



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

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 9, September 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Dynamic and Boosting Flood Prediction through Machine Learning

Rashmi M N, Mr. Srinivasulu M\*

Department of Master of Computer Application, University B.D.T College of Engineering, Davanagere,  
Karnataka, India

**ABSTRACT:** Floods have become the most well-known and lethal cataclysmic events of this century. Absence of a successful flood forecasting framework has brought about grave loss of human existence and infrastructure. This has reiterated on the importance of having in place a flood prediction system. This paper looks at developing the most effective flood determining model. AI calculations and a hearty, productive and precise flood expectation framework will give all the fundamental aid and assistance needed to the residents and government. The main contribution of this paper is to demonstrate the state of the art of ML models in flood prediction and to give insight into the most suitable models. In this paper, the literature where ML models were benchmarked through a qualitative analysis of robustness, accuracy, effectiveness, and speed are particularly investigated to provide an extensive overview on the various ML algorithms used in the field. The performance comparison of ML models presents an in-depth understanding of the different techniques within the framework of a comprehensive evaluation and discussion.

**KEYWORDS:** Flood forecasting • Machine learning • Gradient boost

## I. INTRODUCTION

Among the natural disasters, floods are the most destructive, causing massive damage to human life, infrastructure, agriculture, and the socioeconomic system. Governments, therefore, are under pressure to develop reliable and accurate maps of flood risk areas and further plan for sustainable flood risk management focusing on prevention, protection, and preparedness [1]. Flood prediction models are of significant importance for hazard assessment and extreme event management. Brisk liquefying of ice, outlandish rainfall or a burst dam, can overwhelm a river, spreading over the contiguous land. Ocean front flooding happens when a colossal storm or tsunami makes the ocean flood inland. Floods are considered as the most common natural disaster on Earth, second only to the forest fires. The second aspect is the capability of each ML algorithm, which may vary across different types of tasks. This can also be called a “generalization problem”, which indicates how well the trained system can predict cases it was not trained for, i.e., whether it can predict beyond the range of the training dataset. For example, some algorithms may perform well for short-term predictions, but not for long-term predictions. These characteristics of the algorithms need to be clarified with respect to the type and amount of available training data, and the type of prediction task, e.g., water level and streamflow. In this review, we look into examples of the use of various ML algorithms for various types of tasks. At the abstract level, we decided to divide the target tasks into short-terms.

## II. LITERATURE REVIEW

Felix and T. Sasipraba [2] have developed a flood warning system for which data Yovanhas been collected from remote sensing satellites and ground application. The parameters considered are the amount of rainfall and water-level of nearby water bodies. Gradient Boost Algorithm [3] is used to obtain a non-linear relationship between the total sum of rainfall and runoff, thus reducing the mean-squared error. For the datasets which were not a part of the training dataset, the Decision Tree Algorithm is used to predict floods on datasets which were not used in the training dataset. The architecture is divided into three layers which are interconnected. The three layers are the Physical Layer, Network Layer and Application Layer.

Physical Layer collects the data from sensors and the Application Layer is the user interface which is the Mobile/Desktop Application. The historical dataset was collected from the Meteorological Department and Flood Forecasting Commission. At an interval of an hour, the realtime data is collected and sent to the Network Layer using the GSM Module and stored on the Cloud Server to be passed to the Machine Learning Model. Here, it is also worth emphasizing that, in this paper, the prediction lead-time was classified as “short-term” or “long-term”. Although flash floods happen in a short period of time with great destructive power, they can be predicted with either “short-term” or

“long-term” lead times to the actual flood. In fact, this paper is concerned with the lead times instead of the duration or type of flood.

