International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN :2147-6799 www.ijisae.org Original Research Paper

Towards Efficient Disaster Management: Role of Machine Learning, Deep Learning and WSN Technologies

Rucha Joshi [1], Arya Shinde [2], Supriya Kelkar [3], Mahendra Deore*[4]

Submitted: 26/09/2023 **Revised**: 15/11/2023 **Accepted**: 27/11/2023

Abstract: Disasters are occurrences that have the potential to adversely affect a community via casualties, ecological damage, or monetary losses. Due to its distinctive geoclimatic characteristics, India has always been susceptible to natural calamities. Disaster Management is the management of disaster prevention, readiness, response, and recovery tasks in a systematic manner. This paper reviews various types of disasters and their management approaches implemented by researchers using Wireless Sensor Networks (WSNs) and machine learning techniques. It also compares and contrasts various prediction algorithms and uses the optimal algorithm on multiple flood prediction datasets. After understanding the drawbacks of existing datasets, authors have developed a new dataset for Mumbai, Maharashtra consisting of various attributes for flood prediction. The performance of the optimal algorithm on the dataset is seen by the training, validation and testing accuracy of 100%, 98.57% and 77.59% respectively.

Keywords: *Disaster management, Machine learning, deep learning, prediction, dataset, Wireless sensor networks.*

1. Introduction

The application of WSNs and ML in disaster management systems has attracted significant interest from scholars in recent years. This interest is sparked by the rising number of disasters that occur globally and result in the destruction of a significant number of lives and assets, as well as by the simplicity with which these innovative and affordable solutions may be applied.

WSN is a network of multiple sensor nodes that supports automated data retrieval over a communication architecture. The primary uses of wireless sensors are for disaster preparedness, warning, response, and recovery. Wireless sensors function as monitoring tools during natural and man-made disasters, warning people of impending danger such as fire, landslide, and flood alerts. Machine learning is referred to as an aspect of artificial intelligence that focuses primarily on the modeling of algorithms that enable a computer to learn from data and prior experiences. ML algorithms build mathematical models by training on sample historical data for making predictions or decisions. Smart sensors that can identify disasters such fires or earthquakes can be set up with WSN, and ML techniques can be used to interpret the sensor data and send early alerts to the impacted areas. The location of emergency response teams, vehicles, and equipment can be determined in real-time, using WSN. When planning routes, allocating resources, and making decisions during emergency response operations, ML approaches can analyze this data to optimize the emergency response. ML algorithms can analyze the data that has been acquired from WSN sensors and evaluate the extent of the damage, set priorities for the recovery effort, and assist in making decisions regarding post disaster assessment. Recent advancements in WSNs and Machine Learning has rendered them as one of the most crucial enabling technologies for early-warning systems for natural disasters, extending beyond their initial limitations as basic platforms for continuous, fine-grained monitoring and predictions.

Mumbai, which is situated on India's western coast, is vulnerable to flooding because of its geography, frequent monsoon rains, and growing urbanization. The city faces a significant amount of rainfall during the monsoon season which lasts from June to September. Mumbai is prone to frequent flooding of varying severity due to high rainfall, inadequate drainage systems and incursions of natural water bodies. Mumbai is especially prone to flooding in low-lying neighborhoods, slums, and locations close to rivers, creeks, and the shore. Thus, this paper mainly focuses on Mumbai's demography with the motivation of developing a flood prediction system using WSN and ML algorithms which will help in mitigating losses of lives and property.

The subsequent modules of this study are organized as follows: In Section II presents methodology followed for literature review, Section III discusses literature review,

[1, 2, 3, 4] *Department of Computer Engineering*
MKSSS's Cummins College of Engineering for Women, Pune, India.
[3] *ORCID ID : 0000-0002-4957-2750*
[4] *ORCID ID: 0000-0002-8499-7797*
** Corresponding Author Email: mdeore83@gmail.com*

Section IV formulates the proposed model, Section V discusses the results while Section VI discusses the performance of the ML models and Section VII concludes the article.

2. Methodology

This article presents majorly a review on various implementations of WSN, ML and Deep Learning techniques for disaster management. Stepwise approach was used to identify and choose the articles that were to be included in this work. Initially, all relevant studies were uncovered using keyword streams. The keywords used for the identification process were “Wireless Sensor Networks", “Disaster Management”, “WSN for floods”, “WSN for disaster management”, or “Deep Learning for Disaster Management”. Databases for these articles are Elsevier Science, Springer and IEEE Transactions.

Secondly, the articles obtained from these databases went through a screening process from which duplicates and non-relevant articles were excluded and 44 articles were selected. Lastly, the selected articles were reviewed and then were included for the literature review.

3. Literature Review

A disaster is a phenomenon that causes damage to a neighborhood through the loss of life, environmental harm, or financial losses. The Center for Research on the Epidemiology of Disasters reported that between 1998 and 2017, trillions of dollars were lost by the economy of countries that were disaster-affected. The United States leads the group with losses of almost $1 trillion, followed by India, Japan and China. The UN Refugee Agency reports that during the past 20 years, the frequency of disasters has virtually doubled. The Asia-Pacific region has been among the most exposed to disasters since 1995 [1]. Disaster Management is a methodical approach to managing the responsibilities of disaster prevention, preparedness, response, and recovery. In this section, we review various disasters and their management approaches using WSN, deep learning and machine learning.

A. Floods

Intense rainfall, flood peaks in tributaries and main river, cyclones, landslides leading to hindrance of flow, change in river course and substandard natural drainage system are the primary causes of floods. Various strategies for flood preparedness, emergency response and mitigation are reviewed in papers [2], [3], [4],[5] and [6].

A flash flood is an instantaneous response to precipitation that is of exceptionally high magnitude and occurs quickly. These floods are typically seen in metropolitan areas of civilization when the primary land cannot support. By accumulating rainfall data and utilizing a Multi-Layer Perceptron Classifier (MLP), which belongs to the family of feed forward classifiers of Artificial Neural Networks (ANN), authors of [7] have proposed a system to forecast flash floods in metropolitan areas.

The potential drawbacks for the majority of flood warning systems include internet dependence, delay of information transmission from government to the general public, towers becoming inoperable due to electricity shortages, intricate calculation processes, and immoderate power consumption by the alarm mechanisms. Therefore, authors presented a system that uses less power, functions effectively without internet access, and relies on Zigbee rather than mobile towers to transmit messages instead of alarm messages. All public phones with integrated Zigbee devices have an offline Android app designed for them [8].

In [9] and [10] the authors have used a fusion of social and physical sensing for flood prediction. The model was able to learn the local contextual details efficiently and classified the tweets with 90% accuracy.

