Energy efficient IoT-based cloud framework for early flood prediction

Mandeep Kaur1 · Pankaj Deep Kaur1 · Sandeep Kumar Sood2

Received: 21 January 2021 / Accepted: 28 June 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
Flood is a recurrent and crucial natural phenomenon affecting almost the entire planet. It is a critical problem that causes crop destruction, destruction to the population, loss of infrastructure, and demolition of several public utilities. An effective way to deal with this is to alert the community from incoming inundation and provide ample time to evacuate and protect property. In this article, we suggest an IoT-based energy-efficient flood prediction and forecasting system. IoT sensor nodes are constrained in battery and memory, so the fog layer uses an energy-saving approach based on data heterogeneity to preserve the system’s power consumption. Cloud storage is used for efficient storage. The environmental conditions such as temperature, humidity, rainfall, and water body parameters, i.e., water flow and water level, are being investigated for India’s Kerala region to calibrate the flood phases. PCA (Principal Component Analysis) approach is used at the fog layer for attribute dimensionality reduction. ANN (Artificial Neural Network) algorithm is used to predict the flood, and the simulation technique of Holt Winter is used to forecast the future flood. Data are obtained from the Indian government meteorological database, and experimental assessment is carried out. The findings showed the feasibility of the proposed architecture.

Keywords Internet of Things · ANOVA · Tukey post hoc test · Holt Winter · ANN · Flood

Mandeep Kaur
bajwamandeep96@gmail.com

Pankaj Deep Kaur
pankajdeepkaur@gmail.com

Sandeep Kumar Sood
san1198@gmail.com

1 Department of Computer Science and Engineering, Guru Nanak Dev University Regional Campus Jalandhar, Punjab 144007, India
2 Department of Computer Applications, National Institute of Technology, Kurukshetra, Haryana 136119, India

Published online: 13 July 2021
1 Introduction

Natural disasters are universal incidents and require significant assistance to tackle. Natural calamities such as floods, earthquakes, and hurricanes are events that intensely impact wide zone, distressing populations, affecting goods, and tremble the people on both economic as well as psychological viewpoints (Avvenuti et al. 2017). From these disasters, flood is one of the ordinary disastrous incidents in various countries consistently around the world. Over the past two decades, floods have accounted for 44% of all disaster incidents and have affected 1.6 billion people worldwide (UNDRR 2020). China is the nation most impacted by floods from 2000 to 2019, with an average of 20 floods every year and 900 million people affected (UNDRR 2020). In 2017, floods in the US caused 3019 fatalities and resulted in an overall loss of US$ 95,000 million (https://natcatservice.munichre.com). In India recently, in Gujrat, the floods devastated 6.44 lakh farmers. The cost of crop destruction is expected to be Rs. 867 crore. Assam deals with the displacement of 61,923 people and approximately 160 deaths. In West Bengal, 1.67 lakh citizens were forced to seek refuge in emergency shelters (www.newsclick.in), resulting in part from the lack of development of warning systems and information at the community level of the imminent flood. There is needed to take special procedures to predict it and to manage the situation before it occurs. Flood prediction and detection are done using IoT sensors and cloud computing in a geographical area by providing proficient acquisition, processing, and efficient storage of flood-related information. IoT and cloud computing are in combination become a powerful platform for flood management by monitoring the water bodies from remote sites providing continuous information about flood to decision-making agencies and rescue units.

Flood forecasts and warnings are intended to decrease the risk of loss of life and economic influence. A framework for flood alert provides data accumulation, analysis of data, scrutinize and warning (Sakib et al. 2016). The presented model is based on forecasting and detecting the flood in advance based on environmental factors like temperature, humidity, rainfall, and hydrological parameters like water flow and water level. Nonlinear data are dimensionally reduced by using principal component analysis (PCA) dimension reduction method. In this study, principal component analysis is proposed as a novel pre-processing technique for the flood detection systems to reduce the dimensions of flood-related attributes, and the resulting input representation is trained with artificial neural networks (ANN) for classifying the data. Artificial neural network is a nonlinear computational structure comprising many interconnected processing units (Tur et al. 2010). ANN is used because of its characteristics like high parallelism, fault tolerance, learning, and generalization capabilities. In this study of flood prediction domain, IoT framework uses a comprehensive historical dataset of continuous observations over a span of time to predict the flood and related activities.

1.1 Contributions

The paper contributes to this aspect by

- Using IoT sensors to monitor the study area for various flood-causing attributes.
- Efficient utilization of hardware by employing an energy-conserving mechanism at Fog Layer.
• Maximize network capacity consumption by applying a dimensionality reduction mechanism at Fog Layer.
• Cloud computing assisted flood prediction and forecasting using precise approaches.
• Alerts generation to the water resource and disaster response organizations if a region is particularly vulnerable to floods.
• Sharing information and transmission of evaluated results to the concerned agencies (Water Resource Agencies and CDEOC).

In the proposed system, the general framework of IoT-based flood prediction and forecasting system is described. A survey on various flood monitoring and management systems focused on IoT with different data mining methodologies has been addressed in Sect. 2. Sect. 3 describes key terms relevant to our proposed flood data analysis model and computing framework with alert generation process. In Sect. 4, a complete experimental comparison of the proposed approach with different techniques is described. Comparative analysis and innovation of research are presented in Sect. 5. Section 6 discusses the proposed system’s limitations. Finally, Sect. 7 brings the study to a conclusion with some important discussions.

