Implementation of the Internet of Things for early Floods in Agricultural Land using Dimensionality Reduction Technique and Ensemble ML

[1]Murali Dhar M S, [2]Kishore Kumar A, [3]Rajakumar B, [4]Poonguzhali P K, [5]Hemakesavulu O and [6]Mahaveerakannan R

[1]Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, TamilNadu, India.
[2]Department of Robotics and Automation, Sri Ramakrishna Engineering College, Coimbatore, India.
[3]Department of AI & DS, JNN Institute of Engineering, Chennai, TamilNadu, India.
[4]Hindusthan College of Engineering and Technology, Coimbatore, India.
[5]Department of EEE, Annamacharya Institute of Technology & Sciences, Rajampet, Andhra Pradesh, India.
[6]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India.
[1]msmddhar@gmail.com, [2]kishorekumar.a@srec.ac.in, [3]rajakumarb@jnn.edu.in, [4]poogasanthosh@gmail.com, [5]hkesavulu66@gmail.com, [6]mahaveerakannanr.sse@saveetha.com

Correspondence should be addressed to Murali Dhar M S : msmddhar@gmail.com.

Article Info
Journal of Machine and Comp uting (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202303050
Received 02 June 2023; Revised from 25 August 2023; Accepted 26 September 2023.
Available online 05 October 2023.

Abstract – Due to human activities like global warming, pollution, ozone depletion, deforestation, etc., the frequency and severity of natural disasters have increased in recent years. Unlike many other types of natural disasters, floods may be anticipated and warned about in advance. This work presents a flood monitoring and alarm system enabled by a smart device. A microcontroller (Arduino) is included, and its support for detection and indication makes it useful for keeping tabs on and managing the gadget. The device uses its own sensors to take readings of its immediate surroundings, then uploads that data to the cloud and notifies a central administrator of the impending flood. When admin discovers a crisis situation based on the data it has collected, it quickly sends out alerts to those in the local vicinity of any places that are likely to be flooded. Using an Android app, it alerts the user's screen. The project's end goal is to develop an application that swiftly disseminates flood warning information to rural agricultural communities. Scaled principal component analysis (SPCA) is used to filter out extraneous data, and an ensemble machine learning technique is used to make flood predictions. The tests are performed on a dataset that is being collected in real-time and analysed in terms of a number of different parameters. In this research, we propose a strategy for long-term agricultural output through the mitigation of flood risk.

Keywords – Flood, Monitoring System, Ensemble Machine Learning, Scaled Principal Component Analysis, Microcontroller.

I. INTRODUCTION

Natural disasters account for around 60,000 annual deaths, or 0.1% of all deaths worldwide [1]. Some examples of such natural calamities are floods, earthquakes, hurricanes, and landslides. As much as 40 percent of all natural disasters worldwide occur as a result of flooding [2]. The huge increase in flood hazards [3] is attributable to climate change, hurricanes, high storms [4]. It is estimated that floods have caused dollars in damage and the deaths of thousands of people worldwide [5]. In this era of concern for sustainability and smart cities [6], it is imperative that flood damage be reduced to a minimum [7], as it results in enormous economic losses in addition to the tragic loss of life [8]. Since the turn of the 2000, there has been a significant increase in flood events [9], with losses due to flooding increasing from $6 billion to $10 billion as a direct result of extreme rainfall events [10]. As a result, governments have spent billions on flood prevention projects [11]. The economic growth of the impacted countries is further hampered by the costs of rescue operations, reconstruction, and relief services. An estimated USD 19 billion was lost worldwide due to floods in various geographical locations in 2012 [12]. Many people perish because of delayed response times for aid and recovery services following a flood because of a lack of precise and rapid technologies that might routinely sense the onset of flooding in an area.

Unmanned Aerial Vehicles (UAVs), on the other hand, are typically developed for image-based systems, making them unfit for detecting lava floods. For a lava flood early warning system, machine learning and artificial intelligence are better suited to catastrophic situations [13] or early prediction [14-15]. Several studies, like [16], have integrated Internet of Things (IoT) with machine learning and AI techniques for use in emergency management. Most of these systems are used for detection [17-19]. Fuzzy-based algorithms are the best applicable to monitoring and early warning systems among all methods for ML and AI approaches. Decision-making systems that are based on the theory of fuzzy sets can have the same level of conceptual knowledge as their designers, and there are also such systems that can process many inputs and provide several outputs. [20] Fuzzy logic can be used to create a flood detection system, for instance. Further, a decision tree method combined with IoT was developed for flood detection and alerting, yielding good accuracy and performance [21].

It is common practise for flood monitoring systems to be linked to rainfall intensity schemes because the heavy downpour most common catastrophes caused by flooding. As [22] examined a neural network approach for application in a cheap flood surveillance system in which a water level sensor when a Raspberry PI was connected to evaluate rainfall intensity. For both scientific inquiry and agricultural output, high-dimensional data is everywhere. The vast amounts of information it provides are accompanied by significant difficulties for data mining and pattern detection due to the data's sparseness and redundancy. In a pattern recognition system, dimensionality reduction is a critical stage because it can simplify learning methods, lessen the effects of noise, and boost classification precision. SPCA is utilised here as a dimensionality reduction method, which helps to enhance the ensemble machine learning approach taken in this work. Predicting early flood in agricultural lands using an ensemble model.

The rest of the paper's information is organised into distinct chapters. Section 2 analyses relevant literature. In Sections 3 and 4, we provide context for the integration of ML with the IoT, and we briefly describe the suggested paradigm. In Section 5, we compare the proposed model to the state-of-the-art validation approaches. Section 6 concludes the paper and looks ahead to potential extensions of the research.

II. RELATED WORKS

Susceptibility in Iran and India are generated using an unique hybrid model presented by [23], which combines the autoencoder models. In two scenarios, nine and twelve variables were examined as predictors for flood susceptibility mapping, respectively. The criterion was utilised to evaluate the predictive efficacy of the suggested hybrid model in comparison to the conventional MLP model. In the training phase, MLP and autoencoder-MLP models achieved AUROC curves of 75% and 90% for the Iran and India examples, and 74% and 93% for the testing phase. Based on the findings, a hybrid autoencoder-MLP model is preferable to the MLP perfect for flood susceptibility mapping.

