



Faculty of Computer Science and Information Technology

**Extreme Gradient Boosting (XGBoost) and Random Forest (RF) Hybrid
Ensemble with Bayesian Optimization in Landslide Susceptibility Mapping**

Dorothy Anak Martin Atok

**Master of Science
2025**

Extreme Gradient Boosting (XGBoost) and Random Forest (RF) Hybrid Ensemble
with Bayesian Optimization (BO) in Landslide Susceptibility Mapping

Dorothy Anak Martin Atok

A thesis submitted

In fulfilment of the requirements for the degree of Master of Science

(Computer Science)

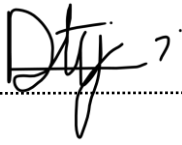
Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK

2025

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



Signature

Name : Dorothy Anak Martin Atok

Matric No. : 21020026

Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak

Date : 11/04/2025

ACKNOWLEDGEMENT

First, I acknowledge the blessings bestowed upon me by the Almighty God that enabled me to be motivated and persistent in this endeavour. Thanks to my amazing parents and family for supporting me financially and morally. I would also like to thank my best friends for constantly boosting my motivation and making me believe I can complete this roller-coaster journey. The research and thesis writing process was not easy, but thanks to my fantastic supervisor, Dr Chai Soo See, who gave me the freedom to explore the subject of my study without any pressure and was always there to guide me. Lastly, thank you to all the Faculty of Computer Science and Information Technology and the Centre of Graduate Studies members who have helped me whenever I have doubts.

Thank You.

ABSTRACT

Many models have been developed in the past for predicting landslide vulnerability. The generated models have certain shortcomings, especially when it comes to the problems of overfitting and overestimation. Therefore, the objective of this study is to assess and enhance the performance of Extreme Gradient Boosting (XGBoost) and Random Forest (RF) in predicting the susceptibility of landslides through the combination of the two models and the use of the Bayesian Optimization (BO) algorithm. The selected study area for this case study is Penang Island, which is a landslide-prone region in Malaysia. The topographical, hydrological, anthropogenic, and external factors that influence landslides were considered in this study. Geographic Information Systems (GIS) have been used to create the spatial databases of all the landslide conditioning elements. With an area under the curve (AUC) of success rate (SR) as high as 100.0% and prediction rate (PR) of 97.1%, the results indicate that optimized XGBoost performed best, followed by random forest (SR:99.3%, PR:95.9%), and stacked models (SR:96.8%, PR:95.6%) The final AUC findings showed that stacking and optimizing the hyperparameters can increase the algorithms' performance accuracy by overcoming the overfitting and overestimation issues. Although there are many applications for landslide susceptibility maps, they are crucial for site selection, engineering structure and disaster avoidance.

Keywords: Bayesian optimization, extreme gradient boosting, geographic information system, landslide susceptibility mapping, random forest

***Penggabungan Extreme Gradient Boosting (XGBoost) dan Random Forest (RF)
dengan implementasi Bayesian Optimization (BO) dalam penghasilan peta kerentanan
tanah runtuh***

ABSTRAK

Banyak model telah dibangunkan pada masa lalu untuk meramalkan kawasan yang berisiko tinggi untuk mengalami tanah runtuh. Model yang dihasilkan mempunyai kelemahan tertentu, terutamanya apabila ia melibatkan masalah ramalan model terlampau sama dengan set data pelatih dan masalah aggaran berlebihan oleh model. Oleh itu, objektif kajian ini adalah untuk menilai dan meningkatkan prestasi Teknik “Extreme Gradient Boosting” (XGBoost) dan “Random Forest” (RF) dalam meramalkan kejadian tanah runtuh melalui gabungan kedua-dua model dan penggunaan “Bayesian optimization” (BO). Lokasi kajian kes adalah di Pulau Pinang, salah satu kawasan yang kerap mengalami tanah runtuh di Malaysia. Faktor topografi, hidrologi, antropogenik dan luaran yang mempengaruhi tanah runtuh telah dipertimbangkan dalam kajian ini. Sistem Maklumat Geografi (GIS) telah digunakan untuk mencipta data spatial bagi semua faktor tanah runtuh. Dengan kawasan di bawah lengkung (AUC) kadar kejayaan (SR) setinggi 100.0% dan kadar ramalan (PR) 97.1%, keputusan menunjukkan bahawa XGBoost yang dioptimumkan menunjukkan prestasi terbaik, diikuti oleh RF (SR:99.3%, PR :95.9%) dan model hibrid (SR:96.8%, PR:95.6%) Penemuan akhir AUC menunjukkan bahawa menyusun dan mengoptimumkan hiperparameter boleh meningkatkan ketepatan prestasi algoritma dengan mengatasi isu ramalan model terlampau sama dengan set data pelatih dan masalah aggaran berlebihan oleh model.

Kata Kunci: *Bayesian optimization, extreme gradient boosting, peta ramalan tanah runtuh, random forest, sistem maklumat geografi*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
<i>ABSTRAK</i>	iv
TABLE OF CONTENTS	v
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xv
CHAPTER 1: INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	4
1.3 Research Question	7
1.4 Research Objectives	7
1.5 Hypothesis	8
1.6 Scope Of Study	9
1.7 Thesis Outline	10
CHAPTER 2: LITERATURE REVIEW	12
2.1 Overview	12

2.2	Fundamentals Of Landslide Susceptibility Mapping	13
2.3	Landslide Conditioning Factors	14
2.3.1	Common Landslide Conditioning Factors	14
2.3.1.1	Digital Elevation Model (Dem)	15
2.3.1.2	Slope Angle	16
2.3.1.3	Slope Length	16
2.3.1.4	Plan And Profile Curvature	17
2.3.1.5	Normalized Difference Vegetation Index	17
2.3.1.6	Distance From Road And Stream	19
2.3.1.7	Stream Power Index, And Topographic Wetness Index	19
2.3.1.8	Selection Criteria And Data Sources	21
2.3.2	The Importance And Challenges In Selecting Landslide Factors	23
2.3.3	Landslide Conditioning Factors Selection Techniques	24
2.3.3.1	Multicollinearity Test	24
2.3.3.2	Boruta Model	25
2.3.4	Frequency Ratio (Fr) Analysis	27
2.3.5	Landslide Inventories	28
2.4	Evolution Of Landslide Susceptibility Mapping Models	29
2.4.1	Qualitative Approaches	30
2.4.2	Semi-Quantitative Approach	31

2.4.3	Quantitative Approaches	31
2.5	Machine Learning Techniques As Landslide Susceptibility Prediction Models	33
2.5.1	Logistic Regression	34
2.5.2	Support Vector Machine	35
2.5.3	Deep Learning	36
2.5.4	K-Nearest Neighbor	37
2.5.5	Ensemble Machine Learnings	38
2.5.5.1	Random Forest	39
2.5.5.2	Extreme Gradient Boosting	46
2.5.5.3	Stacking	49
2.5.6	Summary Of Strengths And Weaknesses Of MLmodels	52
2.6	Hyperparameter Optimization Techniques	53
2.6.1	Manual Search	53
2.6.2	Grid Search	54
2.6.3	Bayesian Optimization	55
2.6.4	Summary Of The Strength And Weakness Of Various Optimization Methods.	56
2.7	Selection Of Landslide Susceptibility Prediction Models	57
2.8	Challenges In Developing Landslide Susceptibility Maps	59
2.8.1	Summary On The Challenges In Developing Lsm And The Proposed Solution	62
2.9	Performance Evaluation Methods	64

2.9.1	Area Under The Curve Of Receiver Operating Characteristics	64
2.9.2	Confusion Matrix And Statistics	66
2.10	Conclusion	68
CHAPTER 3: MODELLING AND IMPLEMENTATION		70
3.1	Overview	70
3.2	Research Layout	70
3.3	Study Area	71
3.4	Landslide Conditioning Factors Spatial Databases	72
3.4.1	Topographical Factors	72
3.4.2	Hydrological Factors	74
3.4.3	Environmental and Anthropogenic Factors	78
3.5	Landslide Inventories Map	82
3.6	Modelling	84
3.6.1	Multicollinearity Test	84
3.6.2	Landslide Conditioning Factors Selection With Boruta	85
3.6.3	Frequency Ratio	87
3.6.4	Comparative Analysis	88
3.6.5	Hyperparameter Optimization	90
3.6.5.1	Selection Of Hyperparameter Optimization Technique	90
3.6.5.2	Bayesian Hyperparameter Optimization	91

3.6.6	Random Forest	94
3.6.6.1	Structure	94
3.6.6.2	The Out-Of-Bag (OOB) Error Estimate	95
3.6.6.3	Hyperparameters	96
3.6.7	Extreme Gradient Boosting	99
3.6.7.1	Structure	99
3.6.7.2	Hyperparameters	100
3.6.8	Stacking	103
3.6.8.1	Meta-Classifer	103
3.6.8.2	Implementation Of Stacking	104
3.7	Final Landslide Susceptibility Map	106
3.8	Conclusion	107
CHAPTER 4: RESULTS AND DISCUSSION		109
4.1	Overview	109
4.2	Feature Selections	109
4.2.1	Multicollinearity Test	110
4.2.2	Landslide Conditioning Factors Selection Via Boruta	111
4.2.3	Frequency Ratio	113
4.3	Modelling	116
4.3.1	Performance Comparison Between XGboost, RF, LR, KNN and SVM	116

4.3.2	Sensitivity Analysis	119
4.3.3	Random Forest with Bayesian Optimization	121
4.3.4	Extreme Gradient Boosting with Bayesian Optimization	123
4.3.5	Stacked Extreme Gradient Boosting and Random Forest	125
4.3.6	Model Comparisons	126
4.4	Final Landslide Susceptibility Maps	133
4.4.1	Practical Implementation of LSM	135
4.5	Conclusion	137
	CHAPTER 5: CONCLUSION AND RECOMMENDATIONS	139
5.1	Findings And Conclusions	139
5.2	Limitations And Recommendations	140
5.3	Summary	142
	REFERENCES	145

