

Dam Hazard Prediction Using an AIoT Based System

Nabila Aissani¹, Abderraouf Messai¹, Salheddine Sadouni¹

¹LSIACIO Laboratory, Frères Mentouri Constantine 1 University, Constantine, Algeria.

Article Info

Article history:

Received Oct 13, 2025

Revised Mar 24, 2026

Accepted Apr 16, 2026

Keywords:

Security

Dams

Artificial Intelligence of Things

Heavy rainfalls.

ABSTRACT

Dams are a very important infrastructure because they are one of the main water and electricity supplies for most countries; However, it is susceptible to demolition due to several factors such as aging and floods. This not only causes damage to water supplies and economical loss, but also human lives. And with the current climate change and global warming, heavy rainfalls and storms leading to floods are rapid and abrupt; existing studies tend to focus on one aspect rather than a total security system. Which drives us to look for the answer to the most important question; How can we deploy a system that is capable predicting coming floods and estimating the impact giving us enough time for a reaction? Hence, in this paper, we propose an Artificial intelligence of things based system to predict potential floods in dams by overtopping and potential breaching, their time, and estimated threatened distance. This system includes mainly 1) several types of sensors like temperature, humidity, and water level 2) AI models like one class support vector machine, isolation forest, hybrid models, and regression models. The results were promising for improving dam security.

Copyright © 2026 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Nabila Aissani

LSIACIO Laboratory, Electronic Department, Faculty of science and technology, University of Constantine 1 Constantine, Algeria

Email: nabila.aissani@doc.umc.edu.dz

1. INTRODUCTION

Floods are a natural disaster that has been occurring for ages; and there is not one particular source for it. Be it seas tsunamis, rivers or heavy rainfalls:

they all have horrible consequences: Loss of lives, demolition of structures and uncountable physical damage [1]; like the one that happened in western Europe in July 2021 [2].

But dam flooding is more likely to happen due to climate change and heavy rainfalls and storms like the one that happened in Libya 2023, these rainfalls are not always predictable which means we cannot always see them coming and have to be ready for it when it happens abruptly. This disaster can cause human losses and huge physical damage to the dam and other structures[3] [4], therefore, a control and monitoring system is necessary to prevent the dam from breaking and reduce the terrible consequences to the minimum range. Knowing that floods can come in different scenarios and can extend to several area depending on the flow, and seeing we cannot always predict the natural causes of the floods; the important questions raise as to how we can predict and help prevent the breaking of the dam and reduce the damage caused by the flooding that comes from unexpected heavy rainfalls. From another part, technology is in an nonstop evolution where we can see different tools emerging like the Internet of things' solutions, which is a simple but effective idea of connected devices that communicate with each other[5] [6] to help monitor and manage a zone . Also, the most common and evolving technology is the artificial intelligence which basically takes over human's time and energy consuming tasks and allowing machine to perform them where AI and IoT are commonly seen in the field like mentioned in [7].

1.1. Related works

The seriousness of floods and their consequences pushed researchers and specialists to think hard for solutions to this dilemma like [8] who used dual module design to address inflexibility and computational challenges in previous solutions. We could see that the use of internet of things recently is a very common tool like the work of [3], and when integrated with artificial intelligence; the solutions are more abundant and promising as seen in the work of [9] who provided a IoT framework cloud based to forecast floods and integrates AI as well, and [1] where A prototype utilizing ML and IoT was developed for hydro-meteorological data collection in order to detect floods and then classify the categories of that flood, finally; alert authorities and locals before the disaster takes place. Another work focused on evaluating the meteorological conditions, when the rainfall was higher than average, this to apply an accurate analysis of floods exacerbation during that period of above average rainfall heaviness. The causes and effects extracted by this work can allow a rational discussion of doable solutions which can be affiliated to government agencies allowing a minimization of the general damages of this hazard [10]. The [11] insisted that not having real data about the weather which can help us know patterns and connect the dots; leads to a lack of effectiveness in the prediction of this disaster; as a solution, the utilization of IoT to collect the necessary information of the surroundings and transfer them is proposed. Simple but effective methods like monitoring and warning via text messages were also introduced to minimize the damage of this hazard like proven in [5]. There also has been a Development of a new algorithm to effectively measure and monitor flood levels and rate changes to trigger early flood warnings, it was Demonstrated to be effective and functional through real-life implementation. [12]. As already noticed; Many methods took place as [4] emphasized because Flood Modeling is extremely important for anticipating and mitigating flood impacts effectively, and this as applied in their own work includes multiple tools such as hydrology and artificial intelligence. and [13] developed a system for a precise flood forecasting which can estimate the timing of flood in advance. And the basic image of AIoT in this side of security was seen in a work study that used support decision system to monitor floods by monitoring, predicting, and controlling them. This is done by analyzing historical data and clustering precipitation and flood records then estimating flood disasters based on rainfall with the help of artificial intelligence. They classified the data of each group is according to the amount of rainfall, flood risk and safe status.[14]

When it comes to dams part, it is worth to mention that a dam in its turn can have different security issues; it can break due to age [3], structure failure or floods as well . I.e, when there is a heavy water flow against the dam wall it can cause breaching or even the whole dam demolition. There is a back and forth relation between dams demolition and floods; a heavy flood can cause breaking of a dam and in return a dam break can cause a flooding of reservoir water. Like mentioned in [15] work which focused on studying the consequences in case of dam failure, where they highlighted the fact that dams serve various purposes but their failure can lead to catastrophic events like floods. In case of floods in a dam, it's crucial to predict inundation depth, flood wave arrival, coverage area, and water velocity for emergency action planning. And seeing all these reasons, it is safe to say that it's really crucial to address this issue and put light on the security of the dams, like the paper [16] that Developed DB-CDS model for rapid analysis of overtopping breaches in check dam systems. Model includes dam system spatial topology method and advanced dam breach flood calculation method. Their simulation results show low error rate, another work built an AI applications in hydropower and dam engineering, including predictive modeling, real-time monitoring, optimization, and case studies of machine learning techniques [17]. An alternative solution was given by [18] which considers various breach widths and reservoir levels to assess uncertainties to generate probabilistic maps that eventually provide quantitative information on the uncertainty of dam-break flood hazard. A different approach focused on self supervision to detect deformations in dam [19]. [20] evaluated three machine learning approaches and proposed an anomaly detection method that helps reduce misclassifications, and [21] used deep encoders for automatic anomaly detection in water levels. To make the existing systems even better [22] work suggested the inclusion of autonomous operation and statistical data-filtering to help the prevention of any malfunction along with other key functions. When [23] suggested a machine learning framework to forecast flooding in embankment dams; this method helped the adaptability and reduction of systematic overestimation. [24] proposed a new approach for better water resource management especially in the recent times when climate change is overwhelmed by uncertainty; their method confirms the efficiency and reliability of machine learning models

