *Logistic Regression*

Around the beginning of the twentieth century, logistic regression was used for the first time in the bioscience. Later,It was later used in various social apps. Logistic regression isemployed whenever the target is classified.
Variables in Regression Analysis

1) *Logistic Binary Regression*

There were only 2 feasible outcomes for the categorized reaction. For instance, weather this is spam or not?

2) *Logistic Multinomial Regression*

Three or much more classifications that are not ordered. determining the food choices is an example (Veg, Non-Veg, Vegan)

3) *Logistic Ordinal Regression*

At least three categories, each with a different ranking. Take,for instance, the 1–5 scale used to evaluate films.

E. *Random Forest*

Random Forest is well-known machine-learning algorithm that can be utilized in categorization and regression jobs. This is a technique for group learning which brings together manydecision trees so that it generate more accurate forecasts..
Random Forest algorithm works as follows:

1) choosing a portion of the data as randomly (with replacement).
2) Select a subset of features to consider for splitting at random for each node in the decision tree.
3) Split the node according to the characteristic that, given a certain criterion such as information gain or Gini impurity provides the optimal split.
4) Steps 1-3 should be repeated to create multiple decision trees.
5) When making a prediction, use all of the decision trees in the forest to make a prediction, and then choose the mode (for classification) or mean (for regression) of the predictions as the final prediction.

IV. RESULT

The suggested task would be a method for evaluating the dataset regarding rainfall in order to expect flash floods with greater accuracy using algorithms. This study shows a performance tuning point selection inheritable technique is shown in the table. The below steps demonstrate its suggested model offers an simple and systematic strategy for floodprediction:

1) Step 1: The rainfall dataset is preprocessed.
2) Step 2: The rainfall dataset is randomly divided into testingand training.
3) Step 3: dataset was learned using the xgboost, LogisticRegression, Decision Tree, and KNN algorithms.
4) Step 4: The model is built with the highest accuracy usingthe xgboost and DT algorithm.
5) Step 5: Run the prediction model on the test data andvalidate the results.

Table: Performance analysis

Algorithm	Accuracy
XgBoost	0.9937888198757764
Logistic Regression	0.9875776397515528
Decision Tree	0.9937888198757764
KNN	0.9875776397515528

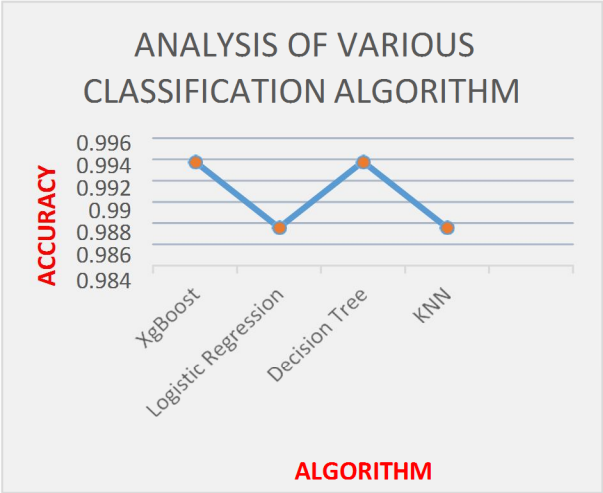


Fig. Analyzing variety categorization of algorithm forflood prediction

V. CONCLUSION

A flood prediction system based on machine learning has thepotential to greatly benefit flood-prone communities. Machine learning algorithms can accurately predict the likelihood and severity of a flood by using historical data andreal-time monitoring, allowing authorities to take preventative measures to minimize the impact on the community.