This paper surveys the prior research conducted in the field of flash flood warning systems and compares different algorithms such as detection methods like extended Kalman filtering, fuzzy logic, support vector machines, feed-forward neural networks, NNARX-based neural network autoregressive models, adaptive neuro fuzzy inference systems and particle swarm optimization [11]. Unattended ground sensors are transducers like passive infrared, seismic, acoustic, etc. that are installed to detect and distinguish between different movements (UGS) [11].

This paper proposes a search and rescue management method for locating victims post floods using aerial deployment of WSN by implementing MLEACH protocol. It has been modified for efficient routing to ensure little data loss and quick transmission of datagrams to a base station, along with Dijkstra’s algorithm and TSP for data transfer [12].

In addition to a supporting hardware platform, authors deliver a sophisticated next-generation middleware framework designed to facilitate environmental monitoring based on WSN [13].

In [14], an automated unsupervised coastline detection method is proposed using Sentinel-1 time series data, that standardizes the active fluctuations of coastal regions over a constrained amount of time. Images from Sentinel-2 are then used to qualitatively analyze the predicted SAR coastline.

In order to forecast cyclone tracks utilizing geographical locations and several meteorological parameters, authors in [15] suggest a unique data-driven DL algorithm. This model consists of three layers: a CNN layer, a GRU layer, and a multidimensional feature selection layer.

Work in [16] and [17] employ interpolation and data augmentation techniques in order to enhance the temporal resolution of images obtained from satellites. These artificially enhanced images are provided as training data to a CNN model to classify them as either cyclone or no cyclone and to locate the vertex of the cyclone. Authors in [18] propose a model that analyzes weather data and synchronous urban flood data collected by the IoT system and feeds it to a deep reinforcement learning algorithm as training data to generate simulations of the storm surge floods in urban areas near coastal regions.

This work proposes a WSN decision-based model for detection of floods by comparing the collected meteorological data to historical data by using the SVM model. This model transmits the binary decision (flood/no flood) to a cloud server which is secured to monitoring rooms where further decisions are taken [19].

A hybrid deep learning algorithm known as ConvLSTM is presented in [20] which is developed by integrating CNN and LSTM in order to forecast future occurrences of floods from a rainfall dataset.

According to Table 1, it is seen that the majority of papers have utilized techniques such as WSN, DL and Neural Networks whereas the most commonly used tools are Python packages, Zigbee hardware and IOT sensors. The most commonly used input data sources are rainfall data, temperature, humidity, wind speed, urban flood data, etc. From this study, authors of this paper have detected some drawbacks, for instance in [1] it is challenging to synchronize and choose an ideal sampling rate because every sensor has a distinct sample rate. Therefore, there is no common ground available to fuse data. The frequent update of DSDV and its routing tables in [12] leads to increased energy consumption. And the study [15] has not taken into account the possibility of multiple tropical cyclones and intensities of tropical cyclones. While the YOLO model in [12] is proposed for detecting cyclones where it struggles in identification of smaller objects that appear in groups.

B. Earthquakes and Landslides

Earthquakes of high magnitude have occurred in various parts of India viz. Andaman and Nicobar Islands, Kutch, Himachal and the North-East. Himalayan areas are particularly susceptible to earthquakes [21]. Landslides can occur in mountainous areas of India viz., the Himalayas, the Nilgiris, North-East India and the Eastern and Western Ghats. To minimize the loss of property and life, sensor networks can be used in crucial applications like landslide prediction and early warning systems. Due to the widespread use of sensors in locations vulnerable to landslides, clustering is a suitable technique to eliminate unnecessary communication from co-located sensors. [22] proposes CAMP and HBVR as two distributed clustering and multi-hop routing algorithms where CAMP is a novel clustering and routing technology, while HBVR is a BVR with HEED improvement.

The objective of [23] is to design an early warning system for mountain mass landslides. The causes of mountain mass landslides namely earthquakes, rainstorms, landslide displacement, soil texture, human activity and slope gradient, may be identified by employing a back propagation (BP) neural network approach.

According to Table 2, it can be seen that earthquake disaster management studies have utilized techniques such as CAMP+TEEN, HBVR+TEEN, TEEN, LEACH, WSN and BP Neural Networks using sensors with input data sources such as slide surface slope gradient, free slope gradient, soil texture, etc. In [21], authors were yet to investigate the issues in fault-tolerance, scalability, and intra-cluster multihop communication of CAMP - TEEN. As a result, it is impossible to implement the recommended protocols in terms of uniform wear leveling of the network, fault tolerance and scalability in real time. The research conducted in [22] lacked a comprehensive experimental data analysis as the acquired findings were not convincing enough because the proposed algorithm ignored several of the aspects that were challenging to predict during the forecast.

Table 1 Summary of research on floods

Reference No.	Purpose	Techniques	Tools	Input Data
[3][4][5][7]	Flood preparedness, emergency response and mitigation	Remote Sensing,	GIS	Satellite Images, Surveys, Statistics of casualties and damages caused by floods
[6]	Flash flood prediction	Multi-Layer Perceptron (MLP)	Matplotlib	Annual rainfall

[8]	Flood warning system	IoT and android	Zigbee hardware, android	Water level
[9][10]	Fusion of Social and Physical Sensing for flood prediction	Neural networks, WSN, Logistic Regression	AWS, @umbc floodbot	Twitter, Howard County Government website for rain water
[11]	Comparing various flash flood warning systems	Fuzzy Logic, FFNN, SVM, ANFIS, PSO, NNARX, GPS	Transducers, UGS (unattended ground sensors)	Water level, precipitation amount, velocity, acceleration, pressure, altitude, wind speed, and wave current configuration
[12]	Search and rescue management method for locating victims post floods	TSP, MLEACH, WSN	NS3	Heat signature, flow of flood
[13]	Middleware platform for WSN	Multiparadigm WSN programming, DiSent middleware platform	REDE sensor nodes, solar cells	River monitoring
[14]	Coastline detection	Unsupervised learning	Open Street Maps	Satellite images from Sentinel1, SAR and Sentinel2
[15]	Tropical cyclone forecasting	CNN, GRU	Python 3, scikit-learn, keras and TensorFlow	Dataset of real-world tropical cyclones
[16][17]	Cyclone prediction	Deep Learning (YOLO model, R-CNN model)	Python, Google Collaboratory	Datasets of ISRO, NASA, KALPANA-I, Meteorological Oceanographic Satellite Data Archival Center (MOSDAC) and India Meteorological Department (IMD)
[18]	Storm Surge flood predictions for urban areas	Markov Decision Process (MDP) method, SWAN model for numerical simulation, StormWater Management Mode (SWMM) mode	IoT sensors, Python 3.5 packages such as scipy, numpy, copy, datetime, itertools, random	Weather and Urban flood data
[19]	Flood detection	Radial basis function (RBF) kernel used by SVM	Model B and MATLAB Simulink, Arduino Uno R3, Raspberry Pi 3	Sensors at various locations in the area under consideration collected data on the following: temperature, humidity (DH11), water level, air pressure, wind speed, and precipitation (0/1) We also obtained rainfall and air pressure at sea level

