Journal of Computer Science

Research Paper

Random Forest and Extreme Gradient Boosting with Bayesian Hyperparameter Optimization for Landslide Susceptibility Mapping in Penang Island, Malaysia

Dorothy Anak Martin Atok[1], Soo See Chai[1], Kok Luong Goh[2], Neha Gautam[3], Kim On Chin[3]

[1]*Faculty of Computer Science and Information Technology, University of Malaysia Sarawak, Malaysia*
[2]*School of Science and Technology, International University College of Advanced Technology Sarawak, Malaysia*
[3]*Faculty of Computing and Informatics, Universiti Malaysia Sabah (UMS), Sabah, Malaysia*

Article history
Received: 03-09-2024
Revised: 6-02-2025
Accepted: 24-03-2025

Corresponding Author:
Soo See Chai
Faculty of Computer Science and Information Technology,
University of Malaysia Sarawak,
Kota Samarahan, Sarawak,
Malaysia
Email: sschai@unimas.my

Abstract: Landslide susceptibility models often face challenges of overfitting and overestimation. This research focuses on improving the predictive capabilities of the Extreme Gradient Boosting (XGBoost) and Random Forest (RF) algorithms by applying Bayesian Hyperparameter Optimization (BayesOpt). Penang Island, a region in Malaysia prone to frequent landslides, was chosen as the study area. Ten Landslide Conditioning Factors (LCFs), including elevation, slope angle, NDVI, and proximity to streams and roads, were derived using Geographic Information Systems (GIS). From the total of 886 landslide and non-landslide data points, a 70:30 split was employed for training and testing, respectively. BayesOpt-RF emerged as the top-performing model among all those assessed with an AUC of 99.50% (Success Rate) and 95.80% (Prediction Rate). RF (SR: 100.00%, PR: 95.60%), XGBoost (SR: 100.00%, PR: 95.20%), and BayesOpt-XGBoost (SR: 96.70%, PR: 93.00%) followed. While BayesOpt did not consistently improve prediction performance, it effectively minimized overfitting and ensured optimal model operation. For effective site selection, the generated landslide susceptibility maps are significant, infrastructure planning, and disaster mitigation.

Keywords: Bayesian Hyperparameter Optimization, Extreme Gradient Boosting Landslide Susceptibility Mapping, Random forest.

Introduction

Penang Island, located on the northwest coast of Peninsular Malaysia, is characterized by a unique topography where densely populated urban areas coexist with hilly terrains and steep slopes. This geographical setting has made the region highly susceptible to landslides, exacerbated by heavy and prolonged rainfall. Significantly, the Tanjung Bungah landslide on October 21, 2017, claimed 11 lives, underscoring the susceptibility of urban regions to slope failures (The Sun Daily, 2018). Similarly, on October 19, 2018, the Bukit Kukus landslide led to the deaths of nine workers, underscoring the critical need for effective landslide risk mitigation strategies (Malay Mail, 2018).

In response to these incidents, landslide susceptibility mapping (LSM) has emerged as an essential tool for identifying high-risk zones and implementing targeted mitigation measures. LSM facilitates the development of strategies such as slope stabilization, enforcement of land-use zoning regulations, establishment of early warning systems, and initiation of community awareness programs. The performance of LSM is contingent upon the detailed assessment of key landslide conditioning factors (LCFs), which encompass terrain features, vegetation cover, and hydrological influences such as elevation, slope, curvature, and proximity to infrastructure and water bodies.

Advancements in geospatial technologies and machine learning algorithms have significantly enhanced the precision of LSM. Techniques such as Random Forest, Extreme Gradient Boosting, and Support Vector Machines have been successfully applied in various studies, demonstrating superior performance in modelling landslide susceptibility. Geographic Information Systems (GIS) serve as a fundamental tool for the integration, management, and spatial analysis of geospatial datasets, thereby enhancing the precision and robustness of landslide susceptibility mapping outputs.

Even with recent improvements, achieving a balance between predictive precision and generalization in LSMs remains difficult. Model outcomes are still heavily affected by data quality, overfitting risks, and the selection of relevant LCFs. This study develops a landslide susceptibility map (LSM) for Penang Island by leveraging a well-structured set of landslide conditioning factors (LCFs) and implementing machine learning models to generate reliable and interpretable predictions. The findings aim to support the design of effective mitigation strategies and enhance comprehension of landslide risk in tropical urbanized regions.

Previous Studies

Recently, numerous machine learning (ML) approaches have been applied to predict landslide susceptibility in targeted regions. The evolution of these models has enhanced their capability to accurately capture the intricate relationships between landslide events and their triggering factors (Wang et al., 2020). Accessible satellite photos, remote sensing data, landslide history, and GIS all contribute to applying ML effectively in landslide susceptibility mapping.

Nevertheless, comparative studies conducted by Youssef & Pourghasemi (2021) and a comprehensive literature review on ML models in LSM by (Merghadi et al., 2020) show that each machine-learning model has its strengths and weaknesses. One of the most common conventional ML used in LSM studies is the Logistic Regression (LR). LR is good at addressing the non-linear relationship between the LCFs and the occurrence of landslides. Nevertheless, Zhou et al. (2018) and Wubalem and Meten (2020) stated that LR is unsuitable for addressing complex non-linear problems and has limitations of generalization and simplification of landslide factors, affecting its prediction performance.

Next is the Support Vector Machine (SVM) algorithm upholds the structural risk minimization principles, which rely on statistical learning (Vapnik, 1999). SVM holds promise in addressing prediction challenges associated with a smaller number of samples, nonlinearity, and high-dimensional variables (Zhou et al., 2018). However, SVM exhibits limitations in suitability for LSM studies dealing with huge datasets, making it less suitable for regional LSM (Zhao et al., 2021).

Then, the K-Nearest Neighbor (KNN) is also commonly used in LSM studies. KNN has shown good prediction performance but as the number of samples increases, computation costs increase (Avand et al., 2019; Hussain et al., 2022).

Moreover, advanced machine learning techniques, including deep learning and ensemble models, have been increasingly applied in landslide susceptibility mapping (LSM) research. Studies by Khosravi et al. (2020) and Yi et al. (2020) demonstrated that deep learning architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) outperform conventional ML approaches in predicting landslide risk with higher accuracy. However, they are only ideal for big datasets (Khosravi et al., 2020; Yi et al., 2020). The CNN model has a high capability in capturing images, audio, and text by modelling spatial-temporal locality (Zhang et al., 2019).

Finally, ensemble learning refers to the combination of multiple models or learners to improve prediction accuracy, robustness, and generalization. Techniques such as RF and XGBoost are prominent examples of ensemble learning, as they combine the outputs of multiple decision trees. Ensemble methods are widely recognized for their high robustness and ability to handle complex datasets effectively (Nachappa et al., 2020).

Comparative studies provide evidence of the advantages of ensemble methods. For instance, Hussain et al. (2022) and Sahin (2020) demonstrated that RF and XGBoost consistently achieved superior prediction accuracy compared to other ML models such as SVM, LR, CNN, and KNN. Similarly, Youssef and Pourghasemi (2021) highlighted the strong generalization capability of these models in various geospatial prediction tasks, while Zhang et al. (2019) reported their effectiveness in capturing complex relationships in high-dimensional data.

However, despite these strengths, ensemble models like RF and XGBoost are not without limitations. A key concern is the risk of overfitting, particularly when the models are optimized to achieve exceptionally high success rates. Overfitting arises when a model becomes excessively attuned to the training data, thereby compromising its generalizability to unfamiliar datasets. This issue has been noted in several studies, where RF and XGBoost exhibited high accuracy during training but experienced marginal declines in predictive performance on independent test data (Hussain et al., 2022). Overfitting can undermine the reliability of landslide susceptibility mapping (LSM) applications, especially when applied to real-world scenarios where data variability is high.

Therefore, this research explores Bayesian optimization as a strategy to mitigate overfitting, aiming to enhance the model's generalization capability while preserving strong predictive performance. By tackling the limitations of ensemble models, the research provides valuable insights that strengthen the integration of machine learning approaches in geospatial risk assessment.

Problem Statement

Even though several ML models have demonstrated their efficiency in landslide susceptibility prediction, there are still notable obstacles where the MLs' predictive performance differs between different study areas. This variation can be attributed to differences in landslide conditioning factors (LCFs) and the frequency or distribution of landslide occurrences across various regions.

Furthermore, as demonstrated by Park & Kim (2019) and Wang et al. (2021), model complexity can cause overfitting even in cases where ML models yield excellent accuracy. Therefore, the requirement for efficient hyperparameter optimisation approaches considerably complicates the performance of ML models. Thus, an efficient hyperparameter optimization method is important to ensure the optimal performance of ML models.

Research Objectives

To overcome these challenges, a novel methodology is proposed by implementing and evaluating ensemble machine learning algorithms, Random Forest (RF) and Extreme Gradient Boosting (XGBoost) for the development of a landslide susceptibility map (LSM) in Penang Island. These models are selected for their demonstrated robustness and accuracy in previous studies. Furthermore, this research implements Bayesian hyperparameter optimization (BayesOpt) to fine-tune the models, ensuring their optimal performance. By focusing on these advanced techniques, this study aims to overcome the limitations of existing models and contribute to more reliable and accurate LSM in diverse geographical contexts.

Study Area

Penang Island is located northwest of Peninsular Malaysia in the Strait of Malacca. The island is around 301.248 km^2 (Figure 1). The study region is between 5°15'N and 5°30'N and 100°10'E and 100°20'E. The island has steep hills covering half the region at 18 m. The Timeline News (*Channel News Asia,* 2022) reported many landslide cases in Penang Island, including on October 19, 2022, when severe weathe r triggered tiny landslides and fallen branches on Penang Hill. Moreover, on June 21, 2021, severe rainfall runoff triggered a flash flood of muddy water in front of the Surin Condominium in Tanjung Bungah, Penang. A steep hill was cut when Solok Tanjung Bungah was extended along a 250-foot contour. On November 21, 2018, a minor landslide occurred on Jalan Paya Terubong between Green Garden and Jalan Tun Sardon.

Methodology

The development of a landslide inventory map for landslides and LCFs' spatial databases, LCFs' importance evaluation, landslide prediction modelling, models' performance evaluation and final susceptibility map development were the important phases in this research study (Figure 2).

Fig. 1: Penang Island.

