Challenges in Sustainability
https://www.acadlore.com/journals/CIS

Predictive Landslide Risk Assessment Using Synthetic Data and Machine Learning

Fathey Mohammed[1*], Narishah Mohamed Salleh[1], Aw Kay Rong[1], Chan Jun Cong[1], Kenneth Haw Mun Ban[1], Kua Er Shiun[1], Teo Chuan Wei[1], Syafiq Ashraf Ahmad Khalid[2]

[1] Department of Business Analytics, Sunway Business School, Sunway University, 47500 Subang Jaya, Malaysia
[2] Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, 71800 Nilai, Malaysia

[*] Correspondence: Fathey Mohammed (fathey.m.ye@gmail.com)

Received: 03-12-2026 **Revised:** 04-05-2026 **Accepted:** 04-10-2026

Citation: Mohammed, F., Salleh, N. M., Rong, A. K., Cong, C. J., Ban, K. H. M., Shiun, K. E., Wei, T. C., & Khalid, S. A. A. (2026). Predictive landslide risk assessment using synthetic data and machine learning. *Chall. Sustain.*, *14*(2), 339–359. https://doi.org/10.56578/cis140208.

Abstract: Landslides remain one of the most critical natural hazards, posing significant threats to infrastructure, the environment, and human life. Traditional approaches to landslide risk prediction, such as rainfall threshold models and image-based classification, often face limitations including data imbalance, low generalizability, and poor performance in capturing medium- to high-risk scenarios. This study introduces a predictive framework that integrates synthetic data generation with a multiple logistic regression model to improve landslide risk assessment in the Malaysian context. The model was trained on balanced datasets and evaluated through confusion matrices, performance metrics, and validation using unseen data across three distinct scenarios. Results demonstrate that a multiple-logistic-regression model trained on this balanced data achieved an overall accuracy of 0.73, precision of 0.73, recall of 0.73, and a Receiver Operating Characteristic-Area Under Curve (ROC-AUC) of 0.80. In three validation scenarios using unseen data from 2015–2024 (three months before, during, and three months after known landslide events), the model correctly identified medium and high-risk periods when other machine-learning models defaulted to low-risk predictions. The study highlights the trade-off between accuracy and generalization in machine-learning-based early warning systems and underscores the importance of class-balancing and rigorous validation for real-world applicability. Our findings, therefore, demonstrate that the logistic-regression model, when paired with synthetic data augmentation, can serve as a cost-effective, interpretable pre-screening tool for regional landslide risk assessment in Malaysia.

Keywords: Landslide prediction; Machine learning; Logistic regression; Synthetic data generation; Rainfall threshold; Soil wetness; Landslide risk assessment

1. Introduction

Landslides are one of the most devastating natural disasters globally, causing significant loss of life, property damage, and environmental degradation (Alcántara-Ayala, 2025; Turner, 2018). In Malaysia, the occurrence of landslides is frequent and unpredictable due to the country's tropical climate, especially heavy rainfall during the monsoon seasons. Malaysia experiences heavy rainfall throughout the year, especially during the Northeast Monsoon and Southwest Monsoon seasons (Syafrina et al., 2017). High volumes of rain increase soil saturation, which significantly affects slope stability, often triggering landslides. When rainwater infiltrates the soil, it reduces its cohesion and leads to increased pore pressure, making slopes prone to collapse. Historically, events such as the Highland Towers collapse in 1993, as well as recent landslides in residential and recreational areas, highlight the urgent need for effective landslide management and prevention strategies (Akter et al., 2019; Rosly et al., 2022). As climate change intensifies, the likelihood of landslides increases, especially in unusually heavy rainfall (Zhang et al., 2023). Anthropogenic pressures further elevate landslide susceptibility. Rapid urban expansion into hilly terrains, deforestation for agriculture, and inadequate land-use planning weaken natural slope stability and expose communities to elevated risk (Alcántara-Ayala, 2025; Froese & Schilling, 2019).

Despite these recurring incidents, landslide monitoring systems remain fragmented and reactive, with most efforts emphasising post-disaster response rather than early detection. Many high-risk rural and semi-urban areas also lack appropriate geospatial monitoring technologies (Kumari et al., 2025; Zahri et al., 2025). Several persistent challenges limit Malaysia's ability to predict and mitigate landslides effectively:

- Existing landslide detection and monitoring mechanisms are often insufficient to provide timely warnings, especially in the face of high rainfall (Ligong et al., 2022; Zahri et al., 2025).
- While rapid urban expansion into hilly areas and the clearing of forests for agriculture destabilize slopes, increasing the potential for landslides, there is a lack of comprehensive land-use planning and the weakening of natural defences, such as forests, further heighten the risk (Alcántara-Ayala, 2025; Froese & Schilling, 2019).
- The complex interaction of geographical and environmental factors, such as soil type, terrain properties, rainfall patterns, and vegetation cover, makes it difficult to predict landslides accurately. Traditional statistical models often fail to comprehend this complexity (Casagli et al., 2023; Li & Duan, 2024).
- Many high-risk areas in Malaysia are not equipped with the latest geospatial monitoring and early warning systems. This limits the ability to predict landslides effectively, especially those triggered by intense rainfall (Kemarau et al., 2025; Kumari et al., 2025).

Traditional approaches, including rainfall threshold models and expert-based or physically based methods, provide valuable insights but suffer from several limitations. Expert-based models rely heavily on subjective judgment, which can reduce reproducibility and scalability. Physically based models offer a mechanistic understanding of slope stability and hydrological processes but often require detailed geotechnical data and complex calibration, limiting their applicability at regional scales (Guo et al., 2025; Ye et al., 2025). Despite efforts to mitigate these risks, landslides continue to cause significant loss of life. The lack of an advanced predictive model tailored to Malaysia's unique landscape, rainfall patterns, and urbanization challenges leaves many high-risk areas vulnerable to strategies (Akter et al., 2019; Rosly et al., 2022; Zhang et al., 2023).

In recent years, machine learning (ML) approaches have been increasingly adopted in landslide prediction due to their ability to model nonlinear relationships among environmental variables and process large-scale datasets (Lima et al., 2022; Liu et al., 2023; Prakash, 2025). These models enable the integration of rainfall patterns, soil conditions, and other environmental indicators to improve predictive performance (Tehrani et al., 2022; Vung et al., 2023). Such capabilities support the development of data-driven early warning systems, particularly in rainfall-induced landslide contexts relevant to Malaysia (Alqadhi et al., 2024; Prakash, 2025; Sharma et al., 2022). ML has been widely applied in landslide prediction and often outperforms traditional statistical approaches (Ghayur Sadigh et al., 2024; Huang et al., 2020). However, key challenges remain, including severe class imbalance in landslide datasets, limited temporal validation across unseen scenarios, and reduced generalizability for operational early warning. Although rainfall and soil wetness are recognized as primary triggering factors, existing models frequently rely on simplified indicators or lack systematic validation across different temporal phases.

This study develops a predictive model that leverages ML to provide actionable insights for landslide risk assessment by integrating precipitation, soil wetness, forest loss, and environmental features. To address the scarcity of landslide samples, synthetic data augmentation is incorporated to balance the minority class and improve generalisability. The emphasis on interpretable modelling, rather than highly opaque deep-learning architectures, ensures that the resulting predictions remain practical and usable for regional authorities. By explicitly addressing data imbalance, integrating hydrologically relevant variables, and rigorously validating model performance across unseen datasets, this study contributes a practical, interpretable, and generalizable early warning framework that bridges the gap between predictive analytics and real-world landslide risk management. To achieve this aim, the research is guided by the following objectives:

- To identify and analyse the key temporal factors contributing to landslides in Malaysia, with particular emphasis on accumulated rainfall and soil wetness dynamics.
- To develop a ML–based probabilistic prediction model for temporal landslide risk assessment using rainfall and soil wetness data.

Beyond methodological contributions, landslide risk prediction has important implications for sustainable development, particularly in rapidly urbanising and environmentally sensitive regions such as Malaysia. Accurate and timely identification of landslide risk supports disaster risk reduction, protects vulnerable communities, and informs land-use planning and infrastructure development. In this context, developing reliable and interpretable predictive models is essential not only for technical performance but also for enabling practical, cost-effective, and sustainable early-warning systems.

2. Related Works

Recent literature highlights the challenges inherent in rainfall-driven and image-based landslide early-warning systems. Guzzetti et al. (2020), after reviewing 26 LEWSs worldwide, concluded that most schemes rely on simple rainfall thresholds and therefore do not generalise well across different terrains and climatic zones. Ha et al. (2023)

extended this idea by combining a spatial multi-criteria evaluation model with a rainfall threshold and historical landslide inventory; however, they acknowledged that empirical thresholds lack transferability and may miss factors such as soil saturation, seismic activity, or human interventions.

Ligong et al. (2022) attempted to calibrate rainfall thresholds for Malaysia using up-to-date datasets but noted that precipitation alone cannot capture the complex interactions of geology, land use, slope, and soil conditions, particularly when rainfall and landslide records are sparse. Qin et al. (2021) employed deep learning via distant domain transfer learning to compensate for a lack of labelled images. However, their results underscored that landslide susceptibility is highly context-specific and that limited imagery and data imbalance hinder robust validation. Collectively, these studies show that rainfall thresholds and image classification, while useful, suffer from poor transferability and data limitations. These limitations motivated our decision to pursue a machine-learning model augmented with synthetic data, which aims to overcome class imbalance and improve generalisation across unseen conditions.

Guo et al. (2025) expanded the traditional threshold concept by integrating multi-timescale rainfall metrics and effective recharge, demonstrating improved sensitivity to prolonged rainfall conditions. However, their model remains constrained by a rainfall-only perspective and scenario-based simulation, limiting its transferability to diverse geomorphological settings. Similarly, Huang et al. (2020) explored the use of LSTM deep-learning models to capture temporal dependencies in rainfall-induced landslides. However, the approach remains heavily dependent on large, high-quality time-series datasets and offers limited interpretability. These studies highlight that rainfall-triggered landslide prediction still faces challenges related to data availability, generalizability, and the integration of multi-source environmental factors, reinforcing the need for more robust and adaptable prediction frameworks.

Zulkafli & Abd Majid (2024) employed logistic regression with geospatial factors to map landslide-prone areas in Kuala Lumpur. Their work focused solely on spatial mapping without addressing temporal risk or class imbalance. Al-Najjar et al. (2021) used multiple ML models, including logistic regression, XGBoost, and ANN, combined with feature-transformation preprocessing, improving susceptibility prediction performance. However, their study remained limited to spatial susceptibility and did not consider temporal dynamics or interpretability of complex models. Similarly, Nhu et al. (2020) applied ensemble ML methods such as AdaBoost with remote sensing and terrain data for tropical landslide susceptibility mapping. However, their work emphasized spatial prediction only and did not evaluate temporal risk patterns or provide an interpretable framework suitable for early-warning applications.

