VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2024, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 12, Number 1, January-March 2024 pp:210-224

Flood Prediction System Using IOT & Artificial Neural Network

Muhammad Wajid [1], **Muhammad Kamran Abid** [1], **Asif Raza** [2], **Muhammad Haroon** [3*]
Abdul Qadeer Mudasar [1]

[1]Department of Computer Science, NFC IET Multan, Pakistan; [2] Department of Computer Science, BZU Multan, Pakistan;[3]Department of Computer Science, Xi'an University of Technology, Xi'an, Shaanxi, China

Keywords: *Machine Learning, IOT, Artificial Neural Network, Classification Tree, Back Propagation Neural Network, Regression Tree.*

Journal Info:
Submitted:
December 25, 2023
Accepted:
March 20, 2024
Published:
March 31, 2024

Abstract

Floods pose significant challenges as one of nature's most devastating disasters, making the development of accurate forecast model's complex. This issue has led to severe consequences such as crop loss, population displacement, damage to infrastructure, and disruption of essential services. Advanced research on flood prediction models has played a crucial role in providing policy recommendations, mitigating risks, reducing human casualties, and minimizing property damage caused by floods. In this context, we propose an Internet of Things (IoT)-based flood prediction and forecasting model that prioritizes energy efficiency. Given the limited battery and memory capacity of IoT sensor nodes, we employ an energy-saving strategy within the fog layer, leveraging data diversity to minimize energy consumption. Additionally, cloud technology offers an effective storage solution. To accurately calibrate flood phases, we investigate climatic factors such as humidity, temperature, rainfall, as well as water body parameters including water flow and elevation. Neural networks are commonly used in constructing forecast systems, as they can replicate the complex calculations involved in flood physical processes, resulting in improved performance and cost-effectiveness. In our approach, the Artificial Neural Network (ANN) technique is employed for flood forecasting, and the effectiveness of different algorithms, such as Logistic Regression and Decision Tree, is assessed by comparing them to ANN. Accuracy values are computed using a classification report assessment, and graph parameters are carefully evaluated. Ultimately, our proposed system utilizes the ANN technique to train a predictive model by examining the dataset. This model generates real-time flood risk forecasts through a user-friendly graphical interface.

***Correspondence author email address:** 1201214002@stu.xaut.edu.cn
DOI: 10.21015/vtse.v12i1.1603

1 Introduction

Floods are destructive natural disasters that affect hundreds of millions of people worldwide, causing significant damage to livelihoods, infrastructure, and agriculture. According to the World Meteorological Organization (WMO), floods are the third-largest catastrophe globally [1]. Climate change has led scientists to estimate a potential sea level rise of 4 inches by 2030, which could result in severe flooding in many parts of the world [2]. Research conducted by the Environmental Studies Institute suggests that over 60% of cities worldwide are at risk of flooding in the next 30 years due to rising sea levels.

Floods are the most common expected catastrophe caused by an excess of water overwhelming typically dry land. They can result from heavy rain, rapid snowmelt, or coastal events like hurricanes, cyclones, or tsunamis. In the past decade, floods, tropical hurricanes, droughts, extreme heat, and strong storms have accounted for 80-90% of recognized natural disaster-related tragedies. Climate change is contributing to the increasing frequency and severity of floods, with more intense rainfall events expected to continue in the future. (Source: www.who.com)

Different types of studies in various fields, such as flood data gathering, flood forecasting, flood early warning systems, and flood data visualization, aim to decrease the impact of flood catastrophes by providing timely warnings to affected communities. Technological advancements in cloud computing, machine learning, and data science enable the development of a comprehensive flood disaster management system. This system would efficiently alert areas at risk of flooding by integrating sensing systems and leveraging computational techniques. By utilizing these advancements, it becomes possible to enhance flood preparedness and response, ultimately reducing the negative consequences of floods. [3]

The Internet of Things (IoT) is integral to flood management systems, providing real-time monitoring and early warning capabilities. While it cannot prevent disasters, IoT features are essential for disaster recovery, offering valuable data for preparedness and proactive measures. This technology aids in the creation of flood risk management strategies.

This research proposes a real-time flood forecasting system that utilizes IoT sensors and artificial neural networks (ANN) for data processing. The study focuses on a specific area and collects sensor data using a short-term power device. The accuracy of flood forecasts, the time required for forecasting, the reliability of communication networks, and the cost of system deployment and maintenance are crucial for effective flood detection systems. To address these challenges, the researchers examined three machine learning models and trained them using historical data from the study areas. The goal was to develop models capable of accurately predicting water levels on a daily basis for flood management [4]. By leveraging real-time data and machine learning, the system aims to improve flood disaster management efforts.[5]

Floods occur worldwide each year due to various reasons, making it challenging to completely eliminate this natural catastrophe. However, improving the accuracy of prediction models and implementing effective pre-warning systems can help reduce and even prevent losses caused by floods. Unfortunately, in many emerging countries, the flood prediction systems operated by meteorological and hydrological departments are inadequately equipped and lack the use of IoT technologies. There is also a lack of proficiency and insufficient resources. Consequently, a significant number of people in different countries are affected by floods. Even in well-organized countries, the adoption of effective and economical flood prediction systems is limited. As a result, flood disasters continue to occur, causing harm to life and property without adequate measures to mitigate their impact.

The main concerns regarding natural catastrophes are property damage and loss of lives. To mitigate the effects of floods, an IoT system with sensors is used to measure weather parameters and collect data. Artificial Neural Networks (ANNs) are suitable for estimating flood conditions due to their high-speed data processing capabilities and ability to derive general solutions from specific datasets. Flooding occurs when river water exceeds normal levels, resulting in overflow. Our work focuses on using IoT-based sensors to predict and provide pre-warnings for flood threats by analyzing sensor data, Physically based techniques are complex models that accurately represent system intricacies, while data-driven models require time for improvement and careful consideration of results. Data-driven AI models establish relationships between data features and water levels, whereas physically constructed models aim to replicate the hydrological process precisely. Flood prediction models can be broadly categorized into two approaches: physically-based and data-driven.

The main contributions of this paper are as follows:

- The paper introduces an information-driven model that is useful for continuous prediction in various fields. By utilizing relevant data and applying suitable algorithms, this model is able to generate accurate predictions of the expected outcomes.
- The paper focuses on the application of the information-driven model for real-time flood prediction. It proposes a framework that combines IoT sensing and artificial neural networks to predict floods in real-time. This approach takes advantage of sensor data and leverages the power of edge devices, resulting in a more efficient and timely prediction system.
- The paper emphasizes the implementation of the proposed framework on a low power edge device. This is significant because it allows for resource-efficient prediction capabilities, enabling real-time flood prediction even in areas with limited power availability.
- The proposed model aims to predict flood catastrophes and improve flood prediction using low-energy IOT sensors and wireless technology. It utilizes an ANN to enhance accuracy and incorporates a low-energy gateway node for efficient hardware energy usage. The model also focuses on predicting and controlling sensor sleep intervals to reduce data dimensions and improve energy efficiency.

