# LANDSLIDE PREDICT: MACHINE LEARNING AND SENSOR-BASED EARLY WARNING

**Dr. Smitha B[1], Aparna Prasad K[1], Arya Ashok[2], Adithya R[4]**

Department of Electrical and Electronics Engineering, NSS College of Engineering Palakkad, Kerala, India

## ABSTRACT

Landslides are natural disasters that pose significant threats to human life and infrastructure. This project presents a machine learning (ML) and sensor-based early warning system to predict landslides effectively. The system integrates real-time sensor data with ML algorithms to enhance prediction accuracy and provide timely alerts. Key hardware components include soil moisture sensors, rain gauges, vibration sensors, and tilt sensors, all interfaced with an ESP32 micro- controller for data acquisition. The collected data is analyzed and processed using a trained machine learning model, which predicts landslide-prone areas based on historical patterns and real-time conditions. The trained model has achieved 98By combining sensor-based monitoring and machine learning-based predictions, this system enhances early warning capabilities, thereby reducing the risk of landslide- related damages.

## 1. INTRODUCTION

Landslides are a major natural disaster that cause significant loss of life and property, especially in regions with unstable terrain. Predicting and mitigating these disasters is crucial to reducing their impact. This project focuses on developing a Machine Learning (ML) and Sensor-Based Early Warning System for Landslide Prediction. The proposed system integrates machine learning algorithms and real-time sensor data to identify areas prone to landslides. By collecting and analyzing satellite images, the ML model predicts landslide risks with high accuracy. Additionally, hardware components such as soil moisture sensors, vibration sensors, rain gauges, and tilt sensors which include both accelerometer and gyroscope (MPU6050) are used to continuously monitor environmental conditions. The collected data is processed using an ESP32 microcontroller, which plays a key role in data acquisition and communication. A web-based platform is also developed to display real-time sensor readings and risk levels, ensuring that critical information is accessible to both the public and relevant authorities. Furthermore, an SMS alert system is implemented to notify authorities in case of potential landslide risks, enabling timely action. This report provides a comprehensive overview of the project, including its design, hardware components, software implementation, working mechanism, and outcomes. The goal is to create a reliable and efficient landslide prediction system that can aid in disaster preparedness and mitigation efforts.

### 1.1. RELATED WORKS

Landslide prediction and monitoring have been crucial for research due to the severe threats they pose to human life, infrastructure, and environment. Various techniques have been developed to detect, predict, and mitigate landslides using geotechnical, remote sensing, and computational methods. Early research primarily relied on empirical and statistical models, while recent advancements have integrated machine learning, deep learning, and IoT-based monitoring systems to enhance prediction accuracy and real-time monitoring. Initial landslide studies were largely based on geological and geotechnical surveys that assessed soil properties, slope stability, and environmental factors such as rainfall intensity, seismic activity, and vegetation cover. Statistical approaches such as the Frequency Ratio (FR) model, Logistic Regression (LR), and Analytical Hierarchy Process (AHP) were widely used for landslide susceptibility mapping. These models provided a probabilistic estimation of landslide-prone areas but were limited by their dependence on historical data and inability to adapt to real-time conditions.

Our project addresses these shortcomings by integrating IoT-based real-time monitoring with machine learning-driven predictive analytics. Unlike traditional methods that depend solely on historical data or AI models that require preprocessed inputs, our approach continuously collects real- time data using a network of IoT sensors. These sensors measure critical parameters such as soil moisture, ground displacement, rainfall intensity, and seismic activity, which are then processed using ML algorithms to detect early signs of slope instability. This enables an immediate response system that provides early warnings, reducing the risk of catastrophic landslides. Additionally, our project is more cost-effective than conventional AI-based methods, as it optimizes sensor placement and employs edge computing techniques to minimize computational overhead.

Another key advantage of our approach is scalability and adaptability. While existing statistical and AI models often struggle with variations in terrain and climatic conditions, our system is designed to dynamically adjust to different environmental settings. The integration of IoT and ML al- lows for continuous learning and adaptation, ensuring higher accuracy and robustness in diverse geographic regions. Moreover, by combining multiple data sources, including remote sensing, geotechnical analysis, and AI-based classification, our project provides a comprehensive and reliable landslide monitoring system that outperforms existing methods in terms of accuracy, efficiency, and real-time response.