Recent advances in ML have significantly improved the accuracy of landslide prediction. As shown in Table 1, studies using Random Forest (Zhang et al., 2023), Support Vector Machines, and hybrid neural networks demonstrate strong predictive capability when high-quality inventories are available. Deep learning models such as CNN and U-Net have also been applied for landslide detection in imagery-based datasets (Liu et al., 2023). More recent efforts include multi-timescale rainfall and effective recharge indicators to refine threshold-based prediction (Guo et al., 2025) and LSTM models to capture temporal dependencies in rainfall-induced landslides (Huang et al., 2020). Complementing these works, several studies in Malaysia and tropical environments have applied ML for landslide susceptibility mapping: logistic regression with geospatial factors for Kuala Lumpur (Zulkafli & Abd Majid, 2024), ML models with feature transformation in the Cameron Highlands (Al-Najjar et al., 2021), and ensemble methods using remote sensing and terrain data (Nhu et al., 2020).

Table 1. Thematic summary of related landslide prediction studies

Research Theme	Studies	Key Contributions	Common Limitations
Rainfall Threshold & Hydrological Models	Guo et al. (2025); Ha et al. (2023); Ligong et al. (2022)	Develop rainfall-based indicators and empirical thresholds for early warning; incorporate multi-timescale rainfall and recharge effects	Focus primarily on rainfall; limited integration of soil or geotechnical factors; static thresholds; limited generalizability; weak real-time adaptability
Operational Landslide Early Warning Systems	Guzzetti et al. (2020)	Global review of LEWS implementation and rainfall-based operational systems	Limited real-time capability; coarse spatial resolution; poor transferability across terrains
Deep Learning & Image-Based Detection	Huang et al. (2020); Qin et al. (2021)	Apply CNN/LSTM for spatial and temporal landslide detection; capture nonlinear patterns	Require large labeled datasets; class imbalance; black-box nature; limited interpretability; weak cross-regional validation
Spatial Susceptibility Mapping (ML-based)	Al-Najjar et al. (2021); Nhu et al. (2020); Zulkafli & Abd Majid (2024)	Use ML and geospatial conditioning factors to generate susceptibility maps	Focus on spatial zoning rather than temporal prediction; limited validation on unseen time periods; data imbalance often unaddressed; reduced operational interpretability

Despite substantial progress, the existing literature reveals three key limitations. First, rainfall-threshold and

hydrological models primarily rely on precipitation variables, limiting their ability to capture the complex interactions among geological, environmental, and anthropogenic factors. Second, many machine-learning and deep-learning approaches focus on spatial susceptibility mapping, with limited attention to temporal validation and real-world predictive deployment. Third, data imbalance and scarcity—particularly in regions such as Malaysia—remain insufficiently addressed, often leading to poor detection of medium- to high-risk events. To address these gaps, this study proposes a temporally validated probabilistic prediction framework that integrates tabular environmental data with synthetic data augmentation to mitigate class imbalance. By adopting an interpretable multiple logistic regression model and validating performance on unseen temporal scenarios, the proposed approach aims to provide a more generalisable, operationally relevant, and cost-effective solution for landslide risk assessment in data-constrained environments.

3. Methodology

This study adopts a three-stage methodological framework that integrates multi-source environmental datasets into a unified landslide prediction pipeline tailored to Malaysia's climatic and geomorphological context. The framework is designed to support temporally structured, event-based probability estimation under highly imbalanced conditions. Figure 1 presents a three-stage ML workflow for predictive modelling. The Data Collection stage sources environmental and geospatial data from NASA Power Data Access Viewer, Global Forest Watch, and NASA's Open Data Portal. The Data Preparation stage involves exploratory data analysis (EDA) with heatmaps, Conditional Tabular Generative Adversarial Network (CTGAN)-based data augmentation to address class imbalance, and dataset balancing. Finally, the Model Development & Evaluation stage implements multiple algorithms (Logistic Regression, XGBoost, LightGBM, CatBoost), compares their performance, and validates generalization capability.

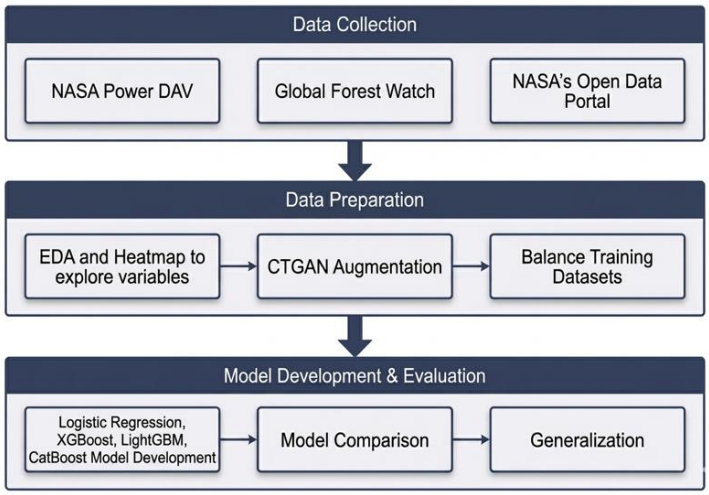

Figure 1. A three-stage methodological framework

3.1 Data Collection

This study integrates multiple environmental data sources to construct a unified landslide prediction dataset for Malaysia. While the use of rainfall, soil moisture, forest loss, and historical landslide information is well established in landslide research, this study focuses on how these commonly used features are operationalized within a district-level, temporally driven predictive framework. In contrast to prior Malaysian studies that predominantly rely on single-source inputs or static susceptibility mapping, the integrated dataset is used to support event-based probability estimation across time windows under conditions of sparse and imbalanced observations.

As supported by the literature review, this study uses precipitation (mm) and soil wetness data to determine the risk of landslides in a particular district in Malaysia. For forest cover loss data, it is collected from Global Forest Watch (https://www.globalforestwatch.org/) based on each district in Malaysia listed on the website. This is totalled at 144 districts, ranging from 2002 to 2023, which is a total of 22 years. District-level forest loss values were extracted for 144 districts in Malaysia, covering the 22-year period from 2002 to 2023. GFW provides annual tree-cover loss derived from Landsat satellite imagery, making it suitable for long-term environmental change assessment. Forest disturbance has been identified in previous ML-based landslide studies as a significant anthropogenic factor, and its integration into the dataset enhances the predictive realism of the model.

Daily precipitation (mm) and soil-wetness variables were sourced from NASA Power Data Access Viewer

(https://power.larc.nasa.gov/data-access-viewer/). NASA POWER is commonly used in environmental machine-learning applications due to its consistent satellite-derived climatic measurements. Three soil-wetness indices were collected to capture moisture conditions at different depths:

- GWETTOP – Surface soil wetness (0–5 cm)
- GWETROOT – Root zone soil wetness (0–100 cm)
- GWETPROF – Profile soil moisture extending to bedrock

Data were retrieved for the same 144 districts, covering 1 January 2007 to 31 December 2014. This time window was selected to align with the availability of Malaysia's historical landslide inventory, ensuring temporal consistency across datasets.

Historical landslide event data were obtained from NASA's Global Landslide Catalog, accessed through NASA's Open Data Portal (https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog-Not-updated-/h9d8-neg4). The Global Landslide Catalog contains 6,788 global landslide records from 2007 to 2014, of which 112 events occurred in Malaysia. These 112 events were geocoded and mapped to their corresponding districts to serve as the ground-truth labels (landslide = 1, no landslide = 0) required for supervised machine-learning training. Although the dataset is severely imbalanced, this characteristic reflects Malaysia's real-world landslide frequency and motivates the use of synthetic minority oversampling in later stages of the methodology.

3.2 Data Preparation

3.2.1 Data integration and restructuring

The integration of a globally recognised landslide inventory addresses the need to employ data sources consistent with modern ML-based susceptibility studies (e.g., Random Forest, SVM, CNN, hybrid models). By harmonising climatic, hydrological, forest-disturbance, and historical landslide indicators, this study constructs a comprehensive dataset that is compatible with state-of-the-art machine-learning approaches in natural hazard prediction. The integration of multi-depth soil-wetness variables differentiates this study from most Malaysian landslide research, which typically uses only rainfall thresholds. This layered approach is consistent with recent ML literature showing that subsurface moisture conditions significantly improve landslide prediction accuracy.
Three datasets—(i) historical landslide events from NASA's Global Landslide Catalog, (ii) precipitation and soil-wetness indicators from NASA POWER, and (iii) forest-loss data from Global Forest Watch—were merged into a unified analytical dataset. Because each source provides data in different temporal resolutions and formats, all records were harmonised by State, District, Year, and Day-of-Year (DOY). The resulting structured dataset contains the following key variables:

- Climatic: PRECTOTCOR (Daily Precipitation), 3-day and 15-day accumulated rainfall
- Hydrological: GWETTOP (Surface Wetness), GWETROOT (Root-zone Wetness), GWETPROF (Profile Moisture)
- Environmental: Accumulated Forest Loss
- Target: Landslide Occurrence (1 = landslide, 0 = no landslide)

Because several districts recorded more than one landslide on the same date, events were consolidated, reducing the 112 original cases to 100 unique daily occurrences.

3.2.2 Exploratory data analysis and handling imbalanced data

An EDA was conducted to understand the landslide scenario in Malaysia. Figure 2 established an overview of Malaysia's landslide-related environmental data (2007–2014). This analysis reveals significant spatial and temporal variability. Sabah and Kuala Lumpur recorded the highest event counts. At the same time, seasonal analysis demonstrates a strong monsoon-driven pattern: landslide frequency peaks during November–January and May, coinciding with maximum 15-day accumulated rainfall (~170 mm) and elevated surface soil wetness (GWETT > 0.94). Conversely, June–August represents a distinct low-hazard period with minimal rainfall (~100 mm) and reduced soil moisture (~0.88). These findings underscore the critical role of rainfall-triggered soil saturation in landslide occurrence, offering a foundation for spatially targeted and seasonally adaptive hazard management and early warning systems in tropical regions.

A comprehensive EDA was conducted on 420,778 records, within which only 100 correspond to actual landslide events. This severe class imbalance highlights the rarity and unpredictability of landslides, but the EDA also reveals consistent spatial and temporal patterns. Landslides cluster in states such as Sabah, Kuala Lumpur, and Sarawak, and frequently occur during monsoon-intensified months such as February, June, November, and December. The rainfall correlation variable (PRECTCORR) trends upward during these months, reinforcing the well-established linkage between accumulated rainfall and slope failures. These data-driven observations validate existing geotechnical knowledge and underscore rainfall as the primary landslide driver in the Malaysian context.

The dataset exhibits extreme class imbalance, with landslide days representing a very small fraction of total observations (Figure 3). Such an imbalance can bias conventional classifiers toward majority-class predictions, limiting their ability to detect medium- and high-risk scenarios. These findings motivate the need for structured

imbalance-handling strategies prior to model training. Accordingly, a temporally constrained synthetic minority augmentation framework was implemented.

