



Article

A Smart Framework for Managing Natural Disasters Based on the IoT and ML

Fares Hamad Aljohani ^{1,*}, Adnan Ahmed Abi Sen ^{1,*}, Muhammad Sher Ramazan ¹ , Bander Alzahrani ¹ 
and Nour Mahmoud Bahbouh ²

¹ Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

² Department of Information and Communication Sciences, Granada University, 18071 Granada, Spain

* Correspondence: fnahezaljohani@stu.kau.edu.sa (F.H.A.); adnanmmn@hotmail.com (A.A.A.S.)

Abstract: Natural disasters greatly threaten our lives in addition to adversely affecting all activities. Unfortunately, most solutions currently used in flood management are suffering from many drawbacks related to latency and accuracy. Moreover, the previous solutions consider that the whole city has the same level of vulnerability to damage, while each area in the city may have different topologies and conditions. This study presents a new framework that collects data in real-time about bad weather, which may cause floods, where the framework has a proposed classification algorithm to process sensed data to determine the level of danger in each area of the city. In case of a threat, the framework will send early alerts to users and rescue teams. The framework depends on the Internet of Things (IoT) and fog computing coupled with multiple models of machine learning (Rain Forest, Decision Tree, K-Nearest Neighbor, Support Vector Machine, Logistic Regression, and Deep Learning) to enhance performance and reliability. In addition, the research suggests some assistant services. To prove the efficiency of the framework, we applied the proposed algorithm to real data for the city of Jeddah, Saudi Arabia, for the years 2009 to 2013 and for the years 2018 to 2022. Then, we depended on standard metrics (accuracy, precision, recall, F1-score, and ROC curve). The Rain Forest and Decision Tree achieved the highest accuracy, exceeding 99 percent, followed by the K-Nearest Neighbor. The framework will provide flood detection systems that can predict floods early, send a multi-level warning, and reduce financial, human, and infrastructural damage.

Keywords: Internet of Things; fog computing; machine learning; flood risk prediction; wireless sensor networks; reliability



Citation: Aljohani, F.H.; Abi Sen, A.A.; Ramazan, M.S.; Alzahrani, B.; Bahbouh, N.M. A Smart Framework for Managing Natural Disasters Based on the IoT and ML. *Appl. Sci.* **2023**, *13*, 3888. <https://doi.org/10.3390/app13063888>

Academic Editors: Muhammad Babar, Saleem Iqbal and Aftab Khan

Received: 11 February 2023

Revised: 13 March 2023

Accepted: 16 March 2023

Published: 18 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Machine learning and the Internet of Things (IoT) are rapidly changing the world into an intelligent space leading to changes in how things get done in various fields. According to Al-Barazanchi et al. [1], the Internet of Things enables continuous data collection through devices scattered all over, and the data is then sent from these devices to the cloud for analysis, processing, inferring knowledge, determining behaviors, and designing of more intelligent technologies to allow for seamless services for users.

The IoT incorporates wireless network sensors (WSNs) and radio frequency identifications (RFID) for data capturing and transfer. The WSNs capture data related to physical conditions in human beings' environment, such as temperature, pressure, heart rate, illumination, humidity rate, wind speed, the amount of precipitation, and much more. The RFID radio identifiers give objects unique identifiers. In IoT, RFID radio identifiers enable the identification of objects and capturing of data related to the objects remotely [2].

There are vast applications of wireless network sensors in the Internet of Things technologies in different majors and systems such as health, learning, and crowd and disaster management to name a few. Thus, these systems aim to apply the technology for data collection in the real time, then analyze it by machine learning algorithms for more sophisticated services [3,4].

Disaster management systems are among the most important systems in which the Internet of Things and artificial intelligence have played significant roles. Floods are among the most common natural disasters combated using technology, notably through hydraulic modeling [5]. Hydraulic modeling is often less effective when it comes to flood prediction, as proved in countries already implementing various models but are still faced with the issue of flooding. To counter the ineffectiveness of hydraulic modeling in flood risk management, studies on how recent technologies, such as the application of machine learning and IoT, can aid in disaster management and have a useful role in relaxing the effects of some types of disasters, such as floods [6].

Floods are frequent and are accompanied by massive destruction, ranging from loss of life to agricultural damages and poor economic conditions. The United Nations [7], through the UN Institute of Water, Environment, and Health, states that floods affect more than half a billion people every decade. Asia is primarily affected by floods per the United Nations; notably, Saudi Arabia has recently witnessed at least a dozen flood-related deaths and injuries following the torrential rains. Intelligent flood detection systems are becoming necessary in the disaster management plans of cities [8]. Feasibility studies carried out when the authorities explore the possible solutions to floods point to incorporating technology into the legacy methods already in use. The most feasible technologies revolve around the Internet of Things approach based on wireless network sensors and machine learning [9].

WSNs have been heavily utilized recently due to the ease of use and maintenance coupled with their affordability and high data accuracy; wireless sensors' power usage is relatively low, thus they have more durability [10]. The wireless network sensors connect to a base station wirelessly, the central processing unit of a wireless sensor network. Many recent wireless sensors connect using wireless technologies, such as LoRa, ZigBee, NFC, Ultra-wideband, and Bluetooth [11].

Bluetooth, NFC, and Zigbee technologies consume less power but are preferable for short-range connections and, thus, are unfavorable for applications that cover vast areas. Notably, the prementioned wireless technologies are not cellular network technologies, including 2G, 3G, 4G, and 5G technologies. Cellular technologies can cover broad areas but are faced with the problem of high-power consumption; thus, they are unsuitable for sensitive systems of delay [12].

The WSNs and RFID connect to the cloud through fog computing. Fog computing is a system-level horizontal architecture that distributes computing, storage, resources, management, and networking services across the range from Cloud to Things [13]. Fog computing, in this case, minimizes the bandwidth needed and the back and onward contact between the cloud and the sensors, which could affect the entire system's functionality. Fog and cloud computing then connect, and thus refined data collected from the sensors reaches the cloud for operation. According to collected data in the cloud, and in case of any event, few governmental organizations send warning messages (SMS) to the phones for the public. This method of notifying is not effective and does not support real-time awareness for users [14].

The contributions of this research will be as follows:

- First: A review study on flood early detection, the solutions presented in that, and the negatives and advantages of those solutions.
- Second: presentation of a proposed approach platform base on a hybrid solution by integrating the IoT and machine learning algorithms.
- Third: Building a smart algorithm to classify and predict the degree of risk in advance. We replaced the traditional prediction of the rainfall amount by RT classification of the level of threat according to many weather variables and previous situations.
- Fourth: An actual study of actual data in the city of Jeddah, Saudi Arabia, during the previous years (2009 to 2013) and years (2018 to 2022) to prove the effectiveness of the proposed algorithm.
- Fifth: Building an application to alert users and educate them in advance in the event of danger and alert the Civil Defense to take the necessary measures and a decision.

Jeddah city in Saudi Arabia is our study area. It is situated on a coastal plain. The west border is the Red Sea, and the east is Hijaz Mountains. Usually, valley areas are more threatened by flooding than others, especially after heavy rains [15]. Given the city's heavy floods in 2011, 2012, 2016, and most recently, last November 2022, it was necessary to provide an efficient solution to reduce the effect of these cases. That can be achieved by a smart system providing early notifications about any potential floods. Figure 1 shows damages resulting from the last flood in 2022 in Jeddah.



Figure 1. Jeddah floods, 2022.

Following are the main and sub-sections of this paper. Firstly, a review study of the current solutions for managing disasters, especially floods. The research presents the proposed framework and classification algorithm in detail. After that, the research discusses the implementation of a real dataset and results according to standard metrics. Finally, the conclusion and future trends.

2. Related Work

In this section, various literature on the applications of the IoT and ML in disaster management is discussed.

2.1. Internet of Things in Disaster Management

Beltran et al. [13] discussed an Arduino-based disaster management system whose effectiveness is weighed through experiments. The system collects data using sensors, after which users are notified through messages about potential disasters. The system utilizes GSM wireless connections, which are not energy efficient. Additionally, the system does not incorporate machine learning for disaster forecasting. Adeel et al. [16] conducted a survey on the roles of Wireless Sensor Networks and the Internet of Things in disaster management systems. The survey investigates systems applying the models above and ascertains the effectiveness of such systems. However, the systems experience high packet loss and lack robust, fair data rate allocation.