2 RELATED WORK

The Internet of Things (IoT) is a network of linked embedded devices that communicate without human involvement [6]. It is widely used for monitoring resources and providing early warnings for disasters, such as floods and fires, to prevent loss of life and property damage . IoT features, including information perception, data exchange, and smart data processing, play a crucial role in intelligent control and management. By leveraging IoT-based early warning systems, disaster damage can be minimized [7].

Technology allows for recording elements such as rainfall and storm trajectories. Water level and flow rate, on the other hand, require sensors. The Internet of Things (IoT) enables the transmission of data from a wireless sensor network to control centers, which can then analyze the data and devise flood-prevention strategies [8].

Disaster Management with the Internet of Things (IoT) [9] involves ongoing research projects using IoT technologies to combat weather events, particularly in the field of flood monitoring. These projects utilize IoT-based Integrated Information Systems (IIS) for efficient ecological checking and controlling tasks. Real-time data streaming, processing, and visualization are facilitated through IoT, enabling continuous delivery of information over internet communication networks.[10]

One project suggests using an embedded system based on IoT and machine learning to predict flood likelihood in a river basin. Another project presents a real-time water monitoring system that broadcasts alerts on social media platforms when water levels reach a specified threshold. Both projects utilize wireless sensor networks (WSN) to measure environmental [11] parameters such as water levels, temperature, humidity, and vibration, and employ data mining algorithms to estimate the probability of future disasters.

These initiatives showcase the potential of IoT [12] in disaster management, offering effective monitoring and prediction capabilities to address the challenges posed by weather-related events. WSNs (Wireless Sensor Networks) are a fundamental component and the backbone of IoT systems, requiring a network of interconnected devices. They utilize low-power sensors and microcontrollers to provide AI capabilities for environmental sensing. Examples of WSNs commonly used in IoT include Wi-Fi, BLE, and ZigBee. Communication standards like IEEE 802.15.4, 6LoWPAN, and CoAP are emerging to improve sensor power efficiency and standardize WSNs. However, deploying WSNs in real-world domains presents challenges such as harsh environments, security threats, and power quality issues. The market value of WSNs is expected to grow from 0.45*billionin*2012*to*2 billion in 2022 [13]. WSNs have distinct characteristics in terms of power consumption, data transmission range, and bandwidth. BLE, for example, is a low-power, short-range communication technology crucial for future IoT deployments.

IEEE 802.15.4 is a commonly used protocol in WSNs for the Physical and MAC Layers [14]. Its major characteristics include low power, relatively low cost, short distance, and low transfer speeds. The standard defines two types of edge devices: FFDs and RFDs. FFDs can serve as controllers in PANs or WSNs, interacting with any device. RFDs have limited abilities and are typically used as end devices, relaying sensor data when they wake up from sleep mode. WSNs based on IEEE 802.15.4 [15] can be designed as peer-to-peer or star networks. In peer-to-peer networks, nodes join randomly based on connectivity, while in star networks, one FFD acts as the network controller, with other nodes connecting to it. IEEE 802.15.4 performs well in hostile electromagnetic environments and extends sensor battery life, making it suitable for field applications like flood forecasting.

Bluetooth Low Energy [16] is a low-energy variant of Bluetooth introduced in Bluetooth 4.0. It utilizes a wide spectrum with adjustable frequency hopping for communication. BLE devices can be masters or slaves, with masters being able to handle multiple connections with slave devices simultaneously. BLE communication conserves energy by keeping devices inactive most of the time, resulting in a low energy usage and favorable energy-to-bit-transmitted ratio. This makes BLE suitable for flood early warning systems and other low-power, short-range messaging applications.

The Internet Protocol (IP) is the most widely used and accepted protocol in computer networks. The Wireless Sensor Network (WSN) community has increasingly adopted the IP protocol architecture for WSNs due to its advantages. Using the IP protocol allows sensors to connect to popular IP networks without the need for interpretation gateways. The IP protocol is well-established, documented, and easily accessible to anyone. "IPv6 over Low Power Wireless Personal Area Network" or 6LoWPAN [17] is a protocol designed to make IPv6 more accessible to WSNs. It is often used with the IEEE 802.15.4 PHY and MAC layers.

6LoWPAN offers benefits such as popularity, adaptability, a larger address space, dynamic address setup, and IPv6 security features. However, there are challenges in porting IPv6 to sensing devices, including the larger size of IPv6 data packets compared to the maximum packet size allowed by IEEE 802.15.4. To address this, a fragmentation and maintenance layer was added to enable IPv6 in 6LoWPAN. Additionally, 6LoWPAN employs tunneling and address translation techniques. Another proposed protocol for low-cost devices is the IPv6 Routing Protocol for Low-Power and Loss Networks (RPL), which simplifies the configuration and maintenance of dynamically generated networks.

Instead of using a local server or a PC, a system of reserved servers housed on the internet is used to collect, archive, and process data [18]. Cloud computing can be leveraged in various applications to enhance the benefits of Wireless Sensor and Actor Networks (WSANs) for flood monitoring. Artificial Neural Networks (ANNs) are effective scientific modeling systems [19] that simulate genetic neural networks using interconnected neuronal units. ANNs, particularly Feedforward Neural Networks (FFNNs) [20] and Recurrent Neural Networks (RNNs), are widely used for flood modeling due to their adaptability and accurate estimations. However, there are challenges in network construction, data management, and interpreting the physical aspects of the system being studied [21]. Support Vector Machines (SVMs), introduced by Vapnik in the 1960s, have gained popularity as effective classifiers in various domains. SVMs offer appealing features and promising empirical performance in fields such as electrical engineering, civil engineering, mechanical engineering, medical science, and economics.

DT's ML approach is a contributor to prescient modeling in flood simulation. It utilizes a tree structure with branches leading to leaf objective values. In classification trees (CT), the final factors involve discrete values representing class labels or combinations of attribute names. Regression trees (RT) are used when the objective variable has continuous values, and a group of trees is employed. DTs are fast algorithms, making them popular for flood modeling and forecasting [22].

The CART (Classification and Regression Tree) [17] is a commonly used DT in machine learning for flood modeling. Random forests (RF) are another prominent DT method for flood prediction, which includes multiple tree predictors generating response predictor values based on independent variables. Several studies have focused on flood prediction using wireless sensor networks (WSNs). One study proposed a flood alarm system based on neuro-fuzzy controllers utilizing WSN and the IEEE 802.15.4 protocol. They used the Raspberry PI as a gateway to transmit sensor data to the central server, where flood alarms were generated using historical and sensor data through a neurofuzzy controller approach [23].

Another flood forecasting system utilized WSN with ZigBee [10] for communication and artificial neural networks (ANN) for flood predictions in the river basin. The WSN was built on a ZigBee mesh topology to provide alternative paths for sensor nodes, and data was sent to the central server through GPRS. A ZigBee-based hydrology monitoring system with a star topology was also presented. The authors recommended using ZigBee for sensor communication and GPRS for data transmission to a central hub. Lastly, a flood monitoring system was described, which utilized energy-efficient sensors to measure the quantity of flooding. The system incorporated cloud computing and data streaming as part of the IoT paradigm.