LIST OF TABLES

	Page
Table 2.1: Data sources for every LCFs	22
Table 2.2: Past studies investigating effect of LCFs on models' prediction performance.	23
Table 2.3: XGBoost Hyperparameters Listings	47
Table 2.4: Summary of strengths and weaknesses of ML models	52
Table 2.5: Summary of strengths and weaknesses of various hyperparameters optimization methods	57
Table 2.6: Summary of strengths and weaknesses of various hyperparameters optimization methods	63
Table 3.1: Pseudocode for variance inflation factor (VIF) and tolerance (TOL)	85
Table 3.2: Pseudocode for Boruta model	86
Table 3.3: Pseudocode for KNN, LR, SVM, RF and XGBoost	89
Table 4.1: Variance Inflation Factor (VIF) and Tolerance (TOL) of the Landslide Conditioning Factors	111
Table 4.2: Boruta Model's result of importance of variables	112
Table 4.3: Cross-validation and prediction accuracies of all MLmodels	119
Table 4.4: Sensitivity Analysis for RF and XGBoost models	120
Table 4.5: Hyperparameters Bayesian Optimization of Random Forest	122

Table 4.6:	Validation Dataset Confusion Matrix and Statistics for RF	123
Table 4.7:	Success rate and Prediction rate of RF	123
Table 4.8:	Hyperparamters optimal values for XGBoost	124
Table 4.9:	Validation Datasets Confusion Matrix and Statistics of XGBoost	125
Table 4.10:	Success Rate and Prediction Rate for XGBoost	125
Table 4.11:	Testing Datasets Confusion matrix and Statistics of Stacked model	126
Table 4.12:	Success Rate and Prediction Rate for Stacked model	126
Table 4.13:	DeLong Test results on models AUC	129

LIST OF FIGURES

	Page
Figure 1.1: Implementation Process of hazard mapping assessment by Malaysia NSMP (2019-2023)	2
Figure 2.1: LSM Approaches (Shano et al., 2020)	30
Figure 2.2: Random Forest Structure (Mumtaz et al., 2023)	41
Figure 2.3: XGBoost Algorithm Structure	46
Figure 2.4: Stacking algorithm.	50
Figure 2.5: Confusion matrix	66
Figure 3.1: Research Methodology Flowchart	71
Figure 3.2: Penang Island	71
Figure 3.3: Extraction of elevation points from Google Earth Pro	73
Figure 3.4: Exported Elevations shapefile and the attribute table with the elevation values	74
Figure 3.5: (a) Fill layer (b) Flow direction layer (c) Flow accumulation layer	76
Figure 3.6: (a) Stream link (b) Stream order (c) Stream network	76
Figure 3.7: (a) Radian slope layer (b) Tanslope layer (c) Scaled flow accumulation layer	77
Figure 3.8: Open Street Map polyline road data	78
Figure 3.9: Landsat 8 Image from the USGS Earth Explorer	79

Figure 3.10:	Penang Island Landsat 8 image of Band 4 and 5 in ArcMap	79
Figure 3.11:	LCFs (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	81
Figure 3.12:	Grid Unit Extraction via LaGriSu Tool packs process	83
Figure 3.13:	Stacked model of current research.	105
Figure 4.1:	Boruta model's chart plot on the importance of variables	113
Figure 4.2:	Frequency Ratio of landslide conditioning factors	114
Figure 4.3:	Success Rate AUC result	128
Figure 4.4:	Prediction Rate AUC results	129
Figure 4.5:	Overfitting Rate of all models	132
Figure 4.6:	Penang Island LSMs via (a) RF (b) XGBoost (c) Stacking	133
Figure 4.7:	Percentage of susceptibility areas covered by LSM generated by all models	134
Figure 5.1:	Summary of the research problem, questions, objectives, methods, results, and findings.	142

LIST OF ABBREVIATIONS

AB	AdaBoost
ADTree	Alternating decision tree
AHP	Analytical Hierarchy Process
AUC	Area under curve
BGAM	Boosted Generalized Additive Model
BO	Bayesian Optimization
CF	Certainty Factor
CNN	Convulational Neural Network
DEM	Digital Elevation Model
FOP	Factor Optimization Process
FR	Frequency Ratio
GP	Gaussian process
IDW	Inverse Distance Weighted
KNN	K-Nearest Neighbor
LaGRiSU	Landslide Grid and Slope Unit
LR	Logistic Regression
LSM	Landslide Susceptibility Mapping
ML	Machine Learning
NDVI	Normalized Difference Vegetation Index
NSMP	National Slope Master Plan
OOB	Out of Bag
RF	Random Forest

RFE	Recursive Factor Elimination
RNN	Recurrent Neural Network
ROC	Receiver operating characteristics
RUSLE	Revised Universal Soil Loss Equation
SPI	Stream Power Index
SVM	Support Vector Machine
TOL	Tolerance
TWI	Topographic Wetness Index
UCB	Upper Confidence Bound
USGS	United States Geological Survey
USLE	Universal Soil Loss Equations
VIF	Variance Inflation Factor
XGB	Extreme Gradient Boosting

CHAPTER 1

INTRODUCTION

1.1 Background

Landslides are natural hazards that present a major threat to the property, human life, and the environment where they occur because of their potential for destruction (Ayalew & Yamagishi, 2005; Chen et al., 2018). Penang Island is a region that consistently grapples with the occurrence of landslides. Penang Hill, situated in a hilly terrain, stands out as a construction project subject to ongoing, vigilant monitoring by regulatory authorities. The intensified scrutiny is a proactive measure undertaken in response to the increased susceptibility of Penang Hill to potential landslide events. Minor landslides have been observed to occur frequently on Penang Hill, which is particularly during periods of heavy rainfall.

To lower risks and losses from landslides, one of the first tasks completed by the Slope Engineering Malaysia upon its establishment in 2004 was the creation of a National Slope Master Plan (NSMP). A few of the NSMP action plans (2009-2023) that have been put into practice involve developing susceptibility, hazard, and risk maps, which are then followed by tasks like inventorying slopes, executing public awareness and education campaigns, and establishing agency committees to create efficient policies and collaboration for slope mitigation (Kementerian Kerja Raya, 2009).

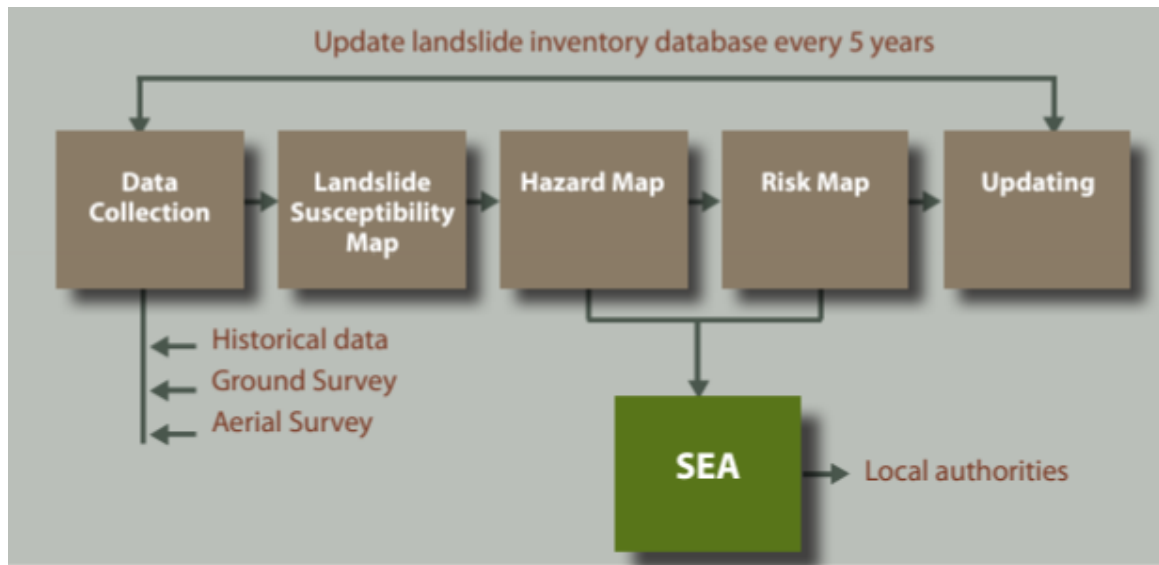


Figure 1.1: Implementation Process of hazard mapping assessment by Malaysia NSMP (2019-2023)

The Public Works Department (PWD) of Malaysia was tasked with cataloguing slopes and creating hazard and risk maps in high-risk regions, particularly cities, and towns where development has occasionally occurred without appropriate planning, design, and construction, according to NSMP. The budget for further hazard and risk mapping for landslide-prone areas, which spans an estimated 1,700 km², has been granted and encompasses approximately 95% of Malaysia's high-risk landslide-prone areas, which include major cities such as Penang Island and Kuala Lumpur (Kementerian Kerja Raya, 2009). In addition, the landslide susceptibility map (LSM) has also been included in the production of landslide risk assessments and mapping in guidelines prepared by the Association of Southeast Asian Nations (ASEAN) (ASEAN, 2021).

In addition, the Penang State Government has identified highland areas with an elevation over 76 m and a slope greater than 25° as a sensitive environment where development needs to be limited in their Structure Plan for 2030 (Jabatan Perancang Bandar dan Desa Negeri Pulau Pinang, 2019). The state's population is growing along with its

economy, which has increased demand for housing and infrastructure. As a result, development has spread into more vulnerable geohazard-prone locations, including hillsides and low-lying coastlines. Land development inside high-land and sloped areas has proceeded despite the inherent risks, particularly for residential construction, which raises the likelihood and potential effects of landslides (Lee & Pradhan, 2006).