### III. ALGORITHMS

**3.1 Logistic Regression:** Logistic Regression is a machine learning algorithm that predicts the probability of a categorical dependent variable. It is a statistical way of analyzing a set of data that comprises more than one independent variable that determines the outcome. The outcome is then measured with a dichotomous variable. The goal of this algorithm is to find the best model to describe the relationship between a dichotomous characteristic of interest and a set of independent variables. In this algorithm, the dependent variable is a binary variable that contains data coded as 1 or 0. In other words, the logistic regression.

**3.1.2 Support Vector Machines:** SVM uses a classifier that categorizes the data set by setting an optimal hyperplane between data. This classifier is chosen as it is incredibly versatile in the number of different kernel functions that can be applied, and this model can yield a high predictability rate. Support Vector Machine is one of the most popular and widely used clustering algorithms. It belongs to a group of generalized linear classifiers and is considered as an extension of the perceptron. It was developed in the 1990s and continues to be the desired method for a high-performance algorithm with a little tuning.

**3.1.3 K-Nearest Neighbor (KNN):** K-Nearest Neighbor is one of the supervised machine learning algorithms that stores all instances corresponding to training data points in an n-dimensional space. For real-valued data, the algorithm returns the mean of k nearest neighbors, and in case of receiving unknown discrete data, it analyses the closest k number of instances which is saved and returns the most common class as the result of the prediction. In the distance-weighted nearest neighbor algorithm, the contribution of each of the k neighbors is weighed according to their distance, giving higher weight to the closest neighbors. The K-Nearest Neighbor algorithm is a classification algorithm and is robust to noisy data as it averages the k-nearest neighbors. The algorithm first takes a bunch of labeled points and analyses them to learn how to label the other points. Hence, to label a new point, it looks at the closest labeled points to that new point and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point. This algorithm makes predictions about the validation set using the entire training set. Only by searching through the entire training set to find the closest

instances, the new instance is predicted. Closeness is a value that is determined using a proximity measurement across all the features involved.

### 3.2 Data Validation/ Cleaning/Preparing Process:

This process involves variable identification by data shape, data type, and evaluating the missing values, duplicate values. The process of data cleaning varies from dataset to dataset. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making. A sample of data, known as the validation dataset, is held back from training the model, that can be used to make the best use of validation and test datasets when evaluating your models, giving an estimate of model skill while tuning models and procedures.

**3.3 Data Pre-processing:** Pre-processing is the process where the data is transformed prior to feeding it to the algorithm. This is done in order to convert the raw data into a clean dataset for better results after feeding it to the algorithm. As the data gathered are from different sources, and not uniform, it is not feasible for analysis. To achieve better results from the applied model in the Machine Learning method of the data has to be in a proper manner. Some specified Machine Learning models need information in a specified format; for example, the Random Forest algorithm does not support null values. Therefore, to execute random forest algorithms, null values have to be managed from the original raw data set.

**3.4 Exploration Data Analysis of Visualization:** Data Visualization is considered as a very important skill in applied statistics which focus on quantitative descriptions and estimations of data, and machine learning. Data visualisation provides an important suite of tools for gaining a qualitative understanding which can be useful when exploring and getting to know a dataset which can help with identifying patterns, corrupt data, outliers and much more. Data visualization can also be used to demonstrate key relationships in plots and charts that are more visceral.

**3.5 Outlier Detection Process:** Many machine learning algorithms are known to be sensitive to range and distribution of attribute values in the input data. Outliers in those data can mislead and skew the training process of the algorithms which will result in longer training times, less accurate models and ultimately poorer results. Even before the process of training the model, the outliers can result in misleading representations hence resulting in misleading interpretations of collected data. Outliers are capable of skewing the summary distribution of attribute values.

statistics like mean, standard deviation and in plots such as scatter plots, histograms, compressing the body of the data. Outliers are also capable of representing examples of data instances which are relevant to the problem such as anomalies in the case of fraud detection and computer security. It couldn't fit the model on the training data, and this can affect the model to work inaccurately in case of real data. Hence, we must ensure that our model gets the correct patterns from the data, and not receiving too much noise. Cross Validation is a process where we train our model using the subset of the dataset and then evaluating using the complementary subset of the dataset. Cross validation is advantageous as it provides a more accurate estimate of out of sample accuracy and also for its more efficient use of data as every observation is used for both training and testing.