in forecasting water levels in dam reservoirs. Similarly [25] focused on water level forecasting but using other method where they showed that Hybrid CNN-LSTM model had the best performance. [26] gave an innovative way to solve computational efficiency when it comes to the traditional methods for flood mapping. By using a transformer-based surrogate model to forecast flood propagation in very short time which helps early warnings. Another work combined two methods for a new solution in hybrid prediction-to-map to cover the gaps in traditional flood forecast who rely on either numerical models only or machine learning models which achieved high accuracy [27]. But works did not only stop at the prediction and forecasting, some of them along with prediction, took interest in the efficiency of the solution from aspects like energy consumption like [28]

The goal of this paper is to define a robust system for dam control and monitoring. It includes anomaly detection and flood prediction to reduce potential damage. This paper also suggests measures of security that align with the situation; First: Opening of spillways to reduce pressure against the dam wall and lower the water level. Second: In case of continuous danger, estimation of time and distance for alert and evacuation. Hence. We focus on the utilization of the artificial intelligence of things (AIoT) in two main phases:

- Anomaly detection with hybrid model that combines one class support vector machine (SVM) and isolation forest.
- Prediction of flood using the combination of random forest and linear regression.

The IoT part is responsible for the collection and transfer of the data within the network and across the different blocks of the architecture, while the AI part is responsible for the processing, prediction, and decision making in convenience according to the given data.

1.2. Paper structure:

This paper is composed of the following parts:

1. Introduction
2. System architecture
3. Methods and materials
4. Results and discussion
5. Work limitations
6. Future work
7. Conclusion

1.3. Paper Contributions and originality:

Although previous studies showed good performance and help improve security, a complete system is still missing. The main focus is binary detection, generic flood alerts, or dam wall failure simulation, as well as the lack of combined spatial and temporal hazard estimation, which eventually helps in risk management. To address these gaps. This work proposes the following parts:

1. The exploitation of multiple weather features surrounding the dam for better view of the environment
2. Use of a hybrid model stacking: One class SVM and isolation forest for anomaly detection
3. Conception of an AIoT pipeline for flood prediction
4. Time before flood calculation and threatened distance from the dam estimation with analytical computation instead of hydrological simulation to reduce latency
5. Hazard level classification
6. Sample experiment for system logic work using pedagogic equipments

1.4. System overview:

The system aims to enhance the dam's security from overtopping and potential breaching. It focuses on protecting the dam and human life in case of floods. Data from the area around the dam are collected using different types of sensors such as humidity and water level, communicated via Zig-Bee inside the network and MQTT on the outside, the processing pipeline employs one class SVM, isolation forest for anomaly detection, random forest and linear regression for prediction. Analytical formulations are used to compute and estimate the distance threatened in a numerical output rather than a map for latency reduction.

2. SYSTEM ARCHITECTURE

2.1. Data flow sequence:

The flow chart 1 displays with simplicity the flow sequence; Specifies the main nodes through which the data pass.

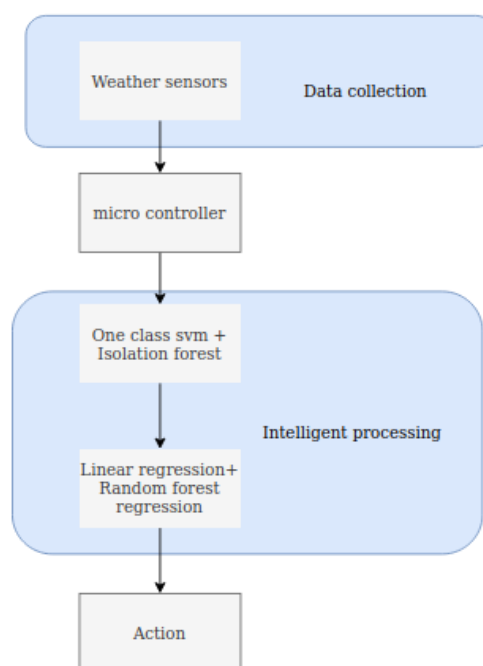


Figure 1. Data flow sequence

Illustrative system scenario: To understand how the system architecture operates; Consider a scenario where the system collects data in real time from different sensors like weather conditions and structural parameters such as pressure against the dam walls, the anomaly detection module -based on one class SVM and isolation forest model- learns the normal patterns and standard case scenarios. When an anomaly is detected, it is immediately flagged, which will activate the flood hazard prediction module -based on linear regression and random forest regressor- of the pipeline. In the first place, water level and flow data are used to estimate the time left before reaching the warning threshold; if exceeded, additional data are considered to define structural stress, helping to predict the severity of the situation. In the critical scenario where water continues to increase with high hydrodynamic intensity despite mitigation, the system evaluates the flow velocity and the increase rate of the water level to estimate the arrival time of the flood, and the threatened distance with analytical computation instead of hydrological simulation. During this procedure, the IoT communication framework guarantees dependable data transfer, employing Zig-Bee for local sensor networking and MQTT for external interactions, facilitating immediate responsiveness and synchronized decision-making.

2.2. Data collection module

for data collection, the architecture consists of different kinds of sensors:

- Water level sensor: Serves two different purposes
 1. Measurement of the water level of the dam at different times
 2. The calculation of the water inflow and outflow
- Wind direction sensor and wind speed sensor: Help us define how the water in the dam is to be affected by the wind.
- Atmospheric pressure sensor: Typically, decreasing atmospheric pressure corresponds to cloudy, rainy weather; the information provided by this sensor can help us estimate the chances of having rainfall
- Soil humidity sensor: Helps to determine the presence of rainfall, its heaviness, and also detect faulty water level sensors
- Water pressure and velocity sensors: Help to establish patterns between weather and water condition and the pressure against the wall, which can lead to potential dam breaches
- Air humidity and temperature sensors: Helps in the detection of any sudden change in the pattern of the climate that can lead to abrupt heavy rainfalls

For the data collected to be more reliable, the deployment of several sensors helps to distinguish whether any sensor is giving wrong or erroneous data. Along with that, using filters like Kalman filter will significantly reduce the noise or estimate the incomplete data.