The effects of flood and drought disasters (FDD) on farming were examined by [24] in terms of their spatial and temporal distribution. This research proposes a comprehensive index of FDD based on the impacted and destroyed crop rate and combines multi-methods to analyse the disaster FDD on agriculture in China. What we learned from the data is as follows: (1) Droughts have a greater temporal and geographical impact on crops than floods. The effects of FDD on farms have been declining since 1978. (2) Droughts were most severe in the Northeast, Northwest, and North China, while floods had the greatest impact on agriculture in the Northeast. The Southwest, Central-south, and East regions of China all felt the effects of these natural disasters. (3) The transfer pathways of the gravity centres of the FDD's impact on agriculture were diametrically opposed; the gravity centres of flood disasters moved to the southwest and the northeast, whereas those of droughts moved to the northeast and the southwest. (4) Largest sections of the Northwest, Southwest, Central-south, and North of China experienced droughts, whereas the Southern regions of China saw the most benefit from human activity in reducing agricultural losses due to floods. Benefits for agriculture management and effective FDD governance are anticipated from the study's findings.

Utilising UAVs, [25] creates an automated flooded areas in aerial photographs. In the study, the Haar cascade classifier was used to analyse UAV aerial photos for recognisable structures like roads and buildings, as well as to locate flooded areas. After the landmarks have been retrieved, they are included in a deep learning model's training dataset. The experimental results demonstrate a 91% and 94% success rate in identifying buildings and highways in photographs, respectively. Dividing flooded and non-flooded properties areas from the input case study photographs was 91% accurate overall. Promising findings were obtained using pre- and post-flood classes of test photos, indicating that the technique is effective. Using this technique, rescuers can swiftly pinpoint flooded areas and reach stranded citizens.

In [26] introduces AgriFloodNet, an improved convolutional neural network model for mapping agricultural fields impacted by floods in remote sensing photos. Sentinel-1's bi-temporal Synthetic Aperture Radar (SAR) photos are used to efficiently learn permanent and flood water features, while Sentinel-2's cloud-free multispectral images (MS) are used to effectively learn agricultural land features. When it comes to depicting the effects of flooding on agricultural lands, a change map is generated by fusing the processing results of multisensory pictures at the decision level. With an accuracy of 98.75%, the proposed AgriFloodNet excels at FLOOD dataset. Single-sensor solutions, on the other hand, achieve an accuracy of 96.88% for SAR images and 91.11% for MS images. The typical has also been tested for a new flood occurrence in Patna, India. A total of 61% of the test site's agricultural fields were damaged or destroyed when a flood occurred on what was thought to be 75% of the site's total area.

After a storm hits [27] creates a framework to locate the locations most affected by flooding, (2) maps the regions most affected by the high rainfall and thunderstorms, and (3) evaluates the storm's primary influence on the land cover through

the flood. The wave of the severe rainfall and thunderstorm event indicated that about 2255.67 of the study area was inundated. Seventy percent of built up areas and roads were flooded during this disaster, with another 50 percent of shrubs getting wet. There was less damage to sand dunes than to other land covers, but overall, about 30% of agricultural land, tree canopy land, and barren ground was flooded. These findings not only show that land cover is vulnerable to flooding, but also highlight the benefits of using Google information on its platform to monitor and track natural hazards through time.

With climate data spanning 30 years, from 1991 to 2020 [28] seeks to provide a deep forecasting model based on optimised (Gated Recurrent Unit) GRU neural network to predict rainfall in Pakistan. For accurate forecasting, the climatic variables were first retrieved, and then refined by removing outliers and extreme values. After that, we used data normalisation techniques to convert all the numbers to a common scale, without sacrificing any of the variance or precision. The suggested model outperformed state-of-the-art rainfall forecasting models in terms of prediction accuracy by keeping NMAE and NRMSE to very low levels. A correlation and regression analysis was performed on the climatic variables that were employed in the forecasts. The data also showed that the air quality factors had a positive link with rainfall throughout the year. Compared to the first and second quarters, when there was a weaker or nonexistent relationship between air quality factors and rainfall, the third quarter and the end of the year revealed a strong correlation. Additionally, the data demonstrated a robust relationship between climatic variables and monthly rainfall occurrence throughout the year.

The [29] uses a revised universal soil loss equation (RUSLE) model to try to forecast the likelihood of soil erosion. To run the LSTM model, we used daily rainfall data collected from five agro-meteorological sites in the Central Highlands of Sri Lanka between 1990 and 2021. Precipitation values for the next 36 months at each station were input into the LSTM model, which was trained with a time series of monthly rainfall data. The 2024 rainfall erosivity map layer was generated using geo-informatics technologies. Soil erosion rates in the Highlands are expected to average 11.92 t/ha/yr, according to the RUSLE model. About 30 percent of the land area is predicted to fall into the moderate to extremely high soil erosion susceptibility classes, according to the soil erosion susceptibility map. Previous soil erosion map layers from the years 2000, 2010, and 2019 were used to verify the accuracy of the final product. A satisfactory level of prediction ability is demonstrated by the AUC-ROC value of 0.93 for the soil erosion susceptibility map.

III. ML AND INTERNET OF THINGS IN FLOOD DETECTION

The Internet of Things (IoT) has numerous ramifications for research and important meetings in the present day. Web and things are the first sources for the concept of IoT. Now, the fixed meaning of these two phrases describes the structure of dependent variations that may be mapped onto established correspondence traditions. The Internet of Things (IoT) was first proposed in the 1980s as a means of enhancing cutting-edge remote developments like RFID, sensors, actuators, and cellphones. Disk-to-disk communication has become restricted to a narrowly tailored system where a subset of the nodes act as a handoff or portal administrator during the course of regular framework failure. When the affected area gains access to live innovations like Internet, satellite, or a functional This hand-off administrator will use a conventional cell coordinate connect the area to the rest of the world. Using IoT, tasks are distributed over various devices in order to verify device-to-device communication and facilitate rapid, objective data collection and analysis inside their own frameworks. In this case, IoT is a diverse framework for categorising the associated gadgets in their own unique ways. The inexpensive to moderate cost of the numerous sensors allows for a high data accuracy rate and moderate hardware dependability. Specifically, Internet of Things technologies are in high demand in emergency management initiatives.