3 PROPOSED METHODOLOGY

The proposed flood prediction methodology utilizes an IoT architecture, as shown in Figure 1. It employs cloud technology to perform flood forecasting on a low-power computing device within the IoT Wireless Sensor Network(WSN). The data acquisition layer collects environmental data from different locations using sensors such as rain sensors, temperature devices, water movement sensors, humidity sensors, and water level sensors. The data is then analyzed at the fog layer to adjust sampling frequency and undergoes preprocessing steps like handling missing values, categorical values, and incorrect results. The fog layer also utilizes Principal Component Analysis (PCA) to reduce data dimensionality before sending it to the cloud layer. The proposed system employs an Artificial Neural Network (ANN) for analysis and is trained on a real dataset using a cloud-based system and a Raspberry Pi. The trained model is used to predict flood water levels over time. The short-term flood prediction system, depicted in Figure 1, involves periodically reading sensor data using BLE wireless communication. The data is filtered at the fog layer to determine the mode of values for water level, humidity, temperature, and rainfall over an hour. The filtered data is then reframed, scaled, and placed at the top of an array. The ten most recent values of water level and rainfall from the array represent the past ten hours of time series data used for prediction.

4 Data Acquisition layer

The data acquisition layer collects a significant amount of data, particularly related to floods and their characteristics in a local region. Sensor nodes in the Internet of Things (IoT) are responsible for gathering data on various climatic and hydrological variables that contribute to flooding. These variables are crucial for accurate flood prediction and forecasting. Table 1 provides information on the properties and corresponding sensors involved in this data collection process.

5 Metrological Parameters

Flooding is influenced by meteorological conditions, particularly during the rainy season. Increased rainfall

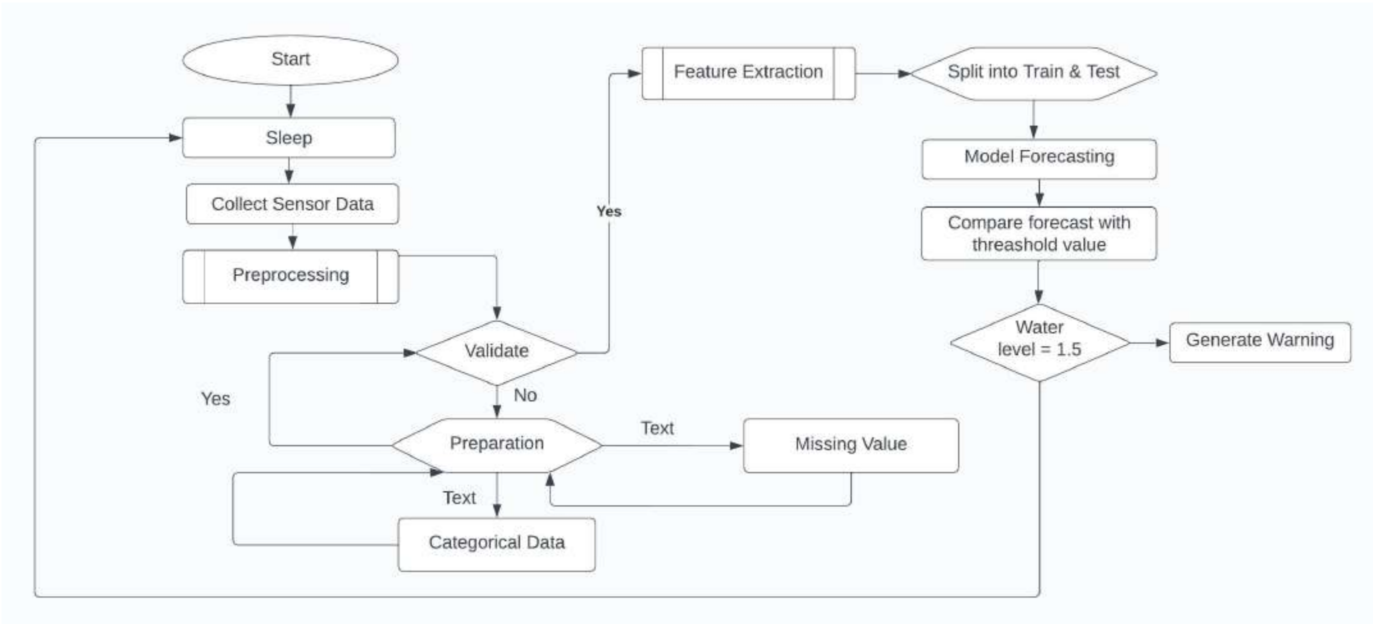

Figure 1. The proposed flood forecast is depicted in flowchart

Table 1. Flood Causing Parameter

Metrological parameter	Explanation	Sensors
Temperature Humidity Rainfall	Moody air temperature Water vapor in air Amount of precipitation	Temperature sensors, humidity sensors Rain gauge precipitation sensors
Season	Winter,pre-Monsoon,Monsoon,Post-Monsoon	
Water level Water flow	Level of water Volume of water flow	Water level sensor Water flow sensor

intensity during this time can result in flood events. Factors such as temperature, humidity, and the monsoon season further contribute to the incidence of flooding.

6 Hydrological parameters

Flood conditions are determined by two main factors: the water level in the bodies of water present in the study area and the rate at which water flows over time. These two elements play a crucial role in understanding and predicting the occurrence and severity of flooding in an area.

7 IOT Framework

The Raspberry Pi can support multiple connections and connect numerous Arduino Nano 33 BLEs through BLE. It processes sensor data and can transmit early warning signals via the Internet, local networks, or an alarm system connected to the edge gateway. The Raspberry Pi used in this study has an internet connection and sends data to ThingSpeak, a cloud-based IoT systems monitoring platform (https://thingspeak.com). ThingSpeak can disseminate early warning messages on various digital platforms like Twitter, Facebook, and others. SMS can also be used for distributing early warning alerts [39].

8 Alternative solutions to IOT Wireless Technologies

Figure 3 provides a list of commonly used wireless technologies in early warning systems for IoT. These technologies are selected based on their suitability for different applications within the diverse IoT ecosystem.

