ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)

Available Online at www.ijetcr.org

Volume 14; Issue 03; 2026; Page No. 1-9

AI-Driven Sensor Networks for Early Flood Detection and Risk Mitigation

Vikash Kumar[1], Sarvesh Kumar[2], Arvind Kumar Mishra[3]

[1,2,3]Amity Institute of Information Technology, Amity University, Patna

vikashpaper1991@gmail.com[1], skumar8@ptn.amity.edu[2], akmishraw@gmail.com[3]

Conflicts of interest: Nil

Corresponding author: Vikash Kumar

Abstract

The increasing frequency and severity of both extreme weather conditions and in the first-place floods in recent years developed into the long-awaited realization that there is a need to have effective monitoring and notification systems. The importance of preparatory measures in improving preparedness and response can be explained by the floods that are described as having severe implications on human beings, infrastructures as well as the natural environment. One of the significant dangers to the world population is floods; therefore, there is the necessity to have effective early warning systems that would guarantee people are evacuated and mitigated in good time. The research proposal will analyse the uses of machine learning algorithms and sensor technology in enhancing flood prediction systems and flood warning systems. The proposed system uses machine learning models as predictive analytics based on real-time data of different sensors (rainfall gauges, river level, weather stations, etc.). Regression, classification, and ensemble models are some of the models used in this study and are trained on Historical data to predict flood occurrences with better precision and lead time. The architecture of the system makes it easy to constantly acquire data, preprocess it, train the model and deploy alerts in the form of mobile applications and emergency communications channels. Precision, recall, and F1-score are evaluation metrics that show the effectiveness of the approach compared to traditional ones. The results help to highlight the prospects of incorporating superior technologies to develop the preparedness and response to floods, which helps to reduce the risks and helps to reduce the number of damage-ages and victims of floods. Future trends will involve scalability, resistance to various environmental factors, and the incorporation of new IoT systems to provide holistic disaster management systems.

Keywords: Flood Prediction, Machine Learning Algorithms, Sensor-Based Monitoring, Early Warning Systems, Real-Time Data Analytics, Internet of Things (IoT), Disaster Management and Mitigation

Introduction

The drastic weather occurrences with an augmented frequency and severity have been an irritating pattern in the globe during the past few years and the floods have become one of the most dramatic and disastrous inputs. These incidences pose an enormous challenge to society in the whole world as they risk life, infrastructure and environment. To address these challenges, it is necessary to have tough monitoring and alert systems that will assist in the provision of early alerts and have good response mechanisms. It possesses some potential prospects to enhance flood forecasting and precursors of floods based on the combination of innovative technologies, specifically, machine learning and sensor networks. The models of machine learning used in this paper are numerous: K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree Classifier, Random Forest Classifier, and a Voting Classifier ensemble strategy applying the Kerala flood dataset of 2018**[1].** Such models are trained in past data of significant environmental variables such as rain patterns and rivers heights.

By strict assessment and comparison, the study is expected to determine effectiveness of each model to predict the occurrences of floods with accuracy and reliability. The combination of machine learning algorithms does not only improve the accuracy of predictions, but it also makes the process of making decisions in a timely manner to reduce the effects of floods on communities and infrastructure possible. This research paper aims at achieving the following objectives by investigating the methodologies employed, comparing the performance of different models, and implications to future research to improve the system of flood prediction with the help of more refined data analytics and sensor devices **[2].**

Literature Survey

Floods are experienced when the amount of precipitation exceeds the capacity of reservoirs and river channels. The influence of climate change. An environment can be regarded as the primary cause of a shift in the composition of floods, their intensity, and size. The popularity of machine learning models is increasing as they can process complicated data and provide a relationship between different physical parameters, which would be very tedious with old statistical modeling tools. Flood prediction models have played a major role in the reduction of risks, policy recommendations, loss of human life minimization, and minimization of property damage because of floods. Machine learning (ML) techniques have been significant in this development in the last 20 years, having been adequately used to recreate the complicated mathematical equations of physical phenomena of floods. Such techniques have improved the effectiveness of prediction systems, which provide superior and more economical solutions. A vast number of developed flood prediction systems use machine learning and image processing **[3]**. Nevertheless, the explicit evaluation of different methods and analysis in respect of remote sensing in flood forecasting is not much explored. It is easy to conceive an amalgamation of sensing systems, communication networks, cloud computing, machine learning and data analytics, and so on, as this will allow creating an integrated flood disaster management system that will be able to provide alerts to the regions facing the flood. Internet of Things (IoT) technology is increasingly growing in popularity, and it becomes crucial to tackle a range of issues on the interdisciplinary front. To do this, an IoT application is required to bridge the linkage between citizens and sensor nodes placed in the areas of vulnerability via the use of internet. This will allow the citizens to have access to current and precise information regarding the prevailing weather conditions at any time. Therefore, the new technologies can save lives in the case of flood disasters **[4]**.