In conclusion, compared to existing projects that either rely on static statistical models or purely on machine learning without real-time capabilities, our project offers a more holistic and effective solution. By integrating IoT for live data collection and ML for predictive analysis, we ensure faster response times, improved accuracy, cost-effectiveness, and greater adaptability in landslide-prone areas. This approach significantly enhances disaster preparedness and mitigation strategies, making it a valuable advancement over conventional landslide prediction and monitoring methods.

### 1.2. RESEARCH AREA

The study of machine learning for landslide prediction is important, as advanced algorithms such as Convolutional Neural Networks (CNN) and U-Net can fully process satellite imagery and real-time sensor data in order for identifying particularly high-risk areas with higher accuracy. Real-time monitoring integration using sensors improves forecasts by using IoT-based sensors like soil moisture, vibration, tilt, and rainfall sensors, which constantly gather ecological data. For more analytical understandings, the ESP32 chip processes and then sends that data to such digital cloud services. In addition, specific edge computing and real-time alert systems may be studied to lower latency in disaster response, allow- ing early warnings via GSM and web-based notifications. AI-driven geospatial analysis is another key research area; hybrid approaches combining satellite data with ecological parameters such as rainfall intensity, soil type, as well as seismic activity can improve on landslide prediction models. Additionally, multiple energy-efficient sensor networks can be studied in order to optimize power consumption in re- mote landslide-prone areas, ensuring long-term operation of monitoring systems. These specific research directions aim to truly improve disaster preparedness via integrating artificial intelligence, IoT, and also real-time monitoring, thereby making landslide prediction systems far more efficient and extremely scalable.

## 2. METHODOLOGY

### A. SYSTEM ARCHITECTURE AND DATA COLLECTION

1. *SOFTWARE SETUP*: ML is integrated to predict the landslide prone areas by collecting satellite images of various landslide affected places and then they are used for training the machine. The trained machines after acquiring a particular accuracy is ready to detect landslide. The data set required for this collected from the open source Kaggle. The satellite images consist of 6 channels which include RGB, NDVI, Slope, Elevation and Mask. At testing phase, the trained machine will provide output in an image format and the landslide prone areas are show in particular color. After plotting the landslide prone areas sensors are kept on these areas and real time data is collected. The prediction will be seen in website.

The foundation of this model training lies in an open- source dataset sourced from Kaggle, specifically curated for landslide prediction and monitoring applications. This dataset comprises satellite images capturing landslide- affected regions. Each image is structured with six distinct channels, providing comprehensive environmental data that enhances the model’s ability to detect landslide-prone areas.

The various channels used are RGB (Red, Green, Blue) channels which capture the visible spectrum of the landscape, providing essential visual information on vegetation, water bodies, and exposed soil. These channels help the model identify the surface conditions and distinguish between different land cover types. NDVI (Normalized Difference Vegetation Index) channel which is a critical indicator of vegetation health and density, derived from infrared and red bands. A healthy vegetation cover often suggests a stable land surface, while low or disturbed NDVI values can indicate landslide activity or areas vulnerable to erosion.

Slope in this channel it measures the gradient or steepness of the terrain, which is a major factor influencing landslide susceptibility. Steeper slopes generally increase the risk of landslides, as they are less stable and more prone to failure under certain conditions. Elevation channel it provides insights into the topographical characteristics of the region, such as high- altitude areas, valleys, and mountainous terrain. This information helps the model understand altitude-related variations in landslide risk. Mask channel serves as a binary indicator to distinguish landslide-affected areas from unaffected regions. It helps in training the model by clearly marking regions with landslide incidents, which improves its ability to learn and identify patterns associated with landslides.

The dataset includes images from various geographical locations, capturing a wide range of topographical features, climates, and land use patterns. This diversity enriches the model's robustness by exposing it to different terrain types and climatic conditions, which enhances its predictive accuracy across diverse landscapes. The dataset's high-resolution satellite images allow for detailed analysis of each channel, ensuring the model receives fine-grained spatial information. This enables the model to capture subtle variations in terrain and vegetation that might be indicative of landslide susceptibility.