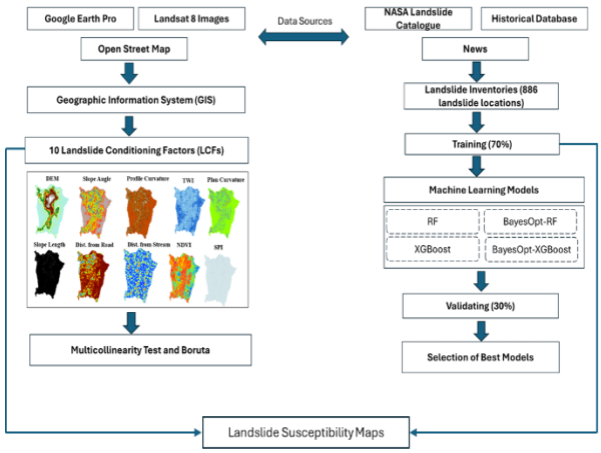


Fig. 2: Research layout.

Landslide Inventories

Landslide inventories record geological hazard locations, times, and types in an area (Pašek, 1975). The landslide data for both study areas was collected from the open-source NASA Landslide Catalogue website (https://gpm.nasa.gov/landslides/index.html, accessed on March 30, 2021) and historical geohazard literature (Elmahdy et al., 2016; Han et al., 2021; Lee & Pradhan, 2006; Nhu et al., 2020). The landslide inventory map included recent news reports from The Stars and The Channel NewsAsia. Notably, Penang Island has 443 landslides.

Grid Unit Extraction

The study extracted binary landslide point features in QGIS 3.18 using the LaGriSu tool pack (Al-Thuwaynee et al., 2023) and a Digital Elevation Model (DEM) layer. Non-landslide samples were randomly drawn from stable areas to avoid introducing class imbalance (Chen et al., 2017; Park & Kim, 2019; Sahin, 2020). Sameen et al. (2020) recommended that non-landslide data points be selected at least 500 meters away from known landslide locations to ensure spatial independence. Next, there must be a minimum of 100 meters separating any two non-landslide samples.

The LaGRiSU tool pack created by Al-Thuwaynee et al. (2021) has been applied in previous LSM studies such as Al-Thuwaynee et al. (2023), Guo et al. (2022) and Nnanwuba et al. (2022). The LaGRiSU helped in the extraction of binary (1 for landslides and 0 for landslide-free locations). It consists of a grid unit, which divides the territory into regular squares of a specified size. The grid unit exhibits superior automation and a simple structure (Nam & Wang, 2020; Zhang et al., 2022). A total of 886 data points, comprising both landslide and non-landslide locations, were identified following the extraction process (Figure 3). Subsequently, landslide susceptibility values corresponding to ten conditioning factors were assigned to these datasets.

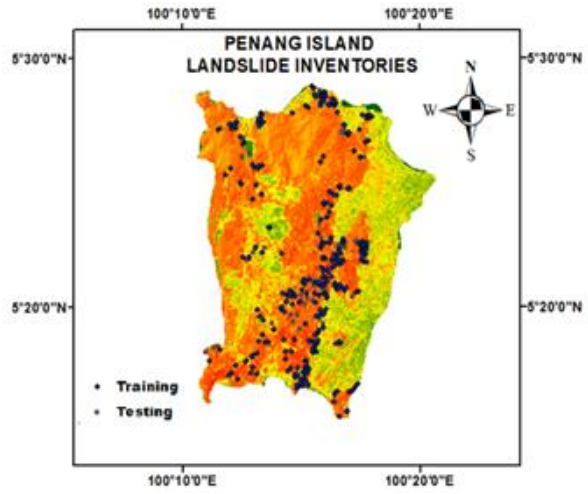


Fig. 3: Penang Island landslide inventories.

Landslide Conditioning Factors

One of the key prerequisites for achieving reliable landslide susceptibility predictions is the careful compilation and refinement of the LCF dataset. The LCFs included in this study were chosen based on common LCFs used in earlier research in similar tropical environments as Penang Island (Bag et al., 2022) and also LSM studies conducted in Penang Island (Han et al., 2021).

In this study intrinsic influencing variables known as the topographic factors including elevation or Digital Elevation Map (DEM), slope angle, slope length, plane curvature, Stream Power Index (SPI), profile curvature, Topographic Wetness Index (TWI), and environmental factors which includes Normalized Difference Vegetation Index, distance from roads, distance from rivers were being used as shown in Figure 4 (Zhou et al., 2021). This study also utilised GIS software known as ArcMap 10.8 to develop the LCFs' spatial databases, and all the maps were projected to an output cell of 30 m x 30 m. The coordinate system of the layer was set to the UTM zone 47N.

The selection of the 10 landslide conditioning factors (LCFs), namely the slope angle, slope length, distance from road, distance from stream, NDVI, TWI, SPI, plan curvature, and profile curvature was guided by a two-step process: literature review and feature selection analysis.

At the initial stage, the selection was based on an extensive review of studies on landslide susceptibility mapping. These factors were consistently identified as influential in previous research within tropical regions, particularly areas with similar geophysical and environmental characteristics to Penang Island. Besides, at the initial stage, the selected data also depending on its availability.

After confirming the chosen LCFs in the initial phase, then to refine the selection and confirm the relevance of these factors, the Boruta algorithm was applied. This algorithm is a robust wrapper method designed to identify all relevant variables by comparing their importance against randomized features (shadows). The results validated the significance of the 10 factors, ensuring that only influential variables were retained for the modelling process.

Digital Elevation Model (DEM)

Digital Elevation Model (DEM) development frequently uses elevation points. A DEM, a digital cartographic dataset, represents a continuous topographic elevation surface through several cells. It makes it possible to identify watersheds, establish stream networks, and characterize a location's physical functions (Sidaway, 2022). The majority of landslide susceptibility studies have commonly used it.

In the study by Devkota et al. (2013), it was found that higher elevations are thought to be more susceptible to landslides than lower elevations. In comparison to lower elevations or foothills, the rocks are stronger at higher elevations (Chen et al., 2018) because dehydrogenases are activated by organic materials, specifically carbon and nitrogen in the soil, at high

elevations (Błońska et al., 2018). In this study elevation point of Penang Island was extracted from Google Earth Pro and utilized to develop a DEM map via ArcMap 10.8 software. The DEM map shows that Penang Island elevations range from 0 m to 797.41 m (Figure 4a).

Slope Angle

Landslides occur more frequently on slopes involving steeper slope angles. Landforms with a larger slope angle are more prone to collapse. The slope gradient is crucial for groundwater flow and affects the soil moisture, both of which are directly related to the occurrence of landslides (Meena et al., 2022), which is closely correlated to landslides. The slope angle of the current study was derived from DEM with a resolution of 30 m, ranging from a range of 0° to 60.44° (Figure 4b).

Slope Length

Slope length influences the hydrological processes and soil loss in hilly areas through its relationship with slope angle (Pourghasemi & Rahmati, 2018). The LS factor in the present research spans a range from 0 to 124.02 (Figure 4c). Derived from the DEM, the LS factor was calculated in ArcMap according to Equation 1 (Wischmeier & Smith, 1978):

$$LS = \left(\frac{\lambda}{22.13}\right)^{m} \times (65.41\ sin^2\theta + 4.56\ sin\theta + 0.0065) \quad (1)$$

By referring to Equation 1, λ is the slope length in meters (m), θ is the angle of slope in percent (%), and m is the constant dependent on the value of the slope gradient as discussed above.

Plan and Profile Curvature

Plan curvature affected the surface runoff dispersion and convergence (Aghdam et al., 2016). On the other hand, by controlling the materials' acceleration or deceleration on a slope, the profile curvature influences the deposition of material (Xiao et al., 2019). In this work, ArcMap 10.8 was used to extract plan and profile curvatures from the DEM. A range of -5.08 to 4.86 was recorded for both plan and profile curvatures (Figure 4d) and -7.59 to 8.29 (Figure 4e), respectively.

Distance from Road

One significant anthropogenic element influencing the structure of natural slopes is road development, which involves excavating and cutting slopes (Xiao et al., 2019). Proximity to roads is considered a potential landslide-inducing factor. In this study, the distance from roads, ranging from 0 to 1561.15 meters (Figure 4f), was calculated using the Euclidean Distance tool in ArcMap 10.8.

Distance from Stream

Landslide tragedies are often associated with surface water. The river constantly erodes and hollows out the slope toe, causing slope instability (Li et al., 2021). A range of -5.08 to 4.86 was recorded for both plan and profile curvatures. (Figure 4g).

Normalized Difference Vegetation Index

It plays an essential role in water retention, which subsequently enhances the cohesion and shear strength of lithological substrates (Sidle & Ochiai, 2013). This study applied the Normalized Difference Vegetation Index (NDVI) to assess vegetation density, where elevated NDVI values reflect areas with thicker vegetation cover. The index typically spans from -1 to 1, with greater values indicating increased vegetation presence.

The Landsat 8 image of Bands 4 and 5 from the USGS Earth Explorer, accessed on March 3, 2021, was used to create the NDVI layer for Penang Island (Figure 4h). For a better capture, it was ensured that the cloud cover was less than 20%. The NDVI map was then calculated using Equation 2:

$$NDVI = (BAND\ 5 - BAND\ 4)/(BAND\ 5 + BAND\ 4) \quad (2)$$

Topographic Wetness Index and Stream Power Index

The Stream Power Index (SPI) serves to characterize the potential erosion caused by flowing water at a specific location on the topographic surface. As both the catchment area and slope gradient expand, there is a corresponding increase in the volume of water from upstream areas and the speed of water flow, leading to heightened SPI values and an elevated risk of erosion (Florinsky, 2017). SPI considers data on catchment area and slope gradient as in Equation 3:

$$SPI = ln(CA.\tan G) \quad (3)$$

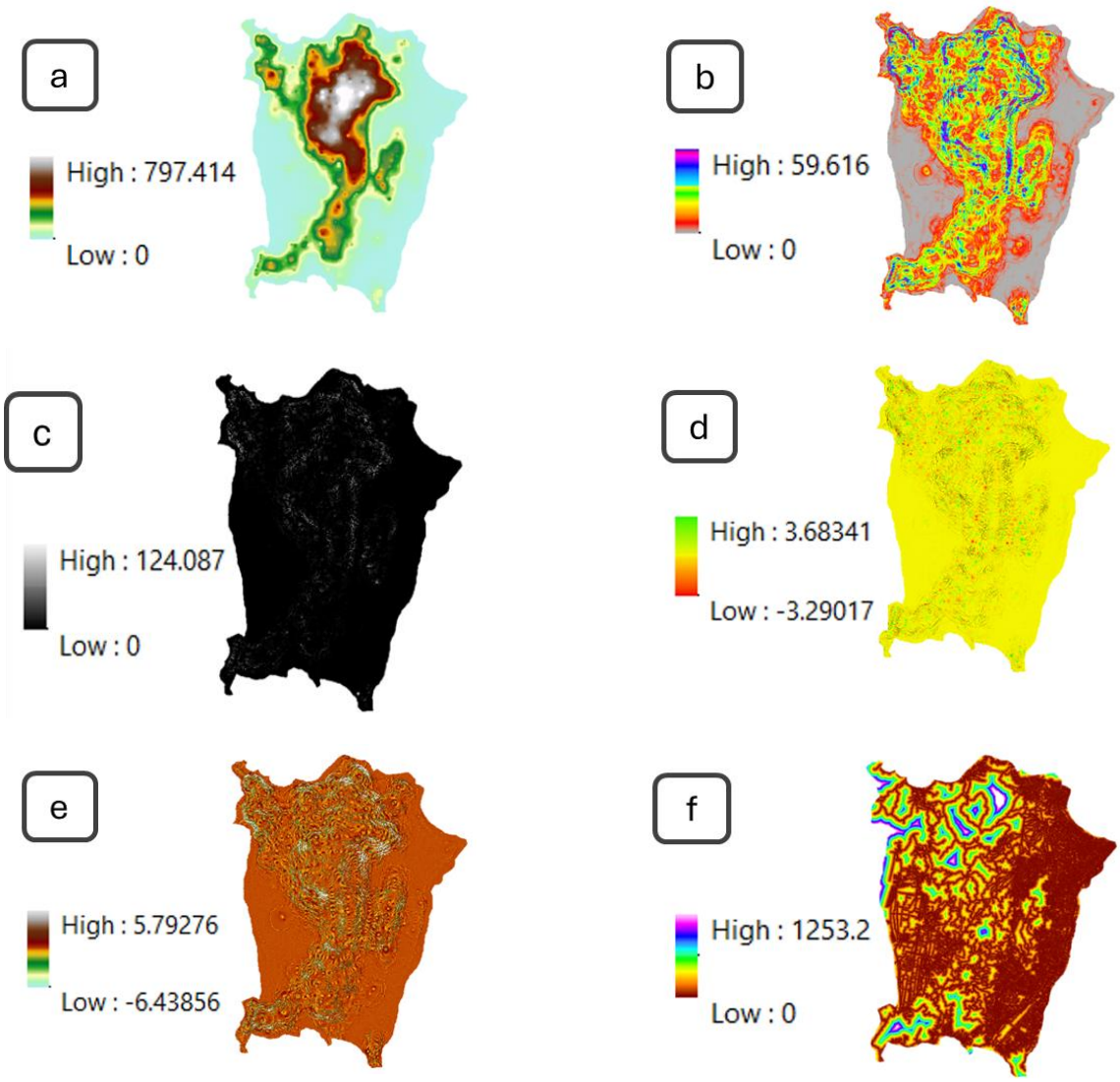
Based on Equation 3, CA and G are catchment area and slope gradient, respectively. Topographic Wetness Index (TWI) TWI describes the soil moisture spatial pattern (Yilmaz, 2010). Like other combined morphometric variables, SPI and TWI can be derived from a digital elevation model (DEM). Penang Island's SPI is in the range of 0 to 117,894 (Figure 4i) and the TWI range is 2.97 to 25.95 (Figure 4j).

Multicollinearity Test

Multicollinearity arises in regression analysis when independent variables exhibit high intercorrelation, which can distort parameter estimates and inflate standard errors. The Variance Inflation Factor (VIF) assesses the degree of variance inflation caused by multicollinearity, while Tolerance (TOL) represents the extent to which a predictor's variance is independent of other predictors (Rabby & Li, 2020; Saha & Saha, 2022). High VIF is typically greater than 10, indicating a strong correlation between predictor variables; TOL values near 1 indicate low multicollinearity, while values near 0 indicate high multicollinearity.

Landslide factors importance evaluation with the Boruta algorithm

RF classification uses the Boruta algorithm, an attribute selection technique that employs shadow characteristics as duplicates of original features (Amiri et al., 2019). These shadow characteristics are utilised to create decision trees, where the accuracy loss resulting from randomly assigned attributes is used to determine the relative value of each attribute. The Z-score is a useful tool for evaluating the impact of input variable accuracy on the quality of the model prediction and for determining significant LCFs (Ahmadpour et al., 2021).



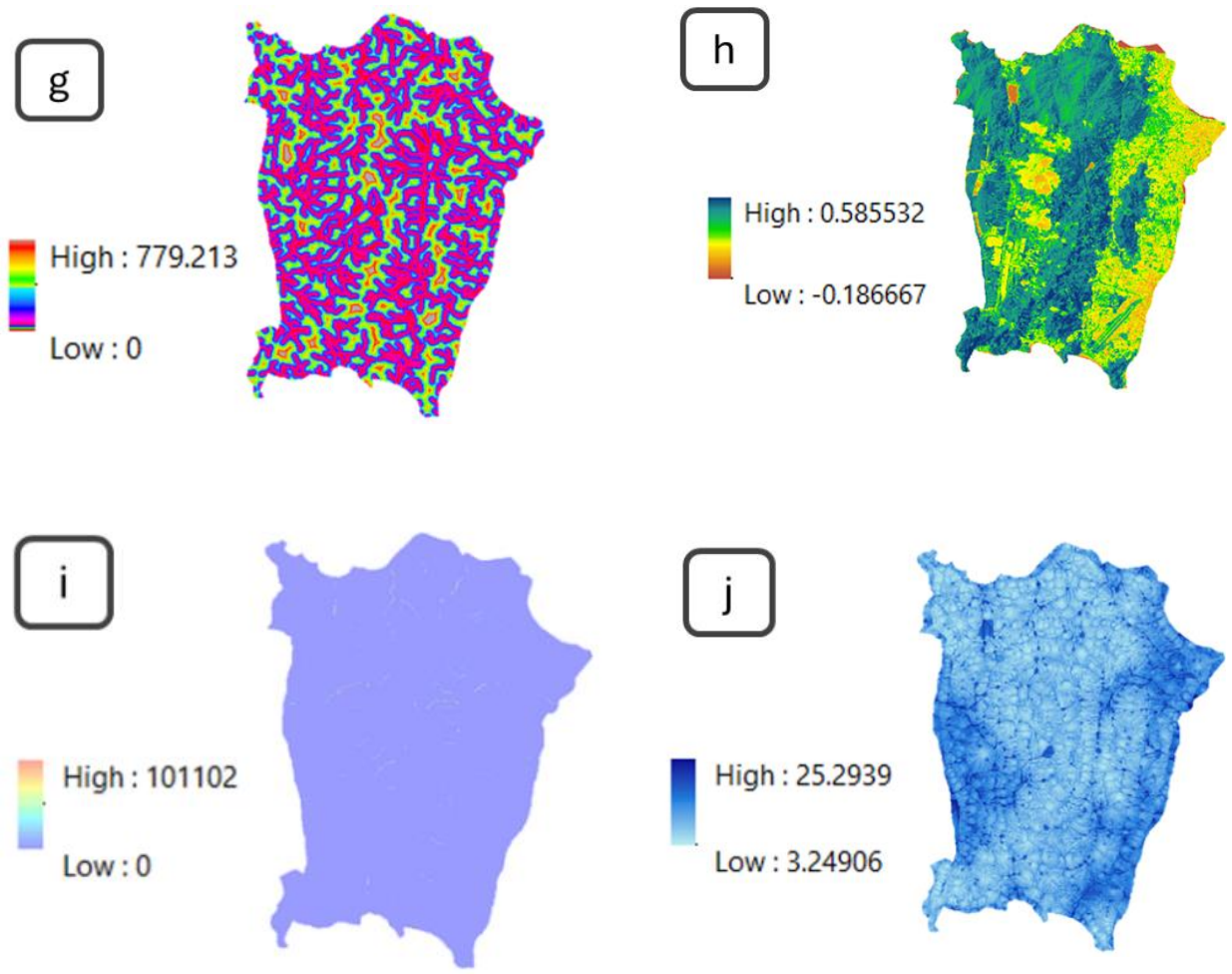


Fig. 4: LCFs (a) DEM (b) Slope angle (c) Slope Length (d) Plan Curvature (e) Profile Curvature (f) Distance from Road (g) Distance From Stream (h) NDVI (i) SPI (j)TWI.

Machine Learning Modelling

In the field of landslide susceptibility assessment, machine learning (ML) has emerged as an essential part of problem-solving for spatial modelling (Sevgen et al., 2019). Several advantages result from machine learning techniques: (a) it can automatically extract information from big databases; (b) it can adapt its structure in response to the available landslide data that and (c) it can develop classification and regression to produce an accurate landslide model. These models can be extended to large-area analysis and are more rapid and affordable than traditional approaches (Kavzoglu et al., 2019). The effectiveness of two advanced machine learning algorithms, which differ in terms of complexity, was assessed in the current study's landslide susceptibility mapping. They include RF and XGBoost. The modelling process was conducted using RStudio Software version 4.3.0.

Data Partitioning

The dataset consisting of both non-landslide and landslide points were split into training (70%) and testing (30%) datasets at random. The decision to partition the dataset into 70% training and 30% testing was based on standard practices observed in the landslide susceptibility mapping previous studies such as Akinci & Zeybek (2021), Chen et al. (2020) and Hussain et al (2022), as well as the general aim of achieving a balance between model training and validation. Several studies, including Achour & Reza (2020), Hussain et al. (2021), Saha et al.

(2021), and Sahin (2022), stated that there is no explicit criterion or a standard ratio for dataset splitting.

Therefore, in this study, the dataset was sufficiently large to allow for a 70:30 split, providing enough samples for training without compromising the integrity of the testing phase. This ratio was chosen to maximize the learning capability of the model, while still ensuring the model's ability to generalize well on the testing data. Unwanted attributes in the data were eliminated, and the data was scaled as part of the preprocessing stage. Following data preprocessing, the RF and XGBoost models were built, incorporating Bayesian hyperparameter optimization.

