Machine Learning Based Real Time-Heuristic Sensor Data Analytics For Early Warning Prediction

Shelendra Pal (shelendrapal12@outlook.com)
Teerthanker Mahaveer University
Divyendu Kumar Mishra
Veer Bahadur Singh Purvanchal University
Anandakumar Haldorai
Sri Eshwar College of Engineering
L. Rama Parvathy
Saveetha School of Engineering
S. Janupriya
K Ramakrishnan College of Engineering
D.Vijendra Babu
Aarupadai Veedu Institute of Technology

Research Article

Keywords: Landslide, Warning prediction, Artificial Neural Networks (ANNs), Back-Propagation Neural Network (BPNN)

Posted Date: October 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1012679/v1

Machine learning based real time-heuristic sensor data analytics for early warning prediction

[1]Dr Shelendra Pal, [2]Dr. Divyendu Kumar Mishra, [3]Dr. Anandakumar Haldorai, [4]Dr. L. Rama Parvathy

[5]S. Janupriya, [6]Dr. D.Vijendra Babu

[1]Assistant Professor, Teerthanker Mahaveer University (TMU), Moradabad U.P.

shelendrapal12@outlook.com

[2]Assistant Professor, Department of CSE, Veer Bahadur Singh Purvanchal University, Jaunpur, Uttar Pradesh, India.

divyendu01mishra@gmail.com

[3]Professor (Associate), Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India-641202

anandakumar.psgtech@gmail.com

[4]Professor, CSE Department, Saveetha School of Engineering, SIMATS, Chennai-602105.

ramaparvathyl.sse@saveetha.com

[5]Assistant Professor, Dept of ECE, K. Ramakrishnan College of Engineering, Kariyamanickam Road, Samayapuram, Tiruchirappalli, Tamil Nadu -621112

janupriya.ece@krce.ac.in

[6]Professor, Department of Electronics & Communication Engineering, AarupadaiVeedu Institute of Technology, Vinayaka Mission's Research Foundation, Paiyanoor-603 104. Tamil Nadu, India.

drdvijendrababu@gmail.com

Abstract

Landslides have the potential to cause significant property damage as well as fatalities. Landslides are identified by real-time heuristic data analysis acquired using wireless sensor networks (WSNs) in changing environments. People can abandon dangerous locations earlier when landslides are forecasted. In this paper, the early warning prediction system developed using machine learning; Artificial Neural Networks (ANNs) provide precise predictions. The weight coefficients of ANN can be adjusted exactly enough by network functional training. In the case of unbalanced data distribution, the proposed ANN model is unable to learn the sample data pattern. This results in incorrect prediction and therefore, a switching method is utilized to switch between alternative predictors based on the current environmental state. Furthermore, the proposed model has been developed to forecast and compensate for errors during the prediction phase. Thus, the proposed

model can enhance precise prediction, and an early warning prediction system of landslides can issue warnings 44.2 minutes before a landslide occurrence.

Index Key: Landslide, Warning prediction, Artificial Neural Networks (ANNs), Back-Propagation Neural Network (BPNN).

1. INTRODUCTION

Natural calamities are wildly unpredictable occurrences caused by natural phenomena. Natural catastrophes can result in death, property destruction, and financial losses. Since 1900, depending on data supplied by international health agencies, there have been approximately 14 million victims of all sorts of catastrophes [1]. Children are among the most sensitive individuals who are directly affected by catastrophes. Disaster victims originate from a variety of nations, and calamities can put many children's lives in jeopardy [2]. Disaster risk minimization is a theory that describes that communities may decrease disaster-related impact and casualties. Recognizing the disaster vulnerability is an instance of risk mitigation as per the Sendai paradigm [3] to reduce disaster developed by United Nations International Strategy. As per Goswami's study [4], the goal of disaster administration is to reduce the number of casualties, to preserve victims quickly, to remove people to safe locations, to rehabilitate the damages quickly, and to provide first aid instantaneously. Big data techniques, machine learning techniques, and deep learning techniques are the current modern technologies. Big data, as per Hashem's study, is a collection of methodologies and technologies that necessitates a new type of unification to discover enormous hidden values generated from complicated, diversified, and massive databases. The benefit of employing big data is that it may be used to discover patterns derived from data analytics as well as the development of hidden data [5].

Almost the previous year, catastrophic disasters including floods, tsunamis, earthquakes, cyclones, landslides, as well as numerous pandemics, have affected over 2.7 billion people. In the earlier, catastrophes have claimed the lives of many people; the worst disaster in New Guinea would be an earthquake that displaced about 58,350 people based on the displacement tracking matrix [6]. The flooding in China during July 1931 claimed the lives of 4, 000, 000 people, making it the worst natural catastrophe in history. Pandemics relate to the fast disease spread across a large zone, while disasters are generally physical ecological alterations. Several pandemic epidemics have also occurred across the universe. These scenarios are unpredictable and complicated, putting human life, the ecosystem, and a country's economics in danger. As a result, every country would want to use the most effective and precise methods to cope with such situations. The approach employed to forecast the catastrophe or pandemic's expected repercussions is crucial to its control. The resources are employed for more precise predictions and knowledge. Computer science advancements have made a vast number of data available to disaster management agencies. Because such data is frequently unstructured, cleaning and processing large amounts of information can be difficult.

Many individuals have been severely harmed as a result of the absence of a competent pandemic and disaster management system. It is impossible to foresee a disaster accurately, and victims have been neither relocated from the disaster region promptly. People have not been given mitigating measures after the catastrophe. Moreover, in the pandemic, effective measures to control the subsequent spread of the virus are not taken. To address these concerns, this article presents a thorough examination of all available processes and strategies that may be used throughout the post-disaster and pre-disaster periods to reduce damages to the bare minimum. The systems created to aid in hazard prediction must be able to withstand the obstacles which a hazard managerial system may face [7]. For example, in the event of a hurricane or sandstorm, a system's eyesight may be harmed by dusty particles or automated driving systems must be reliable in foggy circumstances [8].

Because of communication lack in case of disaster is the most important issue to be faced. Several difficult tasks involve optimizing the proportion of people secured while a pandemic and disaster condition, trying to evacuate people at the appropriate time, recognizing vulnerable locations for pandemic distribution, approaching the most influenced people/location and offering them adequate resources, calculating the economic loss, and several others [9]. Huge amounts of data are frequently presented to decision-makers, who must make forecasts and choices as fast as feasible [10, 11]. For medical uses, deep learning algorithms may be utilized for picture categorization and three-dimensional segments [12]. Machine learning (ML) is lately evolved as among the most important computer techniques, and it is progressively being used in everyday routine and a variety of industries [13]. ML is a type of artificial intelligence (AI) that utilizes algorithms to generate forecasts based on the given data features.