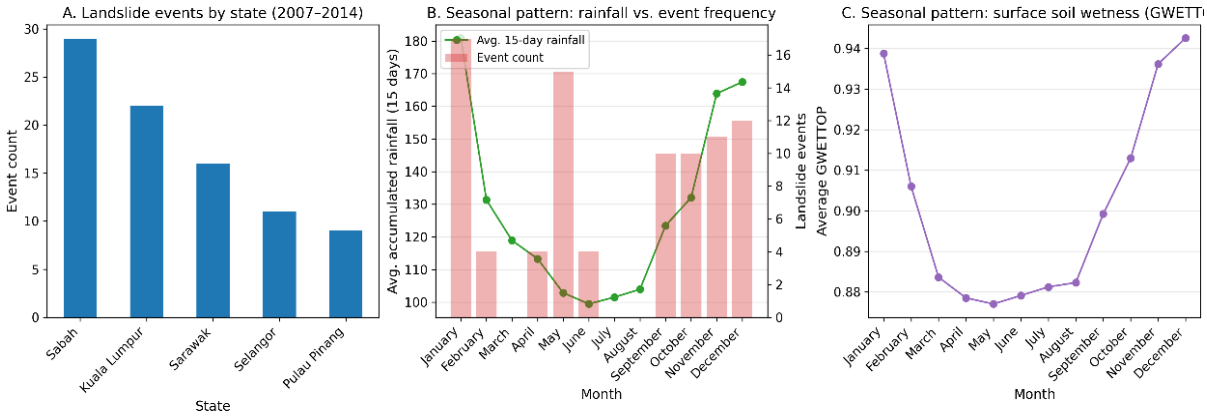

Figure 2. Exploratory overview of Malaysia's landslide-related environmental data (2007–2014)

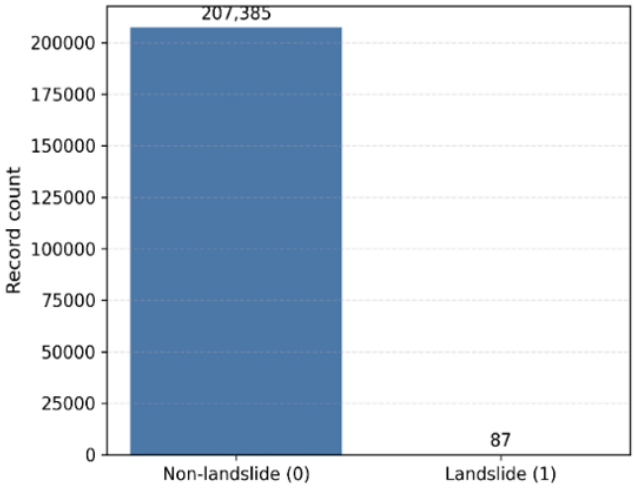

Figure 3. Class distribution (original dataset)

Such an imbalance is common in rare-event prediction and has been shown to cause:

- poor recognition of minority classes,
- overfitting to the majority class, and
- deceptive accuracy despite poor recall (Li et al., 2022).

The EDA also demonstrated that landslide occurrences are spatially concentrated in only a few states—Sabah, Kuala Lumpur, Sarawak, Selangor, and Pulau Pinang—while many districts have zero historical events. This suggests that retaining all states would amplify noise and hinder the model's ability to learn meaningful patterns from the minority class.

The first step to address this data imbalance is to filter out the other states and retain only the top 5 states with the highest landslide occurrence. These include Sabah (29 cases), Kuala Lumpur (22 cases), Sarawak (16 cases), Selangor (11 cases), and Pulau Pinang (9 cases). This significantly changes our dataset from the originally 420,678 records with no landslide and 100 records with landslide to 207,298 records with no landslide and 87 cases with landslide. Although both records have decreased, the class imbalanced ratio has improved from 4207:1 to 2383:1.

Despite earlier steps such as state filtering and feature accumulation, the dataset remained severely imbalanced, with the minority class (landslide occurrence) underrepresented. To address the class imbalance problem, this study employed the CTGAN model, a GAN variant specifically designed for tabular data (Sauber-Cole & Khoshgoftaar, 2022). Before GAN training, all categorical variables ("State", "District", and "Month_Name") were encoded using LabelEncoder, while continuous variables were preserved in their original scale because CTGAN internally models continuous distributions using mode-specific normalization, eliminating the need for manual scaling.

A CTGAN (a GAN variant for tabular data) was used, which follows the same adversarial training principle but includes additional mechanisms such as conditional sampling and mode-specific normalization. Through this

adversarial training process, the generator gradually produces synthetic data points that closely resemble the minority-class distribution. This means the synthetic samples carry realistic combinations of rainfall, soil wetness, and environmental conditions associated with actual landslides. Once the GAN model was trained, synthetic samples were generated until both classes were balanced at 207,298 records each, resulting in a final dataset of 414,596 samples for model development. As illustrated in Figure 4, the dataset was first partitioned chronologically into training and testing subsets prior to any augmentation procedure. All synthetic data generation was restricted to the training partition to ensure that validation and testing sets remained composed exclusively of authentic, unseen observations.

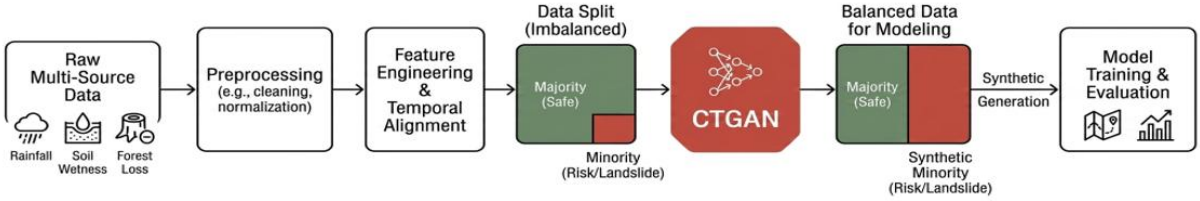

Figure 4. Addressing class imbalance in landslide prediction via Conditional Tabular Generative Adversarial Network (CTGAN) augmentation

The CTGAN architecture follows the default structure, consisting of a generator and discriminator trained adversarially using the Adam optimizer, with the following configuration: embedding dimension = 128, generator dimension = (256, 256), discriminator dimension = (256, 256), batch size = 500, epochs = 300, and pac = 10 (used to stabilize discriminator training). These parameters allow the model to learn nonlinear feature distributions that characterize minority-class (landslide) events. The training dataset contained 165,907 majority-class samples (no landslide) and 70 minority-class samples (landslide). The number of synthetic minority samples generated was matched exactly to the majority class (n_to_generate = 165,837), using ctgan.sample(). The resulting synthetic records were assigned the class label "1" and concatenated with the real samples to form a fully balanced training dataset of 331,814 instances, maintaining all 14 explanatory features. Importantly, synthetic generation was performed only on the training set to prevent data leakage. After balancing, categorical encoders fitted on the training set were applied consistently to the test set.

To provide explicit evidence of the similarity between real and synthetic data, Table 2 presents a comparison of summary statistics for key explanatory variables. The results show that soil wetness variables (GWETTOP and GWETROOT) exhibit closely aligned mean and standard deviation values between the real and synthetic datasets, indicating strong preservation of these environmental characteristics. For rainfall-related variables, the synthetic data show higher mean and variance compared to the real dataset. This reflects the CTGAN model's ability to generate more diverse and extreme rainfall patterns, which are important for representing rare landslide-triggering conditions in an otherwise highly imbalanced dataset.

Table 2. Summary statistics comparison between real and synthetic data

Feature	Dataset	Mean	Standard Deviation
GWETTOP	Real	0.901	0.070
GWETTOP	Synthetic	0.922	0.062
Accumulated Rainfall (15)	Real	128.28	75.52
Accumulated Rainfall (15)	Synthetic	222.43	176.61
Accumulated Rainfall (3)	Real	25.66	24.41
Accumulated Rainfall (3)	Synthetic	48.65	72.29
GWETROOT	Real	0.911	0.087
GWETROOT	Synthetic	0.941	0.085

Overall, these results suggest that the synthetic data preserve the core structure of the real data while enhancing variability in key predictive features. This is further supported by the model's stable performance on unseen real-world datasets (Section 4), indicating that the synthetic data contribute to improved generalisation rather than overfitting.

This detailed pipeline ensures (i) transparent preprocessing steps, (ii) clear GAN architecture disclosure, (iii) reproducible sampling logic, and (iv) a fully documented feature engineering and model-building workflow—directly addressing reviewer concerns regarding missing methodological details. As illustrated in Figure 4, a GAN consists of two competing neural networks, Generator and Discriminator. The generator takes a latent random vector as input, learns to create synthetic observations that mimic real landslide instances, and attempts to "fool" the discriminator. Discriminator receives both real and synthetic samples, learns to differentiate between real and

fake data, and provides feedback that fine-tunes the generator.

This approach provides two key benefits:

- Improved representation of rare events: GANs preserve nonlinear interactions between precipitation, soil moisture, and forest loss—critical for landslide prediction.
- Enhanced model generalisation: The balanced dataset ensures the machine-learning models learn from sufficient examples of both classes, reducing bias toward the majority class.

3.2.3 Accumulation for forest loss and precipitation

Landslides are not immediately triggered by the condition of a single day but often several days or even weeks of conditions, therefore it is more suitable to accumulate the factors of forest loss and precipitation to provide better explanatory variables for our ML model to train. As the data collected from Global Forest Watch only contains primary forest loss (ha) for every single year, and forest is supposed to be observed and accumulated not recalculated every single year, this study accumulates the primary forest loss starting from the year 2002 to reflect long-term vegetation disturbance—a factor frequently highlighted in ML-based landslide studies (Steger et al., 2021). Thus, the new accumulated forest loss for each of the year is used as the accumulated forest loss.

A single day's rainfall condition is not suitable and practical to be used to predict the landslide occurrence, according to Guzzetti et al. (2007). Observing 3 days and 15 days precipitation before a landslide is the most widely used day count and has the highest correlation to the landslide occurrence. Accumulated precipitation of 15 days is used to assess the long-term impact of rainfall towards the occurrence of landslide such as how long-term rainfall will affect the soil toughness and properties, which is a crucial factor of landslides, while accumulated 3 days is used to assess the short-term impact of rainfall towards the occurrence of landslide like immediate heavy rainfall or short thunderstorms.

In addition, landslides are influenced by a multitude of interacting factors, making their relationships highly complex. Only observing the correlation of a single factor in isolation may not properly capture its role or reflect a direct causal relationship with landslide occurrences. Therefore, a correlation heatmap of features may not reliably evaluate their performance, leading to misleading conclusions about their importance (Steger et al., 2021). For that reason, we make a new explanatory variable by combining the precipitation factor and primary forest loss factor and rename it into Combined Factor to observe the relevance and correlation of this factor toward the response variable:

$$\text{Combined\_Factor} = \text{Accumulated Rainfall}_{15d} \times \text{Accumulated Forest Loss}$$

The multiplicative formulation reflects the conceptual understanding that forest loss reduces slope stability, increasing susceptibility of exposed soil, and prolonged rainfall increases deep soil saturation, elevating pore-water pressure. This means that the risk may amplify when both conditions occur simultaneously. Such interaction terms are commonly used in environmental ML to capture nonlinear, synergistic geophysical processes.

