



Addressing overfitting and overestimation challenges in landslide susceptibility modeling: a case study of Penang Island, Malaysia

Dorothy Anak Martin Atok¹ · Soo See Chai¹

Received: 11 December 2023 / Accepted: 23 April 2025
© The Author(s) 2025

Abstract

In the realm of landslide susceptibility prediction, the challenge of overfitting and overestimation has persisted despite various modeling attempts. This study aims to elevate the predictive capabilities of the Extreme Gradient Boosting (XGBoost) and Random Forest (RF) models for landslide susceptibility assessment through the innovative application of Bayesian Optimization (BO). Using data from Penang Island in Malaysia, we comprehensively incorporated topographical, hydrological, human, and environmental factors influencing landslides. Leveraging Geographic Information System (GIS) tools, we meticulously constructed spatial databases encompassing all pertinent landslide conditioning elements. Our findings unveil the remarkable performance of the optimized XGBoost model, achieving an astounding 100.0% Success Rate (SR) and an impressive 97.1% Prediction Rate (PR). In comparison, the optimized RF model achieved an SR of 99.7% and a PR of 96.3%, while the stacked models followed closely with an SR of 96.8% and a PR of 95.6%. These conclusive results underscore the transformative potential of addressing overfitting and overestimation challenges through the strategic combination of stacking and hyperparameter optimization. The improved accuracy of these algorithms bears immense significance, extending to applications in site selection, engineering structure health monitoring, and disaster mitigation, thus elevating the importance of Landslide Susceptibility Maps (LSMs) in safeguarding communities and infrastructure.

Keywords Extreme gradient boosting · Geographic information system · Hybrid · Landslide susceptibility · Random forest

✉ Soo See Chai
sschai@unimas.my

Dorothy Anak Martin Atok
dorothymartinatok@gmail.com

¹ Faculty of Computer Science and Information Technology, University of Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

1 Introduction

In Malaysia, landslides rank as the second most catastrophic natural disaster following floods. Despite the absence of significant earthquakes, the country continues to experience large-scale landslides primarily triggered by gravity and relentless heavy rainfall. The rapid development since the 1980s has made locating suitable low-lying areas for urban expansion increasingly challenging. This has led to heightened development in highland and steep terrain regions adjacent to densely populated cities, consequently elevating the risk of landslides for urban residents.

In Malaysia, landslides, such as shallow slides, mudflows, debris flows, and rockfalls, are common occurrences during or after heavy rainfall (Rahman and Mapjabil 2017). To mitigate the economic impact of these events, hazard mapping, historical data analysis, and prioritized hazard reduction planning are crucial. Therefore, developing a Landslide Susceptibility Map (LSM) is essential for assessing landslide risk and implementing risk reduction strategies, as it identifies potential landslide-prone areas (Senouci et al. 2021). Furthermore, landslide susceptibility reflects the likelihood of landslides occurring due to various factors (Reichenbach et al. 2018).

Machine learning algorithms have significantly improved landslide prediction accuracy and adaptability (Bui et al. 2018). These techniques encompass traditional models like Logistic Regression (LR), Artificial Neural Networks (ANN), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). While LR is widely utilized but performs best in less complex scenarios with smaller sample sizes (Yilmaz 2010), SVM excels with small datasets but is impractical for large-scale data (Fang et al. 2020). ANNs, despite their complexity, may struggle with certain landslide comparisons (Agrawal & Dixit 2023). At the same time, KNN is straightforward but costly as the sample size increases (Mutlu et al. 2019).

Hybrid models, integrating multiple models or optimization techniques, have gained popularity for their improved prediction performance (Ado et al. 2022). They employ methods like Artificial Bee Colony (ABC) Optimization, Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Bayesian Optimization (BO). Deep learning algorithms like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have also shown promise in landslide susceptibility analysis (Paoletti et al. 2018). CNN automatically extracts meaningful features, while RNN is suitable for national-scale analysis but not regional-scale (Zhu et al. 2017).

Ensemble machine learning techniques, such as boosting and bagging, enhance landslide vulnerability assessment by combining multiple classifiers (Bui et al. 2016). Boosting sequentially adds classifiers to control bias and variance, while bagging generates multiple decision trees (Pham et al. 2020). Random Forest (RF) is a popular bagging ensemble method, and Extreme Gradient Boosting (XGBoost) is a well-known boosting method.

The selected technique must be effective in dealing with the nonlinear relationship between the landslide conditioning factors and the occurrences of landslides. It is one of the few characteristics that must be considered in choosing an appropriate approach to be employed (Mezaal et al., 2018). Next, the method of choice must manage geographic heterogeneity, consume numerous input parameters, and manage numerous output parameters such as classification and regression (Taalab et al. 2018). It is also crucial to consider prediction accuracy as well as non-linearity issues. The advantage of choosing many machine learning models over just one is that the hybrid-ensemble technique emphasizes each model's strengths while minimizing its flaws (Chen & Li 2020).

Overestimation and overfitting are common challenges in LSM (Agrawal & Dixit 2023). Overestimation reduces map precision and wastes resources (Rabby et al. 2020). Maintaining model accuracy while simplifying complexity is challenging, and addressing overfitting requires further research. Hence, hyperparameter optimization can help in creating an accurate yet less complex model.

The nonlinear relationship between the landslide conditioning factors and the landslide susceptibility can be addressed by the "black box" model RF (Sahin 2022). RF is more capable of addressing the overestimation problem. On the other hand, the XGBoost model outperforms the RF in coping with class imbalance and is more resistant to overfitting due to its structure (Wang et al. 2020). Thus, establishing heterogeneous ensemble learning through stacking the two classifiers can further improve the capacity to predict homogeneous learners (Susan et al. 2021).

Therefore, this research study aims to address overestimation and overfitting using a hybrid ensemble of RF and XGBoost with BO Hyperparameter techniques to develop an LSM for Penang Island, Malaysia. There were ten additional conditioning factors and model performance was evaluated using a confusion matrix, statistics, and the area under the Receiver Operating Characteristic (ROC) curve (AUC). Accordingly, the resulting LSM is valuable for urban planning, infrastructure development, and land use planning to reduce landslide hazards (Susan et al. 2021).

2 Study area

This research focused on Penang Island, located in Malaysia (Fig. 1). The study area lies to the northwest of the peninsula. The island is 306 km² in area, and a river separates it from the mainland. The island is located between latitudes 5°15'N and 5°30'N and longitudes

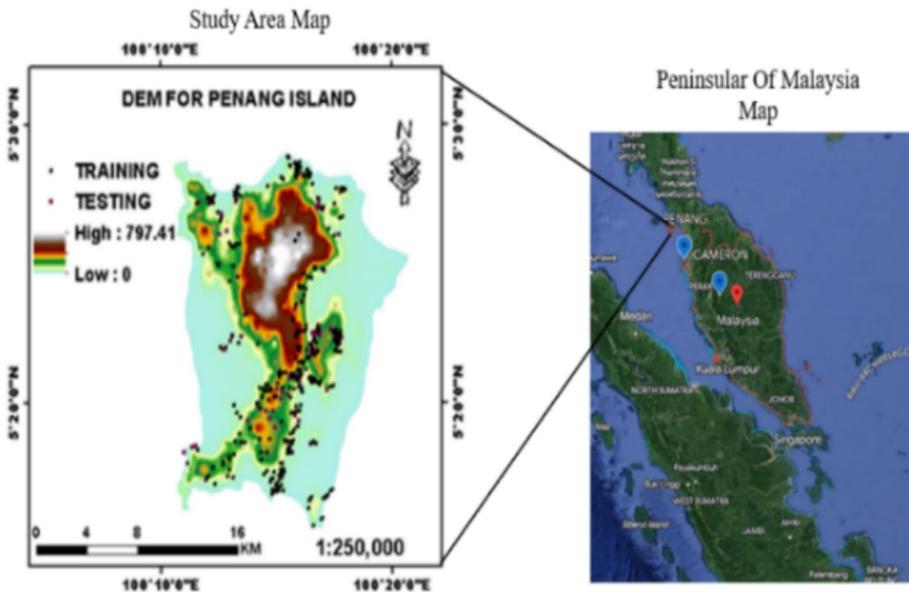


Fig. 1 Penang Island study areas in Malaysia

100°10'E and 100°20'E, where the study area is located. Most of the bedrock in the research area is composed of granite. As a result, 443 landslide locations have been discovered on Penang Island.

3 Methodology

3.1 Data collection

The initial phase of this methodology was data collection, followed by creating a landslide inventory map, establishing spatial databases for landslide conditioning factors, assessing the significance of these factors, modeling and validating landslide susceptibility models, and ultimately producing the final susceptibility map (Fig. 2). Multiple data sources were utilized in this study, with a strong reliance on open-source software such as RStudio and Geographic Information Systems (GIS), including ArcMap 10.4 and QGIS 3.8.

Data for this study was primarily sourced from publicly available resources. Google Earth Pro was employed to determine altitude information for the study areas. GPS Exchange Format (GPX) file format features generated from these elevation points were imported into ArcMap to create a Digital Elevation Model (DEM) at a 1:250,000 scale, utilizing ArcMap's interpolation capabilities. Simultaneously, Landsat 8 imagery with spatial resolutions of 30 m in multispectral and 15 m in panchromatic bands was obtained from the United States Geological Survey (USGS) Earth Explorer. Meanwhile, road network data was extracted from OpenStreetMap. Landslide locations were identified using the National Aeronautics and Space Administration (NASA) Global Landslides Catalogue. Subsequently, all datasets were projected to UTM-Zone 47N with the WGS84 Datum.

3.2 Landslide inventories

Landslide inventory mapping is fundamental in landslide prediction and susceptibility mapping (Pradhan et al. 2010; Roslee et al. 2012). This study employed a pre-existing landslide inventory map, augmented with the latest landslide data from NASA Landslide Catalogues. To acquire non-landslide grid cells, we explored three techniques from the literature: (i) the seed cell approach, (ii) random selection from landslide-free areas, and (iii) selecting sites with slopes less than 2° (Huang et al. 2017). Our approach utilized Grid Unit Extraction to randomly select areas within the study region, aligning with the second method.