2 Related work

This section analyzes various flood management systems and data mining techniques used for flood data. Firstly, different frameworks have been discussed related to flood management systems, followed by data mining techniques.

2.1 Flood management system

In Ray and Mukherjee (2017) presented the model based on IoT to experience the significant issues with disasters, for example, remote monitoring and real-time analysis, cautioning, data analysis, information accumulation. A complete dialog is presented on state-of-the-art situations to deal with devastating incidents. Moreover, IoT-based guidelines and market-prepared deployable products were reviewed to tackle disaster problems. It is concluded that the Internet of Things-based technologies are suitable for disaster management and handling disastrous situations. This survey reveals key challenges and research patterns in IoT-enabled emergency response systems. In Afzaal and Zafar (2016) proposed a model in which sensors were installed to observe water levels in different water bodies.

Gateways are used as an interface with the cloud and enable actors based on information handled by the cloud. The Vienna Development Method-Specification Language (VDM-SL) is used for the configuration and execution of the system. VDM-SL is utilized to create potential test scenarios to minimize device failures and omissions. The result shows that there is no error in the specification. In Hernández-Nolasco et al. (2016) offer a sensor for calculating water levels in rivers, reservoirs, lagoons, and streams. A prototype is designed using a micro-model installed on a basic open circuit based on a water level measurement sensor. Netduino Plus 2 is used to perform the micro-model. In Kumar et al. (2015) presented an innovative approach to India’s recurrent issue of floods.

The system is an integrated flood prediction, and alert generation system developed using Internet of Things techniques. The system uses an innovative approach to calculate and monitor several flood-related parameters at different locations to forecast river flooding.
in real-time reliably. In Yusoff et al. (2015) suggested that flood control and early warning monitoring can be tackled by cloud computing. The study confirms that GreenCloud supports crucial functions for the growth of smart cities. In Lo et al. (2015) presented an image processing techniques-based model to determine the flood conditions. The experimental results indicate the reliability of the visual sensing approach. In Hadid et al. (2020) presented a method for predicting river stream levels using the hybrid model and Dempster–Shafer algorithm for PWARX (Piecewise Auto-Regressive eXogenous) model.

2.2 Meteorological data analytics

Several techniques in hydrology and water resource studies had been described in recent decades. The techniques such as genetic algorithm, ANFIS (Adaptive Neural based Fuzzy Inference System) (Hong et al. 2018), ARIMA (Razak et al. 2018), SARIMA (Machekposhti et al. 2018), OLAP (Ma et al. 2010), ANN (Goymann et al. 2019), dimension reduction techniques (Widmann and Schar 1997) enable users to arrange the enormous amount of hydro-meteorological data. A variety of data analytical tools are used to evaluate and optimize different techniques used in flood prediction, forecasting, and management. ArcGIS (Gilandeh et al. 2020) is used to interpret maps and compile geospatial information. HEC-RAS (Ongdas et al. 2020), HEC-HMS (Romali et al. 2018), WMS (Hoseini et al. 2016) are flood mapping and modeling software.

In Zhu et al. (2019) conduct a survey on various forecasting techniques and their applications in multiple domains. In Owotoki et al. (2006) proposed a model for integrated flood management (IFM) that employs the data mining techniques in a three-tier web-based framework dedicated to sustainable development for stakeholders. In 2003, Marhaba et al. proposed a principal component analysis-based system to monitor water quality’s spatial and temporal variations. In Widmann and Schar (1997) presented a method for analyzing daily precipitation in Switzerland using principal component analysis technique. A mathematical model is developed to observing precipitation patterns due to shifts in the frequency or precipitation behavior of Alpine weather groups.

In Rehman et al. (2019) presented various flood vulnerability assessment methods, including ANN. In Aziz et al. (2013) described artificial neural network implementation for regional flood inundation mapping in the Australian case study. In Wu et al. (2010) applied modular ANN technique to forecast the rainfall time series. In Chu et al. (2018) proposed a modified principal component analysis (MPCA) method for assessing environmental variables to track ecological changes in coastal recovery areas. In Ghadim et al. (2018) discuss using the Holt-Winters time series model’s additive and multiplicative types to forecast environmental variables one year in advance.

3 Proposed model

Figure 1 explains the proposed model for flood prediction and forecasting. It consists of three layers: data acquisition layer, fog and cloud. The data acquisition layer gathers environmental data from several sensors such as rainfall sensors, temperature sensors, water flow sensors, humidity sensors, water level sensors at different locations and water bodies. The data collected are analyzed at the fog layer to examine the data variation to adapt the sampling frequency of the sensor nodes. The data dimensionality is decreased further by doing the principal component analysis (PCA) on the fog layer and forwarding it to the
cloud layer. Data are maintained in a cloud-based repository from which valuable information is extracted for efficient processing and effective decision-making.

### 3.1 Data acquisition layer

The data acquisition layer gathers a large amount of data. IoT sensor nodes are responsible for collecting data on flood events and related parameters in the local area. Successful flood prediction and forecasting focus on information on the various meteorological and hydrological attributes that cause flooding. The overview of flood attributes and accompanying sensors is shown in Table 1.
1. Meteorological attributes: Flood is greatly dependent on meteorological conditions of a particular location. Meteorological attributes comprise information about temperature, humidity, precipitation, and monsoon season significantly escalate the occurrence of maximum rainfall that causes flood conditions.

2. Hydrological attributes: Water level in water bodies of the study area and water flow regarding time is considered an essential factor for flood conditions.

