International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Aug-2022)

https://dx.doi.org/10.12785/ijcds/120149

Adaptive IoT Enabled Flood Management System: An Artificial Intelligence Approach

Himanshu Rai Goyal[1], Kamal Kumar Ghanshala[2] and Sachin Sharma[3]

[1,2,3] *Department of Computer Science aand Engineering, Graphic Era (Deemed to be University), Dehradun, India*

Received 9 Oct. 2021, Revised 19 Jul. 2022, Accepted 31 Jul. 2022, Published 6 Aug. 2022

Abstract: Flood control is one of the most important areas where practical measures have been implemented to minimise damage. We are developing an IoT-enabled adaptive AI technique for efficient fluid management in this research project. In recent years, floods in Kerala, India, have caused significant damage to people, infrastructure, and the environment. Precipitation rates, temperature, and other data were collected by various IoT devices. Data on floods in Kerala is collected and sorted into two categories: 80% for training and 20% for testing. Six different machine learning models are trained, and the proposed hybrid model achieves a maximum accuracy of 99.64%. We were able to avoid losses as a result of integration.

Keywords: Artificial Intelligence, Flood management, IoT, Kerala floods, Machine Learning

1. Introduction and Overview

Wireless communications services are in high demand, and they require a high data rate with reliable connectivity, which is difficult to offer because the wireless mobile system is being deconstructed multi-way [1]. Human people have always attempted to change their fate, whether by nature or by choice. It was frequently a failure, but the tests and experiments always resulted in improvements and solutions. A calamity is defined as a collection of events that disrupt the financial or other critical communities. For decades and millennia, reliving the place and its reason in the past has been a virtually impossible task. In today's environment, innovative experiments employing the same rejuvenation plan have been done. The traditional plans for safe storage spaces alter the efficient operation of entire cities merely to increase their long-term viability and stability in the event of a disaster. A situation like this is an occurrence that must be thoroughly addressed. And the pressure of obligation is frequently too much for people or our old network equipment to handle. Today's connectivity solutions can provide a breath of fresh air for these management systems. IoT-based auto-activating systems have impacted the majority of researchers. It can be programmed as a wireless self-activation system to help us with the tasks and issues we face. IoT negates the physical infrastructure that lowers management and upstream costs. Forceful barging is less likely to occur when network control is virtualized. It is more secure because only authorised users have access. In space and genetic identification, additional safety precautions, such as the control of potentially high-end equipment, can be used. It also protects it from catastrophic occurrences by removing embedded systems from the affected area. Devices are becoming less reliant on local hardware and more reliant on data center-based devices. This combination of IoT and Cloud creates a system that is both present and secure enough to prevent data breaches, as well as current enough to rely on without the danger of human error. To create a prediction system, neural networks are sometimes integrated with IoT-based technologies. The structure of NN is similar to that of a mammalian neuron. It contributes to the development of data sensing, collection, and computation paradigms. This paper discusses in depth the usage of available best adaptive AI technologies in flood management with IoT. The following is the rest of the paper in order: The literature review is defined in Section 2. The flood management system is described in detail in section 3. The suggested model's architecture is presented in Section 4. Section 5 contains explanations and evaluations of the implementation. The sixth section contains concluding remarks.

2. Literature Review

The technological advantages will be understood by referring to the literature available and the technology used in different stages of flood management such as Mitigation, Early warning system, and so on. Yuewei Wang et al. discuss possible simulations of the concerned area to be run. The simulation would give the people an idea of the shortfalls of the city before the disaster itself. However, creating a simulation of a city with accuracy requires many live variables to be updated frequently, which is only a possibility in live reporting via an IoT management and data

E-mail address: himanshugoyal19@gmail.com, kamalghanshala@gmail.com,sxsharma88@gmail.com

collection system [1]. Sunitha Kuppuswamy et al. presented work on the disaster warning paradigm. The paradigm collects data from the past cyclones of similar and at times the same areas such as Andhra Pradesh, Odisha, Tamilnadu, and West Bengal was quoted from the research work. The data is then analyzed to create a prediction on present conditions recorded by the IoT system. If an alert was issued, the news is distributed via another IoT system and also by the traditional media [2]. Also, the analysis can be done by a paradigm too Yuewei Wang et al. discuss this very idea. They integrate an IoT-based data collection system with a DNN paradigm to correlate data collected and pass it through a decision-making algorithm trained on old flood data to predict annual floods. Systems like this require quite a lot of prior data to be even nearly correct. A disaster like an annual flood fulfills that need for data to be analyzed [1]. Mudy Solehman et al. also discussed another decision tree algorithm in their work. The accuracy came up to be a performance at 100% in partition data is 80%: 20% [3]. Rupali Patil et al. describe a cheaper system analyzing live data and rainfall prediction. It helps in getting proper and timely water management, reducing the risks of rain-fed floods [4]. The affordable price is the credit of a better WSN system. Suparna Saha Biswas et. al. explained integrating AI and IoT to obtain a data prediction paradigm [5]. Disasters, manmade or natural can result in concerning circumstances for the infrastructure present near and in the affected area. The destruction may lead to no connectivity lines and cut in direct communication. The ruble from the infrastructure itself may make the task of locating the victims a nightmarish job to do. A better method of communication is of urgency between the areas affected and the areas providing support. Md Kamruzzaman et.al. discussed and promote the prospect of replacing or adding on to the traditional telecommunication system with an IoT-based live reporting network [6]. Sunitha Kuppuswamy discussed prospects of communication to the people at times of disasters like a cyclone [2]. They also go into details of integrating the paradigm with a learning paradigm that observes the data of past such cyclones, to work upon a warning/alerting system. J. A. Hernandez-Nolasco et al. also explained how-to focuses on the dissemination of information before the disaster. It alerts the possible victims, caution being better than regret. It uses sensors to measure the live water level and alerting at imminent danger. It's a condition-based resulting algorithm [7]. Mieke et.al. discusses post-disaster management. Even after taking the most caution, there would be minimal damage. A restoration system is of necessity. The Netherlands has restored its flora and fauna as well as most of its ecological balance through its infamous restoration system. This paper discusses such plans and systems applied by IoT for data sensing, collection, and prediction. The restoration of ecological balance also helped in restoring the landscape to recent floods, cities got moved to safe planes predicted. Low lying areas have built infrastructure to help the worst of the flood that can be generated in that specific area based on IoT predictions [8]. Semih Sami Akay et al. deal with managing physical damage of the river coast after a disaster. Many landscape and ecological systems depend on the river. Being one of the largest and dependable sources of freshwater, many ecosystems thrive on these lands. The changes happening due to man-made or natural disasters can wipe out whole species from the river basin. Restoring the morphological damage in the basin or other coastline is quite a responsibility. The visualization of the changes allows the researchers to know the results of such ecological catastrophe a little bit better [9]. IoT is used in many applications. It would make sense for it to be bested and updated whenever possible. Samira P G et al. discussed one such development. Flooding and Synchronization techniques that are widely used in most systems needed to be rationalized for working efficiently. They discussed a method of speeding up the transmission of data and reducing and even negating drift clock error [10].