				using the Google API
[20]	Flood forecasting	A hybrid deep learning algorithm (ConvLSTM) that combines an LSTM network and a CNN (convolutional neural network)	Scikit-Learn, Tensorflow and Keras, MATLAB packages in python	Data on rainfall collected over a 30-year period at nine flood-prone areas in Fiji from the Fiji Meteorological Service

Table 2 Summary of research on earthquake and landslides

Reference No.	Purpose	Techniques	Tools	Input Data
[22]	Routing Protocols for Landslide Prediction	CAMP+TEEN, HBVR+TEEN, TEEN, LEACH, WSN	N/A	N/A
[23]	Early warning system for mountain mass landslides	BP Neural Network	Sensors	Slide surface slope gradient, free slope gradient, soil texture, etc.

C. Fire

In the middle ranges of Himalaya, India, forest fires are frequent due to lightning, high heat, or negligence on the part of natives. When animals are exposed to fire in the forest, disaster management is crucial for their survival. Timely notifications and responses are crucial for minimizing the damage and protecting the animals. Authors in [24] have presented a potential paradigm for leveraging wireless sensor networks (WSN) for disaster management.

The research in [25] aimed to develop a fire detection system in buildings using WSN by deploying sensors in each room. These sensors collected data such as gas content level, smoke level, temperature level, count of people per room and amount of water available. The model used a K-nearest neighbor algorithm to categorize the conditions as normal, early warning and fire.

According to Table 3, it can be observed that the studies have utilized techniques such as WSN, using classification algorithms such as KNN and sensors, while the most commonly used input data sources are temperature, smoke and detection of gasses. From this study, authors have detected some drawbacks, for instance in [24], not all parameters required to detect fire and alert nearest people of any damage are taken into consideration. Whereas in [25], KNN would not be effective on a sizable data set as the amount of data that is processed increases and more data has to be stored.

Table 3 Summary of Research on Fire Disasters

Referen ce No.	Purpose	Technique s	Tools	Input Data
[24]	Forest fire mitigati on	WSN	Senso r	Temperatur e, Smoke
[25]	Fire detectio n	KNN	Senso r	Temperatur e, Smoke, Gas

D. Drought

Drought is a unique natural disaster that affects both the economy and nature. Drought mitigation necessitates a focused strategy that includes preventative and remedial actions. Thirteen states of India have regularly been identified as drought-prone, while all of India's regions namely regions of Bundelkhand, Orissa and Karnataka experience drought occurrences with various periodicities. Paper [26] provides a critical assessment of the drought monitoring, data management, effect, and mitigation techniques from an Indian viewpoint.

In [27],[28], the author reviews a mechanism to forecast drought by means of WSN using DFAS. The Drought Forecast and Alert System enables the required employees to take preventative actions, such as adjusting agricultural water, to minimize loss.

According to Table 4, the observation is that the studies have utilized techniques such as WSN, GIS, RS and BP Neural Networks whereas the most commonly used input data sources such as rainfall, moisture, cultivated crop, daily mean temperature and soil moisture are collected using sensors.

Table 4 Summary of Research on Drought

Reference No.	Purpose	Techniques	Tools	Input Data
[26]	Drought monitoring	Regression based models for data analytics, GIS, RS	GPS	Rainfall, moisture, crops cultivated,
[27],[28]	Drought forecasting	(DFAS), BP neural network algorithm	Sensors, Decision System (ID2S)	Daily mean temperature, rainfall and soil moisture

E. Use of WSN and ML in the management of disasters other than floods, earthquakes, fires and drought

From the perspective of various disaster management of flood, earthquake, landslide, fire, tsunami etc., the studies are based on different aspects such as risk assessment which is then further divided into two research methods - single and hybrid in [29].

Authors of [30] suggest a disaster response process that can successfully identify potential helicopter landing sites using deep learning autoencoders trained using satellite images, thus reducing cognitive load on an expert image analyst by 70%.

At the moment, accidents at construction sites due to natural disasters such as earthquakes are more frequent with greater severity, and the resulting financial losses are also escalating. This is due to the development of high-rise structures and complicated constructions, as well as a growth in the size of construction sites. In order to minimize the capital losses at construction sites, [31] attempts to use the deep learning algorithm DNN to design a system for predicting financial losses.

There are numerous research projects towards the advancement of wireless sensor networks (WSNs). However, very few of these extensive research projects specifically take disaster mitigation and post-disaster rescue operations into consideration. In [32] a WSN based methodology for collecting data for disaster mitigation, search and rescue operations is proposed.

An overview of current efforts that employ WSN for data collection in disaster zones is presented in [33]. Disaster management projects such as SENDROM, INSYEME, WINSOC, USN4D, AWARE are based on WSN as their basic architecture to collect and communicate useful data.

Natural disasters may compromise the structural integrity of bridges, which may result in accidents, fatalities, and property damage. Use of a structural health monitoring (SHM) system allows for both detection and monitoring of structural health. A WSN based SHM system is proposed in [34]. An ultrasonic sensor measures the water level whereas an accelerometer sensor is utilized to quantify the tilting angle of the bridge pillars. Sensor data is uploaded via visual basic programming to a server or static IP address. To notify the user with data regarding the scenario or condition, an android application must also be developed.

Data on high priority structures, sound and air quality, weather conditions and other sorts can be gathered using wireless sensor nodes. Such information would also be helpful for surveillance and monitoring [35]. [36] proposes a sample integrated system that may be utilized to monitor meteorological parameters including temperature, rain, humidity as well as air quality for the detection of harmful gasses, depending on the situation. Additionally, a load cell has been added to offer ongoing surveillance of flyovers and bridges in order to avoid accidents or structural disasters. Users would utilize this data to manage their daily activities, and the municipal government could use it for any significant planning and decision-making.

Big data applications for disaster management are presented in a comprehensive literature review by [37]. Disaster early warning systems can benefit from big data produced by geoinformatics and remote sensing platforms. Disasters may be predicted by global positioning systems (GPS), and environmental monitoring devices and cloud-based geographic information systems (GIS).