Random Forest

Based on decision trees, the RF algorithm aggregates predictions from multiple trees, significantly enhancing both decision tree and bagging models (Breiman, 2001; Youssef et al., 2016). The "mtry", "ntree", and "min_n" were the essential hyperparameters of the RF. The "min_n" is the minimum number of occurrences in the node to split further, and the "mtry" is the number of sampled predictors at each stage. Additionally, the RF model's "ntree" defines how many trees should be built (Hussain et al., 2021; Rabby et al., 2020).

Table 1: XGBoost hyperparameters listings

Hyperparameters	Definition
"Eta"	This parameter defines the learning rate, acting as a step-size control during model updates to help mitigate overfitting.
"Max_depth"	The maximum depth of each tree in the model.
"Min_child_weight"	This parameter sets the lowest sum of observation weights a child node must have, serving as a mechanism to reduce overfitting.
"Subsample"	It refers to the proportion of data samples randomly selected to train each individual tree within the model.
"Gamma"	Gamma specifies the minimum loss reduction required to make a split.
"Colsample_bytree"	The subsample ratio of columns when constructing each tree.
"Nrounds"	The number of trees.

Extreme Gradient Boosting

XGBoost is an advanced implementation of the gradient tree boosting algorithm, recognized for its computational efficiency in machine learning tasks. In this approach, each successive tree is trained on the residual errors of the preceding trees. Unlike RF, which relies on majority voting for its final output, XGBoost generates predictions by aggregating the outputs of all individual trees through summation. (Friedman, 2001, 2002). This approach offers superiority over standard gradient boosting by incorporating a regularization parameter, preventing the model from overfitting (Cao et al., 2020). The performance of the XGBoost depended on the tuning of the hyperparameters. The common hyperparameters of XGBoost are tabulated in Table 1.

Bayesian Hyperparameter Optimization

Bayesian optimization serves as an efficient technique for hyperparameter tuning in machine learning, as it leverages probabilistic models, most commonly a Gaussian Process (GP) to construct a surrogate model of the objective function, thereby enabling robust and automated parameter selection (Rana et al., 2017). As the iterative process unfolds, the algorithm carefully manages the trade-off between exploration and exploitation, considering its evolving understanding of the target function. In this study, GP was tailored to the observed data points and coupled with an exploration strategy such as the Upper Confidence Bound (UCB) method (Srinivas et al., 2012).

Performance Evaluation

Susceptibility class distributions over the research region were generated by building machine learning models using the available landslide training datasets. As suggested by Youssef and Pourghasemi (2021), the natural breaks (Jenks) classification strategy was used to classify these resulting landslip susceptibility maps (LSMs) into four risk levels: none, low, high, and extremely high. Both statistical and visual evaluations, such as receiver operating characteristic (ROC) curves, success and prediction rate graphs, and confusion matrices together with the corresponding performance indicators, were used to validate the model.

Confusion Matrix and Statistics

Based on the statistical results shown in Table 2, "False Positive" (FP) and "False Negative" (FN) represent the number of landslide instances that were misclassified by the model, while "True Positive" (TP)

and "True Negative" (TN) indicate correctly identified cases. Metrics such as accuracy, precision, sensitivity, and specificity expressed as values between 0 and 1 or as percentages reflect model performance, with higher values indicating better predictive capability.

Table 2: Statistics used for model performance evaluation.

Statistics	Equation
Accuracy	(TP+TN)/(TP+TN+FP+FN)
Precision	TP/(TP+FP)
Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)

ROC

The Area Under the Curve (AUC) graph has index values from 0.5 (50%) to 1 (100%). A minimum AUC threshold of 60% is recommended. Note that AUC values of 90% imply excellent model classification, 80%-90% good, 70%-80% moderate or satisfactory, and less than 60% poor (Pradhan, 2013). Effectiveness of the developed model in landslide susceptibility assessment is reflected through the Success Rate (SR) curve (Pham et al., 2017). It is generated by comparing the number of landslides included in the training dataset with the LSMs (Pradhan et al., 2010). Prediction Rate (PR) curves, on the other hand, gauge the likelihood of landslide occurrences, serving as a means to validate the predictive capability of a model (Brenning, 2005). These curves are constructed using the testing dataset (Fabbri et al., 2002).

Additionally, factors such as distance from roads, TWI, SPI, distance from streams, slope length, plan curvature, and profile curvature were identified as significant, consistent with numerous prior studies (Pham et al., 2022; Sun et al., 2020). These factors have consistently been recognized as dominant conditioning factors in various research endeavours.

Results

Multicollinearity test

Table 3: Multicollinearity Assessment of Landslide Conditioning Variables Using Tolerance and Variance Inflation Factor Metrics.

LCFs	Tolerance	Variance Inflation Factor
Slope	0.56	1.78
Plan Curvature	0.58	1.72
TWI	0.64	1.57
Profile Curvature	0.66	1.51
NDVI	0.72	1.38
DEM	0.74	1.36
Slope length	0.78	1.29
Distance from Road	0.88	1.14
Distance from Stream	0.90	1.12
SPI	0.90	1.11

Table 3 summarizes the multicollinearity analysis of the whole Penang Island dataset. The Penang Island landslide factor, slope has the greatest VIF of 1.78 and the lowest TOL of 0.56. The SPI had the highest TOL of 0.90 and the lowest VIF of 1.11. All LCFs had TOLs below 0.1 or VIFs above 10 proving that all the ten landslide factors did not demonstrate significant collinearity.

LCFs Importance Analysis Via Boruta

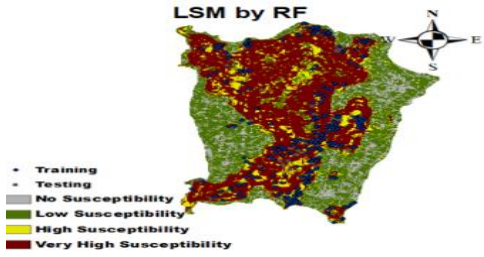
The Boruta model was used to evaluate the impacts of the ten contributing factors on disaster occurrences within the study area. As shown in Table 4, all ten LCFs were found to be influential in the occurrence of landslides across Penang Island. Consistent with previous LSM research conducted in Iran's Abha Basin and Bangladesh's Muzaffarabad district, slope and digital elevation model (DEM) emerged as the primary contributors to landslide susceptibility (Hussain et al., 2022; Youssef & Pourghasemi, 2021).

Table 4: Landslide conditioning factors selection via Boruta.

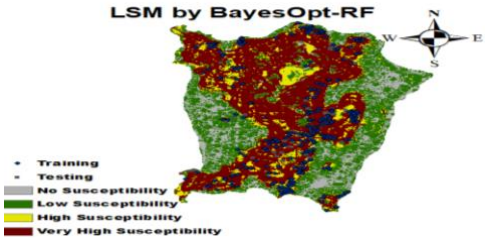
Landslide Conditioning Factors	Importance	Norm Hits	Decision
Slope	50.72	1.00	Confirmed
DEM	40.08	1.00	Confirmed
TWI	28.55	1.00	Confirmed
SPI	19.49	1.00	Confirmed
Plan Curvature	15.58	1.00	Confirmed
NDVI	15.60	1.00	Confirmed
Slope Length	13.73	1.00	Confirmed
Profile Curvature	13.36	1.00	Confirmed
Distance from Road	12.17	1.00	Confirmed
Distance from Stream	9.19	1.00	Confirmed

Final LSMs

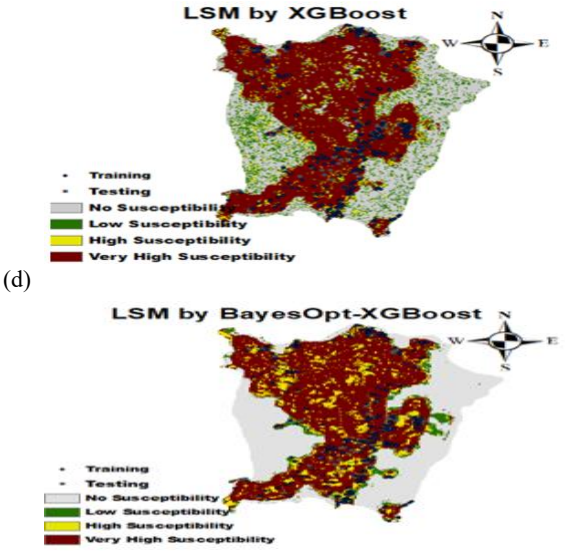
(a)



(b)



(c)



(d)

Fig. 5: Landslide susceptibility maps by (a) RF (b) BayesOpt-RF (c) XGBoost (d) BayesOpt-XGBoost.

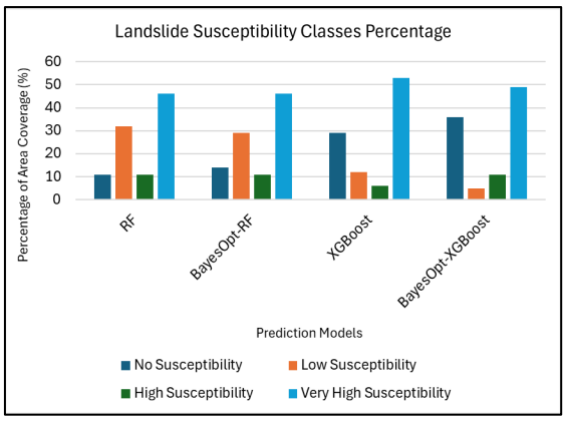


Fig. 6: Landslide Susceptibility Classes Percentage.

The reclassification process grouped the landslide susceptibility maps (LSMs) into four distinct classes namely "No Susceptibility", "Low Susceptibility", "High Susceptibility", and "Very High Susceptibility", based on the Jenks natural break technique. (Jaafari et al., 2014; Chen et al., 2013). This classification approach offers a logical and statistically robust framework for delineating landslide-prone zones, aiding in the interpretation of spatial risk distribution. The results from the four models were evaluated and compared, as shown in Figures 5 and 6, to assess their practical utility.

Both the Random Forest (RF) and Bayesian Optimized Random Forest (BayesOpt-RF) models consistently identified areas of "High Susceptibility" and "Very High Susceptibility", which closely correspond to historically landslide-prone regions. These models yield dependable results that are applicable to practical scenarios, including disaster risk management, urban planning, and infrastructure development. Areas categorized as "Very High Susceptibility" may be designated as high-priority zones for immediate mitigation measures, such as slope reinforcement, enhancement of drainage infrastructure, or vegetation control.

