



# ENRICHING FLOOD AND LANDSLIDE PREDICTION THROUGH DEEP LEARNING MECHANISM

Mehul Chavan<sup>1</sup>, Shraddha paliwal<sup>2</sup>, Reetika Raina<sup>3</sup>, Prajakta Londe<sup>4</sup>, Pournima Kale<sup>5</sup>,  
Dr.Pramod Jadhav<sup>6</sup>

<sup>1</sup>IT Dept., G.H Raison Institute of Engineering and Technology, Pune

<sup>2</sup>IT Dept., G.H Raison institute of Engineering and Technology, Pune

<sup>3</sup>IT Dept., G.H Raison institute of Engineering and Technology, Pune

<sup>4</sup>IT Dept., G.H Raison institute of Engineering and Technology, Pune

<sup>5</sup>IT Dept., G.H Raison institute of Engineering and Technology, Pune

<sup>6</sup>IT Dept., G.H Raison institute of Engineering and Technology, Pune

## ABSTRACT

Despite their widespread distribution, regular occurrence, and numerous, geographical, and devastating qualities, landslide catastrophes have inflicted immeasurable damages to the country's economy, lives, and properties. India has intrinsic and environmental causes for landslide incidence due to its tropical humid long rainy season and unique geographical position. The rainfall-induced landslide's realization is of tremendous importance. Because the fundamental circumstances for both calamities are identical, landslides frequently precede floods or vice versa. Because of its crucial significance in decreasing economic and lives losses, flood prediction is among the most challenging, complicated, and significant challenges in engineering. Forecast accuracy has improved in recent years has resulted of advances in data collecting via satellite observations, as well as advances in technology and computational methods for uncertainty analysis and interaction. Therefore an approach for landslide and flood prediction is the need of the hour. For this purpose the proposed work is incorporates the LSTM neural network which is basically a time based model to predict the flood and landslide. This process is catalyzed by the use of K- nearest neighbor classification model which is ensemble with LSTM to produce good results of root mean square error.

**KEYWORDS:** K Nearest Neighbors, Long Short Term Memory, Flood and Landslide prediction.

## I. INTRODUCTION

Variations in the climate associated with global warming are producing extreme meteorological scenarios including such heavy rain and flooding. Numerous locations throughout the world are dealing with a dramatic surge in water depth, which has resulted in infrastructural destruction and civilian deaths. This circumstance, characterized as a flood event, becomes even more critical for emerging and much less advanced nations, wreaking havoc on their economies. On a worldwide platform, floods are said to have resulted in a massive loss of life and impacted a large number of people.

Researchers concentrate on the utilization of Sensor Networks amongst some of the various ways was using to observe flooding. Numerous variables may be observed in real-time, including such water table, temperature, and

radiation. This massive amount of data is evaluated in order to estimate the likelihood of a flood, and other such estimates can save lives. As a result, it is critical to simplify this operation in order to reduce human involvement and eliminate unwanted inefficiencies.

Flooding tragedies continuing to strike numerous regions of the world, resulting in a large quantity of deaths and property damages. Flooding predictions must be made in a reasonable manner in order to establish an early warning signals, notifying the people to the impending catastrophe. By giving adequate time for evacuating or infrastructure modifications, ample warning can still save lives and livelihoods. Sophisticated computational modelling can produce a significant time change. Various versions are developed for anticipating multiple possibilities such as floods, precipitation, and so on. Machine intelligence approaches such



as the Artificial Neural Network and the SVM Classifiers have indeed been established as effective and adaptable computing tools. The reluctance to give extremely precise findings owing to values outside the spectrum, on the other hand, has demonstrated to be a significant constraint.

Weather prediction is only possible for a duration of 5 days. Nevertheless, in the event of a rapid shift in the environment, this information may alter. This information may not be readily available. The Kerala flooding were indeed a significant calamity in which a lack of data demonstrated to be a serious impediment for flood forecast. The dams' discharge velocity could not have been estimated, resulting in massive infrastructural damage. This demonstrates the necessity for a reliable flood forecast system that considers elements such as precipitation, river flow velocity, and volume of water. Floods may be predicted using these three major variables. When working with additional real-world data, nevertheless, the findings showed lower predictive performance.