The initial phase in the procedure is to gather real-time data from sensors such ultrasonic water flow/level sensors, temperature/humidity sensors, and rain gauges. A microcontroller receives all the information. With the help of numerous software components running in the cloud, the fog, or at the edge of the network, IoT can analyse data in real time. Future flood hazards can be predicted using a variety of ML and AI methods for data mining by comparing raw data to known parameters. To coordinate evacuation and relief activities, this data can be broadcast via social media, other broadcast modes, and warning messages sent via the short messaging service (SMS). The Internet of Things (IoT) is used to keep tabs on the environment and alert authorities when danger is imminent.

IV. PROPOSED MODEL

As seen in **Fig 1**, the proposed model's process flow is shown. A variety of sensors, including water level sensors, temperature sensors, rain gauge sensors, and vibration sensors, are used to keep an eye on the farms. These readings are transmitted to an Arduino UNO, which stores them in the cloud. When SPCA combined with ensemble machine learning successfully identifies flooding on the field, the resulting data are delivered to android applications like WhatsApp or SMS to warn the farmers.

Client Upstream From Agricultural Land

The first and important stage of the proposed system is client upstream from agricultural land. This stage consist of several sensors, which has helps to identify the weather conditions. The sensors such as, rain gauge sensor measurements, temperature & humidity measurements, water level measurement, vibration measurements and image sensor. With the help of Arduino UNO sensors measured data's are collected and transfer to cloud/client downstream. The working of the all sensors are described in the below subsections.

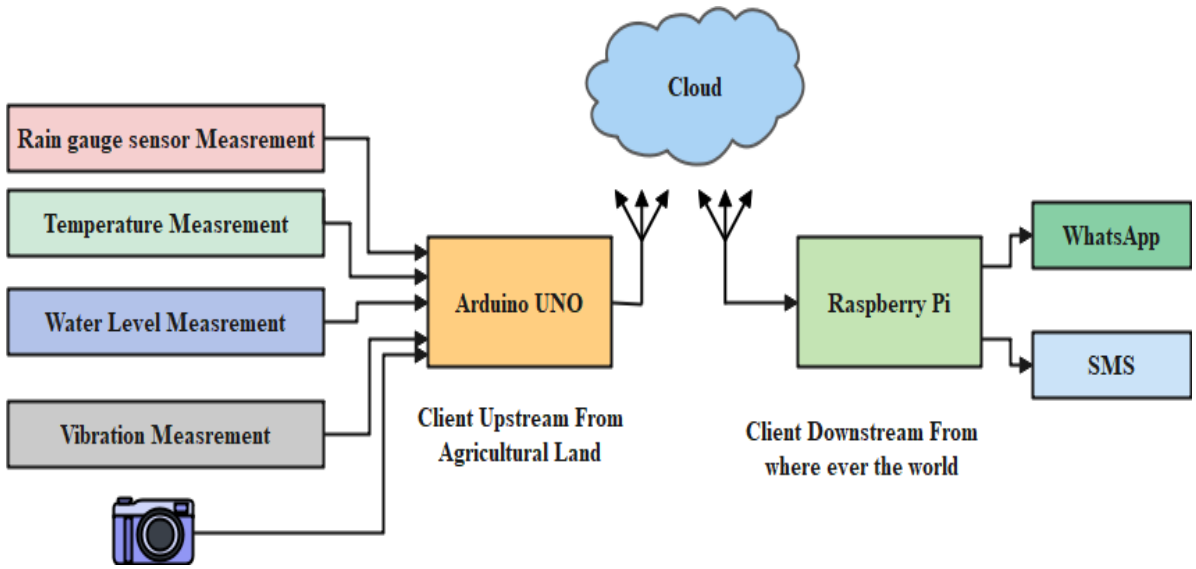

Fig 1. Proposed System model.

Arduino UNO

The tmega328p microcontroller is an integral part of the Arduino development package, which is used in this project. It has both digital and analogue I/O pins. Values will be transmitted to the server. Some users are signed up for the mobile apps in order to see the secure locations in their immediate vicinity via Google Maps as part of the experiment. Any participant in the experiment who has already registered for the same application can see the results as soon as they are available.

Rain Gauge Sensor Measurements

The sensor for the rain gauge, as well as the integrated microcontroller for the ESP, and the Internet of Things make up the backbone regarding the gadget for measuring the intensity of the rainfall's architecture. A mechanical and electronic system work together to form the sensor located on the rain gauge. In terms of the mechanical system, the following are the details: Dimensions of 42 x 40 cm, sensitivity of 1 mm/tip, 100 cm2 funnel cross-sectional area, 10 ml water capacity per tip. There are no constraints on the design of the tipping bucket system, which is constructed without rainwater reservoirs. Using a reed switch, we can determine how much water is entering the electronic system by tallying the number of times the bucket is tipped over. Establishing a connection between the ESP microcontroller and the rain gauge sensor allows for the conversion of counted tips over the course of an hour into a mm/h measurement of rainfall intensity.

The graphic shows that the ANN's output is split across five different roots labelled O1, O2, O3, O4, and O5. However, it takes in just one piece of information at a time: the sensor's output. The output variable, rainfall intensity, is categorised using information found in [30]. With no rain, O1 is the output every hour or every day. O2 precipitation falls at a rate of 0.1 to 4.9 millimetres per hour, or 0.1 to 199 millimetres per day. When the O3 indicator is present, precipitation rates of 5–9.9 mm/hour are typical. O4 is characterised by intense rain, with a precipitation rate of 10-20 mm/hour. Rainfall rates of 20 mm/h or more (or 100 mm/day) qualify as O5 rainfall.

Temperature & Humidity Measurements

The DHT11 sensor model is employed in this work. It is a cheap digital sensor that measures the relative humidity and temperature of the air or water surrounding it and outputs this information as a digital signal to a microcontroller's data pin.

Water Level Measurement

In order to determine how deep the water is, a float sensor is used. The normal operating position serves as a complete opening and closing circuit when it is not in use. When the water level is now lower than the predetermined minimum, power flows through the warning circuit, and then on to the rotor and hall effect sensor after that housed in the current plastic body, re-routing the supply water to the same rotor as before. Changes in velocity are visible, coupled with variations in flow rates.