Furthermore, the accurate identification of "High Susceptibility" zones could help inform zoning regulations, guiding land-use planners to restrict construction activities or implement stringent building codes in these areas. By focusing on these high-risk zones, resource allocation for landslide prevention and mitigation could be optimized, reducing overall disaster risks and potential economic losses.

On the other hand, the results from the XGBoost and Bayesian Optimized XGBoost (BayesOpt-XGBoost) models indicate a tendency to overestimate areas classified as "Very High Susceptibility." While this overestimation might be beneficial in creating a conservative buffer for landslide risk management, it may reduce the maps' practicality for stakeholders who rely on precise spatial information to allocate resources effectively. For example, overestimated high-risk zones might lead to excessive resource allocation in areas that may not require immediate intervention, thus diverting resources away from genuinely critical zones.

This limitation was similarly reported in the landslide susceptibility study of Rangamati Hill, Bangladesh, where XGBoost overestimated high-susceptibility areas, making it less practical for real-world decision-making despite its strong predictive performance. Overestimation experienced by the current XGBoost and BayesOpt-XGBoost, potentially caused by spatial autocorrelation. Spatial autocorrelation refers to the tendency of geographically proximate units to exhibit similar values more frequently than would be expected in a random distribution, indicating a lack of independence among observations.

Implementing spatially stratified cross-validation is an effective strategy to mitigate the effects of spatial autocorrelation in landslide susceptibility mapping. This approach involves dividing the study area into distinct geographic regions, ensuring that training and testing datasets are spatially independent. By doing so, the model is less likely to overfit to spatially clustered data, leading to more accurate and generalizable predictions. For instance, Li et al. (2024) demonstrated that incorporating spatial adjacency information in slope units significantly improved the predictive accuracy of landslide susceptibility models, highlighting the importance of accounting for spatial dependencies in model validation. Besides, Rabby & Li (2020), also

suggested that by stacking two different ML models can overcome the problem of overestimation experience by the respective ML model such as XGBoost.

Discussion

From a broader practical perspective, the LSMs generated in this study can serve as an essential tool for multiple stakeholders, including government agencies, urban planners, disaster risk managers, and infrastructure developers. In the context of disaster risk management, areas identified as "High Susceptibility" and "Very High Susceptibility" can be prioritized for mitigation efforts, including slope reinforcement. Erosion control and the installation of monitoring equipment to detect early signs of landslides.

Next, it is for urban and infrastructure planning where the susceptibility maps can guide decision-making for future development projects by identifying zones that are unsuitable or require additional safeguards for infrastructure construction, such as roads, bridges, or buildings.

Furthermore, the maps can be integrated into early warning systems to inform evacuation plans and emergency responses in regions prone to landslides, particularly during periods of heavy rainfall or seismic activity.

Besides, policymakers can use the LSMs to enforce zoning regulations that limit or prohibit construction activities in highly susceptible areas, thereby reducing the likelihood of human and economic losses.

Finally, the maps enable better prioritization of resources by identifying areas requiring immediate attention versus those with lower susceptibility levels, ensuring a more cost-effective approach to landslide risk management.

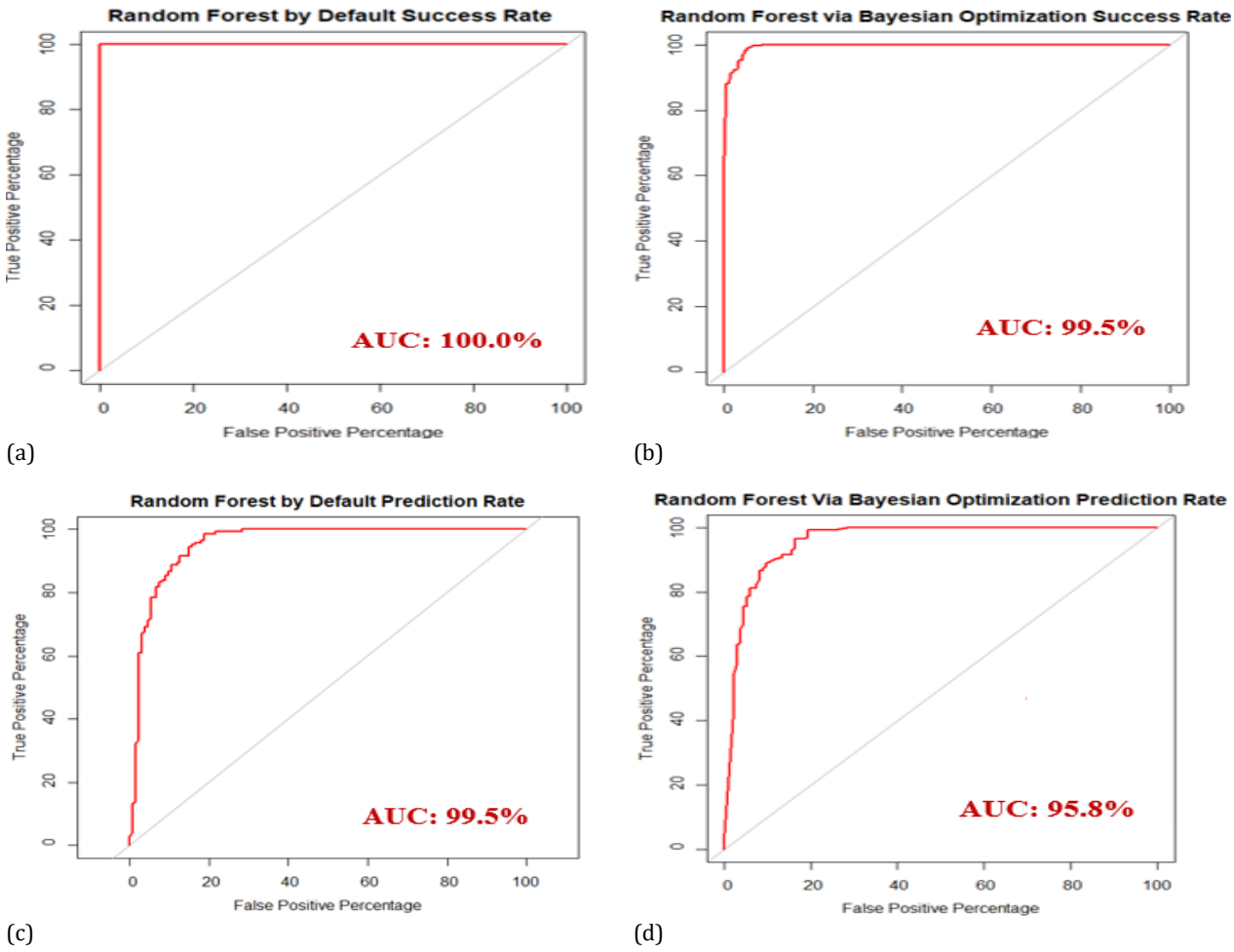


Fig. 7: Random Forest and BayesOpt-RF AUROC Curve of Success and Prediction Rate.

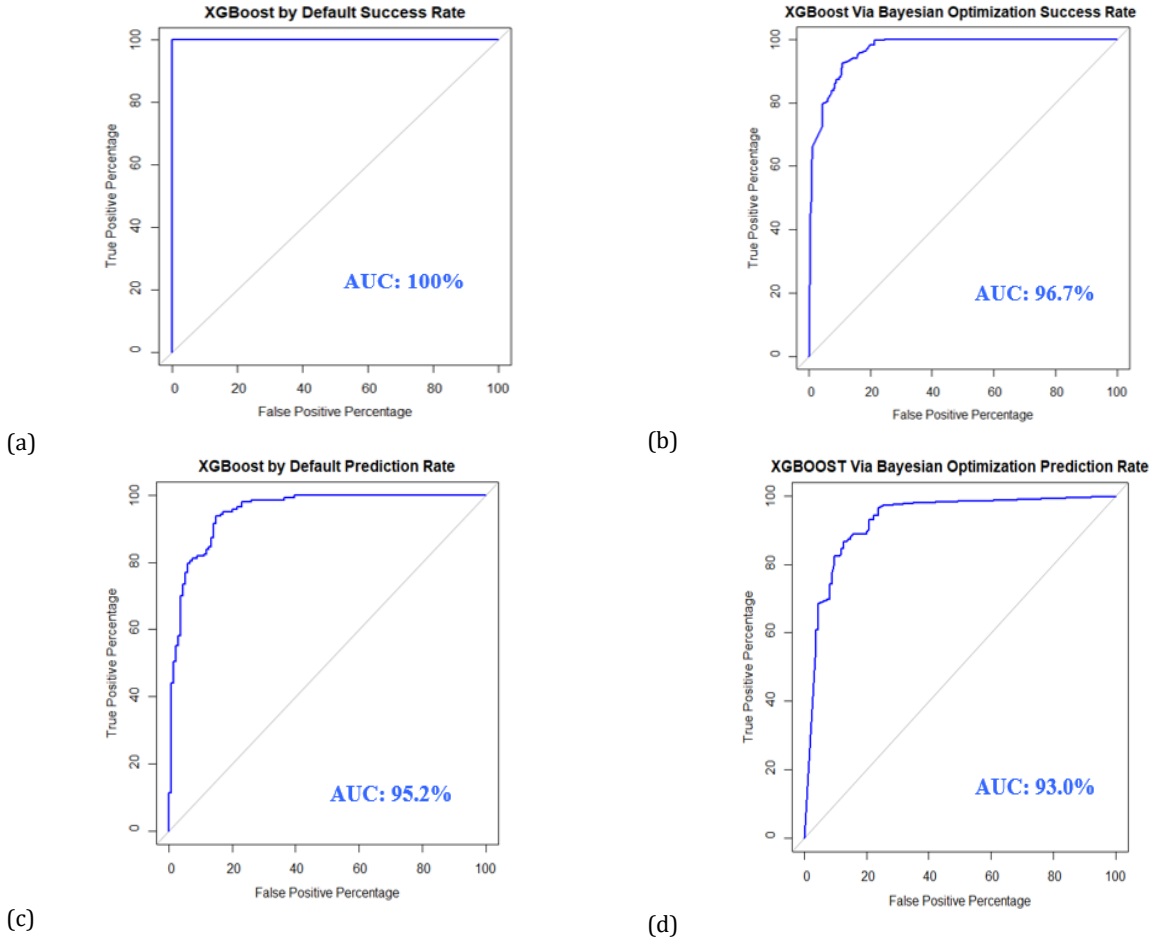


Fig. 8: XGBoost and BayesOpt-XGBoost AUROC curve of Success and Prediction Rate.