3.3 Landslides Conditioning Factors (LCFs)

The choice of landslide conditioning factors significantly influences the predictive capabilities of machine learning algorithms for landslide susceptibility. Previous studies on LSM utilizing machine learning approaches (Chen et al. 2018a, b, c; Moayedi et al. 2019) have employed various combinations of these factors. However, the selection criteria for these factors should consider their (i) relevance to landslide occurrences, (ii) measurability, (iii) non-redundancy, and (iv) alignment with the geomorphological characteristics of the study area (Ayalew & Yamagishi 2005). In this study, we have identified ten variables for inclusion, drawing from existing literature and the expertise of subject-matter specialists.

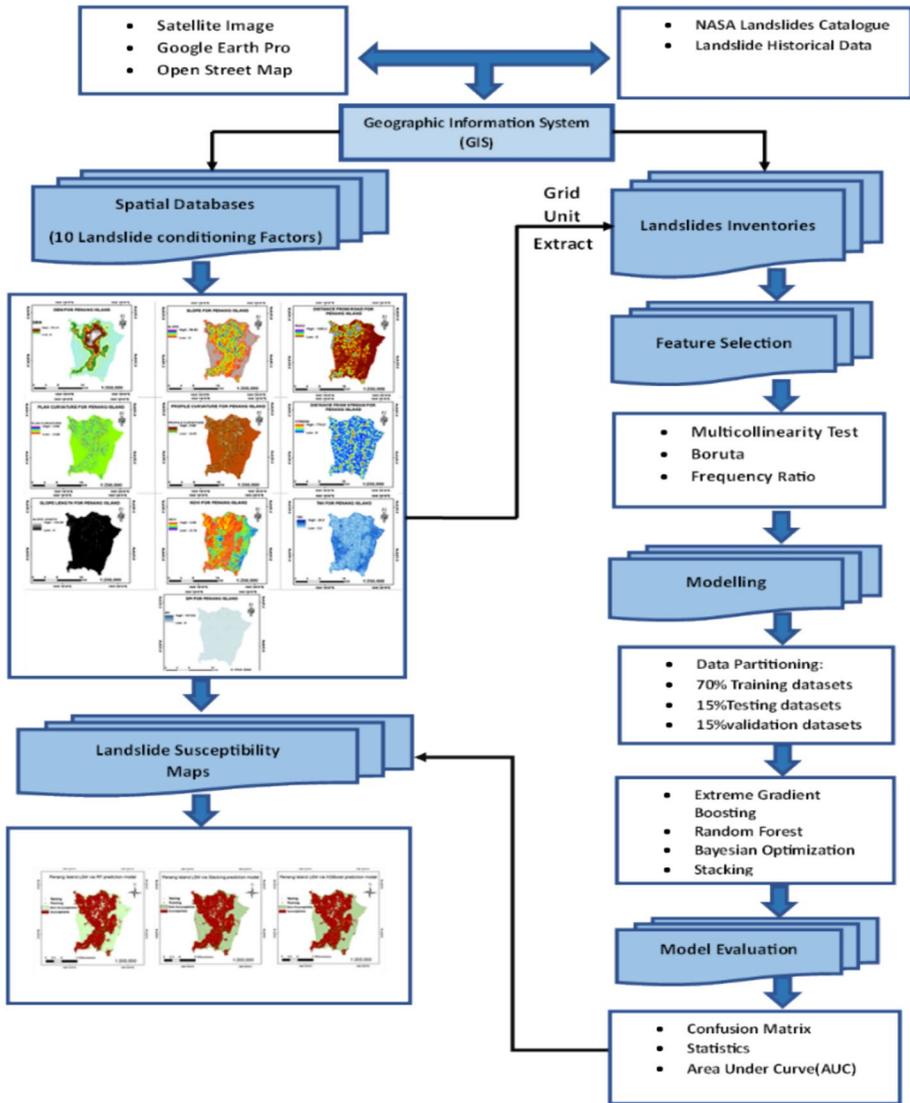


Fig. 2 Research methodology flowchart

3.3.1 Topographical factors

Topographical characteristics encompass slope angle, slope length, profile curvature, and plan curvature, all derived from DEMs. Elevation plays a pivotal role, impacting both human activities and various biophysical attributes, consequently influencing landslide occurrences. Studies have even pinpointed specific elevation levels as prone to landslides in certain basins (Gómez & Kavzoglu 2005). Moreover, elevation influences other

landslide-related factors like slope, curvature, and Stream Power Index (SPI) (Chen et al. 2018a, b, c).

A DEM is a 3D representation of terrain elevation. DEMs are critical for LSM, providing elevation-based data for deriving landslide conditioning factors. In this study, our DEM was constructed from Google Earth elevation point data using the "Interpolation Tool," specifically Inverse Distance Weighted (IDW), resulting in a 30 m × 30 m grid. The elevations on Penang Island range from 0 to 797.41 m (Fig. 3a).

Slope angle is a significant factor influencing landslides due to its impact on moisture concentration, pore pressure, and hydraulic continuity at a regional scale (Ayalew & Yamagishi 2005; Duman et al. 2006). Each process plays a role in slope instability (Chen et al. 2017a, b). On Penang Island, slope angles vary from 0° to 59.62° (Fig. 3b). Curvature is another influential factor, directly controlling water flow velocity and erosion processes (Chen et al. 2017a, b; Duman et al. 2006). Plan curvatures on Penang Island range from -3.29 to 3.68 (Fig. 3c), while profile curvatures vary from -6.43 to 5.80 (Fig. 3d).

The slope length gradient (LS) component combines slope length (L) and slope steepness (S) to determine soil erosion rates. Additionally, longer slopes accumulate more runoff, and steeper slopes exhibit higher runoff velocities, leading to erosion (Das et al. 2022). In addition, slope lengths on Penang Island span from 0 to 124.09 (Fig. 3e).

3.3.2 Anthropogenic and environmental factors

This study considers additional anthropogenic, environmental, and geographic factors, including the distance to a road, distance to a stream, and the Normalized Difference Vegetation Index (NDVI). Proximity to highways is a significant determinant of landslide locations, as both natural and man-made slopes near roads are susceptible to this hazard (Rozos et al. 2011). Anthropogenic factors, such as road cuts, excavation, and additional load, can induce soil instability, potentially triggering landslides (Pourghasemi et al. 2012). We employed the Euclidean distance technique to create a distance-to-road layer, with values ranging from 0 m to 1253.2 m for Penang Island (Fig. 3f).

Streams can impact an area's stability by eroding slopes, making the presence of natural drainage systems a potential indicator of landslide-prone locations (Ahmed et al. 2020). On Penang Island, stream elevations range from 0 m to 779.21 m (Fig. 3g). NDVI maps were generated using Landsat 8 satellite images from 2021, sourced from the USGS Earth Explorer. Moreover, NDVI plays a vital role in water retention and improving the shear resistance and soil cohesiveness of lithological masses (Sidle & Ochiai 2013). For Penang Island, the NDVI values range from -0.19 to 0.59 (Fig. 3h).

3.3.3 Hydrological factors

This study incorporates two key hydrological indicators: the SPI and the Topographic Wetness Index (TWI). SPI serves as an essential predictor of landslides by assessing a surface's erosion potential (Chen & Li, 2020b; Moayedi et al. 2019; Pirasteh & Li 2017). While Positive SPI values indicate minimal erosion risk, negative values signify a higher potential for erosion (Pirasteh & Li 2017). Across the research regions, SPI spans from 0 to 101,102 for Penang Island (Fig. 3i).

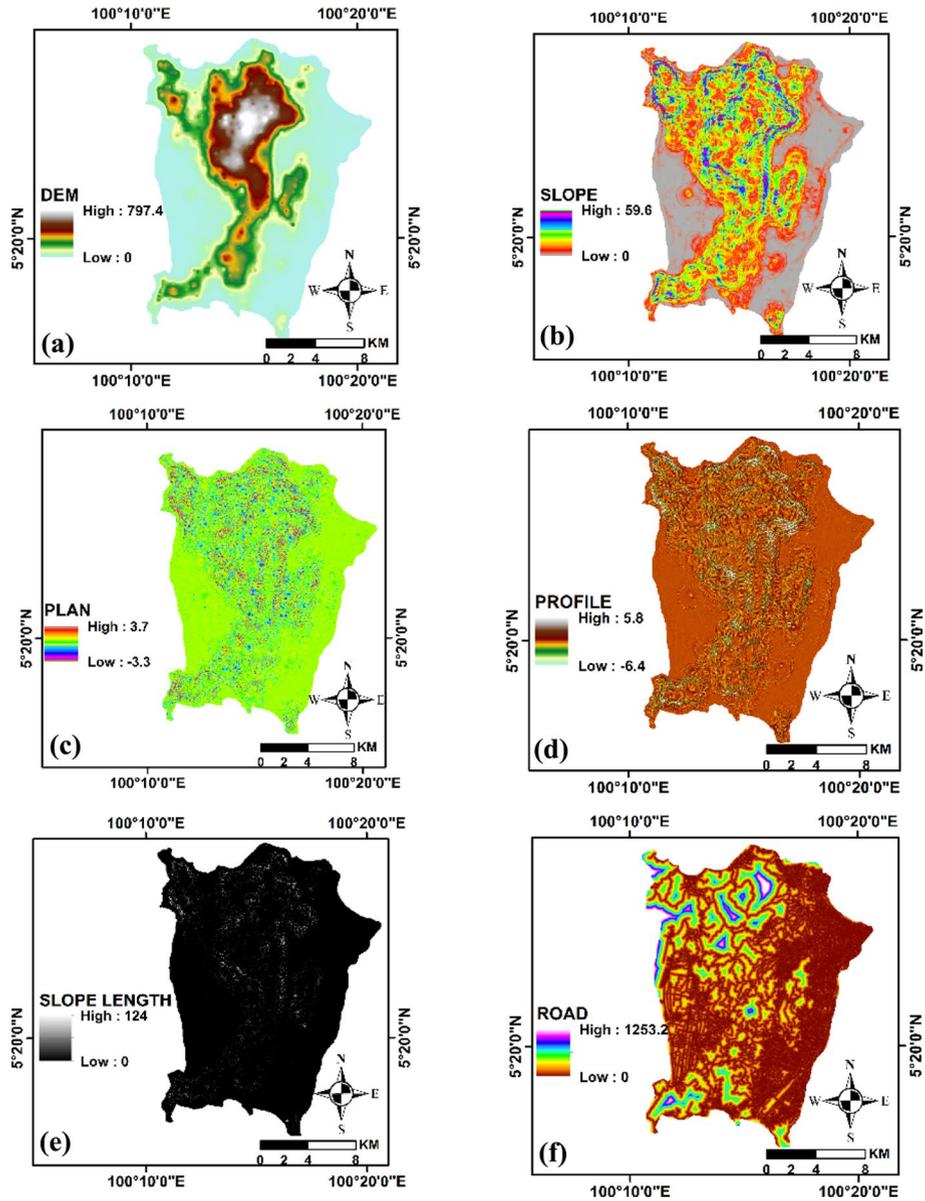


Fig. 3 Landslide conditioning spatial databases: **a** DEM, **b** slope angle, **c** plan curvature, **d** Profile curvature, **e** slope length **f** distance from Road, **g** distance from stream, **h** NDVI, **i** SPI, **j** TWI