### 3.2 Fog layer

The fog layer is the transitional layer between cloud and the data acquisition layer. The raw data obtained from the data acquisition layer are preprocessed at this layer. The fog layer performs a two-step preprocessing of data such as ANOVA and Tukey post hoc tests to save energy and PCA-based data dimensional reduction. Table 2 describes various notations used in the fog layer.

#### 3.2.1 ANOVA model and Tukey’s post hoc test

Fog layer receives datasets information from the data acquisition layer. Batteries operate the sensors nodes, and power restricted sensors require an energy-efficient data collection strategy. A method for measuring the active and sleep duration based on sensor data has been presented in the proposed model. To avoid redundant information, sensor active and sleep durations of sensors are analyzed using the ANOVA (Analysis of Variance) method with Tukey Test. The one-way ANOVA model is used to measure the total heterogeneity \( \mathbb{H}_t \) of data generated by sensor nodes in N time slots. Total heterogeneity \( \mathbb{H}_t \) is calculated as the measure of the heterogeneity within duration \( \mathbb{H}_{\text{within}} \) and heterogeneity between duration \( \mathbb{H}_{\text{btw}} \), illustrated as:

\[
(\mathbb{H}_t) = (\mathbb{H}_{\text{within}}) + (\mathbb{H}_{\text{btw}})
\]

Each sensor node captures a new data value for every flood attribute:

| Table 1 Flood causing attributes |
|----------------------------------|
| Meteorological attributes       | Description                        | Sensors                        |
| Temperature                      | Atmospheric air temperature         | Temperature sensors, humidity sensors |
| Humidity                        | Water vapors in air                 |                                |
| Rainfall                        | Amount of precipitation             | Rain gauges, precipitation sensors |
| Season                          | Winter, pre-monsoon, monsoon, post-monsoon |                                |
| Hydrological attributes         | Description                        | Sensors                        |
| Water level                     | Level of water in water bodies      | Water level sensors            |
| Water flow                      | Volume of water flowing             | Water flow sensors             |
Here, \( f_{ac} \) is \( a \)th observation taken by sensor node in \( c \)th duration; \( n_c \) indicates the number of observations in \( c \)th duration; \( N \) indicates a total number of durations; Mean\(_c\) denotes mean of data values captured in \( c \)th duration; Mean denotes mean of data values captured across all \( N \) time durations. The one-way ANOVA findings aid to assess whether or not the means of separate datasets obtained over subsequent time durations differ significantly. Further, the Tukey post hoc test is implemented to assess whether the variance between data values from distinct durations surpasses a particular threshold \( \tau_{th} \). An update accompanies this on the active and sleep duration of the sensor node to make more efficient use of their energy.
\[ \theta_i = \sum_{a=1}^{n_1} f_{1a}^2 + \sum_{a=1}^{n_2} f_{2a}^2 + \sum_{a=1}^{n_3} f_{3a}^2 + \cdots + \sum_{a=1}^{n_N} f_{Na}^2 - \frac{\left(\sum_{a=1}^{n_1} f_{1a}^2 + \sum_{a=1}^{n_2} f_{2a}^2 + \sum_{a=1}^{n_3} f_{3a}^2 + \cdots + \sum_{a=1}^{n_N} f_{Na}^2\right)^2}{n_N} \] (4)

\[ \theta_{bps} = \frac{\left(\sum_{a=1}^{n_1} f_{1a}^2\right)^2}{n_1} + \frac{\left(\sum_{a=1}^{n_2} f_{2a}^2\right)^2}{n_2} + \frac{\left(\sum_{a=1}^{n_3} f_{3a}^2\right)^2}{n_3} + \cdots + \frac{\left(\sum_{a=1}^{n_N} f_{Na}^2\right)^2}{n_N} - \frac{\left(\sum_{a=1}^{n_1} f_{1a}^2 + \sum_{a=1}^{n_2} f_{2a}^2 + \sum_{a=1}^{n_3} f_{3a}^2 + \cdots + \sum_{a=1}^{n_N} f_{Na}^2\right)^2}{n_N} \] (5)

\[ \theta_{wps} = \theta_i - \theta_{bps} \] (6)

\[ \mathfrak{F}_{bps} = \frac{\theta_{bps}}{N - 1} \] (7)

\[ \mathfrak{F}_{wps} = \frac{\theta_{wps}}{n_N - N} \] (8)

\[ \tau = \frac{\mathfrak{F}_{bps}}{\mathfrak{F}_{wps}} \] (9)

where \( \theta_{bps} \) denotes the sum of squares between durations; \( \theta_{wps} \) denotes the sum of squares within durations; \( \mathfrak{F}_{bps} \) denotes the mean square between durations; \( \mathfrak{F}_{wps} \) denotes the mean square within durations.

The threshold value \( \tau \) is calculated using the value of two most subsequent durations. The future active and passive time durations of nodes are modified based on the calculated value. If \( \tau \) exceeds \( \tau_{th} \), the time periods for are set to constant values \( T_a \) and \( T_p \).

\[ T_{\text{Active}}^{c+1} = T_a \] (10)

\[ T_{\text{Passive}}^{c+1} = T_p \] (11)

If the value of \( \tau_{th} \) is greater than \( \tau \), passive time is raised and active time decreased.

\[ T_{\text{Active}}^{c+1} = T_a \left( \frac{\tau_{th} - \tau}{\tau_{th}} \right) \] (12)

\[ T_{\text{Passive}}^{c+1} = T_p \left( \frac{\tau_{th} - \tau}{\tau_{th}} \right) \] (13)

where \( T_{\text{Active}}^{c} \) is \( c \)th active time duration of node; \( T_{\text{Passive}}^{c} \) is \( c \)th passive time duration of node. Algorithm 1 depicts the operation of the proposed platform’s energy saving mechanism.
3.2.2 Dimension reduction