#### IV. PERFORMANCE ANALYSIS METRICS

**True Positive:** It is an outcome where the model correctly predicts the positive class. The outcome is considered as true positive when the system can correctly predict that an incident has indeed occurred. **True Negative:** It is an outcome where the model correctly predicts the negative class. The outcome is considered as true negative when the system can correctly predict that the particular incident has not occurred.

**False Positive:** False Positive is an accuracy measure where the model mispredicts the positive class. The outcome is considered as False Positive when the system cannot correctly predict that the particular incident has occurred.

**False Negative:** False Negative is an accuracy value where the model mispredicts the negative class. The outcome is considered as False Negative when the system cannot correctly predict that the particular incident has not occurred.

**Sensitivity:** Sensitivity is a measure of the proportion of true positive values, that is, the actual number of positive cases that are correctly predicted as positive.

#### V. RESULTS

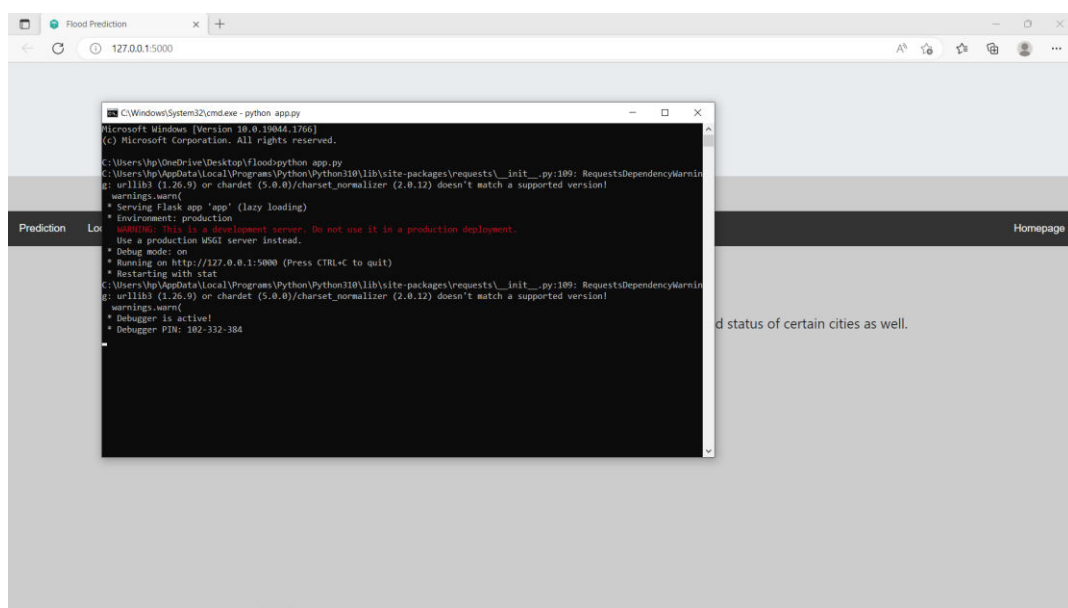


Fig:1 Command Prompts



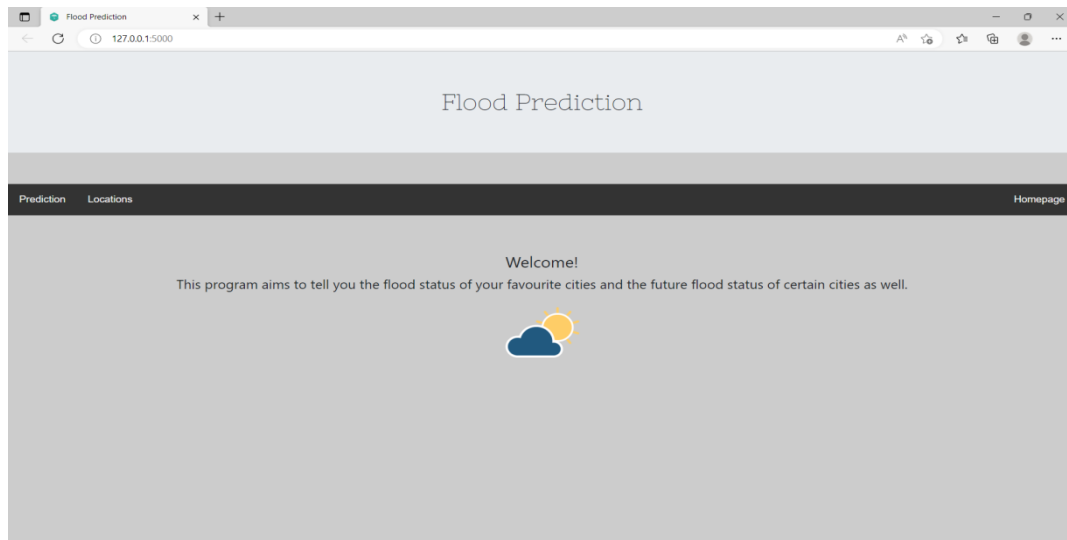


Fig:2 Home page lets welcome the program for flood prediction.

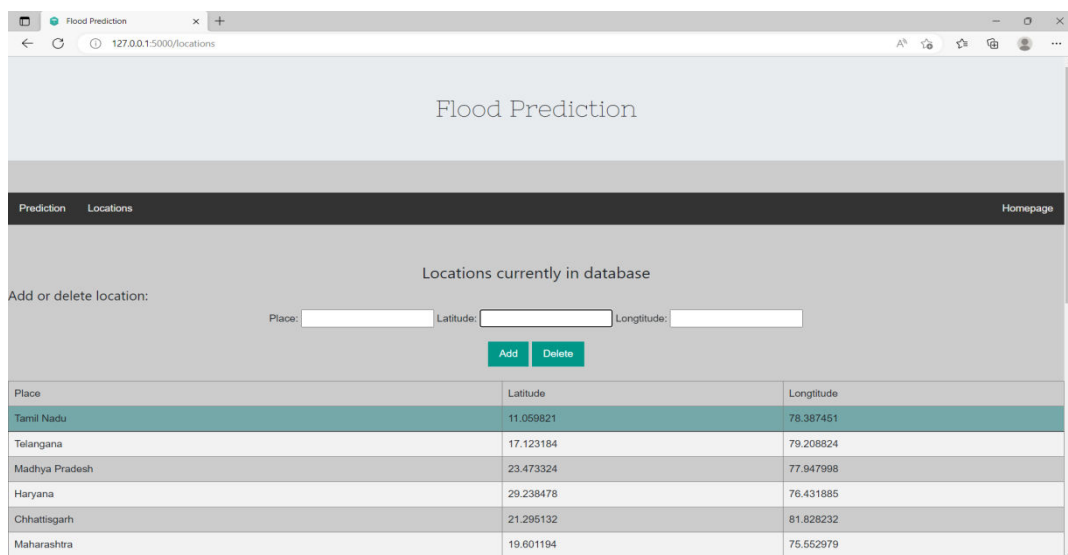


Fig:3 This image shows that the currently location for database in place,latitude and longitude add these three.