Sciullo and Di Felice [17] designed a LoRa-based mobile emergency management system through which users can request help in case of disasters. The system effectively enables long-range multi-hop communication using the Internet of Things and smartphones, which could be influential in the case of disasters. Xu, Ota, and Dong [18] designed a UAV-mounted mobile edge computing system based on the LoRa technology for data collection. Although the system is effective, it is faced with the problem of low data transmission rates. The results of a study by Ejaz et al. [19] show the enormous impact on the performance of IoT platforms for wildfire management. Notably, the Internet of Things is valiantly applied in disaster management, and the performance of the frameworks applying IoT, in this case, can handle most of their requirements.

2.2. Machine Learning in Disaster Management

Chamola et al.'s [20] survey suggests that machine learning is applied in disaster management through machine learning algorithms useful for predicting disasters and handling post-disaster data. The study concludes that algorithms are helpful in disaster forecasting and impact analysis if applied appropriately. Another work [21] supports the argument by stating that “machine learning and deep learning facilitates disaster management tasks in visualizing, analyzing, and predicting natural a disaster”. The application of machine learning discussed in the two literary works revolves around machine learning in data analysis, visualization, and disaster predictions.

Ranasinghe and Ilmini [22] reviewed different Machine Learning methods and algorithms such as Artificial Neural Networks (ANN), Support Vector Machine (SVM), Multilayer Perception (MLP), and Long Short-Term Memory (LSTM). They are all effective models that are used to forecast floods. Widiyari and Nugroho [23] discussed the application of the Multilayer Perceptron (MLP) model of machine learning in disaster management. Multilayer Perceptron is the most widely used neural network in time series data forecasting. Halgamuge et al. [24] also discussed neural networks and neural network optimizers and their essence in disaster management. Álvarez and Morales–Esteban [25] analyzed multiple approaches for spatial and temporal massive data analysis regarding natural disasters, and conclusively, neural networks aid forecasts as they help find the linear and nonlinear relationship data.

2.3. Predicting Precipitation

Nourani et al. [26] developed a system that utilizes ensemble modeling based on artificial intelligence to predict precipitation. The ensemble modeling combines the results of different artificial intelligence-based models applied in precipitation prediction. The modeling compared includes feed-forward neural network (FFNN), adaptive neural fuzzy inference system (ANFIS), and most miniature square support vector machines (LSSVM). The forward-feed neural network is considered the best for predicting precipitation in the study prior to applying ensemble modeling. FFNN is a type of artificial neural network that avoids loops.

Kang et al. [27] developed a long short-term memory model to aid precipitation prediction. The model is effective as it can eliminate inputs that show weak correlations. However, the system only works in the presence of extensive data. Yao et al. [28] developed a coupled flood forecasting approach based on the GXM model and WRF forecasts to help predict precipitation and floods. However, the uncertainties associated with the coupled approach are unestablished; thus, it is not as practical. In their study, Bukhari et al. [29] implemented the NARX model to predict precipitation. The model is effective as the authors end up stating, “Neural networks is the most suitable technique certainly for predicting different climate conditions as other parameters, such as temperature, relative humidity, and atmospheric pressure, can be included in the neural networks modeling for more precise forecasting”. Precipitation prediction is best achieved using artificial intelligence and machine learning models.

2.4. Flood Risk Management Systems

Samikwa et al. [30] designed a flood prediction system using IoT and ANN where the prediction computation is carried out on a low-power edge device. The system monitors real-time rainfall and water level time series and uses the temporal correlative information for ahead-of-time prediction of flood water levels using the LSTM model. The system is efficient but may be slightly ineffective in case of processing lesser data. Hassan et al. [31] conceptualized a system that warns users about potential floods. The system transmits data using a GSM modem. The system lacks the abilities of deep learning and consumes a lot of energy, thus it may be inefficient.

Esposito et al. [32] proposed a flood early warning system that is not dependent on mobile towers for alert message broadcasting. The system operates on low power and

works well even in the absence of the Internet using ZigBee. The system was designed to monitor a flood occurrence on a dam and provides real-time alerts in case of a flooding event while operating on low power. However, the system needs to include more data analysis for ahead-of-time prediction of a flooding event. Acosta-Coll et al. [33] proposed an early warning system that alerts on the potential of floods occurring. The system utilizes the Lora WAN communication protocols for the wireless sensor network. It also has sensors for data collection and a cloud for data processing. Data is sent directly to the cloud and that risks the system's functionality. Notably, the proposed system in this paper utilizes fog computing to eliminate such threats to the system.

Table 1 presents the summary of the previous contributions in the floods management.

Table 1. Summary of the literature review.

Research	The Objectives	Used Model	Result	Limitation
[13]	To create a system that alerts users of possible disasters.	Arduino coupled with GSM and ZigBee	The developed system alerts users of possible disasters via messages.	The solution does not involve artificial intelligence and smart technologies.
[16]	To survey the role of roles of IoT and wireless Sensor Networks in disaster management.	IoT and Wireless Sensor Networks.	A comprehensive study detailing the application of IoT and WSN in disaster management is developed.	The systems discussed that utilize 4G suffer high packet loss. The systems using Lora WAN lacks robust fair data rate allocation and power control methods to address existing
[17]	To deploy next-generation phone-based ECS by using the LoRA technology for long-range, multi-hop communication among people involved in the emergency and rescue teams.	LoRa	A platform named LOCATE is developed for the purpose of deploying phone-based ECS. LOCATE is a self-localization algorithm implementing a trilateration technique, and a multi-hop dissemination scheme which combines probabilistic DTN and biased-contention mechanisms.	The LoRa communication model applied is characterized by huge delays thus the platform cannot support real-time requirements.
[18]	To develop a system that can collect and data while a in the air by means of aerial vehicles and long-distance transmission to offer new technical support for disaster management.	UAV Technology and mobile edge computing.	The study discovers the potential to combine UAV technology and Lora WAN communication into the implementation of MEC services for disaster management.	The UAVs have a higher energy consumption rate thus lower flying time. LoRa technology applied has delays thus may affect real time data relay and processing
[20]	The main objective of the paper is to provide a literature review of state-of-the-art machine learning algorithms for disaster and pandemic management	N/A	The study gets to discuss various ML algorithms, and models that can be used in different phases of disaster management to construct deployable models.	N/A
[22]	To review and choose a suitable model among different machine learning methods and algorithms, including artificial neural networks (ANN), support vector machine (SVM), multilayer perceptron (MLP), and long short-term memory (LSTM) which are used to forecast floods	Long short-term memory model of recurrent neural networks.	The study establishes that long short-term memory networks are better than ANN, MLP, and SVM because long short-term memory models can learn long-term patterns better	Lack of big data limits experimenting further on the model.
[23]	To compare neural network models used in hydrology data processing in flood prediction systems.	Multilayer perceptron (MLP) model	The research establishes that when compared to the multiple regression linear, MLP has better results in forecasting using hydrology data.	Lack of big data limits experimenting further on the model
[24]	To propose new approaches for spatial and temporal massive data analysis regarding natural disasters.	Neural network optimizers.	The resulting work provides a data collection model for disaster prediction, which could be utilized to collect climatic characteristics and topographical characteristics in with larger samples.	The model is unsuitable for less data.
[25]	To analyze new approaches for spatial and temporal massive data analysis with regards to natural disasters.	N/A	The paper proposes new approaches for data handling related to natural disasters	N/A
[28]	To apply a one-way coupled hydrometeorological modelling approach by using the WRF precipitation to drive a distributed hydrological model for flood simulation and forecasting.	Weather research and forecast model, Grid-Xinjiang model	The study demonstrates the importance of temporal and spatial patterns in predicting the timing and magnitude of incoming floods	The uncertainties of coupled flood forecasting approach based on the GXM model and WRF forecasts remain unestablished.