Chirag Panchal et al. present another example of IoT development. WCS for Electric cars which makes electric cars easier to handle and more dependable. The increase in electric cars replacing motor oil cars decreases the carbon emission drastically [11]. N.N. Srinidhi's work explained the basic network optimization for all the blocks of IoT. It also explained the challenges realized while optimizing an IoT network. It also contains the author's study about these systems and their results [12]. Disaster management has been quite a concern. Scientists and poets have also thought of ways and written about magical measures of weather control. We haven't yet fastened the ride on control but surely, we are getting closer to managing the best out of its outcomes. IoT has bettered the sensory and communication system to a new limit. It has been integrated with the cloud for security and AI for predictions at the time of need. This study shows AI with IoT's role and application in saving lives, financial realities, and whole ecosystems. Table 1 shows the pros and cons of different methods. The methods followed in flood management are not efficient and adoptive based on the type of flood situation. This was observed in the above literature. So, AI techniques can be explored to improve flood monitoring and management system. IoT techniques can be implemented cyber security algorithms for security so that it can be improved and also, we can find out the target people using AI to distribute the alert message so that it will not increase the burden on the rescue team.

3. Overview of the flood management system

Floods are one of the most severe and common forms of natural disasters. A flood is a run-off volume of water that immerses land. Floods are normal and expensive natural catastrophes, which are generally local, short-lived events that could occur quickly, occasionally with little or no alert. Floods influence equally communities and individuals and have enormous social, economic as well as environmental significance. The impact of floods may be both negative and positive, moreover, its effect differs significantly depending on the place and degree of flooding, and the susceptibility and cost of the native and created ecosystems they influence.

TABLE I. Pros and Cons of Different Methods

Methods	Pros	Cons
Communication and Tracking of Victims in Terms of Crucial Disasters [1]	IoT not succumbing to destruction makes connectivity much more possible in these grave scenarios.	It still is a post-destruction approach. Loss of life isn't avoided. Infrastructure is destroyed.
Disaster Warning System [2] [3] [4] [5]	Deal with protecting before the disaster. Concerns taken early enough can save people, at times infrastructure and crop yields.	The prediction may vary in accuracy. A false alarm can cost a ton of preparation and financial load.
Simulated Data Prediction via IoT [6]	It usually holds more accuracy than warning systems without simulations.	Still has varying accuracy and can cause false alarms.
Post-Disaster Management via IoT [7] [8] [9]	Restoration of livelihoods saves the unsaved from disaster. Restoration of ecology and river basins saving many ecosystems.	Again a post-destruction approach. Loss of life isn't avoided. Infrastructure is destroyed.
Betterment of IoT Systems [10] [11] [12]	Developments have been a boon, not only for disaster management but most of the industrie	Project in development. The update can be resource-consuming to the system if the changes vary too much. This defeats the very purpose of network optimization

Floods may be caused by the following reasons: Heavy monsoon: Excess rain also causes flood problems, especially in those areas where the drainage system is poor. This is generally occurring in plain areas. Cyclones: Cyclones lead to flood problems in coastal areas. Glacier burst: melting of glacier and glacier burst suddenly increases the water volume in a particular area. All the above-mentioned issues are also dependent on climate change and global warming [13].

A. *Impact of flood*

The impacts of floods, as mentioned above could be both negative and positive, differ significantly which depending upon their locality, period, extent, flow, and depth. Uncontrolled floods influence both people and societies and cause social, economic, and ecological outcomes [13].

1) *Consequences of floods on individuals and communities*

As most of us are well aware, that the instant and direct impacts of flooding caused human life loss, destruction to belongings, crops destruction, livestock loss, and worsening of health due to waterborne ailments. As telecommunication systems and infrastructure such as roads, flyovers, and power plants are harmed and disordered, some of the economic endeavors might come to a halt, people have to abandon their homes resulting in disruption of normal life. Likewise, damage to industries can lead to livelihood loss. Infrastructure damage also leads to long-term after-effects, such as clean water supply disruptions, wastewater treatment, transport, electricity, communication, health care, and education system. Livelihood's loss, purchasing power reduction, and damage to land value in the flood-prone area can leave people economically week and susceptible. Floods can also devastate and shock sufferers along with their families for long terms. The loss of beloved ones has profound impacts. Dislocation from home, property loss, and business disruption may lead to unending trauma. For some individuals, the emotional effects could be long-lasting [14].

2) *Social and economic impacts of floods*

Flooding in the main agrarian production areas can result in extensive crop and fence damage and loss of domestic animals. Excessive downpour leads to loss of crops, soil water logging, and harvesting delays are further exaggerated by poor transportation glitches owing to the flooded and damaged infrastructure. Damage to roads, railways, and shipping ports could have a significant influence on the local and national economies. Reduced agrarian production may also impact beyond the production areas, since food shortage supply may result in increased food prices. Oppositely, floods do have long-standing advantages to agricultural produce by reviving the groundwater water resources, also recharging water storage ponds, especially in arid and semi-arid, inland areas, and wetlands. Silt deposition by floodwater also rejuvenates soil fertility. Other environmental plus points of floods could be enhanced production of fish, thereby helping the fishermen economy. Recharge of inland ponds and lakes resources results in recreational environments and water sport activities.

B. *Challenges in flood management*

Traditional flood management systems which are currently being used have many shortcomings. So many new Flood Management Systems are being developed using innovative technologies. But these new systems still face a lot of challenges.

1) Population growth

In early times people used to live on higher grounds away from flood plains to avoid the floods but population growth and food shortages have forced societies to utilize floodplains, and therefore to control floods with structures. This is one of the major challenges faced in flood management. This puts a lot of pressure on the natural system and increases the damage potential on flood plains.

