Prediction of Flash Flood using Rainfall by MLP Classifier

Vinothini A, Kruthiga L, Monisha U

Abstract*: Flood are one of the unfavorable natural disasters. A flood can result in a huge loss of human lives and properties. It can also affect agricultural lands and destroy cultivated crops and trees. The flood can occur as a result of surface-runoff formed from melting snow, long-drawn-out rains, and derisory drainage of rainwater or collapse of dams. Today people have destroyed the rivers and lakes and have turned the natural water storage pools to buildings and construction lands. Flash floods can develop quickly within a few hours when compared with a regular flood. Research in prediction of flood has improved to reduce the loss of human life, property damages, and various problems related to the flood. Machine learning methods are widely used in building an efficient prediction model for weather forecasting. This advancement of the prediction system provides cost-effective solutions and better performance. In this paper, a prediction model is constructed using rainfall data to predict the occurrence of floods due to rainfall. The model predicts whether "flood may happen or not" based on the rainfall range for particular locations. Indian district rainfall data is used to build the prediction model. The dataset is trained with various algorithms like Linear Regression, K- Nearest Neighbor, Support Vector Machine, and Multilayer Perceptron. Among this, MLP algorithm performed efficiently with the highest accuracy of 97.40%. The MLP flash flood prediction model can be useful for the climate scientist to predict the flood during a heavy downpour with the highest accuracy.

Keywords: Prediction, MLP classifier, Flash flood, Rainfall.

I. INTRODUCTION

Every year, India is the topmost flood-prone disaster place in the world. Mostly water logging in urban cities occurs in low-lying areas. Moreover, the increase in water logging is due to some fundamental points such as surface runoff, relative altitude, and not enough path of the water to drainage [2, 8]. So, flood forecasting is essential at these places. In a recent year, there were many parts of countries which are prone to flood like Assam, Bihar, Goa, Odisha, Pune, Maharashtra, TamilNadu, Karnataka, Kerala, and Gujarat.

Revised Manuscript Received on April 16, 2020.
***** Correspondence Author
Vinothini A *, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India. Email: vinothini.a@rajalakshmi.edu.in
Kruthiga L, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India. Email: kru98log@gmail.com.
Monisha U, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India. Email: nmonishaumapathi@gmail.com.

In the year 2015 rainfall, Chennai received 1049 millimeters (mm) of rainfall in November. Since 1918, 1088 mm of precipitation was the best recorded in November. Between October and December, the average rainfall in Kanchipuram district is 64 cm. It received the heaviest rainfall of 181.5 cm, which is 183% higher against average precipitation. In the Tiruvallur district, the average rainfall is 59 cm but recorded 146 cm of rain [14].

There was much research for prediction of flood ahead, but not many methods give the estimate with high accuracy. The flood prediction analysis majorly uses Machine Learning (ML). There are many methods in machine learning to predict the problem with higher accuracy [2].

In this work, we have proposed to estimate the flash flood to prevent places that are prone to flood risk. The approach is to the establishment of the ML algorithm model. It incorporates the flood factor to estimate short term prediction in an urban area with higher accuracy.

II. LITERATURE SURVEY

The flash flood explored a small but ample flood forecasting research work carried out. Akshya et al. [1] proposed the classification of the flood-hit region using the aerial images from the satellite. A hybrid ML approach, SVM classifier with k-means provides better accuracy but after flood region classified. The prediction system designed using IoT and Neural Networks for smart prediction of the flood by Bande et al., [27] suggested that the data collected from the sensor using IoT and Wi-Fi with an Artificial Neural Network (ANN) approach was used for communication of the data analysis in flood prediction. The Wi-Fi is highly impossible due to network during rainfall, the high cost of IoT, and communication. Cruz et al. [8] studied the flood prediction system by Multi-Layer ANN. Rain gauge, soil moisture, and water level are the parameters used in this model. Moreover, the actual setup tested and the model showed 2.2645 of Root Mean Square Deviation rate across the whole data set, which implies an overall small difference between the flood predicted, and actual flood level. The Multi-layered artificial neural network gives very well validation, but this study requires more parameters for calculation of prediction.

The coastal cities flood loss prediction by Cui et al. [20] designed the AHP-GM-ANN model would reduce the predictive error to get reliable results significantly.