The primary working hypothesis in prognostic techniques based on historical landslide events is that future landslides will have similar distribution patterns to previous ones (Youssef & Pourghasemi, 2021). An LSM can be created based on the information available to predict the occurrence of future landslides. Statistical, analytical, and soft computing tools for generating landslide susceptibility models exist.

It is essential to select suitable techniques for assessing landslide hazards and zoning. Geographic Information Systems (GIS) and machine learning (ML) models are powerful tools for LSM studies. ML model is a computational algorithm which includes Random Forest (RF) or Extreme Gradient Boosting (XGBoost) that is fed data to spot trends and extrapolate future trends. More specifically, within the context of landslide susceptibility study, the term refers to the predictive models which have been developed to analyze the landslide susceptibility by learning from data that is fed into them such as landslide inventories and landslide conditioning factors and producing results in the form of classification of susceptibility. As with the rest of machine learning, these models are part of it but only as tools to be applied to resolve certain issues through obtaining results from given data.

ML models, such as Random Forest and Extreme Gradient Boosting, are used in landslide susceptibility mapping (LSM) to predict areas at risk for future landslides. These

models learn patterns from data and are trained on historical occurrences and environmental variables. Studies like Youssef and Pourghasemi (2021) and Pham et al. (2021) have successfully created accurate susceptibility maps using these models. Despite a decade of comprehensive studies employing GIS and ML models, there remains ample opportunity for refinement within the current models due to certain drawbacks such as overfitting and overestimation which can affect the models' performance (Agrawal & Dixit, 2023; Park & Kim, 2019). This enhances the model's accuracy and performance in developing a highly accurate and reliable LSM. Thus, this research study will focus on addressing issues of the existing models used in LSM and proposing potential solutions to improve the performance of the models.

1.2 Problem Statement

Machine learning (ML) models have been used by researchers to generate landslide susceptibility maps in the past decade. Somehow, the variability in geo-environmental conditions across various study areas is also affecting the performance of the machine learning models (Dou et al., 2019). Each region has different characteristics of topography, hydrology, and environmental factors. Different models generate different evaluation results because each model has its own implementation method, benefits, and limitations (Merghadi et al., 2020). Besides the quality of landslide inventory, the landslide conditioning factors spatial databases play a great role in mapping landslide susceptibility (Tien Bui et al., 2012).

The existing data-driven susceptibility mapping models, despite a decade of study utilizing both GIS and ML models, face limitations in global applicability. This constraint arises from the substantial variability observed in landslide conditioning parameters. Challenges are further compounded by restrictions in accessing up-to-date and accurate data,

particularly in areas that are inherently difficult to reach. These constraints collectively impede the seamless development of a universally applicable landslide susceptibility map (Can et al., 2021). Thus, it is very important to choose a robust machine learning model with consistent performance across various study areas with the various landslide conditioning factors less affected by the available data quality. This will save more time, cost, and resources if there is a robust model that can predict and able to produce a reliable landslide susceptibility map with minimum input of landslide conditioning factors (Dou et al., 2019).

Furthermore, an overlooked challenge in landslide susceptibility mapping, evident in prior research such as Agrawal and Dixit (2023) and Rabby et al. (2020), is the issue of overestimation. In the context of LSM, overestimation in ML model is the model's tendency to predict more areas as susceptible to landslides than are actually at risk, potentially inflating the number of high-risk zones. This occurs when the model becomes too sensitive to certain patterns or features in the training data, leading to an exaggerated number of false positives (areas predicted as landslide-prone but are not). Overestimation tends to occur even when the ML model has high prediction performance (Agrawal & Dixit, 2023). This problem will produce a less reliable and accurate landslide susceptibility map (Rabby et al., 2020). Thus, it will also cause cost and time inefficiency for the town planner and local authorities.

Moreover, with contemporary machine learning prediction techniques, striking a balance between model simplicity and predictive accuracy proves challenging. This is because machine learning tends to perform better when its complexity increases, and this will lead to the occurrence of overfitting (Wu et al., 2019). The persistent issue of overfitting remains a critical concern, necessitating further refinement and the development of more reliable solutions. Developing an optimum model that is less complex but accurate is very

important. This can be achieved through a good hyperparameter optimization technique (Sun et al., 2020).

Numerous ML models have been developed over the years, but ensemble learning models such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost) have consistently outperformed them. XGBoost and RF are two effective machine-learning techniques employed in LSM research. The XGBoost is the boosting method of ensemble learning that has recently been employed in constructing landslide susceptibility models. The RF bagging strategy has been widely used for the past seven years.

Although the XGBoost and RF have been proven to be reliable models in predicting landslide susceptibility, they still have drawbacks which require improvement, especially in dealing with overfitting (Park & Kim, 2019) and overestimation issues (Agrawal & Dixit, 2023; Rabby et al., 2020). Hence, integrating two different ensemble learning models can further enhance the single homogeneous models' ability to make predictions and deal with geo-environmental variability (Susan et al., 2021). Moreover, the overestimation problem can be overcome by stacking two different models (Rabby & Li, 2020).

However, the practical applications of the proposed hybrid model may vary across different geo-environmental contexts. While the hybrid model could provide better adaptability and robustness in diverse environments, its limitations must also be considered. These include potential challenges in handling highly complex or large-scale datasets, variations in model performance across different geographical regions, and the need for careful selection and preprocessing of landslide conditioning factors to ensure accurate predictions (Liu & Li, 2020).

1.3 Research Question

1. What are the important factors contributing to the landslide occurrences?
2. How do individual RF and XGBoost models perform in landslide susceptibility mapping, considering the various landslide conditioning factors?
3. What are the optimal hyperparameter settings for individual RF and XGBoost to reduce overfitting and improve model accuracy in the context of landslide susceptibility mapping?
4. How can overestimation issues in individual RF and XGBoost models be identified and mitigated to enhance the reliability of landslide susceptibility maps?

1.4 Research Objectives

1. To analyze and understand the relationship between landslide conditioning factors using techniques like the Boruta algorithm for factor selection.
2. To evaluate and compare the performance of individual Random Forest (RF) and Extreme Gradient Boosting (XGBoost) in producing accurate landslide susceptibility maps (LSMs) by applying model validation techniques like cross-validation and performance metrics such as AUC score, accuracy, sensitivity, specificity, precision and kappa value.
3. To employ and assess the effectiveness of hyperparameter optimization methods, particularly Bayesian Optimization, for tuning the parameters of individual RF and XGBoost models, with steps including defining the

hyperparameter search space, selecting optimization criteria, and evaluating performance improvements in reducing overfitting.

4. To investigate and develop strategies for mitigating overestimation in individual RF and XGBoost models, such as model stacking, and assess their impact on improving the accuracy and reliability of landslide susceptibility maps.

1.5 Hypothesis

The hypothesis, based upon the problem statement, research questions and research objectives are presented as follows:

1. The importance of every landslide conditioning factors a can be determined effectively by factor selection techniques such as Boruta algorithm.
2. Next, the individual Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models are having good and reliable prediction performance, with prediction accuracy dependent on the quality and selection of landslide conditioning factors (LCFs).
3. Hyperparameter optimization using advanced methods like Bayesian Optimization will improve the performance of individual RF and XGBoost models by minimizing overfitting and enhancing prediction accuracy.
4. The hybrid ensemble of single RF and XGBoost models via stacking method will address limitations of the individual models, especially overestimation issue, resulting in improved accuracy and robustness in landslide susceptibility mapping.

1.6 Scope of Study

This research focuses on predicting landslide susceptibility mapping within a tropical region, specifically Penang Island (located between latitudes 5°15'N and 5°30'N, and longitudes 100°10'E and 100°20'E). The study examines internal landslide conditioning factors, such as topographical, hydrological, anthropogenic, and external factors, utilizing publicly available data sources, including Google Earth Pro, OpenStreetMap, the NASA landslide catalogue, and USGS data.

While the primary focus is Penang Island, the methodologies and findings developed in this study are applicable to other tropical regions with similar environmental and geophysical conditions. This broadens the potential relevance of the study, offering insights for landslide prediction and risk assessment in other parts of the world.

The research aims to assess and enhance the performance of hybrid machine learning models, specifically combining Random Forest (RF) and Extreme Gradient Boosting (XGBoost), to address the issues of overfitting and overestimation. The proposed solutions focus on reducing overfitting due to model complexity and overestimation caused by spatial correlation. These solutions, such as model stacking and hyperparameter optimization, are specifically designed for these issues and may not be effective for other factors that cause overfitting or overestimation. The models will be optimized through advanced hyperparameter optimization methods, such as Bayesian Optimization, to improve accuracy and reliability in landslide susceptibility predictions.

Several challenges are anticipated in this study, including variations in the quality of the publicly available data, such as inconsistencies in the datasets sourced from open-access platforms as well as high cloud cover for the landsat 8 images extracted from United States

Geological Survey (USGS) Earth Explorer. Additionally, model interpretability could present challenges when understanding the decision-making process of complex machine learning models, potentially affecting the model's practical application and trustworthiness. Generalization of the models across different regions with varying environmental conditions might also require further adjustments.

This study contributes to the advancement of landslide susceptibility mapping by integrating hybrid ensemble models, which combine the strengths of both RF and XGBoost. By utilizing model optimization techniques, such as Bayesian Optimization, and addressing overfitting due to model complexity and overestimation caused by spatial correlation, this research enhances the accuracy of landslide susceptibility maps. The findings and methodology presented here will significantly contribute to improving landslide risk assessment and mitigation strategies, which are crucial for urban planning, disaster management, and site selection in landslide-prone regions.