When it comes to the communication inside the IoT network, between sensors and micro-controllers; We propose the use of Zig-bee as it is efficient for this inside network task with lower power consumption.

2.3. Data storage and treatment module

In the architecture, we can use the capabilities of cloud computing to provide efficient and scalable data storage solutions. Also; we can analyze and treat the data either in real time or after storing them and using them like historical data for learning and training.

For out network communication (cloud and micro-controller), we suggest the use of MQTT to have good scalability and high efficiency in out network communication over an ad hoc network which is more suitable for dynamic and harsh environments. The ad hoc network will help the data to be transferred within the network flexibly without a fixed infrastructure, and its self organization gives more resilience and assures connectivity in case of a link failure. The figure 2 displays the utilization of communication technologies in and outside the IoT network.

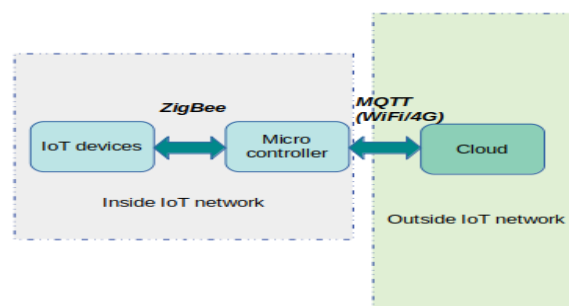


Figure 2. Communication technologies in and outside the network

2.4. Artificial Intelligence choice:

Our work is divided into two different phases:

2.4.1. Anomaly detection:

Seeing that this part focuses on detecting an anomaly out of myriad case scenarios; we tested the learning abilities for two options: First, a one-class algorithm which is One-class SVM for its known theoretical attributes: that only takes normal case scenario as data to learn along with the non-linear mapping due to the randomness of the values of our data, it also gives the possibility of parameter settings and adjustment according to the needs to give the best results possible. Second, a hybrid model that combines both one class SVM and isolation forest which is good for multi-dimensional data, suitable for unsupervised learning; in order to have better anomaly detection.

2.4.2. Flood hazard prediction:

In this part, three case scenarios are studied: normal, danger, and critical. Two regression models are used for this purpose:

- Linear regression model: Helps to draw the patterns between any linear values
- Random forest regressor: Used to learn the non linear patterns
- The final model is a linear regression model that takes the output of the previous predictions and uses them as input for the final output

The parts and hybrid models are all called in a pipeline function along with the data to be tested either in normal scenarios or anomaly / danger scenarios

The model starts with water level and inflow to estimate the time in which the water will reach the defined thresholds if it continues that rate.

The model assumes continuity and predicts accordingly. For flood prediction: We had to separate the dataset for every scenario, we focused for the normal scenario on the water inflow and water level. For the second situation, when the water level surpasses a certain threshold, the model focuses on other features like water pressure against the wall and wind speed, these two features can give insight about the intensity of the water movement and the potential demolition of the dam. In addition, this scenario is supposed to be simultaneous with the opening of the dam to help reduce the pressure on the wall and protect it from breaking or overflowing. Finally, if the water still rises despite the opening of the dam and reaches the critical threshold, the model uses the rise of the water level and the velocity of the flow to tell the approximate time of the flood and the threatened distance for evacuation alerts or any other measures. The model will use the real data after giving the approximate values to tell if that scenario is what is happening or not.

System latency: The proposed system is made to have low latency in order to estimate threats and react in the shortest time. For communication, the adopted protocols Zig-Bee for data acquisition and MQTT for data transfer have a low communication delay (usually less than 100 ms per hop). The machine learning models employed are specifically chosen to be simple and lightweight , —One-Class SVM, Isolation Forest, Random Forest, and Linear Regression—allow operations and predictions to be computed in under 1 second on standard hardware. the output of estimated threatened distance is chosen numerical rather than spatial to reduce generation latency. Thus, the end-to-end system latency, which includes both data processing and transmission, is below operational thresholds for timely hazard assessment, making it suitable for real-time applications compared to more computationally intensive approaches.

3. METHODS AND MATERIALS:

3.1. Software choice:

Our work was simulated on a computer station running on a Linux operating system; We trained our model on GoogleColab with Pytorch; Visual studio and arduino IDE were used for experimental purposes. Some of our datasets were downloaded from the meteostat and Kaggle website and others were generated in the seasonal ranges for testing and training purpose, mainly, these ranges depend on the available data of the Oroville dam in the USA like the website National Weather Service. The experimental data was collected from the sensors. The table 1 is a sample of the simulated dataset.

Table 1. Dataset sample

Soil H (%)	Air H (%)	Temp (°C)	Water L (m)	Wind Dir (°)	Wind (m/s)	Sp	Pressure (hPa)
72.0	72.6	20.7	2.0	227.6	6.1		1014.1
71.1	65.5	21.3	2.1	346.9	6.2		1013.8
67.7	64.8	23.3	1.9	236.9	5.6		1013.5
74.1	69.1	25.2	1.9	127.3	7.2		1013.7
66.2	67.7	24.3	2.0	256.2	8.8		1013.2

3.2. System electronic architecture:

Theoretically, the system's architecture is composed of: Water level, wind direction, wind speed, atmospheric pressure, soil humidity, air humidity, and water pressure, and water flow velocity sensors. Micro-controller for data transfer. Cloud for processing. Actuators for reaction. Like shown in 3