Vibration Measurements

Vibration detection can begin with a number of different inputs, including material volume, satellite imagery, and infrared mapping. According to this research, the frequency of vibration is used as the starting point parameter or primary input in order to identify cool lava floods. The movement and flow of material or silt from the peak of the mountain causes oscillations. The identification of this vibration allows for the detection of cool lava floods. As can be seen in Fig 2, This is what the cold lava substance vibration sensor looks like an instrumentation design.

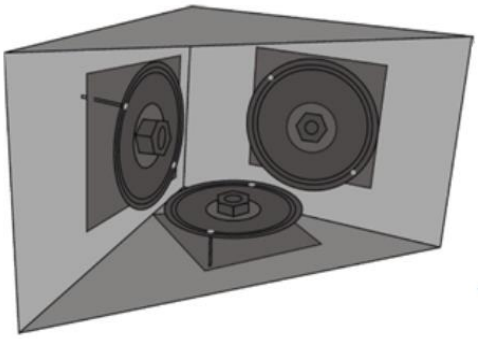

Fig 2. Cold Material Vibration Instrumentation.

Image Sensor

An ultra-compact UAV known as River-map was chosen for monitoring land regions and detecting flooding. aerial pictures with a high resolution and the RGB color space of the case study region were taken with a Go-Pro® digital camera. As can be seen in **Fig 3(a-b),** on September 2, 2021, real-time surveillance of the area under investigation was carried out using this setup, drawing attention to the devastation as a result of the agriculture flood in this area. The pictures demonstrate that the agricultural field has been flooded.

(a) With flood (b) With flood

Fig 3. Sample Collected Images

Client Downstream From where ever the world

This is the second most important part of the proposed work. Client Downstream section has detected the flood, that information is transferred to the farmers via SMS, and what's app. In the received signal, consist of five different data's such as rain gauge sensor data, temperature & humidity data, water level data, vibration data and image data. In that, rain gauge, temperature, humidity, water level and vibration data are directly considered as features and it's fused. Another Sensor data is an image data, which is feature extracted and dimensionality-reduced data is fused with the other sensor data and it's given to the ML technique.

Feature Extraction

In order to maintain both the low and high frequency info, images are typically fragmented into a sum of sub-images at varying resolution levels. Texture data can be retrieved from photos with the use of the DWT property. f(u) is the inner product of the square-integral function f(u) and the wavelet transform w, and (u) is the original-valued function. The wavelet function is given in the Eq. (2).

$$w[f(s,\tau)] = \left(f, \psi_{s,t}^{k}\right) = \int_{\infty}^{\infty} f(u)\psi_{s,t}^{k}(u)du \tag{1}$$

$$Where, \psi_{s,t}^{k}(u) = \left(\frac{1}{\sqrt{s\psi_{s,t}^{k}}}/s\right) denotes \tag{2}$$

scale (s), translation (t), and orientation (k), all in the context of the wave family. Vertical, horizontal, and diagonal orientation are indicated by the h, v, and d parameters. During the parameters s=2j and =2j,n,j,nz, a dyadic wavelet decomposition is obtained. A scalable example of DWT, the dyadic wavelet decomposition follows 2. We use perfect reconstruction filter banks to implement dyadic wavelets for the wavelet decomposition that follows. Through the use of Eq. (3)'s wavelet function (u) and the scaling function (u), we may (4). The wavelet atoms can be described using scaling and the three mother atoms h, v, and d. These primordial nuclei are found by multiplying the 1-dimensional tensors (u) and (u), which are represented by the symbols in Eq. (5), and the 2-dimensional tensors (u) and (u), respectively (6)

$$\psi_{j,n}^{k}(u) = \frac{1}{\sqrt{2^j}}\psi^{k}\left(\frac{u-2^j.n}{2^j}\right) \tag{3}$$

$$\varphi_{j,n}^{k}(u) = \frac{1}{\sqrt{2^j}}\psi^{k}\left(\frac{u-2^j.n}{2^j}\right) \tag{4}$$

$$\varphi(u) = \varphi(u_1)\varphi(u_2), \psi^{k}(u) \tag{5}$$

$$\varphi^{v}(u) = \varphi(u_1)\psi(u_2), \psi^{d}(u) \tag{6}$$

Combining down samplers with digital filter banks allows for the realisation of DWT on a two-dimensional scale. Both a low-pass and a high-pass digital filter make up the filter bank. The wavelet configuration structure is used to categorise the bank's number into groups that correspond to the functions that are sought. Next, a 1-dimensional wavelet transform is performed on the rows and columns of the underwater image independently to get a 2-dimensional wavelet coefficient. Frequency domain decomposition of the original A (2(j+1)) fat resolution photos yields four subband images. Resolution in the vertical, horizontal, and diagonal directions between the 4-subband images originates from the original 3-subband images, such as D (2i)h f, D (2i)h f, and D (2i)h f. Since A (2i) f is a low-resolution approximation image, the whole underwater image is given by the symbol A (2(j+1)) f in the Eq (7).

$$A_{2^{j+1}}f = D_{2^i}^{h}f, + D_{2^i}^{v}f + D_{2^i}^{d}f + A_{2^i}f \tag{7}$$

The 2-dimensional orthogonal wavelet represents the separated image components. Wavelet decomposition of an image yields four orthogonal sub-bands, designated by the notations D (2i)h f,+D (2i)v f+D (2i)d fandA (2i) f for Low-Low (LL), Low-High (LH), High-Low (HL), and High-High (HH), respectively. Extraction of LL and LH wavelet highlights from each sectioned image serves as the work's foundational phase of component extraction. Gray Level Co-event Matrix (GLCM) is utilised and the component values are extracted with the aid of wavelet highlighting. Highlights tally up like us, Dissimilarity, Autocorrelation, Cluster Prominence, Cluster Shade, and Contrast. Entropy of energy, There must be homogeneity, homogeneity, homogeneity. Highest possible probability, the number of cubes, Variance, Statistical mean, average, change, entropy, dispersion, transformatio n entropy, Statistics such as the Evidence Ratio, the Information Ratio, the Inverse Difference, the Inverse Difference Standardized, and the Converse Alteration Moment Normalized.