The network has become increasingly autonomous in the recent period of many other developing techniques including IoT, Unmanned Aerial Vehicles (UAVs), and satellite-based innovation. Many local considerations, including power control, bandwidth choice, data rate choice, and user connection to a base station, are required in these kinds of systems. During uncertain as well as unpredictable situations, we may utilize machine learning methods to solve these challenges and reduce human intervention. To conclude, ML algorithms outperform current technology in the following ways:

• ML algorithms may quickly analyze large amounts of data and discover patterns. Furthermore, machine learning techniques make it simple to analyze diverse sorts of data. The use of machine learning in everyday life, including traffic forecasts, video monitoring, and online customer service, has boosted its appeal.

• In machine learning, rule-based techniques can aid in the detection of false communications. Human interaction and decision-making are reduced when machine learning algorithms are used. Such methods aid in the prevention of rumors, particularly in the instance of man-made calamities.

• When the amount of data grows, the performance of machine learning techniques continues to improve. If the amount of information in an earthquake forecasting system grows, the algorithm's capacity to detect grows as well.

• Multi-dimensional data may be handled by ML algorithms, which can also discover outliers within data collection. Outlier evaluation is a useful approach in circumstances with high risks. Instead of eliminating every outlier, we must pay particular attention to them until attempting to forecast very improbable occurrences such as pandemics or disasters.

In such circumstances, a variety of machine learning techniques may be used to produce quick and accurate choices. The next parts of this paper deal with how such algorithms may be used to produce more accurate judgments. This model may be used in a wide range of situations and can assist to reduce human intervention [14]. Even though numerous studies have been conducted on the usage of machine learning algorithms, few have focused on their applications in a pandemic as well as disaster management. In this paper, a natural disaster landslide is predicted in the earlier stage and the warning is provided using machine learning, Artificial Neural Network (ANNs). The considered landslide is evaluated using physical learning such as rainfall and soil moisture and the safety factor is estimated for the corresponding data.

2. STATE-OF-THE-ART

We can establish the trend and fix the issue using this data analytics. By forecasting the incident or even using it as an early warning system. We have to employ data sources, and an algorithmic model for training and testing the data to tackle this issue. The majority of the research papers concentrate on the disaster managing stage and through Manzhu's study the corresponding data resources are acquired, but they don't discuss how the model/algorithm performed in the evaluations. The data source analysis as well as the methodology used to address the issue is the most significant component of earlier recognition and prediction.

In the assessment of communal membership for datasets, categorization algorithms are employed to categories a data element under any one of the various predefined groups [15]. Since categorization is a well-established paradigm in machine learning, typically suffers from challenges like lacking data management. Missing information collection values can cause issues in both the training process and categorization process. Techniques like disregarding skipped data or replacing extremes with plausible data can help solve this problem [16]. Before categorization, weights can be applied to the closest neighbors to improve the KNN approach's accuracy. This algorithmic approach is simple to use and possesses a short execution time. It is independent of a specific requirement to build a model, tweak a few parameters, or make any additional assumptions. The approach can be utilized for categorization followed by regression. The KNN slows down the growth of data volume is the main disadvantage. KNN was contrasted to various algorithms for detecting acute respiratory illnesses by the researchers of [17]. KNN was also utilized by [18] to identify patients having influenza. In addition, the researchers of [20] utilized KNN to forecast the movements of infected individuals using data location from user's devices. To save money, the researchers of [20] utilized SVM to select the most significant crowd that perfectly represents pilgrim behavior for subsequent processing using fuzzy logic. Different kernels are chosen from SVMs may represent the nonlinear judgment limits. It is extremely resistant to overfitting, particularly in high-dimensional space.

SVMs are frequently employed to validate a calamity, as seen in [21]. Via sensor data, you may apply SVM to categorize Edge Histogram (EH) emergency level. When compared to different algorithms for identifying if a person is dealing with ISPA, SVM demonstrated the highest efficiency in the model in [17]. In addition, [22] discovered that SVM seems to have a higher precision for user position confirmation compared to wireless networks which rely on channel features information to function. In addition, the researchers of [23] propose using SVM to divide aerial photos into a massive flood as well as non-flood-affected regions. In addition, the researchers of [24] employed SVM to predict a disaster epidemic in a particular region.

The logistic regression functional cost is sigmoid, and it is dependent on probability. The functional cost is in the range of 0-1. Logistic regression was employed by the researchers of [25] and [26] to retrieve relevant post-disaster data from tweets. A relevant tragedy is represented by 1 in its functional cost, while its non-relevance is represented by 0. In addition, the authors of [27] employed logistic regression to calculate the probability of people surviving a calamity. It is estimated the connection among the survival rate as well as the disaster's severity. Internal nodes contain characteristics, branching provides formulas, while leaf nodes provide results in a decision tree. Many classification metrics exist; however, entropy, as well as the Gini index, is the most extensively employed. The researchers of [28] employed decision trees to identify sandstorms. In this research, two groups are identified: sandstorm and no. Moreover, decision trees were employed by the creators of [19] to establish the user’s position in an epidemic.

A random forest is an ensemble approach in which numerous decision trees have been trained simultaneously to generate a single result. Bagging is the term used to describe the convergence of decision trees. The researchers of [29] utilized a random forest to identify post-disaster alterations. Likewise, the researchers of [30] were employed to identify building issues. Random forests are utilized for more than only land cover categorization [31] since it is utilized for flood forecasting [32]. In estimating the number of individuals affected by influenza in common spots, the random forest algorithm in [33] outscored the others with 95% precision. It achieves great precision as an outcome of its advantage in merging the results of whole decision trees. Furthermore, according to [34], the random forest algorithm outperforms KNN, decision tree, and SVM with a precision of 77.8%.

Gradient boosting was employed by the researchers to forecast the number of individuals affected with Influenza [38]. In addition, the researchers of [35] utilized gradient boosting to forecast individual's return patterns after leaving the disaster-affected zone. With 86.4 % precision, this classifier surpassed all others. DNNs have the benefit of being able to learn features over time and adjust their output regularly. DNN was employed by the researchers to estimate a crowd evacuation path, and the network model was found to be 78 % accurate [36]. Furthermore, the researchers of [37] present a deep neural framework that is utilized to evacuate crowds using UAVs. DNN is used by the researchers of [38] as well as [39] to forecast the number of persons in a given location. Later, we'll look at a sort of DNN called a Convolutional Neural Network (CNN) that is frequently utilized for image/video analysis.

ANN is utilized to detect or compute the storm intensity [40] and it is involved in detecting the user's position when combined with IoT technique [41]. This allows for the improved identification of an infected user's area and the implementation of social distancing techniques. Neural networks are used to confirm a user's position [22]. The findings in [22] suggest that ANN is employed for categorization while data is inadequate for wireless network positioning confirmation. ANN is being used to enhance the precision of its model for predicting the number of persons in a given location [42]. The pixel orientation in an image is occasionally lost by ANN. It can make it difficult to categories aerial photos captured using UAVs. In comparison to ANN, CNN is more accurate at classifying and recording orientation. CNN emphasizes the risk in disaster-affected regions. Input, convolution includes many filtering tensors, pooling, as well as fully-connected layers that have been included in CNN models [43].