1.7 Thesis Outline

Chapter 1: Introduction

This chapter provides an overview of what is being studied and an explanation of its significance to set the scene for the study. It also provides an overview of the research's scope, goals, and issue description.

Chapter 2: Literature Review

This chapter presents a review of previous studies conducted in the field of landslide susceptibility assessment and prediction. It discusses in detail the tools and techniques used by researchers in the recent past. The review focuses on the overview of data sources,

landslide inventories, factors, spatial map development, landslide prediction models, and performance validation techniques.

Chapter 3: Modelling and Implementation

In this chapter, the models and methods used for conducting this research have been described. The chapter contains an explanation of the area of study, the methodology used to implement this research, and how it has been implemented.

Chapter 4: Results and Discussion

This chapter presents the results which have been generated as a result of the implementation of the methodology. It also explains how the results are interpreted and discusses the implications. The results and implications are analyzed to develop the proposed methodology's strengths and weaknesses.

Chapter 5: Conclusions

This chapter presents a complete summary of the research as well as the outcomes that were produced. It derives conclusions and comments upon the insights gained from the research. This chapter provides suggestions for further improvements and recommends the course of action to take in the future. In the end, this chapter describes the research findings and contributions to the current body of knowledge that this research seeks to provide.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

The previous chapter largely described the issues regarding landslide susceptibility prediction and mapping and the necessity of employing ensembles of various machine learning algorithms in this process. This chapter discusses the three important phases of developing a landslide susceptibility map. The first phase reviews the development of landslide inventories and the triggering factors of landslide events. The topic of discussion includes several kinds of research that have data collection methods, spatial database generation, and data usage for training and modelling algorithms. Next, since the focus of this research is prediction utilizing machine learning techniques, the relevance of various algorithms used for this goal in various studies is explored. Comparative reviews of the commonly used machine learning model in the past 10 years were conducted to accomplish this. The comments on observed strengths and drawbacks in the literature are made. Finally, it addresses the challenges in the process of developing the landslide susceptibility map and the potential solution extracted from the recommendations and findings of previous research. The study gap is then highlighted at the end of the review section.

2.2 Fundamentals of Landslide Susceptibility Mapping

Susceptibility maps are a vital beginning towards hazard assessment and mitigation measures in places with significant landslide risk at regional or catchment scales (Dou et al., 2019; Li et al., 2020). Landslide susceptibility is described by Brabb (1985) as the likelihood of a landslide occurring in each area based on the topographical and environmental characteristics provided. A landslide susceptibility mapping (LSM) technique is used to identify places where landslides are likely to occur (Guzzetti et al., 1999).

LSM evaluation is used by decision-makers and local authorities to divide the land surface into zones with varying levels of stability and instability. This procedure, referred to as "landslide risk zoning," is crucial to administering, controlling, and implementing preventative measures meant to lessen the hazards associated with past and present landslides. LSM requires the application of Geographic information system (GIS) and machine learning for the development of reliable landslide susceptibility maps (Merghadi et al., 2020).

It is efficient to identify and designate landslide-prone locations to construct a geographic database of landslide occurrence, or landslide inventory. Landslide conditioning factors (LCFs), which measure the geographical characteristics of landslide locations that may influence potential slope stability, can be assembled into a database using GIS to help in handling spatial data (Budha et al., 2018; Lee, 2019). The landslide inventories and LCFs spatial databases will be the input of the machine learning for the landslide susceptibility prediction. Hence, the identification of landslide inventories, the consideration of key LCFs and the meticulous selection of optimal machine learning models are imperative steps in the

development of a highly accurate and reliable landslide susceptibility map for a given study region (Hearn & Hart, 2019).

2.3 Landslide Conditioning Factors

Landslides occur due to highly complicated interactions between various landslide conditioning factors (LCFs), including topographical, hydrological, geological, and environmental factors (Kadirhodjaev et al., 2018; Sevgen et al., 2019). The topographical, geological, and hydrological elements are typically classified as static because they are primarily stable. Environmental elements, on the other hand, are typically dynamic and constantly changing (Pham et al., 2017). As a result, it is crucial to identify and examine the relationship between the LCFs and the landslide occurrence.

2.3.1 Common Landslide Conditioning Factors

There are two sorts of elements that affect a region's landslide risk. Geology, slope gradient, elevation, vegetation cover, and long-term drainage patterns are intrinsic factors that affect landslide susceptibility. Extrinsic factors that tend to cause landslides in an area of particular susceptibility include heavy rainfall and earthquakes (Atkinson & Massari, 1998; Wu & Sidle, 1996). The probability of landslide occurrence depends on both intrinsic and extrinsic variables.

The intrinsic influencing variables known as the topographic factors including elevation or Digital Elevation Map, slope angle, slope length, plane curvature, profile curvature Stream Power Index, Topographic Wetness Index, and environmental which includes Normalized Difference Vegetation Index, distance from roads, distance from rivers have all been used frequently (Zhou et al., 2021).

Choosing the influencing elements according to the actual circumstances is required to build a system for assessing landslide susceptibility in each research region. To create highly accurate landslide susceptibility models, Sahin (2020) claims that it is necessary to combine relevant landslide conditioning components. This is because accurate information about the characteristics of the study areas that could potentially increase the probability of landslides occurring is needed.

2.3.1.1 Digital Elevation Model (DEM)

An elevation or topographic map is defined by large-scale detail that shows detailed ground topography, drainage, forest cover, administrative areas, transportation paths, and other man-made features. Elevation maps have become increasingly important in developing urban settlements, contemporary infrastructure, and risk assessment (Hawker et al., 2022).

Digital Elevation Model (DEM) development frequently uses elevation points. A DEM, a digital cartographic dataset, represents a continuous topographic elevation surface through several cells. Each cell depicts a feature's height (Z) at its X and Y coordinate surfaces. It makes it possible to identify watersheds, establish stream networks, and characterize a location's physical functions (Sidaway, 2022). The DEM refers to the "elevation" of the ground surface, excluding vegetation, man-made structures, or other items above the surface.

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).

2.3.1.2 Slope Angle

The slope angle is an input parameter in susceptibility studies. It explains that it is the expression of the rate of the vertical distance to the horizontal distance between two specified points with the tangent angle (Çellek, 2022). The slope is important for forming and developing and is susceptible to landslides. The slope factor was considered in the study conducted by Meena et al. (2022), and it was found that landforms with a higher slope angle are more susceptible to collapse, which is closely correlated to landslides.

2.3.1.3 Slope Length

Slope length was the second most important factor contributing to landslide occurrence in Abha Basin (Youssef & Pourghasemi, 2021). The longer the slope length, the greater the amount of cumulative runoff, and the steeper the slope of the land, the higher the velocities of the runoff that contribute to erosion which resulted in landslide (Das et al., 2022).

For soil erosion prediction, empirical models such as the Universal Soil Loss Equation (USLE) (Wischmeier & Smith, 1978) and the Revised Universal Soil Loss Equation (RUSLE) (Renard et al., 1994) are widely employed. The slope length gradient (LS) component reflects the effect of topography on soil erosion rate by combining the effects of slope length (L) and slope steepness (S). The LS factor is the most significant parameter for estimating soil loss and managing soil using the USLE and RUSLE.

The LS factor, in this case, indicates the ratio of soil loss on a specific slope length and steepness to soil loss on a slope with a length of 22.1 m and a steepness of 9% (Renard et al., 1994; Wischmeier & Smith, 1978). The LS factor was calculated using the distance between each grid point, the height difference between the maximum and least pixel from the DEM, and the slope. The LS factor was computed (Wischmeier & Smith, 1978) using the Equation 2.1:

$$LS = \left(\frac{\lambda}{22.13} \right)^m \times (65.41 \sin^2 \theta + 4.56 \sin \theta + 0.0065) \quad \text{Equation 2.1}$$

By referring to Equation 2.1, λ is the slope length in meter (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.

2.3.1.4 Plan and Profile Curvature

Profile and plan curvatures are used for hill-slope and land- slide analysis (Abd Majid & Rainis, 2019). The degree of slope degradation is determined by the curvature of the plan and the profile, which regulate the speed at which water flows (Aghdam et al., 2016). The curvatures had a fair contribution to the occurrence of landslides in many studies, such as Wang et al. (2020), Nakileza and Nedala (2020) and Pham et al. (2018).

2.3.1.5 Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) had been an important factor for the LSM study in Lushui, China (Hu et al., 2020). The most susceptible class of the NDVI in Lushui region falls in the range of – 1 to 0.12. This is because bare land is more prone to erosion.

NDVI assesses vegetation by comparing the amount of near-infrared light reflected by vegetation to the amount of red light absorbed by vegetation. The NDVI is vital in immobilizing huge amounts of water and enhancing the lithological mass's shear resistance and soil cohesion (Sidle & Ochiai, 2013). The addition of grassland and roots is well-acknowledged to play an important effect in soil strength (Hammad Khaliq et al., 2022). The NDVI influences landslide susceptibility modeling (Althuwaynee et al., 2012). The NDVI is calculated using Equation 2.2:

$$NDVI = \frac{(NIR - R)}{(NIR + R)} \quad \text{Equation 2.2}$$

In Equation 2.2, NIR and R are the electromagnetic spectrum's near-infrared and red bands, the outcome of this calculation is a value between -1 and +1. A high NDVI value will result from low reflectance (or low values) in the red channel and high reflectance (or high values) in the NIR channel. Overall, NDVI is a standardized method of determining healthy vegetation. Higher NDVI readings indicate better vegetation. The NDVI scale always falls in between -1 to +1 ranges:

- Negative one (-1) value for areas other than vegetation cover such as snow, water, or clouds where red reflectance is greater than Near-infrared.
- For bare soils, with reflectance values, roughly the same in red and near-infrared, the NDVI has values close to 0.
- One (1) means vegetation formations that have positive NDVI values, generally between 0.1 and 0.7.