On the other hand, TWI provides insights into terrain characteristics, runoff volume, and soil conditions (He et al. 2019). In this study, TWI was selected as an additional conditioning factor due to its ability to elucidate the interplay between topography and

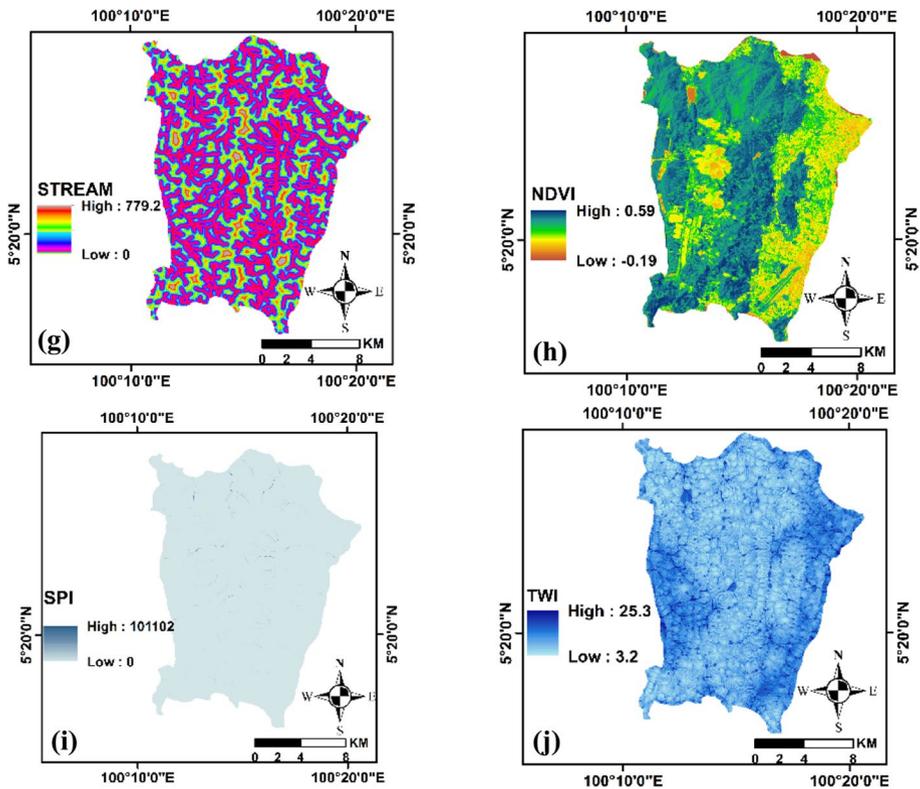


Fig. 3 (continued)

moisture (Z. Wang et al. 2020). TWI value on Penang Island ranges from 3.2 to 25.3 (Fig. 3j).

3.4 Feature selection

In the feature selection process, variables or attributes that hold greater significance and relevance to the target class are retained, while those that lack significance are eliminated. This feature selection process offers several key advantages, including mitigating the curse of dimensionality, expediting learning, enhancing generalization capabilities, and producing comprehensible models and explanations (Ao 2008). In this study, the Multicollinearity test and the Boruta model were employed to ascertain the significance of each landslide conditioning component.

To assess multicollinearity among the variables, we utilize the Variance Inflation Factor (VIF) and Tolerance (TOL) approaches. These methods have been frequently employed in prior studies (Ado et al. 2022; Amiri et al. 2019; Chen et al. 2018a, b, c; Gao et al. 2020; Rabby & Li 2020; Saha & Saha 2022). The "olsrr" package and the "ols_vif_tol()" function within RStudio are used to execute the multicollinearity tests for VIF and TOL concerning the landslide conditioning variables (Hebbali & Hebbali 2017). As a result, these tests

effectively identify redundant components that should be eliminated (Devkota et al. 2013; O'Brien, 2007).

However, it is essential to note that the absence of a direct relationship between certain factors does not necessarily imply their lack of significance in influencing landslides (Chan et al. 2022). Hence, we employ the Boruta model for feature selection in this project. Boruta, based on wrapper-based approaches, serves as a robust Factor Optimization Process (FOP) (Amiri et al. 2019). The Boruta model is implemented using the "Boruta" package in RStudio (Kursa & Rudnicki 2010).

3.5 Correlational analysis

After confirming the chosen conditioning factors, a more in-depth analysis of the relationship between the parameters selected and the occurrence of landslides was conducted. The correlational study was performed using Frequency Ratio (FR) analysis, similar to previous studies (Huang et al. 2020; Rabby & Li 2020; Xiao et al. 2019).

3.6 Modeling

The modeling process was conducted using RStudio software, version 4.3.0. Both the landslide datasets and the ten spatial databases were imported into R. The datasets for each study area were randomly divided into three sets: 70% for training, 15% for testing, and 15% for validation. It is worth noting that there is no specific criteria or universally accepted rule of thumb for determining the dataset splitting ratio, as observed in previous studies (Achour & Reza 2020; Hussain et al. 2021; Saha et al. 2021; Sahin, 2022b).

3.6.1 Extreme gradient boosting (XGBoost)

In recent years, the open-source gradient boosting method, XGBoost, has been employed in data science. The XGBoost adheres to the gradient boosting principle, which combines a set of weak learners' predictions to produce a strong learner using an additive training technique (Chen & Guestrin 2016). The XGBoost model requires several parameter selections to generate predictions. However, the model's effectiveness always depends on choosing the best parameters.

3.6.2 Random forest

For classification, regression, and unsupervised learning, RF is regarded as a potent ensemble learning technique (Chen et al. 2017a, b), while LSM has extensively used this technique (Chen et al. 2018a, b, c; Sun et al. 2020). In ensemble models, numerous weak learners are typically trained, and their combined outputs are then used to produce more accurate predictions. The RF algorithm builds weak learners in the form of decision trees.

3.6.3 Bayesian hyperparameter optimization

Enhancing the accuracy of machine learning models hinges on effective hyperparameter optimization. This procedure seeks to identify the optimal hyperparameter values based on an evaluation index (Sun et al. 2020). Widely adopted methods for hyperparameter

optimization encompass grid search, random search, and Bayesian optimization (BO) (Sameen et al. 2020). Furthermore, this study employed BO for hyperparameter optimization, selecting the best-performing hyperparameters for final training and testing of the respective machine-learning models.

BO had been widely used in previous LSM studies conducted in Fengjie County, China (Sun et al. 2020), Wuqi County, China (Wang et al. 2021) and Anhua County, China (Yang et al. 2023). All the stated previous LSM studies had only emphasized the improvement of machine learning models (RF, XGBoost and logistic regression) prediction performances because of the BO. However, similar studies did not highlight the effectiveness of hyperparameter optimization using BO in minimizing the occurrence of overfitting in the machine learning models.

Besides, it is crucial to determine which acquisition function to use in the BO for optimal results in the hyperparameter optimization process. Somehow, in the previous studies of LSM by Sun et al. (2020), Wang et al. (2021) and Yang et al. (2023) they did not address which acquisition functions of BO namely the upper confident bound (UCB), probability improvement (PI) and expected improvement (EI) suitable to be used in the BO.

Therefore, in this research, the focus of implementing the BO as a method of optimizing the RF and XGBoost models was to minimize the occurrence of overfitting due to the models' complexity and to determine the most suitable function acquisition for BO.

3.6.3.1 Upper confident bound (UCB) Bayesian optimization creates a posterior distribution of functions, called the Gaussian process (GP) that best describes the function that needs to be optimized (Rana et al. 2017). Posterior distribution improves as more observations are collected, and the algorithm becomes more convinced about which parameter space areas are worth exploring (Wu et al. 2019).

As the iteration process continues, the algorithm balances its needs for exploration and exploitation while considering what it learns about the target function. A GP is fitted to the known samples (points already investigated) at each step, and the posterior distribution, together with an exploration method like Upper Confidence Bound (UCB) (Srinivas et al. 2012).

The UCB, one of the most used acquisition functions. Therefore, by using this UCB acquisition function in this study an optimized RF and XGBoost model named RF_BO_UCB and XGBoost_BO_UCB will be developed.

3.6.3.2 Probability improvement (PI) Function probability improvement (PI) tries to explore near the current optimal value point to find the points most likely to prevail over the current optimal value (Kushner 1964). The search process continues until the number of iterations of the algorithm reaches the upper limit (Wu et al. 2019). Thus, by using this PI as acquisition function in this study an optimized RF and XGBoost model named RF_BO_PI and XGBoost_BO_PI will be developed.

3.6.3.3 Expected improvement (EI) The expected improvement (EI) acquisition function balances exploration and exploitation by quantifying the expected value of improvement over the current best observation. If the improvement of the function value is less than the expected value after the algorithm is executed, then the current optimal value point may be the local optimal solution, and the algorithm will find the optimum value point in

other positions of the domain (Wu et al. 2019). By using this EI acquisition function in this study optimized RF and XGBoost models named RF_BO_EI and XGBoost_BO_EI will be developed.

3.6.3.4 RF and XGBoost hyperparameters For RF, the BO process focused on tuning two key hyperparameters: "mtry," representing the number of sampled predictors at each step, and "min_n," indicating the minimum number of instances required in a node to enable further splitting. Additionally, number of trees (ntree) and nodelist were set to default value which is 500 and to 14 respectively (Hussain et al. 2022; Sun et al. 2021; Wang et al. 2021; Wu et al. 2019). The "tune_bayes" function was employed for BO in RF (Duval et al. 2022).