For the transportation and distribution of natural gas, oil, water, sewage, etc., pipeline systems are widely employed. These systems are old, which frequently results in leaks and pipe breaks. Hence, to detect unforeseen issues like leaks early and save them from becoming a major catastrophe, such systems must be regularly monitored. A comprehensive optimization paradigm centered on the Multi-objective Chaotic Ant Swarm Optimization (MCASO) is proposed by [38] with the goal of increasing network performance and WSN energy efficiency. The clustering procedure is carried out using the K-Means++ algorithm, while MCASO is employed for optimization.

[39] explores the possible uses of WSN in the mining industry, particularly in relation to aid rescuers locate trapped workers in an emergency situation. Disaster scenarios are acted out in an artificial mine to evaluate the viability of a proposed prototype sensor node which operates at 433MHz for tracking the health of the miners and localizing them.

According to Table 5, it is seen that majority of papers have utilized techniques such as WSN, GIS and routing mechanisms whereas the most commonly used tools are Python packages, IBM Statistical Package for the Social Sciences (SPSS) and the most commonly used input data such as pressure, temperature, humidity, precipitation, air quality, pH, etc. using sensor nodes. From this study, authors have detected some drawbacks, for instance the acquired financial losses in [31], can be shown in a desktop application or an online application. The simulations conducted in [32] are compared only with the SENDROM model. Only two types of sensors are used in [34] for data collection. Also, the results are displayed on the android application only when a person refreshes the application. Data gathered at the time of disasters from a variety of diverse sources is especially noise-prone in [35]. The experiments conducted in [37] were carried out in a mock mine and therefore the scenario in a real mine might differ. In [38], the proposed routing mechanism has low performance in areas where nodes are likely to collapse leading to increase in probability of packet loss.

Table 5 Summary of WSN and ML applications in other disasters

Reference No.	Purpose	Techniques	Tools	Input Data
[31]	Forecasting financial losses	DNN	Python3.7, NATHAN (Munich Reinsurance Company's Natural Risk Assessment Network), IBM Statistical Package for the Social Sciences (SPSS) V23, CURIE 23	Claim Payouts insurance company
[32]	Disaster mitigation and search and rescue operations	UWSN, terrestrial WSN, ad hoc relay station, SENDROM	ARS	Location ID, Node Type ID
[34]	Structural health monitoring	SHM(Structural health monitoring)	Visual basic software, android application, Accelerometer sensor and ultrasonic sensor	Water level and tilting angle of the bridge
[35]	Monitoring meteorological	Fault tolerance routing mechanism,	Sensor nodes	Temperature, humidity, rain and air quality

	parameters	AODV as routing mechanism, WSN		
[37]	Applications of big data in disaster management	GIS	GPS	Satellite images
[38]	Increasing network performance and WSN energy efficiency	K Means++, MCASO	Python 3.6	Pressure, flow, temperature, pH, conductivity, turbidity.
[39]	Locating trapped miners	Contiki-OS	Sensor nodes	Location

F. Review IoT Architectures for the Disaster Management and Disaster Selection for the Proposed Work

During the literature review, authors of this paper found that multiple papers proposing an IoT architecture for flood data management that can serve as the basis for the implementation of IoT infrastructure that collect, transmit and manage flood related data. The architecture is generally divided into three layers that are perception, network and application layers. In [40], a flash flood management system is proposed, to handle flash flood disasters more effectively. Real-time updates and notifications for flash flood situations can be sent to the community directly by utilizing IoT technologies.

In [41-42], it is suggested to leverage IoT and ANN to anticipate short-term floods, with the prediction computation taking place on a low-power edge device. Using lengthy short-term memory, the system keeps track of rainfall and water level sensor data in real-time and forecasts flood water levels in advance.

[43] provides a review on the research works that utilize IoT for flood data management. The paper then proposes an IoT architecture for the same. Here, architecture is divided into three layers that are perception, network and application. In order to design and assess a flood forecasting model to predict the future occurrence of flood authors of [44] have developed an early warning system that uses LoRaWan network implemented on Arduino and raspberry pi. The model sends predictions with an accuracy of 98% to a cloud server connected to monitoring rooms.

After the extensive literature review, authors found that flood has been and is the most frequent type of natural disaster causing heavy losses to life and property. Thus, authors of this paper decided on focusing on flood management. Mumbai is vulnerable to flooding because of its geography, frequent monsoon rains, and growing urbanization. The city faces a significant amount of rainfall during the monsoon season which lasts from June to September and thus is prone to frequent flooding of varying severity due to high rainfall, inadequate drainage systems and incursions of natural water bodies. Mumbai is especially prone to flooding in low-lying neighborhoods, slums, and locations close to rivers, creeks, and the shore. Thus, this paper mainly focuses on Mumbai's demography with the motivation of developing a flood prediction system using WSN and machine learning algorithms which will help in mitigating losses of lives and property.

4. Proposed Model

A. Proposed Architecture

The proposed architecture consists of 4-tiers as depicted in Fig.1. The perception layer consists of sensors that collect data from surroundings and transfer it to the edge layer. An edge layer is essential for WSN architecture as it can provide data preprocessing, low latency, energy efficiency, fault tolerance, scalability and network autonomy. The edge layer preprocesses the data collected by the perception layer by involving techniques such as filtering, aggregation, compression or simple data analysis, thus reducing the amount of raw data that needs to be transferred to the central server. By processing the data locally edge nodes can respond quickly to events without waiting for data to be transferred to the central node. The pre-processed data from the edge layer is then provided to the central server which uploads the data on a cloud application for analysis and prediction.

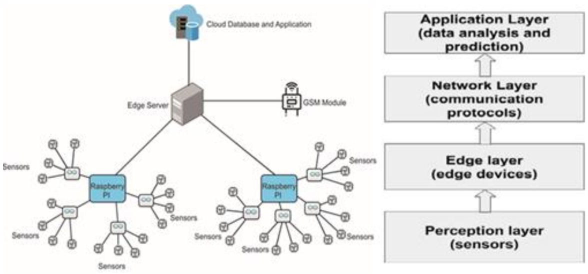


Fig. 1 Proposed 4-Tier WSN Architecture for Flood Management

In the first tier, the sensing nodes namely the water level sensors, temperature and humidity sensors and water flow sensors are distributed all across the water body and continuous data collection takes place with a time delay of a few minutes. The sensors are connected to Arduino devices via wireless communication, which together form the perception layer. Arduino devices are used for the data collection in order to aggregate data from multiple sensors at one location. A cluster consists of multiple sensors and an Arduino device to collect data from those sensors. Furthermore, data is sent from the Arduino device to the raspberry pi device via wireless communication where the data is analyzed in order to decide whether there would be an early warning. All the data collected at the various raspberry pi devices from all the clusters will be stored on an edge server or the central server. Therefore, the edge server acts as the data collection center for the sensor data from all the clusters. In case of an early warning the GSM module will be activated and alerts would be sent to the local authorities. Otherwise, the pre-processed data from the edge layer is then provided to the central server which uploads the data on a cloud application for analysis and prediction.