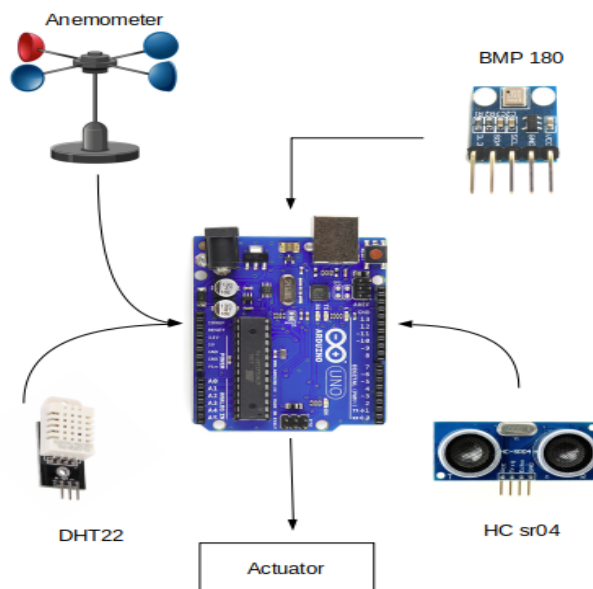


Figure 3. Architecture devices

On a smaller scale for experimental purposes ; We used the following:

- water level sensor
- Micro-controller board
- humidity sensor
- temperature sensor
- buzzer

These equipments are pedagogic and only for testing. For a more environmentally suited and real life application we can switch to other devices.

3.3. System logical flowchart:

This flowchart 4 summarizes the logic of this system.

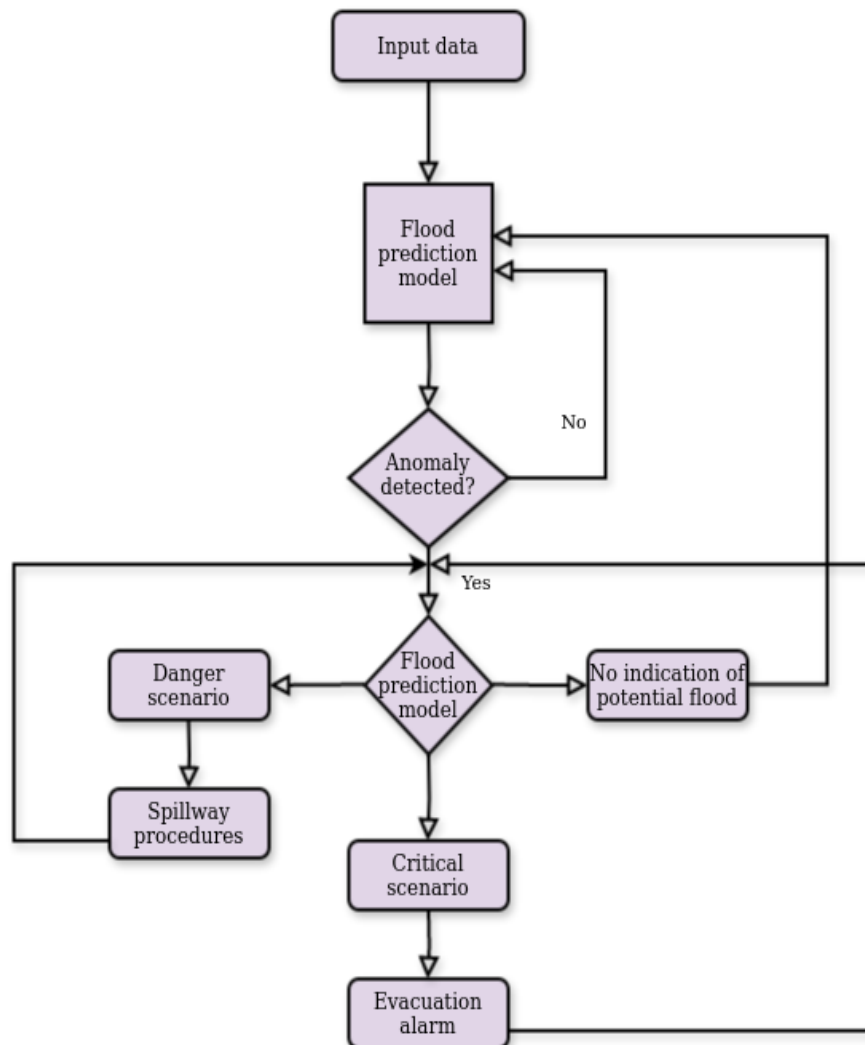


Figure 4. System logical flow chart

This flowchart displays the logic that the system uses to go from one node of the pipeline to another. The chart is an enclosed loop of real time data analysis and treatment. In case of safety, the system will continue reading and analyzing it; Once a potential hazard is detected, the system goes down node by node until it makes a final decision without stopping to read the necessary data.

3.4. Pseudo code

This is a pseudo-code for the experimental part

Algorithm 1 Sensor Data Ingestion and Flood Hazard Detection

```

Initialize: Arduino port  $\leftarrow$  'arduino_port'
Initialize empty lists:  $water\_level \leftarrow []$ ,  $Temperature \leftarrow []$ ,  $Humidity \leftarrow []$ 
▷ Read and append data from serial port

 $water\_level \leftarrow water\_level + [float(parts[0])]$ 
 $Temperature \leftarrow Temperature + [float(parts[1])]$ 
 $Humidity \leftarrow Humidity + [float(parts[2])]$ 
▷ Load trained models

 $svm\_model \leftarrow joblib.load("...")$ 
 $iso\_model \leftarrow joblib.load("...")$ 
 $normal\_model \leftarrow joblib.load("...")$ 
 $danger\_model \leftarrow joblib.load("...")$ 
 $critical\_model \leftarrow joblib.load("...")$ 
 $scaler \leftarrow joblib.load("...")$ 
▷ Create data frame

 $data \leftarrow$  DataFrame with columns:
  'Air Humidity (
  'Temperature (°C)'  $\leftarrow Temperature$ 
  'Water Level (cm)'  $\leftarrow water\_level$ 
▷ Run hybrid anomaly detection

 $is\_anomalous \leftarrow$ 
 $hybrid\_anomaly\_predict(data, iso\_model, svm\_model, numeric\_features, window\_size)$ 
 $num\_zeros \leftarrow count(is\_anomalous = 0)$ 
 $num\_ones \leftarrow count(is\_anomalous = 1)$ 

if ( $is\_anomalous == 1$ ) contains any true then
  Call  $flood\_hazard\_pipeline(data, normal\_model, danger\_model, critical\_model)$ 
else
  Print "No Anomaly detected"
end if