Dimension reduction module is used to acquire extracted features of the flood-related geographical attributes and sensor dataset, which is substantially smaller in size, thus far intimately preserving the accuracy of the original data. Specifically, mining on smaller dataset will be more effective and deliver the same (or nearly identical) outcomes. In this proposed system, principal component analysis (PCA) method is applied to sensed data and geographical attributes for dimension reduction according to Algorithm 2.

$$c$$ Eigenvectors generate the dataset in new representation with reduced dimensions. The PCA restricts all new values to lower dimensionality and updates the database. Now that data are gathered and pre-processed, it must be evaluated for determining the flood level severity on the basis of the data received.

### Algorithm 2: Dimension reduction of flood attributes by using PCA

**Input:** Data set: $s_1^{(1)}, s_2^{(2)}, \ldots, s_n^{(p)}$; $n$: number of flood attributes, $p$: number of observations.

1 Normalize the original data: Calculate the mean ($\mu_i$), Variance ($\sigma_i$) and Covariance ($\text{Cov}(s_{jk})$).
2 Find the correlation coefficient ($c_{jk}^{(i)}$) and correlation matrix ($C$) of the normalized data.

$$c_{jk}^{(i)} = \frac{\text{Cov}(s_{jk})}{\sigma_j \cdot \sigma_k}$$

3 Determine the Eigen values ($\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$) from equation $|C - \lambda I| = 0$ and Eigen vectors ($e_{11}, e_{12}, e_{13}, \ldots, e_{mn}$) from equation $|C - \lambda I| e_{ij} = 0$ for the correlation coefficient matrix.
4 Sort the Eigen values and corresponding Eigen vectors so that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n$.
5 Choose the first $c \leq n$ eigenvectors and create the data set in new format.
3.3 Cloud layer

The flood-related environmental sensory IoT data from different locations is stored in Cloud. The data are pervasively sensed and periodically collected at different time intervals. Therefore, for further analysis, data are stored in cloud servers. The flood-related activities are adequately categorized based on reduced attributes by using artificial neural networks. Table 3 describes the notations used in the cloud layer.

3.3.1 Flood prediction sub-layer and alert generation

The artificial neural network (ANN) method is adopted to classify the dimensionally reduced data by the principal component analysis algorithm. ANN model is composed of 3 layers of node, i.e., the input layer, the hidden layer, and the output layer. Every node has several inputs (from dimensionally reduced flood-related attributes) and several outputs (according to flood sensitivity factor). The structure of the multilayered feed-forward neural network is shown in Fig. 2. Circles represent the nodes, and lines represent the relations. Each input \((c_1, c_2, c_3, ..., c_n)\) is multiply by a relation parameter known as weight \((W_i)\) and combined to generate a single value. A transfer function then regulates this value. The cumulative output value of a node can be expressed as follows:

\[
S_j = f(C \times W_i - T_j)
\]

Alert generation components are reasonable for transmitting alert notifications to decision-making agencies and citizens of a particular flood detected area (Algorithm 3).

| Table 3 | Notation table for the Cloud Layer |
|---------|-----------------------------------|
| Notation | Description                        |
| \(C\)     | Flood attributes                   |
| \(S_j\)   | Sensitivity factor                 |
| \(W_i\)   | Weight associated with each node   |
| \(T_j\)   | Threshold value for each flood attribute |
| \(L_t\)   | Level component of Holt Winter’s method |
| \(T_t\)   | Trend component of Holt Winter’s method |
| \(Q_t\)   | Seasonality component of Holt Winter’s method |
| \(a, \beta\) and \(\gamma\) | Model parameters |
| \(S_t\)   | Flood stage at time \(t\)          |
| \(x\)     | Length of season                   |
| \(L_0\)   | Initial value of level component   |
| \(T_0\)   | Initial value of trend component   |
| \(Q_0\)   | Initial value of seasonality component |

\[\text{Springer} \]
**Algorithm 3:** Classification and Alert Generation

**Input:** Dimensionally reduced flood attributes \( C = c_1, c_2, \ldots, c_n \).

**Output:** Alert generation according to possibility of flood.

1. Calculate Flood sensitivity factor \( S_j \) according to various dimensionally reduced flood attributes. As \( S_j = f(C \cdot W_t - T_j) \); Where, \( C = c_1, c_2, \ldots, c_n \) input to ANN model, \( W_t \) is weight associated with each node, \( T_j \) threshold value for each flood attribute.

2. **if** Flood sensitivity factor \( S_j \) > Predefined Threshold **then**
   
3. **else**
   
4. Flood is detected and immediate alert is generated to decision making agencies \((WRA,CDEOC)\) and citizens of that area.