For XGBoost, the "ParBayesianOptimization" package facilitated the BO of hyperparameters. It necessitated the definition of an objective function that the optimizer could utilize, along with input from the hyperparameters themselves. The study tuned seven XGBoost hyperparameters: "nrounds," "eta," "gamma," "min_child_weight," "max_depth," "subsample," and "col_sample_bytree" (Can et al. 2021; Wang et al. 2021). Following the definition of the objective function, parameter boundaries were established to guide the optimizer's search process. The optimal values of hyperparameters of both RF and XGBoost are summarized in Table 1.

3.7 Performance evaluation

To assess the performance of the landslide susceptibility models, a well-established method known as the ROC curve and its associated AUC were employed (Ahmed et al. 2020; Chen et al. 2018a, b, c). The ROC curve indicates a binary classifier system's performance, illustrating sensitivity in relation to the false positive rate. Meanwhile, sensitivity measures a model's ability to correctly identify true positives among the sum of true positives and false negatives. Conversely, specificity gauges the ratio of true negatives to the sum of true and false positives. Moreover, the ROC curve is constructed by plotting the true positive percentage on the y-axis against the cumulative distribution function of the false positive percentage on the x-axis.

Table 1 Summarizes the hyperparameters optimal values obtained from different Acquisition Functions of BO in RF and XGBoost

Classifier	Hyperparameter	Optimal values		
		UCB	PI	EI
XGBoost	nrounds	69	25	21
	colsample_bytree	0.9040207	0.8259445	0.8259445
	subsample	1	0.942043	0.942043
	max_depth	10	2	2
	gamma	0.1	0.06856842	0.08526783
	eta	0.2	0.2	0.2
	min_child_weight	1	24.51712	24.51712
RF	ntree (default)	500	500	500
	node size (default)	14	14	14
	mtry	10	10	10
	min_n	1	1	1

The calculated AUC value falls into various classifications: poor (50%-60%), average (60%-70%), good (70%-80%), very good (80%-90%), and excellent (90%-100%) (Adnan et al. 2020a, b). In addition to the ROC and AUC assessment, a confusion matrix was generated to quantify several statistical indices, including overall accuracy, precision, and recall (Adnan, et al. 2020a, b; Ahmed et al. 2020).

4 Results and discussion

4.1 0.1.1 Multicollinearity test and Boruta

Table 2 summarizes the findings of the multicollinearity diagnostics study conducted on the entire dataset of Penang Island, including the training and test datasets. For Penang Island, the slope was discovered to have the highest VIF value (1.78) and the lowest TOL value (0.56). On top of that, the SPI has the highest TOL (0.90) and the lowest VIF (1.11). All landslide conditioning factors had $TOL < 0.1$ or $VIF > 10$ values. As a result, the findings revealed no collinearity issues among the ten landslide factors.

Moreover, as illustrated in Fig. 4 and detailed in Table 3, all the landslide conditioning factors have been verified as significant for Penang Island. This aligns with findings from LSM in the Abha Basin (Youssef & Pourghasemi 2021) and the Muzzafarabad district (Hussain et al. 2022), where both slope and the DEM exhibited greater influence on risk incidence compared to other contributing factors. Consequently, employing the Boruta Model, these ten criteria were selected for the subsequent round of feature selection, reaffirming their significance. Notably, these ten factors have also been included as contributory variables in previous Malaysian studies (Han et al. 2021; Nhu et al. 2020; Pradhan 2013; Shahabi & Hashim 2015).

4.2 Random forest

The optimized RF models utilizing different BO's acquisition functions, RF_BO_UCB, RF_BO_EI and RF_BO_PI yielded the same the Out-of-Bag (OOB) error. By corresponding to hyperparameter values of $mtry = 10$ and $min_n = 1$ with 500 trees and ten

Table 2 Variance inflation factor (VIF) and tolerance (TOL) of the landslide conditioning factors

Variables	Tolerance (TOL)	Variance Inflation Factor (VIF)
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

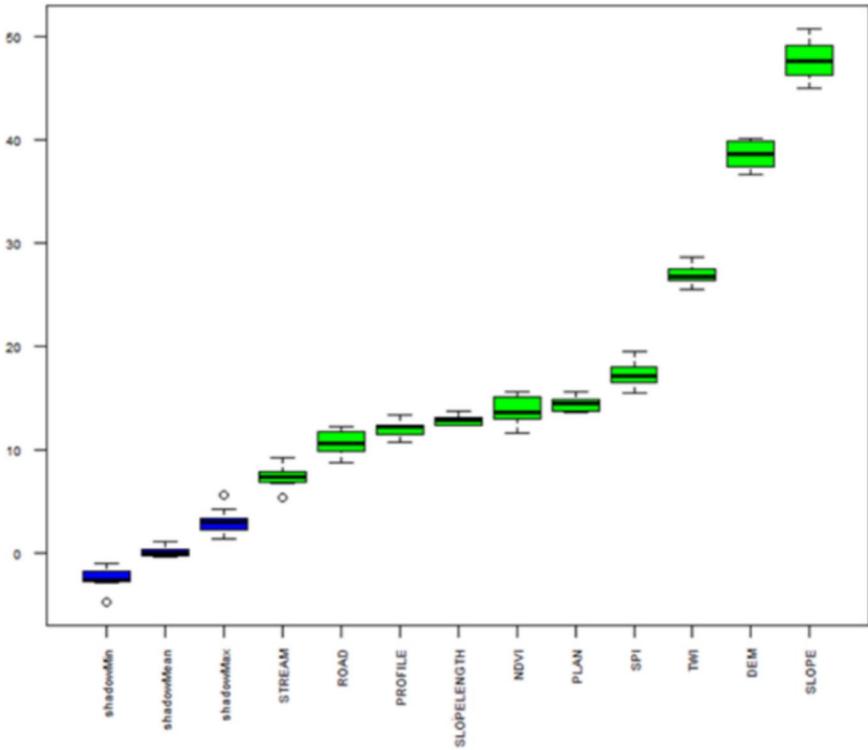


Fig. 4 Boruta model’s chart plot on the importance of variables

Table 3 Feature selection by Boruta model

Landslide conditioning factors	Minimum importance	Maximum importance	Decision
Slope	44.91	50.72	Confirmed
DEM	36.59	40.08	Confirmed
TWI	25.52	28.55	Confirmed
SPI	15.47	19.49	Confirmed
Plan Curvature	13.56	15.58	Confirmed
NDVI	11.56	15.60	Confirmed
Slope Length	12.30	13.73	Confirmed
Profile Curvature	10.69	13.36	Confirmed
Distance from Road	8.77	12.17	Confirmed
Distance from Stream	5.34	9.19	Confirmed

variables considered at each split, the OOB error was 9.26%, resulting in a trained model OOB accuracy of 0.9073726 which is equivalent to 90.74%. Additionally, K-fold cross-validation was applied. The tenfold cross-validation results for the Penang Island optimized RF models exhibited models’ accuracy of 0.901 and a kappa value of 0.802.

Table 4 Statistics for optimized and unoptimized RF prediction on validation datasets

Statistics	RF	Optimized RF		
		RF_BO_UCB	RF_BO_PI	RF_BO_EI
Accuracy	0.9015	0.8906	0.8906	0.8906
Kappa	0.8027	0.781	0.781	0.781
Precision	0.9483	0.8696	0.8696	0.8696
Sensitivity	0.9552	0.9153	0.9153	0.9153
Specificity	0.8462	0.8696	0.8696	0.8696

Table 5 Success rate and Prediction rate of optimized and unoptimized RF

Models	Performance evaluator AUC (%)	
	Success rate	Prediction rate
RF	100.00	96.06
RF_BO_UCB	99.70	96.30
RF_BO_PI	99.70	96.30
RF_BO_EI	99.70	96.30

Meanwhile, for the unoptimized RF model, the cross-validation accuracy and kappa are 0.8965574 and 0.783373 respectively. Besides, the OOB error of the unoptimized RF is 10.18% resulting in OOB accuracy of 89.82% which is slightly lower than the optimized RF. These results signify the RF model's strong predictive accuracy. The sensitivity, specificity and precision values of the optimized RF and unoptimized RF models are in the range between 0.8–1.0 indicating that the model are robust in prediction effect of the landslides samples.

Tables 4 present statistical findings from the prediction of both unoptimized and optimized RF models on the validation datasets. Regardless of utilizing different function acquisition in the RF_BO_UCB, RF_BO_PI and RF_BO_EI the same results for Success Rate (SR) and Prediction Rate (PR). Notably, the difference between the training dataset's AUC value, representing the SR, and the testing dataset's AUC value, representing the PR, was a mere 3.38% meanwhile the difference of SR and PR of the unoptimized RF model was 3.94% (Table 5). Therefore, it can be observed that the optimization process had reduced as much as 0.56% of the overfitting rate.

The identical results for the performance of RF_BO_UCB, RF_BO_EI and RF_BO_PI suggest that the RF model's performance is driven more by the dataset characteristics and RF's intrinsic robustness than by differences in acquisition functions. The optimization landscape for these hyperparameters is likely smooth and well-behaved, leading to convergence on the same values regardless of the acquisition function (Probst et al. 2019).

As a conclusion, both optimized and unoptimized RF models had exhibited great performance in predicting the landslide occurrence. Somehow, based on overall result including the excellent cross-validation accuracy, OOB accuracy, the AUC and the minimal overfitting rate the optimized RF model will be chosen to produce the LSM for this study.