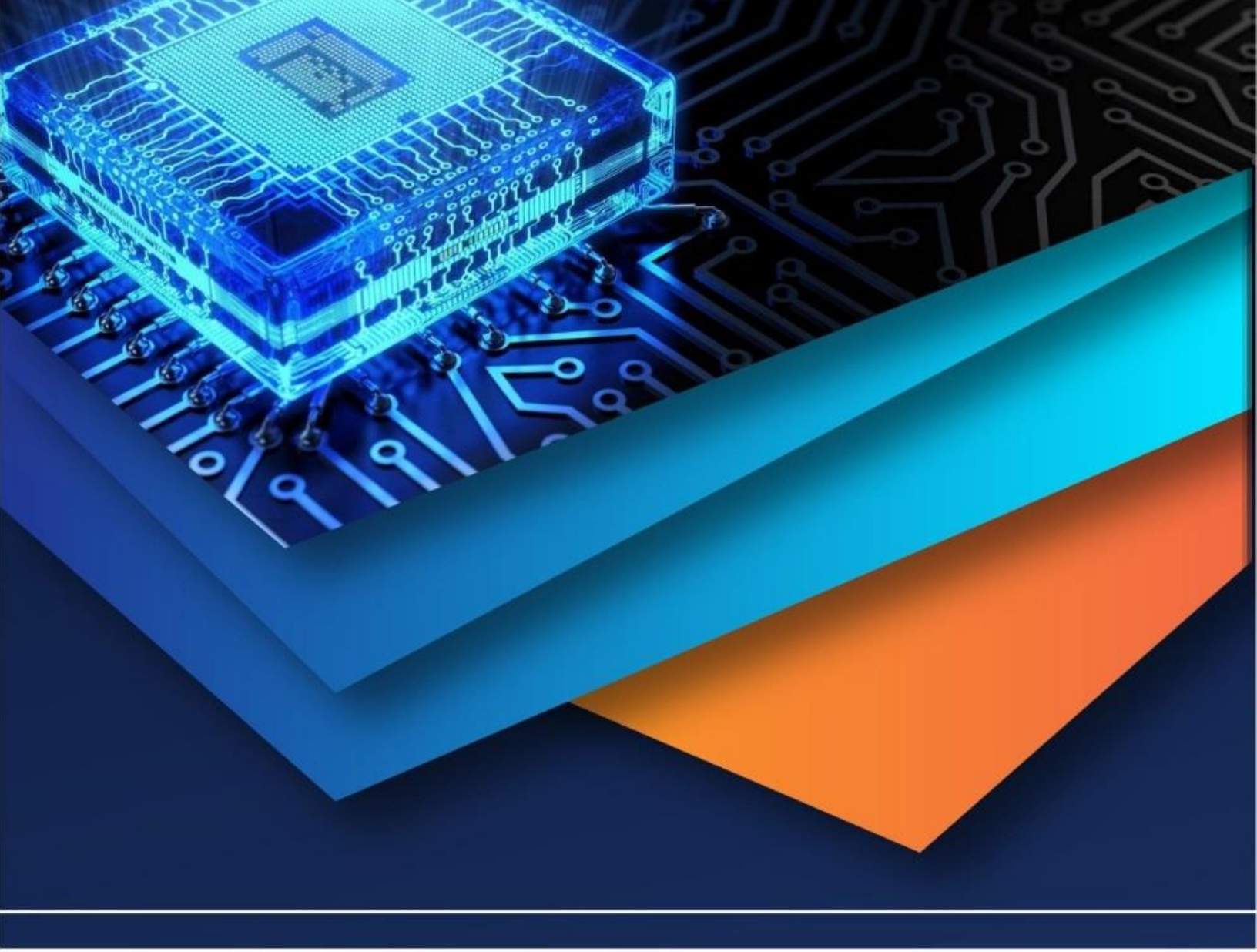
Machine learning models, such as Random Forest, can be trained on large datasets of historical flood data, weather data, topographical information, and other relevant factors to accurately predict the likelihood of future floods. Implementing a flood prediction system based on machine learning can save lives, reduce property damage, and improve emergency response efforts. Overall, the potential benefits of such a system make it a worthwhile investment for any flood- prone community.

REFERENCES

[1] J. Akshya and P. L. K. Priyadarsini, "A Hybrid Machine Learning Approach for Classifying Aerial Images of Flood- Hit Areas," 2019 International Conference on ComputationalIntelligence in Data Science (ICCIDS), 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862138.
[2] A. B. Ranit and P. V. Durge, "Different Techniques of Flood Forecasting and Their Applications," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), 2018, pp. 1-3, doi: 10.1109/RICE.2018.8509058.
[3] F. A. Ruslan, K. Haron, A. M. Samad and R. Adnan, "Multiple Input Single Output (MISO) ARX and ARMAX model of flood prediction system: Case study Pahang," 2017IEEE 13th International Colloquium on Signal Processing &its Applications (CSPA), 2017, pp. 179-184, doi:10.1109/CSPA.2017.8064947.
[4] F. R. G. Cruz, M. G. Binag, M. R. G. Ga and F. A. A. Uy,"Flood Prediction Using Multi-Layer Artificial Neural Network in Monitoring System with Rain Gauge, Water Level, Soil Moisture Sensors," TENCON 2018 - 2018 IEEE Region 10 Conference, 2018, pp. 2499-2503, doi: 10.1109/TENCON.2018.8650387.

[5] 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.
[6] J. Su, Y. Zhang, and J. Li, "Flood Prediction Based on Machine Learning Models: A Case Study of the Yangtze
[7] River Basin," Journal of Hydrology, vol. 568, pp. 824-834, 2019, doi: 10.1016/j.jhydrol.2018.11.059.
[8] 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.
[9] 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.
[10] A. 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.
[11] Vinothini A and Baghavathi priya A, "Survey of Machine Learning Methods for Big Data Applications," in International Conference on
[12] Computational Intelligence in Data Science, 2017.
[13] H. Zhang, Y. Zhang, Z. Liu, and G. Liu, "Flood Prediction Based on Machine Learning Approaches: A Case Study in the Luanhe River Basin, China," Water, vol. 10, no. 8, 2018, doi: 10.3390/w10081051.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)