5. Flood is not detected and no alert is generated

6. Exit.

### 3.3.2 Flood forecasting sub-layer

This sub-layer anticipates the potential occurrences of flood events by evaluating the current and historical values generated by the flood prediction layer. For this task, the Holt-Winters forecasting approach is used, which is the most frequently used exponential smoothing approaches. Holt Winter’s method considers three components to determine the future flood. The three components are given as follows:

1. Level \( (L_t) = \alpha \frac{S_t}{L_{t-1}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \)
2. Trend \( (T_t) = \beta(L_t - L_{t-1}) + (1 - \beta)(T_{t-1}) \)
3. Seasonality \( (Q_t) = \gamma \frac{S_t}{L_{t-1}} + (1 - \gamma)Q_{t-x} \)

where \( L_t, T_t, \) and \( Q_t \) are level, trend and seasonality components at time t. \( \alpha, \beta \) and \( \gamma \) are model parameters. \( S_t \) is flood stage at time t and x is seasons’ length. Flood stage for
$t + p$ is determined as: $S_{t+p} = (L_t + pT_t)Q_{t-x+p}$. The initial values for Holt Winter’s components are given as follows:

\[
L_0 = \bar{S}_1 - \frac{n}{2}T_0 \quad (14)
\]

\[
T_0 = \frac{\bar{S}_n - \bar{S}_1}{(n-1)x} \quad (15)
\]

\[
Q_0 = \frac{\bar{S}_y}{\bar{S}_y - \left(\frac{x-1}{2} - z\right)T_0} \quad (16)
\]

$z = 1, 2, 3, ..., x$ and $y = 1, 2, 3, ..., n$, where $\bar{S}_y$ is arithmetic mean of predicted flood stages for $y$th year, $n$ is the cumulative number of years considered.

Algorithm 4: Flood Forecasting sub-layer procedure

1. Input: Flood prediction dataset
2. Initiate the level, trend and seasonal components using $L_0$, $T_0$, $Q_0$.
3. Determine level, trend and seasonal components’ revised values using $L_t$, $T_t$, $Q_t$.
4. Evaluate the forecast value for $t=t+p$ by using $S(t+p)$.
5. Exit.

4 Performance evaluation

This portion of the paper discusses the implementation findings and addresses the reliability assessment of the proposed approach. The phases are addressed ahead:

4.1 Data accumulation and integration

Data cannot be derived directly from the environment for applying the proposed model but can be gathered from multiple official sources of data. We have generated flood attributes systematically by collecting data from India’s different government sites (http://www.imd.gov.in, http://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-datalist-rain-stations-india-metlogicaldepartment) and dataset repository (https://www.kaggle.com/biphili/india-rainfall-kerala-flood) for the Kerala region. The environmental dataset is created in such a way that all possible flood-related attributes are considered. The dataset contains 180 cases (about 14 districts of Kerala in 2018) with information on five attributes, i.e., season, temperature, relative humidity, rainfall, water level. The relationships between the flood attributes are presented by using the correlation matrix. Figure 3 depicts the correlation matrix of 5 independent variables. The correlation coefficient of variables is shown in Table 4.
4.2 Energy conservation using ANOVA and Tukey’s post hoc test

Water level sensor data and temperature data are retrieved from the sensor dataset (Lyons et al. 2011) for several lakes in Alaska to determine the performance of the proposed energy-conserving mechanism. The dataset contains hourly data for the attributes of water level and temperature. 24 h water level sensor data are considered for implementation of ANOVA and Tukey’s test. Considered data are divided into 6 intervals with 4 h in each interval. The result of ANOVA and Tukey post hoc test is depicted in Fig. 4. The result shows the maximum overlap of mean intervals.
4.3 Dimension reduction

The dimensionality of the final dataset is minimized by the principal component analysis (PCA) approach. All modern statistical analysis packages use programs to measure the vector pair (eigenvalue-eigenvector) of the sample correlation matrix. The PCA algorithm is available in Minitab and is applied directly to on generated dataset. Figure 5 shows the scree plot of flood attributes corresponding to eigenvalues and principal components. The first two principal components are selected out of 5 principal components since their corresponding eigenvalues are greater than unity. Table 5 specifies the loading values of two chosen PCs. These loadings illustrate how each variable contributes to a principal component. Almost all variables are significantly contributed to PC1, while WL is major contributed to PC2. Selected principal components clarified 78.4% of the overall variable heterogeneity in PCA (Table 6). The plot of first principal component against second principal component (Fig. 6) shows that samples were divided.

| Variable | PC1     | PC2     |
|----------|---------|---------|
| MONTH    | 0.324   | 0.154   |
| TEMP     | −0.499  | 0.363   |
| RH       | 0.566   | 0.242   |
| RAIN     | 0.570   | −0.008  |
| WL       | −0.001  | 0.887   |

| Variable   | PC1  | PC2  | PC3  | PC4  | PC5  |
|------------|------|------|------|------|------|
| Eigen value| 2.2988 | 1.1234 | 0.8827 | 0.4430 | 0.2521 |
| Cumulative variance | 46% | 78.4% | 86.1% | 95% | 100% |
These two principal components are directed to the cloud layer for forecasting and prediction of floods.

### 4.4 Flood prediction analysis

Two PCs, which explain 78.4% of the total variance, were extracted to utilize the ANN technique for flood prediction. The generated dataset is split into two subgroups. The
The first subgroup (70% of data) is used for training purposes, and the remainder 30% is used for testing purposes. The accuracy of the predictive model is assessed in terms of confusion Matrix ROC curve analysis. The results of ANN show an overall 94.2% accuracy. Table 7 displays the predicted and observed accuracy of analyzed data. The ROC curve of the predictive model (ANN) is shown in Fig. 7. The area under curve is 0.9807. The tenfold cross-validation of ANN result in mean absolute error (MAE) 0.21321, root-mean-square-error (RMSE) 0.427, relative absolute error (RAE) 18.75%, root relative squared error (RRSE) 28.75%, and Durbin–Watson statistic 1.55234.