Proposed methodology

Machine learning models are critical to the flood prediction field as they are used to make more accurate and efficient evaluations and predictions of the occurrence of floods using data-driven algorithms. This study used the Kerala flood dataset of 2018 to test a wide range of machine learning models such as K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree Classifier, Random Forest Classifier and ensemble learning. These models were trained and evaluated using historical information that encompasses significant environmental variables aiming at enhancing our ability to predict and reduce disasters of floods with great precision**.**

Machine Learning Models Several machine learning models were employed in this study.

KNN, or K-Nearest Neighbors: KNN is a machine learning method which involves most of the nearest data in the feature space to classify fresh data points. Flood prediction with the 2018 Kerala dataset by using KNN, the data needed to be prepared by scaling the features of the data first and then splitting into training and test data sets. The KNN model was then trained on the training data with the forecasting of flood events being based on previous environmental factors such as the rainfall and the river levels. Evaluations of the model accuracy and performance were conducted to give an insight into the performance and predictability of the model on flood risk management **[5].**

The K-Nearest Neighbors (KNN) algorithm identifies the risk of floods by the distance between the feature vectors of the test samples and the training samples. The most widespread one is the Euclidean distance.

Distance Calculation (Euclidean Distance Formula)

$X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ in an n-dimensional space, the Euclidean distance between them is given by:

$$\mathbf{d(X, Y)} = \sqrt{\sum_{i=1}^{n} (\mathbf{x_i} - \mathbf{y_i})^2}$$

Where:

- x_i and y_i are the feature values (for example, rainfall, water level, temperature, etc.).
- n is the number of features.
- d (X, Y) represents the distance between the two points.

In simple words, the Euclidean distance calculates how far apart two points are by finding the square root of the sum of the squared differences between their corresponding features.

Finding K Nearest Neighbours

- Compute d (X, Y) d (X, Y) d (X, Y) for all training samples.
- Sort the distances in ascending order.
- Select the K nearest neighbours.

Majority Voting for Classification

Let $C_1, C_2, \ldots, C_k$ be the classes of the K nearest neighbours. The classification C of the test point is determined by:

$$\mathbf{C} = \mathbf{arg}\ \max_{\mathbf{c}} \sum_{i=1}^{K} \mathbf{I}(\mathbf{C_i} = \mathbf{c})$$

Where:

- I (C_i = c) an indicator function that returns 1 if the i^{th} neighbour belongs to class c, otherwise it returns 0.
- The summation counts how many of the K nearest neighbours belong to each class.
- arg $\max_{(c)}$ selects the class c that has the highest count.

In simple words, the test point is assigned to the class that appears most frequently among its K nearest neighbours' class with the highest count is assigned to the test point.

Logistic Regression: If the variable of interest (target) is nominal and is one of two categories (i.e. whether an event occurs or does not occur), logistic regression is a statistical model that is applied in binary classification problems. It applies the logistic function to determine the likelihood of occurrence of a binary event to ensure that output probabilities are in the range between 0 and 1. Through training the model to estimate the likelihood of flood episodes based on historical data, i.e. rainfall and river levels, Logistic Regression was employed in flood prediction. It upholds effective methodologies of risk management and provides valuable data concerning the variables that influence the occurrence of floods.

Decision Tree Classification: This method of supervised learning can be applied to classification and regression problems; this is known as Decision Tree Classification. Flood simulation and predictive modeling use it extensively. The method forms a tree-like format through recursion that divides the data into subsets based on the value in the attributes. The three nodes, which are the internal ones, indicate the decisions, and the leaf nodes are the class names. Decision trees were applied to predict floods using the Kerala data published in 2018, based on such variables as water levels and rainfall. Such a strategy aids in effective risk evaluation and risk reduction strategies through providing concise explanations of the decision-making processes.