Table 1. *Cont.*

Research	The Objectives	Used Model	Result	Limitation
[27]	To construct a LSTM model used in prediction of precipitation.	LSTM model	The recurrent neural network models, combined with meteorological variables, could predict the precipitation accurately in Jingdezhen City and provide sufficient time to prepare strategies against potential related disasters.	The developed model only works with large data thus unsuitable for less data.
[29]	To analyze the nonlinear autoregressive network with exogenous inputs (NARX) model for a time series to evaluate the pattern of precipitation	Nonlinear NARX network model	The research concludes that the performance of the NARX model is viable to capture nonlinear behavior with a high value of correlation	The implemented NARX model needs further improvements by accommodating evolutionary and fractional metaheuristic algorithm.
[34]	Verification of SVM, KNN, Bayes, and deep neural networks for flood prediction accuracy and error rate.	SVM, KNN, and Naïve Bayes	The deep neural network can be used efficiently and accurately for flood prediction based on monsoon parameters only prior to flood occurrence	The dataset used in research considered old data, dating back to 1990 to 2002
[35]	Proving that machine learning techniques can accurately map and predict flood-prone areas and can be used to develop flood mitigation policies and plans	ANN and LR	We can quickly identify areas with high flood susceptibility using machine learning techniques and promote the development of policies and infrastructure that can reduce the potential effects of flooding.	The problem of availability of data between locations where flood historical, environmental, and hydrological data are scarce
[36]	Creating flood susceptibility maps by using ML models	RT, FDA, GLM, and MDA	Flood susceptibility areas have been identified in the study area	The model does not include the Internet of Things (IoT) to send early alerts.
[37]	ML models can predict the occurrence of urban pluvial flooding	ML models	Based on machine learning (ML) approaches, use a rainfall threshold to classify flood vs. non-flood events.	The model lacks using some techniques such as the Internet of Things for risk classification and sending early alerts to users and rescue teams

Briefly, by reviewing the previous literature on disaster management, especially floods, we have noticed that most of the solutions suffered from a few problems, which are:

- Delay in response to the event without attention to its threat level and real time (RT) notification. It is a major weakness in critical systems that are related to people’s lives, which are based on cloud computing and web services only.
- Most of the studies focused on the whole city in general instead of regions within the city. Flood effects may differ from one area to another according to the topology of each one.
- Papers that used the ML approach depended on the prediction the rainfall instead of RT classification for the current situation or were applied to a small dataset, which affects the reliability of results.
- Most of the studies depended on physical solutions based on hardware only, such as IoT devices. It is better to integrate both in addition to depending on more than one ML model, as we did in this research. This paper has addressed all the above issues as we explained and proved in the next section.

3. Methodology

The proposed framework is based on IoT and machine learning algorithms. Sensors will be distributed in the target region for data collection. The multiple sensor nodes deployed will form a wireless sensor network connected to fog nodes that connect to the cloud. Utilizing fog provides many advantages, such as dividing the cities to cells where each cell will be monitored independently. Moreover, fog will apply the roles and outputs of the proposed ML algorithms on the new data, which means reducing the load on the cloud in addition to fast detecting any threat without delay. The proposed framework and its layers, in addition to the proposed algorithms, will be discussed in detail by the following sub-sections.

3.1. Proposed Framework

It is an integrated and effective framework for early warning of disasters related to weather factors, especially floods resulting from rain or problems resulting from strong winds.

Figure 2 shows the general view of the proposed framework, including the main layers and objects. It shows the main four layers, which are the sensing layer, fog layer, cloud layer, and applications layer. In addition, Figure 2 includes icons that simulate objects of each layer. The first layer has many types of sensors, such as temperature, rainfall level, wind speed, etc. Moreover, the smart vehicle can be used as mobile sensors, and drones can be utilized also to collect information for locations that are difficult to use fixed sensors there. While Figure 3 depicts the detailed view and the functions of each layer. Weather disasters such as floods are related to many factors not limited to the amount of rain only. For example, the different topologies of area (low or high) have different degrees of threat, wind speed, humidity level, and temperature degree).

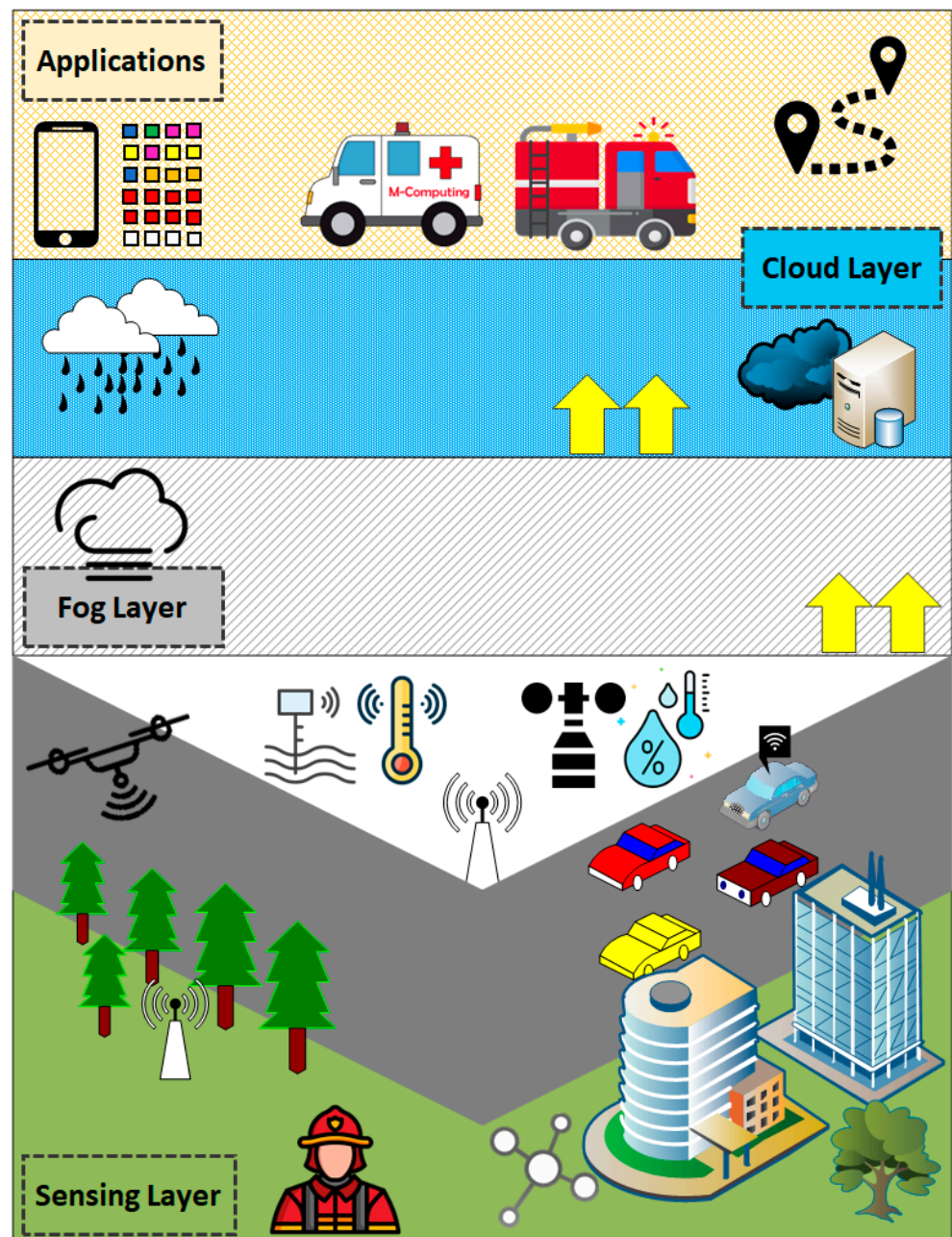


Figure 2. Proposed framework—general view.