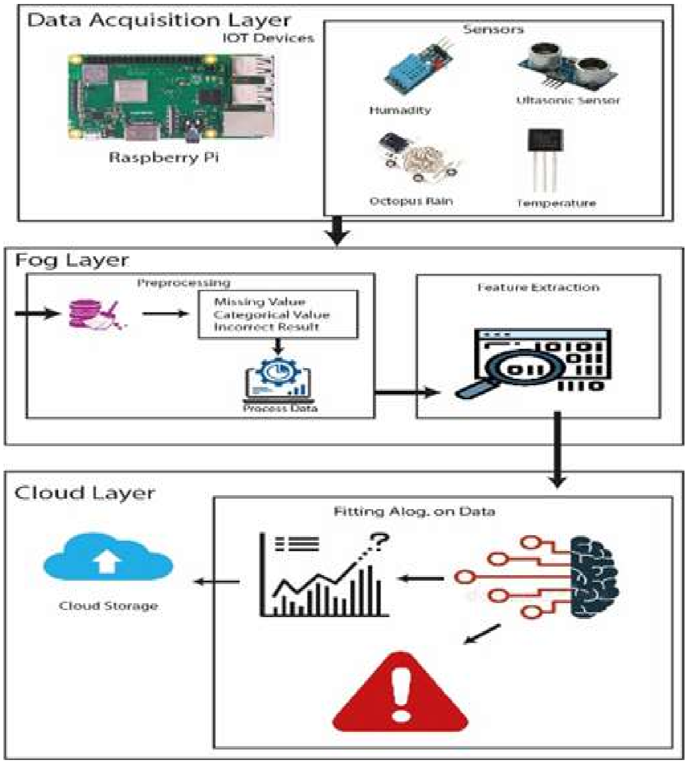

Figure 2. Proposed model

8.1 ZigBee

Zigbee is a wireless technology based on IEEE 802.15.4 that enables short-range communication of sensor information through multiple nodes. It offers data speeds of up to 250 kbps and is suitable for IoT applications within a range of 100 meters. Zigbee's mesh design [40] ensures power efficiency and is commonly used in medium-range IoT applications such as home automation. For instance, it was employed in a flood early warning system to measure water levels and flow, transmitting the data using Zigbee transceivers to nearby locations. [41]

8.2 Wi-Fi

Wi-Fi is a wireless networking technology that provides fast transmission speeds using the IEEE 802.11 standard [42]. It is widely used in applications that require high data rates, like multimedia. However, Wi-Fi is not viable for large networks of power IoT sensors in medical and early warning systems due to its high energy consumption. In a flood predictive model, sensor data from IoT nodes was transported

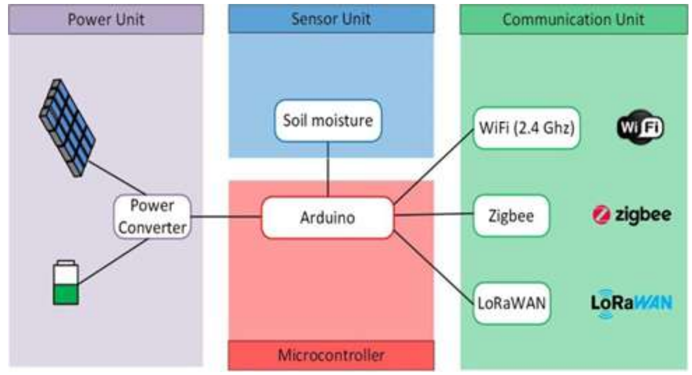

Figure 3. Alternative wireless technologies

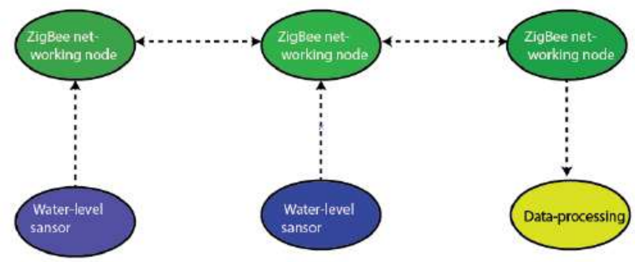

Figure 4. ZigBee Application

through Wi-Fi to a cloud-based analysis system on the Internet. However, the system can be costly to establish and maintain due to the limited battery life of the wireless technology used [43].

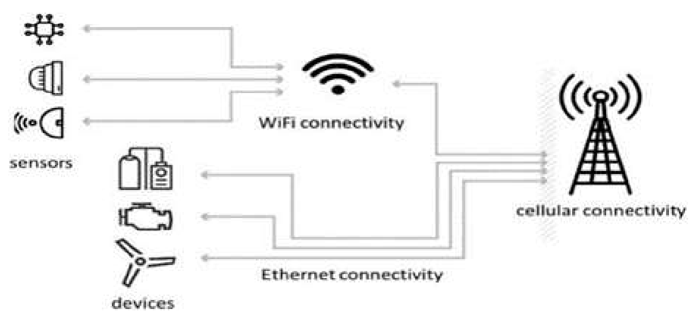

Figure 5. Wi-Fi Technology

8.3 Bluetooth Low Energy (BLE)

BLE (Bluetooth Low Energy) is a variant of Bluetooth technology ideal for battery-powered IoT applications. It operates within a short range of up to 100 meters, providing data speeds of up to 1 megabit per second

[16]. BLE's mesh networking capability allows multiple devices to communicate over a wider area without relying on a single gateway. BLE is well-suited for low-power IoT nodes and edge devices that can operate on battery power. It offers superior energy efficiency compared to Zigbee, as studies have shown BLE to be more effective in terms of bytes exchanged per Joule. This efficiency enables BLE devices to function on battery power for extended periods, potentially reducing the overall cost of implementing systems like flood prediction. Despite having lower bandwidth than WiFi, BLE is adequate for transmitting the required statistical data from flood sensors, such as rainfall and water levels. Its lower bandwidth does not pose a limitation as the quantity of data generated by flood sensors is manageable within BLE's capabilities.

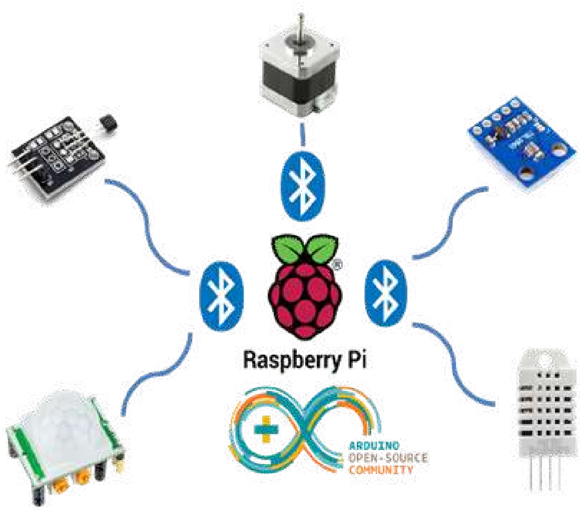

Figure 6. Bluetooth Connectivity

8.4 Raspberry pi 3B+

The Raspberry Pi 3 Model B+ is the final iteration in its series. It features a 1.4GHz Broadcom BCM2837B0 Cortex-A53 (ARMv8) 64-bit SoC, 1GB LPDDR2 SDRAM, and supports 2.4GHz and 5GHz wireless LAN, Bluetooth 4.2, and BLE. It uses a 5V/2.5A DC power supply, supports Power-over-Ethernet (PoE), and can connect to the internet wirelessly. It also has BLE for IoT sensor nodes and can run on battery power. Figure 8 shows the Raspberry Pi 3 Model B+.

Figure 7. Raspberry Pi

8.5 Arduino Nano 33 BLE

The Arduino Nano 33 BLE is a 3.3V low-power Arduino board measuring 45x18 mm. It features a 64 MHz nRF52840 32-bit ARM Cortex M4 CPU, which supports Bluetooth pairing through NFC and offers ultra-low power depletion modes. The board's transportation chipset can function as a BLE and Bluetooth client and host device. This allows for mesh networking between multiple Nano 33 BLE boards without the need for a central server, making it advantageous for our project. Figure 8 displays an Arduino Nano 33 BLE panel.