3. SYSTEM MODEL

If the shear stress in the down-slope is high, landslides happen. The Safety factor (SF) relates to the soil's stability, as illustrated in Equation (1). It is considered into account physical variables such as slope, rainfall, and soil conditions. With a pattern of each metric, it can readily anticipate landslides. Thus, these features are connected with the SF formula to forecast landslides. Depending on the SHALSTAB method, three sections of the SF value have been constructed to discriminate among the unsafe slope ranges. The stability class, as well as unstable class, includes the constant area as well as marginally actively stable in the non-constant area. Various training samples have been utilized to train distinct neural network predictions depending on the classes.

$$SF = \frac{C+\left(1-\frac{R\ \ \alpha\ \ \rho_w}{Tsin\theta\ \rho_s}\right)\rho_s gZcos^2\theta tan\phi}{\rho_s gZcos\theta sin\theta} \quad (1)$$

Here, the C represents the optimum coefficient; R represents the intensity of rainfall; T represents the soil transmissivity; The depth of soil is denoted as Z; the water density is represented as ρ_w ; the soil density is represented as ρ_s ; φ represents a material friction angle internally; θ represents the slope gradient; as well as the particular contributing area.

This paper proposes a whole method for landslide prediction including early warnings. The structure of the proposed Early Warning Prediction (EWP) is displayed in Figure 1. It is divided into two sections and they are physical elements and computation components. The physical elements contain the heuristic sensor data comprises real-time environmental data including soil moisture, rainfall, and slope. The real-time heuristic sensed data is combined by coordinating nodes and transmitted to the computation components via Zigbee transmitters. As explained below, the computation components include a SHALSTAB unit, a switch-based prediction unit, and finally, the precise early warning prediction.

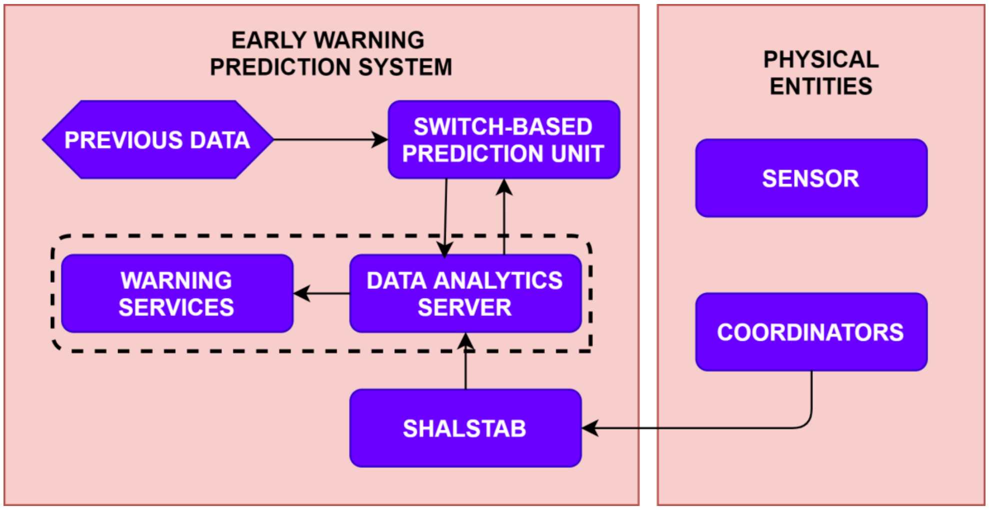

Figure 1: Landslide prediction architecture.

SHALSTAB Unit: The SHALSTAB unit uses Formula (1) to estimate the safety factor (SF) based on the heuristic sensed atmospheric information from physical elements, such as soil moisture, rainfall, and slope. The SF values are stored throughout time as $SF_{actual} = \{SF_0, SF_1, \dots, SF_t\}$ is feed as an input to the precise early prediction of a natural disaster such as landslides.

Switch-based Prediction Unit: The proposed early warning prediction system uses two prediction units to establish two distinct data patterns acquired from previous environmental informational data are the stable and unstable pattern. A proposed Neural Network classification is meant to forecast the upcoming class to switch among the multiple prediction units. Once the neural network classification switches the prediction units correctly, then its switch-based prediction unit can enhance prediction efficiency.

Early Warning Prediction System: The real-time heuristic sensed data analysis server, as well as alert services, makes up the precise early warning prediction system. The switch-based prediction unit described above is used by the analysis server to forecast natural disasters like landslides. $SF_{predict} = \{SF_{t+1}, SF_{t+2}, \dots, SF_{t+n}\}$ is a function that uses the input data feature FS_{actual} for future SF values prediction. There is a disparity among the expected SF as well as the actual SF obtained utilizing Formula (1) for each anticipated SF number. This discrepancy is referred to as prediction error. Depending on the specific trend of prediction mistakes, the analysis server of real-time heuristic sensed data will analyze the usability of the prediction unit. When the error reaches a certain threshold, it indicates that the prediction unit is not appropriate for the current environment, and therefore, the projected outcomes have substantial prediction errors. Further, the prediction unit is re-trained depending on a predefined error threshold and error metrics. When the predicted SF value SF_{t+k} is less than 1, thus, a landslide is expected to take place around k periods. As a result, alert services can send an alarm ahead of time.

A feed-forwarding Back-Propagation Neural Network (BPNN) is used as the prediction unit to anticipate the future Safety Factor (SF). The Widrow-Hoff learning principles are generalized into a multi-layer using a non-linear divergent transferring network function to develop a BPNN, which is a formidable computation system. Because of the intricate connection network, BPNN's great learning and thinking abilities are used to solve issues of high difficulty. The flowchart of a BPNN-based prediction system is shown in Figure 2.

To present a fundamental understanding of the sort of Artificial Neural Network (ANN), the fundamental computational technique of a BPNN is presented. The fundamental construction of a time-series oriented BPNN and contains three different layers namely the input layer, the hidden layer, and finally, the output layer. Within the time duration of t-n to t, time-series input corresponding to a particular characteristic including safety factor, soil moisture, and rainfall is used as input data being feed into the input layer of the network model. Each node pairs are linked via the weight in neighboring levels. The prior layer value of every node is being multiplied with weight and further aggregated to feed as an input to the subsequent layer's nodes. The incoming data are then sent through an activation function, which calculates the node's return value. The ultimate result is obtained by performing the aforementioned actions layer by layer.