Efficiency in data classification concerns the classification of data instances into various groups using the artificial neural network technique. Different statistical methods are used to test and evaluate the classification efficiency of the proposed model. These include accuracy, sensitivity, specificity, and F-measure. Various baseline classifiers are used for comparative study. We have used three separate classifier models as a baseline classifier model for comparison, namely KNN, decision tree, BBN. Results have been obtained for various classifier models and are shown in Fig. 8.

| Event | Predicted Event | Predicted No Event | % correct |
|-------|----------------|--------------------|-----------|
| Event | 73             | 2                  | 97.3      |
| No event | 8              | 97                 | 92.4      |
| All   | 81             | 99                 | 94.4      |
Fig. 8  ANN performance analysis

Table 8  Accuracy assessment parameters

| Statistical parameters            | One month | Seasonal | Average  |
|----------------------------------|-----------|----------|----------|
| Mean absolute deviation          | 0.13889   | 0.13889  | 0.13889  |
| Mean square deviation            | 0.27778   | 0.55556  | 0.41669  |
| Mean absolute percentage error   | 1.208871% | 3.267913%| 2.238392%|

Fig. 9  Flood forecasting for one month
4.5 Flood forecasting analysis

Holt-Winters forecasting model is used to estimate the future trends in flood stages. Minitab software is used for implementing Holt-Winter’s model using. The results of ANN are used as feedback for the model. The result of flood forecasting for a time span of one month is shown in Fig. 9, and seasonal forecasting is depicted in Fig. 10. The results show variation in observed and forecasted values. The accuracy assessment parameters, i.e., mean square deviation, mean absolute deviation and mean absolute percentage error, are shown in Table 8.

5 Comparative analysis and innovation of research

Based on six characteristics, i.e., (i) contribution (ii) application domain (iii) innovative technology used (iv) energy saving (v) real-time aspect (vi) decision-making model, the conceptual model is compared to several relevant research in the flood forecasting domain, as shown in Table 9. It is noted that the proposed system comprises various aspects of flood management, including detection and alert generation, prediction, and forecasting. The studies under consideration are applicable for only one aspect of flood monitoring with limited smart technologies. The proposed framework comprises various smart technologies and covers more flood monitoring aspects. In addition, the study provides a real-time aspect of the conceptual framework with a novel mechanism of energy saving within sensor nodes. Results of proposed prediction and forecasting models compared to similar studies are shown in Tables 10 and 11.

6 Limitations

The proposed approach considers only specific attributes for which IoT sensors may collect data for flood prediction and forecasting. Several additional flood-related parameters such as river attributes (river channel, riverbed’s gradient, and channel shape), soil, vegetation,
| Authors                  | Application Domain                                      | Smart technology used | Authors Contribution                                                                 | Authors Contribution                                                                 | Technology Real-time aspect | Application Domain | Authors Contribution                                                                 |
|-------------------------|---------------------------------------------------------|-----------------------|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|----------------------------|-------------------|--------------------------------------------------------------------------------------|
| Lo et al. (2015)         | Flood monitoring                                        | Sensors              | Visual sensing for urban flood monitoring and decision-making model                    | Threshold-based approach                                                                | No                         | Energy saving       | Sensors, No, Yes, Threshold-based approach                                             |
| Ray and Mukherjee (2017) | Cloud computing based flood detection and management    | sensors              | Cloud computing based flood detection and management of urban flood                      | Decision-making model                                                                  | No                         | Energy saving       | Flood detection and threshold-based approach                                         |
| Abdurrahman et al. (2017)| Flood forecasting                                       | Sensors              | Flood Forecasting using threshold-based approach                                        | Decision-making model                                                                  | No                         | Energy saving       | Flood detection and threshold-based approach                                         |
| Suparni et al. (2019)    | Flood forecasting                                       | N/A                  | Study of hybrid neural-fuzzy inference system for forecasting flood event vulnerability  | Deep learning model                                                                    | No                         | Energy saving       | Flood forecasting and decision tree                                                   |
| Vinothini and Jayanthly (2019)| Flood detection and notification                        | N/A                  | IoT-based flood detection and notification system                                       | Deep learning model                                                                    | No                         | Energy saving       | Flood detection and decision tree                                                   |
| Haidi et al. (2020)      | Data-driven modeling                                    | N/A                  | A real-time flood detection system based on machine learning                           | Decision-making model                                                                  | No                         | Energy saving       | Flood detection system based on machine learning                                     |
| Hashi et al. (2021)      | A real-time flood detection system based on machine learning | N/A                  | A real-time flood detection system based on machine learning                           | Decision-making model                                                                  | No                         | Energy saving       | Flood detection system based on machine learning                                     |

Table 9: Analysis of the proposed framework in comparison to previous research.
| Authors       | Contribution                                                                 | Application Domain                  | Smart technology used | Energy saving | Real-time aspect | Decision-making model                                                                 |
|---------------|------------------------------------------------------------------------------|-------------------------------------|-----------------------|---------------|-----------------|----------------------------------------------------------------------------------------|
| Moishin et al. (2021) | Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm | Flood forecasting                     | N/A                   | No            | No              | Convolution Neural Network (CNN), Long-short term memory (LSTM)                        |
| Proposed Model | Energy-efficient IoT-based cloud framework for early flood prediction        | Flood detection, alert generation, prediction and forecasting | Internet of Things (IoT), Fog Computing, Cloud Computing | Yes           | Yes             | Principal component analysis (PCA), Artificial Neural Network (ANN), Holt Winter’s forecasting |
and catchment characteristics (size, shape, and land use) have not been taken into account. Assimilation of remote sensing technologies and GIS can further extend the proposed framework.

