FLOOD PREDECTION USING MACHINE LEARNING

Pruthvi N [1],Manjula Sanjay Koti[2]

[1] *Student, MCA,Dayananda sagar academy of technology and managment,Karnatka,india*
[2] *Prof&Head,MCA, Dayananda sagar academy of technology and management,Karnataka, india*

ABSTRACT

One of the undesirable natural catastrophes are floods. Significant property and life losses can result from flooding.Additionally, it could harm crops and trees that have been planted in agricultural areas. Flooding may be brought on by prolonged rainfall, inadequate precipitation drainage, surface runoff from melting snow, or dam failure. Today, the natural water storage places, such as rivers and lakes, have been destroyed and turned into construction sites. Flash floods can happen quickly within a few hours, as opposed to a typical flood. Research on flood prediction has evolved in order to reduce the amount of fatalities, property losses, and other concerns caused by flooding.Machine learning techniques are commonly used to build efficient prediction models for weather forecasting. Due to advancements in forecasting technology, more advanced and efficient solutions are now available. Utilizing the precipitation data from this research, a prediction model has been developed to anticipate the occurrence of floods triggered by rainfall. The model generates a forecast on the likelihood of a "flood occurrence" based on the expected range of precipitation for specific locations. Data on rainfall from various districts in India has been employed to construct the forecast model. The dataset has been utilized to train different algorithms, including Multilayer Perceptron, Support Vector Machine, and Linear Regression. Among these algorithms, the Multilayer Perceptron (MLP) classifier demonstrated satisfactory performance, achieving the highest accuracy rate of 97.40%. Consequently, a climate scientist can reliably predict the occurrence of a flood during heavy rainfall events using the Machine Learning Predection flash flood prediction model.

Keyword : - *precipitation, rapid inundation, MLP classifier*

---

1. TITLE-FLOOD PREDECTION USING MACHINE LEARNING

The worst flood catastrophes occur every year in India. Low-lying areas in large cities are where water logging occurs most commonly. The water level incresed logging is also attributed to a number of basic issues, height, and a lack of adequate drainage routes [2, 8]. In these regions, flood forecasting is essential. The present years, areas that are prone to floods include Assam, Bihar, Goa, Odisha, Pune, Maharashtra, Tamil Nadu, Karnataka, Kerala, and Gujarat. Chennai had 1049 mm of rain in 2015. In November, it rained a lot. Since 1918, November's rainfall total was greatest at 1088 mm. Normally, the months of October through December provide 64 cm of rain to the Kanchipuram district. There, during the period of the highest rainfall, 181.5 cm, or 183% more precipitation than usual, was recorded. The Tiruvallur district received 146 cm of rain as opposed to the normal of 59 cm.An abundance of research has been carried out on flood forecasting, but only a few methods offer forecasts that are very accurate. The flood prediction study mainly relies on machine learning (ML). A range of techniques using machine learning are available for more precise issue prediction . To prevent floods, we recommended employing flash flood predictions. A model for the ML algorithm is what we're trying to build. For more precise short-term predictions in an urban environment, it includes the flood component.

2. LITERATURE REVIEW

A Modest yet Significant Exploration of Flood Forecasting: Focus on Flash Floods .In a notable study by Akshya et al. [1], satellite aerial photos were employed to categorize flood-affected regions. To enhance accuracy, a hybrid ML technique combining SVM classifier with k-means was utilized for flood zone classification. Bande et al.'s smart flood prediction system incorporating neural networks and IoT facilitated data collection from sensors using IoT and Wi-Fi, while an Artificial Neural Network (ANN) approach enabled data analysis for flood prediction. However, challenges arise due to Wi-Fi network disruption during rainfall and the high cost of IoT communication. Cruz et al. [8] proposed a Multi-Layer ANN system for flood prediction, incorporating rainfall gauge, soil moisture, and water level as factors. The model demonstrated a relatively low Root Mean Square Deviation rate of 2.2645, indicating close alignment between predicted and actual flood levels. To further enhance forecasting, future studies should explore additional factors.