Random Forest Classification Random: Forest is a decision tree classification method-based ensemble learner. It enhances precision of flood forecasts by using the joint expertise of multiple decision trees, since it can determine complex relationships in the data. The random forest model uses arbitrary subsets of training data to form several decision trees to reduce the risk of overfitting. Random Forest is a great algorithm in identifying the complicated relationships among environmental variables, e.g. rainfall patterns, to create even more precise predictive floods. The ensemble nature of this method diminishes the biases that arise on individual decision trees, offers unbiased information of elements of flood risks, and improves disaster management plans.

Ensemble Learning: Ensemble learning involves several machine learning models, which are joined together to make predictions more resilient and more accurate. Such an approach enhances generalizability, minimizes the weaknesses of a single model, and enhances the flood prediction systems to multiple factors of uncertainties and data variability due to diversity of predictions. A strategy based on ensemble learning made use of KNN, Random Forest, and Logistic Regression and enhanced the overall prediction accuracy and ensured more correct flood forecasts **[6]**.

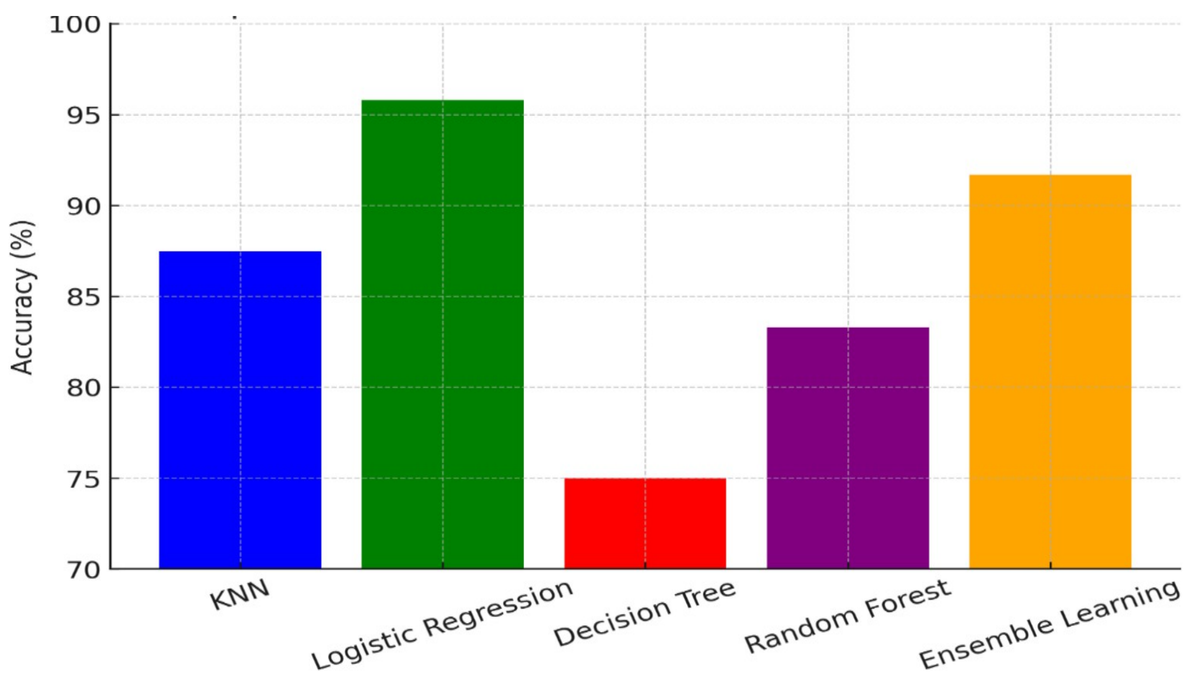

Fig. 1. Comparison of the Model Accuracies for Flood Prediction

Model Architecture

A flood prediction hardware model is typically a model that involves the utilization of diverse environmental sensors to monitor the important factors of the floods like intensity of rainfall, the condition of the water levels in real time. These sensors are strategically located in the flood prone regions to collect the continuous data that are then processed by embedded systems or microcontrollers.

