IOT BASED FLOOD PREDICTION SYSTEMUSING EDGE COMPUTING

M. Goudhaman., M.E., (Ph.D)
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai, India

Saradha Devi B[1]
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai, India

Suwetha M[2]
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai, India

Vasundhara S K[3]
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai, India

ABSTRACT-- Natural calamities like floods have a huge impact on the society. It is proven that we have reached a huge height in the technological aspect. Why not we use that in situation likethis, so that the intensity of impact can be reduced. Our IOT based flood prediction system can be implemented to record, analyze the levels of water resources using ultrasonic sensors. The data collected through the sensors are analyzed, processed and filtered and only the necessary data is sent to the cloud, as the system works under the edge computing technology. Once the data gets to the cloud, our system has an automated tool, PLX-DAQ which transforms any form of data into data sets. These data sets are analyzed using the embedded machine learning algorithm and if any abnormalities are found, the respective authorities are notified so that the necessary actions can be taken. So, once the authorities are notified, evacuation ofpeople, safeguarding of assets and livelihood can be achieved to a greater extent. Moreover, the transparency and integrity between people and the government can be achieved through this system.

Keywords- Flood prediction system, Internet of things, Edge computing, Machine Learning

I. INTRODUCTION

This chapter deals with the general introduction, existing system and the proposed system of the project. Overview gives a broad introduction about the project. Proposed system provides a solution to overcome the limitation faced by the existing systems. In our work, we have predominantly used edge computing integrated with the IOT and a machine learning algorithm for flood prediction. Nature sometimes gets aggressive and flips its side from being tender to a monster, so that we can recall whatever we have done to destroy it and such calamities results in a huge collateral damage and all livelihood of people is shaken down to a very ground level i.e., we even require a decade to overcome that. These natural calamities include earthquakes, volcanic eruptions, floods and several to list out. We cannot completely avoid the damage caused by them but can definitely reduce the intensity of the loss and this is achieved with the help of technology. Being an agricultural country, agriculture and cultivation plays a crucial role in our nation's economy and definitely an inseparable component of people's life. Around 70 percent of our country's people relies on agriculture and this is completely destroyed during flood times. Moreover, floods not only have an impact on agriculture but also the disturbs the livelihood of people in all possible ways that includes spreading of water borne diseases, damage to property and assets and so on. Hence it is definitely an alarming situation of all to find a solution to this and it comes through the IOT based flood prediction system using edge computing. IOT based flood prediction system is pretty much an existing system in our society but even then, the after effects of floods is never reduced to a considerable amount. Hence our proposed system is integrated with edge computing and moreover to ensure reliability and efficiency, our system has a machine learning algorithm that analyses the data collected and classifies them into flood and non-flood situations.

There are plenty of real time flooding examples to showcase the effect of flooding situations. One of such real time examples is the event of Cyclone Vardah, which had a massive damage in Chennai region, Tamil Nadu during the timeline of 6th December 2016 to 13th December 2016. Though the time period is about 8 days, the after effects of Vardah went about several years and the collateral damage hit the roof. To study the impact of floods, we can highlight the day 1 of Vardah, where around 10 people were killed as a result of flood-related activities and this count includes 3- year-old child and four women and around 258 trees were uprooted and the major damage was caused due to the uprooted trees and water logging in low laying areas. Several water-borne diseases spread amongthe people which even more the impact as they could have done various precautions if they were aware of the alarming situation. So, if there was an efficient system for prediction, the state situation must have turned into a great positive aspect. The first and foremost need of people was of medical camps as a precautionary measure and the damage caused in low laying areas can be minimized. The main goal of our system is to make sure that people and authorities get to know the seriousness of the situation and we believe in the saying that "precaution is better than cure" and when acting as per the saying, we can make a lot of difference in people's life and the society can be even more upgraded.

A. EDGE COMPUTING

Our flood prediction system works on edge computing technology, in which the data collected is processed in the same place where the data is collected, thereby reducing the transmission delay and latency. Systems like flood prediction must ensure high transmission rate as this impact on the life of people. Many existing flood alert systems works on cloud-based technology and which is the main reason for transmission delay. As per the saying**, "A pie baked at the main location would get cold on its way to a distant customer"** by Michael Clegg, Vice President and General Manager of IoT and Embedded at Supermicro, our system sends only the necessary data to the cloud, significantly improves the time requiredto make a decision based on the data, which is critical for many usecases that utilize real-time decisions like floods and so on. timebut also reduces the costs of building the system and reduces the resources requirement, thus making our system easy to implement in real time. Moreover, when spotlighted upon the security and privacy aspect, edge ensures 100 percent data security and privacy. As an example, to the statement, the privacy of Apple's users is known to the entire world and the Apple Corporation entirely works on edge and all their devices including wrist bands works on edge technology making our system even more reliable.