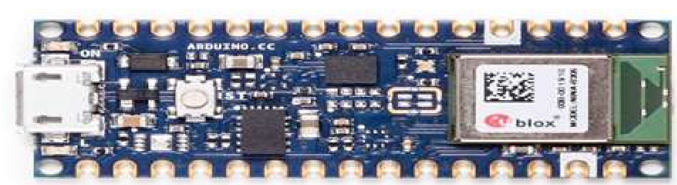

Figure 8. Arduino Nano 33 BLE

9 Sensors

9.1 Water Level Sensor

The water level is measured using the HC-SR04 ultrasonic sensor. This sensor utilizes ultrasonic sound waves to determine the distance to an object. It can measure distances between 2 and 400 cm without physical contact, with a range accuracy of up to 3 mm. The Nano 33 BLE pins are directly connected to the sensor.

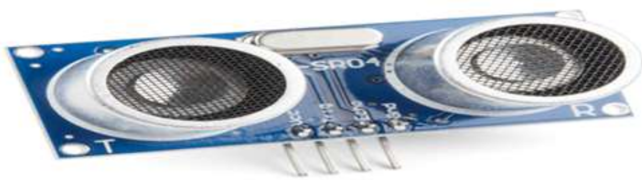

Figure 9. Water level Sensors

9.2 Rain Sensor

The Octopus rain sensor is used to measure the quantity of rainfall. It is connected to the Arduino IO Extension shield, which is linked to the Pi Supply Octopus Rain/Steam Sensor. As the moisture on the sensor surface increases, the output voltage also rises. The sensor can be powered by either 3.3V or 5V.

Figure 10. Rain Sensor

9.3 Humidity Sensor

A humidity sensor, such as the DH11, detects and measures humidity levels in its environment, providing accurate readings through an electrical output. It is commonly used in HVAC systems, weather stations, and industrial processes for efficient monitoring and control.

Figure 11. Humidity Sensor

10 Fog Layer

The fog layer acts as a bridge between the cloud and data gathering layers, pre-processing acquired raw data. It undergoes two steps of data preprocessing using ANOVA and Tukey post hoc tests, which improve energy conservation and reduce data size. PCA is also used to optimize data processing. control.

11 Data Validity / Parameter Identification

Validation procedures determine the model's failure rate, aiming to match the genuine failure rate of the dataset. In real-world scenarios, working with representative data samples is necessary. Specifications for duplicated values, data types, and missing values help identify and address issues. Adjusting model hyper parameters requires unbiased evaluation through sample selection. Incorporating a validity dataset into the model can skew the evaluation. A justification set is used to regularly test the model structure. Data identification aids in understanding the data and selecting the prediction model construction method.

12 Process for validating/cleaning/ and preparing data

This method involves identifying variables based on data structure and type, as well as assessing missing and duplicate values. The data cleaning process varies depending on the dataset. The main objective of data cleaning is to identify and rectify errors and anomalies to enhance the data's value for analytics and decision-making. During model evaluation, a portion of the data, known as the validation dataset, is withheld from training. This data is used to maximize the authenticity and effectiveness of testing when evaluating models.

13 Data Pre-processing

Pre-processing is the transformation of data before feeding it into a procedure to improve outcomes. It involves cleaning up unprocessed data from diverse sources, making analysis difficult due to its lack of uniformity.

14 Visualization Data Analysis Exploration

Data visualization is a critical skill in statistics and machine learning, helping to gain a qualitative understanding of a dataset and identify patterns, outliers, and more. It enables the presentation of crucial relationships through visual charts and graphs.

15 Process for Detecting Outliers

The variety and delivery of feature standards in the input data affect machine learning techniques. Outliers can deceive and skew the algorithm's training phase, leading to slower training epochs, less precise models, and poorer outcomes. In fraud detection and computer security, outliers can signify anomalies that the model was unable to fit on the training data, causing incorrect performance in real-world situations. Cross-validation is a method where we train the model on one subset of the dataset and evaluate it on another, providing a more precise estimate of out-of-sample correctness and better utilization of data.

16 Dimension Reduction

The dimension reduction unit extracts feature from flood-related data, resulting in a smaller, yet accurate data set. Mining this smaller data set yields successful and similar results.

17 Cloud Layer

Flood-related ecological sensory IoT data is stored in the cloud. Data is collected at different time intervals and analyzed using artificial neural networks to categorize flood-related activities.

18 Results And Experiments

Data for implementing the proposed model is collected from official data sources in India's Kerala area. The dataset includes flood-related attributes such as season, temperature, relative humidity, rainfall, and water level. The flood prediction system uses deep learning neural network algorithms (ANN) as well as decision trees (DT) and logistic regression. The implementation is done in Python 3 using different machine learning libraries, with Google Colab serving as the computational resource.

19 Exploring Data

The data in the CSV file has been preprocessed and transformed from nominal to numerical values. Monthly rainfall in Kerala's flood-prone environment is used as an input parameter. The dataset was divided into training data and testing records.

```
data_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1118 entries, 0 to 1117
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   YEAR            1118 non-null   int64
 1   JAN             1118 non-null   float64
 2   FEB             1118 non-null   float64
 3   MAR             1118 non-null   float64
 4   APR             1118 non-null   float64
 5   MAY             1118 non-null   float64
 6   JUN             1118 non-null   float64
 7   JUL             1118 non-null   float64
 8   AUG             1118 non-null   float64
 9   SEP             1118 non-null   float64
 10  OCT             1118 non-null   float64
 11  NOV             1118 non-null   float64
 12  DEC             1118 non-null   float64
 13   ANNUALRAINFALL 1118 non-null   float64
 14  FLOODS          1118 non-null   int64
dtypes: float64(13), int64(2)
memory usage: 14.0 KB
```

Figure 12. Data information

The proposed model requires data that cannot be directly obtained from the environment. However, it can be acquired from official data sources. The dataset consists of 15 columns, including monthly and annual rainfall, with the last column indicating whether floods occur (categorized as "Yes" or "No").

Graphical representation of data using python libraries to show monthly rainfall during the monsoon season. The monsoon regime impacts the climate, with rainfall intensity and duration varying by location. The wettest areas are often along the south-west coast. Figure 13 displays the visualized data. The data history above includes monthly rainfall, annual rainfall, and flood. The heatmap uses colors to rep-