The Combined_Factor was included during the exploratory evaluation phase and was tested alongside individual predictors using Correlation analysis, Model-based feature importance, and predictive performance sensitivity checks. However, this factor did not emerge among the top four predictors that demonstrated the strongest and most consistent contribution to landslide classification. For this reason, it was not retained in the final model input set. Although excluded from the final model, constructing this feature was an important diagnostic step that enriched the understanding of rainfall–vegetation interactions within Malaysian geomorphological conditions.

3.2.4 Feature selection

Following synthetic data generation using GAN, feature selection was conducted using correlation analysis to identify variables most strongly associated with the response variable. The top four explanatory variables were selected based on their correlation strength: GWETTOP (Surface Soil Wetness) (0.43), Accumulated Rainfall (15 days) (0.40), Accumulated Rainfall (3 days) (0.35), and GWETROOT (Root Zone Soil Wetness) (0.34). In addition to ranking variables by correlation, pairwise relationships among the selected predictors were examined to ensure that no severe multicollinearity was present. While some degree of association is expected among hydrological variables, no excessively high correlations were observed that would compromise model stability. The selection, therefore, balances predictive relevance with model simplicity, supporting robust and interpretable model development.

3.3 Model Development and Evaluation

Given the relatively limited and imbalanced nature of landslide datasets in Malaysia, model selection prioritised stability, interpretability, and robustness over complexity. Logistic regression was therefore considered a suitable baseline approach, as it performs well on small-to-moderate tabular datasets, provides probabilistic outputs for risk interpretation, and is less prone to overfitting compared to more complex models when data are sparse. In

addition, its transparent structure supports operational use in early-warning systems, where model interpretability is essential. To evaluate whether increased model complexity would yield performance gains, logistic regression was compared with several gradient boosting algorithms.

Six different models have been developed for comparison purposes (Table 3). Top 4 features identified in the correlation heatmap are applied for model training. A total of 4 gradient boosting algorithms and 2 logistic regression models of different situations are chosen for model development.

Table 3. Machine learning (ML) algorithms summary

Model Development	Description
Logistic Regression (Before Balance)	A simple linear model is used to predict the probability of an event occurring using a logistic function. When applied before balancing the dataset, it may be biased toward the majority class due to class imbalance.
Logistic Regression (After Balance)	The same as above, but applied to a balanced dataset using techniques of synthetic data generation of GAN. This helps improve the model's ability to predict the minority class.
LightGBM (GBM) (After Balance)	A highly efficient and fast boosting algorithm designed for large datasets. It uses techniques like leaf-wise tree growth and histogram-based learning for speed and scalability.
XGBoost (GBM) (After Balance)	An optimized version of gradient boosting provides state-of-the-art performance in many ML competitions.
CatBoost (GBM) (After Balance)	A gradient boosting algorithm specifically designed to handle categorical data effectively without the need for extensive preprocessing.
Historical Gradient Boosting (GBM) (After Balance)	A version of gradient boosting that speeds up the training process by binning continuous variables into histograms, which reduces the computational complexity.

To systematically assess the trade-off between model simplicity and predictive performance, logistic regression was evaluated alongside four gradient boosting models—LightGBM, XGBoost, CatBoost, and Histogram-Based Gradient Boosting were trained on the balanced dataset to evaluate whether more complex, nonlinear algorithms could offer improved performance. These models were selected because they are widely used in landslide-susceptibility and hazard-prediction research, particularly in environments where predictor interactions may be nonlinear. All models were trained using consistent cross-validation and hyperparameter settings to ensure comparability.

This modelling setup allows systematic comparison between simple interpretable models and more expressive boosting-based classifiers, enabling evaluation of both predictive accuracy and temporal generalisation under real-world conditions. Performance assessment and validation strategies are detailed in the following section.

Figure 5 shows that there is a calculation of the top 4 features in a correlation heatmap to contribute as a "*z*" value towards our Logistic Regression model through this multiple linear regression equation. After the *z* value is fitted into the logistic regression model, the results are represented as a *P* value, and it will be categorised into three different risk levels. When *P* value is lower than 0.4, the risk of having a landslide will be categorised as low risk; when *P* value is within the range of 0.4 to 0.6, it is categorised as medium risk. Lastly, when *P* value is higher than 0.6, it will be categorised as a high risk of having landslides.

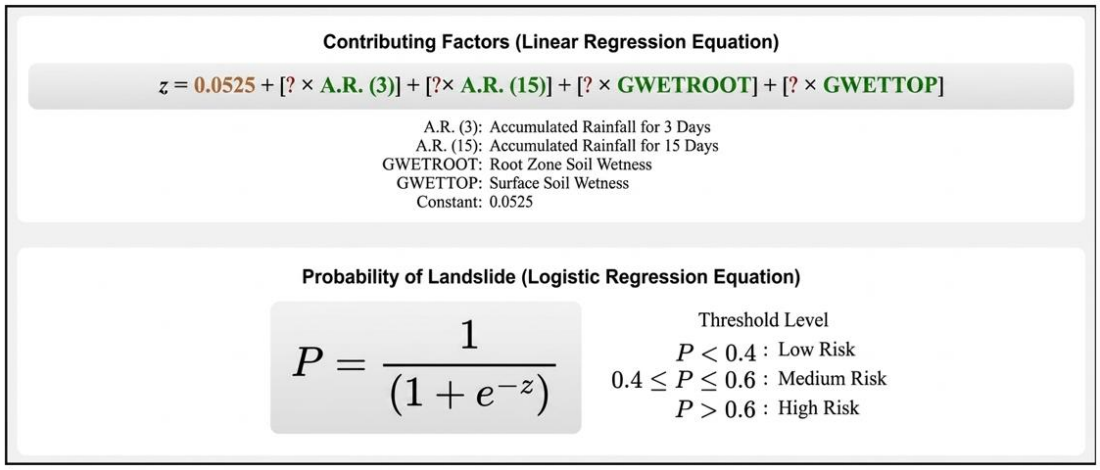

Figure 5. Logistic regression probability formulation and risk categorisation
Source: Based on Huangfu et al. (2021).

The risk thresholds used in this study were adopted from Huangfu et al. (2021) to ensure consistency with established practices in landslide risk classification. While these thresholds were not specifically calibrated for the Malaysian context, they provide a reasonable baseline aligned with prior studies. Small variations in threshold values are unlikely to materially affect the overall conclusions, as the key findings are driven by consistent differences in predicted probability patterns across models rather than by exact cut-off values.

To ensure rigorous evaluation and avoid information leakage, data partitioning was performed using a two-stage procedure:

- an internal train–test split for model development, and
- three independent unseen datasets for external validation.

4. Experiments and Results

After class balancing using CTGAN, the dataset contained 414,596 samples with equal landslide and non-landslide classes. An 80/20 split was used; the training set includes 331,814 real and synthetic samples, and the testing set includes 41,495 real samples. Synthetic GAN samples were used exclusively in the training set to address class imbalance, whereas the test set contained only real events to ensure unbiased performance assessment.

The analysis of the model evaluation metrics highlights the performance of several ML models, including LightGBM, XGBoost, HistGradientBoosting, CatBoost, and Logistic Regression (both before and after GAN balancing). This section compares the performance metrics of all six ML models and identifies the most suitable algorithm for this study's expected outcomes. The evaluation focuses on classification performance using standard metrics, including accuracy, precision, recall, and Receiver Operating Characteristic (ROC)-based measures, to assess the ability of each model to distinguish between landslide and non-landslide events.

Table 4 shows that XGBoost achieved the best overall performance with an accuracy of 99.86%, precision of 0.9971, and recall of 1.0000, making it the most effective model for identifying landslide occurrences. LightGBM and HistGradientBoosting also performed comparably well. Logistic Regression (Pre-GAN) showed perfect accuracy but extremely poor precision and recall due to its inability to handle imbalanced data. After GAN balancing, Logistic Regression improved significantly, though it still lagged behind the gradient boosting models. Overall, XGBoost is the best-performing model offering the highest accuracy, precision, recall, and variance explanation. While Logistic Regression (Post-GAN) showed noticeable improvements compared to its pre-GAN counterpart, it is still not competitive with the gradient boosting models.

Table 4. Classification metrics

Algorithm	Accuracy	Precision	Recall
Logistic Regression (Pre-GAN)	1.0000	0.5000	0.5000
Logistic Regression (Post-GAN)	0.7300	0.7300	0.7300
LightGBM	0.9997	1.0000	1.0000
XGBoost	0.9986	0.9971	1.0000
HistGradientBoosting	0.9984	0.9968	1.0000
CatBoost	0.9971	0.9943	1.0000

4.1 Diagnostic Experiment: Verifying Model Behaviour Using a Generative Adversarial Network-Balanced Dataset

To further examine whether the near-perfect performance of the gradient boosting models in Experiment 1 (Original Imbalance Datasets) was due to genuine predictive capability or overfitting, a second diagnostic experiment (CTGAN Balance datasets) was conducted using a fully balanced dataset generated via CTGAN. In this experiment, the minority class (landslide occurrences) was synthetically expanded to match the majority class, resulting in an equal representation of both classes while preserving their joint feature distributions.

After retraining all models on this balanced dataset, the gradient boosting models (LightGBM, XGBoost, CatBoost, and HistGradientBoosting) continued to report extremely high training accuracy (>99%), a pattern already observed in Experiment 1. However, unlike the inflated accuracy metrics, the ROC curves and Precision-Recall curves in Experiment 2, as shown in Figure 6, revealed far more modest separability between the positive and negative classes. In particular, the PR curves showed very low average precision scores, indicating that the models struggled to correctly detect minority-class events even when trained on artificially balanced data.

To further examine model generalisation, we compared model behaviour across training-based metrics and external validation scenarios. While gradient boosting models achieved near-perfect accuracy (>99%) on the balanced training and test datasets, their performance degraded substantially when evaluated on unseen temporal data, where they consistently failed to identify landslide events and defaulted to low-risk predictions. This discrepancy between in-sample performance and out-of-sample behaviour indicates limited generalisation and suggests that the models are capturing dataset-specific patterns rather than robust predictive signals. In contrast,

Logistic Regression (Post-GAN) maintained consistent performance across both evaluation settings, demonstrating greater stability and reliability under real-world conditions.

This divergence between accuracy and discrimination metrics suggests that the high performance reported by the boosting models in Table 4 was not indicative of true generalization, but rather a consequence of overfitting and model complexity. The synthetic balancing procedure removed the class imbalance, yet the models still failed to produce reliable probabilistic separation, demonstrating that the core limitation lies in model expressiveness rather than class scarcity. In contrast, Logistic Regression (Post-GAN), although not achieving the same inflated accuracy levels, displayed more stable ROC and PR behaviour in both experiments and continued to generalize better on external unseen datasets (Table 5). This indicates that Logistic Regression provides more reliable decision boundaries for this environmental dataset, reinforcing its selection as the primary operational model for landslide-risk prediction.