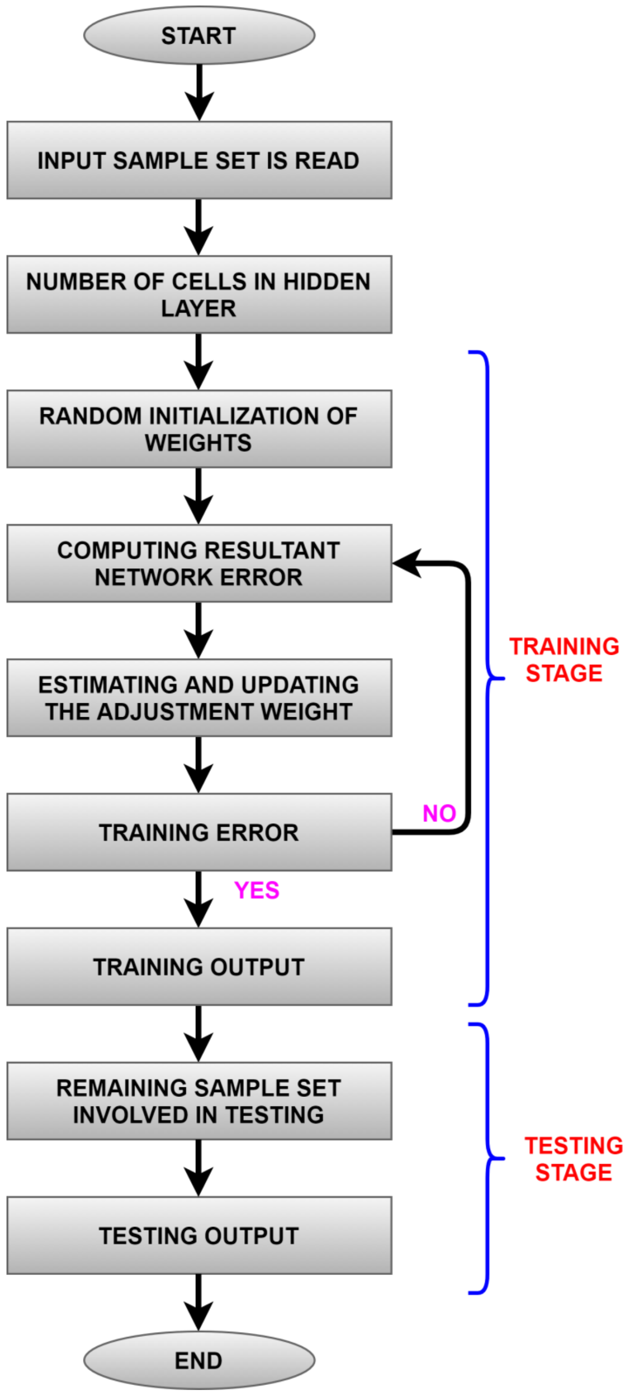

Figure 2: BPNN model.

A BPNN model in ANN is utilized for training and testing the sensed data for future SF prediction using the prior SF. The proposed back-propagation approach is used to train as well as to keep updating the weights before actual prediction to efficiently decide acceptable weights for distinct SF. The main mathematical formalism of a conventional one hidden layer feed-forwarding network containing N hidden neurons can be seen in Formula (2) for the training sample data $T_{data} = (x_i, p_i), i = 1, \ldots, N$, wherein $x_i = [x_{i1}, x_{i2}, \ldots., x_{in}] \in R^n$ is considered as an impact factor as well as $p_t = [p_{i1}, p_{i2}, \ldots., p_{ik}] \in R^k$ is considered as the training goal.

$$o_j = \sum_{i=1}^{\tilde{N}} g(w_i \cdot x_i), \qquad j = 1, \ldots.., N \tag{2}$$

Here, jth BPNN output is given as $o_j = [o_{j1}, o_{j2}, \ldots., o_{jk}]$, w_i represents the weight of the link among neurons available in the input layer and hidden layer, where g(x) is an activation function that defines how much modification the neuron's outcome must be depending on the total inputs. Formula (3) shows the sigmoid function, which is considered as an activation function employed in the proposed BPNN system. The sigmoid function is often known as the logistic function with a 0-1 output range.

$$g(x) = \frac{1}{1+e^{-x}} \tag{3}$$

The mismatch between the predicted and actual outcome is referred to as the prediction error, then the weights must be changed to minimize the prediction error rate. The Levenberg-Marquardt (LMA) approach converges the quickest and yields the smallest mean square error. Thus, the functional training of LMA is chosen to estimate the outcome networking error and modification weight W. Formula (4) illustrates it:

$$W_{k+1} = W_k - [J^T J + uI]^{-1} J^T \delta \tag{4}$$

In the kth iteration process, the matrix weight is denoted as W_k, Jacobian matrix containing the network error for corresponding weight is denoted as J and the 1^{st} order weight difference, the unit matrix is denoted I, constant is denoted as u, the resultant network error is denoted as δ, and LMA modifies the constant u effectively to minimize the resultant network error δ.

Formula (5) is used to estimate the training error E well following the training process to see if the training stage having gained integration. The training stage is resumed when it is larger than the threshold value of training error, $E_{threshold}$.

$$E = \frac{1}{N} \sum_{i=1}^{N} (T_i - O_i)^2, \tag{5}$$

Here, T_i represents the goal SF value of training 'i' sample; O_i represents the BPNN outputs SF value of training 'i' sample and N represents the total training samples.

To prevent a protracted training period in the recommended re-training method, the number of outputs i.e. Num_{output} is fixed to one. It indicates that by using the previous SF value as input, just a single prediction outcome will be achieved for each iteration process.

The hidden layer with the number of neurons is represented as Num_{neuron} is calculated using the below expression:

$$Num_{neuron} = \sqrt{Num_{input} * Num_{output}} \tag{6}$$

3.1. Early Warning prediction:

An early warning prediction system is built as illustrated in Figure 3 to ensure that the neural networks prediction model depending on the switch can be exact in environmental changes. It is separated into two phases namely a learning-based re-training flow and tuned horizon flow prediction.

3.1.1. Learning-based Retraining flow:

The error estimation is used to determine whether or not to retrain. The pairing average error $\left(SF_{actual}, SF_{predict}\right)$ in the error-estimation window (EEW) is calculated by the error determination processing. EEW_{now} as well as EEW_{prev} are used to estimate the average error (AVGE) of EEW, as well as the accumulation error (ACCE) mainly to verify the re-training constraints of both EEW. The two criteria within the prediction modeling require to be re-trained are given by expression (7) and (8), wherein C_{IE} denotes the interval error coefficient used to determine the short-term tolerated error limit, which in our approach is equivalent to the horizon prediction size. Because of the significant variation of the input pattern which the actual prediction modeling could not forecast, the prediction modeling will be re-trained whether the AVGE difference between two consecutive EEW is too significant. C_{AE} is the cumulative error coefficient that is utilized to define the long-term tolerated error limit. When the ACCE is extremely great when contrasted with the prediction modeling containing an average error, $AVGE_{model}$, then the prediction system will be re-trained because the prediction outcomes are becoming erroneous. This also means that the surrounding is altering throughout time, necessitating adaptation. Formulae (9) and (10) are used to determine $AVGE_{now}$ respectively.