2) Climate change

Other than population growth, one of the major challenges faced in today's world is climate change. It has escalated the hydro logical cycle, increasing the magnitude and frequency of extreme flood events. It has also changed the seasonality of floods. All of this has made it hard to make predictions on the collected past data. Due to climate changes, sea levels have risen increasing the risk of flooding in the coastal regions.

3) Deforestation

Deforestation is one of the main reasons that cause floods to escalate to a dangerous level. It results in higher sediments across the banks and bed of the river which in turn reduces the discharge capacity of the river. With the increase of population, there has been a rise in need of farming land, and urbanization which are the key reasons for deforestation.

4) Land shortage

Reallocating people is one the easiest way to prevent losses during floods, but for many countries including India, it is not feasible. Considering the huge population of India there is a shortage of alternative land that can provide people with all their needs while protecting them from floods. Most of the affected population includes poor and vulnerable people. They have no choice but to settle in flood-prone areas. This makes the task of flood management more challenging. Although absolute flood protection is neither technically economically nor ecologically feasible, we have to find a new approach to flood management that reconciles the development of the country and its people with its flood security. Since it is not always possible to avoid flood-prone areas, we must make a sustainable system that can make life safe in the flood plains minimizing the loss of lives and maximizing the benefits from flood plains by preserving the environment.

4. PROPOSED HYBRID MODEL

The IoT-based flood management consists of the flood management devices, network, cloud IoT flood management core, data analysis, proposd hybrid model, and managing the IoT cloud applications and services which is shown in Figure 1.

A. IoT Flood management devices

The IoT flood management devices consist of sensors, actuators, and controllers. These devices are explained as follows [15].

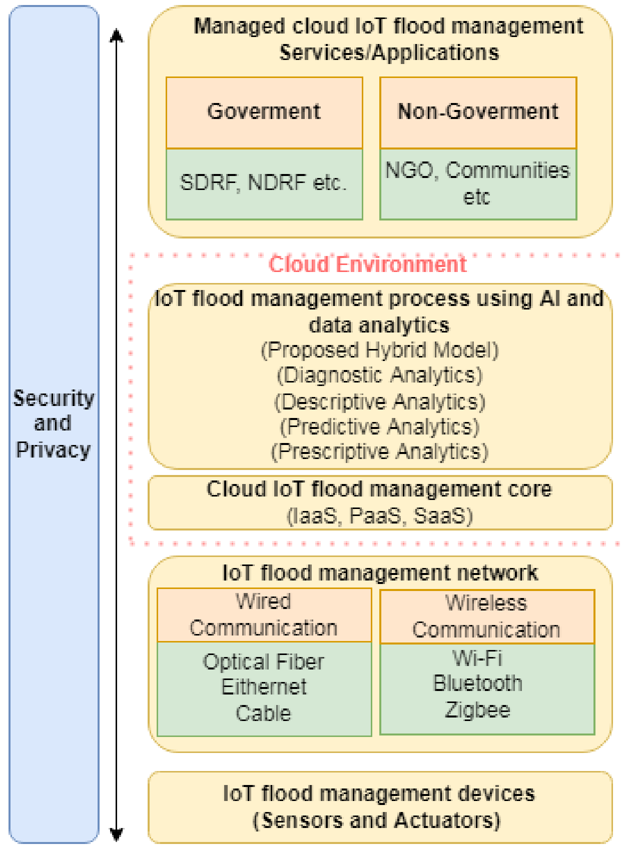

Figure 1. Architecture of IoT based flood management system

1) Arduino Uno

It is a microcontroller used to control different sensors, motors, and so on. These controllers are widely used because of their low cost and good performance as compared to any other type of microcontroller. The Arduino Uno controller used in this study and is based on ATmega328. The programming of this controller is simpler and only we required a computer with the Arduino computer program to make this controller work as we required. The power consumption is also less.

2) Water Flow Rate Sensor

A water flow tare sensor is commonly used where the flow rate of fluids is required to measure especially water. This device consists of a Hall Effect sensor that senses the flow rate based on the mechanism provided in the device. Based on the water pressure or force it converts that signal or senses and sends it back to the microcontroller.

3) Ultrasonic Sensor

It uses sonar to determine the distance of any object, obstacle detecting in robots, and so on. The range of detection varies from 1 inch to 13 feet. The device consists of a transmitter and a receiver end. In our study, we use this sensor to find out the wastewater level to find out whether the water level is normal or critical for further.

4) Wi-Fi module

This is used to control the wastewater disposal system on a real-time basis. This synchronized system helps to data collection from stations (station for all devices/ sensors and so on). The data flow from the cloud to the stakeholders. This is integrated with a microcontroller.

5) GPS Module

The GPS module is used along with all the devices to send the location where the wastewater level reaches the critical level in the water disposal system. The stakeholder will access the data and location send by the device for further processing of the data to avoid the flood or any damage.

6) DC Motor

DC motors are widely used in many applications. These applications vary from robot chaises motor to industrial applications. In our study, we are using a 5V DC motor to the opening of valves for the wastewater disposal system when the water level reaches a critical level.

B. IoT flood management network

The type of network used to connect these IoT devices is most important. The network can be wired or wireless connection. But during flood situations, the wired connections will not be feasible and in a wireless connection, Bluetooth and Zigbee have a low range so the only option is to connect these devices through a Wi-Fi network.

C. Cloud IoT flood management Core

It helps to manage securely and easily connect and ingest data from all the devices connected. The IoT cloud services are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS). These services are required to securely connect to the devices. It uses Infrastructure, Platforms, and software as a service in Cloud IoT-based flood management.

D. IoT flood management processing using data analytics

Predictive analysis is used here. We prefer it because the flood intensity from the available Cloud IoT server data is sufficient to mitigate losses. For IoT flood control, therefore, predictive analytics are ideally suited. The SVM and Logistic regression have been using in the prediction of flood pattern recently. The proposed hybrid model is the integration of SVM and Logistic regression. This model will be more accurate and gives the best possible forecasting in the flood management after analyzing the vague data obtained from the smart IoT flood devices. Figure 2 shows the proposed hybrid model the input data of real time flood variables along with the past flood data is supplied. Flood related features are extracted, and these features will be like past total rainfall and present total rainfall data. This data is divided in to 80% training the hybrid model and 20% testing it. The forecasting output which is obtained after testing are finalized to alert all the stakeholders.