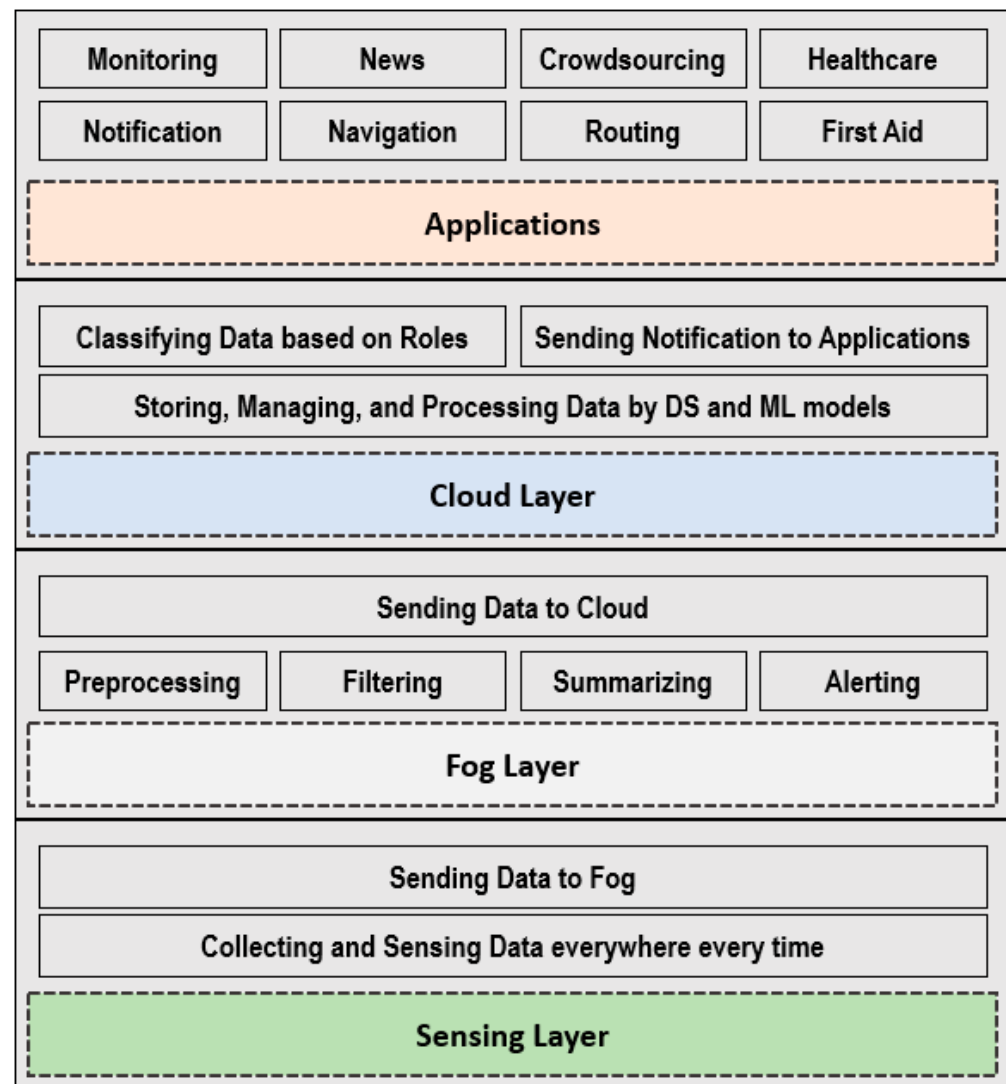


Figure 3. Proposed framework—functional view.

Therefore, several types of sensors were used in our perception layer (First Layer). A large number of sensors were deployed to cover the whole city. Sensors have limited resources, so the second layer in the proposed framework was fog computing. Sensors and IoT objects send their captured values to the fog periodically (e.g., every 6 s). Fog manages the flow of data on behalf of the cloud especially the fog node manages a small area. Thus, fog nodes will collect data from the sensing layer and apply some important functions which are:

- Caching data;
- Detecting abnormal data or sensor;
- Calculating the average of captured data from many sensors to enhance reliability and reduce the number of connections to the cloud;
- Normalizing values and creating inputs of ML model;
- Applying the ML model to classify the inputs data;
- Sending notifications in real time if there is any threat;
- Sending feedback and forwarding summarized data to the cloud every few minutes.

The third layer is the cloud. This layer receives data from the fog and stores it permanently. After a period, important historical data is created. Thus, the cloud applies the proposed ML algorithm to the historical data to form a smart classifier, in addition to important knowledge about the weather in each area.

Finally, the fourth layer is the applications one. In this layer, we proposed eight services and applications. However, it is available to add more services to any service provider. The eight applications are:

- Alerts application, which sends an immediate alert to citizens of a specific area in case of danger;
- News application, which sends the update of trusted information about the weather conditions or the status of the regions in the city;
- Routing application, which helps drivers to choose the best and safest route during the bad weather;
- Navigation application helps the civil defense to reach the affected or threatened places quickly;
- First aid application, which is useful for volunteers to know the procedures that must be in emergency cases and circumstances to help others and the civil defense groups;
- Crowdsourcing application allows people to evaluate the current situation and thus improve the accuracy of estimation cases in addition to results of the proposed algorithm;
- Health application, which enables users in bad situations to request ambulance or request online medical consultations if the movement is danger and forbidden;
- Monitoring application is useful to monitor the level of citizens' commitment to government instructions, especially in the cases of a high threat indicator.

3.2. Proposed Algorithm

This paper proposed a classification algorithm applied to historical data in the cloud. The results of algorithms will be as rules. The cloud sends these rules to the fog nodes to classify any new sensed data from the first layer. Thus, fog nodes will detect and notify any potential threat in RT. In the case of a threat, an alert is sent to the citizens to take precautions, and to the civil defense to prepare for any emergency. Moreover, the results from fog nodes will be used as feedback to the proposed algorithm in the cloud for continuous enhancement over time.

The proposed algorithm relied on periodic classification. This classification gives an early notification of the threat and then alerted people. Moreover, with the continuation of the threat indicator over time, the algorithms can estimate the level of threat. In addition, the proposed algorithm depended on the best three accuracies of the machine learning models that were tested. Where the three algorithms are adopted together, the majority is taken in the test case, which will enhance the reliability of classification.

Figure 4 explains the main steps of proposed algorithm in the case of training and testing. The steps are:

In the phase of training

- Reading dataset of historical data in cloud;
- Preprocessing data by filtering a nominal value;
- Removing the missing value or replaced by average value of the feature;
- Normalizing values of each feature;
- Converting the date value to month number;
- Selecting features depending on correlation value with threat column (Y);
- Applying SMOTE function to address the unbalanced data issue;
- Dividing data to training and testing (80, 20);
- Applying ML models (we selected five models RF, SVM, DT, KNN, and LR);
- Comparing results based on the confusion matrix;
- Selecting the best three models.

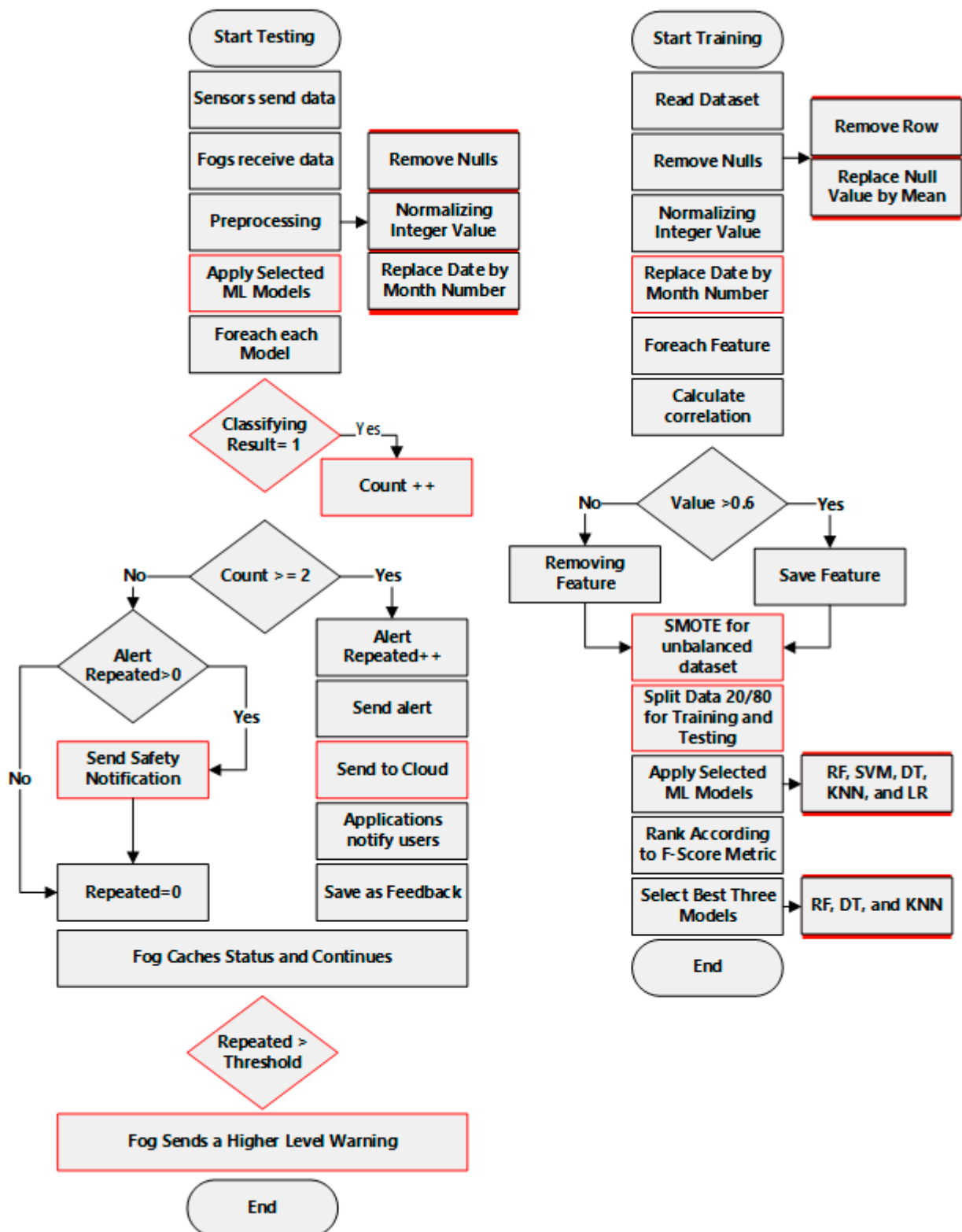


Figure 4. Proposed algorithm.