B. MACHINE LEARNING

Our flood prediction system has a machine learning algorithm embedded to make sure that the results are very accurate and thereby reliable. Machine learning is a branch that comes under the artificial intelligence roof, which makes the system to learn from the data collected and improve their performance and accuracy without being reprogrammed. In machine learning, algorithms are trained to find patterns and features in a huge volume of data in- order to make decisions and predictions based on new data. Experts say that, "The better the algorithm, the more accurate the decisions and predictions will become as it processes more data". And in our system, we have embedded logistic regression algorithm.

II. PROBLEM STATEMENT

Flooding is a natural phenomenon which cannot be avoided to the fullest. The effects and effects of flooding are thoroughly studied and there are certain measures laid down to reduce the severity. The major problems we say in the grounds of flooding are insufficient system to monitor to the water resources and next, it unless the waters hit their roofs. Hence, it is clear that the society needs a system for flood prediction to be relied on.

III.OBJECTIVE

As per our research, it is clear that any flood prediction system can be efficient if it fulfills the following requirements. Hence it becomes our objective to build a system which satisfies the following:

- High transmission rate, which is why we have built the system using edge technology.
- Reliable and Accurate, to ensure these we have embedded a machine learning algorithm.
- Cost efficient, only limited resources are used to build the entire system
- Easy to implement and maintain.

IV. REQUIREMENT ANALYSIS

Requirements are the basic constrains that are required to develop a system. Requirements are collected while designing the system. The following are the requirements that are to be discussed.

1. Functional requirements
2. Non-Functional requirements
3. Environment requirements
   i. Hardware requirements
   ii. Software requirements

A. Functional Requirements

The software requirements specification is a technical specification of requirements for the software product. It is the first step in the requirements analysis process. It lists requirements of a particular software system. The following details to follow the special libraries like Sklearn, pandas, NumPy, matplotlib and seaborn.

B. Non-Functional Requirements

- Process of non-functional steps,
- Problem definition
- Preparation of data
- Evaluation of algorithm
- Improvisation of results
- Prediction of result

C. Hardware Requirements

- Ultrasonic sensor
- Rain detection sensor
- Humidity sensor – DHT11
- 16x2 LCD display screen
- Buzzer
- ESP-12E based NodeMCU
- Arduino UNO

D. Software Requirements

The following are the software requirements required to build an efficient Flood prediction system and they are,

- Arduino IDE
- Embedded C
- Python

V. EXISTING SYSTEM

The existing system was proposed by Eric Samikwa, during the time period of June 18, 2020 in the RISE research Institutes of Sweden. Even this system is built under the IOT domain with a touch of edge technology. The system uses IOT sensors to collect environmental parameters in real time and makes a short- term forecast of the water level using Artificial neural network and various other environmental factors like humidity, temperature, air pressure, which can directly have an impact to the rainfall is also measured, so that the prediction accuracy is improvised when using ANN.

However, rainfall and water level data has proven to be adequate in the prediction of flood water levels using time series ANN analysis with good performance. But when compared with the systems in past, this system utilizes edge computing as the flood prediction is carried on a low power edge computing device which is within the IOT WSN. This means the raw sensor data is not sent over the Internet at all times for the real time prediction of water levels but rather the output of the prediction at relevant intervals. This system operates on low power as it uses low power devices for IoT sensing and edge computing, and utilizes an ultra-low power WSN communication between the sensing nodes and the computing edge platform. Thus, this system can be deployed as a battery powered IoT flood prediction system where Internet access is provided to the edge computing platform and not all sensor nodes. Lastly, unlike the other systems, this system uses a different ANN analysis for better performance and the model has been evaluated on a real dataset.