Landslides are one of the many natural disasters that may devastate management and facilities lives. Precipitation is one of the main reasons of the landslides. Landslides are indeed prevalent in steep terrain as a result of excessive rainfall. The precipitation cycle, volume, and duration have been shifting from day to day as the climate is affected. Similarly, the severity and frequency of landslides caused by downpours are rising over the world. To minimize the hazard to human lives, continual landslide observation and forecast are essential. Landslide vulnerability is the probability of a landslide happening in a given location based on regional geological factors, and it may be used to forecast wherever landslides will strike in the research area.

Contributing variables include natural or man-made elements that might influence landslide vulnerability. Environmental variables and activating elements are two types of contributing elements. External parameters include topological aspects such as aspect, angle, height, elevation, etc., geophysical characteristics such as fracture, rock properties, etc., land use / cover whether there are roads, structures, foliage, etc., drainage characteristics such as spillage, riverbed, etc., and so on. Activating elements, such as rainfall and seismic activity, have a disjointed and abrupt influence on the occurrence of landslides.

There is indeed a plethora of study material on the subject of landslides. Landslide surveillance, forecasting, and mitigation are the focus of several research organizations. Multiple machine learning systems, like as neural network models, Bayesian networks, and evolutionary algorithms, can also be used to forecast the likelihood of landslides. These techniques are not real-time since they demand a lot of calculation time to forecast the outcome. Therefore, a number of different researches for the purpose of prediction of flood and landslides are effectively outlined in this survey paper.

This Research paper segregates the section 2 for the evaluation of the past work in the configuration of a literature survey, Section 3 elaborates the techniques used in the implementation and Section 4 discusses the obtained results. Finally, section 5 provides the conclusion and the future work.

## II. RELATED WORKS

Mujie Li [1] expresses that to explore the impact of premature precipitation on collapses in Bazhou district, the system retrieved remotely sensed 3B42 rainfall output information since the day of incidence to the very first 4 days following rainfall-induced landslides. The precipitation data from the most recent persistent precipitation of the rainfall-induced collapse is then utilized. The precipitation I-D criterion model was built using information to determine the model equation components. Furthermore, relying on the rainfall model equation functional images, the researchers proposes a technique for proactively forecasting landslide rates. By combining the landslide vulnerability research technique with the rainfall approach, the researchers suggest a climatological advance warning framework of rainfall-induced landslides in Bazhou area. The scientists used actual information from the database to develop the quantitative methodology, which is determined by the number of probabilistic landslides and precipitation. Finally, a model for climatic advance detection was developed.

Darmawan Utomo [2] narrates that the architecture of the system is divided into two parts: Physical Entity and Computation Components, according to the author. Atmospheric information is collected by sensor network in the Physical Entities. Moisture content and rainfall amount are examples of ecological information. The supervisor then integrates the perceived data from the sensors and sends it to the Computation Components. Damage Detection Component, Pre - processing stage Subsystem, Deep Neural Network-based Information Prediction Component, Information Compensation Subsystem, and Statistical Data are included in the Computing Components. By incorporating large datasets (such as annual rainfall, soil humidity, gradient, and gradient orientation) in an LSTM model to measure the FS, an implementation of Temporal Data Restoration framework for landslide susceptibility mapping has been postulated. When prepared employing Tanh activation function with numerous epochs, the LSTM framework produces better performance than the LE.