B. Machine Learning Implementation

The problem of making predictions falls under two categories namely classification and regression depending on the type of input data. Regression algorithms are used to determine continuous outputs such as precipitation values whereas classification algorithms predict discrete data. In this study, initially, authors have used a Kaggle dataset [45] as the training data for developing a prediction model. The dataset consists of input attributes as daily average temperature and precipitation which is mapped to the output attribute of alerts. The alerts vary between a range of four colors, namely, green, yellow, orange and red where, green represents pleasant weather conditions while red represents extreme weather conditions such as heavy rainfall. Thus, the prediction model authors built required to map the input data to discrete alerts making it a classification problem. Various machine learning and deep learning techniques were studied to build this flood prediction model.

ML models are trained using historical data by identifying patterns within the attributes of the dataset based on which a self-learning mathematical prediction model is generated. Therefore, the ML algorithms implemented in this paper come under supervised learning and so the models implemented in this study are supervised machine learning techniques.

Upon research on several machine learning and deep learning approaches, authors have chosen the following four methods to generate the prediction model.

1) ML and Deep Learning Techniques used:

- *Supervised Machine Learning Techniques Used*

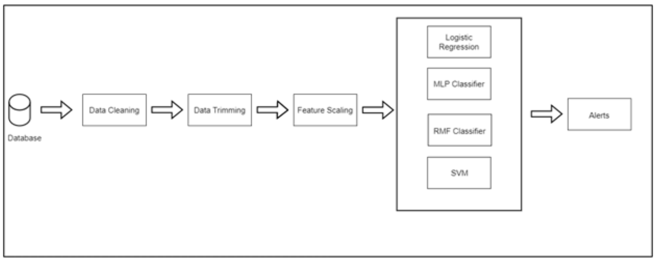
1. Logistic regression: Logistic regression predicts categorical variables by calculating the likelihood that an event will occur or not. It maps output of linear regression function as an input for the sigmoidal function. The sigmoid function generates probability of the occurrence of an event.
2. Support Vector Machine: SVM algorithm distinctly classifies data points by finding a hyperplane in N-dimensional space where N represents the number of features.
3. Random Forest Classifier: It is based on the concept of ensemble learning. It forms multiple decision trees on different subsets of the dataset and then takes their average to increase prediction accuracy of the model. An increase in the number of decision trees leads to an increase in accuracy and eliminates overfitting.

- Deep learning techniques:

4. Multi-Layer Perceptron: It is a fully connected dense network consisting of an input layer where number of nodes is proportional to number of input attributes, multiple hidden layers and an output layer. Every node uses the sigmoid activation function which takes input and maps it to a value between 0 and 1.

Fig.2 illustrates the stages of the prediction model proposed by the authors.

Fig. 2 Proposed Machine Learning model



2) Data Preprocessing

The data passes through multiple data-preprocessing phases before it is used for training the prediction model. Data cleaning, analysis, exploration, and manipulation are done using the Python package namely, pandas. Initially the collected data has to be cleaned in order to increase efficiency of the models in making accurate predictions due to higher quantity of data. This is done by removing unwanted data from the dataset. On removing unwanted data, the rest of the dataset passes through the data trimming phase. In this phase we remove the outliers by using the Interquartile Range method. The final data-preprocessing stage is the feature scaling. Feature scaling is a technique for standardizing the independent features present in the data within a predetermined range. Machine learning models such as linear regression, logistic regression, etc., which use gradient descent techniques require the data to be scaled within range. The data preprocessing stages enhance the model's functionality and training stability.

5. Experimental Results and Discussion

The aim of this study is to provide alerts based on precipitation data. To accomplish this aim, various models have been designed and contrasted with the goal to determine the optimal prediction model.

C. Experimental Results using Kaggle Dataset

Authors have used Kaggle dataset [45] which consists of only two attributes namely daily precipitation and average temperature from the years 1990 to 2022. This dataset is used for initial experiments.

1) Monthly Precipitation Model using Single Attribute of Kaggle Dataset

The Kaggle dataset [45] used by the authors consists of only two attributes namely daily precipitation and average temperature from the years 1990 to 2022. Initially, the daily precipitation data was aggregated by the authors to form the monthly precipitation dataset which consisted of precipitation data for months of June to October for each year along with a binary attribute of yes/no stating whether floods occurred in that year or not. All the selected prediction models were trained on this monthly dataset. The results are represented in Table 6.

It can be observed that none of the models were able to efficiently find patterns between the attributes thus resulting in lower accuracy. The lower accuracy may be due to the fact that average of all the precipitation values was taken leading to loss of valuable data. Furthermore, the aim of the study is to provide alerts for a particular day instead of a particular month of the year. Thus, in the next step the models were trained on the original Kaggle dataset [45] which includes both the attributes, daily precipitation and average temperature.

Table 6 Evaluation of Monthly Precipitation Model Using Single Attribute

Models	Accuracy	MSE	RMSE
Logistic Regression	57.14	0.42	0.65
SVM	28.57	0.71	0.84
RMF	57.14	0.42	0.65
MLP	57.14	0.42	0.65

2) Daily Precipitation Model Using Single Attribute of Kaggle Dataset

All the prediction algorithms were trained on the precipitation attribute of the Kaggle dataset [45] leading to results represented in Table 7. From the table, it can be concluded that all the models were able to capture the patterns in the dataset and give efficient predictions.

Table 7. Evaluation of Daily Precipitation Model Using Single Attribute

Models	Accuracy	MSE	RMSE
Logistic Regression	97.47	0.17	0.41
SVM	99.59	0.026	0.161
RMF	99.69	0.022	0.148
MLP	99.59	0.026	0.161

3) Daily Precipitation Model Using All The Attributes of Kaggle Dataset

These models are not applicable in real world scenarios due to lower complexity; thus, authors implemented the prediction model using all the attributes of the Kaggle dataset [45] namely precipitation and temperature as input to the model and output as alerts attribute. The results are represented in Table 8. It can be observed that the accuracy for each model has been updated due to the addition of temperature attributes. It can be seen that Random Forest Classifier (RMF) is providing the highest accuracy out of all the four algorithms.