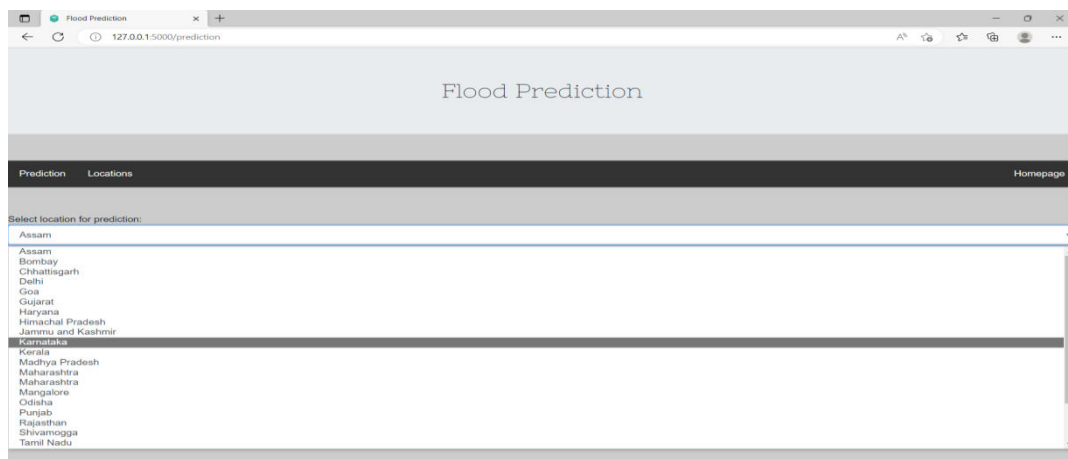


Fig:4 This image shows selected the state.

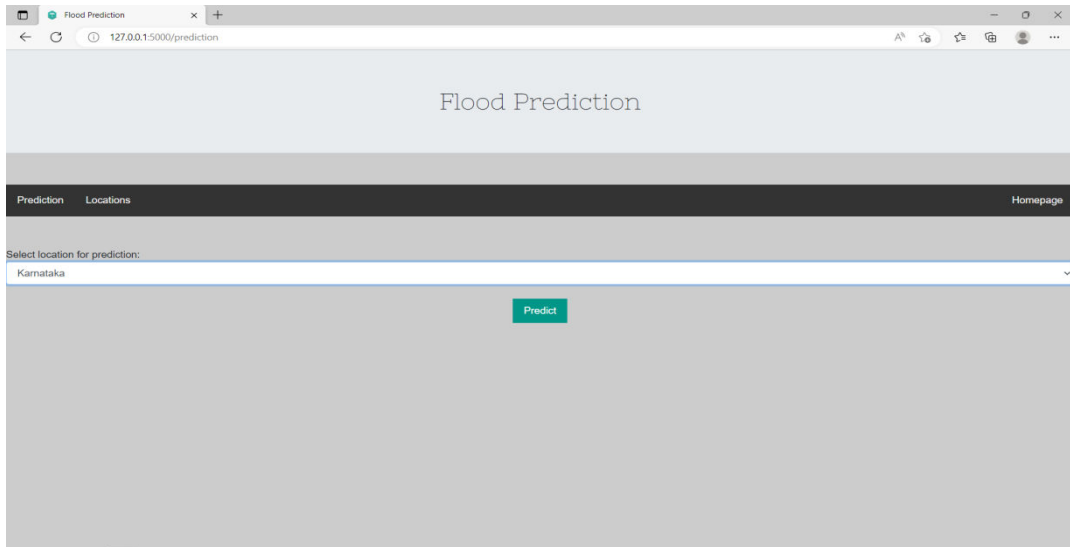


Fig:5 This image selected the location for prediction any state can be predicted.