Chaw Chaw Khaing [3] The landslide warning mechanism, which uses the internet of things and weather patterns, is competent of incredible and constant monitoring of various landslide characteristics such as annual rainfall, relative humidity relevance, moisture content, and heat, as well as alerting local residents to evacuate the area in the event of an approaching landslide. The technique can assist various land sliding risky zones with much the same route, although the based on soil value may change. It would also aid in the



development of a precise and dependable landslide detection and surveillance system relies on empirical analysis of various rainfall and groundwater factors, as well as other categorization methods. This technique aids in the protection of those who live near land-slide prone regions. One of the most difficult study topics in the realm of seismic investigation is surveillance and alerting the landslide mechanism.

Shu Sun [4] addresses the need of studying the evolutionary status and management of landslides as among China's largest catastrophic geological calamities. This work provides a unique control approach based on the estimation of development condition, taking into account the complexities and non-linearity of landslide systems. To begin, the MTLSTLM model is developed to anticipate the system's state level based on characteristics associated and to determine whether the mechanism needs to be regulated. Furthermore, when the level forecast is hazardous, the precipitation frequency estimator relying on bootstrapping is intended to find the safe frequencies of the control variability. Then, at the durations, the acceptable rainfall quantity is picked and re-verified. The findings of a simulation on a genuine landslide instance show that the down-level management strategy is effective.

Haoran Zhang [5] For the landslide vulnerability evaluation in the Pearl River Delta, the information fusion technique centered on LSI, AHP, and multifactor space superposed assessments is utilised to incorporate past and present landslide statistics, proposed mathematical assessment, and professional knowledge. The research region was categorized into five vulnerability classifications depending on landslide hazard: extremely low vulnerability, low vulnerability, moderate vulnerability, high vulnerability, and very high vulnerability. The most significant influence on landslides appears to be topography characteristics, preceded by engineered geology lithology and land use categories. The relatively high vulnerability classifications were mostly found in the areas of STD of height, ruggedness, and curvature, sand-shale material and coal, layered clastic bedrock, and bedding epimetamorphic boulder, and forest and public and private terrain, respectively. It indicates that human activities has a significant impact on slope collapse in the Yangtze River Delta.

Kamal Das [6] describes a landslide estimation using a hub and various sensors. A shaft with a rain sensor probe, humidity sensor, weight sensor (electronic weighing gauge device), and gyroscope is inserted into the ground. The precipitation monitoring sensor detects amount of rainfall, whereas the humidity sensor detects the relative moisture level. The pressure fluctuation at a specific spot is measured by the load cell transducer. Any displacement of the shaft over which the gauges are mounted is detected by the accelerometer. Furthermore, employing a Sensor Network, the technique for threshold-based landslide monitoring and

forecasting. The acquired sensor data is normalized using a min-max normalization algorithm in the hub's processing unit. Hub then creates an index and correlates it to a predetermined value. If the index exceeds a specified criterion, the hub transmits an alert message to the distant remote server, warning of a probable landslide.

Kusala Munasinghe [7] The prospect of applying the RFE approach for constructing a precise landslide forecasting models is explored in this work, which fills a research need. The landslide inventory of several landslide-prone locations in Sri Lanka are used to evaluate the forecasting models. However, because the forecasting model's performance is heavily dependent on location sensor information, the model must be fine-tuned for each new site using variable identification and model evaluation. Because of its site-specific character, the findings of two independent estimation techniques can indeed be contrasted even if they were developed using the identical landslide catalogue (same site). As a result, the authors are now testing the effectiveness of estimation techniques constructed employing Support Vector Machines, Random Forest, AdaBoost, and RFE in attempt to grant a full evaluation amongst the most prevalent features extraction approaches.

Shuai Chen [8] The OCSVM model's usefulness in landslide risk under the situation of inadequate landslide catalogue was investigated. For the building of the OCSVM model for the investigation of landslide hazard in Sichuan province District, SU is being used as the cartography component, and seven significant parameters (elevation, perspective, rock properties, and proximity from the riverbed, soil characteristics, earthquake strength, and agricultural use) were chosen. In contrast to the classic Svm classifier, OCSVM just requires one type of data sample. It is not overly reliant on the whole landslide registry and has little effect on the selection of landslide-free examples. OCSVM demonstrated the benefits of consistent accuracy and repeatable prediction findings. Finally, given the difficulty of getting a comprehensive landslide inventory post-earthquake, the suggested methodology may quickly generate a vulnerability map as well as provide critical recommendations for disaster response, evacuation, and post-seismic land use management.