2.3.1.6 Distance from Road and Stream

The construction of mountain roads demands engineering activities such as cutting or excavating slopes, which alter the original geological conditions and weaken the natural support of rock slopes. These activities have a significant and negative impact on the occurrence of landslides (Wang et al., 2016). Road building, such as cutting and excavation is a major anthropogenic element influencing the structure of natural slopes (Xiao et al., 2019). Road distance is the 4th most influential LCF in the research conducted by Youssef and Pourghasemi (2021) and Nhu et al. (2020) had shown that an area located within 150m of the road had experienced multiple landslide events.

Next, 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). Thus, the distance from the river had been used for the LSM in Mahrloo lake, Iran (Amiri et al., 2019) and Shicheng County, China (Huang et al., 2020). The distance from the river had an important role in landslide events in Shicheng County, whose contribution was significantly proven by all of the prediction models used in the study.

2.3.1.7 Stream Power Index, and Topographic Wetness Index

Stream Power Index (SPI) can be used to describe potential flow erosion at the given point of the topographic surface. As catchment area and slope gradient increase, the amount of water contributed by upslope areas and the velocity of water flow increases. Hence, the stream power index and erosion risk increase (Florinsky, 2017). Topographic Wetness Index (TWI) TWI describes the soil moisture spatial pattern (Yilmaz, 2010).

TWI is ranked 3rd, and SPI is ranked 6th crucial factor out of all the 13 LCFs in the Chittagong, Bangladesh landslide studies (Rabby & Li, 2020). Similarly, in the LSM study in Nanchuan county, China (Chen & Li, 2020), the TWI and SPI also had shown important contribution as the causal factors of landslide .

Like other combined morphometric variables, SPI and TWI can be derived from a DEM by the sequential application of methods for local and nonlocal morphometric characteristics, followed by an arithmetic combination of the results of these calculations. SPI considers data on catchment area and slope gradient as in Equation 2.3:

$$\mathbf{SPI = ln(CA. tan G)} \quad \text{Equation 2.3}$$

By referring to Equation 2.3, CA and G are catchment area and slope gradient, respectively.

2.3.1.8 Selection Criteria and Data Sources

Selection Criteria

There are no general rules for how many impact factors are used and the methods used to select factors for susceptibility evaluation. The selection criteria should be based on data availability, degree of affinity with landslide locations, measurable factors, non-redundant factors and geomorphological characteristics of the area under study (Arabameri et al., 2022; Hakim et al., 2022; Hussain et al., 2021)

The extrinsic variables are not considered in this study is because the variables may change over a very short time span and are thus very difficult to estimate. If extrinsic variables are not considered, the term “susceptibility” could be employed to define the likelihood of occurrence of a landslide event. The spatial distribution of the intrinsic variables within a given area determines the spatial distribution of relative landslide susceptibility in that region (Dai et al., 2001).

Data Sources

The raw data of the LCFs can be obtained from public sites or authorities (Khaliq et al., 2022; Nakileza & Nedala, 2020; Sahin, 2020), as shown in Table 2.1

Table 2.1: Data sources for every LCFs

Category	LCFS	Data Type	Data Sources
Topographic	DEM	DEM	National Survey and Mapping Department, Google Earth
	Slope angle		
	Slope length		
	Profile curvature		
	Plan Curvature		
Hydorological	TWI		
	Distance from stream		
	SPI		
Anthropogenic	Distance from road	Road networks	Open Street Map
	NDVI	Landsat 8-satellite image	United States Geological Survey (USGS) Earth Explorer

2.3.2 The Importance and Challenges in selecting Landslide Factors

Landslides are the result of a combination of many factors, and there is no single factor or single natural factor that affects landslides. The probability of landslides could increase or decrease depending on how different elements interact. One of LSM's most significant future challenges is considering the interactions between various landslide factors. An improved understanding of landslide occurrences and the available conditioning factor maps for LSM may result from focusing on the significance of the landslide conditioning factors (Zhou et al., 2021).

Table 2.2: Past studies investigating effect of LCFs on models' prediction performance.

Studies	Number of initial LCFs	Number of eliminated LCFs	Model AUC before LCFs elimination (%)	Model AUC before LCFs after elimination (%)
(Meena et al., 2022)	14	9	RF= 90.2	RF= 90.6
(Sahin, 2020)	15	10	LR= 73.26	LR= 74.73
(Zhou et al., 2018)	12	10	SVM= 89.3	SVM= 91.2
(Zhang et al., 2017)	34	12	RF=96.0	RF=97.0

Several studies, including those conducted in the Three Gorges Reservoir in China (Zhang et al., 2017; Zhou et al., 2018), Belluno, Italy (Meena et al., 2022) and Ayancik, Turkey (Sahin, 2020) have shown that elimination of less important and significance LCFs, can improve the model performance accuracy (Table 2.2). The are under the curve (AUC)

of a prediction model can be improved by 1% to 2%. This insight clarifies that using only the vital conditioning factors will suffice for LSM and result in time and cost savings.

Selecting the global landslide conditioning factors presents several obstacles. First, for various reasons, not all conditioning factor maps will be available globally or sometimes locally. These include non-compliance with data sharing and unavailability. For this reason, it can be valuable to identify which of the general conditioning factors are crucial to LSM. Secondly, the least significant factors in one location may not repeat the same behaviour in other parts of the world, and the model's capability varies depending on the region (Al-Najjar et al., 2019; Alqadhi et al., 2022).

2.3.3 Landslide Conditioning Factors Selection Techniques

Selecting the appropriate variables for landslide modelling is crucial during the evaluation of landslide susceptibility. This is because conditioning factors within the training dataset may not be independent of one another, which might create noise and confuse the model (Chen et al., 2017). Multicollinearity test and Boruta are the methods used to perform the feature selections.

2.3.3.1 Multicollinearity Test

Multicollinearity describes a linear relationship between two or more variables. This data issue could seriously impair the accuracy of the model parameter estimates (Alin, 2010). This algorithm has been widely used in selecting important landslide conditioning factors (LCFs) in several studies such as landslide susceptibility prediction mapping (Prasad et al., 2021), gully erosion prediction mapping (Ahmadpour et al., 2021; Amiri et al., 2019), flood

assessment mapping (Habibi et al., 2023; Prasad et al., 2022) and streamflow forecasting (Masrur Ahmed et al., 2021).

The independent variables are prone to correlate in analysing a linear regression model. Multicollinearity is an issue that will lower the model's prediction accuracy and perhaps cause the model to fail (Pourghasemi et al., 2018; Rahmati et al., 2017). Thus, examining the multicollinearity of landslide-influencing components is essential before analyzing landslide susceptibility (Bui et al., 2016; Zhou et al., 2021).

Multicollinearity test consist of variance inflation factor (VIF) and tolerance level (TOL). In general, it is thought that the two indicators have a logical correlation, and the model has a significant correlation issue when $VIF > 10$ and $TOL < 0.1$ (Chen et al., 2018; Xia et al., 2022). Variance inflation factor (VIF) and tolerance (TOL) methods have widely being used in landslide studies conducted by previous researchers such as (Chen et al., 2019; Fang et al., 2020; Li et al., 2022).

2.3.3.2 Boruta Model

In machine learning models, the determining factor optimization process (FOP) step is one of the essential and necessary prerequisites for determining LCFs in the optimum modelling of classification and regression problems to reduce the high computational time and the probability of overfitting the models, which leads to improving the accuracy of the models (Liu et al., 2021). FOP algorithms have eliminated less influential predictors to improve model accuracy.

The Boruta algorithm is one of the wrapper-based algorithms developed based on the random forest classifier (Ahmadpour et al., 2021) and is a feature selection approach within

the RF classification method (Liu et al., 2021). This algorithm follows an all-relevant variable selection approach, which considers all features related to the output variable, considers multi-variable relationships, and can investigate interactions between variables (Ahmadpour et al., 2021).

Boruta had been efficiently used in the study of multiple hazards including the landslides, flood and forest fire hazard susceptibility mapping conducted by Pourghasemi et al. (2020) in four strategic watersheds located in Fars province, Shiraz City, Iran. Pourghasemi et al. (2020), had been using the Boruta model to assess the importance of causal factors for landslide, flood and forest fire occurrence in the Fars Province watersheds and built RF model to predict the multi hazard susceptibility. Boruta model had proven to be efficient in selecting the important causal factors for each hazard and the RF AUC scores in flood, landslide and forest fire susceptibility maps are 83.40%, 93.90%, and 94.30% respectively indicating a high prediction performance. Thus, the Boruta algorithm has been proven to contribute efficiently to the machine learning models' performances in the predicting landslide, flood and forest fire susceptibility.

In addition, Boruta had also efficiently used in flash flood susceptibility assessment in Neka-Haraz watershed, Iran done by Habibi et al. (2023). The study of Neka-Haraz watershed had compared the efficiency of Boruta and recursive feature selection (RFE) in selecting the flood causal factors. Boruta and RFE were combined with extreme gradient boosting (XGBoost) for better prediction performance. Finally, the result had revealed that hybrid of XGBoost-Boruta had the produced better performance with AUC score of 94.3% compared to XGBoost-RFE (AUC=93.8%) and individual XGBoost (AUC=91.1%) which

indicate that Boruta is efficient in selecting important LCFs for building the prediction model.

2.3.4 Frequency Ratio (FR) Analysis

After the multicollinearity test and confirming the chosen LCFs, a more in-depth analysis of the relationship between the subclasses of LCFs selected and the occurrence of landslides need to be conducted. Frequency Ratio (FR) analysis had been used in previous LSM studies done by Huang et al. (2020), Rabby and Li (2020) and Xiao et al. (2019).