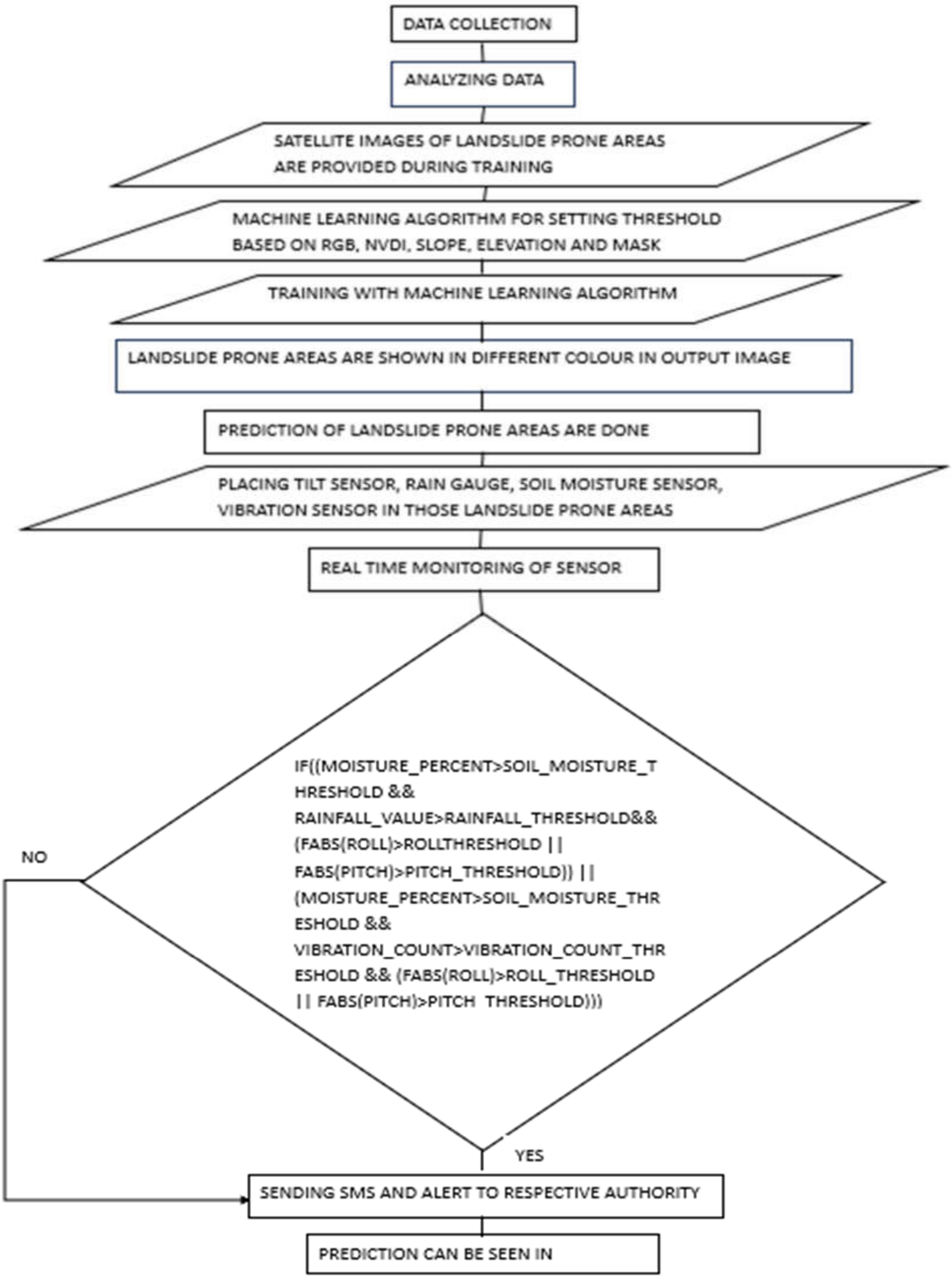


Figure.1. Flow Chart of prediction of landslide

2. *HARDWARE SETUP*: To complement the machine learning model, a sensor-based hardware system was developed using ESP32 as the microcontroller. The system con- sists of multiple sensors that continuously monitor environ- mental parameters in real-time. The sensor-based approach significantly enhances real-time monitoring, reducing dependency on satellite imagery alone. The thresholds set for each sensor were determined based on historical landslide data and environmental research. The system successfully captured real-time environmental variations, allowing for immediate alerts and decision-making. The hardware setup includes multiple sensors to monitor environmental parameters such as soil moisture, rainfall, vibrations, and tilt. These sensors are interfaced with an ESP32 microcontroller, which acts as the central processing unit. The collected data is used to determine landslide-prone conditions.