$$AVGE_{now} > C_{IE} \times AVGE_{prev} \tag{7}$$

$$ACCE_{now} > C_{AE} \times AVGE_{Model} \tag{8}$$

$$AVGE_{now} = \frac{\sum\left|FS_{actual} - FS_{predict}\right|}{SIZE_{EEW}} \tag{9}$$

$$\begin{aligned} ACCE_{now} &= ACCE_{prev} + \Delta AVGE \\ &= ACCE_{prev} + \left(AVGE_{now} - AVGE_{prev}\right) \end{aligned} \tag{10}$$

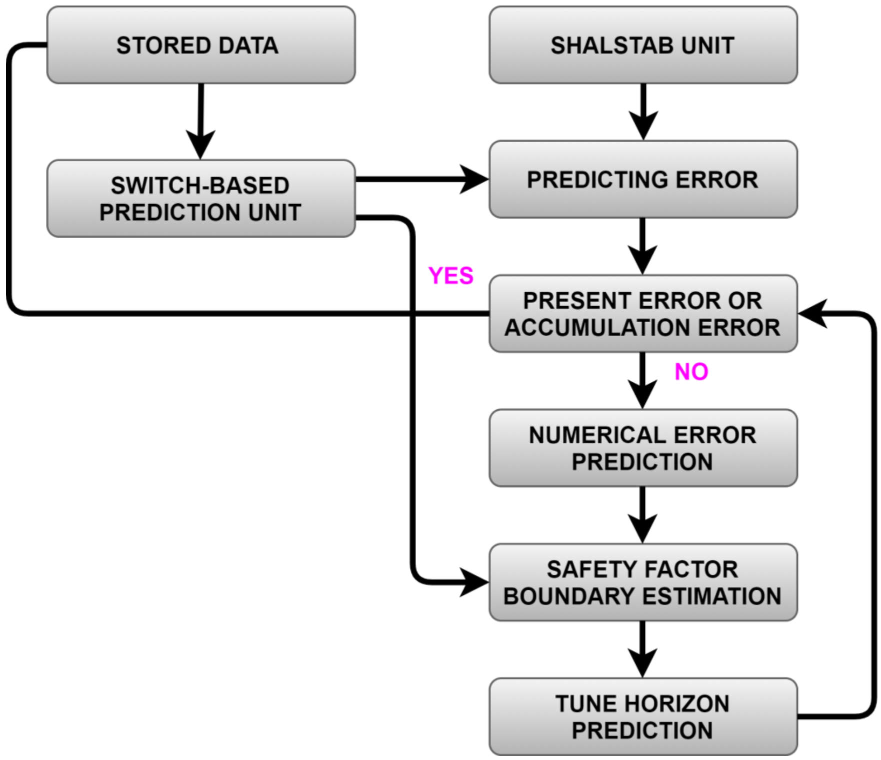

Figure 3: The flowchart of the proposed model.

3.1.2. Tune Horizon Prediction

The error-estimation data are used to fine-tune the horizon prediction in the re-training process. The following are the benefits of a variable-length horizon prediction: Landslides can effectively forecast ahead of time, and the percentage of incorrect predictions is minimized.

The prediction system is used to understand the inherent patterns for predicting the forthcoming mistakes, $ERR_{predict}$, because prediction errors are typically non-linear. The horizon prediction size is adjusted in this manner to forecast the landslides that occurred in the earlier stage. Equation (11) is used to predict the future goal spans, P_{future}. When the $P_{future} < 1$ span is large enough, the system may issue notifications ahead of time.

$$SF_{predict} - |ERR_{predict}| \leq P_{future} \leq SF_{predict} + |ERR_{predict}| \tag{11}$$

Error boundary assessment is required to determine if the horizon prediction size is tuned. Expression (12) may be used to determine the expected $Bound_{low}$ which is the lower bound, because the program already possesses the forecast error, $ERR_{predict}$, as well as the predicted SF, $SF_{predict}$. The below-mentioned criterion is involved in tuning is done, depending on the forecasted outcomes:

When there is neither of $Bound_{low}$ is smaller in contrast with a stable group with a minimum bound, the horizon prediction size is improved by 1 for each point time in the horizon prediction; When there is neither of $Bound_{low}$ is smaller in contrast with a stable group with a minimum bound, the horizon prediction's size is restored to a default value with each point time.

$$Bound_{low} = SF_{predict} - |ERR_{predict}| \tag{12}$$

We made our methods a bit more adaptable for the landslide forecasting event with variable-length horizon prediction based on the aforementioned rules.

4. EXPERIMENTAL DATASETS AND ANALYZATION

The proposed early warning prediction system for the landslide is discussed and the experimental databases that are utilized in the tests are provided initially. The outcomes of an experiment are then displayed. All experiments are done on a PC having 16GB RAM run using 64-bit $Windows^{®}$ 10 operating system, 3.40GHz CPU, i7-3770 $Intel^{®}$ $Core^{TM}$, as well as the $MATLAB^{®}$ software platform.

The Shen-Mu station is chosen as an experimental study for periodic environmental monitoring information. Rainfall, as well as soil moisture, have the most impact on landslides and likelihood can be predicted utilizing SF. Figures 4 and Figure 5 depict the monitoring data of rainfall and soil moisture from the Shen-Mu station. The correlations SF for corresponding monitoring data rainfall and soil moisture will be provided. The SF diminishes and then the slope gets unstable if rainfall and soil moisture increases.

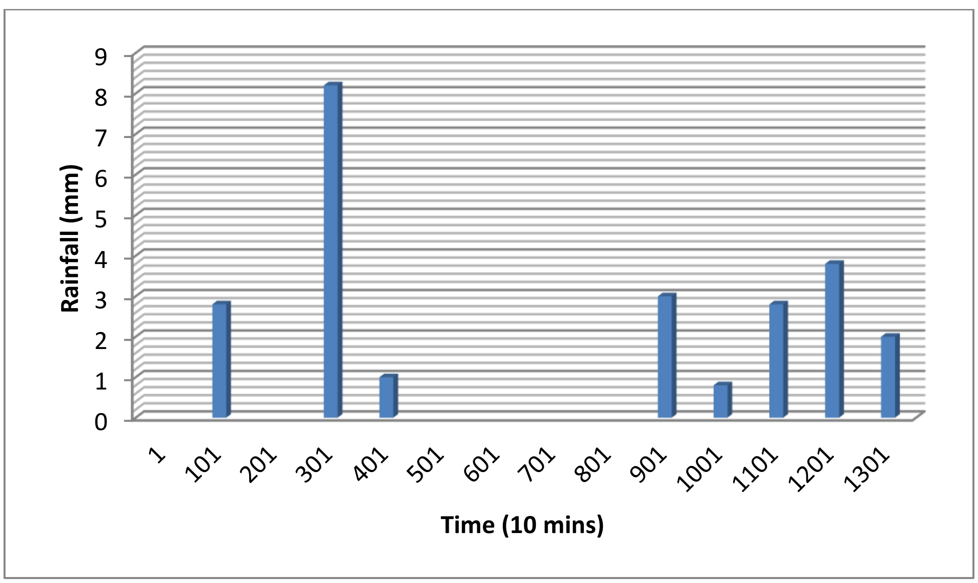

Figure 4: Rainfall prediction.