FR is often used as in indicating the strength of relation between the LCFs and landslide occurrence. The LCFs are divided into multiple subclasses and the FR analysis is then utilized to calculate the FR values of each subclass of conditioning factor on landslides (He et al., 2012; Wu et al., 2013). For an example, the study of LSM in Araniko highway, Nepal by Acharya and Lee (2019), the prediction rate had shown that the slope angle had significant contribution towards the landslide and the FR shows in detail that the slope which is more than 45° is highly susceptible to landslide.

Similarly, the prediction rate in the Chittagong, Bangladesh LSM study by Rabby and Li (2020) indicated that the slope angle played a significant role in the landslide, and the detailed finding from the FR indicates that slopes between 31.2° and 68.4° are the most susceptible to landslides. Thus, it can be concluded that FR plays an essential role alongside the VIF, TOL and Boruta in determining the critical LCFs for landslide susceptibility modelling.

2.3.5 Landslide Inventories

A landslide inventory map records the geographic location of landslide occurrence (Al-Zuhairi et al., 2019; Chen et al., 2020). The precise detection of landslide emplacements is crucial for predicting and assessing the landslide susceptibility models. A new generation of satellite images, like Word view and Geo-eye, aided the quick reproduction of landslide inventory maps (Guerriero et al., 2019; Tofani et al., 2014). The records of landslide events can also be obtained from historical databases, field verification, published data, and news (Khaliq et al., 2022). The landslide inventories and the LCFs are the main input for the landslide susceptibility modelling.

The prediction models needed both landslides and non-landslides information which was obtained from the landslide inventory map. Thus, class imbalance might occur if the landslides points are lesser than the non-landslide points. To avoid the occurrence of class imbalance in landside inventory, the landslide grid and slope unit extraction (LaGRiSU) tool pack was used (Al-Thuwaynee et al., 2021).

- **Class imbalance**

In preparing landslide inventories, the number of landslide data must be balanced with the non-landslide data to avoid class imbalance. The class imbalance usually occurs when the number of landslide data is much smaller than the non-landslide data, such as in the study of Ataturk Dam, Turkey (Can et al., 2021). To avoid class imbalances, the extracted non-landslides data from the study area is ensured to be equal to the number of landslide datasets (Arabameri et al., 2021; Chen et al., 2018; Park & Kim, 2019; Rabby et al., 2020; Sun et al., 2021). Thus, extraction of equal number of landslides and non-landslides data is achievable using the LaGRiSU toolpack.

- **Grid unit extraction**

According to Sameen et al. (2020), the non-landslide sample should be at least 500m away from landslides. Next, the distance between any two non- landslide samples must be more than 100m. These requirements help to prevent sampling in areas that have experienced landslide events. The LaGRiSU tool pack created by Al-Thuwaynee et al. (2021) has been used in several 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 both the grid unit, which divides the territory into regular squares of a specified size, and the slope unit, which delineates the territory into independent slopes based on ridge and valley lines. However, the grid unit exhibits superior automation and a simpler structure compared to the slope unit (Nam & Wang, 2020; Zhang et al., 2022).

2.4 Evolution of Landslide Susceptibility Mapping Models

The accuracy and precision of susceptibility models rapidly evolve from statistical learning and opinion-driven models to greater usage of machine learning approaches. Numerous methods and procedures have been presented to assess landslide susceptibility and hazard zonation (Corominas et al., 2014; Reichenbach et al., 2018). Broadly, all these techniques or approaches may be further classified into qualitative and quantitative (Figure 2.1).

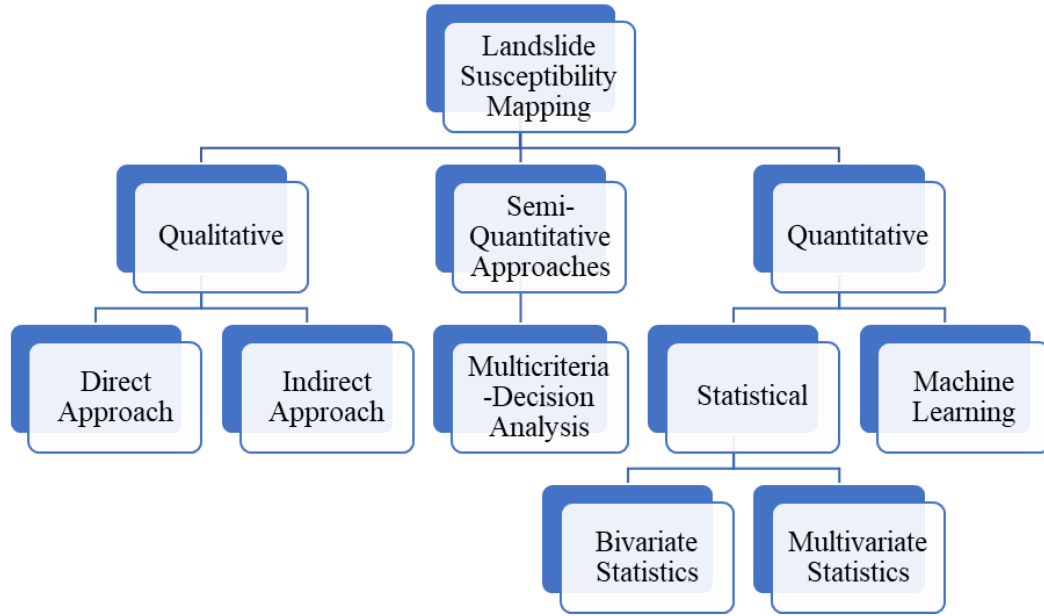


Figure 2.1: LSM Approaches (Shano et al., 2020)

2.4.1 Qualitative Approaches

According to Corominas et al. (2014) and Raghuvanshi et al. (2014), a heuristic evaluation and a geomorphic technique are two examples of qualitative methodologies. These techniques rely on the evaluator's expertise and experience. Landslide evaluation and zonation are done indirectly using the heuristic evaluation technique. These methods are founded on data, mostly collected in the field, and they are firmly supported by an evaluator's expertise and judgement (Raghuvanshi et al., 2014). Expert opinions can significantly differ when assessing the hazard and susceptibility to landslides in a particular area. Nonetheless, the ease of application of these methods contributes to their popularity (Ayenew & Barbieri, 2005).

The direct methods of landslide susceptibility and hazard zonation mapping are the geomorphic techniques. The process of geomorphic mapping landslide susceptibility requires the use of satellite photos or aerial photography (Mandaglio et al., 2016). These

techniques frequently require the development of a landslide inventory map that shows the locations of landslides. Landslide inventory mapping is thought to be simple and necessary for most susceptibility and hazard zonation methodologies (Dai & Lee, 2002), either to verify the predicted future landslide events or to develop general guidelines for hazard prediction. These geomorphic approaches typically do not consider the causes or triggers of the landslides.

2.4.2 Semi-Quantitative Approach

A multi-criteria decision-making approach called the Analytic Hierarchy Process (AHP) is used to assess landslide hazard susceptibility and hazard zonation mapping. According to Pardeshi et al. (2013) and Saaty (2008), AHP is a systematic strategy that includes defining the problem, determining the objectives and natives, creating a pairwise comparison matrix, determining weights, and determining an overall priority. The AHP technique can be used to measure the relationship between landslides and causal factors. There are also some disadvantages to using AHP. The computed output priorities for decision alternatives can introduce uncertainty due to theoretical issues. Additionally, AHP requires a thorough understanding of its fundamentals and may not be suitable for all types of complex decisions (Saaty, 2008).

2.4.3 Quantitative Approaches

Numerous quantitative approaches and techniques have been put forth and effectively used for landslide modelling, which helps to understand landslide patterns and their triggering mechanisms (Dou et al., 2019; Pham et al., 2016). Prior scholars employed bivariate and multivariate statistical approaches for landslide susceptibility evaluation and zonation (Ayenew & Barbieri, 2005; Kanungo et al., 2006). Bivariate statistical analyses

independently assess every causal factor's contribution to causing a landslide in a particular area.

The density ratio is used to quantify the relative contribution of different causal elements to the occurrence of landslides. Each layer of the causative factor is layered over the landslide incident distribution layer in an overlay analysis that yields the density ratio. The overlay analysis will provide a ratio between the landslides and landslide-free zones within each causative factor class (Hamza & Raghuvanshi, 2017). The assessment of these factors and their correlation with previous landslides in the region could serve as a foundation for predicting potential regions that may experience landslides in the future (Chimidi et al., 2017).

Meanwhile, multivariate statistical model is based on the relationship of relative contribution of each causative factor to the total landslide susceptibility in the area (Nandi & Shakoor, 2010; Pardeshi et al., 2013; Saaty, 2008). The percentage of landslides for each pixel is determined in multivariate methods for landslide susceptibility analysis, and a data layer on the existence or absence of landslides is created by statistical analysis. Statistical models have benefited the most from the last ten years of landslide susceptibility mapping studies.

Finally, the current and most used approach for landslide susceptibility prediction is The Machine Learning (ML) approach. The ML approach is a systematic method for creating predictive models, consisting of three main stages mainly the preprocessing, where raw data is cleaned, and normalized, and relevant features are selected, for instance, using Boruta to ensure they are suitable for modelling. Next is experimentation, which includes testing ML models like Random Forest and XGBoost, followed by optimization of the model's

hyperparameters such as through Bayesian optimization. Finally, the performance of the models is evaluated using metrics such as accuracy and AUC to assess the models' reliability and applicability. This iterative process is designed to enhance predictions and ensure they are practically relevant.