Soil Moisture Sensor which measures soil moisture levels to detect increased water content, which weakens soil stability. Connected to the ESP32's analog input pins. Its threshold value is given as 60%. Rain Gauge which measures the amount of rainfall in millimeters. Uses a tipping bucket mechanism with a Hall effect sensor to count pulses and determine rainfall. Its threshold value is given as 5mm. Vibration Sensor detects ground vibrations, which are key indicators of landslides. Connected to ESP32's digital pins for real-time monitoring. Threshold Value: 5 vibrations within a given period. Tilt Sensor monitors ground movement and tilting, indicating shifts in landmass. Interfaced with ESP32 via I2C communication for continuous tilt data acquisition. lit sensor system for landslide detection using the MPU6050, a sensor that combines both a gyroscope and accelerometer. The system will continuously monitor the pitch and roll of the ground to detect any significant shifts in slope. Using the accelerometer and gyroscope data, the system will track changes in the ground's orientation, specifically measuring pitch (forward or backward tilt) and roll (sideways tilt). If either the pitch or roll exceeds a 10° threshold, the system will trigger an alert to indicate potential landslide conditions. Its threshold Value is given as 10° pitch and 10° roll. ESP32 Microcontroller is central processing unit responsible for collecting and processing sensor data. Features Wi-Fi connectivity for data transmission. Communicates with the GSM module to send alerts. GSM Module Sends SMS alerts to authorities when landslide risk is detected uses AT commands to interface with ESP32. Buzzer LED Indicators Provide local alerts when a landslide threat is detected. A buzzer can be installed at landslide-prone areas to provide immediate audible alerts when sensors detect critical conditions (e.g., soil movement, vibration, or heavy rainfall). It helps local residents and workers take necessary precautions or evacuate in case of high landslide risk. The buzzer can be triggered by microcontrollers based on sensor readings.

3. *WEBSITE DEVELOPMENT*: A web-based interface is developed for real-time landslide monitoring. The website provides live sensor data visualization (moisture, rainfall, vibration, tilt). Risk level is shown in website where all people can access it. SMS and alerts to authorities when risk is high. Real time data display sensor data is updated every few seconds.

B. MACHINE LEARNING MODEL:

To ensure that the landslide prediction model is reliable and performs effectively in real- world applications, a rigorous evaluation process is conducted. This phase validates the model's ability to generalize across diverse terrains and environmental conditions, crucial for practical deployment in various geographical regions. Evaluation metrics such as accuracy, F1 score, recall, and precision are thoroughly analyzed, but beyond these metrics, specific steps are taken to assess the model's adaptability to different landscapes, land cover types, and climates. Machine Learning Libraries (TensorFlow, Keras) are Used for building and deploying the ML model. Using platforms like TensorFlow and Keras, the system incorporates Convolutional Neural Networks (CNNs) to analyze terrain images for visual signs of instability. The model has been trained on satellite images, leveraging RGB channels, NDVI (Normalized Difference Vegetation Index), slope, and elevation data.

The model is tested on high-altitude mountainous terrains, where steep slopes, elevation, and rugged topography create an elevated landslide risk. In these areas, the model must accurately interpret slope gradients and elevation changes, distinguishing between stable and potentially unstable slopes. The inclusion of slope and elevation channels in the dataset helps the model adapt effectively to these challenging terrains. In densely vegetated regions, such as tropical rainforests, the model's performance is validated against complex vegetation cover. High NDVI values (indicating dense vegetation) could mask early signs of soil instability, so the model must balance vegetation analysis with other factors like slope and elevation. Here, the mask and NDVI channels help the model to recognize subtle signs of land movement beneath dense vegetation.

In urban and suburban areas, where human activity has altered natural landscapes, the model must adapt to detect landslide-prone zones in the presence of roads, buildings, and other structures. This setting often involves unique drainage patterns and potential erosion along road cuts, and the model's ability to generalize in these areas is validated by checking how well it detects risk despite man- made modifications. In dry, less vegetated areas with sandy or rocky soil, landslides can still occur but might present differently than in wetter regions. Here, the model's interpretation of elevation, slope, and minimal NDVI is critical to accurately assessing landslide risk. Evaluation in arid environments helps ensure the model doesn't overly rely on vegetation indicators and can assess landslide risk in the absence of extensive plant cover.