SPCA Method for Dimensionality Reduction

Let's pretend there are N predictors, and their product is given by X t= (X (1,t),,X (N,t))', where T is the total number of observations and I is any of the integers from 1 to N, inclusive. The goal is to use these predictors to make predictions about a target variable y (t+h) at a forecast horizon of h time units into the future. Each X (i,t) is a useful but limited indicator of the outcome of interest. In light of this, it is improbable that a small subset of them will adequately depict the target's dynamics. However, the curse of dimensionality causes problems with both in-sample and out-of-sample overfitting as well as performance differences were found when every single predictive factor are used in a traditional multivariate regression. It is usual practise to impose a factor structure on the predictors and then extract the latent components from that structure, which results in a reduction in the dimensionality of the data.

In this study, we focus on the N predictors X t and the outcome y (t+h) by considering a latent component model of their joint dynamics.

$$X_{i,t} = \mu_i + \lambda'_i f_t + e_{i,t} = \mu_i + \phi'_i g_t + \psi'_i h_t + e_{i,t} \tag{8}$$

$$y_{t+h} = \alpha + \beta' g_t + \epsilon_{t+h} \tag{9}$$

where g t are r 1-dimensional the aim y (t+h), and ht are (r r 1)-dimensional irrelevant factors, and f t=(g t',h t')' are r-dimensional unobserved factors. '_i=(i I represents the loadings on f t for each predictor I = 1,,N. Any variable in f t can predict y (t+h) in their setup. In contrast, our framework makes greater sense when applied to real-world data since it considers just a subset of ft (factors g t) to be important to the objective. Since the majority of theoretical conclusions rely on all elements, this work clearly analyses the scenario when the irrelevant factors h t are removed from Eq (8) for predicting.

The PCA is a common approach for estimating the latent factors ft due to the factor structure. To be more precise, the PCA computes an estimate for f t=(g" t,h" t)'as T times the eigenvectors associated with the r greatest eigenvalues of M xx, where M xx=1/N _(i=1). ^ The sample covariance matrix N(X I X I has TT dimensions and is defined as ((X i=(X (i,1),) X (i,2),...,X _(i,T))) and (X (i,t)=X (i,t)-1/T _(s=1)TX (i,s)). Let F= (G,H), where G=(g 1,g 2,...,g T)' and H=(h 1,h 2,...,h T)'. The factors Fb can be thought of as a shorthand for much of the variety in X t. It is possible to do forecasting based on either the full set of factors (g t and h t) if one chooses a low bias, or on only some of the factors (g t) if one prefers a parsimonious model and thinks the estimated partial factors are adequate. For either scenario, rN allows for dimension reduction.

However, the PCA has the drawback of not considering the goal data when the model is described by Eq. (8) and (9). In particular, there is no guarantee that the first r-1 main components can provide the most accurate prediction of the target, and the PCA fails to distinguish between target-relevant and irrelevant latent factors when the factors are high. If the factors aren't strong enough, the PCA might not be able to separate them out from the other noise, resulting in inaccurate predictions even when considering the full set of variables. The SPCA is intended to improve upon the PCA by including the target information in the factor extracting process, hence mitigating these shortcomings. Using the SPCA, we make a two-stage prediction of the target: Assemble a set of scaled predictors (_1 X (1,t),..., _N X (N,t)), where the scaled coefficient it is the predicted slope obtained from regression analysis of the target on the i-th (standardised) predictor: Use principal component analysis on (_1 X (1,t),..., _N X (N,t))to derive r factors, and then put them to use to make predictions

about the goal. To be more precise, one must determine the TT matrixM XX°=1/N _(i=1)N I (X i) (I (X i))', where (X i) is predictor i's devalued vector.

To get the sPCA factors, F sPCA, we multiply the r greatest eigenvalues of the matrix M XX° by T, where T is the number of factors in the matrix. Consider the first r 1 columns of F sPCA to be G sPCA, and the transposition of the t-th row of G sPCA to be g tsPCA. Regressing y (t+h) on a constant term and the estimated factors fbtsPCA yields a full-factor based forecast, while regressing y (t+h) on a constant term and the estimated factors gtsPCA yields a partial-factor based forecast. If so, here is what we can expect for y (t+h) according to the SPCA:

$$\hat{y}_{t+h}^{sPCA} = \hat{\alpha}^{sPCA} + (\hat{\pi}^{sPCA})' \hat{f}_t^{sPCA} \tag{10}$$

$\hat{\alpha}^{sPCA} + (\hat{\pi}^{sPCA})'$the respective slope estimates of the above two predictive regressions.

It is worthwhile mentioning that the scaled predictors $(\gamma_1 X_{1,t}, \ldots . \gamma_N X_{N,t}$ follow a latent factor structure as $\gamma_i X_{i,t} = \gamma_i \mu_{i,t} + \gamma_i \lambda_i' f_t + \gamma_i e_{i,t}$ where γ_i is the probability limit of gbi for each i. So the scaled predictors actually share the same factors f with the original predictors. Since the forecasted target $y_{\mathrm{t+h}}$ is related to the factors instead of the loadings, This, of course, raises the question of how the SPCA forecast can be more accurate than the PCA forecast, especially when taking into account each and every component in the forecasting process $y_{\mathrm{t+h}}$. The solution to this issue is that the SPCA eliminates the predictors that are unimportant by giving them decreasing weights. This is how they do it. This process is particularly significant because, in contrast to the signals caused by strong variables, the signals caused by weak factors do not typically dominate the noises to the same degree. It is possible that the typical PCA won't be able to separate the signals from the enormous quantity of noise if there isn't a signal-strengthening process included.

Classification

Decision Tree (DT) One way to do this is with DT, which can be used to demonstrate an algorithm consisting only of if-then expressions. Commonly used supervised classification algorithms are compiled into the DTs. They are effective at classifying data, their decision-making process is simple to understand, and the algorithm for creating (training) them is quick and easy to grasp.

Support Vector Machine (SVM) SVM is a supervised algorithm. This model works wonderfully on a tiny data set with only a few extreme observations. Find the hyperplane that will be used to divide the data. Separate regions of two spaces will be created using the hyperplane. Data in this category will be organised similarly. The following equation represents the decision state of the support vector machine.

$$||Y|| = \sqrt{y_1^2 + y_1^2} + \cdots + y_n^2 \tag{11}$$

Ensemble Learning (EL) Method.