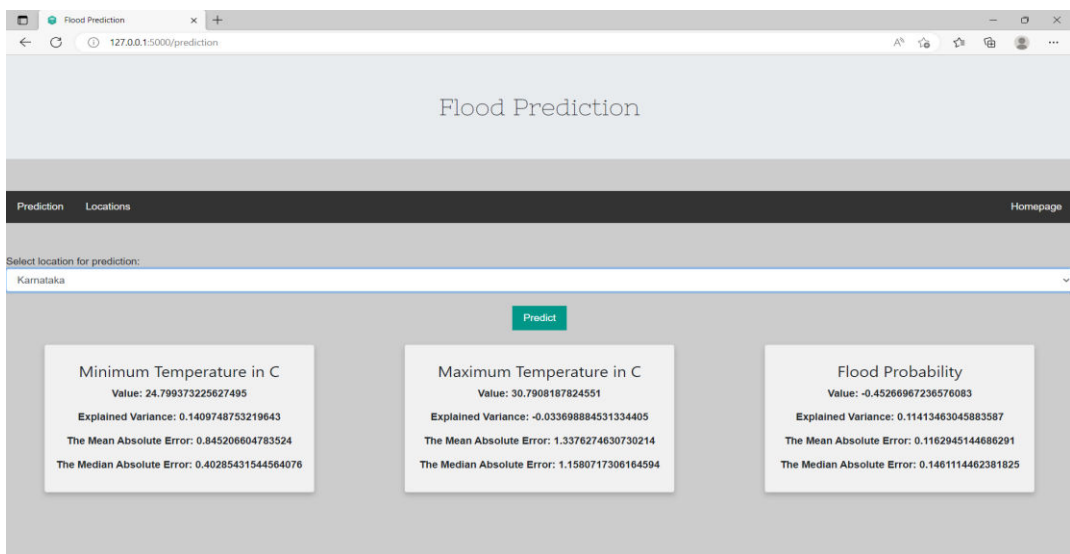


Fig:6 The result shows that predict the location then it display for minimum, maximum temperature and flood probability.

## VI. CONCLUSION

Floods are among the world's most pervasive and destroying catastrophic events. At the point when water is immersed over typical dry land, it causes a flood. It is brought about by a flood of water from a lake, stream, or sea that lowers adjoining land. Due to occasional vacillations in input water volume and vanishing rate, water levels in lakes and streams change emphatically. Floods can happen where downpour falls, despite the fact that downpour isn't the main source. Floods typically work over a time of hours or days, giving local people a lot of opportunity to planning or emptying the impacted region. Others show up out of the blue and abruptly. Quick streams, now and again alluded to as "Streak Flood", can clear out everything in their way, including individuals and property.

This prediction version may be utilized by catastrophe control branch to forecast flash flood. In future, we goal to apply different synthetic intelligence strategies to enhance the prediction accuracy. The system may be automatic via way of means of showing the end result of prediction in web site or laptop utility.



## REFERENCES

- [1] M. Kabir and M. N. Hossen, "Impacts of flood and its possible solution in bangladesh," Disaster Advances, vol. 12, pp. 48–57, Oct. 2019.
- [2] N.G Society, Flood, Nov. 2011. [Online]. Available: <http://www.nationalgeographic.org/encyclopedia/flood/>.
- [3] T. Luo, A. Maddocks, C. Iceland, P. Ward, and H. Winsemius, "World's 15 countries with the most people exposed to river floods," Mar. 2015. [Online]. Available: <https://www.wri.org/insights/worlds-15-countries-most-people-exposed-river-floods>.
- [4] G. FitzGerald, W. Du, A. Jamal, M. Clark, and X.-Y. Hou, "Flood fatalities in contemporary australia (1997–2008)," Emergency Medicine Australasia, vol. 22, no. 2, pp. 180–186, 2010, issn: 1742-6723. doi: 10.1111/j.1742-6723.2010.01284.x. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1742-6723.2010.01284.x>.
- [5] E. Holden, "Flooding will affect double the number of people worldwide by 2030," The Guardian, Apr. 2020, issn: 0261-3077. [Online]. Available: <https://www.theguardian.com/environment/2020/apr/23/flooding-double-number-people-worldwide-2030>.



INNO  SPACE  
SJIF Scientific Journal Impact Factor  
Impact Factor: 8.118



ISSN  INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details