The evaluation process also includes testing the model against data representing different seasons, such as monsoon seasons in tropical areas or snowmelt periods in mountainous regions. During these periods, landslide risks may increase due to heavy rainfall or melting snow, both of which can destabilize slopes. Evaluating the model under these seasonal conditions ensures it can adapt to varying environmental dynamics that affect landslide susceptibility. A sensitivity analysis is performed to assess the model's responsiveness to variations in key channels, such as slope, NDVI, and elevation. This involves slightly modifying values within these channels to observe if and how the model's predictions shift. Such analysis helps validate that the model's predictions are based on relevant geophysical features and that it can distinguish between changes that indicate landslide risk versus normal terrain variation. To further confirm its robustness, cross-validation techniques are applied, splitting the dataset into multiple subsets to ensure that the model's high accuracy is not overly dependent on a specific part of the data. This reduces any potential biases and further validates the model's ability to generalize across different conditions. Through this comprehensive evaluation process, the model's reliability across diverse terrains, climates, and seasonal variations is confirmed, ensuring its practical applicability in landslide monitoring systems worldwide. This multi-dimensional approach to validation helps ensure that the model is not only accurate but adaptable and ready for deployment in a wide range of environments.

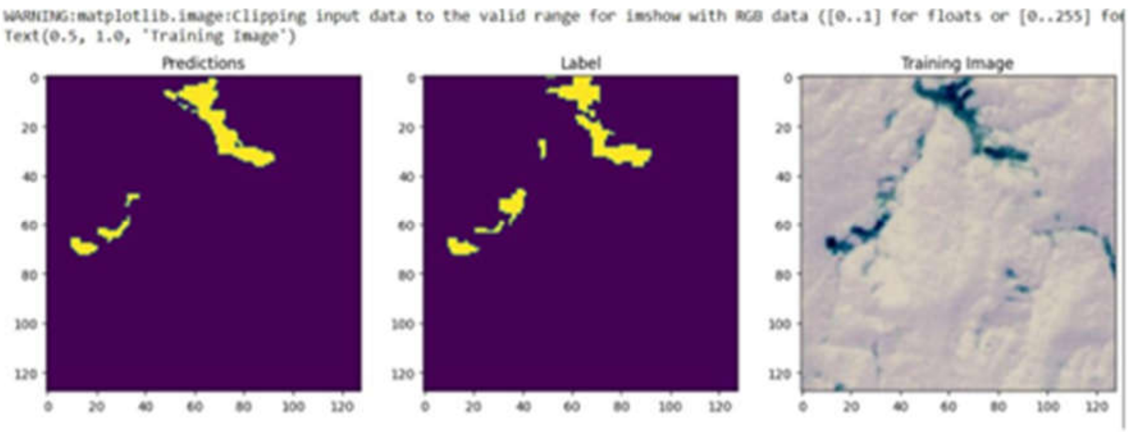


Figure. 2. Training Model

## 3. RESULTS AND DISCUSSION

### A. PERFORMANCE ANALYSIS

The CNN architecture is designed to process and learn from the rich, multi-channel satellite images in the dataset, which include RGB, NDVI, slope, elevation, and mask channels. The model's training process is carefully optimized to achieve both high accuracy and effective generalization across diverse landscapes.

- Training Process and Epochs: The CNN model is trained over 75 epochs, an iterative process where the model progressively refines its weights based on feedback from the dataset. Through each epoch, the model adjusts its filters and neurons to capture increasingly complex features and pat- terns within the satellite imagery. During training, the model encounters a variety of landslide-affected and unaffected areas, learning to differentiate between subtle differences in color, texture, vegetation health, and topographical indicators associated with landslide risk. The model's performance improves with each epoch as it minimizes the error in its predictions, gradually reaching a high level of accuracy and stability in its output.

- Optimization Techniques: To enhance learning and avoid overfitting, the model utilizes techniques such as data augmentation (e.g., rotations, flips) to expose it to different variations of the images, which simulates real-world variability. Additionally, dropout layers and batch normalization are applied to prevent the model from relying too heavily on specific features, further improving its generalization to unseen data.

- Performance Metrics: Accuracy (98%): The model achieves an impressive 98% accuracy, indicating that it correctly classifies landslide and non-landslide areas in the vast majority of cases. Accuracy measures the overall correctness of the model's predictions but can sometimes be influenced by class imbalances.

- F1 Score (68.08%): The F1 score combines precision and recall to give a balanced metric, especially useful when dealing with imbalanced classes. In this context, the F1 score highlights how well the model detects true landslide regions versus false positives, even if one class (non-landslide) is more common. Although lower than accuracy, the F1 score is valuable for assessing performance in real-world applications.

- Loss (3.7%): The low loss percentage (3.7%) demonstrates the model's success in minimizing the difference between its predictions and the actual labels. Loss serves as an internal measure for training feedback, indicating the extent of prediction error that the model aims to reduce over each epoch.