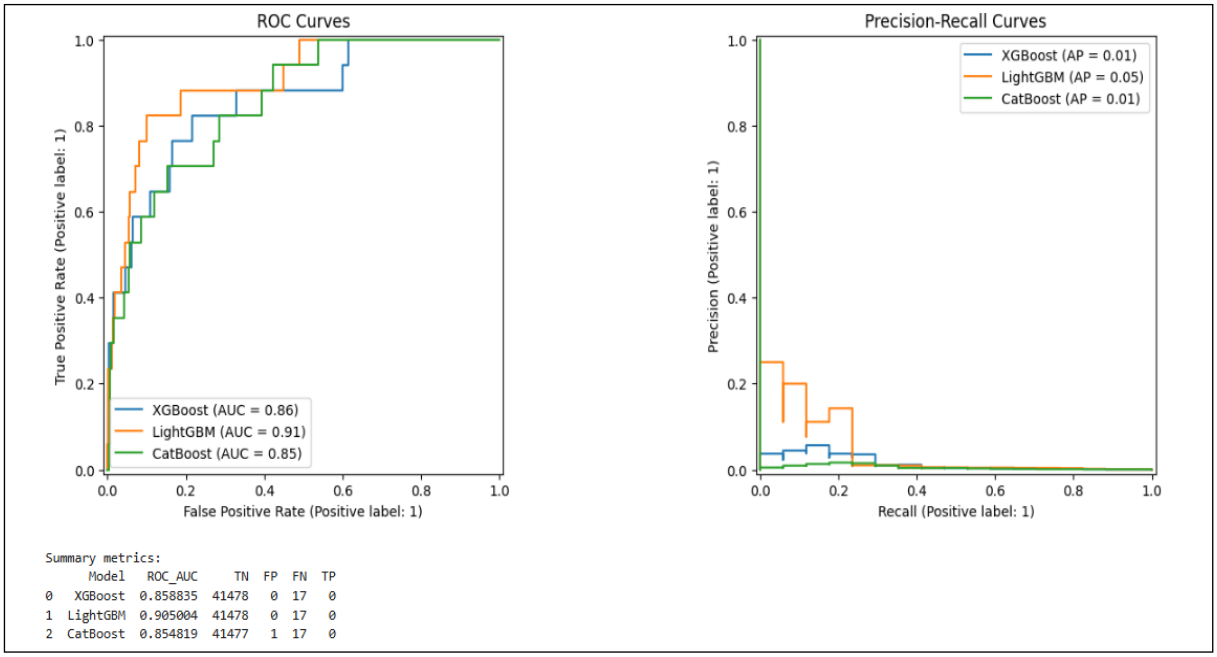

Figure 6. Receiver Operating Characteristic (ROC) and Precision-Recall curves

Table 5. Comparison analysis Experiment 1 vs. Experiment 2

Model	Experiment 1 Accuracy	Experiment 1 Issues	Experiment 2 Accuracy	Experiment 2 ROC-AUC	Experiment 2 Precision-Recall-AUC	Interpretation
Logistic Regression (Post-GAN)	~0.73	Poor handling of imbalance; stable generalization	Similar (~0.73)	Moderate, stable	Moderate	Best overall generalization; consistent behaviour across experiments
LightGBM	>0.999	Clear overfitting; perfect predictions implausible	>0.999	Modest (~0.85–0.90)	Very low	Severe overfitting; poor Precision-Recall-AUC despite high accuracy
XGBoost	~0.998	Overfitting; fails on unseen landslide cases	~0.998	Modest (~0.86)	Very low (~0.01)	Cannot distinguish classes well despite perfect accuracy
CatBoost	~0.997	Overfitting; sensitivity collapsed	~0.997	Modest (~0.85)	Very low	Over-relies on majority trends; limited minority detection
HistGradientBoosting	~0.998	Inflated accuracy; no recall on minority	~0.998	Similar (~0.84–0.88)	Very low	Suggests boosting-family overfitting behaviour

The results from the diagnostic experiment strengthen the interpretation that boosting methods are prone to overfitting on this type of environmental tabular data, whereas logistic regression maintains robustness, interpretability, and superior generalizability.

The comparison highlights that accuracy alone is insufficient for evaluating models trained on imbalanced environmental datasets. Discrimination metrics and validation across multiple experimental conditions reveal that simpler, interpretable models like logistic regression can outperform more complex boosting algorithms in real-world generalizability.

4.2 Validation with Unseen Data

To evaluate temporal robustness, predictive stability, and generalizability, three unseen datasets were constructed from real-world samples that were entirely excluded from model training and testing:

(1) Unseen Batch 1 (±6-Month Window): Contains landslide and non-landslide events occurring within six months before and after historical Malaysian landslide dates. This data was used to assess sensitivity near real hazard periods.

(2) Unseen Batch 2 (Before / During / After Event Segments): Includes samples from three months before the event, the actual event date, and three months after the event. This data was used to evaluate temporal generalization around an event lifecycle.

(3) Unseen Batch 3 (Random Non-Landslide Sample: Contains only non-landslide records sampled from diverse months and districts. This data was used to assess false-positive behaviour under normal conditions.

K-fold CV was not adopted because temporal mixing would allow pre-event and post-event samples to appear in both training and validation folds, which violates the principles of realistic landslide forecasting. Instead, temporal separation preserves the real-world operational context in which future conditions differ from past training data. All unseen batches were temporally independent from the 80/20 split and were used exclusively for external validation.

As shown in Table 6, the Logistic Regression (Post-GAN) model correctly predicts three out of five landslide cases, whereas the other models fail to identify any, consistently defaulting to a low-risk prediction. For non-landslide cases, however, all models—including Logistic Regression (Post-GAN) perform accurately. This suggests that the other five models may be biased toward predicting low risk.

Table 6. First validation of unseen data (+/-6 months window)

Machine Learning	**Logistic Regression (Before Balance)**		**Logistic Regression (After Balance)**		**LightGBM**		**XGBoost**		**CatBoost**		**HistGradient Boosting**	
Actual	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**
1	0.0003	LOW	0.321	LOW	-0.0029	LOW	-0.0003	LOW	-0.0051	LOW	-0.0056	LOW
1	0.0004	LOW	0.5700	**MEDIUM**	-0.0029	LOW	-0.0003	LOW	-0.0051	LOW	-0.0056	LOW
1	0.0004	LOW	0.7399	**HIGH**	-0.0043	LOW	0.0011	LOW	-0.0063	LOW	-0.0056	LOW
1	0.0004	LOW	0.5191	**MEDIUM**	-0.0029	LOW	-0.0003	LOW	-0.0051	LOW	-0.0056	LOW
1	0.0003	LOW	0.3662	LOW	-0.0029	LOW	-0.0003	LOW	-0.0051	LOW	-0.0056	LOW
Predicted Number	0/5		3/5		0/5		0/5		0/5		0/5	
Actual	***P***	**RISK**			***P***	**RISK**			***P***	**RISK**		
0	0.0002	**LOW**	0.2466	**LOW**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0008	**LOW**	0.7003	HIGH	0.0028	**LOW**	0.0025	**LOW**	-0.0158	**LOW**	-0.0043	**LOW**
0	0.0009	**LOW**	0.7702	HIGH	-0.0109	**LOW**	0.0004	**LOW**	-0.0199	**LOW**	-0.0076	**LOW**
0	0.0005	**LOW**	0.5983	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0002	**LOW**	0.3278	**LOW**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
Predicted Number	5/5		3/5		5/5		5/5		5/5		5/5	

Note: Low Risk: $P < 0.4$; Medium Risk: $0.4 \leq P \leq 0.6$; High Risk: $P > 0.6$; Highlighted Risk Value = Correct Predicted.
Source: Based on Huangfu et al. (2021).

To investigate whether these patterns persist across different temporal contexts, the second unseen batch applies a more structured time-window evaluation. Rather than combining all samples within a ±6-month range, Batch 2 separates data into three distinct temporal phases—the event day, three months before, and three months after the landslide. This design allows us to assess each model's ability to recognise precursory conditions, detect event-day signatures, and respond to post-event stabilisation, thereby offering a comprehensive assessment of temporal generalisation across the landslide lifecycle.

As shown in Table 7, Logistic Regression (Post-GAN) again outperforms the other models, correctly predicting seven out of ten landslide cases. In contrast, the remaining models classify all cases as low risk, even on the actual

landslide day. Logistic Regression (Post-GAN) also maintains strong performance on non-landslide days, demonstrating balanced predictive ability. The third batch contains only non-landslide cases. As shown in Table 8, all models—including Logistic Regression (Post-GAN)- perform exceptionally well, correctly predicting all cases as low risk. This indicates that while all models can handle non-landslide conditions effectively, they struggle with scenarios involving actual landslide events.

Table 7. Second validation of unseen data

Machine Learning	**Logistic Regression (Before Balance)**		**Logistic Regression (After Balance)**		**LightGBM**		**XGBoost**		**CatBoost**		**HistGradient Boosting**	
Actual Event	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**
1	0.0008	LOW	0.7226	**HIGH**	-0.003	LOW	-0.0003	LOW	-0.0051	LOW	-0.006	LOW
1	0.0002	LOW	0.3333	LOW	-0.003	LOW	-0.0003	LOW	-0.0051	LOW	-0.006	LOW
1	0.0007	LOW	0.7713	**HIGH**	-0.003	LOW	-0.0003	LOW	-0.0053	LOW	-0.006	LOW
1	0.0002	LOW	0.1598	LOW	-0.003	LOW	-0.0003	LOW	-0.0051	LOW	-0.006	LOW
1	0.0003	LOW	0.3582	LOW	-0.003	LOW	-0.0003	LOW	-0.0051	LOW	-0.006	LOW
1	0.0006	LOW	0.7487	**HIGH**	-0.003	LOW	-0.0003	LOW	-0.0051	LOW	-0.006	LOW
1	0.0003	LOW	0.519	**MEDIUM**	-0.003	LOW	-0.0003	LOW	-0.0051	LOW	-0.006	LOW
1	0.0004	LOW	0.6766	**HIGH**	-0.011	LOW	-0.0004	LOW	-0.02	LOW	-0.006	LOW
1	0.0035	LOW	0.9003	**HIGH**	0.989	LOW	0.9996	LOW	0.98	LOW	0.992	LOW
1	0.0004	LOW	0.5351	**MEDIUM**	1.989	LOW	1.9996	LOW	1.98	LOW	1.992	LOW
Predicted Number	0/10		7/10		0/10		0/10		0/10		0/10	
Actual (Before 3 months)	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**
0	0.0003	**LOW**	0.3544	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0003	**LOW**	0.3636	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0002	**LOW**	0.1813	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0001	**LOW**	0.1093	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0004	**LOW**	0.6222	HIGH	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0002	**LOW**	0.2061	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0003	**LOW**	0.4972	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0003	**LOW**	0.4436	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0004	**LOW**	0.5325	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0002	**LOW**	0.1622	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
Predicted Number	10/10		9/10		10/10		10/10		10/10		10/10	
Actual (After 3 months)	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**	***P***	**RISK**
0	0.0001	**LOW**	0.0581	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0001	**LOW**	0.1003	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0003	**LOW**	0.4065	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0003	**LOW**	0.4342	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0002	**LOW**	0.3565	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0003	**LOW**	0.4024	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0003	**LOW**	0.3758	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0004	**LOW**	0.1874	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0004	**LOW**	0.5512	**MEDIUM**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
0	0.0004	**LOW**	0.1543	**LOW**	-0.003	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.006	**LOW**
Predicted Number	10/10		10/10		10/10		10/10		10/10		10/10	

Note: Low Risk: $P < 0.4$; Medium Risk: $0.4 \leq P \leq 0.6$; High Risk: $P > 0.6$; Highlighted Risk Value = Correct Predicted.
Source: Based on Huangfu et al. (2021).