In the phase of testing

- Fog receives values from sensors;
- Fog preprocesses data and forms input of algorithm;
- Fog applies data on the proposed algorithm;

- Fog checks the majority of classification results for the best three models;
- In case of a threat, fog calculates time of continues threat to estimate the level of threat;
- Fog sends alert for people and teams of this area to take the precautions according to the threat level;
- Fog forwards data to the cloud to enable this data to services providers;
- Cloud creates complete imagination about the status of areas to support decisions of government.

Briefly, the main advantage of the proposed framework is that the alert will be in part of the city and not the entire city. That will relax the adverse impact of the disaster on the daily commercial, educational, and other activities of citizens within the city.

3.3. Other Advantages of the Proposed Framework Are

- Considering the topology of regions (cell) according to fog node location;
- Reducing the overhead on the cloud significantly, thus improving system performance;
- Providing real-time response and tracking for the emergency's cases;
- Detecting cases of outlier values of sensors by comparing their behavior with other;
- Depending on multi-models of ML to enhance the accuracy and reliability;
- Monitoring the level of threat depending on period of continuous and the rate of change;
- Suggesting most important services for dealing with the emergency situations.

3.4. Challenges of the Proposed Framework and Future Trend

- The cost of deploying wireless network sensors and providing fog nodes in all areas of the city, usually these things are available in smart cities;
- Interoperability between different kinds of sensors, companies, services, service providers, protocols, and data formats, which will affect the capability of integrating and processing. Thus, an effective solution has to be provided;
- Concern about preserving the privacy and security of users' data who collaborate or use the suggested services in the framework;
- Utilize Blockchain as an important solution for providing trusted data and avoiding rumors during crises [38];
- Providing smarter services can be utilized in a disaster situation, such as support awareness, encouraging volunteers, supporting disabled users, etc.

4. Implementation and Results

To test our proposed framework and algorithm, we tested it on real data that we have collected from the National Center of Meteorology and the Jeddah Municipality. We obtain daily weather data from Jeddah city for ten years (from 1 January 2009, to 31 December 2013) and (from 1 January 2018, to 31 December 2022). The dataset contains 3654 rows with eleven features are:

- DT (Date);
- Temperature (deg. c);
- Dry rate;
- Wet rate;
- Vapor rate;
- Humidity rate (%);
- Wind speed KTS/Deg;
- Pressure degree (hPa);
- Mean of sea level;
- Total rain amount (mm);
- Station level.

We received data as a PDF file and then we converted it to CSV to deal with it in python. Moreover, we discussed with experts and consultants of weather about the factors

that have an effect on the flood, then selected thresholds for some features, such as rain amount and temperature degree, in addition to classifying data. In addition, we excluded some unimportant features, such as dry rate, pressure, and sea level, which did not change during dates for the same city. Table 2 shows a subset (10 rows) of the sample data based on the features after initial processing.

Table 2. Shape of dataset of Jeddah city weather for 10 years (2009–2013 and 2018–2022) (e.g., of nine rows).

Date	Month (No)	Temperature (Deg. C) (M)	Humidity (%) (M)	Wind Speed KTS/Deg (M)	Sea Level Mean	Rain-Total Amount. (mm)	Class (Y)
25 November 2009	11	25.5	84	7	1010.8	70.000	1
25 December 2010	12	25.0	53	6	1013.8	33.000	1
18 July 2011	7	34.4	57	7	1002.4	0.001	0
3 November 2012	11	29.5	62	4	1010.4	34.200	1
23 February 2013	2	25.3	65	4	1011.5	0.001	0
17 April 2018	4	29.8	61	5	1008.2	38.000	1
8 May 2019	5	30.0	58	8	1006.8	0.001	0
3 June 2020	6	30.3	59	9	1005.2	0.001	0
3 July 2021	7	42	70	16	1001.1	0.001	0
3 August 2022	8	46	65	9	1000.6	0.001	0

4.1. Data Preparation and Implementation

We applied many steps. Firstly, we addressed the null values in some records. We replaced the null value with the average one. Then, we used standardized and normalizing functions for digital variables (temperature, relative humidity, wind velocity, and rain total amount). Normalization and standardization isolate the effect of the difference in ranges between values of variables (See Equations (1) and (2)). After that, we tested the coalition between each feature (variable) and classification column (Y). We excluded the variables with a correlation of less than 0.6 as sea level mean and humidity. This is logical because all data is related to the same city (Jeddah), so the sea level does not have any impact on records. Moreover, we customize our framework to deal with the small area (fog cells) instead of the city for considering the difference in areas' topology.

In addition, we converted the nominal values, including dates, to numeric, where we replaced the date with the month number because each month has different behavior in Saudi Arabia. Finally, we dealt with the imbalanced data issue, where there are 3652 records classified as "no flood" (0) compared to 92 records as 'flood' (1). Many real-world domains have an imbalanced data problem, and combating imbalanced data is critical because it will negatively affect the machine learning process and drive errors in classification or prediction [39]. We applied the SMOTE technique to solve the imbalanced issue.

The data was divided into 80% for training and 20% for testing, where a different 20% is chosen each time of testing to cover all data and enhance the results' reliability. Then, we selected five of the most popular ML models for classification targets. The selected models are Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), Rain Forest (RF), and Liner Regression (LR). Moreover, we added Deep Learning model (DL). We used Python programming language with the Google COLAB platform (Cloud Platform).

We applied the selected ML models to the final dataset and selected the best three models that gave the highest performance according to many metrics (accuracy, precision, recall, F-Score, and ROC curve). The best three models were DT, RF, and KNN. To enhance the accuracy and reliability of our algorithm, we depend on testing the actual value (input data formed by a fog) on these three models together. That means, the final result will depend on the majority of classification results of the three models. For example, if two

of them give 1 and the third gives 0, then the result will be 1. In addition, we considered threats classification early before they become a disaster.

$$\text{Standardization NewX} = \frac{Xi - \text{Mean}(X)}{\text{Standard Deviation}} \quad (1)$$

$$\text{Normalization NewX} = \frac{Xi - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)} \quad (2)$$

For comparison and evaluation between the models, we used standard metrics, which are confusion matrix and ROC curve (ROC), in addition to learning curve. The confusion matrix helps to calculate four important metrics which are:

- Accuracy: is a popular machine learning model validation method used in evaluating classification problems and it is the rate of correct predictions over all test cases. For binary classification, as in our study, accuracy can also be calculated through pros and cons. However, when the class distribution is imbalanced, overall accuracy in machine learning classification models can be misleading, and correctly predicting the minority class is critical. In this case, the majority class may be correctly predicted, resulting in a high accuracy score, while the minority class is misclassified. This creates the false impression that the model is performing well when in fact it is not.
- Precision: It is called the Precision of Positive Observations because it represents the ratio of expected correct positive observations to expected total positive observations. It aids in measuring the model's ability to classify positive samples. In terms of precision when calculating the precision of a model, we should take into account both positive and negative samples that are classified. A machine learning model's precision is affected by both negative and positive samples.
- Recall (sensitivity): It is the ratio of correctly expected positive feedback to all actual observations (not just the positive). It allows us to determine how many positive samples the ML model correctly classified. When calculating a model's recall, we only need all positive samples and ignore all negative samples. The recall is concerned with correctly categorizing all positive samples.
- F1-Score: It is the weighted average of precision and recall. Therefore, this result considers bogus pros and false cons and is better than the accuracy criterion when the distribution is unequal for the categories. It is also useful when false positives and false negatives have the same cost to the system. Otherwise, each criterion (precision and recall) is best viewed separately.

The area under the ROC (AUC) measures the area underneath the entire ROC curve based on the true positive rate (TPR) and false positive rate (FPR), where TPR is equal to the value of sensitivity or recall, and FPR is equal to 1 minus the specificity value.

4.2. Results

We applied our algorithm to the data set of the first five years (2009–2013), then we extended our data to confirm the stability of the results by adding data from the other five years (2018–2022). Figure 5 shows the performance of the different models (KNN, SVM, RF, DF, Log, and DL) on a dataset of five years (2009–2013). The RF and DT are superior in most of the parameters. DL also achieved a good result, but not better than RF; DL took more than 42 s for training while RF took less than 1 s.

To confirm that, Figure 6 depicts the performance of the same models but in 10 years (2009–2013 and 2018–2022). The results here showed enhancement of DL results; however, it still not better than RF, which preserved the best performance and confirmed its stability. Of note, the DL took more than 120 s for the 8 years of data, while RF took less than 1 s.