- Recall (59.8%): Recall indicates the proportion of actual landslide-affected regions that the model successfully identifies. With a recall of 59.8%, the model captures over half of the true landslide incidents, although this metric reveals potential areas for improvement in detecting all landslide-prone regions, possibly by fine-tuning the model or rebalancing the dataset.

- In addition to the overall accuracy (98%), the model's F1 score, recall, and precision are re-evaluated on specific terrain subsets to understand its performance under various environmental conditions. This granular assessment provides insight into the model's strengths and areas for improvement, particularly its recall score, to ensure it doesn't miss landslide-prone areas in any specific terrain type. After model evaluation got final training accuracy as 99.73% and validation accuracy as 98.74%.

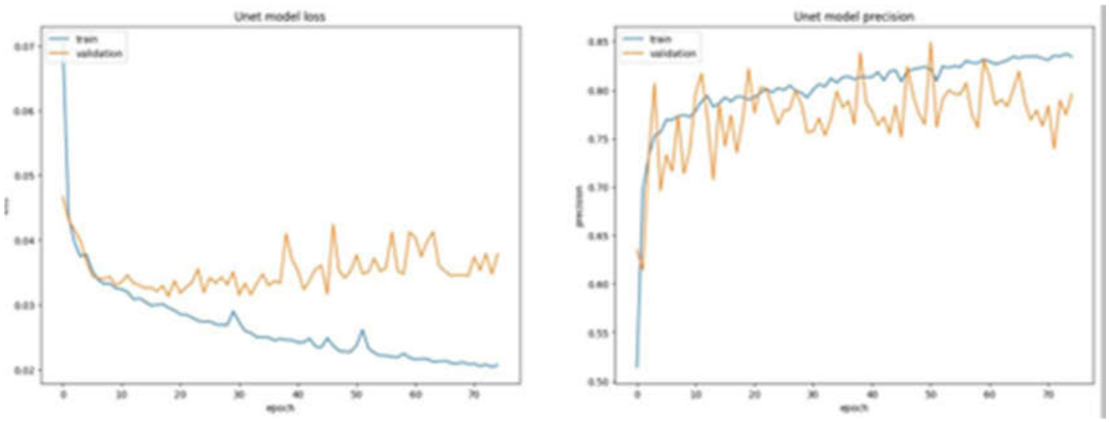

Fig. 3. Loss vs epoch and precision vs epoch of model

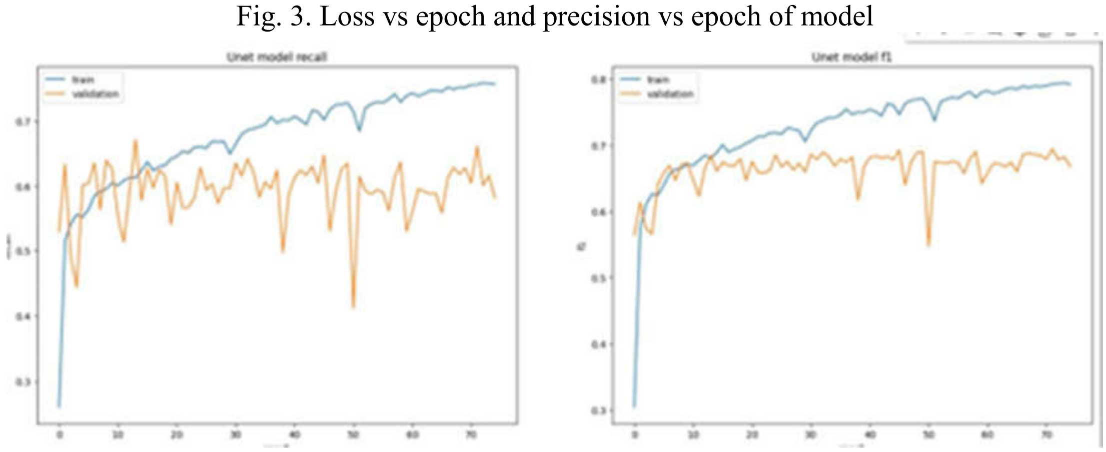

Figure. 4. Recall vs epoch and f1 vs epoch of model

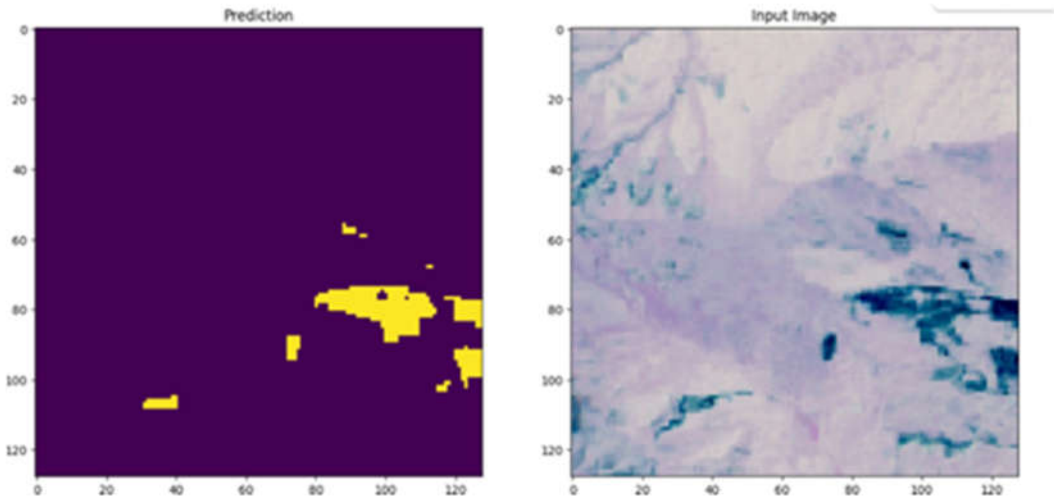


Figure. 5. Output of testing showing landslide prone areas

## B. HARDWARE PERFORMANCE

After plotting the landslide prone areas by machine learn- ing model various sensors like moisture sensor, vibration sensor, rain gauge, tilt sensor are embedded and integrated with ESP32. Real time values are collected by these sensors. Threshold for each sensors value are analyzed and coded into ESP32. Condition is given for predicting landslide risk. When soil moisture percent, rainfall value and roll or pitch of tilt sensor becomes greater than the threshold value or when soil moisture percent, vibration count and roll or pitch of tilt sensor becomes greater than threshold value a message and alert will be sent to respective authorities with the help of GSM module. The ESP32 is integrated with firebase and real time sensor value can be seen on website. From website all people can monitor the real time sensor value and can get the risk prediction.

| SENSOR | PURPOSE | INTEGRATION WITH ESP32 | THRESHOLD SET |
|---|---|---|---|
| Soil Moisture Sensor | Measures the water content in the soil to assess saturation levels | Connected to analog input pins to read real-time soil moisture data | 60% moisture content (beyond which landslide risk increases) |