Table 6 Statistics for optimized and unoptimized XGBoost prediction on validation datasets

Statistics	XGBoost	Optimized XGBoost		
		XGBoost_BO_UCB	XGBoost_BO_PI	XGBoost_BO_EI
Accuracy	0.9474	0.9141	0.9242	0.9167
Kappa	0.8939	0.8277	0.8486	0.8335
Precision	0.9531	0.8986	0.8986	0.8857
Sensitivity	0.9552	0.9322	0.8955	0.8806
Specificity	0.9385	0.8986	0.9538	0.9538

Table 7 Success rate and prediction rate of optimized and unoptimized XGBoost

Models	Performance evaluator AUC (%)	
	Success rate	Prediction rate
XGBoost	100.00	95.01
XGBoost_BO_UCB	100.00	97.10
XGBoost_BO_PI	100.00	96.90
XGBoost_BO_EI	99.00	96.57

4.3 Extreme gradient boosting

The XGBoost models' hyperparameters were also optimized by using BO of different acquisition functions which produced three different models of optimized XGBoost mainly, XGBoost_BO_UCB, XGBoost_BO_PI and XGBoost_BO_EI. Besides, the XGBoost models were trained by 10 K-fold cross-validation strategy. The optimized XGBoost models were then compared to the untuned XGBoost model which only incorporate default value of hyperparameters provided by the *Caret Package*.

The results of cross-validated resampling revealed an accuracy of 0.9451, 0.8925, 0.8656 and 0.8657 for the XGBoost, XGBoost_BO_UCB, XGBoost_BO_PI and XGBoost_BO_EI. Meanwhile, the cross-validation kappa value of XGBoost, XGBoost_BO_UCB, XGBoost_BO_PI and XGBoost_BO_EI are 0.8091, 0.7850, 0.7131 and 0.73134 respectively.

Detailed information, including the statistical metrics for the models' prediction on the validation datasets, can be observed in Table 6. The optimized and unoptimized XGBoost models' sensitivity, specificity, and accuracy values fall between 0.85 and 1.00, suggesting that the models are reliable for predicting the impact of landslides events. Notably, the AUC curve for the SR surpassed the PR by 4.99% for the unoptimized XGBoost model. Meanwhile the optimized XGBoost models had minimized the overfitting range as much as 1.89% to 2.56% from the unoptimized XGBoost model as indicated in Table 7. This discovery effectively highlights how Bayesian hyperparameter optimization mitigates the risk of overfitting in XGBoost models. In conclusion, XGBoost models, both optimized and unoptimized, performed exceptionally well in forecasting the occurrence of landslides. In some way, the XGBoost_BO_UCB model will be selected as the optimized XGBoost model to generate the LSM for this investigation based on the overall results, which include the outstanding cross-validation accuracy, the AUC, and the small overfitting rate.

Table 8 Statistics of stacked model

Model	Statistics	Scores
Stacked XGBoost-RF	Accuracy	0.8693
	Kappa	0.7384
	Precision	0.9000
	Sensitivity	0.8434
	Specificity	0.9000

Table 9 Success rate and prediction rate for stacked model

Model	Performance evaluator AUC (%)	
	Success rate	Prediction rate
Stacked XGBoost-RF	96.8	95.6

The UCB acquisition function performs better than the PI and EI because UCB explicitly controls the trade-off between exploration and exploitation which makes it well-suited for complex, multimodal hyperparameter landscapes like XGBoost. This flexibility allows UCB to explore less-certain regions while still exploiting promising areas, leading to better optimization results (Srinivas et al. 2012). In contrast with RF, XGBoost is more sensitive towards the tuning of its hyperparameters. This causing small changes in its hyperparameters will lead to significant change in its performance.

4.4 Stacked model

The best optimized models of XGBoost and RF were stacked together by using generalized linear model (glm) as the metaclassifier. As the result, the cross-validation resample accuracy and kappa result of the trained stacked model for Penang Island had an accuracy of 0.9063 and kappa of 0.8106. The statistics of the stacked model for validation datasets are presented in Table 8. The stacked model had exhibited extremely good precision value, 0.9 which is equivalent to 90% of the time the model able to correctly predict the landslide points. Notably, the PR AUC exceeded the SR prediction by a mere 1.17%, as indicated in Table 9. This marginal difference signifies effective mitigation of overfitting through the stacking approach.

4.4.1 Correlational analysis with frequency ratio

Table 10 illustrates the FR model results, which assesses the relationship between landslide occurrence and its influencing factors. The FR value serves as an indicator of the overall spatial correlation between the predictor variable and landslides. In the case of Penang Island, the FR analysis revealed a negative correlation between landslide occurrence and elevation, with landslides being most frequent at elevations between 78 and 200 m above sea level. This phenomenon can be attributed to increased precipitation, lower temperatures, and accelerated weathering as elevation rises, all promoting landslide development (Gruber & Haerberli 2007).

On Penang Island, curvatures, particularly profile curvature, exhibited a substantial association with landslides, with landslides occurring in the range of 0.708 to 5.793 for

Table 10 Frequency ratio

Factors	Classes	No. points	% of points	Class area	% of class area	Ratio	FR
Slope	15.898–24.08	42,300	16.10	33,877	10.11	1.59	0.25
	24.08–59.616	16,200	6.16	13,115	3.92	1.57	0.24
	9.352–15.898	76,500	29.11	62,893	18.77	1.55	0.24
	3.741–9.352	88,200	33.56	78,585	23.46	1.43	0.22
	0–3.741	39,600	15.07	146,521	43.74	0.34	0.05
DEM	515.974–797.414	1800	0.68	20,722	6.19	0.11	0.02
	343.983–515.974	10,800	4.11	27,241	8.13	0.51	0.10
	0–78.178	85,500	32.53	177,087	52.86	0.62	0.13
	200.135–343.983	50,400	19.18	46,007	13.73	1.40	0.28
TWI	78.178–200.135	114,300	43.49	63,934	19.09	2.28	0.46
	13.277–25.294	1800	0.68	10,715	3.20	0.21	0.05
	10.338–13.277	13,500	5.14	45,235	13.50	0.38	0.10
	8.350–10.338	39,600	15.07	83,464	24.92	0.60	0.15
SPI	6.534–8.350	116,100	44.18	112,494	33.58	1.32	0.34
	3.249–6.534	91,800	34.93	83,083	24.80	1.41	0.36
	21,013.298–42,819.551	900	0.34	146	0.04	7.86	0.78
	1585.909–7533.069	3600	1.37	3672	1.10	1.25	0.12
	0–1585.909	258,300	98.29	330,504	98.66	1.00	0.10
Plan	7533.069–21,013.298	0	0.00	626	0.19	0.00	0.00
	42,819.551–101,101.718	0	0.00	43	0.01	0.00	0.00
Curvature	0.538–3.683	12,600	4.79	6749	2.01	2.38	0.28
	–3.290 to –0.555	9900	3.77	5437	1.62	2.32	0.28
	–0.555 to –0.145	37,800	14.38	30,241	9.03	1.59	0.19
	0.128–0.538	37,800	14.38	38,641	11.53	1.25	0.15
	–0.145 to 0.128	164,700	62.67	253,923	75.80	0.83	0.10
NDVI	0.404 to 0.586	139,500	53.08	127,634	38.13	1.39	0.39
	0.292–0.404	75,600	28.77	83,301	24.89	1.16	0.32
	0.010–0.168	25,200	9.59	63,241	18.89	0.51	0.14
	0.168–0.292	22,500	8.56	56,527	16.89	0.51	0.14
	0	0	0.00	4020	1.20	0.00	0.00
Slope Length	6.324–15.085	8100	3.08	58 12	1.73	1.78	0.33
	15.085–32.603	1800	0.68	1413	0.42	1.62	0.30
	1.460–6.326	19,800	7.53	23,877	7.13	1.06	0.19
	0–1.460	233,100	88.70	303,746	90.67	0.98	0.18
	32.603–124.087	0	0.00	143	0.04	0.00	0.00
Profile Curvature	–0.251 to 0.133	152,100	57.88	247,240	73.80	0.78	0.10
	–6.439 to –0.827	4500	1.71	4127	1.23	1.39	0.18
	0.1328–0.708	63,900	24.32	53,357	15.93	1.53	0.20
	–0.827 to –0.251	28,800	10.96	22,802	6.81	1.61	0.21
	0.709–5.793	13,500	5.14	7465	2.23	2.31	0.30

Table 10 (continued)

Factors	Classes	No. points	% of points	Class area	% of class area	Ratio	FR
Distance from Road	589.740–1,253.1960	12,600	4.79	6749	2.01	2.38	0.28
	0–68.803	9900	3.77	5437	1.62	2.32	0.28
	68.803–191.666	37,800	14.38	30,241	9.03	1.59	0.19
	353.844–589.740	37,800	14.38	38,641	11.53	1.25	0.15
	191.665–353.844	164,700	62.67	253,923	75.80	0.83	0.10
Distance from Stream	296.406–427.803	44,100	16.78	43,467	12.98	1.29	0.24
	427.803–779.213	16,200	6.16	16,056	4.79	1.29	0.24
	189.456–296.406	51,300	19.52	65,177	19.46	1.00	0.19
	91.672–189.456	63,000	23.97	87,419	26.10	0.92	0.17
	0–91.672	88,200	33.56	122,872	36.68	0.92	0.17

profile curvature and 0.328 to -0.555 and 0.538 to 3.683 for plan curvature. Both profile and plan curvature influence slope erosion by regulating water flow velocity (Aghdam et al. 2016).

Landslides were discovered to be more likely on slopes exceeding 10°, with moderately steep slopes (15°–25°) being the most susceptible, possibly indicating a threshold range (Nakileza & Nedala 2020). Additionally, in the western highlands of Uganda, Nseka et al. (2019) identified additional landslides on slopes ranging from 25° to 35°. Bizimana & Sonmez (2015) suggested that areas with slope angles greater than 14° on convex slopes and more than 41° on concave slopes are most prone to landslides.

The SPI value exhibited the strongest correlation with landslides in the range of 21,013.3 to 42,819.6, while TWI demonstrated significant associations only within the range of 3.2 to 6.5 for Penang Island. Moreover, the data from FR revealed that the correlation between distance from the road and landslides decreases as the distance increases, consistent with the findings of Lei et al. (2020). Although the correlation value is lower in the range of 68.8 m to 589.7 m, it was observed that the correlation value is high in the range of 589.74 m to 1,253.2 m on Penang Island, in alignment with research indicating that most landslides occur within a 1000-m radius of a road (Wang et al. 2022).