For the three batches of unseen data, Logistic Regression (After Balance) demonstrates the best across all unseen datasets, effectively predicting both landslide and no landslide events while maintaining balance across different risk categories. The other five models are strong in terms of metrics on balanced datasets, but fail to capture medium and high-risk scenarios and are prone to predict low-risk scenarios. Logistic Regression After GAN performs better on unseen data, likely due to its simplicity and better generalization, which emphasizes the importance of model selection based on real-world applicability rather than solely based on evaluation metrics.

Despite the relatively limited number of unseen validation cases in each scenario due to the scarcity of well-documented landslide events, the consistent behaviour observed across multiple temporal scenarios provides useful evidence of model generalisation and highlights the practical potential of the proposed approach for real-world applications.

Table 8. Third batch of unseen data

Machine Learning	Logistic Regression (Before Balance)		Logistic Regression (After Balance)		LightGBM		XGBoost		CatBoost		HistGradient Boosting	
Actual Landslide Day	*P*	RISK	*P*	RISK	*P*	RISK	*P*	RISK	*P*	RISK	*P*	RISK
0	0.0002	**LOW**	0.4908	**MEDIUM**	-0.0109	**LOW**	0.0004	**LOW**	-0.0199	**LOW**	-0.0076	**LOW**
0	0.0002	**LOW**	0.3481	**LOW**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0003	**LOW**	0.5833	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0003	**LOW**	0.5909	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0003	**LOW**	0.5276	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0004	**LOW**	0.5555	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0004	**LOW**	0.6375	HIGH	-0.0043	**LOW**	0.0011	**LOW**	-0.0063	**LOW**	-0.0056	**LOW**
0	0.0002	**LOW**	0.4872	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0004	**LOW**	0.5477	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
0	0.0002	**LOW**	0.3711	**MEDIUM**	-0.0029	**LOW**	-0.0003	**LOW**	-0.0051	**LOW**	-0.0056	**LOW**
Predicted Number	10/10		9/10		10/10		10/10		10/10		10/10	

Note: Low Risk: $P < 0.4$; Medium Risk: $0.4 \leq P \leq 0.6$; High Risk: $P > 0.6$; Highlighted Risk Value = Correct Predicted.
Source: Based on Huangfu et al. (2021).

4.3 Spatial Distribution of Predicted Landslide Susceptibility

To complement the temporal validation experiments presented in the preceding sections, the predicted probabilities generated by the Logistic Regression (Post-GAN) model were spatially aggregated at the district level to produce a landslide susceptibility map (Figure 7). Susceptibility values were classified into Low, Medium, and High-risk categories based on predefined probability thresholds. This spatial representation enables geographic interpretation of model-derived probabilities and provides additional insight into the practical applicability of the model beyond numerical performance metrics.

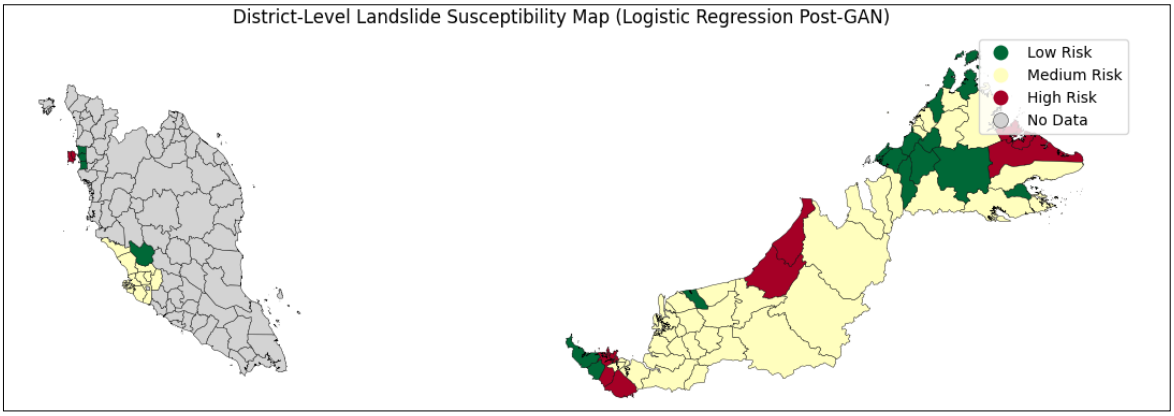

Figure 7. The Malaysia susceptibility map

The spatial distribution reveals coherent clustering of medium- and high-risk districts, particularly within selected regions of Sabah and Sarawak. These districts correspond to areas characterised by elevated rainfall accumulation and soil wetness indices, consistent with the environmental drivers identified in the regression coefficients. The geographic concentration of higher susceptibility zones reflects stable hydrological triggering signals rather than isolated statistical artefacts, suggesting that the model captures physically meaningful relationships between rainfall–soil interactions and landslide occurrence. It was observed that the majority of known landslide occurrences fall within areas classified as medium to high risk by the model, providing preliminary support for the spatial consistency of the predictions. The analysis shows that 80.5% (70 events) of documented landslides occurred in areas classified as “Medium” or “High” risk. Specifically, 25.3% of events were located in High-risk zones (e.g., Miri and Ranau) and 55.2% in Medium-risk zones (e.g., Kuala Lumpur). Conversely, only 10.3% of events fell within Low-risk areas, confirming the model’s predictive accuracy.

Importantly, the spatial pattern aligns with the temporal validation findings reported in Tables 5–7. During unseen event-day evaluations, the Logistic Regression (Post-GAN) model maintained balanced probability outputs and successfully distinguished between landslide and non-landslide scenarios, whereas several boosting-based models exhibited a tendency to default toward low-risk classifications. The presence of geographically distinct medium- and high-risk clusters in Figure 7 further indicates that the model preserves meaningful probability variation across districts, reinforcing its generalisation capability under unseen conditions. Districts not included

in the modelling dataset are displayed as "No Data" to avoid extrapolation beyond observed training support. This conservative representation ensures that susceptibility classifications are restricted to regions with adequate model exposure and prevents overextension of predictive inference.

The integration of temporal validation and spatial susceptibility mapping provides a comprehensive assessment of model robustness and operational relevance. While internal metrics initially favoured boosting algorithms, unseen temporal evaluations demonstrated that Logistic Regression with synthetic balancing offers more stable generalisation under real environmental variability. Spatial aggregation further suggests that predicted risk patterns remain geographically coherent and interpretable at the district level. Together, these findings highlight the importance of evaluating rare-event models across both discrimination performance and spatiotemporal consistency.

5. Discussion

This section focuses on discussing the results and their broader implications for landslide risk assessment and sustainable decision-making.

Table 9 presents the confusion matrix for the logistic regression model. The model correctly identified 45,084 landslide cases (TP) and 45,601 non-landslide cases (TN), demonstrating a strong ability to discriminate between positive and negative events. However, it also produced 17,132 false positives (FP) and 16,641 false negatives (FN). While false positives may lead to unnecessary alerts, they still promote precautionary action, which is generally preferable in high-risk contexts. False negatives, on the other hand, are more critical since they represent missed landslide events; yet, the relatively low FN count suggests the model maintains acceptable reliability for operational deployment.

Table 10 summarizes the evaluation metrics. The model achieved an overall accuracy of 0.73, with balanced precision (0.73) and recall (0.73) for both landslide and non-landslide cases. The ROC-AUC and Precision-Recall-AUC scores (0.80 each) further confirm the model's ability to differentiate between classes while managing the trade-off between sensitivity and specificity. These balanced results indicate that the logistic regression model can provide consistent performance under diverse conditions, an advantage over more complex algorithms that may overfit.

To assess robustness and real-world applicability, the model was validated on two sets of unseen data from 2015 to 2024, comprising both landslide and non-landslide events. For the first validation (6-month window around actual landslides), as shown in Table 11, logistic regression successfully predicted 9 out of 10 non-landslide cases within 3 months before the landslide at the correct risk level of low or medium. As for the actual day of landslide occurrence, the ML model has achieved a 70% accuracy by only wrongly predicting 3 of the cases as a low risk level. Finally, 3 months after the landslide, our model has achieved 100% accuracy in predicting all cases as low- or medium-risk, with no landslide occurring in real life. The average risk probability of pre-landslide and post-landslide days of 0.3478 and 0.2888, respectively, is also included in the risk level of low risk. As for the actual day of landslide occurrence, the average probability of having a landslide falls under medium to near-high risk of 0.5729.

To confirm the model's performance on identifying non-landslide samples and verify the prediction results on the samples collected, a second validation was conducted. The samples collected are all from recent days without having a landslide through a random sampling method. Results in Table 12 show that logistic regression achieved 90% accuracy by only wrongly predicting one of the cases to a high-risk level.

Table 9. Confusion matrix

Confusion Matrix	Positive (1)	Negative (0)
Positive (1)	True Positive (TP) 45084	False Positive (FP) 17132
Negative (0)	False Negative (FN) 16641	True Negative (TN) 45601

Table 10. Logistic regression model performance metrics

Metrics	Performance
Accuracy	0.73
Precision (0)	0.73
Recall (0)	0.72
Precision (1)	0.73
Recall (1)	0.73
ROC-AUC	0.80
Precision-Recall-AUC	0.80

Table 11. First batch of unseen data

3 Months Before Landslide		Actual Landslide		3 Months After Landslide	
Probability	**Risk**	**Probability**	**Risk**	**Probability**	**Risk**
0.3544	Low	0.7226	High	0.0581	Low
0.3636	Low	0.3373	**Low**	0.1003	Low
0.1813	Low	0.7713	High	0.4065	Medium
0.1093	Low	0.1598	**Low**	0.4342	Medium
0.6222	**High**	0.3582	**Low**	0.3565	Low
0.2061	Low	0.7487	High	0.4024	Medium
0.4972	Medium	0.5190	Medium	0.3757	Low
0.4496	Medium	0.6766	High	0.1874	Low
0.5325	Medium	0.9003	High	0.5512	Medium
0.1622	Medium	0.5351	Medium	0.0154	Low
Average Risk	0.3478	Average Risk	0.5729	Average Risk	0.2888

Table 12. Second batch of unseen data

Random Samples Without Landslides	
Probability	**Risk**
0.4908	Medium
0.3481	Low
0.5833	Medium
0.5909	Medium
0.5276	Medium
0.5555	Medium
0.6375	**High**
0.4872	Medium
0.5477	Medium
0.3711	Low
Average Risk	0.5140

Table 13. Comparison with most relevant studies

Study	Study Area / Data Context	Method / Model (s)	Key Performance / Focus	Relevance / Remarks
Zulkafli & Abd Majid (2024)	Kuala Lumpur, Malaysia—urban context	Logistic Regression with geospatial conditioning factors + GIS-based spatial susceptibility mapping	Overall accuracy ≈ 74.1%, sensitivity ≈ 84.7%, specificity ≈ 63.5%	spatial susceptibility mapping only—no temporal risk dynamics. Our study incorporates temporal risk prediction and uses synthetic data balancing.
Al-Najjar et al. (2021)	Cameron Highlands, Malaysia (hilly / highland terrain)	Machine learning (ML) models (XGBoost, Logistic Regression, ANN) + feature-transformation preprocessing	Improved susceptibility mapping performance after feature transformation; exact metrics vary across models	Feature transformation improves ML model performance; our synthetic data approach similarly addresses data imbalance. Focus remains on spatial mapping vs our temporal risk prediction.
Nhu et al. (2020)	Tropical hilly environment (Cameron Highlands, Malaysia)	Ensemble ML models (AdaBoost, decision-tree variants), remote sensing + multiple conditioning factors	Ensemble AdaBoost achieved high AUC (~0.96) in landslide susceptibility mapping	Shows complex ML/ensemble methods give high spatial performance. Our LR + synthetic data emphasizes interpretability, class balance, and temporal risk prediction for early warning.
This study	Malaysia (various regions)	Logistic Regression with synthetic data augmentation to balance classes; evaluated on historical and unseen temporal data (2015–2024)	Accuracy = 0.73, precision = 0.73, recall = 0.73, ROC-AUC = 0.80; validated on unseen data for pre-, during, and post-landslide events	Novelty: combines temporal risk prediction, synthetic data augmentation, and balanced classification. Interpretable, cost-effective pre-screening tool for operational early-warning systems.