| Rain Gauge (Tip- ping Bucket) | Measures amount of rainfall in millimeters (mm)<br>Rainfall =rain Pulse Count* mm Per Pulse | Generates pulses via a hall effect sensor, which are counted to measure rainfall levels. | 5 mm of rainfall (above this, soil becomes more susceptible to landslides) |
| Vibration Sensor | Detects ground vibrations indicating possible movement or tremors | Connected to digital pins to capture ground movement patterns | 5.0 vibrations within a minute (beyond this, an alert is triggered) |
| Tilt Sensor (MPU6050) | Monitors changes in ground inclination to detect possible shifting using pitch and roll by the principle of gyroscope and accelerometer inside MPU6050. | Interfaced with ESP32 via I2C communication for real-time tilt data acquisition | 10 pitch and 10 roll (beyond this angle the area will become more prone to landslide). |
| ESP32 Microcontroller | Acts as central processing unit, collecting sensor data and displaying results | Equipped with Wi-Fi capability for remote data transmission | N/A (handles All data Processing and communication) |

Table.1. Key Hardware Components and Thresholds

## 4. CONCLUSION

The developed landslide prediction and monitoring system effectively combines machine learning techniques with real-time sensor-based data acquisition to identify vulnerable zones and issue timely alerts. The project employed the U-Net algorithm for mapping areas at risk of landslides, while the hardware implementation utilized the ESP32 microcontroller interfaced with multiple environmental sensors, including soil moisture, tilt, vibration, and rainfall measurement modules. The system demonstrated a high accuracy rate of 98% in identifying landslide-prone regions, validating the performance of the machine learning model. Continuous monitoring is enabled through real-time data acquisition, supported by a web platform that provides open access to live sensor readings. This interface allows both residents and authorities to stay informed about current risk levels. In addition, an automated SMS alert mechanism has been integrated to notify concerned individuals when sensor readings exceed predefined safety thresholds, facilitating prompt response and increasing public safety. The blend of predictive modeling and real-time environmental tracking offers a robust early warning system. Looking ahead, the system can be enhanced by scaling up sensor deployment, integrating satellite-based data sources, and adopting more advanced AI algorithms for risk analysis. Overall, this project provides a solid framework for a real-time, adaptable disaster management solution suited for various terrains.