Table 5: Models AUROC and Statistics values.

Models	Confusion Matrix				Accuracy	Precision	Sensitivity	Specificity
	TP	TN	FP	FN	(%)	(%)	(%)	(%)
BayesOpt_RF	115	19	12	131	88.81	90.55	85.82	91.61
RF	117	17	15	128	88.45	88.64	87.31	89.51
XGB	115	19	16	127	87.36	87.79	85.82	88.81
BayesOpt_XGB	116	18	19	124	86.64	85.93	86.57	86.71

Table 6: The hyperparameters optimal values of BayesOpt-XGBoost.

Hyperparameters	Optimal Values
"Eta"	0.1940081
"Max_depth"	2
"Min_child_weight"	25.45416
"Subsample"	0.9594105
"Gamma"	0.9255217
"Colsample_bytree"	0.8103804
"Nrounds'	23

Table 7: The hyperparameters optimal values of BayesOpt-RF.

Hyperparameters	Optimal Values
"mtry"	2
"ntree"	499
"min_n"	8

Comparative Analysis

In this study, four models, Random Forest (RF), Bayesian Optimized Random Forest (BayesOpt-RF), XGBoost, and

Bayesian Optimized XGBoost (BayesOpt-XGBoost) were evaluated and compared based on key performance metrics, including accuracy, sensitivity, specificity, precision, and AUC scores. The results were tabulated in Table 5.

Accuracy

RF achieved an accuracy of 88.45%, while BayesOpt-RF slightly improved to 88.81%. The XGBoost model demonstrated 87.36% accuracy, and BayesOpt-XGBoost reached 86.64%.

While the RF model provided the highest accuracy, the BayesOpt-RF model's marginal improvement suggests that optimizing hyperparameters through Bayesian methods can lead to slight performance gains for RF models.

The XGBoost models, both default and optimized, performed slightly worse in accuracy compared to RF, with BayesOpt-XGBoost showing the lowest accuracy of the four models. This suggests that while XGBoost may be more robust in many applications, its performance in this study might not be as strong as RF for the landslide prediction task.

Sensitivity and Specificity

RF and BayesOpt-RF exhibited strong sensitivity and specificity, both falling within the range of 80% to 100%. Specifically, BayesOpt-RF had slightly better sensitivity at 85.82% compared to the default RF at 87.31%, while specificity remained high for both models (91.61% for BayesOpt-RF and 89.51% for RF).

In comparison, XGBoost and BayesOpt-XGBoost showed similar performance in terms of sensitivity and specificity. BayesOpt-XGBoost showed 86.57% sensitivity, slightly better than XGBoost's 85.82%, and both models demonstrated strong specificity (around 88.81% for XGBoost and 86.71% for BayesOpt-XGBoost).

Overall, RF models had slightly better overall classification performance, especially in terms of specificity and sensitivity, although the XGBoost models performed comparably in classifying non-landslide regions.

Precision

RF and BayesOpt-RF models showed similar precision values (88.64% and 90.55%, respectively), indicating that both models were reliable in predicting landslide zones. XGBoost demonstrated a precision of 87.79%, while BayesOpt-XGBoost performed slightly worse with 85.93% precision. Precision for RF models was consistently higher than for XGBoost, which further emphasizes the RF model's ability to correctly identify landslide zones in this study.

AUC Scores

RF showed AUC values of 100% for the success rate (SR) and 95.60% for the prediction rate (PR), demonstrating that the RF model achieved excellent performance with minimal overfitting. Similarly, BayesOpt-RF exhibited high AUC values (99.50% for SR and 95.80% for PR), with only a slight improvement over the default RF model.

For XGBoost, the SR was 100%, and the PR was 95.2%, suggesting that the model maintained strong performance during training and testing, with minimal overfitting.

BayesOpt-XGBoost, while optimized, showed 96.72% AUC for SR and 93.0% AUC for PR, indicating a slight reduction in performance after hyperparameter tuning. The difference in AUC scores for all models remained small (ranging from 3.7% to 4.8%), demonstrating minimal overfitting across all models.

Overall

RF models generally performed better in terms of accuracy, precision, and specificity, suggesting that Random Forest is better suited for landslide susceptibility prediction in this study.

Bayesian optimization showed slight improvements for RF in all performance metrics, although these improvements were less significant for XGBoost, which was already highly optimized.

The XGBoost models, particularly BayesOpt-XGBoost, performed well in terms of sensitivity and specificity but did not show substantial improvements after Bayesian optimization, which might suggest that XGBoost is more robust to hyperparameter variations compared to RF.

Models Comparisons

Model performance was assessed using five evaluation metrics: AUC, accuracy, precision, sensitivity, and specificity, applied to both training and testing datasets (Table 5). The AUC of the training data indicates model fitting strength, while the test data AUC demonstrates predictive capability (Tsangaratos et al., 2017). Performance validation revealed that all four models achieved excellent and consistent results, each recording an AUC score above 90%. Additionally, the models' accuracy, sensitivity, and specificity ranged from 80% to 100%, confirming their effectiveness and reliability in predicting landslide susceptibility.

A comparative evaluation of the four developed models showed that BayesOpt-RF achieved the highest predictive performance with an AUC of 95.8%, followed by RF (AUC = 95.6%), XGB (AUC = 95.2%), and BayesOpt-XGB (AUC = 93.0%). The overall finding

showed that the RF and XGB are robust prediction models for landslide susceptibility and RF slightly outperforms XGB. This result is similar to the LSM study in Muzaffarabad, Pakistan where RF had outperformed XGB with a slight percentage in the AUC score (Hussain et al., 2022).

Nevertheless, the AUC score of RF and XGB obtained in the Muzzafarbad (Hussain et al., 2022) study is lower than both optimized and unoptimized RF and XGB in this study. Thus, the difference in findings in this current study and the one conducted in Muzaffarabad, Pakistan (Hussain et al., 2022) might probably be due to several reasons such as differences in environmental factors, presence of noise from the raw remote sensing data, hyperparameters used for the models and also the different tuning methods where in the Muzaffarabad study, grid search tuning method was used instead of Bayesian optimization (BayesOpt).

Next, emphasizing on the impact of applying BayesOpt in the current study, it was found that only RF experienced improvements in accuracy and AUC scores of 0.36% and 0.24%, respectively. Meanwhile, XGB experienced a drop in the accuracy and AUC by 0.72% and 2.2% respectively. This had shown that BayesOpt doesn't always significantly improve the accuracy or prediction performance of a model as portrayed in the finding of the LSM study of Wuqi County, China (Wang et al., 2021).

Somehow BayesOpt still can maintain a good and optimal prediction performance of the current models as the BayesOpt-RF and BayesOpt-XGB AUC score are still above 90% indicating that the models are having high performance. Although, the performances of the current BayesOpt-RF and BayesOpt-XGB are better than the BayesOpt-RF and BayesOpt-XGB in the LSM study of Wuqi County, China but the performance drop in the BayesOpt-XGB of the current study contradicted the findings in LSM study of Wuqi County, China study where the finding had shown that RF and XGB had been improved by 3-4% when implementing BayesOpt (Wang et al., 2021).

The observed improvement in performance with BayesOpt for the RF model, but not for XGBoost can be explained by differences in the models' sensitivity to hyperparameter tuning. For RF, the key hyperparameters such as "mtry", "min_n", and "ntree" play a significant role in controlling the model's complexity and generalization ability. Bayesian optimization is effective in fine-tuning these parameters, leading to noticeable improvements in performance, as evidenced by the superior results in accuracy, precision, sensitivity, and specificity for BayesOpt-RF.

In contrast, XGBoost, with its additional hyperparameters like "eta", "max_depth", "min_child_weight", "subsample", "gamma", "colsample_bytree", and "nrounds", is known for its robustness to hyperparameter changes, especially when the default settings are already well-optimized. Given XGBoost's built-in regularization and early stopping mechanisms, the model is less sensitive to small hyperparameter adjustments, explaining why BayesOpt-XGB did not show substantial improvement over the baseline XGBoost model.

Additionally, RF's flexibility with feature importance and its ability to handle various feature types may make it more sensitive to hyperparameter changes compared to XGBoost, which is highly structured and performs well when the parameters are already well-tuned. Ultimately, Bayesian optimization had a more pronounced effect on RF due to its dependence on hyperparameter configurations, while XGBoost's default settings already provided strong performance, resulting in less noticeable improvements.

Regardless of the AUC and accuracy performance of the optimized models, BayesOpt had successfully reduced overfitting percentage in both RF and XGB. Thus, it is proven that the implementation of BayesOpt in this study had achieved its objective which is to ensure that the models performed optimally where it has both good prediction performance and minimal overfitting.

Ultimately, this study's incorporation of Bayesian Optimization (BayesOpt) offered a methodical and effective way to adjust the hyperparameters for RF and XGBoost. By decreasing the OOB error and raising accuracy, sensitivity, and specificity, BayesOpt greatly enhanced the RF model's performance, making it more dependable for landslip susceptibility mapping (LSM).

While the improvements in XGBoost were less pronounced due to its more complex hyperparameter space, BayesOpt still enhanced its robustness by reducing overfitting. Importantly, BayesOpt replaced manual trial-and-error tuning with a data-driven approach, saving computational time and ensuring reproducibility. This study not only highlights the added value of BayesOpt in optimizing machine learning models for LSM but also offers insights into its effectiveness based on model characteristics, advancing methodological innovation in LSM research.

Overall, RF and XGBoost performed exceptionally well in landslide susceptibility modeling, demonstrating their strong capabilities in capturing complex patterns within the data. However, there is potential to enhance model performance further by incorporating additional ensemble techniques or exploring combinations of algorithms. For future work, it would be valuable to consider other ensemble methods such as Gradient Boosting Machines (GBM), AdaBoost, or Bagging, which can complement RF and XGBoost by offering strengths like better generalization and reduced variance

(Kadam & Jadhav, 2020; Sahin, 2022), ultimately addressing limitations such as overfitting and overestimation.

Furthermore, hybrid approaches that combine RF or XGBoost with other machine learning algorithms, such as Support Vector Machines (SVM) or Deep Learning models, may provide improved predictive accuracy. By experimenting with stacking or other ensemble combinations, future research could explore how these techniques can further refine landslide susceptibility predictions and enhance model robustness. While XGBoost and RF have been highly effective in this study, these suggestions for future work could lead to even more reliable and comprehensive models in landslide risk assessment.