```

3.5. Methodological Details:

- Data generation: Currently, there is no public dataset available that includes all the necessary features needed for this work. Therefore, a synthetic dataset was generated, based on online data from one region and a combination of available data, with seven weather features. The seasonal and daily cycles were introduced using the sinusoidal component. Gaussian noise was added for a more natural effect. anomalous variables were added at random points. Data were split into 80% for training and 20% for testing in the second part. Cross validation of 5-folds and 10-folds was used to guaranty robustness. For the experimental part, data were collected from sensors for small scale training and performance testing
- Pre-processing: All continuous features were normalized, "NaN" values were removed.
- Hybrid model: For anomaly detection part, combined one class SVM and isolation forest with rolling statistics
- Evaluation: True anomalies were compared with predictions using the evaluation metrics: precision, recall, and F1-score.
- Prediction: two thresholds were defined: Danger and critical. The time estimation was calculated with this equation:

$$y = \frac{V_{available}}{Q_{inflow}} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 3) \quad (1)$$

Where $V_{available}$ is the volume of water until the danger threshold is reached, Q_{inflow} is the mean inflow and ϵ adds noise

- Final model: We used a stacked regression framework with base models being Random Forest Regressor and Linear Regression
- We performed a hyperparameter tuning for all models to ensure the best possible performance. Except for linear regression due to the absence of tunable hyperparameters

4. RESULTS AND DISCUSSION:

4.1. Study results:

4.1.1. Part one:

When it comes to anomaly detection, we trained two models one class SVM and isolation forest each on its own. The results of the precision, recall and f1 are displayed in table 2:

Table 2. Theoretical detection results

Model	Precision	Recall	F1 score
One class SVM	0.5263	0.80000	0.6349
Isolation forest	0.8889	0.1600	0.2712

However, when we trained a hybrid model that is based on one class SVM and isolation forest models, it showed better performance in detecting anomalies even though there were still false positives in result. The model's performance was evaluated using the precision, recall and f1 score function where it resulted in: 86.6%, 78% and 82.11% in sequence. The figures below display the ROC curve 5 and confusion matrix 6; This shows the discriminatory ability of the model and the detection ability with a rate of 78%

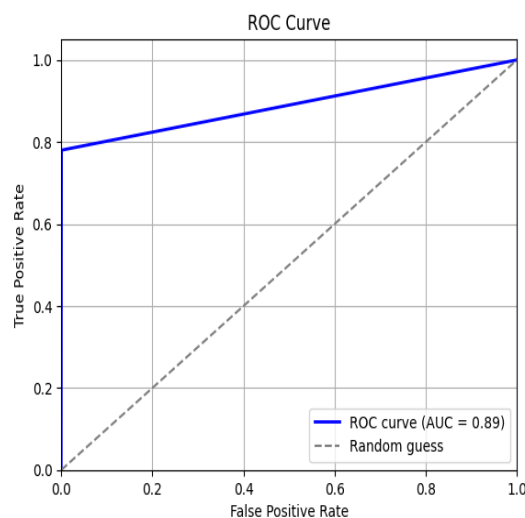


Figure 5. ROC curve

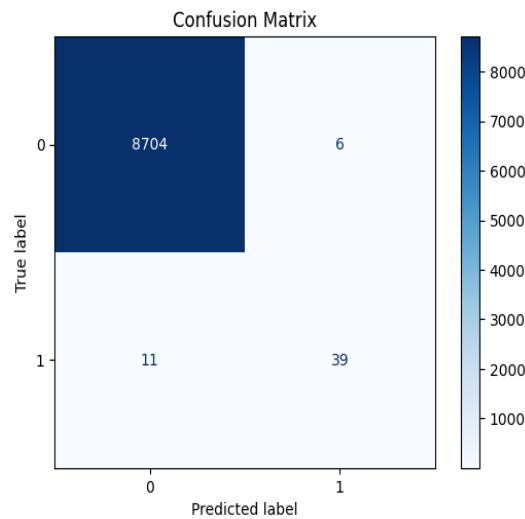


Figure 6. Anomaly detection confusion matrix

4.1.2. Part two:

The models learned separately for three different datasets, depending on the scenario we are building. The combination of linear regression and random forest regression was tested using MSE and R^2 and the results were in the range[0.0025-0.0036] and 0.99 in sequence. The 5-fold cross validation gave the results in table 3

Table 3. Cross-validation MSE results

Metric	Value
MSE scores (per fold)	$[-0.00252, -0.00255, -0.00261, -0.00247, -0.00282]$
Mean MSE	0.00260
Standard Deviation	0.00012

These results show a stable and reliable performance. After cross-validation, we performed a nonparametric test on the three models: Two base models and a stacked model, to see if the differences are statistically significant. The Friedman test was applied across cross validation after increasing the amount to 10 folds, followed by pairwise Wilcoxon signed-rank tests. The results are as follows:

Friedman test p-value: 4.54×10^{-5}

Wilcoxon RF vs. LR: 0.0019

Wilcoxon RF vs. Stacked: 0.0019

Wilcoxon LR vs. Stacked: 0.99

As seen; The Friedman p value < 0.05 which indicates an overall difference. For Wilcoxon, values < 0.05 indicate a significant difference in performance between LR and Rf, RF and stacked. However, the > 0.05 value indicates that the LR and stacked perform similarly.