VI. PROPOSED WORK

The main purpose of studying the existing systems is not only to identify the methodologies used but also to ensure that we uncover the mistakes committed by the existing systems and make sure that our system is flawless. But it is quite difficult to build any system that is 100 percent flawless, but we can try to uncover the errors in the existing system. So, the primary purpose of our flood prediction system is to predict the flood in a very short time with high accuracy and to transmit the same within in a short period of time. Next, we focus on the technical resources, we tried to build oursystem which is less in cost and maintenance with more reliability. Therefore, we implemented the system in IOT domain embedding many sensors to increase the prediction rate, added to itwe have embedded logistic regression algorithm, a machine learningalgorithm, which typically comes under the classification category of supervised learning part of ML.

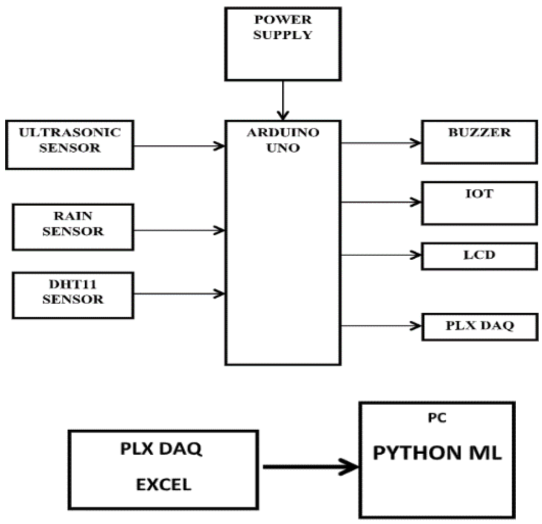

FIGURE 1- Architectural diagram

The IOT sensors embedded are used to monitor the various environmental elements like humidity, temperature, distance which deliberately has a great impact on rainfall which may or may not lead to flood. The data collected fromthese sensors are converted into data sets using an automated tool (PLX-DAQ), which then fed as an input to the logistic regression algorithm for prediction.

VII. SYSTEM MODULES

Modularization is a process that breaks down complex systems into small and solvable parts, in which each part serves for its purpose, to meet the demands. Making any system into modules makes the system easier to debug any errors found in future and also used for reusability. Hence, our flood prediction system is also divided into four modules namely,

- Sensing module
- Data Collection
- Computation Engine
- Flood Prediction

As the main objective of modularity, even in our project, all the modules serve for their own purpose which will be briefed later in the coming sections.

A. Sensing Module

This is the first and foremost step of our flood prediction system, i.e., sensing module. The main objective is to sense and record various environmental elements which has a great impact towards rainfall like the temperature, humidity and so on, using many sensors like DHT11, ultrasonic module HC-SR04. A YL-83 Rain Detector OR RAIN SENSOR is one kind of switching device which is used to detect the rainfall. It works like a switch and the working principle of this sensor is, whenever there is rain, the switch will be normally closed. The YL-83 operates via droplet detection rather than by signal level threshold. An ultrasonic sensor is an instrument that measures the distance to an object using ultrasonic sound waves. An ultrasonic sensor uses a transducer to send and receive ultrasonic pulses that relay back information about an object's proximity. High-frequency sound waves reflect from boundaries to produce distinct echo patterns. This DHT11 Temperature & Humidity Sensor features a temperature & humidity sensor complex with a calibrated digital signal output. By using the exclusive digital-signal-acquisition technique and temperature & humidity sensing technology, it ensures high reliability and excellent long-term stability. All the recordings from these sensors plays as an input to the next module.

B. Data Collection

One of the main problems many systems faced was the data handling, either the system required huge data storage for all the data or the data gets stored in the database but never used them in the future. As our system is built using edge technology, the primary focus is on how we handle the data, only the necessary data is transmitted to the database and they are further more used to train the machine algorithm, thereby the data is handled in a most efficient way. The Sensing devices collects the water level, Pressure of water, humidity, temperature and Rainfall measure to detect the levels of the flood. All collected data will be continuously transmitted, through the Internet communication, infrastructure, to software components designed to compute the water level and to quantify the flood risk.