Geerish Suddul [9] Researchers focus on the utilisation of Wireless Sensor Networks amongst some of the various ways used to monitoring flooding. Different factors may be measured in realtime, such as moisture content, humidity, and sunlight. This massive amount of data is evaluated in order to estimate the likelihood of a flood, and such projections can save lives. As a result, it is critical to streamline this procedure in order to reduce human participation and eliminate unwanted inefficiencies. Numerous Machine Learning algorithms have indeed been employed to achieve this goal, as they can assess pre-processed information and construct learning frameworks to generate predictions on



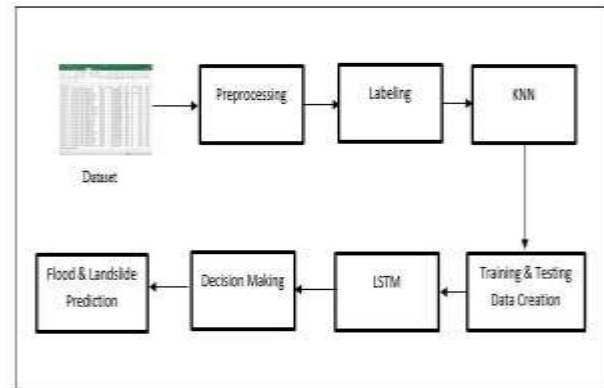
their own. Conventional Machine Learning approaches including such BF Tree, Random Tree, Random Forest, J48, MLP, and Bays Net have also been shown to produce varying levels of prediction performance. Some of them are highly suitable for flood risk assessment, but recent research have shown that hybrid strategies produce even better results than solo strategies in a variety of other fields. With the integration of Genetic Algorithm and nature influenced methodologies, the authors think that more improvement of Machine Learning approaches, particularly Artificial Neural Networks, is potential for precise estimation of flash flood risk assessment.

Nikhil Binoy C [10] The suggested approach is a two multilevel flood forecasting system, according to the description. ANN models are divided into three categories. The dam activation predictive algorithm is the first. Dam reservoir, water level, rainfall, and accumulated rainfall are the four variables that are evaluated to determine whether or not dam will open. The canal data is used for the next level of prediction. The following two ANN models at this level include two sensor inputs: it from an ultrasound device that detects depth and another from a strain gauge output standardized as fluid velocity. Even if a high forecast accuracy was attainable, the outcome demonstrates the unpredictability of the data obtained by the Dam's Central Water commission, implying the presence of additional factors such as erosion, bureaucratic red tape, and so on. The dam gate was utilised to determine the reopening allows the projection. Additionally, a comparable neural network was used to forecast the waterway. A straightforward if-else example might have been utilised, but in the actual world, issues such as runoff water, government interventions, relocation to farms, and so on would be present. These specifics are hard to come by, and modern construction characteristics for them are few. As a result, the authors have left room for varied contributions such as runoff.

Kettner A.J. [11] The DFO Flooding Laboratory has given this information. The first is a computer and browser programme that will not be explored further on this. The other platform is a smartphone version that can be downloaded for free from the Android and Apple application marketplaces. The authors chose to create a 'one-stop-shop' smartphone app because it enables them to engage directly with consumers while providing flood information. The immediate interaction allows students to reflect via an integrated feedback system, alerting us to critical customer demands and complaints, and receiving recommendations on how to make the flooding website and information more usable and easily understandable. Mobile apps also improve data availability, which is especially important when using data while on the go. During or shortly after a flood catastrophe, first-responding organizations such as the International Fund For agricultural development or the Aid Organizations are frequently dispatched on the scene. It is critical to have accurate, up-to-date flood documentation when deciding which places should

be forced to evict immediately, where to implement what threshold of supplies, how and where to deliver these capabilities to the individuals who need them, and where to set up one or more evacuation centers.