4.1.3. Part three:

The third and last part consists in the gathering of the previously mentioned part in a pipeline function where the parameters taken are the new data to test and the pre-trained models; This function was tested in several cases:

- Normal data in normal ranges with noise introduced within: This resulted in the stop at the first part with "Normal behavior" output message

- Anomaly injected data with noise introduced within: This resulted in a jump to the second part for the classification of the potential danger, which itself is divided into different scenarios:
 1. No flood indication: When calling the second part, new data values were introduced to help better classification (water level, water inflow); In this scenario, we introduced these two features in normal ranges. The model gave the results of "No flood indication" message
 2. Danger indication: In this part, after the danger threshold is exceeded, the wind speed and water pressure were also taken into account; The model signals the potential danger along with the estimated time before the disaster begins "WARNING: Flood expected in 30.45 minutes"

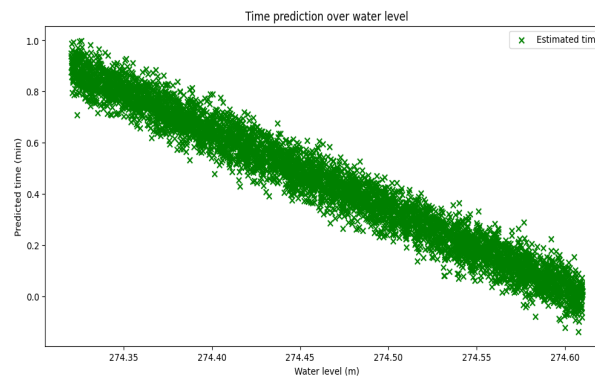


Figure 7. time estimation over water level

It is advised to fully open the spillways at this point to reduce the inflow and water pressure on the dam

3. Critical case scenario: When the water level exceeds the critical threshold: The model was able to identify the threat and resulted in three main outputs: scenario 'critical', time left before the flood which in fewer cases might be a negative value which indicates that the flood had already begun, estimated threatened distance which was 32.4Km in an hour.
in the figure below we can see the time estimation before next level (coming critical situation or flood) for both scenarios:

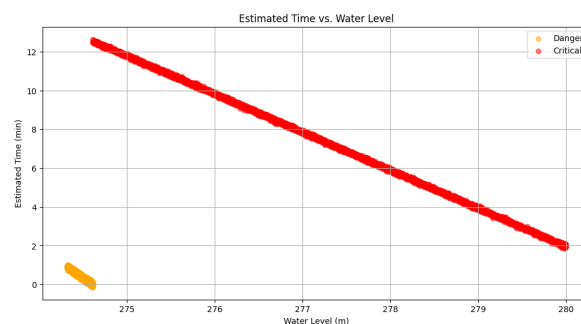


Figure 8. Time estimation with water level

These scenarios are built on the logic of prediction. That is, when anomalous values are triggered, the models predict the potential scenario if those weather features continue in those ranges, which helps to take security measures and counteraction for the safety of the dam and people.

When it comes to the experimental parts, the same models were trained on the real gathered data and then tested on a smaller scale. The results are displayed in the table below for internal comparison.

Table 4. Experimental data results

Model	Precision	Recall	F1 score
One class SVM	0.5524	0.5800	05659
Isolation forest	0.8300	0.8300	0.8300
Hybrid model	0.6309	0.9400	0.7550

The figures below 9 10 11 show the cases of detection of the hybrid model: All danger, No danger and no to predicted danger

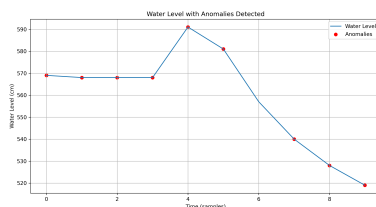


Figure 9. Danger of flood

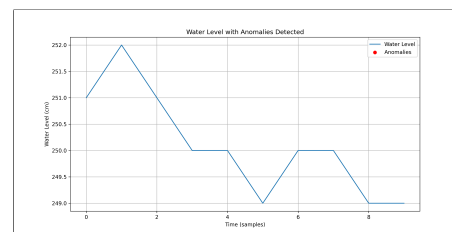


Figure 10. No Danger

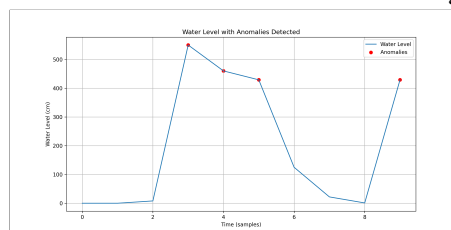


Figure 11. No anomaly to danger

4.2. Comparison and discussion:

4.2.1. Comparative analysis and justification:

The table 5 displays the relevance of our approach going beyond just one task such as detecting anomalies or forecasting the threatened area alone. Our approach combines several important security measures by:

- Classification of the hazard level
- Estimation of time
- Estimation of threatened distance

Which were not addressed in previous studies, either not addressed at all or not all used in one system.

The combination of all the necessary aspects in one system explains why our approach is suitable for dam safety.

Design Choices and Comparison with Existing Approaches: This work is different from many existing studies. By offering a complete system that combines anomaly detection, hazard classification, and temporal prediction, and distance estimation. Unlike approaches based on 3D mapping or flood simulation. The proposed method provides a numerical estimate of the distances threatened.

Table 5. Comparison of existing works and the proposed method.

Method	Core approach	Anomaly detection	Scenario level	Time calculation	Distance estimation	Evaluation metrics
[19]	anomaly detection using STCLP	✓	✗	✗	✗	95% AC, 75% recall, 80% F1 score
[20]	joint analysis of monitoring variables	✓	✗	✗	✗	above 98%
[21]	autoencoder for automatic anomaly detection	✓	✗	✗	✗	99.9% F1 score, 1 AUC
[26]	flood map prediction with transformers	✗	✗	✗	✓	RMSE =10.4, F1=94%
This work	flood prediction	✓	✓	✓	✓	recall 94%, MSE 99%

In addition, lightweight machine learning models are also used instead of transformer or computationally intensive deep learning methods, Which ensures low latency in operations while maintaining reliable performance. Although digital twin approaches offer detailed simulation capabilities, they often introduce significant computational overhead; In contrast, the proposed method prioritizes rapid and efficient estimation suitable for real-time deployment. The estimated distances threatened by the proposed system are calculated on the basis of measured water velocity and hazard classification. Although this approach does not depend on 3D mapping or hydraulic simulators, the resulting estimates are within plausible ranges for dam failure scenarios. This indicates that the analytical approach provides realistic and reliable distance estimates suitable for early warning and real-time deployment.

4.2.2. Anomaly detection:

For the first part; We noticed that the model could learn better as a hybrid due to the excellence covering of both models: One class SVM focuses on the normality and draws the border over it. The isolation forest is better at spotting the different anomalies, also, the one class SVM covers the weakness of the isolation forest towards the dense data; Hence, it is These two models could in a way balance one-an-other's limitations. as seen in difference between the data used in theory and in the small scale experiment, the hybrid model gave good performance in both cases. However; The false positives and negatives were not absent in a complete nor in an almost complete form, and this is due to the introduced seasonal variations that the climate features endure.