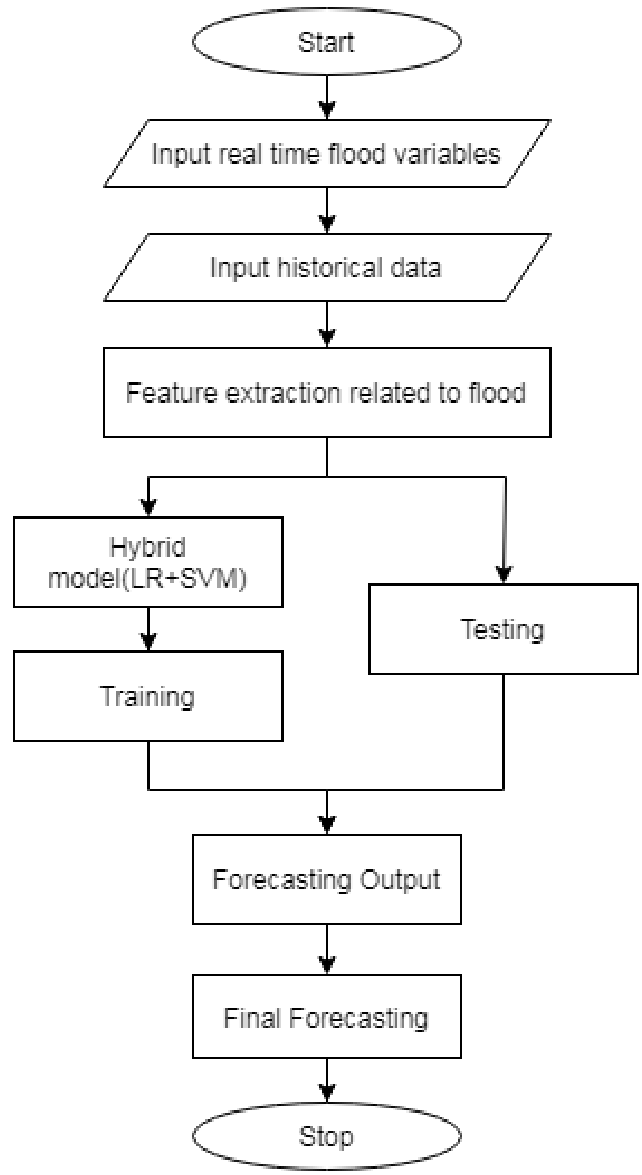

Figure 2. Flowchart of hybrid model

E. Managed cloud IoT flood management applications and services

The stakeholders are told to take the appropriate actions after analyzing IoT cloud data. Government agencies like SDRF, NDRF, etc. will take the requisite steps to reduce damages resulting from expected floods with civil society groups, municipalities, and NGOs, etc.

5. IMPLEMENTATION AND RESULT ANALYSIS

The Dataset used for this ML model is taken from the Kaggle website. It is based on the average rainfall that Kerala received over the last 118 years in different months.

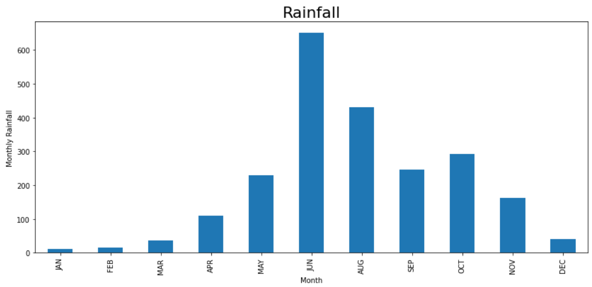

Figure 3. Graph demonstrating amount of rainfall received by Kerala in different months.

A. Data Explanation

On 16 August 2018, Kerala was very much affected by severe floods because of the extreme rainfall during the monsoon season. Nearly 400 casualties were recorded, and approx. thousands of people lost their homes. As a result, we've decided to focus our investigation on Kerala. Kerala is situated between the Arabian Sea and the Western Ghats to the west and east, respectively. It has a tropical monsoon climate and averages 3107 mm of annual rainfall or 7,030 crore m of water. When compared to the national average of 1,197 mm, this is a significant difference. After doing a bit of research, it was found that the major reason for the Kerala floods is the unusually high rainfall during the monsoon. Thus, annual rainfall was chosen as the parameter for doing the flood prediction. This is a time-series data of annual rainfall and flood occurrence from the year 1901 to 2018. For each year, monthly rainfall is shown in each Column. There is also a column for annual rainfall. The reason we included month in our model is that rainfall in India tends to be substantially greater during the monsoon season.[16].

B. Analyzing Data

There is a total of 118 rows and 16 columns in this dataset. After data analysis, it was found that the data is clean and there were no missing values. In the last column, which had the information about the flood occurrence, yes and no were mapped to binary i.e. (1,0) for simplifying further calculations. After visualization of Dataset, it was found that the annual rainfall during June, July, August, and September is much higher as compared to other months which results in increased chances of flood during these months. As shown in Figure 3.

Kerala state has an average precipitation of 3000mm. According to IMD data, Kerala received about 2346.6 mm of rainfall from 1 June 2018 to 1 August 2018 which was 42% more than normal. Other than the year 2018, Kerala faced heavy losses due to floods in 1924 and 1961. During both these years, the monsoon got extremely violent during the mid of July.

The avg. rainfall was almost 56 % more than normal. Hence, we conclude; there are higher chances of floods

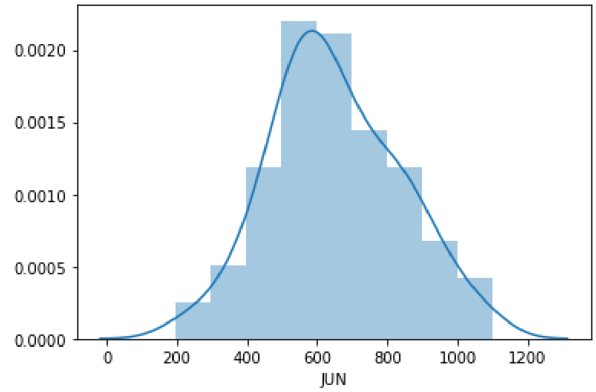

Figure 4. Rainfall received in June month.