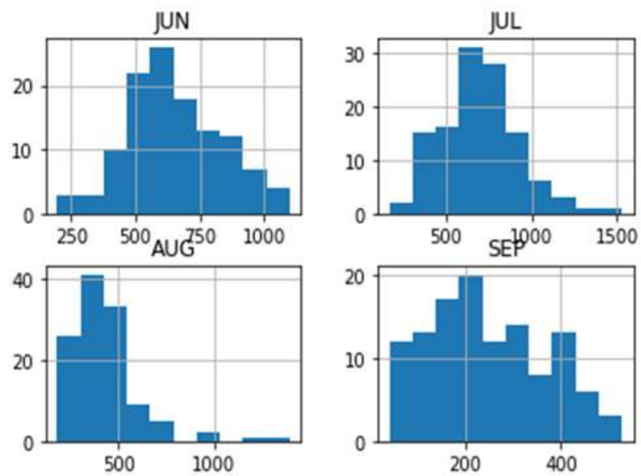

Figure 13. Monsoon rainfall index

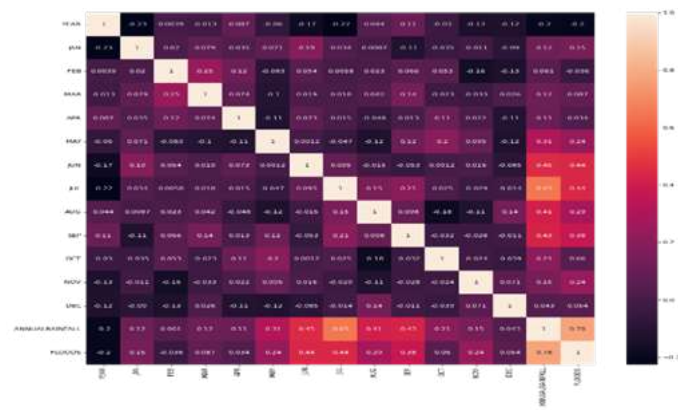

Figure 15. Heat Map Representation Analysis

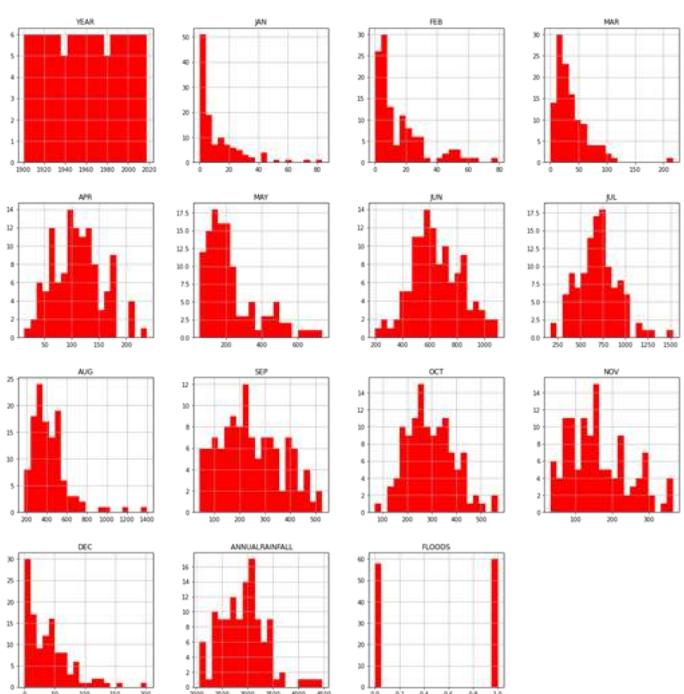

Figure 14. Graphical Representation History

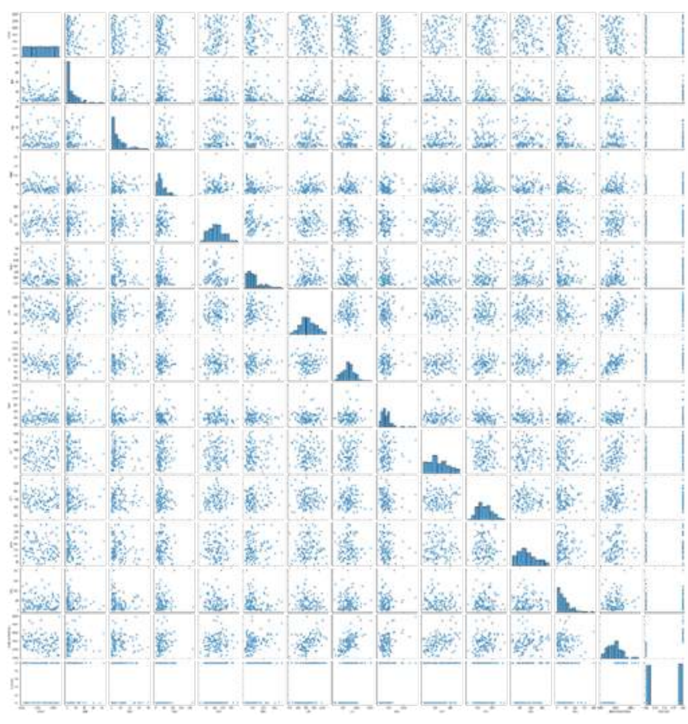

Figure 16. Pair plot

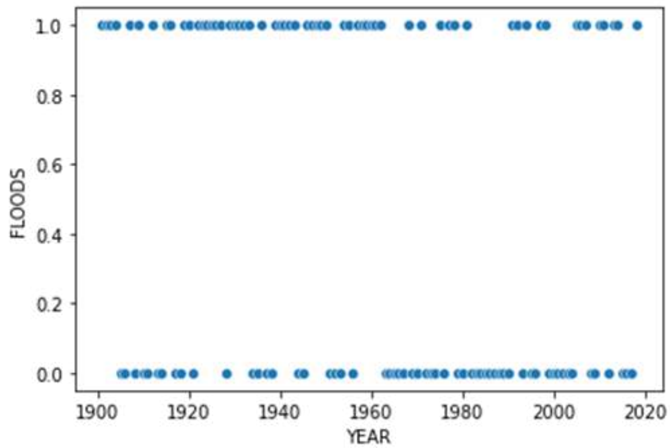

Figure 17. Scatter plot

resent the values of the matrix, with brighter colors indicating more common or higher activities. The graphs form a grid of colored squares, where each square represents the intersection of values from two variables along the horizontal and vertical axes.

A pairs plot is a useful tool for visualizing the distribution of single variables and their relationships. It can help identify patterns and trends for further investigation. Creating pair plots in Python is straightforward. Figure 16 above illustrates a pair plot, with the diagonal representing the distribution of individual variables and the other plots showing joint distributions. Scatterplots are particularly helpful for analyzing quantitative data. Each point in a scatterplot represents a measurement with two coordinates. The X coordinate corresponds to the first variable in the pair, while the Y coordinate represents the second variable and indicates upward or downward movement.

20 Model Visualization

This section describes the implementation of the ANN model using Keras. The model summary provides information about the layers, their sequence, output shapes, and the number of weights in each layer. The ANN is created using the sequential model. Dropout is applied to the data for learning in the hidden layers. Finally, an activation function is applied to obtain the final model. The output shapes and number of weights for each layer are visible. The model has three dense layers with a total of 21,701 trainable parameters.

```
model.summary()

Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 100)               1400

 dense_1 (Dense)             (None, 100)               10100

 dense_2 (Dense)             (None, 100)               10100

 dense_3 (Dense)             (None, 1)                 101

=================================================================
Total params: 21,701
Trainable params: 21,701
Non-trainable params: 0
_________________________________________________________________
```

Figure 18. Layers details ANN Model

21 Result Visualization

The proposed model is trained using deep learning algorithms, specifically an artificial neural network. The dataset is prepared semi-automatically, and two subsets are created: one for training (70%) and one for testing (30%). The model's accuracy is evaluated using a confusion matrix and ROC curve examination, achieving an accuracy of 94.2%.