The model solved the problem in nonlinear relationship variables and also improved the quantitative system accuracy in the predictive method. Kartika et al. [18] implemented the system to predict the flood using Radial Basis Function (RBF). It determines the next month's water level and daily rainfall. But it may be impossible to predict the rainfall due to climate change. Kaur et al. [9] implemented the hybrid algorithm in standalone and cloud environments for efficiency. The automated hybrid algorithms are a Genetic Algorithm (GA) and SVM. The evaluated model shows the results in a cloud environment, which yields the highest accuracy of 86.36%, but in a standalone, it is less inaccurate. Case study of flood prediction system by Ruslan et al. [7] studied the design of prediction models using Multiple Input Single Output (MISO) ARX and ARMAX model structure in Pahang for 7 hours and shows the comparison of the prediction performances. When the prediction time is longer, the model output can be unreliable.

A flood prediction model using LSTM by Widiasari et al. [11] implemented the LSTM model for solving. The model gives context-awareness, which in turn provides information about data processing and real-time predictions of the events to the user. Sardjono et al. [30] applied an ANN for flood systems mitigation in place of Jakarta City. The data patterns were determined as a result of this flood model. The result of Jakarta are expected to use in prediction and flood behavior in the future. Ramli et al. [6] presented the neural network autoregressive model for floods prediction in Kuala Lumpur. The model showed an accuracy of 73.54% as a maximum. The authors suggested that by using different prediction models, the prediction time can reduce further.

As per the literature survey, authors have not combined the classification algorithm and unfocused on the accuracy, which reduces the execution time for optimization. Hence, this research work mainly focuses on predicting the floods accurately with minimum execution time, more accuracy, and best features.

III. MATERIALS AND METHODS

The dataset used for the analysis is an Indian rainfall data, between the periods of 1901 to 2015. The dataset is created in CSV file format in month-wise and also sub-division of the state along with district. The unit used to measure the rainfall is in millimeter (mm). The dataset was collected from the metrological department of 36 locations on a monthly basis. ML method is useful to estimate the future by analyzing historical data like occurrences of flood in earlier times. The ML Performance are measured using algorithms like LR, SVM, KNN, and MLP.

A. Logistic Regression

Classification and Regression problems use logistic regression but mostly used for classification problems. The output of the Logistic Regression problem is only in the form of 0 and 1. Logistic regression is useful, where ever the probabilities between the two classes are essential. Such as "whether it will rain nowadays or not", "either zero or one" and "true or false", etc.

B. K-Nearest Neighbor

K-NN algorithm stores all the existing data and classifies new data information based on the similarity. This states that when new data appears, it can easily classify into a well-set category by using the K-NN algorithm. It is also known as a lazy learner algorithm because it does not learn from the training set, instantly it stores the dataset consequently, and at the time of classification, it acts on the dataset. It will categories the data into similar groups based on similarity.

C. Support Vector Machine

Classification and Regression analysis use Support vector machine (SVM) algorithm for evaluating data. They have two categories in which data sets are trained. However, when new data comes, it can be easily categorized under a particular group. Thus, it is called a non-binary linear classifier.

D. Multi-Layer Perceptron

Multi-Layer Perceptron is the simplest machine learning methods. It can take minimum inputs, each of which weights to signify how important it is, and generate an output decision of "0" or "1". It forms an artificial neural network when combined with many other perceptron. The neural network can theoretically answer any question, given enough training data and computing power.

IV. WORKING MODULES

Waterlogging is vulnerable in the urban city areas similar to sub-pass or a low-lying. Water will accumulate in a short period in flood-prone areas. The fundamental facts in the development of waterlogging are surface runoff and insufficient passage of water to drainage**.** Thus, it is essential for flood forecasting.

The dataset from different sources is to form a generalized dataset on which ML techniques are applied for extracting patterns, and to obtain results with higher accuracy. The figure 1 represent the process of prediction of flood.

A. Rainfall dataset

The dataset consists of 641 records and 19 features. The features are State, Districts, January to December, Annual, Jan-Feb, Mar-May, Jun-Sep, Oct-Dec, of type string and numeric. Dataset derived from https://www.kaggle.com/rajanand/rainfall-in-india.

B. Pre-processing

Dataset undergoes pre-processing, and therefore, it has no missing values, no null values, and no duplicate values. The labels are converted into numeric forms to create an efficient machine-readable format. A new column flood is created, based on meteorological data.