### III. PROPOSED WORK



**Figure 1: Proposed work Overview**

The proposed model for flood and landslide prediction is depicted in the above figure 1. The steps that elaborates each and every module is described in the below steps.

*Step 1: Data Preprocessing-* This is the first step of the proposed model, Here a dataset from the URL : <https://www.kaggle.com/datasets/ggopinathan/rainfall-in-kerala-from-19012017>. for flood and landslide prediction is downloaded in the form of .csv file. The Downloaded dataset contains some attributes like Subdivision, Year, JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC, ANNUAL RAINFALL and FLOOD\_LANDSLIDE. Once this file has been downloaded and then it is subject to preprocessing process. Here in this step the dataset is being reshaped properly into a double dimension list. Later in this process dropping the Column 'SUBDIVISION' with missing values to clean the data. After this process a Pearson Correlation is estimated to calculate the proper correlation between the attributes. Now a proper preprocessed data are prepared to feed to the next process of K-nearest neighbor classification.

*Step 2: K-Nearest Neighbor Classification* – This is the step that eventually produces the best data to feed to the neural network. For this process initially a Euclidean distance is estimated for each row with respect to the all other rows. The obtained Euclidean distance of each row is appended at the end of the row. Now all the rows in the dataset are sorted in ascending order based on the obtained Euclidean distance. Now a K value of 2 is set to half the data, where the first half with the lesser Euclidean distance are considered as the nearest data and then they are considered to feed to the neural network LSTM in the next step of the model. The Euclidean distance is measured using the Equation 1.





$$ED = \sqrt{(\sum (AT_i - AT_j))^2} \quad (1)$$

Where,

ED=Euclidian Distance

AT<sub>i</sub>=Attribute at index i

AT<sub>j</sub>= Attribute at index j

*Step 3: Training and Testing data creation* - Here in this step the dataset attributes are divided into two segments called X and Y. In X Segment some attributes are listed like 'YEAR', all months from January to December along with 'ANNUAL RAINFALL'. On the other hand, in Y segment contains Flood\_Land\_slide attribute is listed. These X and Y lists are split using the split function for training and testing to form X<sub>train</sub>, X<sub>test</sub>, Y<sub>train</sub> and Y<sub>test</sub> lists. A minMax scalar is used to normalize the data for the created lists. After this process the lists are reshaped to feed to the LSTM neural network as explained below.

*Step 4: LSTM Neural Network*- In this step of the proposed model a Long Short term Memory neural network is created for the sequential neural network model using the keras and tensorflow library of the python language. Then a layer of the LSTM neural network is added with 100 kernels along with samples and features. A dense layer with one unit along with “Relu” activation function is added in the neural network model. The LSTM neural network model is optimized using “adam” optimizer with batch size 50 for 100 epochs to produce best prediction rates. The “Relu” activation function is shown in the equation 2.

$$Relu = \max(0, x) \quad (2)$$

Where x is the input attributes values

The architecture of the LSTM Neural Network is shown below

LSTM	
Layer	Layer
LSTM 100	
Dense 1	Relu
Adam Optimizer	
Batch Size	50
Epoch	100

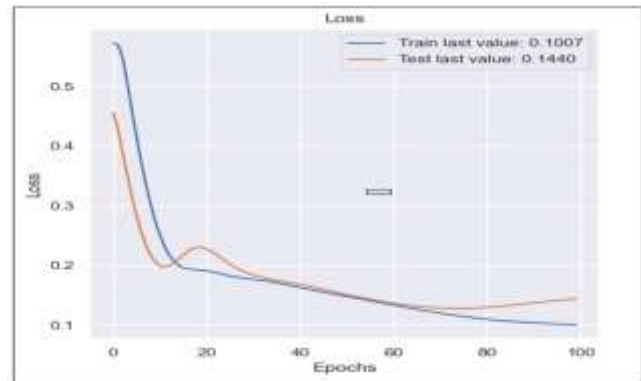
**Figure 2: LSTM Architecture**

*Step 5: Decision making for Flood and Landslide prediction*- The trained model is then used to predict the test data, which is around 33 % of the total data. After predicting the proper prediction then a list of predicted rainfall and landslide data is shown to the user.