The neural network achieves a benchmark performance with a batch size of 64, maintaining an accuracy of 94.2% during training. The training process consists of 20 epochs, where each batch passes through the network multiple times. Increasing the number of epochs typically improves the model's accuracy. The accuracy throughout the training and validation process is shown in Figure 18, with a maximum training accuracy of 98.74% and a maximum validation accuracy of 94.2%.

The Loss Function is crucial in measuring the deviation or imperfection of the neural network's predictions. The loss graph indicates that the model performs similarly on both the training and validation datasets. If the graphs start deviating consistently, it may be appropriate to stop training at an earlier epoch.

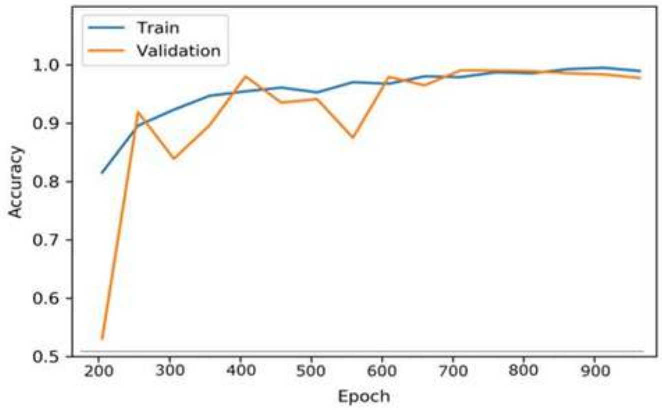

Figure 19. Model Train & Validation Accuracy

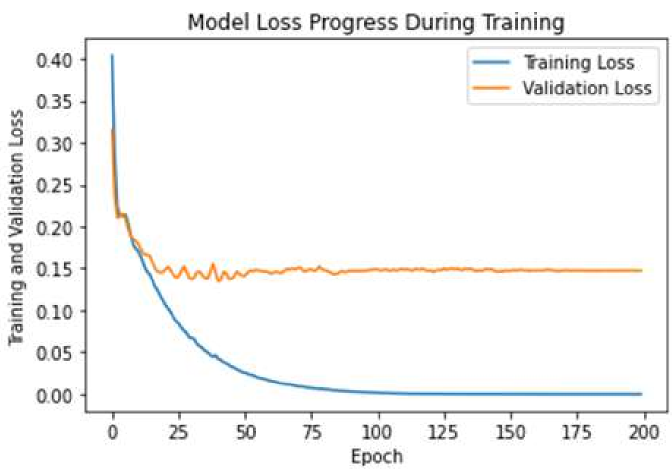

Figure 20. History of Loss

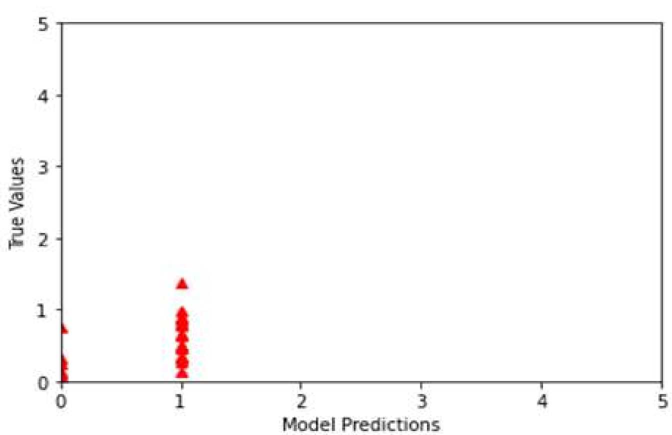

Figure 21. True Value Prediction Flood Come or Not

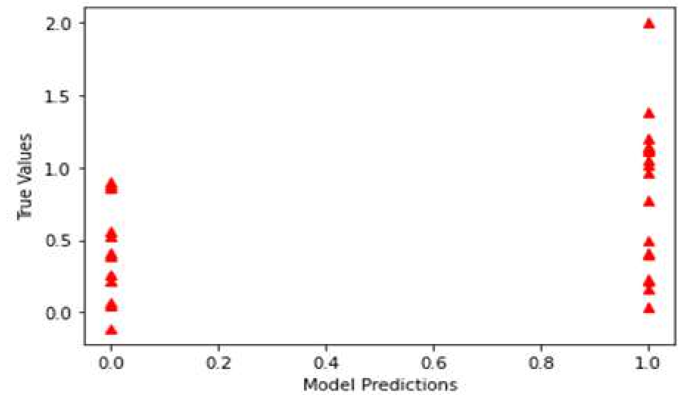

Figure 22. History of Result Accuracy

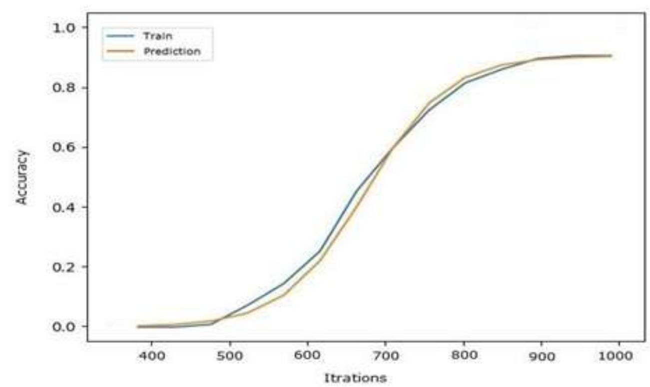

Figure 23. Logistic Regression Model Accuracy

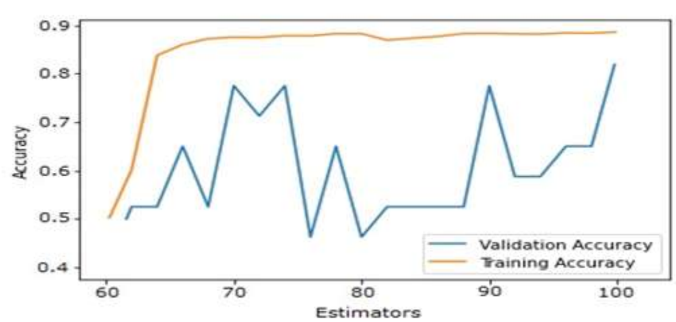

Figure 24. Random Forest Model Accuracy

The overall model shows the flood occurrence probability, with all values ranging between zero and one. Figure 24: Random Forest Model Accuracy Logistic regression demonstrated higher accuracy compared to random forest. LR offers greater efficiency in early flood detection, achieving an accuracy of 89.6% on the training dataset. In contrast, Random Forest achieved an accuracy of 87.9% on the same dataset.