The performance of a model can be greatly enhanced by employing the machine learning technique of ensemble learning. By bringing together several different sets of learners, the model's predictive power is boosted (Base Learners). It is impossible to stress the significance of using the right ensemble for diabetes prediction.

The proposed methodology has incorporated SVM and DT as basic learners with the EL approach. The following equation can be used to express a majority vote.:

$$\sum_{t=1}^{T} d_{t,J} = \frac{c}{max_{j=1}} \sum_{t=1}^{T} d_{t,j} \tag{12}$$

where T is the dataset and C is the class label.

V. RESULTS AND DISCUSSION

Instrumentation Testing

The suggested vibration sensor will undergo instrumentation testing to achieve multiple goals

- Identifying the range of frequencies at which different types of material flows may be detected.
- Establishing the accuracy for this range.
- Locating the range of frequencies at which different types of materials can be detected.

Connecting the vibration sensor to the frequency generator will yield the first test objective, which will be recorded as the frequency registered by the sensor. The sensor performed admirably in all tests, detecting frequencies between 1 and 9000 Hz with a perfect rate of success. After passing the 1000 Hz mark and into the 3270 Hz region, however, accuracy dropped to 99.004%. Additionally, the frequency range at or above 32800Hz is undetectable by the vibration sensor. The suggested vibration sensor can thus be stated to measure frequencies from 1 to 9000 Hz. Multiple subtests will make up the next evaluation. A total of three tests are used to ensure that the cold lava material identification instrument is functioning properly. In the first subtest, stones weighing 30–40 grammes and 20–25 cubic centimetres in volume were employed. Following the placement of the vibration second sensor near the end of the landslide, the stones will be released in a well-orchestrated landslide. Each group of stones (of which there are ten total) was produced by performing ten repeats.

Therefore, we shall perform 100 trials of the subtest. Once the frequencies have been recorded and evaluated, the results will be shown. In this subtest's results, 14 Hz is the lowest and 61 Hz is the highest frequency registered.

In the second section, participants utilised stones weighing 60–90 grammes and measuring 35–40 cm3 in volume. In this section of the test, we may detect frequencies as low as 18 Hz and as high as 87 Hz. In the third experiment, stones weighing 100–140 grammes and measuring 50–75 cm3 in volume were employed. We have a range of frequencies from 25 Hz all the way up to 113 Hz. This research agrees with previous research showing that lahars (cold lava floods) with high sediment content (debris flows) preferentially produce significant signals in the low-frequency range.

Synchronize with the suggested vibration sensor can accurately detect materials in a simulated cold lava flood at frequencies between 14 and 113 Hz. The suggested vibration sensor can thus identify material flows with high accuracy and minimal measurement mistakes. The vibration sensor that has been presented possesses a high sensitivity to lower frequencies while also being capable of detecting material flows throughout a large frequency range. This discovery is essential to the construction of a detection and early warning system for such situations because standard sensors that rely on frequency and acoustic measurements to identify lava flows have limits. As a result, this discovery is critical to the development of a detection and early warning system. It was difficult to calibrate typical geophone instruments because the instrument's high-gain low-frequency channel shows and records signals with low and high frequency independently from one another.

The sensors were put through their paces in field-scale measurement testing after passing the instrumentation tests in the lab. Desa Sumber, Magelang, a small village at the peak of Mount Merapi, was the site of the field trial. Upstream of Kali Senowo, three automatic rain gauges and one vibration sensor were set up. These sensors took readings, analysed them, and then uploaded the results to the cloud-based systems that make up the IoT. When extremely heavy rain was predicted, the system would send out SMS alerts to the FPRB members. Because the vibration sensor measured a frequency of less than 1000 Hz, an evacuation alert was not issued. The following are the outcomes of large-scale flood discharge tests conducted on-site in **Table 1.**

:

Table 1. Field-scale Test Results.

Parameters	Values
Watershed Area	20 km2
Run-off Coefficient	0.9
Rainfall Intensity Category	Very Heavy
Measured Flood Peak Discharge	252 m3/s

Detection Results

Our segmentation and classification results are measured against the challenge evaluation metrics. Criteria including sensitivity (SE), specificity (SP), accuracy (AC), recall (R), and precision (P) are used to assess segmentation results (P). The benchmarks for success are specified as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

$$Precision = \frac{TP}{TP+FN} \quad (15)$$

$$F-Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (16)$$

Where tp, tn, fp and fn signify the number of a true positive, true negative, false positive and false negative.

Table 2. Validation analysis of Projected model with existing techniques

Methodology	Parameter Evaluation			
	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
DT	88.89	79.12	80.92	85.27
SVM	72.32	80.53	83.69	86.07
EML	81.43	82.07	90.06	89.28
SPCA-DT	87.16	81.04	84.17	83.08
SPCA-SVM	94.38	95.43	96.46	96.34
SPCA-EML	**96.90**	**97.84**	**98.20**	**98.67**

In terms of the analysis of precision, the DT, EML, and SPCA-DT all achieved results in the range of almost 81% to 88%, while the SVM achieved results in the range of 72% to 94% and the proposed model achieved results in the range of 96.70%. Existing methods, which are referred to as DT, achieved 79% precision, 80% recall, and 85.27% f-measure, but the proposed model achieved 97% precision, 98% recall, and 98.67% F-measure. The SVM was able to obtain 80%

precision, 83% recall, and 86% f-measure, whereas the SPCA-SVM was able to get 95% precision, approximately 965 recall, and F-measure. The EML had an accuracy rate of 82%, a recall rate of 90%, and an F-measure of 89%, whereas the SPCA-DT had an accuracy rate of 81.04%, a recall rate of 84.17%, and an F-measure of 83.08%. But the proposed model SPCA-EML achieved 97.84% of precision, 98.20% of recall and 98.67% of F-measure. The results of this analysis make it abundantly evident that the proposed model obtained greater performance than the strategies that are already in use. **Fig 4 to Fig 7** presents the graphical comparison of proposed model with existing techniques.