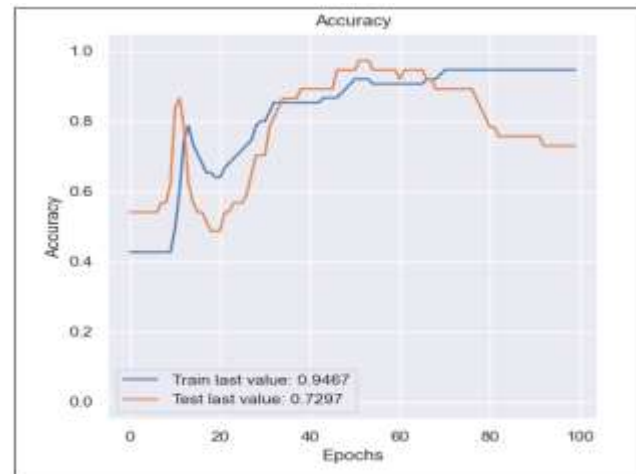
#### IV. RESULTS AND DISCUSSIONS

The proposed model for the flood and landslide detection is deployed using windows based machine with core i5 processor and 8 GB of primary memory. The model is deployed using the python programming language using Spyder as the IDE along with the keras and tensorflow libraries.

The proposed model uses LSTM neural network for the prediction and ran the model for 100 epochs. The achieved accuracy and losses are shown in the below figure 2 and 3 respectively.



**Figure 2: Epoch loss**



**Figure 3: Epoch Accuracy**

These graphs indicates system achieves best accuracy of 0.9467 and loss of 0.1007 as indicated in the figure 2 and 3.

The proposed system utilizes the RMSE( Root mean square error) for the evaluation of the results in between the actual values and predicted values for the flood and landslide prediction. The model is then utilized below mentioned equation 3 for the evaluation of the RMSE.

$$RMSE_{fo} = \left[ \sum_{i=1}^N (z_{fi} - z_{oi})^2 / N \right]^{1/2}$$

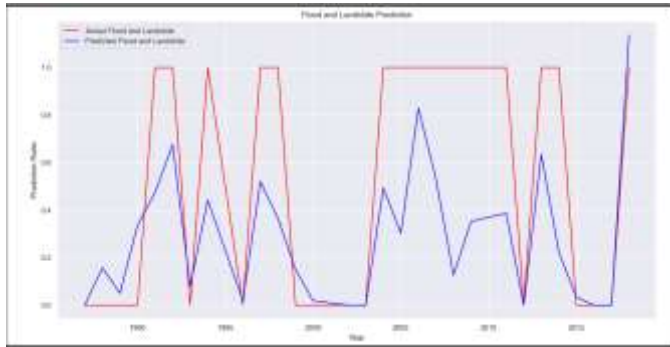
Where,

$\Sigma$  - Summation

$(Z_{fi} - Z_{oi})^2$  - Differences Squared for the expected and achieved Flood and landslide data.

N - Number of predicted data.

The RMSE values are computed for a number of Flood and landslide data and Testing performed through this proposed approach. These values of RMSE are rigorously calculated with the outcomes stipulated in the table 1 and the figure 4 given below.



**Figure 4: RMSE comparison of Actual and Predicted flood and landslide prediction.**

The above figure indicates that the proposed model achieves best RMSE value of 0.3893 for the prediction of flood and landslide values by using the LSTM neural network model. This is considered as the best achievement in the first trail of the experiment.