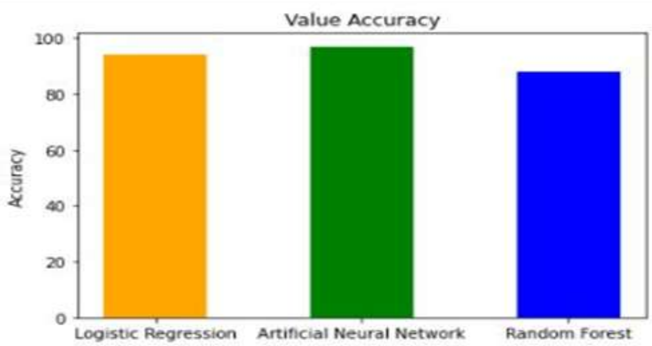

Figure 25. Comparison of LR, ANN, and RF accuracy

22 Discussion

The accuracy scores of Logistic Regression, Random Forest, and Artificial Neural Network techniques are compared in Figure 25. The Artificial Neural Network has the highest accuracy percentage of 94.2% among the three methods, as shown in the graph.

23 CONCLUSIONS

A flood framework is proposed in this dissertation, presenting a cloud-based IoT-based low-powered flood prediction system. Artificial neural networks are utilized to anticipate floods based on real-time rainfall and water data collected by IoT sensors. This architecture contributes to optimistic data development while extending sensor lifetime. Various dimension reduction techniques optimize network bandwidth at the fog layer. The key feature is making predictions using reduced computational infrastructure. Tensor Flow Lite and BLE are used for efficient implementation. The suggested flood prediction system can be installed as a battery-powered IoT device with affordable costs. Comparative analysis indicates that ANN yields better prediction results. Interesting aspects include monitoring additional variables for better predictions, assessing practicality of flood prediction systems, and effective flood alert message delivery.

This system can be made better by extracting other information, such as the speed of the wind and flow of streams, and then feeding those factors into the neural network's design to provide results that are more accurate.

By adding techniques to use location-based services to evacuate communities to the closest safe zone and display alerts or messages on mobile applications, this architecture can be further improved.

Author Contributions

Muhammad Wajid: Conceptualization, Methodology, Software **Muhammad Kamran Abid** : Data curation, Writing- Original draft preparation. **Asif Raza**: Visualization, Investigation. **Muhammad Haroon**: Visualisation and research**Abdul Qadeer Mudasar**: Software, Validation.

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

Funding Information

References

[1] A. A. Ghapar, S. Yussof, and A. A. Bakar, "Internet of things (iot) architecture for flood data management," *International journal of future generation communication and networking*, vol. 11, no. 1, pp. 55–62, 2018.

[2] W. M. Shah, F. Arif, A. Shahrin, and A. Hassan, "The implementation of an iot-based flood alert system," *International journal of advanced computer science and applications*, vol. 9, no. 11, 2018.

[3] B. Choubin, H. Darabi, O. Rahmati, F. Sajedi-Hosseini, and B. Kløve, "River suspended sediment modelling using the cart model: A comparative study of machine learning techniques," *Science of the Total Environment*, vol. 615, pp. 272–281, 2018.

[4] P. Khatri, "Iot strategic research and use case scenario: A direction to the smart life," *Global Sci-Tech*, vol. 10, no. 4, pp. 235–241, 2018.

[5] N.-a. Maspo, A. N. Harun, M. Goto, M. N. Mohd Nawi, and N. A. Haron, “Development of internet of thing (iot) technology for flood prediction and early warning system (ews),” *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 4S, pp. 219–228, 2018.

[6] P.-S. Yu, T.-C. Yang, S.-Y. Chen, C.-M. Kuo, and H.-W. Tseng, “Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting,” *Journal of hydrology*, vol. 552, pp. 92–104, 2017.

[7] C. Chen, Q. Hui, W. Xie, S. Wan, Y. Zhou, and Q. Pei, “Convolutional neural networks for forecasting flood process in internet-of-things enabled smart city,” *Computer Networks*, vol. 186, p. 107744, 2021.

[8] S.-N. Yang and L.-C. Chang, “Regional inundation forecasting using machine learning techniques with the internet of things,” *Water*, vol. 12, no. 6, p. 1578, 2020.

[9] P. P. Ray, M. Mukherjee, and L. Shu, “Internet of things for disaster management: State-of-the-art and prospects,” *IEEE access*, vol. 5, pp. 18 818–18 835, 2017.

[10] M. I. O. Al-lami and M. Ilyas, “Improving scheduling function for ieee802. 15.4 e time slotted channel hopping for wireless sensor networks,” in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE, 2020, pp. 1–6.

[11] S. Jayashree, S. Sarika, A. Solai, and S. Prathibha, “A novel approach for early flood warning using android and iot,” in *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*. IEEE, 2017, pp. 339–343.

[12] P. Singh, G. Bathla, R. K. Singh, and A. Aggarwal, “A novel approach for wireless home automation system using iot,” in *CEUR Workshop Proceedings*, vol. 3058, no. 1, 2021.

[13] S. Bande and V. V. Shete, “Smart flood disaster prediction system using iot & neural networks,” in *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*. Ieee, 2017, pp. 189–194.

[14] Z. Ding, S. Jiang, X. Xu, and Y. Han, “An internet of things based scalable framework for disaster data management,” *Journal of safety science and resilience*, vol. 3, no. 2, pp. 136–152, 2022.

[15] J. Thankappan, K. Mary, D. Raj, D. J. Yoon, and S.-H. Park, “Adaptive momentum-backpropagation algorithm for flood prediction and management in the internet of things.” *Computers, Materials & Continua*, vol. 77, no. 1, 2023.

[16] K. R. Kumar and R. V. Biradar, “An intelligent flood forecasting system using artificial neural network in wsn,” in *Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021*. Springer, 2022, pp. 279–289.

[17] D. Satria, S. Yana, R. Munadi, and S. Syahreza, “Design of information monitoring system flood based internet of things (iot),” in *Proceedings of MICoMS 2017*. Emerald Publishing Limited, 2018, pp. 337–342.

[18] R. Nurpambudi, E. S. P. Wulandari, and R. A. Aziz, “Prediction of flood events in the city of bandar lampung using the artificial neural network,” *Jurnal Infotel*, vol. 15, no. 1, pp. 34–45, 2023.

[19] N. Manohar and A. Archana, “Cloud-based flood prediction using iot devices and machine learning algorithms,” in *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 2021, pp. 754–762.

[20] I. M. Hayder, T. A. Al-Amiedy, W. Ghaban, F. Saeed, M. Nasser, G. A. Al-Ali, and H. A. Younis, “An intelligent early flood forecasting and prediction leveraging machine and deep learning algorithms with advanced alert system,” *Processes*, vol. 11, no. 2, p. 481, 2023.

[21] K. Das, S. Ghosh, and H. S. Dutta, “A machine learning and iot-based novel forecasting approach for flood monitoring and prevention,” *International Journal of Hybrid Intelligence*, vol. 2, no. 2, pp. 175–189, 2023.

[22] W. Dai, Y. Tang, Z. Zhang, and Z. Cai, “Ensemble learning technology for coastal flood forecasting in internet-of-things-enabled smart city,” *International Journal of Computational Intelligence Systems*, vol. 14, pp. 1–16, 2021.

[23] X. Zhang, L. Liu, L. Xiao, and J. Ji, “Comparison of machine learning algorithms for predicting crime hotspots,” *IEEE access*, vol. 8, pp. 181 302–181 310, 2020.