Ultrasonic Sensor: Ultrasonic sensor is a device that uses sound waves, which is ultra-sonic, in determining distances to objects. The ultrasonic sensors may be mounted above water, as in the case of rivers or reservoirs, to make a distance measurement to the water surface. In constant check on the water levels, these sensors would be a source of real time information on water height changes which are essential in flood forecasting and early warning systems. The information that is obtained by the ultrasonic sensors is incorporated into centrally based flood prediction systems. The systems can produce predictive models through continuous monitoring and data analysis to predict the occurrence of floods, act as early warnings, and ensure timely decision-making to minimize the impact of risk to the comma-nighties and infrastructure **[7]**.

Floating Sensor: Floating sensor is the type of sensor that is to be used on water surface or over the water. They are normally buoyant sensors or those which are meant to be floated in water (rivers, lakes or oceans). They can survive the exposure to water, and many are provided with waterproof coverings and floating systems to keep afloat and operational in water. Depending on extensive array of sensor data and Arduino processing, different customized operations have been performed. Such functions are the ability to inform the user about the state of the water level and automatically switch the motor switch off when the tank is full. They apply technologies like ultrasonic or pressure sensors which constantly keep track of how the level of water changes. This information is fundamental in the accurate forecasting and tracking of floods, in real time reporting on the increasing water levels in the rivers, lakes or reservoirs. The information obtained by floating sensors is incorporated into the early warning systems and flood prediction models. With incessant upgrades of predictive algorithms with current environmental data, such systems can predict the occurrence of floods within a shorter timeframe and give prompt warnings to concerned communities and authorities **[8]**.

Arduino Nano: Arduino Nano is a tiny versatile microcontroller board, which is based on the ATmega328P chip. Arduino Nano has the capability of reading sensor data using its digital and analog input pins. It is used to process sensor readings to either be transmitted or processed locally. Arduino Nano would be appropriate in a flood prediction system circuit to interface with sensors installed in flood prone regions, e.g. sensors measuring the water level in rivers or rainfall sensors in urban catchments. It gathers real-time information and conducts local processing and shares important information with the stakeholders or centralized systems to get timely flood notifications and mitigation measures. Arduino nano can perform programmed logic and algorithms to make decisions using sensor information. As an example, it can sound alarms or open actuators (such as pumps or valves) upon flood warnings or a certain environmental condition identified by sensors **[9]**.

GSM Module: GSM (Global System of Mobile Communications) module is a hardware component that provides means of communication between a micro controller or embedded system and the mobile network. A GSM module is a complete GSM system formed by a GSM modem, a microcontroller, a SIM card slot, and communication (either UART or SPI) interfaces. It enables the devices to connect with the mobile network and so transmit and receive information through the SMS (Short Message Service) or the GPRS (General Packet Radio Service). A GSM module may be used in conjunction with Arduino or any other microcontroller-based system within a flood prediction system. It would use SMS or GPRS to send sensor (e.g. water levels) data to a central server or to a cloud platform. The module can also send SMS alerts to local authorities or community in case of any abnormal conditions (such as a sudden increase in water levels) are detected and eliminate timely response and mitigation actions. Since the GSM network has the capability of interconnection and roam throughout the country, and its network power is quite powerful, the user will not require another network **[10].**

Buzzer: Buzzer is a basic electronic appliance that makes a buzzing noise or a beep when an electrical signal is passed to it. In a flood scenario whereby the water level sensors detect the flooding, the microcontroller can use the buzzer to produce a constant sound alerting the residents in the area to seek safety in safer areas. This sound alert is used to supplement other modes of communication so that flood warnings are transferred to all people at risk in the afflicted region even those without mobile phones or digital alerts.

LCD Display: LCD display is a form of screen or optical display which is regulated by liquid crystals and polarization with the help of light. It is also widely applied to flat-panel displays, as part of creating an image through the modulation of light. In the case of a flood, the LCD display will be able to constantly change and display real-time water level information. It can also show warnings and evacuation directions depending on sensor information and assist the residents and emergency responders to make informed choices in real-time.