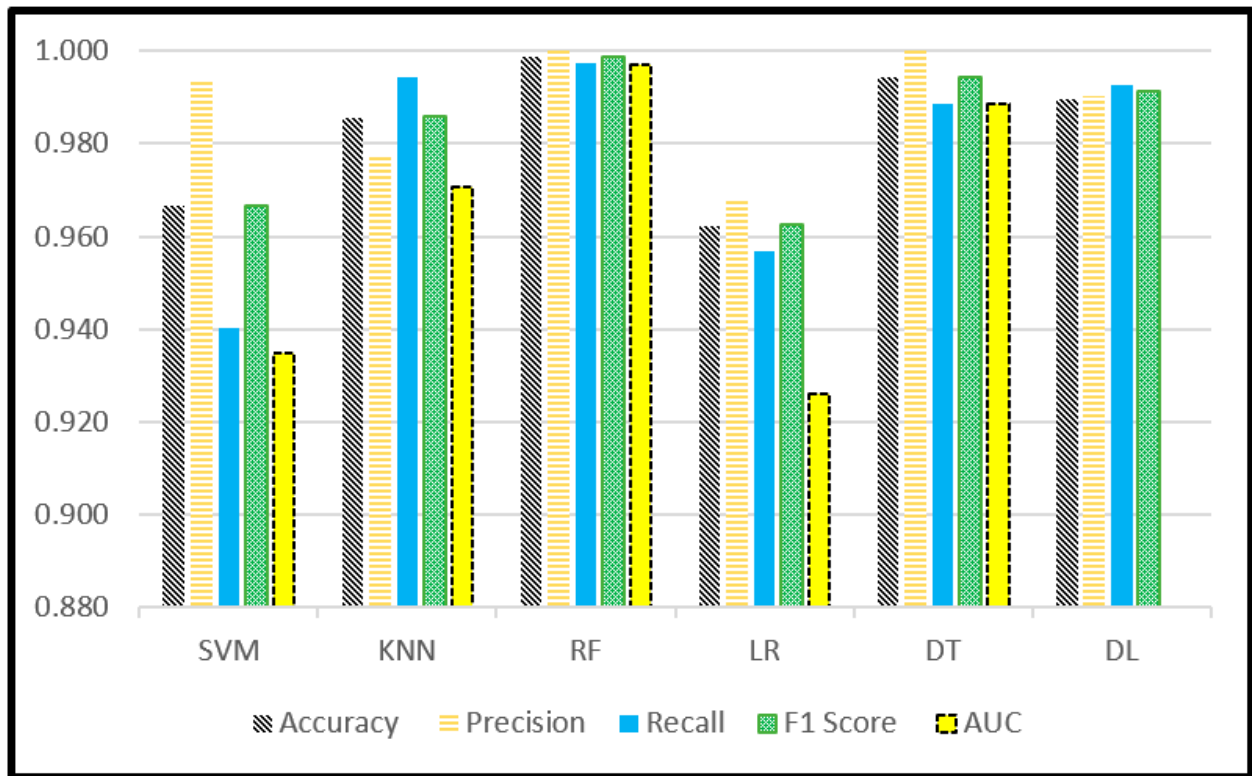


Figure 5. Comparison among the selected models of all criteria (Data 2009–2013).

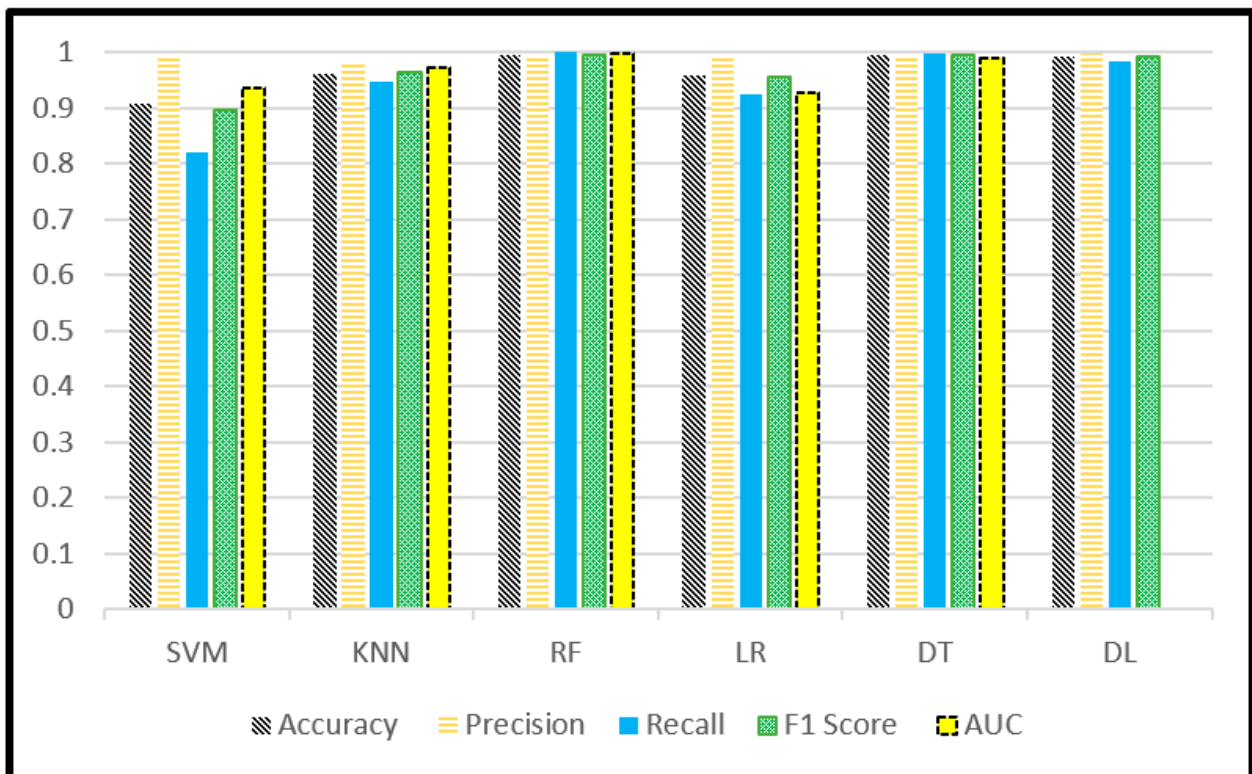


Figure 6. Comparison among the selected models of all criteria (data 2009–2013 and 2018–2022).

From Figure 5, the RF achieved the highest accuracy, “99%”, and a maximum precision of “1”. This means the margin of error is zero, and the probability of classified threat as

not threat is zero also. That is very important in sensitive or critical systems. For the recall value, RF achieved 99%, which means there is a very small margin for error classification for the non-threat case as a threat. There is no effect in this case in the proposed framework. Because the framework repeats the testing process periodically to enhance the accuracy and to find the level of threat in real-time.

The F-score is the common factor that expresses recall and precision together. The results confirmed the superiority of RF and DT with KNN over the other models. Finally, the AUC and ROC present the significant superiority of RF, which is close to the maximum area, as shown in Figure 7. Tables 3 and 4 present the actual metrics values of each model for both datasets (Dataset 1 2009–2013, Dataset 2 2009–2013 and 2018–2022). The time of training per second is added to the tables.

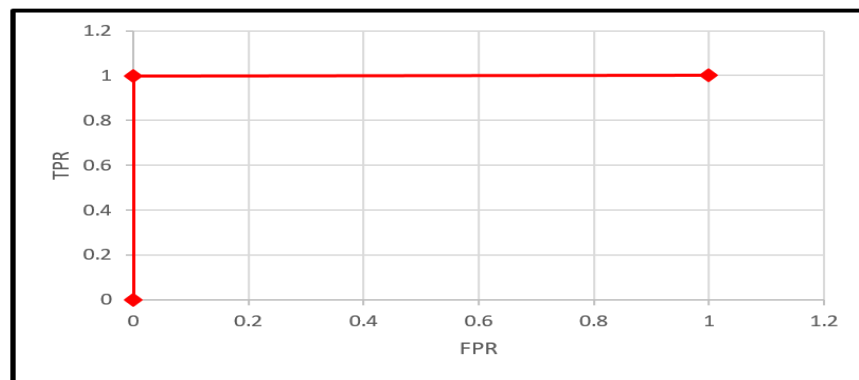


Figure 7. ROC curve for the RF model.

Table 3. Results of performance the different models for data 2009–2014.