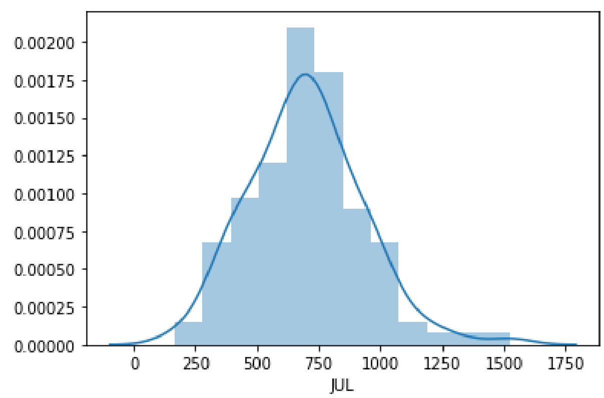

Figure 5. Rainfall received in July month.

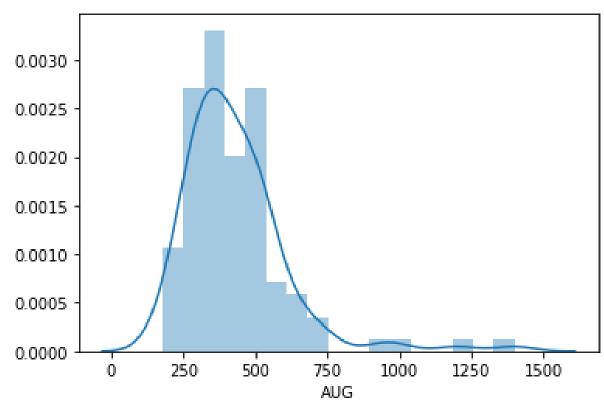

Figure 6. Rainfall received in August month.

during the monsoon most probably from the end of June till the start of August. We have plotted bar plots to visualize the data as shown in Figure 4. 5. 6. The above graphs are the pictorial representation of the amount of rainfall that occurred in Kerala during June, July, and August.This data is further trained and tested to build a suitable model which can precisely predict the floods so that we can further vacate the areas minimizing the loss and saving millions of lives.

C. Different classifiers used

Now we will be explaining the different algorithms we used for prediction:

1) K-Nearest Neighbors (KNN)

KNN or k-nearest neighbors is an algorithm that can be used for flood management classification and regression problems. This algorithm is fair amongst all the parameters. It is mainly used because it's very easy for prediction and has a low calculation time as compared to other algorithms. The classification mainly depends on the majority of votes received. The "K" in KNN states for the nearest neighbor whose vote we want to take for the classification. When the value of K is 1, the object is allocated to the nearest.

KNN can be implemented by using the following steps:
1Importing the dataset.
2Initializing the value of K.
3The value of K is chosen based on the following factors.

- The value of k is the most crucial function in the KNN algorithm. If k is less than 1, our model is overfitted, but if it is greater than 1, the reasoning behind the approach is lost. Thus it is defined by k = sqrt(n).
- We can also use cross-validation for optimizing the results.
- We can also run each value of k for different instances and check for the optimal result.

Distance between test data and each row of training data can be calculated using Eq 1. We have used Euclidean distance as it is the most popular one but other metrics like Manhattan, Chebyshev and Cosine can be used and the probability using Eq 2. Figure 7 shows the Confusion Matrix for the k-Nearest Neighbors algorithm.

$$dist((x, y), (a, b)) = \sqrt{(x-a)^2 + (y-b)^2} \quad (1)$$

$$P(Y = j|X = x) = 1/k \left[\sum_{i=A} I(Y^{(i)} = J \right) \quad (2)$$

2) Decision Tree Classifier (DT)

A Decision Tree is a supervised approach used to tackle classification issues and regression; however, it is mostly utilized in classification. It is a tree-structured classifier with core nodes that resemble data sets, branches that resemble decision-making procedures, and book nodes such as the result. In the decision tree classification, the Decision Node and the Leaf Node are two sorts of nodes. Leaf nodes reflect and do not have any branches the impact of decisions, whereas decision nodes are used for decision-making, and they have several industries. The features of the dataset are controlled by the choices. It's a decision tree because it begins with the root node, which travels to branches and builds a structure like a tree. It's a decision tree. Figure 8 shows the Confusion Matrix for the Decision Tree Classifier. Its working is shown as follows:
1. To begin, it chooses the best attribute for splitting the data.
2. The attribute is then turned into a decision node, and the dataset is divided into smaller kinds subsets.
3. It then begins to construct the tree by recursively repeating the procedure for each child node until one of the following requirements is met:

- Each tuple should be associated with the same attribute value.
- There should be no more characteristics left.
- There should be no further occurrences.

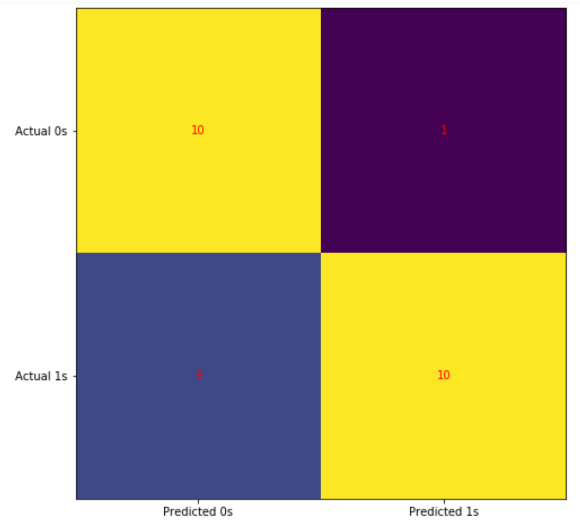

Figure 7. Confusion Matrix for k-Nearest Neighbors algorithm.