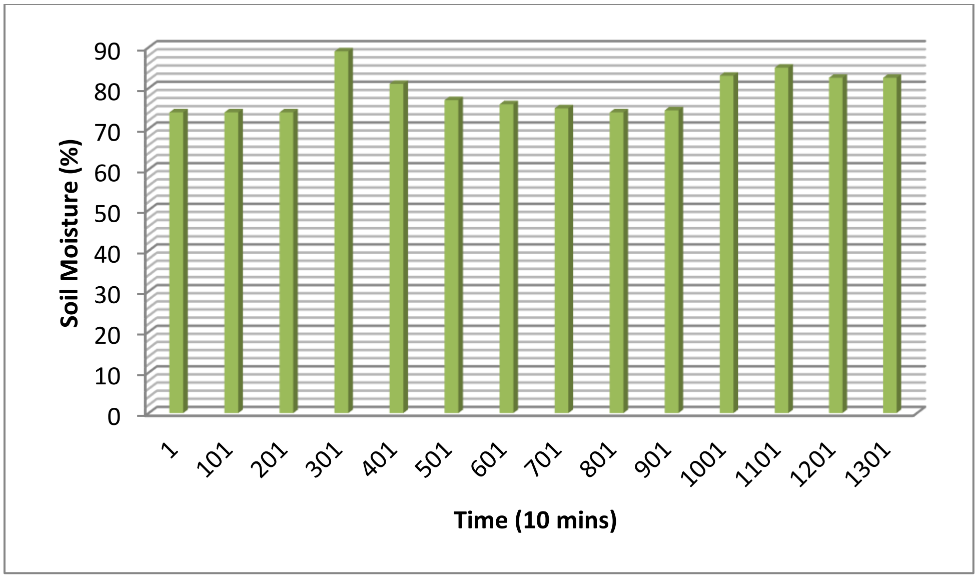

Figure 5: Soil moisture prediction

The Shen-Mu datasets are recorded once every 10 minutes. The system chose 14 sets of samples randomly, each with 1300 sample data. 75 % sample set is used for the training process whereas the remaining 25% is utilized in the testing process. The actual data are utilized as input; they are not normalized, since the SF had to be calculated using the SHALTAB, as shown in Formula (1).

The above-mentioned datasets are used to test the landslide prediction system in the experimental analysis. Various performance indicators are utilized to examine the suggested early warning prediction system model, and they are specified as follows:

Formula (13) defines Mean Absolute Percentage Error (MAPE), wherein n represents the number of prediction data, A_t represents the real value, and the predicted data is denoted as P_t.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|A_t - P_t|}{P_t} \tag{13}$$

Formula (14) gives the expression of Root Mean Squared Error (RMSE), wherein n represents the number of forecasted sample data, y_i represents the real SF value, and the predicted SF data is represented as $\widehat{y_i}$.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{14}$$

Formula (15) defines the Normalized Root Mean Squared Error (NRMSE), wherein $\bar{y}$ represents the mean output of real data.

$$NRMSE = \frac{RMSE}{\bar{y}} \tag{15}$$

For every 10 minutes, Figure 6 illustrates the real data and predicted safety factor. BPNN contains two prediction methods, one to study the stable class pattern whereas the other is to study unstable class patterns are offered. An event-class predictor is built to switch among two BPNN designs and deal with uneven data distribution to forecast the forthcoming class to take appropriate action. Thus, both stable class and unstable class patterns could be learned using an early warning prediction system using machine learning ANN.

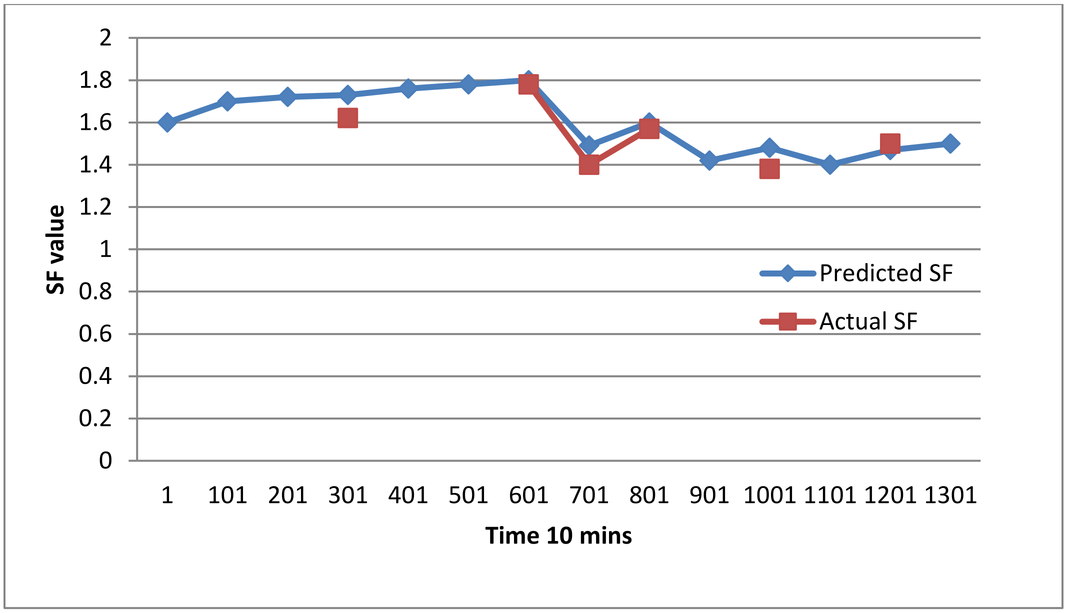

Figure 6: The proposed early warning prediction system.

For the early warning prediction system, the error metrics are assessed. The MAPE, as well as RMSE of our technique, are significantly reduced. The prediction method's forecast precision is higher. The proposed method's NRMSE is likewise higher than others, indicating that our predictions are more accurate. Although the proposed early warning prediction system is more precise, it takes a bit more time and computes resources. The experiment results suggest that the proposed early warning prediction system is a little slower; with a 1.205s time requirement and 23.90% CPU usage and 972kb RAM. This is due to the increased processing time and required resources by the proposed early warning prediction system to cope with the unbalanced data categorization and switching among predictors. The experimental time interval is 10 minutes, so there is plenty of time to cope with the process.