Table 8. Evaluation of Kaggle Dataset Models Considering All the Attributes

Models	Accuracy	MSE	RMSE
Logistic Regression	96.06	0.25	0.50
SVM	98.02	0.12	0.34
RMF	99.69	0.22	0.14
MLP	99.12	0.03	0.19

4) Drawbacks of Kaggle dataset

As mentioned above, the Kaggle dataset [45] consists of only two attributes namely daily precipitation and average temperature from the years 1990 to 2022. Although this is a huge amount of data for training the prediction models efficiently, it lacks complexity due to the lesser number of attributes. This model is not ideal for real life prediction as weather is affected by numerous attributes such as humidity, windspeed, pressure, cloud coverage and much more along with temperature and precipitation.

D. Dataset Preparation by the Authors for the Proposed Model

Due to the lack of complexity of the Kaggle dataset [45], authors searched for a more complex dataset to train the prediction algorithms but were unable to find any. Thus, authors built their own dataset from scratch with the help of web scraping tools from a website which provides the historical data [46]. The data was collected for Mumbai, Maharashtra from the weather station of Chhatrapati Shivaji International Airport Station.

This new dataset consists of attributes such as daily precipitation, average temperature, minimum temperature, maximum temperature, humidity, max wind speed, visibility, sea level pressure, dew point and alerts for each day from 1997 to 2022 for the months of June to September. The alerts consist of four values namely green, yellow, orange and red with green representing pleasant weather conditions whereas red represents extreme weather events such as thunderstorms and heavy rainfall.

E. Experimental Results of ML Models using Proposed Dataset

The proposed dataset is preprocessed and then used to train the machine learning algorithms. The data up to the year 2020 was split into training and validation data in the ratio of 80:20. The data from years 2020 to 2022 is used as testing data. The data was trained and tested on MLP Classifiers, Logistic regression, Random Forest classifier and Support vector classifier.

Accuracy, MSE and RMSE are calculated using the following formulae:

$$Accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y} = y_i) \quad (1)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y} - y_i) \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y} - y_i)} \quad (3)$$

The models were evaluated by comparing training, validation and testing accuracy along with MSE and RMSE for each of the models as shown in Table 9.

Table 9. Evaluation of ML Models Using Proposed Dataset

Models	Training accuracy	Validation accuracy	Testing accuracy	MSE	RMSE
LR	93	93.06	72.57	1.79	1.34
SVM	90.95	89.50	71.23	1.71	1.30
RMF	92.15	92.52	73.57	1.55	1.24
MLP	100	98.57	77.59	1.41	1.19

From Table 9, it can be observed that MLP Classifier is outperforming the rest of the prediction models by providing the highest accuracy with train, validation and test accuracy as 100, 98.57 and 77.59 respectively. The lowest MSE and RMSE are provided by the MLP Classifier as 1.14 and 1.19 respectively.

Therefore, authors used the MLP Classifier model to predict the alerts based on real-time data. An API key was generated in order to obtain the real-time data from the Open WeatherMap API. The key provided only four attributes of the real-time data namely temperature (in Kelvin), humidity, conditions and atmospheric pressure. The temperature was converted from Kelvin to degree Celsius and the atmospheric pressure was converted to sea level pressure (in inHg) Hence, the MLP Classifier was trained again using only these four attributes. For this model, the data from 1997 to 2022 was split into training and testing data in the ratio of 80:20. The model was evaluated and provided an accuracy of 76.15% and the MSE and RMSE values as 1.03 and 1.01 respectively. To obtain real-time predictions the data collected from the API was converted into a dataframe and given as input to the prediction model.

6. Performance Evaluation

The ML algorithms were trained and tested using Kaggle dataset [46] and the proposed dataset prepared by the authors. The results are depicted using the Matplotlib library. Fig. 3 represents the accuracy of all models and Fig. 4 represents the MSE and RMSE values for the models when Kaggle dataset [46] is used. It can be observed that the random forest classifier provides the highest accuracy 99.69% and very low MSE and RMSE values of 0.22 and 0.14 respectively. Fig. 5 and Fig. 6 represent the accuracy, MSE and RMSE values for all the models when trained and tested using the proposed dataset. It can be observed that the MLP Classifier gives the highest training accuracy of 77.59 % and the lowest MSE and RMSE values of 1.14 and 1.19 respectively. From these results it can be observed that the models trained on the proposed dataset are more reliable in real-time prediction as models are trained on multiple real-time attributes.