We could clearly see, as sampled in the figures 12 13 14 15 that most of the false positives and negatives in different features are spotted at the valleys The ROC curve figure shows that the curve is above the random guessing line which is a clear sign of good performance and the AUC = 0.89 that falls in good range

The solution is the addition of the second part itself, where false positives are more than false negatives; They can be dismissed when the data are again processed in the flood hazard prediction; i.e, the false positives will be ignored if the flood prediction model classifies the scenario in the range of "normal"

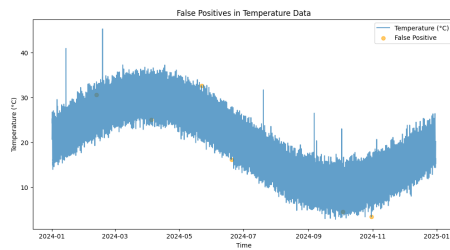


Figure 12. False positives in temperature

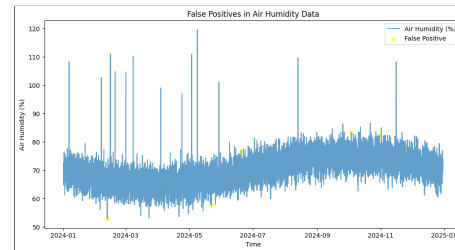


Figure 13. False positives in air humidity

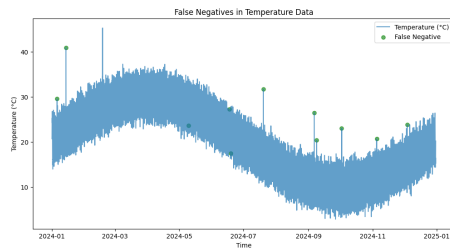


Figure 14. False negatives in temperature

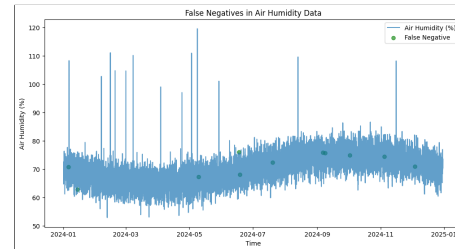


Figure 15. False negatives in air humidity

4.2.3. Flood prediction:

The results shown by MSE and R^2 display a good performance of the models due to the choice of data and its simplicity; The data were separated according to the scenario and, also, only necessary features were studied. The model in the three cases can give a good approximation of the level of flood hazard and the occurrence time.

4.2.4. Pipeline function:

This function takes in the necessary models and data and can not only detect anomalies, but also predict the level of the potential flood hazard. Every potential scenario triggers an additional function. The "danger" level estimates the time for an advance notice and warning, giving time for solutions such as opening dam spillways to reduce the pressure against the wall and slow the water rise, and the "critical" level triggers the threatened distance estimation based on water velocity to help evacuation measures. This work not only protects the dam from demolition in case of unexpected heavy rainfalls but also saves human lives.

5. WORK LIMITATIONS:

5.1. Material wise:

In order to cover a very vast area, The quantity of devices increases the cost of the solution. The possible solutions for this would be to cover only the dam wall, use the sensors only when needed depending on the scenario.

5.2. Smart model wise:

The model needs a large amount of data to learn efficiently, which can sometimes be difficult to obtain and require so much storage. In addition, the climate features tend to change over the year even in the same zone, Thus, the model struggles in those seasonal valleys. The work is meant for future use in real life which will need large storage space for the actual data to be used as historical for further learning and training.

6. FUTURE WORK:

Although our system was detailed and our work was tested on a small scale with pedagogic devices, we still want to introduce other aspects in the system study like latency, energy consumption, and security in more detail and experimental measures; Hence, our future work aims for a real work deployment of this system in which we will use other micro controllers and integrate communication protocols and security measures.

7. CONCLUSION:

Floods are a very dangerous natural phenomenon that can occur unexpectedly, especially due to climate change; One of the threatened structures are the dams; being a sensitive infrastructure their security is needed from different aspects. Hence, in this work, we proposed an AIoT based system to help protecting the dam walls and human lives surrounding it. Where the indication of potential hazard triggers further calculations; If the danger is still flagged with high values of water pressure against the wall. Time is calculated and threatened distance is estimated to protect human lives in signaled perimeters. The results of this work showed a promising path for the learning and application in anomaly detection and hazard classification, especially the use of a hybrid model, who will find the balance for a more accurate and efficient performance. Future works will need to use more real historical data for dam flood cases for a higher accuracy and performance evaluation.