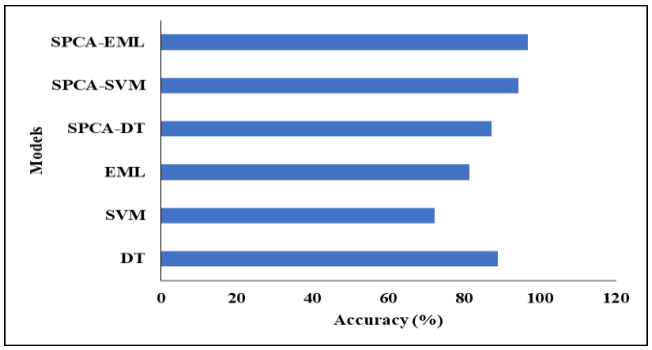

Fig 4. Accuracy Assessment

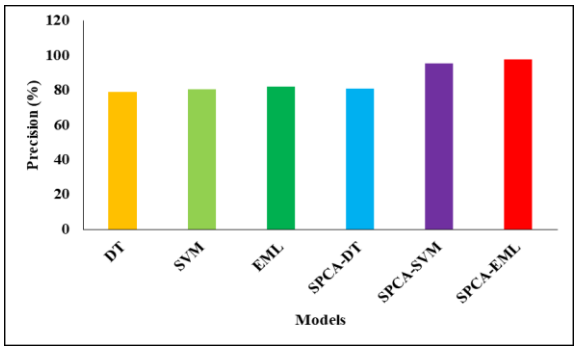

Fig 5. Precision Assessment

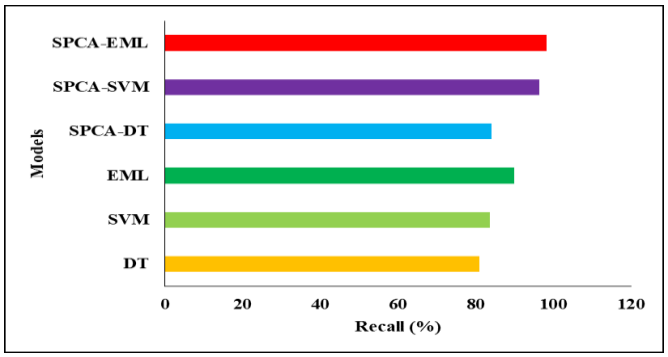

Fig 6. Recall Assessment

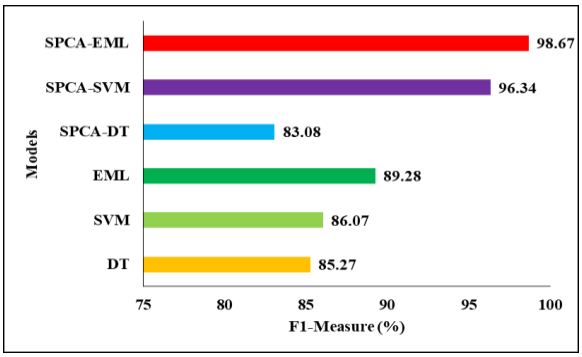

Fig 7. F-measure Assessment

VI. CONCLUSION

In this research, we presented a hybrid model for detecting floods using machine learning. The suggested model was trained using a dataset that included the keys such as temperature, humidity, rainfall, etc. that were detected using feature extraction and SPCA. When compared to the existing methodologies, which obtained about 94% to 96% accuracy and recall for the gathered datasets, the results showed that the suggested model achieved 96% accuracy and 98.20% recall. The population is not detected or assessed in this study so that aid may be provided and any damages can be calculated. Extending the dataset with additional landmarks and attributes will improve the system's accuracy in the long run. A further method for bettering the quality of the dataset is feature selection based on rivers, people, and automobiles. In addition to detection, various deep learning techniques can be investigated to perform flood inundation mapping, such as RNN and LSTM. In areas ravaged by natural disasters, the research can also be combined with methods for determining the most efficient routes for vehicles carrying rescue workers and emergency medical personnel. Any country's flood management, fire, and search and rescue agencies will benefit from this.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1]. G. S, D. T, and A. Haldorai, "A Supervised Machine Learning Model for Tool Condition Monitoring in Smart Manufacturing," Defence Science Journal, vol. 72, no. 5, pp. 712–720, Nov. 2022, doi: 10.14429/dsj.72.17533.

[2]. Munawar, H.S.; Hammad, A.; Ullah, F.; Ali, T.H. After the flood: A novel application of image processing and machine learning for post-flood disaster management. In Proceedings of the 2nd International Conference on Sustainable Development in Civil Engineering (ICSDC 2019), Jamshoro, Pakistan, 5–7 December 2019.

[3]. G. Lenderink and E. van Meijgaard, "Increase in hourly precipitation extremes beyond expectations from temperature changes," Nature Geoscience, vol. 1, no. 8, pp. 511–514, Jul. 2008, doi: 10.1038/ngeo262.

[4]. S. T. Ashley and W. S. Ashley, "Flood Fatalities in the United States," Journal of Applied Meteorology and Climatology, vol. 47, no. 3, pp. 805–818, Mar. 2008, doi: 10.1175/2007jamc1611.1.

[5]. M. Ptak and K. Konarzewski, "Numerical Technologies for Vulnerable Road User Safety Enhancement," Advances in Intelligent Systems and Computing, pp. 355–364, 2015, doi: 10.1007/978-3-319-16528-8_33.

[6]. C. D. F. van Ree, M. A. Van, K. Heilemann, M. W. Morris, P. Royet, and C. Zevenbergen, "FloodProBE: technologies for improved safety of the built environment in relation to flood events," Environmental Science & Policy, vol. 14, no. 7, pp. 874–883, Nov. 2011, doi: 10.1016/j.envsci.2011.03.010.

[7]. Arrighi, M. Pregnolato, R. J. Dawson, and F. Castelli, "Preparedness against mobility disruption by floods," Science of The Total Environment, vol. 654, pp. 1010–1022, Mar. 2019, doi: 10.1016/j.scitotenv.2018.11.191.

[8]. Guha-Sapir, D.; Below, R.; Hoyois, P. EM-DAT: International Disaster Database; Université Catholique de Louvain: Brussels, Belgium, 2015.

[9]. S. Qayyum, F. Ullah, F. Al-Turjman, and M. Mojtahedi, "Managing smart cities through six sigma DMADICV method: A review-based conceptual framework," Sustainable Cities and Society, vol. 72, p. 103022, Sep. 2021, doi: 10.1016/j.scs.2021.103022.