C. Splitting data

The dataset is divided into the testing dataset and training dataset. The training dataset is about 70%, and the test dataset is about 30%.

D. Classification algorithm

The dataset is trained along with algorithm Support Vector Machine, Logistic regression, K-nearest neighbor, and Multi-Layer Perceptron. Parameter used for calculation is:

1) Precision:

Precision is the proportion of positive prediction that observed to be correct.

$$\text{Precision} = TP / (TP + FP) \quad (1)$$

TP refer to True Positive and FP refer to False Positive.

2) Recall:

A recall is the ratio of observed positive values that is correctly predicted.

$$\text{Recall} = TP / (TP + FN) \quad (2)$$

FN refer to False Negative.

3) F1 Score:

F1 score takes the (1) and (2) for the calculation. It gives the average weight of them.

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision} \quad (3)$$

4) Sensitivity:

Sensitivity is also called a Recall.TP becomes higher when the sensitivity value is higher.

$$\text{Sensitivity} = (TP) / (TP + FN) \quad (4)$$

5) Specificity:

Specificity is the ratio of actual negative values.

$$\text{Specificity} = TN / (TN + FN) \quad (5)$$

TN refer to True Negative.

6) Confusion matrix:

The confusion matrix shows the visual representation of the classification methods performance such as TP, TN, FP, and FN rate.

E. MLP model

The MLP algorithm has taken, for further steps. MLP algorithm as the highest accuracy of about 97.40%, so it is used in the prediction of flash flood occurrence.

F. Prediction result

CLASS 1: "FLASH FLOOD MAY HAPPEN":

If the given input is higher than 2400mm, then the flash flood class is set to 1.

CLASS 2: "FLASH FLOOD MAY NOT HAPPEN":

If the given input is lesser than 2400mm, then the flash flood class is set to 0.

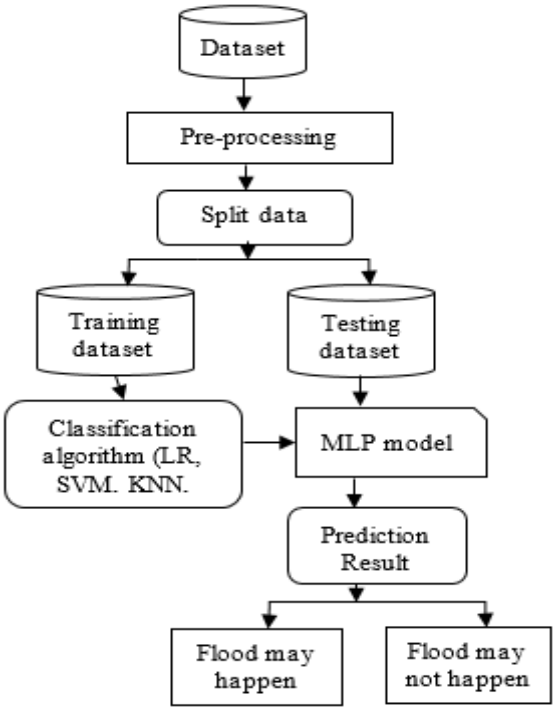

Fig. 1.Process of prediction of flood

V. EXPERIMENT AND RESULT

The proposed work is a way to evaluate the rainfall dataset to predict the flash flood using machine learning techniques with higher accuracy. Table I shows the performance analysis of algorithms. Table II represent the comparison of confusion matrix of algorithm. The following steps stated that the proposed model provides a very easy efficient method for predicting flood:

Step1: The collected dataset of rainfall is preprocessed.

Step2: The dataset of rainfall is randomly partitioned into testing and training.

Step3: The dataset trained with LR, SVM, KNN and MLP algorithm.

Step4: The model is constructed using the MLP algorithm with the highest accuracy and validated with the parameters such as precision, recall, f1-score, sensitivity, specificity, and accuracy.

Step5: Input test data to the prediction model and validate the results.

Accuracy of the algorithm calculated from the Sensitivity and Specificity. From (1) and (2), the precision and recall is measured. The figure 2 represent the accuracy percentage of the classification algorithm.