The collected data is sent directly as a form of dataset (tables) with the help of one of our components, PLX-DAQ, which is an automated tool that straight away converts the data of any form into a tabular 1. The hardware consists of Wi-Fi enabled controller which connects to the server and allows to share the data to through internet. In order to collect datasets for classifying flood data with normal and abnormal condition, raw data has been extracted from IOT sensors. The most important attributes that usually flood monitoring services concentrate on to check flood disaster are temperature, humidity, rainfall and water level. All these above attributes are given as a dataset, to the computation engines to quantify the level of flood. The dataset collected can be viewed in any system and one of such datasets is provided below as an example,

	A	B	C	D	E
1	No.	Temperature	Humidity	Rainfall	Distance Level
2	1	32	30	1	79
3	2	32	30	1	89
4	3	32	30	1	110
5	4	32	22	1	221
6	5	32	27	0	55
7	6	32	33	0	50

FIGURE 2- Sample Dataset

C. Computation Engine

As our system is implemented using Edge computing, which facilitate data processing at or near the source of data generation and in the context of IoT, the sources of data generation are usually things with sensors or embedded devices. The datasets from the data collection module undergoes certain computation of data, and moreover we set the threshold value for flood prediction. All these computations are done within the edge device i.e., piece of hardware that controls data flow, fulfill a variety of roles, depending on what type of device they are, but they essentially serve as network entry-- or exit -- points. In order to reduce the latency and computation time of our system, we only transmit the data to the cloud which is above the threshold value and resulting in abnormality. The computed risks, together with data coming from other sources will be examined by a diagnostic decision system implementing a risk-alert scheduling strategy, able to diagnose the health state of the controlled environment. The sample of such diagnosing is given below, and it is possible to change the indication slogans during implementation. The below table consist of indication field with F (flood) and NF (non-flood) classes.

	A	B	C
1	No.	Distance Level	Indication
2	1	89	NF
3	2	110	F
4	3	221	F
5	4	80	NF
6	5	167	F

FIGURE 3- Sample result after primary diagnose

D. Flood Prediction

In this module, we have embedded a machine learning algorithm to increase the efficiency of the prediction. This module is further more divided into three more modules for better understanding and implementation. The further classified modules are,

- Data Validation and Pre- Processing Technique
- Visualization of Data
- Performance Measurements of ML Algorithm

1. Data Validation and Pre- Processing Technique

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean data set.

2. Visualization of data

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more.

3. Performance measurements of ML algorithm

Logical Regression Algorithm

It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable, in which there are only two possible outcomes, in our system it is Flood or non-flood scenario, which is indicated as F for flood and NF for non-flood. The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable isa binary variable that contains data coded as 1 (yes, success, etc.) or0 (no, failure, etc.). For training this algorithm, we have fed 70% ofdata in learning and 30% of data for testing and this algorithm has an accuracy rate of 96.34%.

VIII. SYSTEM IMPLEMENTATION

A. FROM IOT SYSTEM

Based on the readings of the environmental factors like humidity, temperature, air pressure and the calculated distance using the ultrasonic sensor, our IOT system predicts the flood and non- flood scenario. The rain sensor senses if rain is present or not and gives the following output:

1	No rain
0	Rain

FIGURE 4- Sample output from rain sensor

The ultrasonic sensor which is used in measuring the distance between the obstacle and the sensor, is used in measuring the water level here. So, the logic is that if the rain sensor senses the presence of rain and the distance between the sensor and the water level is too less, there comes the flooding scenario and the result is presented in the 16x2 LCD screen. The various outcomes that can be seen in the LCD screen are,

- Readings from sensors
  Humidity, Temperature, Rainfall, Distance
- Flood Alert, Flood

B. FROM ML ALGORITHM

The algorithm we have implemented is logistic regressionalgorithm, which comes under the classification category of supervised learning. In other words, the logistic regression model predicts P(Y=1) as a function of X. Logistic regression Assumptions: Binary logistic regression requires the dependent variable to be binary. For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome. Only the meaningful variables should be included. The independent variables should be independent of each other. That is, the model should have little. The independent variables are linearly related to the log odds. Logistic regression requires quite large sample sizes. This algorithm gets trained using 70percent of the dataset and uses the rest of 30percent for testing. The output of these type of algorithms is a bivariable, i.e., we can get only two outputs like trueor false, here it is NF or F.

NF	Non-Flood
F	Flood

FIGURE 5- Sample output from ML Algorithm

Based on the rigorous training of the algorithm, the system gets an accuracy of about 96.34percent and since it is a machine learning process, the algorithm's accuracy gets hiked on daily process.

C. FINAL OUTPUT

After the entire prediction is over, it is our sole responsibility to let the people know about the situation, whether it is good or bad scenario. The predicted result gets displayed in the internet and to access the results, we have to access the URL http://www.iotclouddata.tech/233

By publishing the results in the internet, the transparency between the people and the government authorities is established and by accessing it, people can somehow reduce the intensity of damage caused by the flood. And moreover, the government can also take required measures to manage the disaster.