Cui et al. [20] introduced the AHP-GM-ANN model for estimating flood losses in coastal cities, addressing nonlinear connection variables and improving quantitative accuracy in prediction. Kartika et al. [18] developed a flood prediction approach using the Radial Basis Function (RBF) to forecast water levels and daily rainfall for the upcoming month, accounting for the challenges posed by climate change. In the pursuit of optimization, Kaur et al. [9] employed hybrid algorithms combining Genetic Algorithm (GA) and SVM in both standalone and cloud environments. The cloud context exhibited superior accuracy of 86.36%, whereas the standalone environment yielded lower accuracy. Ruslan et al. However, longer prediction periods may result in decreased precision. Addressing this concern, Widiasari et al. proposed an LSTM-based flood prediction model, providing context awareness and real-time event predictions with detailed data processing. Sardjono et al. [30] employed an ANN to mitigate flood risks outside Jakarta City, uncovering valuable data patterns for future flood predictions. Similarly, Ramli et al. [6] [11] introduced a neural network autoregressive model for flood prediction in Kuala Lumpur, achieving a maximum accuracy of 73.54%. The authors recommended exploring multiple prediction models to further reduce forecast time..

3.COMPONENTS AND PROCEDURE

a) **Data Collection:** Gather historical data related to floods, including rainfall, river levels, soil , and other relevant variables. Obtain information on previous flood events, including their extent, duration, and severity. Acquire data from various sources such as weather stations, river gauges, remote sensing, and historical records.

b) **Data Preprocessing**: Clean the collected data by removing outliers, missing values, and inconsistent entries. Normalize the data to bring all variables to a comparable scale. Perform feature engineering to extract relevant features, such as deriving new variables or aggregating existing ones.

c) **Feature Selection**: Utilize statistical methodologies, like correlation analysis, to ascertain the most significant factors for flood prediction. Take into account domain expertise and guidance from experts to determine meaningful variables. Implement dimensionality reduction approaches, such as principal component analysis (PCA), if the dataset comprises a substantial number of attributes.

d) **Data Split**: Divide the dataset into training, validation, and testing sets.The set is used to train the machine learning models.The validation set helps in tuning the hyperparameters and evaluating model performance during training. The set is used to assess the final model's performance and generalization ability.

e) **Model Selection**: Choose suitable machine learning algorithms for flood prediction, such as decision trees, random forests, support vector machines (SVM), or neural networks. Consider the nature of the problem (regression or classification) and the available data. Compare and evaluate different algorithms based on their performance metrics, such as accuracy, precision, recall, or root mean square error (RMSE).

f) **Model Training**: Train the selected machine learning model using the set. Fine-tune the model's hyperparameters through techniques like grid search or random search. Implement cross-validation to assess the model's performance and prevent overfitting.

g) **Model Evaluation**: Evaluate the trained model using the validation set.Measure the model's performance using appropriate evaluation metrics. Adjust the model and repeat the training and evaluation process if necessary.

h) **Model Testing**: Assess the final trained model's performance using the testing set. Compare the model's predictions with the actual flood events. Calculate relevant evaluation metrics to determine the model's accuracy and reliability.

i) **Deployment and Monitoring**: Deploy the trained model in a real-time or near-real-time environment to predict floods. Continuously monitor the model's performance and update it with new data to maintain accuracy. Implement a feedback loop to improve the model over time based on the observed outcomes and user feedback.

j) **Integration with Decision Support Systems**: Integrate the flood prediction model with decision support systems to provide timely alerts and recommendations to stakeholders. Develop user-friendly interfaces to visualize predictions, historical data, and potential flood scenarios.Collaborate with domain experts, emergency response teams, and policymakers to ensure effective utilization of the prediction model for flood management.

3. MODULES

Metropolitan cities, similar to low-lying areas or underpasses, are highly susceptible to waterlogging. These urban areas experience rapid accumulation of water during flood events. Insufficient drainage capacity and surface runoff are the primary factors contributing to the exacerbation of waterlogging. Hence, accurate flood forecasting becomes imperative. To construct a comprehensive dataset that enables effective utilization of machine learning (ML) algorithms for pattern extraction and improved accuracy, it is essential to integrate information from diverse sources. The flood forecasting methodology is depicted in Figure 1.

3.1database of rain Data

The set has 641 records and 19 attributes. State, districts, Jan.-Feb.-Mar.-Jun.-Sept.-Oct.-Dec., annual, of strings and numeric value types. The data source is available at https://www.kaggle.com/raja/rainfall-in-India.