ML Models	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score	Time
SVM	331	334	2	21	0.967	0.994	0.940	0.966	>1 s
KNN	350	328	8	2	0.985	0.978	0.994	0.986	>1 s
RF	352	335	0	1	0.999	1.000	0.997	0.999	>1 s
LR	333	329	11	15	0.962	0.968	0.957	0.962	>1 s
DT	348	336	0	4	0.994	1.000	0.989	0.994	>1 s
DL	2699	1690	27	20	0.989	0.990	0.993	0.991	42 s

Table 4. Results of performance the different models for data 2009–2013 and 2018–2022.

ML Models	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score	Time
SVM	565	697	6	125	0.906	0.989	0.819	0.896	>1 s
KNN	675	666	16	37	0.962	0.977	0.948	0.962	>1 s
RF	691	697	6	0	0.996	0.991	1.000	0.996	>1 s
LR	638	697	6	53	0.958	0.991	0.923	0.956	>1 s
DT	689	698	5	2	0.995	0.993	0.997	0.995	>1 s
DL	3450	3457	6	55	0.991	0.998	0.984	0.991	160 s

Finally, Figure 8 depicts the learning curve of each classification model. This figure proved the superiority of RF where the curve toward 1 and the gap between validation and training curve minimum. In addition, DL, KNN, and DT provided very good performance.

As a justification for the superiority of RF, usually, DT and RF achieve higher performance than other models if the variables of the dataset are categorical or its values have a categorical tendency. That means that the values of the variables have certain thresholds that affect the result, in addition, the values of variables do not have complex relations. That returns to the nature of the weather in KSA. Thus, the DT and RF are suitable for this type of data. Moreover, the RF is used to achieve more stability if the dataset is changed, and

that is exactly what is proven by extending the dataset by an additional five years of data. In addition, the DL can achieve the best results if we provide more and more data. For that, the accuracy of DL is enhanced when we extend data from 5 years to 10 years. However, on another side, the DL required more time for updating the model (in the training phase) (see Tables 3 and 4).

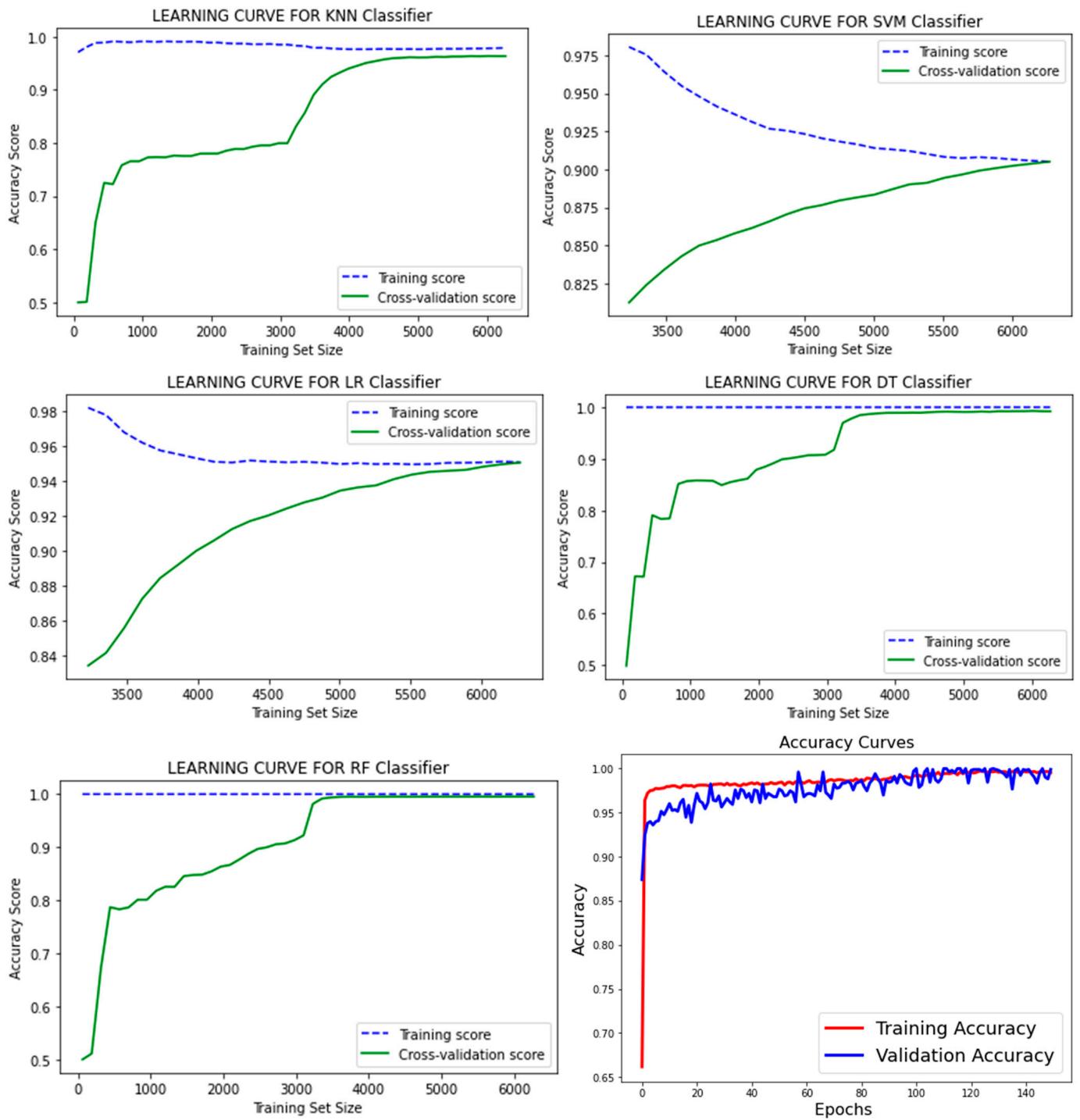


Figure 8. Learning curve of accuracy achievement of all ML models and the DL one.

5. Conclusions

This research presented an effective framework for managing weather disasters, specifically floods. The proposed framework is based on hybrid method by integration between the IoT, which enables the collection of necessary data in real-time and ML. Relying on fog computing helped in processing the sensed data without delay, immediately responding to any emergency change, and reducing the burden on the cloud. The cloud provided historical data from previous experiments to apply a proposed algorithm for early classification of the threat's degree or the possibility of flooding happening. The proposed algorithm is based on three popular machine learning models, which are RF, DT, and KNN, to increase the level of reliability. The accuracy of these models exceeded the 99% threshold. DL also provide good performance but with a long time for training. In the last layer, the proposed framework presented a group of important services for disaster management; however, the details of these applications will be in the next study. Finally, the research presented the challenges that need future solutions, which are related to the security and privacy of users' data, in addition to supporting interoperability. Moreover, in some critical applications, research has to use continuous classification instead of prediction, which will provide a more accurate model.

Author Contributions: Conceptualization, F.H.A. and A.A.A.S.; Methodology, F.H.A. and A.A.A.S.; Software, F.H.A. and A.A.A.S.; Validation, F.H.A. and A.A.A.S.; Formal analysis, F.H.A.; Investigation, B.A.; Resources, F.H.A.; Data curation, F.H.A.; Writing—original draft, F.H.A. and A.A.A.S.; Writing—review & editing, A.A.A.S., M.S.R., B.A. and N.M.B.; Supervision, M.S.R. and B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Exclude this statement.

Informed Consent Statement: Not applicable.