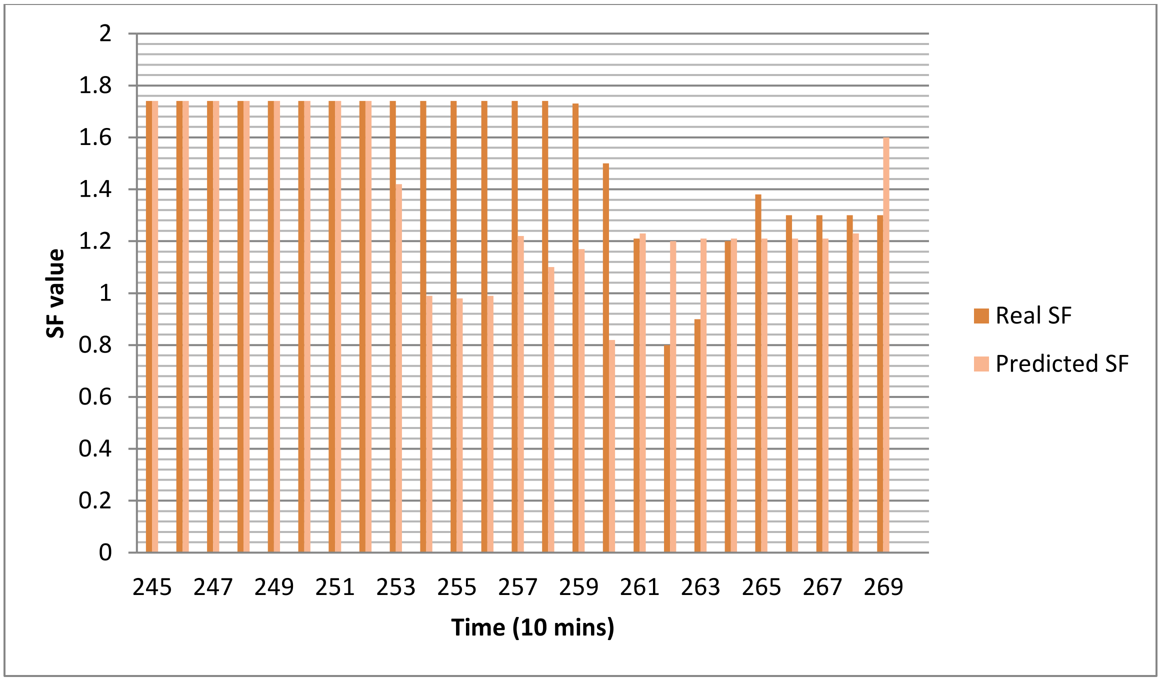

Figure 7: Early warning prediction of landslide occurrence.

We also use a best-case scenario to show how the suggested technology can predict the occurrence of landslides months ahead of time. The 8th prediction outcome at time 255 is SF 1, indicating that landslide may happen after completion of eight units timing, as depicted in Figure 7. The delay among two-point time in our tests is 10 minutes. As a result, the landslide occurrence might be forecast and alerted 80 minutes ahead of time.

5. CONCLUSION

The proposed machine learning ANN is a novel approach for landslide prediction in the earlier stage and provides a warning to take corrective action. This proposed method is capable of solving the challenges of imbalanced data, identifying the horizon prediction using the learned sample data, and enable the model re-training. The stable class and unstable class are balanced and categorized based on a safety factor of SHALSTAB. A BPNN class predictor is presented to address the issue of lower true positive rate, while BPNN with two predictors is built to study the stable class and unstable class patterns. A learning-based retraining flow and tuned horizon prediction are included in the proposed work. These improvements help to achieve the precise early warning prediction of landslides. Furthermore, BPNN is used to build the error model that will be used to forecast upcoming errors of our model and adjust for them during the prediction phase. Thus, the proposed model possesses significantly less MAPE and RMSE implying that it is more precise. Furthermore, our technique's NRMSE is higher indicating that the model is closer to the real situations. According to statistics, the proposed early warning prediction could send alerts 44.2 minutes before a landslide occurred.

Declarations

Conflict of interest : The authors declare that they have no conflict of interest.
Funding : Not Applicable
Availability of data and material: Not Applicable
Code availability: Not Applicable

REFERENCE

1. “Number of reported disasters by type.” [Online]. Available: https://ourworldindata.org/natural-disasters.
2. Tuswadi and T. Hayashi, “Disaster Prevention Education in Merapi Volcano Area Primary Schools: Focusing on Students’ Perception and Teachers’ Performance,” Procedia Environ. Sci., vol. 20, pp. 668–677, 2014.
3. “2015_43291_Sendaiframeworkfordrren_Disaster Reducton 2015-2030,” 2015.
4. S. Goswami, S. Chakraborty, S. Ghosh, A. Chakrabarti, and B. Chakraborty, “A review on application of data mining techniques to combat natural disasters,” Ain Shams Eng. J., vol. 9, no. 3, pp. 365–378, 2018.
5. I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, “The rise of ‘big data’ on cloud computing: Review and open research issues,” Inf. Syst., vol. 47, pp. 98–115, 2015.
6. Gianluca Rampolla del Tindaro, “Resident/Humanitarian coordinator report on the use of CERF funds, Papua new Guinea rapid response earthquake, 2018,” https://cerf.un.org/sites/default/ files/resources/18-RR-PNG-29464-NR01_Papua%20New% 20Guinea_RCHC.Report.pdf, online; accessed 12 September 2020.
7. X. Du, Y. Xiao, M. Guizani, and H.-H. Chen, “An effective key management scheme for heterogeneous sensor networks,” Ad Hoc Networks, vol. 5, no. 1, pp. 24 – 34, 2007, security Issues in Sensor and Ad Hoc Networks. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870506000412.
8. A. Mehra, M. Mandal, P. Narang, and V. Chamola, “Reviewnet: A fast and resource optimized network for enabling safe autonomous driving in hazy weather conditions,” IEEE Transactions on Intelligent Transportation Systems, pp. 1–11, 2020.
9. T. Yaqoob, H. Abbas, and M. Atiquzzaman, “Security vulnerabilities, attacks, countermeasures, and regulations of networked medical devices - A review,” IEEE Commun. Surv. Tutorials, vol. 21, no. 4, pp. 3723–3768, 2019. [Online]. Available: https://doi.org/10.1109/ COMST.2019.2914094
10. L. He, Z. Yan, and M. Atiquzzaman, “LTE/LTE-A network security data collection and analysis for security measurement: A survey,” IEEE Access, vol. 6, pp. 4220–4242, 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2792534.
11. M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, “Deep learning for iot big data and streaming analytics: A survey,” IEEE Communications Surveys Tutorials, vol. 20, no. 4, pp. 2923–2960, 2018.