3) Random Forest Classifier (RF)

Random Forest is a monitored method for machine learning. It is used for both regression and classification. You can use it pretty easily. In contrast to a forest of trees, a forest with more trees is believed to be more powerful. Random forests produce numerous randomly selected data decision-takers, sample it, forecast each tree, and offer the best answer using a voting mechanism. It offers also a great feature indication. Figure 9 shows the Confusion Matrix for Random Forest Classifier algorithm. It works in four simple steps:

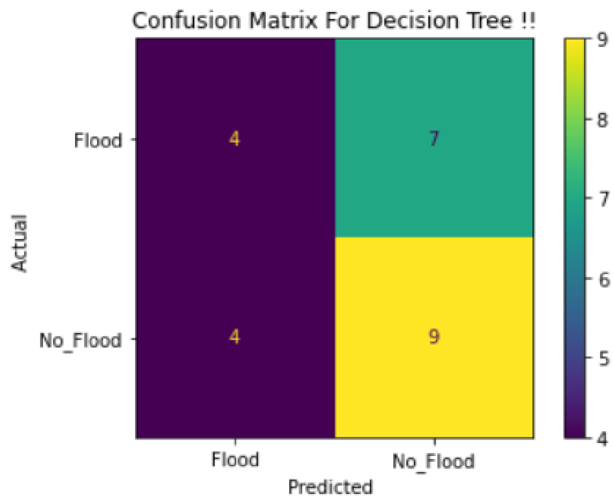

Figure 8. Confusion Matrix for Decision Tree Classifier.

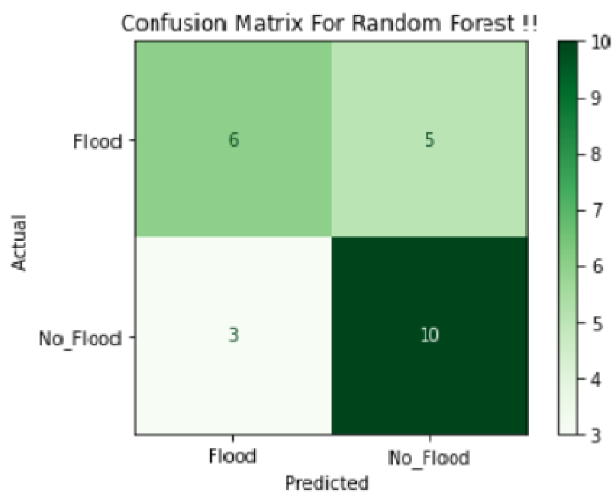

Figure 9. Confusion Matrix for Random Forest Classifier algorithm.

1. It begins by randomly selecting samples from the provided dataset.
2. It then creates a decision tree for each sample, with each decision tree producing a forecast result.
3. It then casts a vote for each of the projected outcomes.
4. Finally, the forecast result with the most votes is chosen as the final prediction.

4) Logistic Regression (LR)

Machine Learning Algorithm with Supervised Learning The likelihood of particular events is calculated using logistic regression. It's used in Binary Outcome Prediction (when there are only two possible outcomes for a problem i.e. 1 or 0). Logistic regression may be divided into two types: binary and multilinear. It's a probabilistic predictive analytic approach for resolving categorization problems. The cost function should be confined to a value between 0 and 1 according to the logistic regression hypothesis. It's a tool for turning probabilities into predictions. Any real

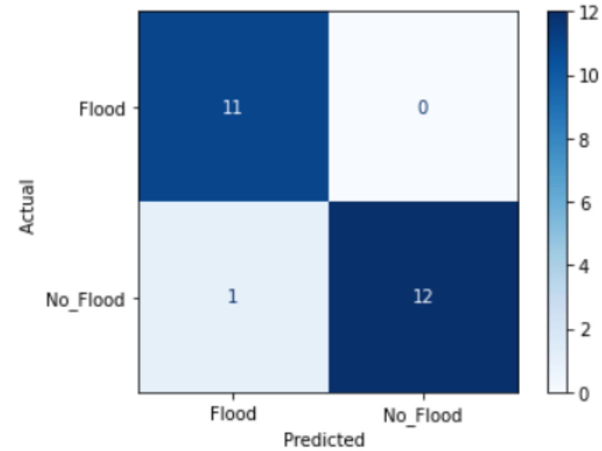

Figure 10. Confusion Matrix for Logistic Regression algorithm.

number is converted to a number between 0 and 1 using this method. Figure 10 shows Confusion Matrix for Logistic Regression algorithm We only have two possible outcomes in our model: flood or no flood, which may be translated to 1 or 0. It is used to anticipate specific events and is mostly utilized in predictive analysis, based on the fact of likelihood. The value is predicted using a logistic sigmoid function. In a logistic model, the dependent variables are categorical and are known as target variables, whilst the independent variables are known as predictors. We use the sigmoid function to predict the categorical values and the threshold decides the outcome. Logistic Regression using Eq 3:

$$p = 1/\left(e^{-(\beta 0+\beta 1X1_{\beta}2X2....\beta nXn)}\right) \quad (3)$$

5) XG Boost Classifier (XG)

XG Boost Classifier is a Python Programming Language algorithm that is a decision tree-based ML algorithm that uses the framework of Gradient Boosting. The prediction problems which have unstructured data, artificial neural networks work best overall with other algorithms. This algorithm has a vast range of applications: It is used to solve regression, ranking, classification, and user-defined type prediction problems. It has high portability, i.e. It runs on all types of OS i.e. Windows, Linux, etc. It also has the support of nearly all programming languages, i.e., C++, R, Python, JAVA, Julia, Scala, etc. It also has cloud integration in it, like it supports these clusters that are AWS, Azure, YARN and works best with Spark, Flink, and other ecosystems.

Figure 11 shows the Confusion Matrix for the XG Boost algorithm Here are the steps to use the XGBoost classifier: 1. First, we do problem definition with our dataset. 2. Then we load and prepare the data. 3. Then we train the XGBoost model. 4. It makes predictions and evaluates the model. 5. Lastly, it ties it all together and runs the example.

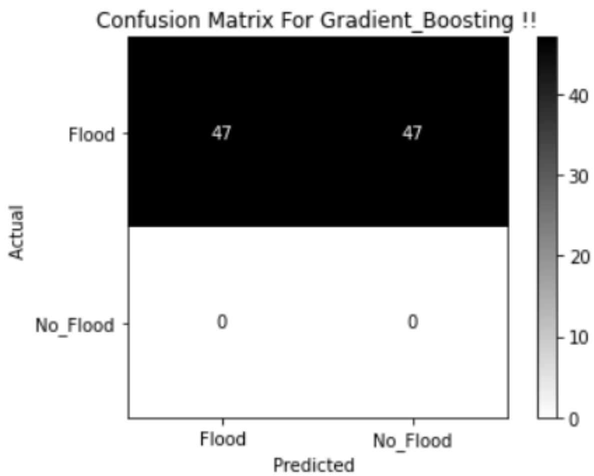

Figure 11. Confusion Matrix for XG Boost algorithm.