Table-I: Performance analysis

PARAMETER	LR	SVC	KNN	MLP
Precision	0.99	0.96	0.97	0.97
Recall	0.96	1	0.98	1
F1-Score	0.98	0.98	0.98	0.99
Sensitivity	0.96	1	0.98	1
Specificity	0.75	0	0.37	0.37
Accuracy (%)	95.3	95.85	95.85	97.40

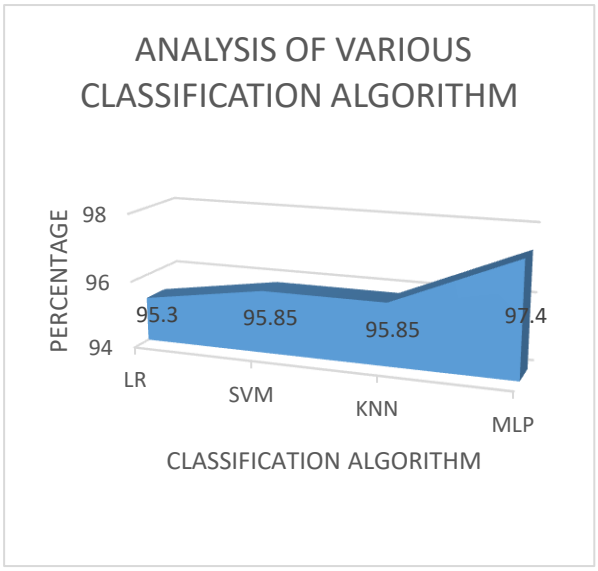

Fig. 2. Analysis of various classification algorithm for flood prediction.

Table-II: Comparison of Confusion matrix parameter

Algorithm	TP	TN	FP	FN
LR	6	178	7	2
SVM	0	185	0	0
KNN	3	182	3	5
MLP	3	185	0	5

VI. CONCLUSION

Damages that occur due to flash flood to living and non-living are very large. In this paper, flash flood prediction model is built. District wise Indian rainfall data collected between the periods of 1901 to 2015 is used for analysis. The pre-processed rainfall data was split into 70% training data and 30% testing data. The dataset is trained with Support Vector Machine, Logistic regression, K-nearest neighbor, and Multi-Layer Perceptron. The performance factors like precision, recall, F1 score, sensitivity, specificity was calculated for each technique. Confusion matrix with TP, TN, FP and FN were calculated. The classification accuracy achieved by LR is 95.3%, SVM is 95.85%, KNN is 95.85%, and MLP is 97.40%. Among the four techniques MLP performed with highest accuracy. The MLP flash flood prediction model predicts whether "flood may happen or not" based on the rainfall range for particular locations. This prediction model can be used by disaster management department to forecast flash flood. In future, we aim to use other artificial intelligence techniques to improve the prediction accuracy. The process can be automated by displaying the result of prediction in webpage or desktop application.