12. G. Bansal, V. Chamola, P. Narang, S. Kumar, and S. Raman, “Deep3dscan: Deep residual network and morphological descriptor based framework for lung cancer classification and 3d segmentation,” IET Image Processing, vol. 14, no. 7, pp. 1240–1247, 2020.
13. Wikipedia contributors, “Machine learning — Wikipedia, the free encyclopedia,” 2020, [Online; accessed 30-April-2020]. [Online]. Available: https://en.wikipedia.org/w/index.php?title= Machine_learning&oldid=954000510.
14. R. Xie, I. Khalil, S. Badsha, and M. Atiquzzaman, “Collaborative extreme learning machine with a confidence interval for P2P learning in healthcare,” Comput. Networks, vol. 149, pp. 127–143, 2019. [Online]. Available: https://doi.org/10.1016/j.comnet.2018.11.002.
15. A. Soofi and A. Awan, “Classification techniques in machine learning: Applications and issues,” Journal of Basic & Applied Sciences, vol. 13, pp. 459–465, 08 2017.
16. Mehmed Kantardzic, “Data mining: Concepts, models, methods, and algorithms,” 2011. [Online]. Available: https://www.academia.edu/18807133/Data_Mining_Concepts_Models_Methods_and_Algorithms.
17. N. L. W. S. R. Ginantra, I. G. A. D. Indradewi, and E. Hartono, “Machine learning approach for acute respiratory infections (ISPA) prediction: Case study indonesia,” Journal of Physics: Conference Series, vol. 1469, p. 012044, feb 2020. [Online]. Available: https://doi.org/10.1088%2F1742-6596%2F1469%2F1%2F012044.
18. R. Yin, V. H. Tran, X. Zhou, J. Zheng, and C. K. Kwoh, “Predicting antigenic variants of h1n1 influenza virus based on epidemics and pandemics using a stacking model,” PLoS ONE, vol. 13, 2018.
19. S.-B. Cho, “Exploiting machine learning techniques for location recognition and prediction with smartphone logs,” Neurocomputing, vol. 176, 05 2015.
20. A. Namoun, A. Mir, A. Alkhodre, A. Tufail, A. Alrehaili, M. Farquad, M. Alwaqdani, T. Alghamdi, and M. Benaida, “A multi-agent architecture for evacuating pilgrims in panic and emergency situations: The hajj scenario,” Journal of Theoretical and Applied Information Technology, vol. 96, p. 20, 10 2018.
21. J. Ren and X. Gao, “Situation assessment model for uav disaster relief in the city,” in 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping, 2011, pp. 1–6.
22. A. Brighente, F. Formaggio, G. M. Di Nunzio, and S. Tomasin, “Machine learning for in-region location verification in wireless networks,” IEEE Journal on Selected Areas in Communications, vol. 37, no. 11, pp. 2490–2502, 2019.
23. J. Akshya and P. L. K. Priyadarsini, “A hybrid machine learning approach for classifying aerial images of flood-hit areas,” in 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1–5.

24. K. Mori, T. Wada, and K. Ohtsuki, “A new disaster recognition algorithm based on svm for eress: Buffering and bagging-svm,” in 2016 45th International Conference on Parallel Processing Workshops (ICPPW), 2016, pp. 22–30.
25. S. Sadhukhan, S. Banerjee, P. Das, and A. K. Sangaiah, “Producing better disaster management plan in post-disaster situation using social media mining,” in Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications. Elsevier, 2018, pp. 171–183.
26. N. Assery, Y. Xiaohong, S. Almalki, R. Kaushik, and Q. Xiuli, “Comparing learning-based methods for identifying disaster-related tweets,” in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019, pp. 1829–1836.
27. V.-M. Le, “Machine learning methods for optimization in multi-agent decision support system : application to sign placement for tsunami evacuation,” Ph.D. dissertation, 12 2016.
28. H. A. Shaiba, N. S. Alaashoub, and A. A. Alzahrani, “Applying machine learning methods for predicting sand storms,” in 2018 1st International Conference on Computer Applications Information Security (ICCAIS), 2018, pp. 1–5.
29. A. R. Joshi, I. Tarte, S. Suresh, and S. G. Koolagudi, “Damage identification and assessment using image processing on post-disaster satellite imagery,” in 2017 IEEE Global Humanitarian Technology Conference (GHTC), 2017, pp. 1–7.
30. A. Cooner, Y. Shao, and J. Campbell, “Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 haiti earthquake,” Remote Sensing, vol. 8, p. 868, 10 2016.
31. V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp. 93–104, 2012.
32. A. Stumpf and N. Kerle, “Object-oriented mapping of landslides using random forests,” Remote Sensing of Environment, vol. 115, no. 10, pp. 2564 – 2577, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425711001969.
33. F. Al Hossain, A. Lover, G. Corey, N. Reich, and T. Rahman, “Flusense: A contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, pp. 1–28, 03 2020.
34. T. Gupta, V. Nunavath, and S. Roy, “Crowdvas-net: A deep-cnn based framework to detect abnormal crowd-motion behavior in videos for predicting crowd disaster,” in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019, pp. 2877–2882.
35. T. Yabe and S. Ukkusuri, “Integrating information from heterogeneous networks on social media to predict post-disaster returning behavior,” Journal of Computational Science, vol. 32, 02 2019.

36. K. Shibata and H. Yamamoto, “People crowd density estimation system using deep learning for radio wave sensing of cellular communication,” in 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2019, pp. 143–148.
37. P. Chhikara, R. Tekchandani, N. Kumar, V. Chamola, and M. Guizani, “Dcnn-ga: A deep neural net architecture for navigation of uav in indoor environment,” IEEE Internet of Things Journal, pp. 1–1, 2020.
38. D. Bega, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, “Deepcog: Cognitive network management in sliced 5g networks with deep learning,” in IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, 2019, pp. 280–288.
39. I. Alawe, A. Ksentini, Y. Hadjadj-Aoul, and P. Bertin, “Improving traffic forecasting for 5g core network scalability: A machine learning approach,” IEEE Network, vol. 32, no. 6, pp. 42–49, 2018.
40. A. Buranasing and A. Prayote, “Storm intensity estimation using symbolic aggregate approximation and artificial neural network,” in 2014 International Computer Science and Engineering Conference (ICSEC), 2014, pp. 234–237.
41. L. Luoh, “Zigbee-based intelligent indoor positioning system soft computing,” Soft Comput., vol. 18, no. 3, p. 443–456, Mar. 2014. [Online]. Available: https://doi.org/10.1007/s00500-013-1067-x.
42. M. Polese, R. Jana, V. Kounev, K. Zhang, S. Deb, and M. Zorzi, “Machine learning at the edge: A data-driven architecture with applications to 5g cellular networks,” IEEE Transactions on Mobile Computing, p. 1–1, 2020. [Online]. Available: http://dx.doi.org/10.1109/TMC.2020.2999852.
43. F. Alidoost and H. Arefi, “Application of deep learning for emergency response and disaster management,” 02 2018.