[10]. R. Sankaranarayanan, K. S. Umadevi, N. Bhavani, B. M. Jos, A. Haldorai, and D. V. Babu, "Cluster-based attacks prevention algorithm for autonomous vehicles using machine learning algorithms," Computers and Electrical Engineering, vol. 101, p. 108088, Jul. 2022, doi: 10.1016/j.compeleceng.2022.108088.

[11]. F. Ullah, S. Qayyum, M. J. Thaheem, F. Al-Turjman, and S. M. E. Sepasgozar, "Risk management in sustainable smart cities governance: A TOE framework," Technological Forecasting and Social Change, vol. 167, p. 120743, Jun. 2021, doi: 10.1016/j.techfore.2021.120743.

[12]. H. S. Munawar, S. Qayyum, F. Ullah, and S. Sepasgozar, "Big Data and Its Applications in Smart Real Estate and the Disaster Management Life Cycle: A Systematic Analysis," Big Data and Cognitive Computing, vol. 4, no. 2, p. 4, Mar. 2020, doi: 10.3390/bdcc4020004.

[13]. M. Peng, S. Garg, X. Wang, A. Bradai, H. Lin, and M. S. Hossain, "Learning-Based IoT Data Aggregation for Disaster Scenarios," IEEE Access, vol. 8, pp. 128490–128497, 2020, doi: 10.1109/access.2020.3008289.

[14]. M. Khalaf et al., "IoT-Enabled Flood Severity Prediction via Ensemble Machine Learning Models," IEEE Access, vol. 8, pp. 70375–70386, 2020, doi: 10.1109/access.2020.2986090.

[15]. P. Pandey and R. Litoriya, "Elderly care through unusual behavior detection: A disaster management approach using IoT and intelligence," IBM Journal of Research and Development, vol. 64, no. 1/2, pp. 15:1-15:11, Jan. 2020, doi: 10.1147/jrd.2019.2947018.

[16]. S. Miau and W.-H. Hung, "River Flooding Forecasting and Anomaly Detection Based on Deep Learning," IEEE Access, vol. 8, pp. 198384–198402, 2020, doi: 10.1109/access.2020.3034875.

[17]. S. Khan, K. Muhammad, S. Mumtaz, S. W. Baik, and V. H. C. de Albuquerque, "Energy-Efficient Deep CNN for Smoke Detection in Foggy IoT Environment," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9237–9245, Dec. 2019, doi: 10.1109/jiot.2019.2896120.

[18]. P. Bo, S. Fenzhen, and M. Yunshan, "A Cloud and Cloud Shadow Detection Method Based on Fuzzy c-Means Algorithm," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1714–1727, 2020, doi: 10.1109/jstars.2020.2987844.

[19]. K. Khuen and A. Zourmand, "Fuzzy Logic-Based Flood Detection System Using Lora Technology," 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Feb. 2020, doi: 10.1109/cspa48992.2020.9068698.

[20]. K. Vinothini and S. Jayanthy, "IoT Based Flood Detection and Notification System using Decision Tree Algorithm," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), May 2019, doi: 10.1109/iccs45141.2019.9065799.

[21]. S. Rani, G. N. Jayalakshmi, and V. P. Baligar, "Low Cost IoT based Flood Monitoring System Using Machine Learning and Neural Networks: Flood Alerting and Rainfall Prediction," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Mar. 2020, doi: 10.1109/icimia48430.2020.9074928.

[22]. M. Ahmadlou et al., "Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks," Journal of Flood Risk Management, vol. 14, no. 1, Dec. 2020, doi: 10.1111/jfr3.12683.

[23]. X. Guan, Y. Zang, Y. Meng, Y. Liu, H. Lv, and D. Yan, "Study on spatiotemporal distribution characteristics of flood and drought disaster impacts on agriculture in China," International Journal of Disaster Risk Reduction, vol. 64, p. 102504, Oct. 2021, doi: 10.1016/j.ijdrr.2021.102504.

[24]. H. S. Munawar, F. Ullah, S. Qayyum, and A. Heravi, "Application of Deep Learning on UAV-Based Aerial Images for Flood Detection," Smart Cities, vol. 4, no. 3, pp. 1220–1243, Sep. 2021, doi: 10.3390/smartcities4030065.

[25]. E. Jenifer, A. Aparna, N. Sudha, and A. Kumar, "AgriFloodNet: a dual patch CNN architecture for mapping flooded agricultural lands via bi-temporal multi-sensor images," Geocarto International, vol. 37, no. 26, pp. 13618–13637, Jun. 2022, doi: 10.1080/10106049.2022.2082549.

[26]. H. A. Zurqani, A. Al-Bukhari, A. O. Aldaikh, K. I. Elfadli, and A. A. Bataw, "Geospatial Mapping and Analysis of the 2019 Flood Disaster Extent and Impact in the City of Ghat in Southwestern Libya Using Google Earth Engine and Deep Learning Technique," Environmental Applications of Remote Sensing and GIS in Libya, pp. 205–226, 2022, doi: 10.1007/978-3-030-97810-5_10.

[27]. S. Fahad, F. Su, S. U. Khan, M. R. Naeem, and K. Wei, "Implementing a novel deep learning technique for rainfall forecasting via climatic variables: An approach via hierarchical clustering analysis," Science of The Total Environment, vol. 854, p. 158760, Jan. 2023, doi: 10.1016/j.scitotenv.2022.158760.

[28]. R. Subha, A. Haldorai, and A. Ramu, "An Optimal Approach to Enhance Context Aware Description Administration Service for Cloud Robots in a Deep Learning Environment," Wireless Personal Communications, vol. 117, no. 4, pp. 3343–3358, Feb. 2021, doi: 10.1007/s11277-021-08073-3.

[29]. Lavigne et al., "Instrumental lahar monitoring at Merapi Volcano, Central Java, Indonesia," Journal of Volcanology and Geothermal Research, vol. 100, no. 1–4, pp. 457–478, Jul. 2000, doi: 10.1016/s0377-0273(00)00151-7.

[30]. Suwarno, A. Ma'arif, N. Maharani Raharja, A. Nurjanah, J. Ikhsan, and D. Mutiarin, "IoT-based Lava Flood Early Warning System with Rainfall Intensity Monitoring and Disaster Communication Technology," Emerging Science Journal, vol. 4, pp. 154–166, Dec. 2021, doi: 10.28991/esj-2021-sp1-011.