After twice the validation with different unseen data, our model has correctly classified 35 out of 40 samples

into the correct risk levels, achieving a prediction accuracy of 87.5% in validating unseen data. The results provide encouraging evidence of the model's effectiveness as an initial screening tool for landslide risk assessment.

To contextualize our work within existing literature, Table 13 summarizes recent studies on landslide susceptibility and risk assessment in Malaysia and tropical environments. While logistic regression and other machine-learning models have been widely applied for spatial susceptibility mapping (Al-Najjar et al., 2021; Nhu et al., 2020; Zulkafli & Abd Majid, 2024), these studies primarily focus on identifying landslide-prone areas without explicitly addressing temporal risk dynamics. Our study extends this body of work by integrating synthetic data augmentation to balance class distributions and applying logistic regression to predict pre-, during, and post-landslide risk over time. Compared to prior works, our approach achieves a balanced performance in precision, recall, and ROC-AUC while remaining interpretable and cost-effective, making it suitable as a pre-screening tool for early-warning systems. This table highlights how our methodology complements and advances existing approaches, demonstrating its practical contribution to landslide risk assessment in Malaysia.

Since the model has been rigorously evaluated through the confusion matrix, performance metrics, and validation on unseen data, the results indicate that our developed model is a potentially useful and interpretable approach for practical applications. It can serve as an effective pre-screening tool for land developers and government agencies to assess the potential risk of landslides in specific areas. The findings further suggest that the multiple logistic regression model shows stable performance within the evaluated scenarios for landslide risk assessment and categorisation.

The findings of this study highlight an important trade-off between model complexity and generalisation in landslide prediction. While advanced gradient boosting models achieved near-perfect performance on balanced datasets, their inability to detect landslide events in unseen temporal scenarios suggests that high in-sample accuracy does not necessarily translate into reliable real-world performance. In contrast, the more interpretable logistic regression model demonstrated more stable behaviour across different evaluation settings, indicating that simpler models may offer greater robustness when dealing with limited and imbalanced environmental data. This observation has broader implications for the application of ML in environmental hazard prediction. In data-constrained settings, increasing model complexity does not always lead to improved predictive reliability and may instead amplify sensitivity to dataset-specific patterns. The results therefore, emphasise the importance of combining appropriate data balancing techniques with models that prioritise stability and interpretability, particularly when the objective is operational deployment rather than purely predictive performance.

Beyond the specific case of Malaysia, the proposed approach may be relevant to other regions facing similar challenges of limited landslide inventories and heterogeneous environmental conditions. The integration of synthetic data generation with interpretable modelling provides a practical pathway for improving rare-event detection in data-scarce environments. From a sustainability perspective, such approaches can support more resilient disaster risk management by enabling earlier intervention, reducing environmental degradation, and minimising socio-economic impacts associated with landslide events.

6. Conclusion

This study introduced a ML–based framework for landslide risk assessment, developed to overcome the limitations of traditional rainfall thresholds and image-based approaches. Through comprehensive evaluation encompassing confusion matrices, probability analyses, ROC, and PR curves, and rigorous validation on multiple unseen datasets, Logistic Regression (After Balance) emerged as the most promising and generalizable model. Unlike gradient-boosting models, which performed exceptionally on the synthetic-balanced dataset but failed to distinguish actual landslide events in real unseen samples, logistic regression maintained meaningful probability variation and produced stable, interpretable risk predictions across diverse temporal conditions.

These findings demonstrate that model simplicity, interpretability, and physical consistency can outperform complex algorithms in geohazard prediction, especially when rare events, distribution shifts, and noisy environmental variables are involved. The results reinforce the importance of prioritizing generalization, robustness, and operational relevance over in-sample accuracy when designing predictive systems for natural hazards.

The proposed framework offers practical value as a pre-screening and decision-support tool for land developers, urban planners, and governmental agencies. By translating environmental variables into actionable low–medium–high risk categories, the system can strengthen multi-criteria early-warning frameworks, guide hillside development approvals, support zoning decisions, and complement existing empirical rainfall-threshold models. With continued refinement, including the integration of geotechnical variables, real-time sensor data, and advanced temporal modelling architectures, the proposed framework holds significant potential to enhance Malaysia's landslide early-warning capabilities. The proposed model can be integrated into a simple decision-support workflow. Environmental inputs such as rainfall accumulation and soil wetness indicators can be collected periodically (e.g., daily or weekly) and fed into the trained model to generate probability-based risk scores. These probabilities can then be categorised into low, medium, or high-risk levels using predefined thresholds. For

example, a sustained increase in predicted risk from low to medium or high levels over consecutive days could trigger early warnings or field inspections by local authorities. In addition, the model outputs can support land-use planning by identifying areas with recurring elevated risk, enabling more informed decisions on infrastructure development and mitigation measures.

This workflow is particularly applicable to regions with limited landslide inventories and imbalanced datasets, where data-efficient and interpretable models are required. The approach is transferable provided that reliable environmental input data (e.g., rainfall and soil moisture indicators) are available, although local calibration of risk thresholds may be necessary to reflect site-specific conditions. As such, the framework offers a practical and scalable solution, while still requiring contextual validation for effective real-world deployment.

From a sustainability perspective, the proposed framework contributes to more resilient and adaptive risk management practices. By improving the detection of medium- to high-risk conditions using a computationally efficient and interpretable model, the approach supports early intervention strategies that can reduce environmental damage, safeguard communities, and minimise economic losses. Moreover, the reliance on tabular environmental data and lightweight modelling makes the framework suitable for deployment in resource-constrained settings, aligning with the principles of sustainable and scalable disaster management.

6.1 Study Implications

The findings of this study carry several important scientific, operational, and policy-relevant implications. First, the Logistic Regression model (After Balance) demonstrates good potential as a cost-effective, interpretable pre-screening tool for identifying landslide-prone conditions before more resource-intensive geotechnical investigations are undertaken. This is particularly valuable for land developers, government agencies, and urban planners who must make data-driven decisions related to zoning, slope management, hillside development approval, and infrastructure placement.

Second, the study shows that ML–derived probability outputs can meaningfully complement existing rainfall threshold or empirical early-warning systems. The model's ability to translate environmental variables into LOW–MEDIUM–HIGH risk categories provides an additional layer of decision intelligence that enhances multi-criteria early-warning frameworks. This hybrid approach aligns with current global trends in disaster risk reduction, where statistical, empirical, and ML-based indicators are combined to increase sensitivity to early hazard signals.

Third, this study highlights the importance of localized, context-specific modelling. Landslide dynamics in Malaysia differ from those in Japan, Europe, or South America due to variation in monsoon-driven rainfall patterns, tropical soil structures, and land-use practices. By training models specifically on Malaysian data—and validating them using unseen samples—the study provides evidence that the value of developing regionalized models rather than relying on generalized or imported systems.

6.2 Limitations and Future Work

Despite promising outcomes, several limitations should be acknowledged. First, the dataset was restricted to the five Malaysian states with the highest historical landslide frequency. While this selection improves the signal-to-noise ratio and mitigates extreme class imbalance, it also limits the spatial generalizability of the model. Areas with differing geological settings, slope profiles, soil structures, rainfall regimes, or land-use patterns may exhibit landslide behaviours not captured in this study.

Second, the study relies primarily on environmental predictors such as rainfall accumulation, soil moisture indices, and forest-loss metrics. The absence of key geospatial and geotechnical variables—including slope gradient, lithology, soil texture, vegetation indices, drainage networks, and anthropogenic modification—may restrict the model's ability to capture deeper causal mechanisms.

Third, although logistic regression demonstrated superior generalization compared to gradient-boosting models, its simplicity also imposes limitations. The model may not fully capture complex nonlinear relationships or temporally evolving precursors inherent in monsoon-driven landslide processes. The study has not yet undergone operational testing. The model has been validated statistically but not evaluated in real-world monitoring environments, where data noise, temporal lag, and sensor variability may influence performance.

Future research will expand the dataset to include additional Malaysian states and varied terrain types to evaluate cross-regional transferability and improve nationwide representativeness. Integrating richer geospatial and geotechnical variables—such as slope angle, lithology, soil type, NDVI, terrain curvature, drainage density, and indicators of human activity—is expected to enhance model robustness and interpretability. Moreover, incorporating real-time data streams from radar rainfall products, in situ sensors, satellite deformation (InSAR), and seismic activity could enable true near-real-time early-warning capabilities. Methodologically, future studies may explore more advanced architectures, including hybrid physics-ML models, temporal neural networks (LSTM, ConvLSTM), attention-based frameworks, and ensemble modelling strategies tailored to local geomorphology.

Finally, while this study focuses on environmental predictors of landslide occurrence, it is important to recognise

that landslide risk is inherently multi-dimensional, encompassing not only hazard likelihood but also social vulnerability and economic exposure. Factors such as population density, infrastructure distribution, and community resilience play a critical role in determining the overall impact of landslide events. The proposed framework therefore represents the environmental hazard component of a broader risk assessment process. Future work could extend this approach by integrating socio-economic indicators to support more comprehensive risk modelling and decision-making aligned with sustainable development objectives.