X. CONCLUSION

Technology has paved ways for human race even to reach Mars, but the essence of technology can only be tasted during the calamities during which we stumbled for a meal. Flood is one of the scenarios when a year-old baby stumbled to have a glass of milk, the oldies struggled to have a dose of insulin, and the affected region gets isolated from the entire world in terms of connectivity. We wish in the upcoming years, through implementing our flood prediction system, we can proudly say that all the difficulties we faced belong to the past time-line and the impacts are reduced to a quite considerable amount as we have built a system of 96.34% accuracy, which includes the process of collecting data, processing and predicting. So, as Engineers we must further more focus on various other socio-oriental issues and build a solution, so that we can say that we have used the technology in a wise and significant way.

X. FUTURE ENHANCEMENT

Though our system is of high accuracy, we have further enhancements which includes predicting the paths in which the flood waters may hit high, alerting every individual people living in those areas through SMS via mobile messaging service. By implementing the above ideas, any system for flood prediction can serve for its purpose to the fullest.

XI. ACKNOWLEDGEMENT

We are very much indebted to (Late) Hon'ble Colonel Dr. JEPPIAAR, M.A., B.L., Ph.D., Our Chairman and Managing Director Dr. M. REGEENA JEPPIAAR, B. Tech., M.B.A., Ph.D.,and the Principal Dr. V. Natarajan., ME., Ph.D., to carry out the project here. We would like to express our deep sense of gratitude to Dr. J. AROKIA RENJIT M.E., Ph.D., our guide Head of the Department and also to M. GOUDHAMAN M.E., (Ph.D) for giving valuable suggestions for making this project a grand success. I also thank the teaching and non-teaching staff members of the department of Computer Science and Engineering for their constant support.

REFERENCE

[1]"P. Mitra et al.", "Flood forecasting using Internet of things and artificial neural networks", 2016 IEEE 7th Annual Information IEMCON Technology, Electronics and Mobile Communication Conference (IEMCON), DOI:10.1109/.2016.7746363, 2016

[2]"E. Devaraj Sheshu", "N. Manjunath", "S. Karthik", "U. Akash", Implementation of Flood Warning System using IoT", Second International Conference on Green Computing and Internet of Things (ICGCIoT), DOI:10.1109/ICGCIoT.2018.8753019, 2018

[3]"Minakshi Roy", "Prakar Pradhan", "Jesson George", "Nikhil Pradhan", "Flood Detection and Water Monitoring System Using IOT", International Journal of Engineering and Computer Science (IJECS), Vol- 9 Issue-7, 2020

[4]"Elizabeth Basha", "Daniela Rus", "Design of early warning flood detection systems for developing countries", 2007 International Conference on Information and Communication Technologies and Development, DOI:10.1109/ICTD.2007.4937387, 2007

[5]"Ledisi Kabari", "Rain–Induced Flood Prediction for Niger Delta Sub- Region of Nigeria Using Neural Networks", European Journal of Engineering Research and Science, Vol-5 Issue-9, 2020

[6]"Thinagaran Perumal", "Md Nasir Sulaiman", "Leong.C.Y", "Internet of Things (IoT) Enabled Water Monitoring System", IEEE 4th Global Conference on Consumer Electronics (GCCE),DOI:10.1109/GCCE.2015.7398710, 2015

[7]"Mohammed Khalaf", "Haya Alaskar", "Abir Jaafar Hussain", "IoT- Enabled Flood Severity Prediction via Ensemble Machine Learning Models", DOI:10.1109/ACCESS.2020.2986090, Vol-8, 2020

[8]"M. Khalaf", "A. J. Hussain", "D. Al-Jumeily", "P. Fergus", "I.O. Idowu", Advance flood detection and notification system based on sensor technology and machine learning algorithm", 2015 International Conference on Systems, Signals and Image Processing(IWSSIP),DOI:10.1109/IWSSIP.2015.7314188, 2015

[9]"Sheikh Haroonsafdar", "Malashree G", "Internet of Things based Flood Detection and Monitoring System Using Raspberry Pi", International Journal of Engineering Trends and Technology (IJETT), Vol- 67 Issue-11, 2019

[10]"Anil Gupta", "Sreeja Nair", "Flood risk and context of land- uses: Chennai city case", Journal of Geography and Regional Planning, Vol-3 Issue-12, 2011