D. Results and discussion

We have implemented different machine learning algorithms on this project like KNN , LR, DT, RF, and XGBoost Classifier. But to check which algorithm is better suited for this model we have to compare them. Based on the following comparative analysis logistic regression was chosen as the most suitable model for this project.

A confusion matrix is used to assess the performance of a classification model. The number of target classes is n in this N*N matrix. It provides a broad overview of the model, including how well it performs and what faults it makes. Depending on the settings, a confusion matrix can have any size. We utilised a binary classifier in this example, which is a 2*2 matrix as seen in Figure 12.

Here,

- True Positive (TP): These are the accurately predicted actual values, indicating that both the actual and predicted classes were correct.
- False Positive (FP): While the model predicted a positive value when the actual class value was false.
- When the actual value is positive but the anticipated value is false, it is called False Negative (FN).
- True Negative (TN): When the actual value and anticipated value were both false.

TP and TN denote the number of times a flood was properly predicted as occurring or not occurring, whereas FP and FN denote the number of times a flood was incorrectly forecasted. We can compute the model's accuracy, precision, recall, and F1 score once we understand these four parameters.

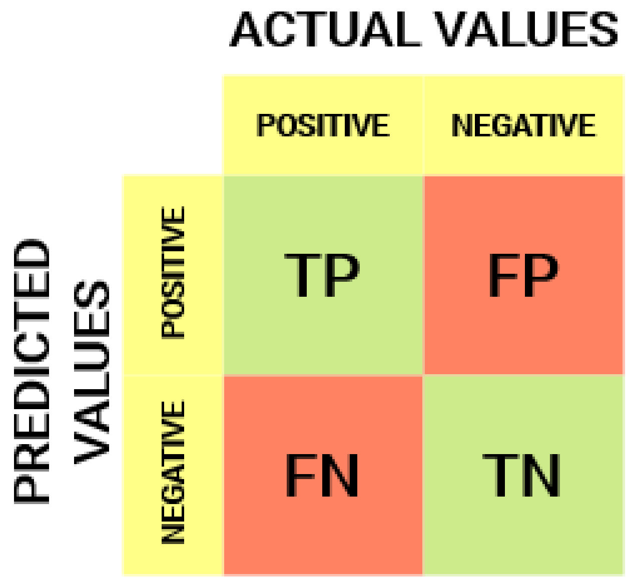

Figure 12. Confusion Matrix.

1) Accuracy

The ratio of accurately anticipated observations to actual observations is known as accuracy. Informally, it's the ratio of correct predictions made by our model. Accuracy = No. of correct predictions/Total Predictions In terms of Binary Classifier

$$Accuracy = TP + TN/TP + FP + FN + TN \quad (4)$$

2) Precision

It's the proportion of accurately anticipated positive observations to total positive observations. It informs us how many accurately anticipated positive numbers turned out to be right.

$$Precision = TP/TP + FP \quad (5)$$

3) Recall

It's the proportion of accurately anticipated positive observations to all of the class's observations. It indicates how many actual true situations the model was able to anticipate.

$$Recall = TP/TP + FN \quad (6)$$

4) F1 Score

It is defined as the weighted average of precision and recall. When we improve the precision of our model then recall goes down and vice versa. It is used to capture both the trends in a single value.

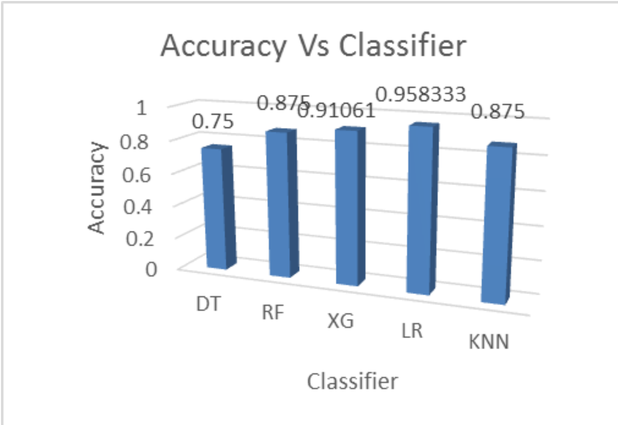

Figure 13. Comparison of accuracies achieved using various classifiers.

TABLE II. Accuracy, Recall, and ROC score of the algorithms

Algorithm	Accuracy	Recall	ROC Score
Decision Tree	75.00	90.09	76.02
KNN	87.50	90.01	0.80
Random Forest	87.50	92.30	87.06
XG Boost	91.06	87.53	89.50
Logistic Regression	95.833	90.90	85.54

$$F1Score = 2*(Recall*Precision)/(Recall+Precision) \quad (7)$$

5) ROC Score

The intensity of categorization issues at various stages. Its value goes from 0.5 to 1, with 0.5 being a bad classifier and 1 denoting an excellent one.

From Table II, it is obvious that most of the forecast results for all ML algorithms are accurate and positive, i.e. faulty results are less accurate when compared to positive results. For logistic regression, i.e. 95.83%, the results' accuracy is maximum. This states that for this algorithm our model performs best as shown by the graph in Figure 13.

Table III examines the different the existing and proposed hybrid models. In comparison to other existing methods, the suggested model is 99.64% accurate in anticipating flood effects. We also discovered that the suggested hybrid model outperforms existing models in terms of accuracy.

6. CONCLUSION AND FUTURE RECOMMENDATION

Flooding was predicted to account for more than 84% of natural catastrophe deaths. Depending on their location, many countries experience different flood events. However, recent years have seen an increase in the frequency and severity of flood disasters due to climatic and global warming trends. Kerala has lately been hit by one of the worst floods in decades. Thousands of homes were affected, and more than 500 people were murdered. India has now become one of the worst-affected countries in the world when it comes to flooding. We constructed an AI-based flood prediction model that forecasts floods for this region based on historical flood data to limit the harm caused by floods. To create this model, we used machine learning algorithms. The dataset was taken from the Kaggle website. It's based on Kerala's average precipitation in different months during the last 118 years. Following that, the data was separated into 80:20 training and test data sets. We then used six different algorithms to train this model. The proposed hybrid method performs with a higher accuracy of 99.64%, indicating that we could correctly predict that the flood will occur most of the time. A flood is a natural calamity, and no natural event can be prevented. But, with the help of technology, we can predict it far ahead of time, giving us enough time to evacuate the region and save millions of lives.