### 7 Conclusion

This paper proposes an energy-efficient cloud system for flood prediction and forecasting based on IoT. The proposed model effectively determines the prediction and forecasting measures for the particular study area (Kerala, India). This framework is contributed for optimistically data generation with efficiently enhance the sensor lifetime. Dimension reduction algorithm is applied at the fog layer to maintain the optimization of network bandwidth. Moreover, ANN predictive algorithm is produced efficient results with 97.3% sensitivity, 92.4% specificity, and future flood stages are forecasted using Holt Winter’s model at the cloud layer. Experimentation results are stored at cloud storage for water management agencies and disaster management groups so that effective measures can be taken on time and reduce the post and during disaster destruction.

### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

### References

Abdurrahman M, Irawan B, Latuconsina R (2017) Flood forecasting using holt-winters exponential smoothing method and geographic information system. In: International conference on control, electronics, renewable energy and communications (ICCREC), 2017, pp 159–163, https://doi.org/10.1109/ICCCREC.2017.8226704
Afzaal H, Zafar N (2016) Cloud computing based flood detection and management system using WSANs. Int Conf Emerg Technol (ICET) 2016:1–6

Avvenuti M, Cresci S, Vigna FD, Tesconi M (2017) On the need of opening up crowdsourced emergency management systems. AI Soc 33:55–60

Aziz K, Rahman A, Fang G, Shrestha S (2013) Application of artificial neural networks in regional flood frequency analysis: a case study for Australia. Stoch Environ Res Risk Assess 28:541–554

Basha E, Ravela S, Rus D (2008) Model-based monitoring for early warning flood detection. In: SenSys’08 - Proceedings of the 6th ACM conference on embedded networked sensor systems. 295-308. https://doi.org/10.1145/1460412.1460442

Chu K, Liu W, She Y, Zulin H, Tan M, Liu X, Gu L, Jia Y (2018) Modified principal component analysis for identifying key environmental indicators and application to a large-scale tidal flat reclamation. Water 10:69. https://doi.org/10.3390/w10010069

Floods - National Disaster Management Authority. [Online]. Available: http://www.ndma.gov.in/en/media/public-awareness/disaster/naturaldisaster/floods.html

FMIS.” [Online]. Available: http://fmis.bih.nic.in/Riverbasin.html

Ghadim HB, Heydari M, Rashidi M, Noori M (2018) Application of Holt-Winters time series models for predicting climatic parameters (case study: Robat Garah-Bil Station, Iran). Polish J Environ Stud. https://doi.org/10.15244/pjes/100496

Gilandeh A, Sobhani B, Ostadi E (2020) Combining Arc-GIS and OWA model in flooding potential analysis (case study: Meshkinhazh city). Nat Hazards. https://doi.org/10.1007/s11069-020-03975-0

Gourbesville P, Batica J, Tigli J, Lavriotte S, Rey G, Raju D (2012) Flood warning systems and ubiquitous computing. La Houille Blanche 98:11–16

Goymann P, Herrling D, Rausch A (2019) Flood prediction through artificial neural networks: a case study in Goslar, Lower Saxony. In: The eleventh international conference on adaptive and self-adaptive systems and applications. Venice, Italy

Hadid B, Duviella E, Lecoeuche S (2020) Data-driven modeling for river flood forecasting based on a Piecewise Linear ARX system identification. J Process Control 86:44–86. https://doi.org/10.1016/j.jprocont.2019.12.007

Hashi AO, Abdirahman AA, Elmi MA, Hashim SZM (2021) A real time flood detection system based on machine learning algorithms. Innovative systems for intelligent health informatics. In: IRICT, (2020) vol 72. Springer, Cham

Hernández-Nolasco J, Ovando MA, Acosta FD, Pancardo P (2016) Water level meter for alerting population about floods. In: 2016 IEEE 30th international conference on advanced information and applications (AINA), pp 879–884

Hong H, Panahi M, Shirzadi A, Ma T, Liu J, Zhu AX, Chen W, Kougiass I, Kazakis N (2018) Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. Sci Total Environ 621:1124–1141. https://doi.org/10.1016/j.scitotenv.2017.10.114

Hoseini Y, Azari A, Alireza Pilpayeh (2016) Flood modeling using WMS model for determining peak flood discharge in southwest Iran case study: Simili basin in Khuzestan Province. Appl Water Sci. https://doi.org/10.1007/s13201-016-0482-4

https://www.kaggle.com/biphili/india-rainfall-kerala-flood

Human cost of disasters: An overview of the last 20 years 2000–2019. Centre for Research on the Epidemiology of Disasters (CRED), UNDRR

India Water Portal, http://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-datalist-rainstations-india-metrologicaldepartment

Indian Meteorological Department, http://www.imd.gov.in

Kumar VA, Girish B, Rajesh K (2015) Integrated weather and flood alerting system. Int Adv Res J Sci Eng Technol 2(6):21–24

Leitão J, Cardoso A, Marques JA, Simões N (2017) Flood management in urban drainage: contributions for the control of water drainage systems using underground barriers. In: 2017 4th experiment international conference, pp 286–291, https://doi.org/10.1109/EXPAT.2017.7984346

Lo SW, Wu JH, Lin FP, Hsu CH (2015) Visual sensing for urban flood monitoring. Sensors (Basel, Switzerland) 15:20006–20029

Lyons EA, Sheng Y, Smith LC, Li J, Hinkel HM, Lenters JD, Wang J (2011) Quantifying sources of error in multitemporal, Landsat-based lake mapping. Int J Remote Sens