Conversely, the correlation between factors and landslide occurrence on Penang Island increases as the distance from the stream grows. The maximum distance from the stream on the island is 427.8 m, which is still considered too close to the stream. Proximity to the river enhances slope base erosion strength, impacting slope stability (Yalcin 2008). Lastly, the FR analysis revealed that, for Penang Island, NDVI values had the most substantial positive association with landslide occurrence, falling within the range of 0.4 to 0.5

4.5 Model comparisons

The performance evaluation of the models on both validation and testing datasets clearly demonstrates that the optimized XGBoost model outperforms both the optimized RF model and the stacked model consisting of both optimized RF and XGBoost. In terms of overall accuracy, the optimized XGBoost model consistently achieved the highest accuracy rate of 0.9141, followed by optimized RF with 0.8906 and the stacked model with 0.8693.

Fig. 5 Success rate AUC of three models

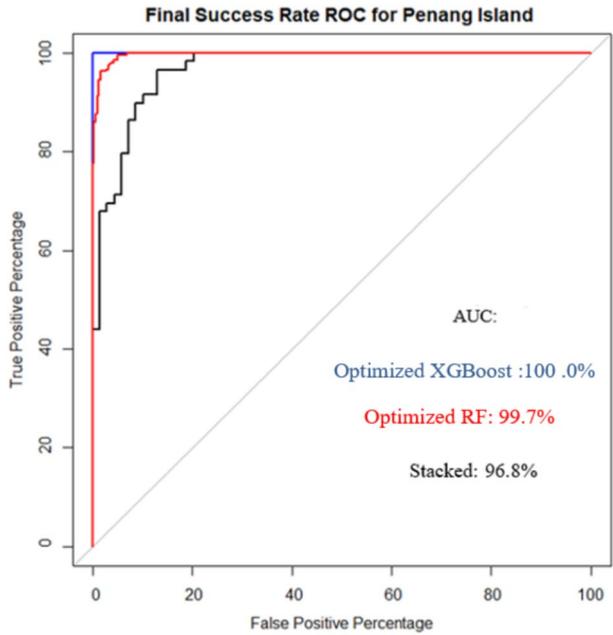
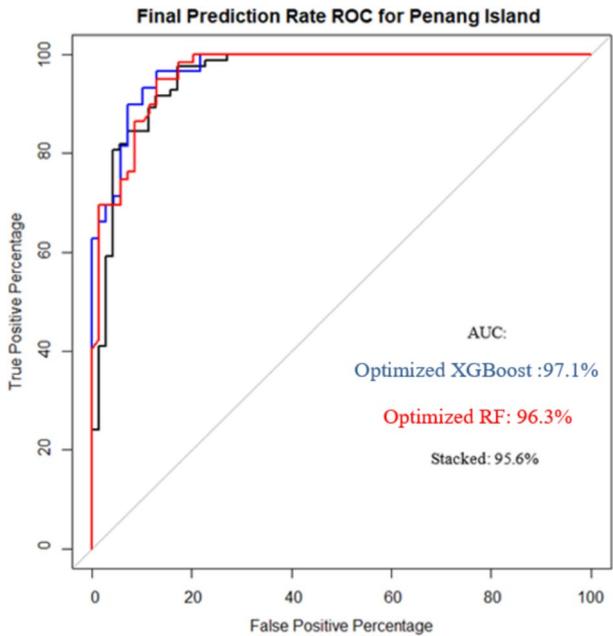


Fig. 6 Prediction rate AUC of three model



For the SR AUC, optimized XGBoost achieved a perfect score of 100%, while optimized RF scored 99.7%, and the stacked model reached 96.8% (Fig. 5). Similarly, in terms of the PR AUC, optimized XGBoost excelled with 97.1%, followed by optimized RF with

96.3%, and the stacked model with 95.6% (Fig. 6). Overall, the final findings had shown that all models exhibited outstanding prediction performance, with accuracy above 0.80 and AUC scores exceeding 90%.

Comparative analysis with previous research by Cao et al. (2020) supports the superiority of the XGBoost method, as it outperformed RF and SVM algorithms with the highest AUC score. Additionally, a study on the Karakorum Highway, Pakistan (Hussain et al. 2022) demonstrated a slight advantage of XGBoost over RF, with only a 0.8% difference in AUC score. While the stacked model performed well, it did not surpass the individual base models. This finding aligns with a study by (Dou et al. 2019), where a stacked SVM exhibited lower performance, highlighting that ensemble machine-learning models may not always yield superior results.

Furthermore, stacking two models may have a limited impact on powerful learners like XGBoost and RF. However, significant improvements can be observed when stacking involves learners with varying performance levels, as demonstrated in a study by Nhu et al. (2020).

Regarding overfitting, the optimized XGBoost and optimized RF models exhibit minimal overfitting compared to the unoptimized models of RF and XGBoost. In comparison, a study on Rangamatti Hill, Bangladesh, by Rabby et al. (2020) achieved reduced overfitting percentages of 4.64% for XGBoost and 4.52% for RF using grid search optimization.

In this study, BO effectively minimized the percentage rates of overfitting in both optimized XGBoost and RF models as shown in Fig. 7. The stacked model of the two optimized models exhibited the lowest overfitting rate at 1.2% compared to the overfitting rate of optimized RF (3.38%) and XGBoost (2.9%). Thus, BO proved to be a successful advanced hyperparameter optimization strategy for XGBoost and RF models.

4.6 Final landslide susceptibility maps

The Janks natural breakpoint approach (Jaafari et al. 2014) was employed to categorize landslide susceptibility into two classes, non-susceptible and susceptible, based on the outcomes of three algorithms. This approach was utilized to produce landslide susceptibility zoning maps for each algorithm. The categorization scores for the three models ranged from 0 to 1, signifying non-susceptible and susceptible zones. Notably, although

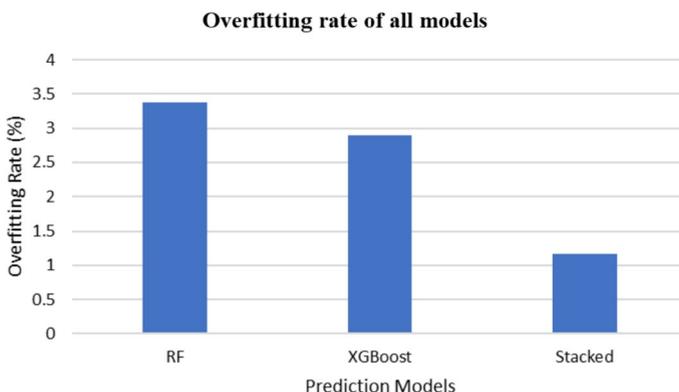


Fig. 7 Overfitting rate of all models

Fig. 8 Penang Island LSM via optimized random forest prediction

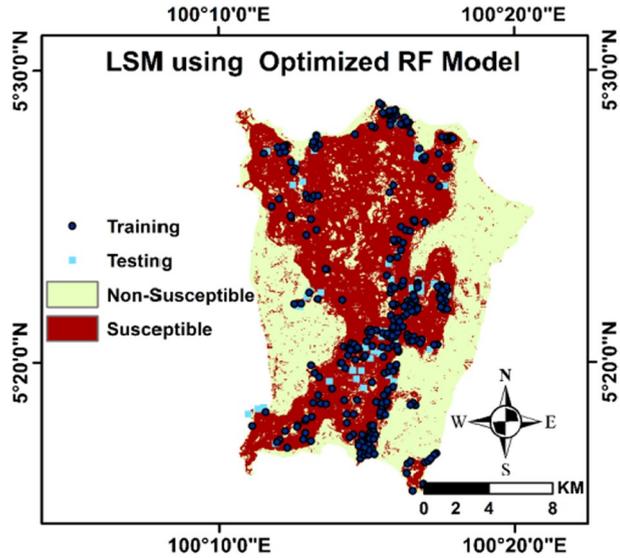
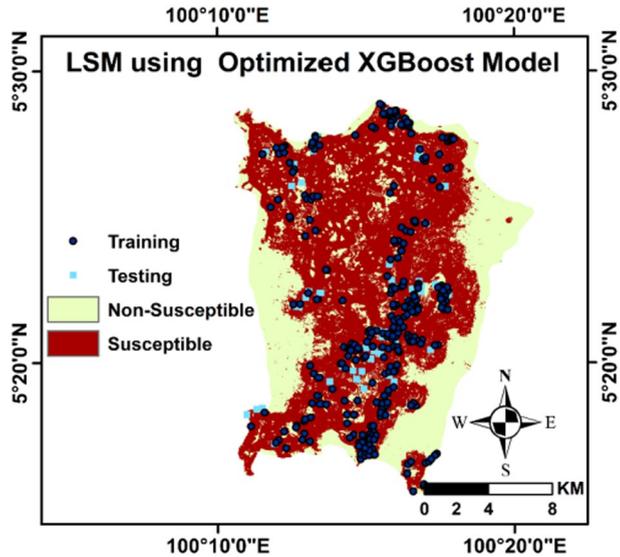


Fig. 9 Penang Island LSM via optimized XGBoost prediction



XGBoost outperformed RF and the stacked models, XGBoost had experienced an over-estimation issue where it overestimates the "non-susceptible" area, as shown in map Fig. 9 compared to the prediction map produced by RF in Fig. 8 and the stacked model in Fig. 10. This aligned with the findings of Rabby et al.(2020), which indicated that the XGBoost model tends to overestimate susceptibility. Interestingly, the XGBoost model designated as high as 65.89% of the study area as a susceptibilitie zone meanwhile the stacked and RF models only classify 56.48% and 56.00% respectively as susceptible zones (Fig. 11). This portrayed an overestimation issue in XGBoost model. According to Abedini & Tulabi (2018), the smallest territory should typically be allocated to