TABLE III. Comparative analysis of proposed model Vs existing algorithms

Algorithm	Accuracy
DNN [17]	91.18
Desision Tree [18]	87.50
SVM [19]	87.50
ANN [20]	91.06
Proposed Hybrid Model	99.64

Advances in neural networks for prediction and other machine learning approaches can lead to more research and improved accuracy by allowing researchers to use more recent datasets. Precipitation and temperature are two more variables that might be used to predict floods.

References

[1] Y. Wang, X. Chen, L. Wang, and G. Min, "Effective iot-facilitated storm surge flood modeling based on drl," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6338–6347, 2020.

[2] S. Kuppuswamy, "A perception study on community response to media technologies for cyclone warning and disaster communication in south india," *Science, technology and society*, vol. 19, no. 3, pp. 399–414, 2014.

[3] M. Solehman, F. Azmi, and C. Setianingsih, "Web-based flood warning system using decision tree method," in *2019 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*. IEEE, 2019, pp. 1–6.

[4] O. Amale and R. Patil, "Iot based rainfall monitoring system using wsn enabled architecture," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2019, pp. 789–791.

[5] P. Mitra, R. Ray, R. Chatterjee, R. Basu, P. Saha, S. Raha, R. Barman, S. Patra, S. S. Biswas, and S. Saha, "Flood forecasting using internet of things and artificial neural networks," in *2016*

IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Ieee, 2016, pp. 1–5.

[6] M. Kamruzzaman, N. I. Sarkar, J. Gutierrez, and S. K. Ray, "A study of iot-based post-disaster management," in *2017 international conference on information networking (ICOIN)*. IEEE, 2017, pp. 406–410.

[7] J. A. Hernández-Nolasco, M. A. W. Ovando, F. D. Acosta, and P. Pancardo, "Water level meter for alerting population about floods," in *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*. IEEE, 2016, pp. 879–884.

[8] M. Van Hemert, "Taming indeterminacy: the co-production of biodiversity restoration, flood protection and biophysical modelling of rivers and coastal environments," *Science, Technology and Society*, vol. 18, no. 1, pp. 75–92, 2013.

[9] S. S. Akay, O. Özcan, and F. B. Şanlı, "Quantification and visualization of flood-induced morphological changes in meander structures by uav-based monitoring," *Engineering Science and Technology, an International Journal*, 2021.

[10] S. P. Gilavan and G. Ekbatanifard, "Eglossy: An energy-efficient glossy-based synchronization and flooding protocol in iot," in *2018 8th International Conference on Computer and Knowledge Engineering (ICCKE)*. IEEE, 2018, pp. 141–146.

[11] C. Panchal, S. Stegen, and J. Lu, "Review of static and dynamic wireless electric vehicle charging system," *Engineering science and technology, an international journal*, vol. 21, no. 5, pp. 922–937, 2018.

[12] N. Srinidhi, S. D. Kumar, and K. Venugopal, "Network optimizations in the internet of things: A review," *Engineering Science and Technology, an International Journal*, vol. 22, no. 1, pp. 1–21, 2019.

[13] N. L. Poff, J. D. Allan, M. A. Palmer, D. D. Hart, B. D. Richter, A. H. Arthington, K. H. Rogers, J. L. Meyer, and J. A. Stanford, "River flows and water wars: emerging science for environmental decision making," *Frontiers in Ecology and the Environment*, vol. 1, no. 6, pp. 298–306, 2003.

[14] M. Buragohain and M. Bhuyan, "Impact of flood and siltation on socio-economy." *Antrocom: Online Journal of Anthropology*, vol. 10, no. 2, 2014.

[15] O. O. Kazeem, O. O. Akintade, and L. O. Kehinde, "Comparative study of communication interfaces for sensors and actuators in the cloud of internet of things," *Int. J. Internet Things*, vol. 6, no. 1, pp. 9–13, 2017.

[16] GGopinathan, "Kerela flood," www.kaggle.com/ggopinathan/rainfall-in-kerala-from-19012017, 2021, [Online; accessed 10-June-2021].

[17] S. Sankaranarayanan, M. Prabhakar, S. Satish, P. Jain, A. Ramprasad, and A. Krishnan, "Flood prediction based on weather parameters using deep learning," *Journal of Water and Climate Change*, vol. 11, no. 4, pp. 1766–1783, 2020.

[18] C. Luu, D.-D. Nguyen, T. V. Phong, I. Prakash, and B. T. Pham, "Using decision tree j48 based machine learning algorithm for flood susceptibility mapping: A case study in quang binh province, vietnam," in *CIGOS 2021, Emerging Technologies and Applications for Green Infrastructure*. Springer, 2022, pp. 1927–1935.

[19] J. Al Qundus, K. Dabbour, S. Gupta, R. Meissonier, and A. Paschke, "Wireless sensor network for ai-based flood disaster detection," *Annals of Operations Research*, pp. 1–23, 2020.

[20] R. Tabbussum and A. Q. Dar, "Performance evaluation of artificial intelligence paradigms—artificial neural networks, fuzzy logic, and adaptive neuro-fuzzy inference system for flood prediction," *Environmental Science and Pollution Research*, vol. 28, no. 20, pp. 25 265–25 282, 2021.

Himanshu Rai Goyal Dr. Himanshu Rai Goyal is an Associate Dean for International Affairs and an Associate Professor in Graphic Era's Department of Computer Science and Engineering. Dehradun, UK, India, Deemed to be University. Wireless sensor networks and the Internet of Things are among his research interests..

Kamal Kumar Ghanshala Prof. (Dr.) K. K. Ghanshala is the current chairman and professor at Graphic Era's Department of Computer Science and Engineering. Dehardun, Uttarakhand, India, Deemed to be University Wireless communication networks, the Internet of Things, vehicular ad hoc networking, and network security are among his research interests.

Sachin Sharma Dr. Sachin Sharma is an Associate Dean for International Affairs and an Associate Professor in Graphic Era's Department of Computer Science and Engineering. Dehradun, UK, India, Deemed to be University Wireless communication networks, the Internet of Things, vehicular ad hoc networking, and network security are among his research interests.