Potentiometer: A Potentiometer (also spelled out as pot) is a simple electro-mechanical product which offers variable resistance. It usually carries three terminals, two external terminals, linked by a voltage source, and a sliding terminal (wiper) sliding along a resistive element within the tensiometer. The output voltage or the resistance between the wiper terminal and one of the outer terminals depends on the position of the wiper. Potentiometers are important in flood prediction systems to adjust the sensor thresholds and other environmental data input calibration. These variable resistors enable one to fine-tune parameters like water level settings, and rainfall or waterfall intensity settings.

Voltage Regulator: Voltage regulators are very important elements in electronic circuits, which provide the supply of voltage to important elements in the circuit at a constant and steady rate. They minimize the effect of voltage changes on vulnerable electronics, enhance the performance of systems since they maximize power usage, control heat generation, and guarantee that different components that necessitate a certain level of voltage are compatible. Voltage regulators are used in flood prediction systems to ensure a successful functionality and correct data capture to prevent possible damage or failure of the system caused by the unstable power conditions **[11]**.

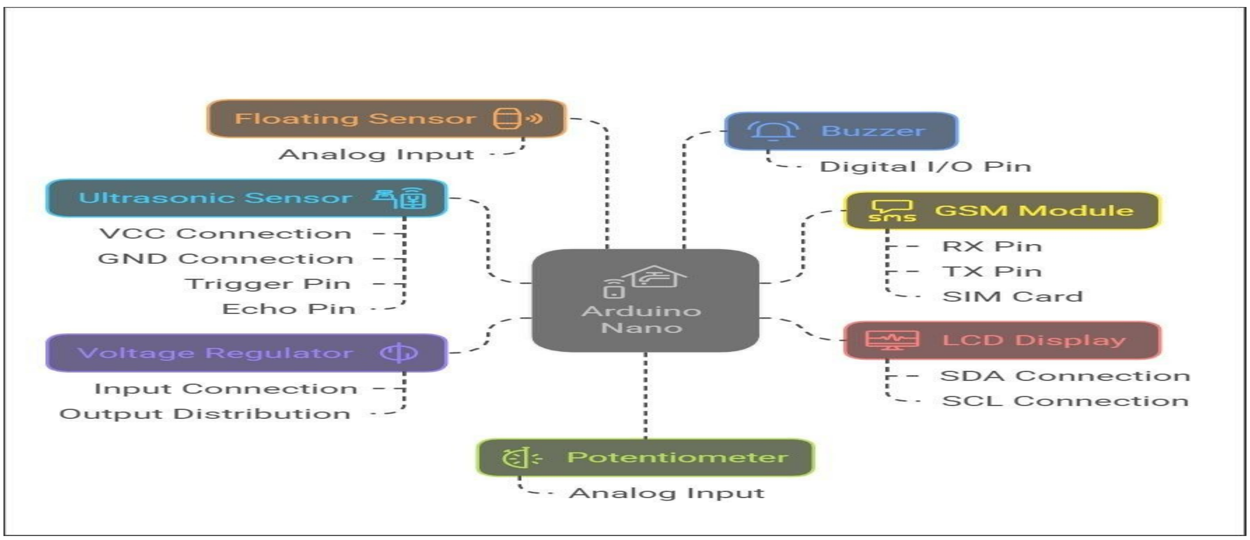

Fig. 2. Model Architecture Diagram

Algorithm for Flood Prediction Using Machine Learning Models

Dataset: Kerala Flood Dataset (kerala_flood_dataset.csv)

This dataset contains environmental parameters used for flood prediction.

Data Preprocessing

Input: Raw dataset (kerala_flood_dataset.csv)

Output: Processed dataset with normalized numerical values **[12]**.

Steps:1

- Load the dataset using pandas. read_csv ().
- Convert categorical labels ('YES' → 1, 'NO' → 0) in the Flood column.
- Split the dataset into features (X) and target (y).
- Normalize the data using Min Max Scaler () to scale all values between 0 and 1.

- Split the dataset into training (80%) and testing (20%) sets using train_test_split ().

Model Training

Input: The scores of various models in terms of accuracy.

Output: Comparative performance in terms of graphs and confusion matrix [13].

Steps: 1 Accuracy of plot models of a bar chart.

Steps: 2 Plot visualization between rainfall and river level through a scatter plot.

Steps: 3 Show a confusion matrix to gain more understanding.