Conclusion

Landslides pose a significant threat to Penang Island, demanding both immediate and long-term mitigation efforts. Landslide susceptibility mapping is essential for identifying high-risk areas and guiding effective land-use planning. This study highlights the value of integrating machine learning techniques with GIS-based spatial data for sustainable risk management. a) Both RF and XGBoost models performed well in predicting landslide susceptibility for Penang Island.

b) The Boruta algorithm identified the most important landslide conditioning factors (LCFs). Among the ten factors considered, slope angle and DEM had the most significant impact on landslide occurrences. Additionally, the multicollinearity test showed no issues among the LCFs.

c) The models were validated using Receiver Operating Characteristic (ROC) analysis, showing good performance with AUC values of 95.8% for BayesOpt-RF, 95.6% for RF, 95.2% for XGBoost, and 93.0% for BayesOpt-XGBoost.

d) Bayesian hyperparameter optimization (BayesOpt) helped optimize both RF and XGBoost. Although BayesOpt led to a slight improvement in BayesOpt-RF and a minor reduction in BayesOpt-XGBoost, it successfully minimized overfitting while maintaining strong prediction performance, with AUC scores above 90%.

e) The landslide susceptibility maps generated in this study can be used by town planners, engineers, and local authorities to identify high-risk areas, guide development planning, and create effective strategies for landslide risk mitigation.

List of Abbreviations

Abbreviation	Full Form
AUC	Area under curve
CNN	Convolutional Neural Networks
DEM	Digital Elevation Model
GIS	Geographic Information System
KNN	K-Nearest Neighbors
LCF	Landslide Conditioning Factors
LR	Logistic Regression
ML	Machine Learning
NDVI	Normalized Difference Vegetation Index
PR	Prediction Rate
RF	Random Forest
SPI	Stream Power Index
SR	Success Rate
SV	Support Vector Machine
TWI	Topographic Wetness Index
XGBoost	Extreme Gradient Boosting

Recommendations

It is recommended to consider incorporating a temporal analysis in future studies to account for changes in landslide susceptibility over time, particularly due to climate or land-use changes. This would provide a more comprehensive understanding of future risks and their potential evolution. Furthermore, while RF and XGBoost are useful in this study, it would be beneficial to explore other ensemble techniques or combinations of algorithms. This could help address issues such as overfitting and overestimation, potentially improving the model's performance and robustness.

Acknowledgement

The authors acknowledge Ministry of Higher Education Malaysia (MOHE) through the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2022/ICT02/UNIMAS/02/2) and Universiti Malaysia Sarawak (UNIMAS) for supporting this project.

Funding Information

The authors acknowledge Ministry of Higher Education Malaysia (MOHE) through the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2022/ICT02/UNIMAS/02/2) and Universiti Malaysia Sarawak (UNIMAS) for supporting this project.

Ethics

This research does not involve human or animal subjects. So, no ethics approval is required for this research.

Authors Contributions

Dorothy Anak Martin Atok: Led conceptualization, methodology design, and initial draft writing.

Soo See Chai: Provided project supervision, critical review, and overall guidance.

Kok Luong Goh: Performed software support, and result visualization.

Neha Gautam: Contributed to resource management.

Kim On Chin: Assisted with manuscript editing.

References

Achour, Y., & Reza, H. (2020). How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geoscience Frontiers, 11*(3), 871–883. https://doi.org/10.1016/j.gsf.2019.10.001

Aghdam, I. N., Varzandeh, M. H. M., & Pradhan, B. (2016). Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). *Environmental Earth Sciences, 75*(553). https://doi.org/10.1007/s12665-015-5233-6

Ahmadpour, H., Bazrafshan, O., Rafiei-Sardooi, E., Zamani, H., & Panagopoulos, T. (2021). Gully Erosion Susceptibility Assessment in the Kondoran Watershed Using Machine Learning Algorithms and the Boruta Feature Selection. *In Sustainability*, *13*(18). https://doi.org/10.3390/su131810110

Al-Thuwaynee, O., Aydda, A., Hwang, I.-T., Kim, S.-W., & Park, H. J. (2021). LaGriSU tool pack for the automatic extraction of grid units and slope units: application to Inje Province (South Korea). *대한지질공학회 학술발표논문집, 2021*(1), 27-27.

Al-Thuwaynee, O. F., Melillo, M., Gariano, S. L., Park, H. J., Kim, S.-W., Lombardo, L., Hader, P., Mohajane, M., Quevedo, R. P., & Catani, F. (2023). DEWS: A QGIS tool pack for the automatic selection of reference rain gauges for landslide-triggering rainfall thresholds. *Environmental Modelling & Software, 162*, 105657.

Amiri, M., Pourghasemi, H. R., Ghanbarian, G. A., & Afzali, S. F. (2019). Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma*, *340*, 55-69. https://doi.org/10.1016/j.geoderma.2018.12.042

Avand, M., Janizadeh, S., Naghibi, S. A., Pourghasemi, H. R., Bozchaloei, S. K., & Blaschke, T. (2019). A comparative assessment of Random Forest and k-Nearest Neighbor classifiers for gully erosion susceptibility mapping. *Water (Switzerland), 11*(10). https://doi.org/10.3390/w11102076

Bag, R., Mondal, I., Dehbozorgi, M., Bank, S. P., Das, D. N., Bandyopadhyay, J., Pham, Q. B., Fadhil Al-Quraishi, A. M., & Nguyen, X. C. (2022). Modelling and mapping of soil erosion susceptibility using machine learning in a tropical hot sub-humid environment. *Journal of Cleaner Production*, *364*, 132428.https://doi.org/10.1016/j.jclepro.2022.132428

Błońska, E., Lasota, J., Piaszczyk, W., Wiecheć, M., & Klamerus-Iwan, A. (2018). The effect of landslide on soil organic carbon stock and biochemical properties of soil. *Journal of Soils and Sediments*, *18*(8), 2727–2737. https://doi.org/10.1007/s11368-017-1775-4

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Brenning, A. (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Sciences*, *5*(6), 853–862.

Cao, J., Zhang, Z., Du, J., Zhang, L., Song, Y., & Sun, G. (2020). Multi-geohazards susceptibility mapping based on machine learning—A case study in Jiuzhaigou, China. *Natural Hazards*, *102*, 851–871. https://doi.org/10.1007/s11069-020-03927-8

Chen, J., Yang, S. T., Li, H. W., Zhang, B., & Lv, J. R. (2013). Research on geographical environment unit division based on the method of natural breaks (Jenks). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*, 47–50. https://doi.org/10.5194/isprsarchives-XL-4-W3-47-2013

Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M., Li, T., Peng, T., Guo, C., Niu, C., Xiao, L., Wang, J., Xie, X., & Ahmad, B. Bin. (2018). GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Science of the Total Environment*, *634*, 853–867. https://doi.org/10.1016/j.scitotenv.2018.04.055

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, *151*, 147–160.https://doi.org/10.1016/j.catena.2016.11.032

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., Dhital, M. R., & Althuwaynee, O. F. (2013). Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. *Natural Hazards*, *65*(1), 135–165. https://doi.org/10.1007/s11069-012-0347-6

Elmahdy, S. I., Marghany, M. M., & Mohamed, M. M. (2016). Application of a weighted spatial probability

model in GIS to analyse landslides in Penang Island, Malaysia. *Geomatics, Natural Hazards and Risk*, *7*(1), 345359.https://doi.org/10.1080/19475705.2014.904825

Fabbri, A. G., Chung, C. F., Napolitano, P., Remondo, J., & Zêzere, J. L. (2002). Prediction rate functions of landslide susceptibility applied in the Iberian Peninsula. *WIT Transactions on Modelling and Simulation*, 31. https://doi.org/10.2495/RISK020681

Florinsky, I. V. (2017). An illustrated introduction to general geomorphometry. *Progress in Physical Geography*, *41*(6), 723–752. https://doi.org/10.1177/0309133317733667

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis, 38*(4), 367–378. https://doi.org/https://doi.org/10.1016/S0167-9473(01)00065-2

Guo, Z., Torra, O., Hürlimann, M., Abancó, C., & Medina, V. (2022). FSLAM: A QGIS plugin for fast regional susceptibility assessment of rainfall-induced landslides. *Environmental Modelling & Software*, *150*, 105354. https://doi.org/10.1016/j.envsoft.2022.105354

Han, G., Shan, F. P., Tien, T. L., & Chin, L. H. (2021). Landslide susceptibility analysis using gradient boosting models: A case study in penang island, malaysia. *Disaster Advances*, *14*(8), 22–37. https://doi.org/10.25303/148da2221

Hussain, M. A., Chen, Z., Wang, R., & Shoaib, M. (2021). Ps-insar-based validated landslide susceptibility mapping along karakorum highway, Pakistan. *Remote Sensing*, *13*(20). https://doi.org/10.3390/rs13204129

Hussain, M. A., Chen, Z., Zheng, Y., Shoaib, M., Shah, S. U., Ali, N., & Afzal, Z. (2022). Landslide Susceptibility Mapping Using Machine Learning Algorithm Validated by Persistent Scatterer In-SAR Technique. *Sensors*, *22*(9), 209–224. https://doi.org/10.3390/s22093119

Jaafari, A., Najafi, A., Pourghasemi, H. R., Rezaeian, J., & Sattarian, A. (2014). GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *International Journal of Environmental Science and Technology*, *11*(4), 909–926. https://doi.org/10.1007/s13762-013-0464-0

Kadam, V. J., & Jadhav, S. M. (2020). Performance analysis of hyperparameter optimization methods for ensemble learning with small and medium sized medical datasets. *Journal of Discrete Mathematical Sciences and Cryptography*, *23*(1), 115–123. https://doi.org/10.1080/09720529.2020.1721871

Kavzoglu, T., Colkesen, I., & Sahin, E. (2019). Machine Learning Techniques in Landslide Susceptibility Mapping: A Survey and a Case Study. *Landslides: Theory, practice and modelling,* 283-301. https://doi.org/10.1007/978-3-319-77377-3_13

Khosravi, K., Panahi, M., Golkarian, A., Keesstra, S. D., Saco, P. M., Bui, D. T., & Lee, S. (2020). Convolutional neural network approach for spatial prediction of flood hazard at national scale of Iran. *Journal of Hydrology*, *591*, 125552. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125552

Lee, S., & Pradhan, B. (2006). Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. *Journal of Earth System Science*, *115*(6), 661–672. https://doi.org/10.1007/s12040-006-0004-0