Data Availability Statement: This link for dataset: <https://github.com/adnanmm/Flood-Dataset>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Al-Barazanchi, I.; Jaaz, Z.A.; Abbas, H.H.; Abdulshaheed, H.R. Practical application of IOT and its implications on the existing software. In Proceedings of the 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI), Yogyakarta, Indonesia, 1–2 October 2020; IEEE: New York, NY, USA, 2020; pp. 10–14.
2. Mezzanotte, P.; Palazzi, V.; Alimenti, F.; Roselli, L. Innovative RFID sensors for internet of things applications. *IEEE J. Microw.* **2021**, *1*, 55–65. [[CrossRef](#)]
3. Sen, A.A.A.; Eassa, F.A.; Jambi, K.; Yamin, M. Preserving privacy in internet of things: A survey. *Int. J. Inf. Technol.* **2018**, *10*, 189–200. [[CrossRef](#)]
4. Bahbouh, N.; Basahel, A.; Sendra, S.; Sen, A.; Ahmed, A. Tokens shuffling approach for privacy, security, and reliability in IoHT under a pandemic. *Appl. Sci.* **2023**, *13*, 114. [[CrossRef](#)]
5. Wang, Y.; Chen, X.; Wang, L.; Min, G. Effective IoT-facilitated storm surge flood modeling based on deep reinforcement learning. *IEEE Internet Things J.* **2020**, *7*, 6338–6347. [[CrossRef](#)]
6. Huang, G.; Shen, Z.; Mardin, R. (Eds.) *Overview of Urban Planning and Water-Related Disaster Management*; Springer International Publishing: Cham, Switzerland, 2018. [[CrossRef](#)]
7. Jung, D.; Tuan, V.T.; Tran, D.Q.; Park, M.; Park, S. Conceptual framework of an intelligent decision support system for smart city disaster management. *Appl. Sci.* **2020**, *10*, 666. [[CrossRef](#)]
8. Nsabagwa, M.; Byamukama, M.; Kondela, E.; Otim, J.S. Towards a robust and affordable automatic weather station. *Dev. Eng.* **2018**, *4*, 100040. [[CrossRef](#)]
9. Mahdavinejad, M.S.; Rezvan, M.; Barekatin, M.; Adibi, P.; Barnaghi, P.; Sheth, A.P. Machine learning for internet of things data analysis: A survey. *Digit. Commun. Netw.* **2018**, *4*, 161–175. [[CrossRef](#)]
10. Khanh, Q.V.; Hoai, N.V.; Manh, L.D.; Le, A.N.; Jeon, G. Wireless communication technologies for IoT in 5G: Vision, applications, and challenges. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 1–12. [[CrossRef](#)]
11. Puliafito, C.; Mingozzi, E.; Longo, F.; Puliafito, A.; Rana, O. Fog computing for the internet of things: A survey. *ACM Trans. Internet Technol. (TOIT)* **2019**, *19*, 1–41. [[CrossRef](#)]

12. Sen, A.A.A.; Yamin, M. Advantages of using fog in IoT applications. *Int. J. Inf. Technol.* **2021**, *13*, 829–837. [CrossRef]
13. Beltran, A., Jr.; Dizon, K.J.; Nones, K.; Salanguit, R.L.; Santos, J.B.; Santos, J.R. Arduino-based disaster management system. *J. Robot. Control. (JRC)* **2021**, *2*, 24–28. [CrossRef]
14. Yamin, M.; Basahel, A.M.; Abi Sen, A.A. Managing crowds with wireless and mobile technologies. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 7361597. [CrossRef]
15. Dano, U.L. Flash flood impact assessment in Jeddah City: An analytic hierarchy process approach. *Hydrology* **2020**, *7*, 10. [CrossRef]
16. Adeel, A.; Gogate, M.; Farooq, S.; Ieracitano, C.; Dashtipour, K.; Larijani, H.; Hussain, A. A survey on the role of wireless sensor networks and IoT in disaster management. *Geol. Disaster Monit. Based Sens. Netw.* **2018**, *3*, 57–66. [CrossRef]
17. Sciullo, L.; Trotta, A.; Di Felice, M. Design and performance evaluation of a LoRa-based mobile emergency management system (LOCATE). *Ad Hoc Netw.* **2020**, *96*, 101993. [CrossRef]
18. Xu, J.; Ota, K.; Dong, M. Big data on the fly: UAV-mounted mobile edge computing for disaster management. *IEEE Trans. Netw. Sci. Eng.* **2020**, *7*, 2620–2630. [CrossRef]
19. Ejaz, W.; Azam, M.A.; Saadat, S.; Iqbal, F.; Hanan, A. Unmanned aerial vehicles enabled IoT platform for disaster management. *Energies* **2019**, *12*, 2706. [CrossRef]
20. Chamola, V.; Hassija, V.; Gupta, S.; Goyal, A.; Guizani, M.; Sikdar, B. Disaster and pandemic management using machine learning: A survey. *IEEE Internet Things J.* **2020**, *8*, 16047–16071. [CrossRef] [PubMed]
21. Arinta, R.R.; Andi, E.W.R. Natural disaster application on big data and machine learning: A review. In Proceedings of the 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 20–21 November 2019; Volume 6, pp. 249–254. [CrossRef]
22. Ranasinghe, R.M.; Ilmini, W.M.K.S. Introducing a LSTM Based Flood Forecasting Model for the Nilwala River Basin with a Mobile Application—A Review. 2020. Available online: <http://ir.kdu.ac.lk/handle/345/3039> (accessed on 1 January 2023).
23. Widiasari, I.R.; Nugroho, L.E. Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. In Proceedings of the 2017 International Conference on Innovative and Creative Information Technology (ICITech), Salatiga, Indonesia, 2–4 November 2017; IEEE: New York, NY, USA, 2017.
24. Halgamuge, M.N.; Daminda, E.; Nirmalathas, A. Best optimizer selection for predicting bushfire occurrences using deep learning. *Nat. Hazards* **2020**, *103*, 845–860. [CrossRef]
25. Martínez-Álvarez, F.; Morales-Esteban, A. Big data and natural disasters: New approaches for spatial and temporal massive data analysis. *Comput. Geosci.* **2019**, *129*, 38–39. [CrossRef]
26. Nourani, V.; Uzelaltinbulat, S.; Sadikoglu, F.; Behfar, N. Artificial intelligence based ensemble modeling for multi-station prediction of precipitation. *Atmosphere* **2019**, *10*, 80. [CrossRef]
27. Kang, J.; Wang, H.; Yuan, F.; Wang, Z.; Huang, J.; Qiu, T. Prediction of precipitation based on recurrent neural networks in Jingdezhen, Jiangxi Province, China. *Atmosphere* **2020**, *11*, 246. [CrossRef]
28. Yao, C.; Ye, J.; He, Z.; Bastola, S.; Zhang, K.; Li, Z. Evaluation of flood prediction capability of the distributed Grid-Xinanjiang model driven by weather research and forecasting precipitation. *J. Flood Risk Manag.* **2019**, *12*, 12544. [CrossRef]
29. Bukhari, A.H.; Sulaiman, M.; Islam, S.; Shoaib, M.; Kumam, P.; Raja, M.A.Z. Neuro-fuzzy modeling and prediction of summer precipitation with application to different meteorological stations. *Alex. Eng. J.* **2019**, *59*, 101–116. [CrossRef]
30. Samikwa, E.; Thiemo, V.; Joakim, E. Flood Prediction Using IoT and Artificial Neural Networks with Edge Computing. In Proceedings of the 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), Rhodes, Greece, 2–6 November 2020; IEEE: New York, NY, USA, 2020.
31. Hassan, W.H.W.; Jidin, A.Z.; Aziz, S.A.C.; Rahim, N. Flood disaster indicator of water level monitoring system. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 1694. [CrossRef]
32. Esposito, M.; Palma, L.; Belli, A.; Sabbatini, L.; Pierleoni, P. Recent advances in internet of things solutions for early warning systems: A review. *Sensors* **2022**, *22*, 2124. [CrossRef]
33. Acosta-Coll, M.; Solano-Escorcia, A.; Ortega-Gonzalez, L.; Zamora-Musa, R. Forecasting and communication key elements for low-cost fluvial flooding early warning system in urban areas. *Int. J. Electr. Comput. Eng. (IJECE)* **2021**, *11*, 4143–4156. [CrossRef]
34. Sankaranarayanan, S.; Prabhakar, M.; Satish, S.; Jain, P.; Ramprasad, A.; Krishnan, A. Flood prediction based on weather parameters using deep learning. *J. Water Clim. Chang.* **2019**, *11*, 1766–1783. [CrossRef]
35. Ighile, E.H.; Shirakawa, H.; Tanikawa, H. Application of GIS and machine learning to predict flood areas in Nigeria. *Sustainability* **2022**, *14*, 5039. [CrossRef]
36. El-Haddad, B.A.; Youssef, A.M.; Pourghasemi, H.R.; Pradhan, B.; El-Shater, A.H.; El-Khashab, M.H. Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. *Nat. Hazards* **2021**, *105*, 83–114. [CrossRef]
37. Ke, Q.; Tian, X.; Bricker, J.; Tian, Z.; Guan, G.; Cai, H.; Huang, X.; Yang, H.; Liu, J. Urban pluvial flooding prediction by machine learning approaches—A case study of Shenzhen city, China. *Adv. Water Resour.* **2020**, *145*, 103719. [CrossRef]

38. Waseem, M.; Khan, M.A.; Goudarzi, A.; Fahad, S.; Sajjad, I.A.; Siano, P. Incorporation of blockchain technology for different smart grid applications: Architecture, prospects, and challenges. *Energies* **2023**, *16*, 820. [[CrossRef](#)]
39. Ali, H.; Salleh, M.N.M.; Saedudin, R.; Hussain, K.; Mushtaq, M.F. Imbalance class problems in data mining: A review. *Indones. J. Electr. Eng. Comput. Sci.* **2019**, *14*, 1560–1571. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.