Machekposhti K, Sedghi H, Telvari A, Babazadeh H (2018) Flood predicting in Karkheh River Basin using Stochastic ARIMA model. Int J Agric Biosyst Eng 12(3):89–96
Ma N, Yuan M, Bao Y, Jin Z, Zhou H (2010) The design of meteorological data warehouse and multidimensional data report. In: Proceedings of second international conference on information technology and computer science. pp 280–283

Moishin M, Chatterjee C, Raghuwanshi N (2009) Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. J Hydrol Eng 14:647–652. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000040

MunichRENatCatSERVICE. Available online: https://natcatservice.munichre.com/

Moishin M, Deo R, Prasad R, Raj N, Abdulla S (2021) Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3065939

Mukerji A, Chatterjee C, Raghuwanshi N (2009) Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. J Hydrol Eng 14:647–652. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000040

MunichRENatCatSERVICE. Available online: https://natcatservice.munichre.com/

Nguyen KT, Chua LHC, Son LH (2014) Flood forecasting in large rivers with data-driven models. Nat. Hazards 71(1):767–784

Noori R, Khakpour A, Omidvar B, Farokhnia A (2010) Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. Exp Syst Appl 37:5856–5862

Ongdas N, Akiyanova F, Karakulov Y, Muratbayeva A, Zinabdin N (2020) Application of HEC-RAS (2D) for Flood hazard maps generation: for Yesil (Ishim) River in Kazakhstan. Water 12:2672

Owotoki P, Manojlovic N, Lindenberg FM, Pasche E (2006) A data mining approach for capacity building of stakeholders in integrated flood management. In: Proceedings of the sixth international conference on data mining (ICDM’06). 0-7695-2701-9/06-2006 IEEE

Puah Y, Yuk Huang, Chua K, Lee T (2016) River catchment rainfall series analysis using additive Holt-Winters method. J Earth Syst Sci. https://doi.org/10.1007/s12040-016-0661-6

Ray PP, Mukherjee M (2017) Internet of things for disaster management: state-of-the-art and prospects. IEEE Access 5:2169–3536

Razak NH, Aris AZ, Ramli MF, Looi LJ, Juahir H (2018) Temporal flood incidence forecasting for Segamat River (Malaysia) using autoregressive integrated moving average modelling. J Flood Risk Manag 11:S794–S804

Rehman S, Sahana M, Hong H et al (2019) A systematic review on flood vulnerability models and methods for flood vulnerability assessment: framework for future research. Nat Hazards 96:975–998. https://doi.org/10.1007/s11069-018-03567-z

Romali N, Yusop Z, Ismail A (2018) Hydrological modelling using HEC-HMS for flood risk assessment of Segamat Town, Malaysia. IOP Conf Ser Mater Sci Eng 318:012029. https://doi.org/10.1088/1757-899X/318/1/012029

Sakhardande P, Hanagal S, Kulkarni S (2016) Design of disaster management system using IOT based interconnected network with smart city monitoring. In: International conference on Internet of Things and applications (IOTA), India

Sakib SN, Ane T, Matin N, Kaiser M (2016) An intelligent flood monitoring system for Bangladesh using wireless sensor network. In: 2016 5th international conference on informatics, electronics and vision (ICIEV), pp 979–984

Shih DS, Chen CH, Yeh GT (2014) Improving our understanding of flood forecasting using earlier hydrometeorological intelligence. J Hydrol 512:470–481

Sood S, Sandhu R, Singla K, Chang V (2017) IoT, big data and HPC based smart flood management framework. Sustain Comput Inform Syst. https://doi.org/10.1016/j.suscom.2017.12.001

Supatmi S, Hou R, Sumitra ID (2019) Study of hybrid neurofuzzy inference system for forecasting flood event vulnerability in Indonesia. Comput Intell Neurosci

Tayfur G, Singh V, Asce F (2006) ANN and fuzzy logic models for simulating event-based rainfall-runoff. J Hydraulic Eng ASCE. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:12(1321)

Tur R, Can B, Koç M (2010) Artificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters. Appl Ocean Res 32:425–433. https://doi.org/10.1016/j.apor.2010.09.005

UttarakhandStat Provides Comprehensive Information on All The Major Sectors of The State Level Consolidated Research Statistical Data. [Online]. Available: http://www.uttarakhandstat.com/default.aspx

Vinodithi K, Jayanthi S (2019) IoT based flood detection and notification system using decision tree algorithm. In: 2019 international conference on intelligent computing and control systems (ICCS), pp 1481–1486. https://doi.org/10.1109/ICCS45141.2019.9065799

Widmann M, Schar C (1997) Principal component and long-term trend analysis of daily precipitation in Switzerland. Int J Climatol 17:1333–1356

Wu CL, Chau KW, Fan C (2010) Prediction of rainfall time series using modular artificial neural networks coupled with data preprocessing techniques. J Hydrol 389(1–2):146–167 www.newsclick.in (Accessed 5 Dec 2020)
Yusoff A, Mustafa IS, Yussof S, Md. Din N (2015) Green cloud platform for flood early detection warning system in smart city. In: The 5th national symposium on information technology: towards new smart world (NSITNSW). 978-1-4799-7626-3/15/2015 IEEE

Zhu X, Zhang G, Sun B (2019) A comprehensive literature review of the demand forecasting methods of emergency resources from the perspective of artificial intelligence. Nat Hazards 97:65–82. https://doi.org/10.1007/s11069-019-03626-z

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.