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).
9. Gurleen Kaur and Anju Bala, "An Efficient Automated Hybrid Algorithm to Predict Floods in Cloud Environment," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 2019.
10. Homar Baez., Idalides Vergara-Laurens., Luz Torres- Molina., and Luis G., Miguel A. Labrador, "A Real-Time Flood Alert System for Parking Lots," 2017.
11. Indrastanti R. Widiasari., Lukito Edi Nugoho., Widyawan., and Rissal Efendi., "Context-based Hydrology Time Series Data for A Flood Prediction Model Using LSTM," in *Proc.of 2018 5th Int. Conf. on Information Tech., Computer and Electrical engineering*, 2018.
12. Joe Marlou A.Opella and Alexander A.Hernandez, "Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework," in *2019 IEEE 15th International Colloquium on Signal Processing & its Applications (CSPA 2019), 8 -9 March 2019*, Penang, Malaysia, 2019.
13. Junio S., Cristiane L.,Marcio R.,Debora F.,Marcia D.,and Wesley P., "Predicting System for flooded areas developed with MLP-type Neural Network," 2018.
14. "loss-due-floods-India-people-killed-crop-houses-damaged- in-65-years," India today, 2019.
15. Kiruthika Devi B., Aravindhan K., Srikanth Kini P.,Adarsh Reddy G, Subbulakshmi T, "A Prediction Model For Flooded Packets In Snmp Networks," in *Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018)*,, 2018.
16. Mohamad Nazrin Napiah., Mohd Yamani Idna Idris., Md Asri Ngadi, and Ismail Ahmedy, "Flood Alerts System with Android Application," 2017.
17. Nikhil Binoy C. Arjun N., Keerthi C., Sreerag S.,and Ashwin, "Flood Prediction Using Flow And Depth Measurement With Artificial Neural Network In Canals," in *Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)*, 2019.
18. Ni Komang Ega Kartika., Muhammad Ary Murti, and Casi Setianingsih, "Floods Prediction Using Radial Basis Function (RBF) Based on Internet of Things (IoT)," in *IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 2019.
19. Pariwat Ongsulee., Veena Chotchaung., Eak Bamrungsi and Thanaporn Rodcheewit., "Big Data, Predictive Analytics and Machine Learning," in *Sixteenth International Conference on ICT and Knowledge Engineering*, 2018.
20. Pengzhan Cui., Yeqing Guan., and Ying Zhu, "Flood Loss Prediction of Coastal City Based on GM-ANN," 2017.
21. Phyo Pa Pa Tun and Myint Myint Sein, "Flood Prediction System for Middle Region of Myanmar," in *7th Global conference on Consumer Electronics*, 2018
22. 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.
23. Qianyu Zhang., Nttha Jindapetch.,Rakkrit Duangsoithong and Dujdow Buranapanichkit., "Investigation of Image Processing based Real-time Flood Monitoring," in *Proc. of the 2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA 2018)*, Songkhla, Thailand, 2018.
24. Ryo Natsuaki and Akira Hirose, "L-Band Sar Interferometric Analysis for Flood Detection in Urban Area –A Case Study In 2015 Joso Flood, Japan," 2018.
25. Shijin LI., Kaikai MA., Zhou JIN., and Yuelong ZHU, "A new flood forecasting model based on SVM and boosting learning algorithms," 2016.
26. Sreehari E and Satyajee Srivastava, "Prediction of climate variable using Multiple Linear regression," in *2018 4th International Conference on Computing Communication and Automation (ICCCA*, 2018.
27. Swapnil Bande and Virendra V. Shete, "Smart flood disaster prediction system using IOT & Neural Networks," in *InternationalConference On Smart Technology for Smart Nation*, 2017.
28. Tibin Mathew Thekkil and prabakaran N, "Real-time WSN Based Early Flood Detection and Control Monitoring System," in *2017 International Conference on Intelligent Computing,Instrumentation and Control Technologies (ICICICT)*,, 2017.
29. Vinothini A and Baghavathi priya A, "Survey of Machine Learning Methods for Big Data Applications," in *International Conference on Computational Intelligence in Data Science*, 2017.
30. Wahyu Sardjono and Widhilaga Gia Perdana, "The Application of Artificial Neural Network for Flood Systems Mitigation at Jakarta City," in *International Conference on Information Management and Technology (ICIMTech)*,Jakarta & Bali, Indonesia, 2019.

AUTHORS PROFILE

Ms.A.Vinothini, is working as an Assistant Professor in the Department of Computer Science and Engineering at Rajalakshmi Engineering College, Chennai. She completed B.Tech- IT from Karpagam College of Engineering, Coimbatore affiliated to Anna University, Chennai and M.Tech- IT from Anna University, Coimbatore. She is currently pursuing Ph.D. in Anna University Chennai. She is a university rank holder. Her areas of interest include Data mining, Machine Learning, Big Data Analytics and Data Structures. She has published various research works in national and international journals.

Kruthiga L, is currently pursuing her Bachelor of Engineering degree in Computer Science and Engineering at Rajalakshmi Engineering College, Anna University. She has completed 3 Mini projects on HTML, CSS and javascript with database connectivity. Her areas of interest include Python, HTML, CSS, javascript, Java, DBMS and IoT. She is currently placed in TCS Ninja. She has presented paper on Smart Transportation at Velammal Engineering College. She has participated in workshop on Ethical Hacking.

Monisha U, is currently pursuing her Bachelor of Engineering degree in Computer Science and Engineering at Rajalakshmi Engineering College, Anna University. She has completed 3 Mini projects on HTML, CSS and javascript with database connectivity. Her areas of interest include HTML, CSS, javascript, Java, DBMS and IoT. She is currently placed in Solartis. She has completed her internship for duration of three months in Solartis. She has presented paper on Smart Transportation at Velammal Engineering College. She has participated in workshop on Ethical Hacking.