ML models extract knowledge from data without relying on rule-based functions, while statistical modelling uses mathematical equations to streamline correlations between variables in the data. Besides, ML modelling is more applicable than statistical models as it can address the complex non-linear relationship between landslide conditioning factors and landslide occurrence (Wang et al., 2020).

In this research, landslide susceptibility prediction was conducted using the ML approach, which falls under the quantitative approach. Unlike qualitative approaches, which rely on expert judgment, or semi-quantitative methods, which combine expert opinion with numerical scoring, the quantitative approach uses numerical data and statistical or computational models to derive predictions. The ML approach, utilizing ML models such as RF and XGBoost, provides a data-driven, objective framework for analyzing the complex relationships between landslide conditioning factors and susceptibility. This ensures greater accuracy and robustness in generating susceptibility maps, making it an ideal choice for addressing the research objectives.

2.5 Machine Learning Techniques as Landslide Susceptibility Prediction Models

Machine learning models predict landslides to increase model accuracy and adaptability to various conditioning factors (Bui et al., 2018). Over the past ten years, the mapping of landslides has significantly improved, mainly due to the application of machine learning approaches in landslide susceptibility prediction and mapping (Shirzadi et

al., 2018). An extensive review of previous and current approaches is necessary to select an effective and suitable modelling technique. Researchers have investigated a variety of machine-learning models used in landslide susceptibility mapping including logistic regression, Support Vector Machine, K-Nearest Neighbor, ensemble and deep learning techniques which will be further discussed in the following subsections.

2.5.1 Logistic Regression

In the field of earth science, logistic regression (LR) is thought to be the most widely employed statistical technique (Yilmaz, 2010). According to Shahabi et al. (2014), LR makes it possible to create a multivariate regression relationship among a dependent variable and several independent variables. In LR, the relationship between the input and output variables is not required to be linear (Belyadi & Haghighat, 2021).

LR is one of the most used multivariate techniques for creating LSM. According to Kavzoglu et al. (2019), they conducted the study to determine the relationship between the dependent variable, which is the presence or absence of a landslide and the independent variables, which are the causative factors. There are eight causative factors employed. LR was used to determine how spatially related the independent and dependent variables were. The finding found that LR is good at addressing the non-linear relationship between the LCFs and occurrence of landslides. Nevertheless, Zhou et al. (2018) stated that LR is unsuitable for addressing complex non-linear problems.

LR also shows a drawback in the study conducted by Wubalem and Meten (2020), where it was found that LR has limitations of generalization and simplification of landslide factors, affecting its prediction performance. In comparative studies of LSM in Kalekoy Dam, Turkey (Sevgen et al., 2019), the AUC score of LR is just 76% and had been

outperformed by artificial neural network (ANN) and RF with score of 84% and 95%, respectively. In addition, the LR had resulted in the conservative result of the landslide susceptibility map for Kalekoy Dam, making it less accurate than the map produced by the RF model. Thus, the final finding of the study had indicated that LR had a fair level of prediction performance and still can be outperformed by other ML models.

2.5.2 Support Vector Machine

A Support Vector Machine (SVM) is a machine learning algorithm that has gained popularity with the development of artificial intelligence and the widespread use of GIS and remote sensing. SVM is a machine learning algorithm based on statistical learning theory that adheres to structural risk minimization principles (Vapnik, 1999). Using a kernel function and an optimal classification hyperplane are the two fundamental concepts of SVM (Yao et al., 2008).

SVM performance was evaluated in the LSM study in Longju in the Three Gorges Reservoir area, China (Zhou et al., 2018). It indicated a very strong prediction performance with AUC score of 93.7% and 91.2% for both success and prediction rates, respectively. The study had incorporated 202 datasets and 12 LCFs. The finding of this study had contradicted the study conducted in Ninqiang County, China (Zhao et al., 2021) that was involving as many as 664 datasets and 10 LCFs. SVM had shown weaker prediction performance, with only 67.6% and 68.0% for its success and prediction AUC score in Ninqiang County LSM.

The LSM studies conducted in Longju within the Three Gorges Reservoir area and Ninqiang County, China, collectively suggest that Support Vector Machine (SVM) holds promise in addressing prediction challenges associated with a smaller number of samples, nonlinearity, and high-dimensional variables. However, it is noteworthy that SVM exhibits

limitations in suitability for LSM studies dealing with extensive datasets, making it less suitable for regional landslide susceptibility mapping. Nevertheless, the outcomes from both studies in Longju and Ninqiang County also attest to the resistance of SVM against overfitting, as evidenced by minimal differences in success and predictive rates. This is because to achieve structural risk minimization, the SVM's loss function comes with regular terms, and there is a balance between training errors and model complexity to prevent overfitting (Zhao et al., 2021).

2.5.3 Deep Learning

In the areas of artificial intelligence, signal processing, and natural language processing, deep learning algorithms that directly learn informative representations from unprocessed data have outperformed conventional machine learning techniques (Paoletti et al., 2018). Convolutional neural networks (CNN) and recurrent neural networks (RNN) have been successfully used by researchers to analyze the susceptibility to natural hazards (Dao et al., 2020; Khosravi et al., 2020; Yi et al., 2020).

According to Zhu et al. (2017), a Convolutional Neural Network (CNN) is an example of a deep learning technique that may automatically learn meaningful features from hierarchical neural networks. One or more convolutional layers, pooling layers, and numerous fully linked layers make up the general CNN architecture. The convolutional layer might use a collection of convolutional kernels to learn efficient representations from the input data.

Additionally, Recurrent Neural Network's (RNN) effectively process tasks involving sequence analysis. In contrast to CNN, it can handle sequential input utilizing recurrent hidden states, which can learn relevant information from both the previous and the current

phases. This indicates that RNNs can extract important information from sequence data (Chen et al., 2015).

These deep learning techniques have been shown to be more accurate predictors of landslide risk than traditional machine learning techniques. These effective examples, nevertheless, have a few drawbacks that restrict the accuracy each algorithm predicts (Korzh et al., 2017; Wang et al., 2011).

Zhang et al. (2019) had a conductive comparative study of LSM, which involved the application of one-dimensional convolutional neural networks (1D-CNN) to analyse and fit the relationship between debris flow triggering factors and occurrence in the study of debris flow susceptibility mapping using machine-learning techniques in the Shigatse area of China. The outcomes show that 1D-CNN had shown such a very good prediction performance with AUC of 97.6%. Somehow XGBoost had been outperforming CNN with an AUC of 98.8% in the Shigatse area LSM. This can be explained because CNN is less efficient than XGBoost in solving tabular data. CNN is efficient in capturing images, sounds, and text via modelling the spatial-temporal locality in second place.

In addition, the LSM research in Iran Ngo et al. (2021), which is a national scale study involving as much as 4069 landslides datasets and 11 LCFs covering 1.648 million km² had proven that both CNN and RNN are only suitable for LSM studies that incorporates large datasets and covering large study region.

2.5.4 K-Nearest Neighbor

When the event (landslide or non-landslide) is mainly represented within its (k) neighbours, the KNN algorithm classifies it. A small positive integer called k is frequently

used (Sameen et al., 2020). A distance metric is used to determine how close the samples are to one another. The distance metric shows how similar or dissimilar any given pair of samples' conditioning factor profiles are. A little feature distance will exist between data points with similar conditioning factors. The model is straightforward in terms of hyperparameters. However, as the number of samples increases, computation costs increase. The distance from each training data point in high-dimensional feature space is used to calculate the landslide susceptibility related to a particular set of values of the conditioning parameters.

These are the k closest data points. The new data point belongs to the dominant susceptibility class among those k nearest neighbors or the class with the most significant number of members among the k members (Cheng et al., 2013). In the research study conducted in Robat Turk, Iran (Avand et al., 2019), KNN had shown a good prediction performance since its AUC score is 80.9% although RF with AUC of 87.4% is still outperforming it. This is because the KNN performance is dependent on the data. Since all the LCFs were incorporated in the model training, regardless of their significant level, which resulted in the degradation of KNN performance due to the presence of noises and irrelevant features.

2.5.5 Ensemble Machine Learnings

Ensemble learning is a popular machine learning technique because of its excellent resilience, accuracy, and generalisation (Sagi & Rokach, 2018). To increase accuracy or stability, ensemble learning integrates many base predictors, combination of several methods (Nachappa et al., 2020). The predictability and robustness of assessing landslide susceptibility have been improved using ensemble machine-learning techniques. When

multiple classifiers are combined to form ensemble learning, the results are much better than when using a single classifier, especially when dealing with "weak classifiers," which are often defined as classifiers that are only marginally better at classification than a random guess (Lessmann et al., 2015).

Machine learning ensembles can be either homogeneous or heterogeneous. Boosting and bagging are two types of homogeneous ensemble techniques. Homogeneous ensemble machine learning is the integration of similar classifiers, such as bagging and boosting. Meanwhile, a heterogeneous ensemble involves the integration of different classifiers, such as stacking. XGBoost represents one boosting strategy within this framework, where classifiers are interdependent and must be executed sequentially. In contrast, bagging algorithms, like RF, do not rely on classifiers and can be run in parallel. In recent years, various ensemble learning techniques have been used to create landslide susceptibility maps to overcome the limitations of a single machine learning algorithm (Pham et al., 2022). To gain a deeper understanding of the RF and XGBoost model, the subsequent subsections provide a comprehensive overview of the RF and XGBoost structures, hyperparameters, strengths, and limitations.

2.5.5.1 Random Forest

- **Model Structure**

Breiman (2001) developed the RF classifier, an ensemble model that combines a few methods to address classification or regression issues (Chen et al., 2018). The decision tree-based RF algorithm integrates the predictions of numerous decision trees. The RF has improved both the decision tree and bagging models (Youssef et al., 2016). One of the most important features of the Random Forest Algorithm is that it can handle the data set

containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. It performs better for classification and regression tasks.