Author Contributions

Conceptualization, N.M.S., A.K.R., and C.J.C.; methodology, F.M., N.M.S., A.K.R., C.J.C., K.H.M.B., K.E.S., and T.C.W.; software, A.K.R., C.J.C., K.H.M.B., K.E.S., and T.C.W.; validation, F.M., N.M.S., and S.A.A.K.; formal analysis, A.K.R., C.J.C., K.H.M.B., K.E.S., and T.C.W.; investigation, F.M., N.M.S., and S.A.A.K.; resources, F.M., N.M.S., A.K.R., C.J.C., K.H.M.B., K.E.S., T.C.W., and S.A.A.K.; data curation, F.M., N.M.S., and S.A.A.K..; writing—original draft preparation, F.M., N.M.S., A.K.R., C.J.C., K.H.M.B., K.E.S., and T.C.W.; writing—review and editing, F.M., N.M.S., and S.A.A.K.; visualization, A.K.R., C.J.C., K.H.M.B., K.E.S., and T.C.W.; supervision, F.M. and N.M.S.; project administration, F.M., N.M.S., and S.A.A.K.; funding acquisition, F.M., N.M.S., and S.A.A.K. All authors have read and agreed to the published version of the manuscript.

Data Availability

The datasets used in this study are publicly available at the following repositories: Global Forest Watch (https://www.globalforestwatch.org/), NASA Power Data Access Viewer (https://power.larc.nasa.gov/data-access-viewer/), and NASA's Open Data Portal ((https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog-Not-updated-/h9d8-neg4).

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of Generative AI Use in Writing

During the preparation of this work, the authors utilized ChatGPT to improve the readability of the article. Afterward, they reviewed and edited the content as necessary and take full responsibility for the publication's content.

References

Akter, A., Noor, M. J. M. M., Goto, M., Khanam, S., Parvez, A., & Rasheduzzaman, M. (2019). Landslide disaster in Malaysia: An overview. *Int. J. Innov. Res. Dev.*, *8*(6), 292–302. https://doi.org/10.24940/ijird/2019/v8/i6/JUN19058.

Alcántara-Ayala, I. (2025). Landslides in a changing world. *Landslides*, *22*(9), 2851–2865. https://doi.org/10.1007/s10346-024-02451-1.

Al-Najjar, H. A. H., Pradhan, B., Kalantar, B., Sameen, M. I., Santosh, M., & Alamri, A. (2021). Landslide susceptibility modeling: An integrated novel method based on machine learning feature transformation. *Remote Sens.*, *13*(16), 3281. https://doi.org/10.3390/rs13163281.

Alqadhi, S., Mallick, J., Alkahtani, M., Ahmad, I., Alqahtani, D., & Hang, H. T. (2024). Developing a hybrid deep learning model with explainable artificial intelligence (XAI) for enhanced landslide susceptibility modeling and management. *Nat. Hazards*, *120*(4), 3719–3747. https://doi.org/10.1007/s11069-023-06357-4.

Casagli, N., Intrieri, E., Tofani, V., Gigli, G., & Raspini, F. (2023). Landslide detection, monitoring and prediction with remote-sensing techniques. *Nat. Rev. Earth Environ.*, *4*(1), 51–64. https://doi.org/10.1038/s43017-022-00373-x.

Froese, R. & Schilling, J. (2019). The nexus of climate change, land use, and conflicts. *Curr. Clim. Change Rep.*, *5*(1), 24–35. https://doi.org/10.1007/s40641-019-00122-1.

Ghayur Sadigh, A., Alesheikh, A. A., Bateni, S. M., Jun, C., Lee, S., Nielson, J. R., Panahi, M., & Rezaie, F. (2024). Comparison of optimized data-driven models for landslide susceptibility mapping. *Environ. Dev. Sustain.*, *26*(6), 14665–14692. https://doi.org/10.1007/s10668-023-03212-1.

Guo, Z., Cheng, M., Wang, Y., Xu, G., Zhang, Y., & Xu, C. (2025). Landslide hazard prediction under an extreme rainfall scenario by considering multiple timescale rainfalls and effective recharge. *Georisk*, *19*(4), 775–803. https://doi.org/10.1080/17499518.2025.2570863.

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., & Melillo, M. (2020). Geographical landslide early warning systems. *Earth-Sci. Rev.*, *200*, 102973.

https://doi.org/10.1016/j.earscirev.2019.102973.
Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorol. Atmos. Phys.*, *98*(3), 239–267. https://doi.org/10.1007/s00703-007-0262-7.
Ha, N. D., Duong, N. H., Khanh, N. Q., Viet, T. T., Van Vung, D., Van, N. T. H., & Ninh, N. H. (2023). Landslide early warning system based on the empirical approach: Case study in Ha Long City (Vietnam). In *Progress in Landslide Research and Technology* (pp. 209–225). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-39012-8_9.
Huang, F., Cao, Z., Guo, J., Jiang, S., Li, S., & Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena*, *191*, 104580. https://doi.org/10.1016/j.catena.2020.104580.
Huangfu, W., Wu, W., Zhou, X., Lin, Z., Zhang, G., Chen, R., Song, Y., Lang, T., Qin, Y., Ou, P., et al. (2021). Landslide geo-hazard risk mapping using logistic regression modeling in Guixi, Jiangxi, China. *Sustainability*, *13*(9), 4830. https://doi.org/10.3390/su13094830.
Kemarau, R. A., Suab, S. A., Eboy, O. V., Sa'adi, Z., Echoh, D. U., & Sakawi, Z. (2025). Integrative approaches in remote sensing and GIS for assessing climate change impacts across Malaysian ecosystems and societies. *Sustainability*, *17*(4), 1344. https://doi.org/10.3390/su17041344.
Kumari, S., Agarwal, S., Agrawal, N. K., Agarwal, A., & Garg, M. C. (2025). A comprehensive review of remote sensing technologies for improved geological disaster management. *Geol. J.*, *60*(1), 223–235. https://doi.org/10.1002/gj.5072.
Li, Q., Zhao, C., He, X., Chen, K., & Wang, R. (2022). The impact of partial balance of imbalanced dataset on classification performance. *Electronics*, *11*(9), 1322. https://doi.org/10.3390/electronics11091322.
Li, Y. & Duan, W. (2024). Decoding vegetation's role in landslide susceptibility mapping: An integrated review of techniques and future directions. *Biogeotechnics*, *2*(1), 100056. https://doi.org/10.1016/j.bgtech.2023.100056.
Ligong, S., Sidek, L. M., Hayder, G., & Mohd Dom, N. (2022). Application of rainfall threshold for sediment-related disasters in Malaysia: Status, issues and challenges. *Water*, *14*(20), 3212. https://doi.org/10.3390/w14203212.
Lima, P., Steger, S., Glade, T., & Murillo-García, F. G. (2022). Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility. *J. Mt. Sci.*, *19*(6), 1670–1698. https://doi.org/10.1007/s11629-021-7254-9.
Liu, S., Wang, L., Zhang, W., He, Y., & Pijush, S. (2023). A comprehensive review of machine learning-based methods in landslide susceptibility mapping. *Geol. J.*, *58*(6), 2283–2301. https://doi.org/10.1002/gj.4666.
Nhu, V.-H., Mohammadi, A., Shahabi, H., Ahmad, B. B., Al-Ansari, N., Shirzadi, A., Clague, J. J., Jaafari, A., Chen, W., & Nguyen, H. (2020). Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. *Int. J. Environ. Res. Public Health*, *17*(14), 4933. https://doi.org/10.3390/ijerph17144933.
Prakash, I. (2025). Conventional and data-driven models for landslide susceptibility prediction: A comprehensive review. *J. Eng. Anal. Des.*, *7*(3), 38–44. https://doi.org/10.5281/zenodo.17718832.
Qin, S., Guo, X., Sun, J., Qiao, S., Zhang, L., Yao, J., Cheng, Q., & Zhang, Y. (2021). Landslide detection from open satellite imagery using distant domain transfer learning. *Remote Sens.*, *13*(17), 3383. https://doi.org/10.3390/rs13173383.
Rosly, M. H., Mohamad, H. M., Bolong, N., & Harith, N. S. H. (2022). An overview: Relationship of geological condition and rainfall with landslide events at East Malaysia. *Trends Sci.*, *19*(8), 3464–3464. https://doi.org/10.48048/tis.2022.3464.
Sauber-Cole, R. & Khoshgoftaar, T. M. (2022). The use of generative adversarial networks to alleviate class imbalance in tabular data: A survey. *J. Big Data*, *9*, 98. https://doi.org/10.1186/s40537-022-00648-6.
Sharma, T., Singhal, A., Kundu, K., & Agarwal, N. (2022). Machine learning/deep learning for natural disasters. In *Applications of Artificial Intelligence, Big Data and Internet of Things in Sustainable Development* (pp. 91–121). CRC Press.
Steger, S., Mair, V., Kofler, C., Pittore, M., Zebisch, M., & Schneiderbauer, S. (2021). Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling—Benefits of exploring landslide data collection effects. *Sci. Total Environ.*, *776*, 145935. https://doi.org/10.1016/j.scitotenv.2021.145935.
Syafrina, A. H., Norzaida, A., & Shazwani, O. N. (2017). Rainfall analysis in the northern region of Peninsular Malaysia. *Int. J. Adv. Appl. Sci.*, *4*(11), 11–16. https://doi.org/10.21833/ijaas.2017.011.002.
Tehrani, F. S., Calvello, M., Liu, Z., Zhang, L., & Lacasse, S. (2022). Machine learning and landslide studies: Recent advances and applications. *Nat. Hazards*, *114*(2), 1197–1245. https://doi.org/10.1007/s11069-022-05423-7.
Turner, A. K. (2018). Social and environmental impacts of landslides. *Innov. Infrastruct. Solut.*, *3*(1), 70. https://doi.org/10.1007/s41062-018-0175-y.

Vung, D. V., Tran, T. V., Duc Ha, N., & Huy Duong, N. (2023). Advancements, challenges, and future directions in rainfall-induced landslide prediction: A comprehensive review. *J. Eng. Technol. Sci.*, *55*(4), 466–478. https://doi.org/10.5614/j.eng.technol.sci.2023.55.4.9.

Ye, C., Wu, H., Oguchi, T., Tang, Y., Pei, X., & Wu, Y. (2025). Physically based and data-driven models for landslide susceptibility assessment: Principles, applications, and challenges. *Remote Sens.*, *17*(13), 2280. https://doi.org/10.3390/rs17132280.

Zahri, R., Md Ali, A. H., Rambat, S., Ghazali, N. H., Ahmad, Y., & Hamzah, M. H. (2025). Flood Early Warning Systems (FEWS) in enhancing disaster risk reduction and community resilience: A systematic review. *Int. J. Law Gov. Commun.*, *10*(40), 425–445. https://doi.org/10.35631/IJLGC.1040031.

Zhang, Z., Zeng, R., Meng, X., Zhao, S., Wang, S., Ma, J., & Wang, H. (2023). Effects of changes in soil properties caused by progressive infiltration of rainwater on rainfall-induced landslides. *Catena*, *233*, 107475. https://doi.org/10.1016/j.catena.2023.107475.

Zulkafli, S. A. & Abd Majid, N. (2024). Urban resilience in the face of natural hazards: Leveraging machine learning to assess landslide risk in Kuala Lumpur, Malaysia. *Int. J. Acad. Res. Bus. Soc. Sci.*, *14*(4), 252–267. http://doi.org/10.6007/IJARBSS/v14-i4/20831.