Li, B., Wang, N., & Chen, J. (2021). GIS-Based Landslide Susceptibility Mapping Using Information, Frequency Ratio, and Artificial Neural Network Methods in Qinghai Province, Northwestern China. *Advances in Civil Engineering*, 2021. https://doi.org/10.1155/2021/4758062

Li, L., Jia, M., Xu, C., Tian, Y., Ma, S., & Yang, J. (2024). Enhancing Landslide Susceptibility Mapping by Integrating Neighboring Information in Slope Units: A Spatial Logistic Regression. *Remote Sensing*, *16*(23), 1–18. https://doi.org/10.3390/rs16234475

Malay Mail. (2018, October 23). Bukit Kukus landslide: SAR stopped after exhaustive search, says district police chief. *Malay Mail*. https://www.malaymail.com/news/malaysia/2018/10/23/bukit-kukus-landslide-sar-stopped-after-exhaustive-search-says-district-pol/1685849

Meena, S. R., Puliero, S., Bhuyan, K., Floris, M., & Catani, F. (2022). Assessing the importance of conditioning factor selection in landslide susceptibility for the province of Belluno (region of Veneto, northeastern Italy). *Natural Hazards and Earth System Sciences*, *22*(4), 1395–1417. https://doi.org/10.5194/nhess-22-1395-2022

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, *207*, 103225.https://doi.org/https://doi.org/10.1016/j.earscirev.2020.103225

Nachappa, T. G., Piralilou, S. T., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, *590*, 125275. https://doi.org/10.1016/j.jhydrol.2020.125275

Nam, K., & Wang, F. (2020). An extreme rainfall-induced landslide susceptibility assessment using autoencoder combined with random forest in Shimane Prefecture, Japan. *Geoenvironmental Disasters*, *7*(1), 6.

https://doi.org/10.1186/s40677-020-0143-7

Nhu, V. H., Mohammadi, A., Shahabi, H., Ahmad, B. Bin, 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. *International Journal of Environmental Research and Public Health*, *17*(14), 1–23. https://doi.org/10.3390/ijerph17144933

Nnanwuba, U. E., Qin, S., Adeyeye, O. A., Cosmas, N. C., Yao, J., Qiao, S., Jingbo, S., & Egwuonwu, E. M. (2022). Prediction of Spatial Likelihood of Shallow Landslide Using GIS-Based Machine Learning in Awgu, Southeast/Nigeria. *Sustainability*, *14*(19), 12000. https://doi.org/10.3390/su141912000

Park, S., & Kim, J. (2019). Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Applied Sciences*, *9*(5). https://doi.org/10.3390/app9050942

Pašek, J. (1975). Landslides inventory. *Bulletin of the International Association of Engineering Geology - Bulletin de l'Association Internationale de Géologie de l'Ingénieur*, *12*, 73–74. https://doi.org/10.1007/BF02635432

Pham, B. T., Tien Bui, D., Pourghasemi, H. R., Indra, P., & Dholakia, M. B. (2017). Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theoretical and Applied Climatology*, 128, 255–273. https://doi.org/10.1007/s00704-015-1702-9

Pham, B. T., Vu, V. D., Costache, R., Phong, T. Van, Ngo, T. Q., Tran, T. H., Nguyen, H. D., Amiri, M., Tan, M. T., Trinh, P. T., Le, H. Van, & Prakash, I. (2022). Landslide susceptibility mapping using state-of-the-art machine learning ensembles. *Geocarto International*, *37*(18), 5175–5200. https://doi.org/10.1080/10106049.2021.1914746

Pourghasemi, H. R., & Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena*, *162*, 177–192. https://doi.org/10.1016/j.catena.2017.11.022

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers and Geosciences*, *51*, 350–365. https://doi.org/10.1016/j.cageo.2012.08.023

Pradhan, B., Lee, S., & Buchroithner, M. F. (2010). A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. *Computers, Environment and Urban Systems*, *34*(3), 216–235. https://doi.org/10.3390/su141912000

Rabby, Y. W., Hossain, M. B., & Abedin, J. (2020). Landslide susceptibility mapping in three Upazilas of Rangamati hill district Bangladesh: application and comparison of GIS-based machine learning methods. *Geocarto International*, *37*(12), 3371-3396.. https://doi.org/10.1080/10106049.2020.1864026

Rabby, Y. W., & Li, Y. (2020). Landslide susceptibility mapping using integrated methods: A case study in the chittagong hilly areas, bangladesh. *Geosciences*, *10*(12), 1–26. https://doi.org/10.3390/geosciences10120483

Rana, S., Li, C., Gupta, S., Nguyen, V., & Venkatesh, S. (2017). High dimensional Bayesian optimization with elastic Gaussian process. *International Conference on Machine Learning*, (pp. 2883-2891).PMLR. https://proceedings.mlr.press/v70/rana17a.html.

Saha, A., & Saha, S. (2022). Integrating the artificial intelligence and hybrid machine learning algorithms for improving the accuracy of spatial prediction of landslide hazards in Kurseong Himalayan Region. *Artificial Intelligence in Geosciences*, *3*, 14–27. https://doi.org/10.1016/j.aiig.2022.06.002

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, *2*(7), 1–17. https://doi.org/10.1007/s42452-020-3060-1

Sahin, E. K. (2022). Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. *Geocarto International*, *37*(9), 2441–2465. https://doi.org/10.1080/10106049.2020.1831623

Sameen, M. I., Pradhan, B., & Lee, S. (2020). Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *Catena*, *186*, 104249. https://doi.org/10.1016/j.catena.2019.104249

Sevgen, E., Kocaman, S., Nefeslioglu, H. A., & Gokceoglu, C. (2019). Photogrammetric Techniques for Landslide Susceptibility Mapping with Logistic Regression. *Sensors*, *19*(18), 3940. https://doi.org/10.3390/s19183940

Sidaway, J. (2022). Working Principles of Digital Elevation Model and its Applications. *Journal of Remote Sensing and GIS*, *11*(6), 1000237. https://doi.org/10.35248/2469-4134.22.11.237.

Sidle, R. C., & Ochiai, H. (2013). Landslides : Processes , Prediction , and Land Use. *Water resources monograph. American Geophysical Union, Washington*, 525 .https://doi.org/10.1029/WM018

Srinivas, N., Krause, A., Kakade, S. M., & Seeger, M. W. (2012). Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, *58*(5), 3250–3265. https://doi.org/ https://doi.org/10.1109/TIT.2011.2182033

Sun, D., Wen, H., Wang, D., & Xu, J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, *362*, 107201.

https://doi.org/10.1016/j.geomorph.2020.107201

The Sun Daily. (2018, November 13). Tanjung Bungah landslide: A man-made tragedy, says Commission of Inquiry. https://www.thesundaily.my/local/tanjung-bungah-landslide-a-man-made-tragedy-says-commission-of-inquiry-DB1311207

Timeline of landslides in Malaysia in 2022. Channel News Asia. https://www.channelnewsasia.com/asia/malaysia-landslides-2022-timeline-3150361

Tsangaratos, P., Ilia, I., Hong, H., Chen, W., & Xu, C. (2017). Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China. *Landslides, 14*, 1091–1111. https://doi.org/10.1007/s10346-016-0769-4

Vapnik, V. (1999). The nature of statistical learning theory. *Springer science & business media*.

Wang, S., Zhuang, J., Zheng, J., Fan, H., Kong, J., & Zhan, J. (2021). Application of Bayesian Hyperparameter Optimized Random Forest and XGBoost Model for Landslide Susceptibility Mapping. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.712240

Wang, Y., Sun, D., Wen, H., Zhang, H., & Zhang, F. (2020). Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang county (Chongqing, China). *International Journal of Environmental Research and Public Health*, *17*(12), 1–39. https://doi.org/10.3390/ijerph17124206

Wischmeier, W. H., & Smith, D. D. (1978). Predicting rainfall erosion losses: A guide to conservation planning (USDA Agriculture Handbook No. 537). U.S. Department of Agriculture, Science and Education Administration.

Wubalem, A., & Meten, M. (2020). Landslide susceptibility mapping using information value and logistic regression models in Goncha Siso Eneses area, northwestern Ethiopia. *SN Applied Sciences*, *2*(5),1–19.https://doi.org/10.1007/s42452-020-2563-0

Xiao, T., Yin, K., Yao, T., & Liu, S. (2019). Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou County, Three Gorges Reservoir, China. *Acta Geochimica*, *38*(5), 654–669. https://doi.org/10.1007/s11631-019-00341-1

Yi, Y., Zhang, Z., Zhang, W., Jia, H., & Zhang, J. (2020). Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *Catena*, *195*, 104851. https://doi.org/10.1016/j.catena.2020.104851

Yilmaz, I. (2010). Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environmental Earth Sciences*, *61*(4), 821–836. https://doi.org/10.1007/s12665-009-0394-9

Youssef, A. M., & Pourghasemi, H. R. (2021). Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geoscience Frontiers*, *12*(2), 639–655. https://doi.org/10.1016/j.gsf.2020.05.010

Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (2016). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, *13*(5), 839–856. https://doi.org/10.1007/s10346-015-0614-1

Zhang, H., Song, Y., Xu, S., He, Y., Li, Z., Yu, X., Liang, Y., Wu, W., & Wang, Y. (2022). Combining a class-weighted algorithm and machine learning models in landslide susceptibility mapping: A case study of Wanzhou section of the Three Gorges Reservoir, China. *Computers & Geosciences*, *158*, 104966. https://doi.org/https://doi.org/10.1016/j.cageo.2021.104966

Zhang, Y., Ge, T., Tian, W., & Liou, Y. A. (2019). Debris flow susceptibility mapping using machine-learning techniques in Shigatse area, China. *Remote Sensing*, *11*(23). https://doi.org/10.3390/rs11232801

Zhao, Z., Liu, Z. yuan, & Xu, C. (2021). Slope Unit-Based Landslide Susceptibility Mapping Using Certainty Factor, Support Vector Machine, Random Forest, CF-SVM and CF-RF Models. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.589630

Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., & Pourghasemi, H. R. (2018). Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. *Computers and Geosciences*, *112*, 23–37. https://doi.org/10.1016/j.cageo.2017.11.019

Zhou, X., Wen, H., Zhang, Y., Xu, J., & Zhang, W. (2021). Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geoscience Frontiers*, *12*(5), 101211. https://doi.org/10.1016/j.gsf.2021.101211