Key Findings: Logistic Regression was found to be the most effective with the accuracy of approximately 95 percent. Voting Classifier was also found to enhance the overall prediction, which made use of several models. Overfitting on Decision Tree resulted in the lowest accuracy.

Future Improvements: - Improve predictions of floods in time-series using Deep Learning (LSTMs, CNNs). –

Add real-time data of IoT sensors to improve flood monitoring. Add more environmental parameters such as wind speed, pressure and soil type [14].

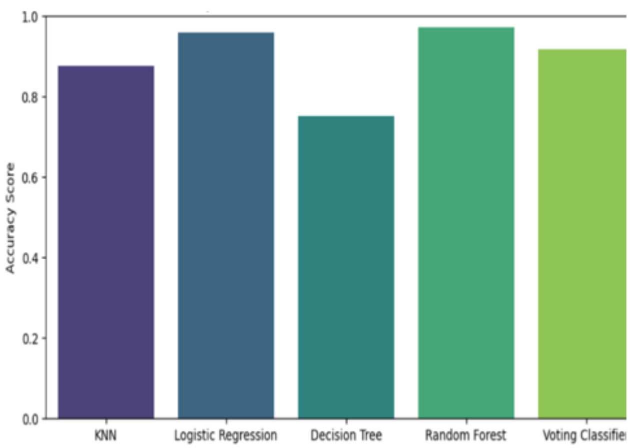

Fig. 3. Comparison of Model Accuracies for Flood Prediction

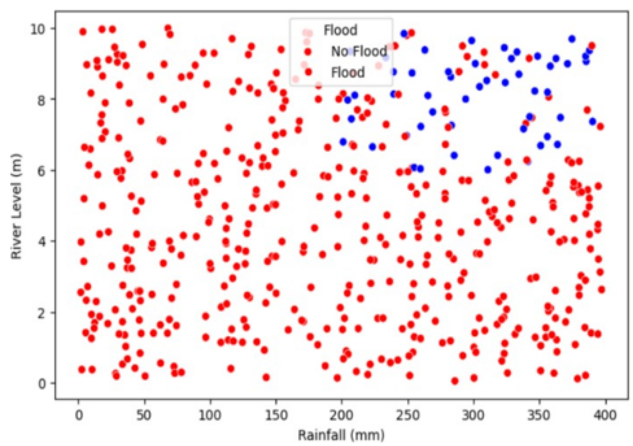

Fig. 4. Rainfall vs. River Level (Flood Occurrence)