## REFERENCES

[1] Akanksha Sharma, Shakti Raj Chopra, Suhas Gajanan Sap- ate.Krishanarora.Mohammad Khalid Imam Rahmani (Senior Member, IEEE), Sudan Jha (Senior Member, IEEE), Sultan Ahmad (Member, IEEE), MD Ezaz Ahmed,Hikmat A. M. Abdeljaber and Jabeen Nazeer., Artificial Intelligence Techniques for Landslides. Prediction Using Satellite Imagery, in IEEE Transactions, September 2024.

[2] Maneesha Vinodini Ramesh.,Design, development, and deployment of a wireless sensor network for detection of landslides,in SciVerse ScienceDirect, Volume 13, Part A, February 2014, Pages 2-18

[3] Hemalatha, Maneesha Vinodini Ramesh, Senior Member, IEEE, and Venkat P. Rangan,Senior Member, IEEE.,Effective and Accelerated Forewarning of Landslides Using Wireless Sensor Networks and Ma- chine Learning,in IEEE Sensor Journal,VOL. 19, NO. 21, November 1, 2019

[4] Hiroki Mizuochi, Kazuhiro Miyazaki, Tomoya Abe, Hideo Hoshizumi, Daisaku Kawabata, Koki Iwao, Moe Matsuoka, Yoshinori Miy- achi,Detection of long-term slope displacement using time-series DInSAR and geological factor analysis for susceptibility as- sessment of landslides in northwestern Kyushu Island towards 5G, ScienceDirect,15May2015, VOL45.

[5] N. A. A. Aziz and K. A. Aziz, “Managing disaster with wireless sensor networks,” in Proc. 13th Int. Conf. Adv. Commun. Technol. (ICACT), Feb. 2011, pp. 202–207.

[6] T. Hemalatha, M. V. Ramesh, and V. P. Rangan, “Adaptive learning techniques for landslide forecasting and the validation in a real-world deployment,” in Proc. Workshop World Landslide Forum. Springer, 2017, pp. 439–447.

[7] F. Bardi et al., “Integration between ground based and satellite SAR data in landslide mapping: The San Fratello case study,” Geomorphol- ogy, vol. 223, pp. 45–60, Oct. 2014.

[8] R. Prabha, M. V. Ramesh, V. P. Rangan, P. V. Ushakumari, and T. Hemalatha, “Energy efficient data acquisition techniques using context aware sensing for landslide monitoring systems,” IEEE Sensors J., vol. 17, no. 18, pp. 6006–6018, Sep. 2017.

[9] G. B. Crosta and P. Frattini, “Rainfall thresholds for triggering soil slips and debris flow,” in Proc. 2nd EGS Plinius Conf. Medit. Storms, Siena, Italy, A. Mugnai, F. Guzzetti, and G. Roth, Eds. 2001, pp. 463–487

[10] R. Bhasin et al., “Landslide hazards and mitigation measures at Gangtok,Sikkim Himalaya,” Eng. Geol., vol. 64, no. 4, pp. 351–368, 2002.

[11] J.H. R. Pourghasemi, A. G. Jirandeh, B. Pradhan, C. Xu, and C. Gokceoglu, "Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran," J. Earth Syst. Sci., vol. 122, no. 2, pp. 349–369, 2013.

[12] V. Singhroy, "Satellite remote sensing applications for landslide detec- tion and monitoring," in Landslides–Disaster Risk Reduction. Springer, 2009, pp. 143–158.

[13] M. V. Ramesh, "Design, development, and deployment of a wireless sensor network for detection of landslides," Ad Hoc Netw., vol. 13, pp. 2–18, Feb. 2014.

[14] V. Tofani, S. Segoni, A. Agostini, F. Catani, and N. Casagli, "Use of remote sensing for landslide studies in Europe," Natural Hazards Earth Syst. Sci., vol. 13, no. 2, pp. 299–309, 2013.

[15] D. Petley, "Global patterns of loss of life from landslides," Geol- ogy,vol. 40, no. 10, pp. 927–930, 2012