REFERENCES

- [1] C. Prakash, A. Barthwal, and D. Acharya, "Floodalert: An internet of things based real-time flash flood tracking and prediction system," *Multimedia Tools and Applications*, pp. 1–27, 2023.
- [2] J. S. Tradowsky, S. Y. Philip, F. Kreienkamp, S. F. Kew, P. Lorenz, J. Arrighi, T. Bettmann, S. Caluwaerts, S. C. Chan, L. De Cruz *et al.*, "Attribution of the heavy rainfall events leading to severe flooding in western europe during july 2021," *Climatic Change*, vol. 176, no. 7, p. 90, 2023.
- [3] R. Martać, N. Milivojević, V. Milivojević, V. Ćirović, and D. Barać, "Using internet of things in monitoring and management of dams in serbia," *Facta Universitatis, Series: Electronics and Energetics*, vol. 29, no. 3, pp. 419–435, 2015.
- [4] V. Kumar, K. V. Sharma, T. Caloiero, D. J. Mehta, and K. Singh, "Comprehensive overview of flood modeling approaches: A review of recent advances," *Hydrology*, vol. 10, no. 7, p. 141, 2023.
- [5] K. Chaduvula, B. R. Markapudi, C. R. Jyothi *et al.*, "Design and implementation of iot based flood alert monitoring system using microcontroller 8051," *Materials Today: Proceedings*, vol. 80, pp. 2840–2844, 2023.
- [6] B. Kaur, S. Dadkhah, F. Shoeleh, E. C. P. Neto, P. Xiong, S. Iqbal, P. Lamontagne, S. Ray, and A. A. Ghorbani, "Internet of things (iot) security dataset evolution: Challenges and future directions," *Internet of Things*, p. 100780, 2023.
- [7] G. Al-Rawas, M. R. Nikoo, M. Al-Wardy, and T. Etri, "A critical review of emerging technologies for flash flood prediction: Examining artificial intelligence, machine learning, internet of things, cloud computing, and robotics techniques," *Water*, vol. 16, no. 14, 2024. [Online]. Available: <https://www.mdpi.com/2073-4441/16/14/2069>
- [8] J. Shi, Z. Yin, A. Leon, J. Obeysekera, and G. Narasimhan, "Fidlar: Forecast-informed deep learning architecture for flood mitigation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 27, 2025, pp. 28 377–28 385.
- [9] H. R. C. Arante, E. Sybingco, M. A. Roque, L. Ambata, A. Chua, and A. N. Gutierrez, "Development of a secured iot-based flood monitoring and forecasting system using genetic-algorithm-based neuro-fuzzy network," *Sensors*, vol. 25, no. 13, 2025. [Online]. Available: <https://www.mdpi.com/1424-8220/25/13/3885>
- [10] J. A. Marengo, M. E. Seluchi, A. P. Cunha, L. A. Cuartas, D. Goncalves, V. B. Sperling, A. M. Ramos, G. Dolif, S. Saito, F. Bender *et al.*, "Heavy rainfall associated with floods in southeastern brazil in november–december 2021," *Natural Hazards*, pp. 1–28, 2023.
- [11] J. K. A. Singh and A. Huong, "Iot flood monitoring system," *Evolution in Electrical and Electronic Engineering*, vol. 4, no. 1, pp. 45–54, 2023.
- [12] M. I. Zakaria, W. A. Jabbar, and N. Sulaiman, "Development of a smart sensing unit for lorawan-based iot flood monitoring and warning system in catchment areas," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 249–261, 2023.

- [13] Y. Hahn, P. Kienitz, M. Wönkhaus, R. Meyes, and T. Meisen, "Towards accurate flood predictions: A deep learning approach using wupper river data," *Water*, vol. 16, no. 23, 2024. [Online]. Available: <https://www.mdpi.com/2073-4441/16/23/3368>
- [14] O. Zabihi, M. Siamaki, M. Gheibi, M. Akrami, and M. Hajiaghahi-Keshteli, "A smart sustainable system for flood damage management with the application of artificial intelligence and multi-criteria decision-making computations," *International Journal of Disaster Risk Reduction*, vol. 84, p. 103470, 2023.
- [15] H. M. Eldeeb, A. Ibrahim, M. H. Mowafy, M. Zeleňáková, H. F. Abd-Elhamid, K. Pietrucha-Urbanik, and M. T. Ghonim, "Assessment of dams' failure and flood wave hazards on the downstream countries: A case study of the grand ethiopian renaissance dam (gerd)," *Water*, vol. 15, no. 8, p. 1609, 2023.
- [16] Q. Zhang, Y. Li, S. Yu, L. Wang, Z. Chen, and J. Zhou, "Rapid quantitative study of check dam breach floods under extreme rainstorm," *Natural Hazards*, vol. 116, no. 2, pp. 2011–2031, 2023.
- [17] M. A. Hariri-Ardebili, G. Mahdavi, L. K. Nuss, and U. Lall, "The role of artificial intelligence and digital technologies in dam engineering: Narrative review and outlook," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106813, 2023.
- [18] C. Rizzo, A. Maranzoni, and M. D'Oria, "Probabilistic mapping and sensitivity assessment of dam-break flood hazard," *Hydrological Sciences Journal*, vol. 68, no. 5, pp. 700–718, 2023.
- [19] Y. Wang and G. Liu, "Self-supervised dam deformation anomaly detection based on temporal-spatial contrast learning," *Sensors*, vol. 24, no. 17, p. 5858, 2024.
- [20] F. Salazar, A. Conde, J. Irazábal, and D. J. Vicente, "Anomaly detection in dam behaviour with machine learning classification models," *Water*, vol. 13, no. 17, p. 2387, 2021.
- [21] I. T. Nicholas, J. R. Park, K. Jung, J. S. Lee, and D.-K. Kang, "Anomaly detection of water level using deep autoencoder," *Sensors*, vol. 21, no. 19, p. 6679, 2021.
- [22] S.-M. Kong, C. Yoo, J. Park, J.-H. Park, and S.-W. Lee, "Aiot monitoring technology for optimal fill dam installation and operation," *Applied Sciences*, vol. 14, no. 3, 2024. [Online]. Available: <https://www.mdpi.com/2076-3417/14/3/1024>
- [23] A. Belay, B. GebreMariam, and G. Geremew, "Machine learning based design flood forecasting for the ribb embankment dam in the tana sub-basin, abay river basin, ethiopia," *Discover Environment*, vol. 3, p. 160, 2025.
- [24] S. Güneş Şen, "Machine learning-based water level forecast in a dam reservoir: A case study of karaçomak dam in the kızılırmak basin, türkiye," *Sustainability*, vol. 17, no. 18, 2025. [Online]. Available: <https://www.mdpi.com/2071-1050/17/18/8378>
- [25] K. M. Khafaji and B. B. Hamed, "Employing deep learning in crisis management and decision making through prediction using time series data in mosul dam northern iraq," *PeerJ Computer Science*, vol. 10, p. e2416, 2024.
- [26] M. Pianforini, S. Dazzi, A. Pilzer, and R. Vacondio, "Real-time flood maps forecasting for dam-break scenarios with a transformer-based deep learning model," *Journal of Hydrology*, vol. 635, p. 131169, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S002216942400564X>
- [27] D. Bao, Z. G. Xue, M. Hiatt *et al.*, "A machine learning-based prediction-to-map framework for rapid and accurate spatial flood prediction," *npj Natural Hazards*, vol. 2, p. 71, 2025.
- [28] M. Wajid, M. K. Abid, A. A. Raza, M. Haroon, and A. Q. Mudasar, "Flood prediction system using iot & artificial neural network," *VFAST Transactions on Software Engineering*, vol. 12, no. 1, pp. 210–224, 2024.