3.2 **Initial processing**

The dataset has been meticulously processed to ensure the elimination of duplicate, null, and missing values. Through this pre-processing stage, the labels have been transformed into numeric representations, facilitating effective machine readability. Furthermore, leveraging weather information, a novel column called "flood" has been generated.

3.3Data division

The data is divided into two subsets: a training dataset and a testing data. The testing data comprises 30% of the total data, while the training dataset contains the remaining 70%. This division allows for evaluating the performance and generalization ability of the trained model on unseen data.

3.4**Classified methods**

The data is trained using the techniques Support Vector Machine, Logistic Regression, K-Nearest Neighbour, and Multi-Layer Perceptron. The computation makes use of the following variable:

1) **Exactness:** The ratio of correctly predicted positive outcomes to the total positive outcomes. (1) False Positive (FP) and True Positive (TP) are the terms used for this measurement.

2) **Think back to**: Recall (2) is the proportion of observed positive values that were correctly predicted. The acronym FN stands for false negative.

3) **Score F**

In order to calculate the F1 score, 1,2 are used. It shows what their typical weight is.

4) **Sensitivity**

Recall is another term for sensitivity.TP rises as sensitivity increases.

5) **Specificity gauges how many actual negatives there are**.

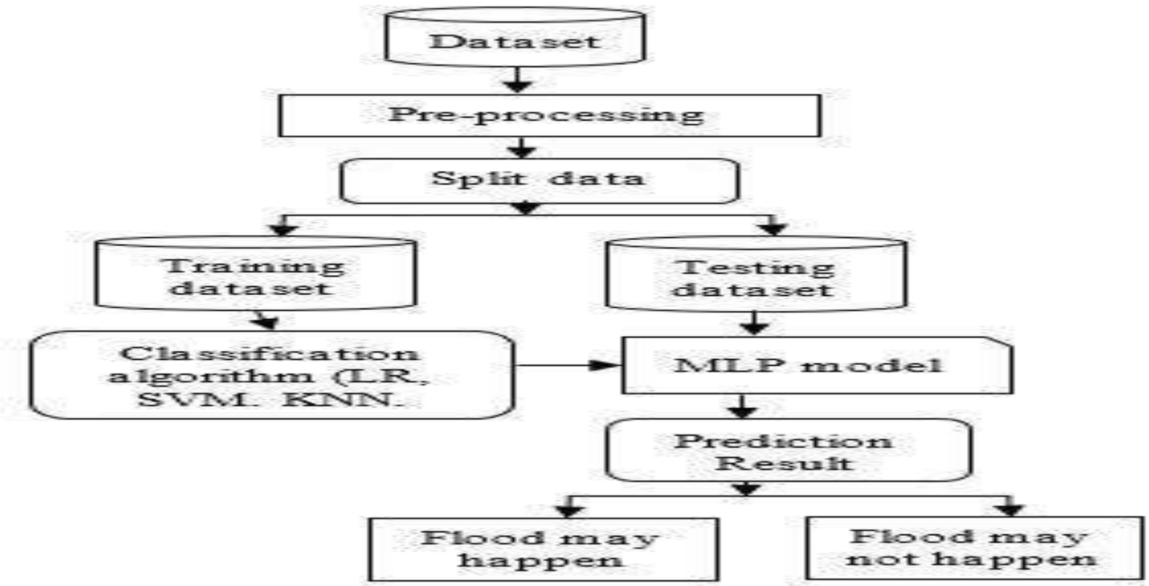
True Network

6) **Confusion chart**

The success of the categorization approaches, are shown visually in the confusion matrix. Model MLP, E. The MLP algorithm has been used for extra stages. Since the MLP approach has the high accuracy (about 97.40%), it is utilised to predicted the onset of flash floods.

3.5 How the forecast turned out

That flash flooding class is set to 1 if the input exceeds 2400mm. CLASS 1: "FLASH FLOOD MAY HAPPEN" If the input has a diameter smaller than 2400mm, class 2 ("FLASH FLOOD MAY JUST HAPPEN") is set for "0."The flash flood class is set to 1 if the provide information exceeds 2400mm.