## V. CONCLUSION AND FUTURE SCOPE

Flooding is more likely to develop in freshwater wetlands and coastline locations, although it may also happen in regions with abnormally lengthy periods of high precipitation. When it rains, the quantity of precipitation that enters the streams is determined by the catchment's features, such as gradient, plant cover, retention capability, and so on. Drainage, tidal, temperature, moisture, and other elements all have an impact on the flooding. These similar situations could also lead to a landslide scenario in hilly or areas with slope. This leads to a lot of destruction and could potentially be hazardous to the inhabitants as both of these disasters happen in a matter of minutes which eliminates any timely rescue operation. As a result, a method for predicting landslides and floods is urgently required. This implementation of the proposed model evaluates the results using the RMSE. The conducted experiments yields RMSE result of 0.3893 for the prediction of flood and landslide values by using the LSTM neural network model.

In the future, this project can be enhanced to work on the live stream data obtained from the weather forecasting department to avoid the loss of lives and properties.

## REFERENCES

1. M. Li et al., "Warning of Rainfall-Induced Landslide in Bazhou District," *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 6879-6882, doi: 10.1109/IGARSS39084.2020.9324416.
2. D. Utomo, L. -C. Hu and P. -A. Hsiung, "Deep Neural Network-Based Data Reconstruction for Landslide Detection," *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 3119-3122, doi: 10.1109/IGARSS39084.2020.9323124.
3. C. C. Khaing and T. L. L. Thein, "Prediction of Rainfall Based on Deep Learning and Internet of Things to Prevent Landslide," *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 2020, pp. 190-191, doi: 10.1109/GCCE50665.2020.9292057..
4. S. Sun, C. Lian and X. Wang, "Down-Level Control Based on Level Prediction of Landslide Evolutionary State," *2021 11th International Conference on Information Science and Technology (ICIST)*, 2021, pp. 48-57, doi: 10.1109/ICIST52614.2021.9440637.
5. H. Zhang, G. Zhang and Q. Jia, "Integration of Analytical Hierarchy Process and Landslide Susceptibility Index Based Landslide Susceptibility Assessment of the Pearl River Delta Area, China," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 11, pp. 4239-4251, Nov. 2019, doi: 10.1109/JSTARS.2019.2938554.
6. K. Das, S. Majumdar, S. Moulik and M. Fujita, "Real-Time Threshold-based Landslide Prediction System for Hilly Region using Wireless Sensor Networks," *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, 2020, pp. 1-2, doi: 10.1109/ICCE-Taiwan49838.2020.9258181.
7. K. Munasinghe and P. Karunanayake, "Recursive Feature Elimination for Machine Learning-based Landslide Prediction Models," *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 2021, pp. 126-129, doi: 10.1109/ICAIIIC51459.2021.9415232.
8. S. Chen, Z. Miao, L. Wu and Y. He, "Application of an Incomplete Landslide Inventory and One Class Classifier to Earthquake-Induced Landslide Susceptibility Mapping," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1649-1660, 2020, doi: 10.1109/JSTARS.2020.2985088.
9. G. Suddul, K. Dookhitram, G. Bekaroo and N. Shankhur, "An Evolutionary MultiLayer Perceptron Algorithm for Real Time River Flood Prediction," *2020 Zooming Innovation in Consumer Technologies Conference (ZINC)*, 2020, pp. 109-112, doi: 10.1109/ZINC50678.2020.9161824.
10. C. Nikhil Binoy, N. Arjun, C. Keerthi, S. Sreerag and A. H. Nair, "Flood Prediction Using Flow and Depth Measurement with Artificial Neural Network in Canals," *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019, pp. 798-801, doi: 10.1109/ICCMC.2019.8819807.
11. A. J. Kettner, G. J. . -P. Schumann and G. R. Brakenridge, "Applying Remote Sensing to Support Flood Risk Assessment and Relief Agencies: A Global to Local Approach," *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 3239-3242, doi: 10.1109/IGARSS39084.2020.9323351.