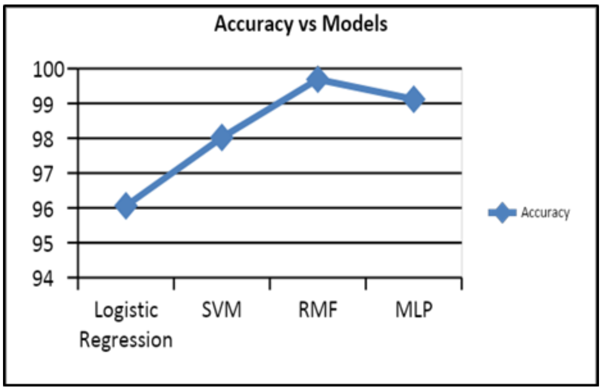


Fig 3 Model Accuracy Comparison Using Kaggle Data

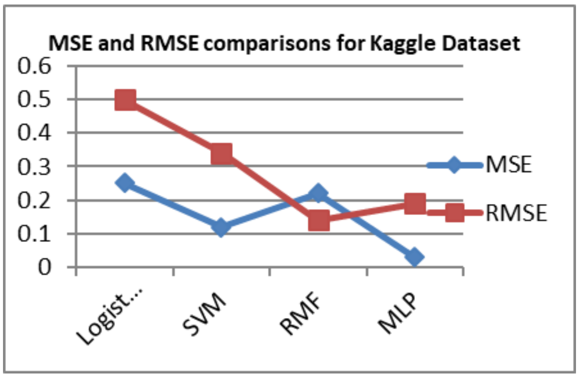


Fig 4 MSE and RMSE comparisons for Kaggle Dataset

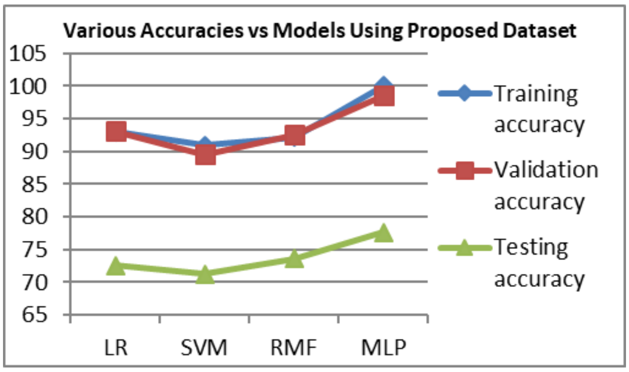


Fig 5. Various Model Accuracies for Proposed Dataset

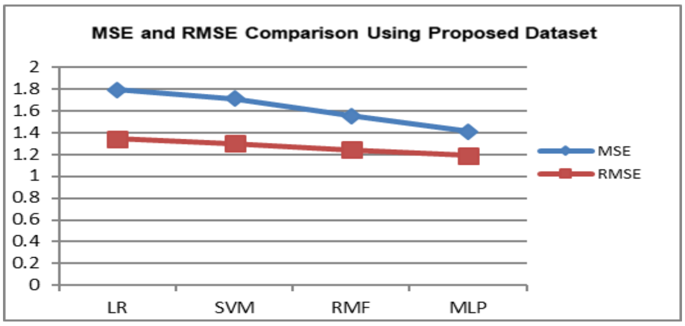


Fig 6. MSE and RMSE Comparison for Proposed Dataset

7. Conclusion

In this paper, authors studied various disaster management systems using WSN, Machine learning and deep learning approaches. Authors summarized the literature and proposed their own 4-Tier architecture for flood prediction and early warning systems. Authors also implemented supervised machine learning and deep learning algorithms to make flood predictions using the Kaggle dataset [46] as the training data. Authors also mentioned the shortcomings of the Kaggle dataset and created their own dataset on which the machine learning algorithms were trained to increase reliability. These algorithms are implemented in the 4th tier of the architecture i.e.: the cloud application.

With the help of the proposed WSN architecture and the newly created dataset, authors propose to design a flood prediction system ideal for forecasting floods and generating alerts for flood prone regions of Maharashtra. The system will collect data from water bodies using wireless IoT devices. This data will be sent to an edge server. The server will analyze the collected data and provide early warnings to users' cellular devices with the help of a GSM module and store the data in a cloud database. The data in the cloud database is given as input to the flood prediction and forecasting algorithms in the cloud application to generate accurate predictions. Authors will also devise an optimal prediction algorithm to be applied in the cloud application by using techniques such as ensemble learning, gradient boosting and time series analysis. Furthermore, authors aim to develop new datasets for other regions apart from Mumbai as well.

References

[1] Abid, Sheikh Kamran & Abid, Muhammad, ‘’Toward an Integrated Disaster Management Approach: How Artificial Intelligence Can Boost Disaster Management. Sustainability’’ 2021.

[2] Neha Singh, "PhD Forum: Monitoring and Detecting Flood by Fusing the Sensor and Social Media Data Streams", 2018.

[3] Javed Alam, "Flood Disaster Preparedness in Indian Scenario",Int. J. on Recent Trends in Engineering & Technology, Vol. 05, No. 03, Mar 2011.

[4] T. Tingsanchali ,"Urban flood disaster management", Procedia Engineering, Volume 32, pp 25-37, 2012

[5] Prakash Tripathi, "Flood Disaster in India: An Analysis of Trend and Preparedness", Interdisciplinary Journal of Contemporary Research, Vol. 2, No. 4, 2015.

[6] Yuliana Rachmawati, Kismartini, and Suharyanto, "The Flood Disaster Management Model in Wonosari Village Semarang City", E3S Web of Conferences 73, 0 ,2018.

[7] K. Ashok Kumar, Dr. Pravin R Kshirsagar, A. Rudra Tapaswi, C. Rohit Yadav, G. Sreeshma, "FLOOD DISASTER PREDICTION USING DEEP LEARNING ALGORITHM", Vol 12, Issue 06, ISSN NO: 0377-9254, 2021.

[8] Jayashree S, Sarika S, Solai A L,Soma Prathibha, "A NOVEL APPROACH FOR EARLY FLOOD WARNING USING ANDROID AND IOT", 2017 Second International Conference on Computing and Communications Technologies (ICCCT 17), 2017.

[9] Neha Singh, Bipendra Basnyat, Nirmalya Roy, Aryya Gangopadhyay, "Flood Detection Framework Fusing the Physical Sensing & Social Sensing", 2020.

[10] Neha Singh, "PhD Forum: Monitoring and Detecting Flood by Fusing the Sensor and Social Media Data Streams", 2018.

[11] Talha Ahmed Khan, Dr. Muhammad Alam, Dr. Zeeshan Shahid, Prof. Dr. Mazliham Mohd Suud, "Prior Investigation for Flash Floods and Hurricanes, Concise Capsulation of Hydrological Technologies and Instrumentation: A survey", 2017 IEEE 3rd International Conference on Engineering Technologies and Social Sciences (ICETSS), 2017.

[12] Harshil Bhatt, Pranesh G, Samarth Shankar, Shriyash Haralikar, "Wireless Sensor Networks for Optimisation of Search and Rescue Management in Floods", IEEE International Conference on Electronics, Computing and Computation Technologies (CONNECT),2021.

[13] Danny Hughes, Jo Ueyama, Eduardo Mendiondo, Nelson Matthys, Wouter Horré, Sam Michiels,Christophe Huygens, Wouter Joosen, Ka Lok Man,Sheng-Uei Guan, "A middleware platform to support river monitoring using wireless sensor networks", Journal of the Brazilian Computer Society volume 17, pages 85–102, 2011.

[14] Ramona Pelich, Member, IEEE, Marco Chini, Senior Member, IEEE, Renaud Hostache, Patrick Matgen, and Carlos López-Martínez, Senior Member, IEEE, "Coastline Detection Based on Sentinel-1 Time Series for Ship- and Flood-Monitoring Applications", IEEE Geoscience and Remote Sensing Letters, VOL. 18, NO. 10, OCTOBER, 2021.

[15] Jie Lian, Pingpong Dong, Yuping Zhang, Jianguo Pan, and Kehao Liu, "A Novel Data-Driven Tropical Cyclone Track Prediction Model Based on CNN and GRU With Multi-Dimensional Feature Selection", IEEE Access (Volume: 8), 10.1109/ACCESS.2020.2992083,2020.

[16] J. Pan, Y. Yin, J. Xiong, W. Luo, G. Gui and H. Sari, "Deep Learning-Based Unmanned Surveillance Systems for Observing Water Levels," in IEEE Access, vol. 6, pp. 73561-73571, 2018.