4. TEST RESULTS AND TESTING

The recommended method uses machine learning techniques to evaluate rainfall data in order to develop a system to more correctly predict flash floods. In Table I, an algorithm performance analysis is shown. Table II displays a comparison of the matrix for algorithmic misunderstanding. The recommended model gives a rather easy and reliable method for forecasting floods, as shown by the steps below: First Step : a preprocessing step is performed on the dataset that was collected on rainfall. Second Step : At random, training and testing groups are created from the dataset on rainfall. Third Step: The dataset trained using the Logistic Regression, Support Vector Machine, K-Nearest neighbor algorithms. Step 4: The approach, which has the highest degree of accuracy. Its correctness is then checked using variables including recall, sensitivity, naiveté, f1-score, accuracy, and precision.(5) The predicted model will be fed test data in step 5, and the results will be verified. calculated utilising the accuracy, specificity, and sensitivity of the algorithm. According to (1) and (2), accuracy and recall are determined. In Figure 1, the accuracy rate is shown as a Performance assessment, Figure I

PARAMETER	LR	SVS	KNN	MLP
Precision	0.98	0.97	0.99	0.96
recall	0.94	0.99	0.94	1
Sensitivity	0.99	0.97	0.99	0.96
Roughness (%)	96.9	93.99	94.99	94.20

5. CONCLUSION

flash floods have devastating consequences for both living and non-living entities. This research project focuses on flash flood forecasting model. The study utilizes district-level rainfall data from India, collected over a period of 1901 to 2015. The data is divided into a testing set (30%) and a training set (70%) after undergoing pre-processing. The dataset is trained using four different machine learning algorithms: K-nearest Neighbors (KNN), Support Vector

Machine (SVM), Logistic Regression (LR), and Multi-Layer Perceptron (MLP). That each approach is evaluated using various metrics such as precision, recall, F1 score, sensitivity, and specificity. Confusion matrices are created to visualize the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The classifi accuracy for SVM, KNN, LR, and MLP is calculated to be 95.85%, 95.85%, and 97.40% respectively.we have four techniques, MLP demonstrates the highest precision, indicating its effectiveness in flash flood prediction. The MLP flash flood prediction model categorizes floods as "may happen or not" based on the observed rainfall range in a specific location. This forecasting model can be utilized by disaster management organizations to enhance preparedness for flash floods. To further improve the accuracy of predictions, additional artificial intelligence techniques can be incorporated. Furthermore, the prediction results can be automated and displayed through a website or desktop program to streamline the process. This would enable easy access to the forecasted outcomes and facilitate effective decision-making in response to potential flash flood events.

6. REFERENCES

[1]. Akshya J and Priyadarsini K, "A Hybrid Machine Learning Approach for Classifying Aerial Images of Flood-Hit Areas," in Second International Conference on Computational Intelligence in Data Science, 2019.

[2]. Amir Mosavi., Pinar Ozturk and Kwok-wing Chau, "Flood Prediction Using Machine Learning Models: Literature Review," *MDPI,* p. 40, 2018.

[3]. Amitkumar B and Durge P, "Different Techniques of Flood Forecasting and their Applications," 2018

[4].Analyn N., Charmaine C., Arnold C., Glenn O., JoseAngelo C., JanRalley P., and Jesse Dave S, "Real-Time Flood Water Level Monitoring System with SMS Notification," 2017.

[5]. Bipendra Basnyat., Nirmala Roy.,and Aryya Gangopadhyay., "A Flash Flood Categorization System using Scene-Text Recognition," in International Conference on Smart Computing, 2018.

[6]. Fazlina Ahmat Ruslan.,Abd Manan Samad and Ramli Adnan, "4 Hours NNARX Flood Prediction Model Using “traingd” and “trainoss” Training Function: A Comparative Study," in 2018 IEEE 14th International Colloquium on Signal Processing & its Applications (CSPA 2018), 9 -10 March 2018, Penang, Malaysia, 2018.

[7]. Fazlina Ahmat Ruslan., Khadijah Haron., Abd Manan Samad and Ramli Adnan, "Multiple Input Single Output (MISO) ARX and ARMAX Model of Flood Prediction System: Case Study Pahang," in IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA 2017), Penang, Malaysia, 2017.

[8]. Febus Reidj G., Matthew G., Marlou Ryan G., and Francis Aldrine A., "Flood Prediction Using Multi-Layer Artificial Neural Network in Monitoring System with Rain Gauge, Water Level, Soil Moisture Sensors," in Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference, (Jeju,Korea, 2018).