Fig. 10 Penang Island LSM via RF-XGBoost stacked model prediction

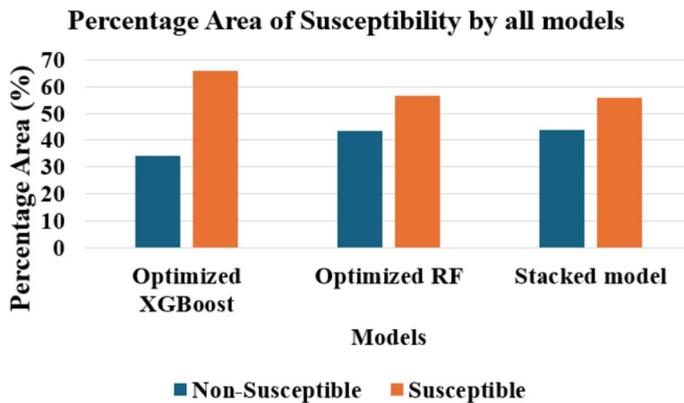
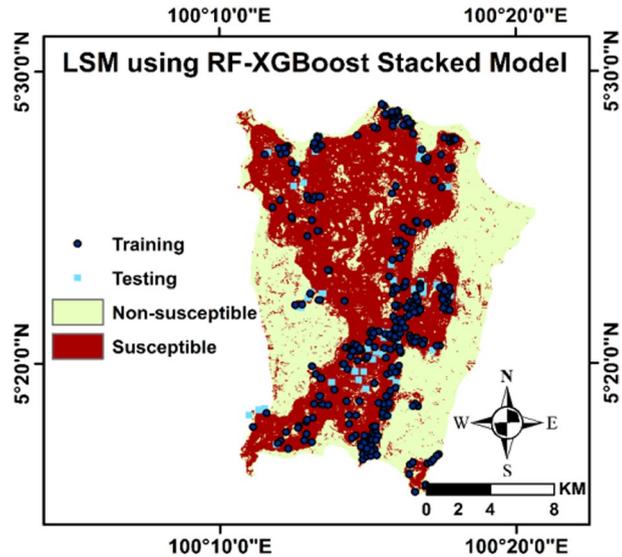


Fig. 11 Percentage of susceptibility areas covered by LSM generated by all models

the zone with the highest susceptibility. Overestimate zoning, even if the susceptibility model exhibits good accuracy, can diminish its practical relevance (Reichenbach et al. 2018).

The stacking of both optimized RF and XGBoost models had managed to overcome the overestimation issues in XGBoost as the percentage difference in the classification of both susceptible and non-susceptible classes between the RF and stacked model are less than 1%. This was achieved by sustaining the true and false positives of both models in check, thus averting overestimation.

Additionally, this finding had been justified by Wang et al. (2020) who discovered in their study that the overestimation issue could be mitigated by stacking two models. This illustrates that the overestimation problem can be minimized through the stacking strategy and the use of optimized models. While optimized XGBoost outperformed other models in this study, it allocated a relatively high percentage of the study area as a susceptible zone, indicating potential

overestimation. Interestingly, the stacking strategy, where two models were combined, slightly reduced the overestimation observed in XGBoost.

5 Conclusion

The results of our experiments offer valuable insights into the prediction efficiency and accuracy of our proposed model, as assessed through various performance measures. We have identified essential internal landslide conditioning factors, including DEM, slope angle, slope length, profile curvature, plan curvature, SPI, TWI, distance from roads, and proximity to streams, based on Boruta and FR analyses.

Among the models evaluated, optimized XGBoost consistently outperformed optimized RF and the stacked model. While all models demonstrated outstanding predictive capabilities, optimized XGBoost achieved the highest overall accuracy, followed by optimized RF and the stacked model, with only slight differences in AUC scores. Furthermore, BO proved to be an effective hyperparameter optimization method, enhancing the performance of both optimized RF and XGBoost models. Moreover, BO, coupled with stacking, effectively reduced the occurrence of overfitting and overestimation.

Our results underscore the potential for further research in this direction. To enhance prediction accuracy, we recommend the development of more high-resolution landslide conditioning spatial databases. Additionally, exploring the integration of other machine learning models to create optimal hybrid models presents a promising avenue for future investigations.

In conclusion, our study demonstrates significant advancements in predictive modeling for landslide susceptibility assessment. By optimizing model performance and mitigating overfitting and overestimation, we have contributed to the evolving field of landslide risk assessment. However, further research and resource allocation is essential to unlock the full potential of this approach and continue improving landslide prediction accuracy.

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

Author contributions Both authors contributed to the study of conception and design. Material preparation, data collection and analysis were performed by DAMA. The first draft of the manuscript was written by DAMA and co-responding authors SSC commented on previous versions of the manuscript. Both authors read and approved the final manuscript.

Funding Open access funding provided by The Ministry of Higher Education Malaysia and Universiti Malaysia Sarawak. The authors acknowledge The Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS/1/2022/ICT02/UNIMAS/02/2) and Universiti Malaysia Sarawak for supporting this project.

Data Availability Data is available upon request from the authors.

Declarations

Conflict of interest The authors declared no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The

images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Abedini M, Tulabi S (2018) Assessing LNRF, FR, and AHP models in landslide susceptibility mapping index: a comparative study of Nojian watershed in Lorestan province, Iran. *Environ Earth Sci* 77:1–13
- Achour Y, Reza H (2020) Geoscience Frontiers How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geosci Front* 11(3):871–883. <https://doi.org/10.1016/j.gsf.2019.10.001>
- Adnan MSG, Rahman MS, Ahmed N, Ahmed B, Rabbi MF, Rahman RM (2020a) Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Remote Sens* 12(20):1–23. <https://doi.org/10.3390/rs12203347>
- Adnan MSG, Talchabhadel R, Nakagawa H, Hall JW (2020) The potential of Tidal River Management for flood alleviation in South Western Bangladesh. In: *Science of the total environment*. Elsevier, Amsterdam, vol 731 <https://doi.org/10.1016/j.scitotenv.2020.138747>
- Ado M, Amitab K, Maji AK, Jasińska E, Gono R, Leonowicz Z, Jasiński M (2022) Landslide susceptibility mapping using machine learning: a literature survey. In: *Remote sensing*, vol 14, Issue 13. MDPI. <https://doi.org/10.3390/rs14133029>
- Aghdam IN, Varzandeh MHM, Pradhan B (2016) Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). *Environ Earth Sci*. <https://doi.org/10.1007/s12665-015-5233-6>
- Agrawal N, Dixit J (2023) GIS-based landslide susceptibility mapping of the Meghalaya-Shillong Plateau region using machine learning algorithms. *Bull Eng Geol Environ* 82(5):170. <https://doi.org/10.1007/s10064-023-03188-2>
- Ahmed B, Rahman MS, Sammonds P, Islam R, Uddin K (2020) Application of geospatial technologies in developing a dynamic landslide early warning system in a humanitarian context: the Rohingya refugee crisis in Cox's Bazar, Bangladesh. *Geomat Nat Haz Risk* 11(1):446–468. <https://doi.org/10.1080/19475705.2020.1730988>
- Amiri M, Pourghasemi HR, Ghanbarian GA, Afzali SF (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(1):55–69. <https://doi.org/10.1016/j.geoderma.2018.12.042>
- Ao SI (2008) Data mining and applications in genomics. In: *Lecture notes in electrical engineering*: vol 25 LNEE
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda–Yahiko Mountains Central Japan. *Geomorphology* 65(1–2):15–31. <https://doi.org/10.1016/j.geomorph.2004.06.010>
- Bizimana H, Sonmez O (2015) Landslide Occurrences in the Hilly Areas of Rwanda, their causes and protection measures
- Bui DT, Shahabi H, Shirzadi A, Chapi K, Alizadeh M, Chen W, Mohammadi A, Ahmad B, Panahi M, Hong H, Tian Y (2018) Landslide detection and susceptibility mapping by AIRSAR data using support vector machine and index of entropy models in Cameron Highlands, Malaysia. *Remote Sens* 10(10):1527. <https://doi.org/10.3390/rs10101527>
- Can R, Kocaman S, Gokceoglu C (2021) A comprehensive assessment of XGBoost algorithm for landslide susceptibility mapping in the upper basin of Ataturk dam, Turkey. *Appl Sci*. <https://doi.org/10.3390/app1114993>
- 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. *Nat Hazards* 102:851–871
- Chen W, Li Y (2020) GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *CATENA* 195(June):104777. <https://doi.org/10.1016/j.catena.2020.104777>
- Chen W, Panahi M, Pourghasemi HR (2017a) Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *CATENA* 157:310–324. <https://doi.org/10.1016/j.catena.2017.05.034>

- Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017b) 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>
- 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 BB (2018a) GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci Total Environ* 634:853–867. <https://doi.org/10.1016/j.scitotenv.2018.04.055>
- Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J, Zhu AX, Pei X, Duan Z (2018b) Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci Total Environ* 626:1121–1135. <https://doi.org/10.1016/j.scitotenv.2018.01.124>
- Chen W, Zhang S, Li R, Shahabi H (2018c) Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Sci Total Environ* 644:1006–1018. <https://doi.org/10.1016/j.scitotenv.2018.06.389>
- Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, 13–17-Aug, pp 785–794. <https://doi.org/10.1145/2939672.2939785>
- Das S, Bora PK, Das R (2022) Estimation of slope length gradient (LS) factor for the sub-watershed areas of Juri River in Tripura. *Model Earth Syst Environ* 8(1):1171–1177. <https://doi.org/10.1007/s40808-021-01153-0>
- Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Dhital MR, Althuwaynee OF (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. *Nat Hazards* 65(1):135–165. <https://doi.org/10.1007/s11069-012-0347-6>
- Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z, Chen CW, Han Z, Pham BT (2019) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* 17(3):641–658. <https://doi.org/10.1007/s10346-019-01286-5>
- Duman TY, Can T, Gokceoglu C, Nefeslioglu HA, Sonmez H (2006) Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey. *Environ Geol* 51(2):241–256. <https://doi.org/10.1007/s00254-006-0322-1>
- Duval F, Boucher J-P, Pigeon M (2022) How much telematics information do insurers need for claim classification? *N Am Actuar J* 26(4):570–590
- Fang Z, Wang Y, Peng L, Hong H (2020) A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *Int J Geogr Inf Sci* 35(2):321–347. <https://doi.org/10.1080/13658816.2020.1808897>
- Gao H, Fam PS, Tay LT, Low HC (2020) Three oversampling methods applied in a comparative landslide spatial research in Penang Island, Malaysia. *SN Appl Sci*. <https://doi.org/10.1007/s42452-020-03307-8>
- Gómez H, Kavzoglu T (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Eng Geol* 78(1–2):11–27. <https://doi.org/10.1016/j.enggeo.2004.10.004>
- Gruber S, Haerberli W (2007) Permafrost in steep bedrock slopes and its temperature-related destabilization following climate change. *J Geophys Res: Earth Surf*. <https://doi.org/10.1029/2006JF000547>
- Han G, Shan FP, Tien TL, Chin LH (2021) Landslide susceptibility analysis using gradient boosting models: a case study in Penang Island, Malaysia. *Disaster Adv* 14(8):22–37. <https://doi.org/10.25303/148da2221>
- He Q, Shahabi H, Shirzadi A, Li S, Chen W, Wang N, Chai H, Bian H, Ma J, Chen Y, Wang X, Chapi K, Ahmad BB (2019) Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *Sci Total Environ* 663:1–15. <https://doi.org/10.1016/j.scitotenv.2019.01.329>
- Hebbali A, Hebbali MA (2017) Package ‘olsrr.’ Version 0.5, 3
- Huang F, Yin K, Huang J, Gui L, Wang P (2017) Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Eng Geol* 223:11–22. <https://doi.org/10.1016/j.enggeo.2017.04.013>
- Huang F, Cao Z, Guo J, Jiang SH, Li S, Guo Z (2020) Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *CATENA* 191(March):104580. <https://doi.org/10.1016/j.catena.2020.104580>
- Hussain MA, Chen Z, Wang R, Shoaib M (2021) Ps-insar-based validated landslide susceptibility mapping along Karakorum highway, Pakistan. *Remote Sens* 13(20):4129. <https://doi.org/10.3390/rs13204129>