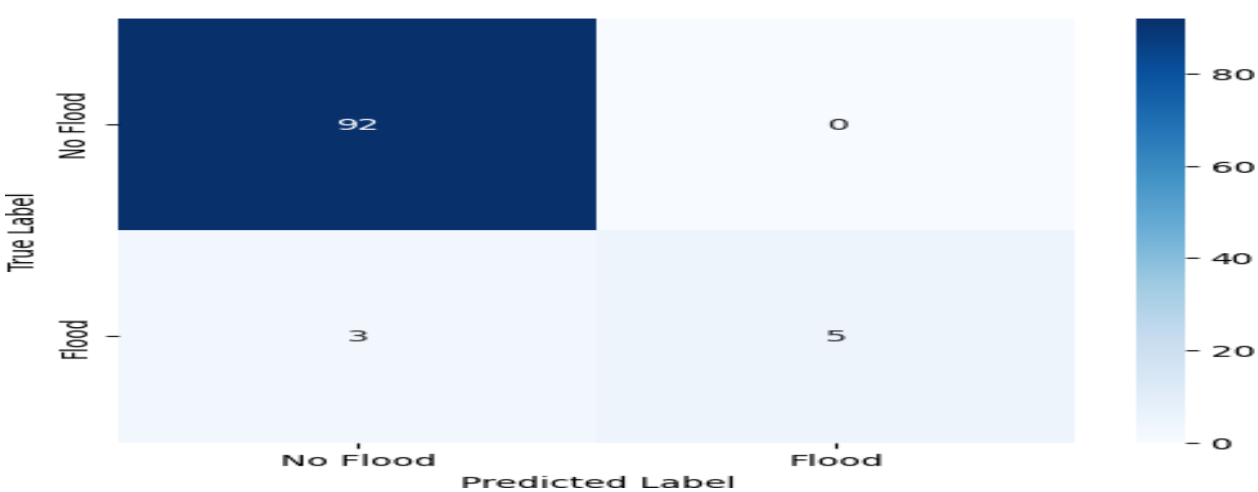

Fig. 5. Confession Matix Predicted Label

Conclusion

The following study demonstrates the way AI-based sensor networks could be groundbreaking in terms of risk mitigation and early forecasting of floods in a timely manner. The suggested solution is very useful in accuracy of floods prediction and future forecasting, as it combines real-time data of multiple environmental sensors with the most advanced machine learning algorithms. The results of the research are informative of the power of the ensemble learning methods in improving the abilities to generate predictions as well as in the assistance of the proactive measures of the response to catastrophes. In addition to the technological success, this paper demonstrates the opportunities of flexible and scalable flood monitoring models which can be integrated into the cloud-based framework and the Internet of Things to process the information and relay the emergency in real times. The potential of such smart systems in protecting communities, preventing destruction of infrastructure as well as minimizing the number of deaths because of extreme weather events are immense. To make the flood prediction systems more reliable and stable, future studies should explore applying deep learning models, real-time combination of IoT sensors, and sophisticated geospatial analytics. With the constant invention and frequent upgrades of such technologies, we will manage to build a stronger and more proactive enterprise of disaster management and, finally, increase the world preparation to floods and other hazards of this nature.

References

1. Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. Augmented Human Research, 5(1), 12.
2. Joshi, A., Kumar, V., Chauhan, N., Kumar, A., & Singh, R. K. (2024, November). The intersection of AI, ML, and industry: A review of emerging trends in real-world applications. In 2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE) (pp. 769-773). IEEE.
3. Mason, V., Andrews, H., & Upton, D. (2010). The psychological impact of exposure to floods. Psychology, health & medicine, 15(1), 61-73.
4. Kumar, V., Kumar, L., Kumar, A., Chauhan, N., & Singh, B. P. (2024). Enhancing autonomous navigation and collision avoidance in drone technology using Deep Reinforcement Learning. International Journal of Technical Research & Science, 9(04), 24-30.
5. Huang, J. C., Ko, K. M., Shu, M. H., & Hsu, B. M. (2020). Application and comparison of several machine learning algorithms and their integration models in regression problems. Neural Computing and Applications, 32(10), 5461-5469.
6. Pandey, D., Niwaria, K., & Chourasia, B. (2019). Machine learning algorithms: a review. Mach. Learn, 6(2), 916-922.
7. Zhmud, V. A., Kondratiev, N. O., Kuznetsov, K. A., Trubin, V. G., & Dimitrov, L. V. (2018, May). Application of ultrasonic sensor for measuring distances in robotics. In Journal of physics: conference series (Vol. 1015, No. 3, p. 032189). IOP Publishing.
8. Tinka, A., Rafiee, M., & Bayen, A. M. (2012). Floating sensor networks for river studies. IEEE Systems Journal, 7(1), 36-49.
9. Kumar, V., Joshi, A., Thakur, G., Chauhan, N., Kumar, A., & Singh, B. P. (2024). Describing the Research Initiative: Unmanned Aircraft Education in Technology. Journal of Recent Innovations in Computer Science and Technology, 1(1), 27-37.
10. Sauter, M. (2010). From GSM to LTE: an introduction to mobile networks and mobile broadband. John Wiley & Sons.
11. Guochen, A., & Zhanyou, S. (2007, August). Programmable voltage regulator design based on digitally controlled potentiometer. In 2007 8th International Conference on Electronic Measurement and Instruments (pp. 1-453). IEEE.

12. Hassan, T., Majeed, S., & Memon, M. Q. (2024, November). Urban Pluvial Flood Prediction Using Machine Learning Models. In 2024 4th International Conference on Innovations in Computer Science (ICONICS) (pp. 1-6). IEEE.
13. Prati, R. C., Batista, G. E., & Monard, M. C. (2011). A survey on graphical methods for classification predictive performance evaluation. IEEE Transactions on Knowledge and Data Engineering, 23(11), 1601-1618.
14. Bukhari, S. A. S., Shafi, I., Ahmad, J., Butt, H. T., Khurshaid, T., & Ashraf, I. (2025). Enhancing flood monitoring and prevention using machine learning and IoT integration. Natural Hazards, 121(4), 4837-4864.