[17] Snehlata Shakya, Sanjeev Kumar, and Mayank Goswami, "Deep Learning Algorithm for Satellite Imaging Based Cyclone Detection", IEEE journal of selected topics in applied earth observations and remote sensing, VOL. 13, 2020.

[18] Yuewei Wang, Xiaodao Chen, Lizhe Wang, Senior, IEEE, and Geyong Min, "Effective IoT-Facilitated Storm Surge Flood Modeling Based on Deep Reinforcement Learning", IEEE Internet of Things Journal,Volume: 7, Issue: 7, July 2020.

[19] Mohammed Moishin, Ravinesh C. Deo, Ramendra Prasad, Nawin Raj, Shahab Abdulla, "Designing Deep-Based Learning Flood Forecast Model with ConvLSTM Hybrid Algorithm", IEEE Access, volume 7, 2021.

[20] Jamal Al Qundus, Kosai Dabbour, Shivam Gupta, Régis Meissonier, Adrian Paschke, "Wireless sensor network for AI-based flood disaster detection",Annals of Operations Research,2020.

[21] Promad Patil, "Disaster Management in India", Indian Stream Research Journal, Vol2, Issue 1, pp 1-4,2017.

[22] Kalyana Tejaswi, Prakshep Mehta, Rajat Bansal, Chandresh Parekh, S. N. Merchant and U. B. Desai, "Routing Protocols for Landslide Prediction using Wireless Sensor Networks", 2006.

[23] Yuan Yan, Muhammad Aqeel Ashraf, "The application of the intelligent algorithm in the prevention and early warning of mountain mass landslide disaster", Arabian Journal of Geosciences,2020.

[24] Mr.Anand.S.Bhosle, Mr.Laxmikant. M.Gavhane, "Forest Disaster management with Wireless Sensor Network", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) ,2016.

[25] Irawan Dwi Wahyono, Khoirudin Asfani, Mohd Murtadha Mohamad, HA Rosyid, AN Afandi, Aripriharta, "The New Intelligent Wireless Sensor

Network using Artificial Intelligence for Building Fire Disasters", The third International Conference on Vocational Education and Electrical Engineering,2020.

[26] Anil Kumar Gupta, Vinay Kumar Sehgal, "Drought disaster challenges and mitigation in India: Strategic appraisal", CURRENT SCIENCE, VOL. 100, NO. 12, 25 JUNE 2011.

[27] N. Bandyopadhyay, C. Bhuiyan, A.K. Saha "Drought mitigation: Critical analysis and proposal for a new drought policy with special reference to Gujarat (India)", Progress in Disaster Science, Volume 5, January 2020.

[28] Dr. Ravinder Singh Sawhney, "WIRELESS SENSOR NETWORKS FOR DISASTER MANAGEMENT", International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 5

[29] Ling Tan, Ji Guo, Selvarajah Mohanarajah, Kun Zhou, "Can we detect trends in natural disaster management with artificial intelligence? A review of modeling practices", Natural Hazards,2021.

[30] Vyron Antoniou, Chryssy Potsiou, "A Deep Learning Method to Accelerate the Disaster Response Process", 12, 544,2020.

[31] Ji-Myong Kim, Junseo Bae, Seunghyun Son, Kiyoung Son, Sang-Guk Yum,"Development of Model to Predict Natural Disaster-Induced Financial Losses for Construction Projects Using Deep Learning Techniques", 2021.

[32] Suman Saha, Mitsuji Matsumoto, " A Framework for Disaster Management System and WSN Protocol for Rescue Operation", TENCON IEEE Region 10th Conference,2007.

[33] Imane Benkhelifa, Nadi Nouali, Samira Moussaoui, "Disaster Management Projects Using Wireless Sensor Networks: An Overview",28th International Conference on Advanced Information Networking and Applications Workshops, 2014

[34] Miss.Pooja Krishnath Patil, Prof.Dr.S.R.Patil, "Structural Health Monitoring system using WSN for bridges",International Conference on Intelligent Computing and Control Systems,2017.

[35] Renshu Wang, Bin Chen, Jing Zhao, "The Wireless Sensor Network (WSN) for Meteorological Monitoring in Transmission Lines", IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference, 2019

[36] Sonar Padwal, Ashwini Holkar, Shubhangi Khote, Prajakta Maral, Vidya Kadam, "Using Wide Area Monitoring WSN", International Conference on Information, Communications & Embedded Systems.

[37] Muhammad Arslan, Ana Roxin, Christophe Cruz, Dominique Ginhac, "A Review on Applications of Big Data for Disaster Management", 2018

[38] Yandja Lalle, Maroua Abdelhafidh, Lamia Chaari Fourati, Jihene Rezgui, " A hybrid optimization algorithm based on K-means++ and Multi-objective Chaotic Ant Swarm Optimization for WSN in pipeline monitoring",15th International Wireless Communications & Mobile Computing Conference (IWCMC), 2019.

[39] Idrees Zaman, Anna Forster, Asad Mahmood,Frederick Cawood,"Finding Trapped Miners with Wireless Sensor Networks",978-1-5386-6638-8/18/$31.00, IEEE

[40] Alex Wai, Dr Mohd Fo'ad bin Rohani, "Flash Flood Management System Using IoT Technology",UTM Computing Proceedings Innovation in Computing Technology and Applications Volume: 2| Year: 2017 |ISBN: 978-967-0194-95-0

[41] E. Samikwa, T. Voigt and J. Eriksson, "Flood Prediction Using IoT and Artificial Neural Networks with Edge Computing," 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), Rhodes, Greece, 2020.

[42] Hassan, H & Mazlan, M & Ibrahim, T & Kambas, M, "IOT System: Water Level Monitoring for Flood Management",IOP Conference Series: Materials Science and Engineering,2020.

[43] Abdul Ghapar, Azimah & Yussof, Salman & Bakar, Asmidar, "Internet of Things (IoT) Architecture for Flood Data Management", International Journal of Future Generation Communication and Networking,2018.

[44] Al Qundus, Jamal & Dabbour, Kosai & Gupta, Shivam & Meissonier, Régis & Paschke, Adrian, "Wireless Sensor Network for AI-based Flood Disaster Detection",Annals of Operations Research. 319. 10.1007/s10479-020-03754-x,2020.

[45] Ritwek Khosla, "Weather data Indian cities (1990 to 2022)", Kaggle, doi:https://www.kaggle.com/datasets/vanvalkenberg/historicalweatherdataforindiancities.

[46] https://www.wunderground.com/weather/VABB