- Hussain MA, Chen Z, Zheng Y, Shoaib M, Shah SU, 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 HR, Rezaeian J, Sattarian A (2014) GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *Int J Environ Sci Technol* 11(4):909–926. <https://doi.org/10.1007/s13762-013-0464-0>
- Kursa MB, Rudnicki WR (2010) Feature selection with the boruta package. *J Stat Softw* 36(11):1–13. <https://doi.org/10.18637/jss.v036.i11>
- Kushner HJ (1964) A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J Basic Eng* 86:97–106
- Le Chan JY, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong ZW, Chen YL (2022) Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics* 10(8):1283. <https://doi.org/10.3390/math10081283>
- Lei X, Chen W, Pham BT (2020) Performance evaluation of GIS-based artificial intelligence approaches for landslide susceptibility modeling and spatial patterns analysis. *ISPRS Int J Geo Inf* 9(7):1–20. <https://doi.org/10.3390/ijgi9070443>
- Moayedi H, Mehrabi M, Mosallanezhad M, Rashid ASA, Pradhan B (2019) Modification of landslide susceptibility mapping using optimized PSO-ANN technique. *Eng Comput* 35(3):967–984. <https://doi.org/10.1007/s00366-018-0644-0>
- Mutlu B, Nefeslioglu HA, Sezer EA, Ali Akcayol M, Gokceoglu C (2019) An experimental research on the use of recurrent neural networks in landslide susceptibility mapping. *ISPRS Int J Geo Inf* 8(12):1–21. <https://doi.org/10.3390/ijgi8120578>
- Nakileza BR, Nedala S (2020) Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda. *Geoenviron Disasters* 7(1):27. <https://doi.org/10.1186/s40677-020-00160-0>
- Nhu VH, Mohammadi A, Shahabi H, Ahmad BB, Al-Ansari N, Shirzadi A, Clague JJ, 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):1–23. <https://doi.org/10.3390/ijerph17144933>
- Nseka D, Kakembo V, Bamutaze Y, Mugagga F (2019) Analysis of topographic parameters underpinning landslide occurrence in Kigezi highlands of southwestern Uganda. *Nat Hazards* 99(2):973–989. <https://doi.org/10.1007/s11069-019-03787-x>
- O'brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 41(5):673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Paoletti ME, Haut JM, Plaza J, Plaza A (2018) A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J Photogramm Remote Sens* 145:120–147. <https://doi.org/10.1016/j.isprsjsprs.2017.11.021>
- Pham BT, Prakash I, Dou J, Singh SK, Trinh PT, Tran HT, Le TM, Van Phong T, Khoi DK, Shirzadi A, Bui DT (2020) A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int* 35(12):1267–1292. <https://doi.org/10.1080/10106049.2018.1559885>
- Pirasteh S, Li J (2017) Global changes and natural disaster management: Geo-information technologies. *Global Changes and Natural Disaster Management: Geo-Inform Technol March*, pp 1–228. <https://doi.org/10.1007/978-3-319-51844-2>
- Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Nat Hazards* 63(2):965–996. <https://doi.org/10.1007/s11069-012-0217-2>
- 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. *Comput Geosci* 51:350–365. <https://doi.org/10.1016/j.cageo.2012.08.023>
- Pradhan B, Sezer EA, Gokceoglu C, Buchroithner MF (2010) Landslide susceptibility mapping by neuro-fuzzy approach in a landslide-prone area (Cameron Highlands, Malaysia). *IEEE Trans Geosci Remote Sens* 48(12):4164–4177. <https://doi.org/10.1109/TGRS.2010.2050328>
- Probst P, Wright MN, Boulesteix A (2019) Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip Rev Data Min Knowl Discov* 9(3):e1301
- Rabby YW, Li Y (2020) Landslide susceptibility mapping using integrated methods: a case study in the chittagong hilly areas, bangladesh. *Geosciences (Switzerland)* 10(12):1–26. <https://doi.org/10.3390/geosciences10120483>

- Rabby YW, Hossain MB, Abedin J (2020) Landslide susceptibility mapping in three Upazilas of Rangamati hill district Bangladesh: application and comparison of GIS-based machine learning methods. *Geocarto Int.* <https://doi.org/10.1080/10106049.2020.1864026>
- Rahman HA, Mapjabil J (2017) Haliza & Jabil. *Health Environ J* 8(1):58–71
- Rana S, Li C, Gupta S, Nguyen V, Venkatesh S (2017) High dimensional Bayesian optimization with elastic Gaussian process. In: International conference on machine learning, pp 2883–2891
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180(March):60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Roslee, R., Jamaluddin, T. A., & Talip, M. A. (2012). Landslide Susceptibility Mapping (LSM) at Kota Kinabalu, Sabah, Malaysia using Factor Analysis Model (FAM). *J Adv Sci Eng Res*, 2(June 2015), pp 80–103. <https://www.sign-ific-ance.co.uk/dsr/index.php/JASER/article/view/134>
- Rozos D, Bathrellos GD, Skilodimou HD (2011) Comparison of the implementation of rock engineering system and analytic hierarchy process methods, upon landslide susceptibility mapping, using GIS: a case study from the Eastern Achaia County of Peloponnesus, Greece. *Environ Earth Sci* 63(1):49–63. <https://doi.org/10.1007/s12665-010-0687-z>
- 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. *Artif Intell Geosci* 3(March):14–27. <https://doi.org/10.1016/j.aiig.2022.06.002>
- Saha S, Roy J, Pradhan B, Hembram TK (2021) Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at East Sikkim Himalayan, India. *Adv Space Res* 68(7):2819–2840. <https://doi.org/10.1016/j.asr.2021.05.018>
- Sahin EK (2022) Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. *Geocarto Int* 37(9):2441–2465. <https://doi.org/10.1080/10106049.2020.1831623>
- Sameen MI, 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>
- Senouci R, Taibi NE, Teodoro AC, Duarte L, Mansour H, Meddah RY (2021) Gis-based expert knowledge for landslide susceptibility mapping (LSM): Case of Mostaganem coast district, west of Algeria. *Sustainability (Switzerland)* 13(2):1–21. <https://doi.org/10.3390/su13020630>
- Shahabi H, Hashim M (2015) Landslide susceptibility mapping using GIS-based statistical models and remote sensing data in tropical environment. *Sci Rep* 5:1–15. <https://doi.org/10.1038/srep09899>
- Sidle RC, Ochiai H (2013) Landslides : processes , prediction , and land use processes , prediction , and land use ro y c . Sidle American Geophysical Union. June 2019. <https://doi.org/10.1029/WM018>
- Srinivas N, Krause A, Kakade SM, Seeger MW (2012) Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Trans Inf Theory* 58(5):3250–3265
- 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>
- Sun D, Xu J, Wen H, Wang D (2021) Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng Geol* 281:105972. <https://doi.org/10.1016/j.enggeo.2020.105972>
- Susan S, Kumar A, Jain A (2021) Evaluating heterogeneous ensembles with boosting meta-learner. *Lect Notes Netw Syst* 145(1):699–710. https://doi.org/10.1007/978-981-15-7345-3_60
- Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* 2(2):159–178. <https://doi.org/10.1080/20964471.2018.1472392>
- Tien Bui D, Ho T-C, Pradhan B, Pham B-T, Nhu V-H, Revhaug I (2016) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environ Earth Sci* 75(14):1101. <https://doi.org/10.1007/s12665-016-5919-4>
- 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. *Front Earth Sci.* <https://doi.org/10.3389/feart.2021.712240>
- Wang Z, Ma C, Qiu Y, Xiong H, Li M (2022) Refined zoning of landslide susceptibility: a case study in Enshi County, Hubei, China. *Int J Environ Res Public Health* 19(15):9812. <https://doi.org/10.3390/ijerph19159412>
- Wang Z, Liu Q, Liu Y (2020). SS symmetry. *Mdpi*, 1–19
- Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H (2019) Hyperparameter optimization for machine learning models based on Bayesian optimization. *J Electron Sci Technol* 17(1):26–40. <https://doi.org/10.11989/JEST.1674-862X.80904120>

- 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>
- Yalcin A (2008) GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. *CATENA* 72(1):1–12. <https://doi.org/10.1016/j.catena.2007.01.003>
- Yang C, Liu L-L, Huang F, Huang L, Wang X-M (2023) Machine learning-based landslide susceptibility assessment with optimized ratio of landslide to non-landslide samples. *Gondwana Res* 123:198–216
- Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ Earth Sci* 61(4):821–836. <https://doi.org/10.1007/s12665-009-0394-9>
- Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region. *Saudi Arabia Geosci Fron* 12(2):639–655. <https://doi.org/10.1016/j.gsf.2020.05.010>
- Zhu XX, Tuia D, Mou L, Xia G-S, Zhang L, Xu F, Fraundorfer F (2017) Deep learning in remote sensing: a review. *IEEE Geosci Remote Sens Mag* 5:8–36. <https://doi.org/10.1109/MGRS.2017.2762307>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.