According to Elith et al. (2008), the architecture of the Random Forest algorithm is illustrated in Figure 2.2 and explained as follows:

- I. **Selection of Subset:** Bagging starts by choosing a random sample, or subset, from the entire dataset.
- II. **Bootstrap Sampling:** Each model is created from Bootstrap Samples, taken from the original data with replacement. This process is known as row sampling.
- III. **Bootstrapping:** The step of row sampling with replacement is referred to as bootstrapping.
- IV. **Independent Model Training:** Each model is trained independently on its corresponding Bootstrap Sample. This training process generates results for each model.
- V. **Majority Voting:** The final output is determined by combining the results of all models through majority voting. The most predicted outcome among the models is selected.
- VI. **Aggregation:** This step, which involves combining all the results and generating the final output based on majority voting, is known as aggregation.

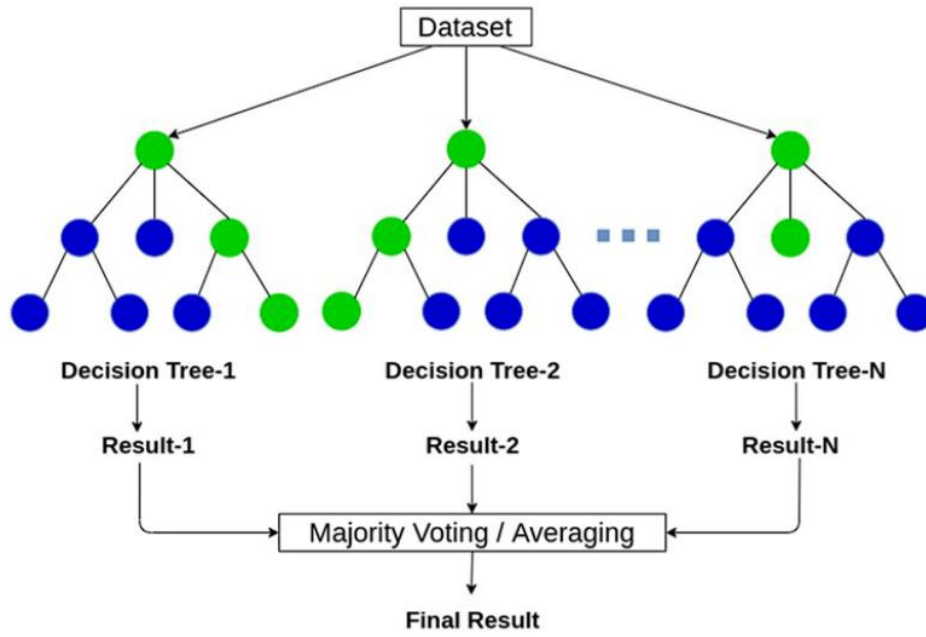


Figure 2.2: Random Forest Structure (Mumtaz et al., 2023)

Branching, leaf nodes, and decision nodes are characteristics of Random Forests. The training observation is questioned, and the primary split is carried out in a decision node (Can et al., 2021). The Random Forest's random operation greatly enhanced the classifier's performance. The rapid construction of each decision tree allows for parallelization in the Random Forest's development, which significantly accelerates the classification process (Parmar et al., 2019).

- **Out-Of-Bag**

The model never uses some observations because the RF approach employs bootstrapping for sampling. Out of Bag Observation (OOB) is the name given to these observations. It is utilized to figure out the OOB error, which is just the model's misclassification rate on the OOB (Ada & San, 2018). As stated by Matthew (2011) and Mohandoss et al. (2021) the advantages of using OOB based on are as follows:

- I. The OOB experience no data leakage during the validation process since the data is unused during the training of the model which resulting in better predictive model.
- II. Since OOB generalization ensures no leakage, so there is no over-fitting of the data and hence least variance and hence it makes a much better predictive model than a model using other validation techniques.
- III. It requires less computation as it allows one to test the data as it is being trained.
- IV. The OOB is widely used in model which having the small to medium-sized datasets. Thus, making it suitable to be used in this study since there were medium sized landslide datasets to be processed.

- **Hyperparameters**

The primary factors affecting the success of the decision tree are feature selection, the predictors used for splitting, and the decision of when to stop the computing process (Shirzadi et al., 2018). The "mtry", "ntree" and "min_n", were the hyperparameters of the RF (Chen et al., 2017; Sun et al., 2021).

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 Random Forest model's "ntree" defines how many trees should be built (Schonlau & Zou, 2020). When we are dealing with many predictors, some predictors will be correlated. In such a scenario, a small value of "mtry" is recommended (Chen et al., 2018).

The common RF hyperparameters optimization techniques are manual tuning (Wang et al., 2020), grid search (Adnan et al., 2020), and Bayesian optimization (BO) (Sun et al., 2020). A detailed explanation regarding each of the stated techniques will be explained in the hyperparameters optimization techniques section.

- **Strength And Limitations**

RF has been proven to outperform several ML models such as SVM, ANN, KNN, and LR various LSMs studies across different study regions such as in Abha Basin, Iran (Youssef & Pourghasemi, 2021), Muzzafarabad, Bangladesh (Hussain et al., 2022) and Lianhua County, China (Hong et al., 2016). Regardless of dealing with different environment factors inducing landslides in different study regions, RF still maintains good prediction performance with AUC score falls in the range of 80%-90% for the studies above.

As the non-linear LCFs were constantly changing and updating within years, the RF had proven to be able to address the non-linear relationship between the LCFs and the landslide occurrence. This has been proven in the research study of LSM created by Khaliq et al. (2022), which involved RF and LR in the District of Hattian Bala in Pakistan's northwest Himalayas. 349, 393, and 735 landslide inventories from 2005, 2007, and 2012, respectively, were used in the experimental design. There were 13 LCFs used. The RF maintained high prediction accuracy across the three years even though the LCFs changed

throughout the 3 different years. The AUC score of RF prediction is 93.57%, 89.24%, and 98.27% for the year 2005, 2007 and 2012 respectively.

Besides, it was also proven that the RF is not experiencing overestimation since it shows a lower coverage area in the very high susceptible zone compared to LR in 2005. The study demonstrates that the increase in area coverage from 2005 to 2007 and 2012 in a highly susceptible zone, as predicted by RF, is comparably lower than the increment rate observed with LR. This finding suggests that RF exhibits resistance to overestimation and proves to be reliable in addressing the non-linear relationship between LCFs and the occurrence of landslides. The resistance of RF towards overestimation, which makes it able to produce a reliable landslide susceptibility map was also proven in the study conducted by Rabby et al. (2020) where its true positive and false positive values are consistent.

Besides, a hybrid method that incorporates the combination of RF and two feature selection techniques, namely GeoDetector (RF-GeoDetector) and recursive features elimination (RF-RFE), had been done by (Zhou et al., 2021) for factors optimization and enhancement of the predictive ability of the landslide susceptibility models. The finding shows that RF-RFE had the highest AUC score, which is 86.3% followed by the RF-GeoDetector (86.0%) and finally the single RF model (85.3%). All these three models had shown excellent performance. Somehow, even when the single RF had the lowest score, the difference of the AUC score with the highest score AUC is only 1% which is just a small difference with the RF-RFEs core. The reason for this occurrence is that RF has multiple decision trees. Given that each tree was constructed to partition the factors in pairs, the resulting prediction score was obtained through the aggregated responses of these trees. This approach enhances the probability of independence among the factors. Even if some of the

trees represented dependent factors, the remaining trees would suppress those and achieve the best results (Akinici & Zeybek, 2021).

The finding in Zhou et al. (2021) study not only portrays the strength of single RF model but also the performance gap of the RF model on the training set (accuracy=0.986) and the test set (accuracy=0.858) indicates that the RF model has limited generalization ability and induces over-fitting. Overfitting issues in RF have also been found in the study conducted by Park and Kim (2019) through their research with the objective of comparing and analyzing the landslide susceptibility in Woomyeon Mountain, South Korea. This is because the prediction rate curve's AUC values of RF were lower than those of the success rate curve, which had AUC values nearly equal to 1. There was a 20% decrease in the prediction rate's AUC value, particularly with the RF model. This outcome demonstrated how the training data overtrained the RF model. This may be related to higher errors for accurate data and poor generalization from training data. Despite the feature selection procedure, the noise was still present in the landslide explanatory variables. Besides, the landslide area is so small compared to the non-landslide area that the model could not learn and forecast the non-landslide area.

Based on the discussed past findings above related to RF, it can be concluded that the strength of RF is having high prediction performance (Hong et al., 2016; Hussain et al., 2022; Youssef & Pourghasemi, 2021), consisting throughout various and changing LCFs (Khaliq et al., 2022), resistance towards overestimation (Rabby et al., 2020) and having high independency towards factors (Akinici & Zeybek, 2021; Zhou et al., 2021). Somehow, it does experience the overfitting issues (Park & Kim, 2019).

2.5.5.2 Extreme Gradient Boosting

- **Structure**

XGBoost was derived from the efficient machine learning technique called the gradient tree boosting algorithm. Figure 2.3 shows the XGBoost structure where it is an iterative decision tree algorithm with multiple decision trees. Every tree is learning from the residuals of all previous trees. Rather than adopting most voting output results in Random Forest, the predicted output of XGBoost is the sum of all the results (Friedman, 2001, 2002). It is superior to standard gradient boosting since it includes a regularization parameter that prevents the model from becoming overfitting (Cao et al., 2020). Decision trees develop progressively in the standard boosting method. The boosting model minimizes the error by fitting decision trees on the residuals from the preceding tree rather than the result (Duan et al., 2021).

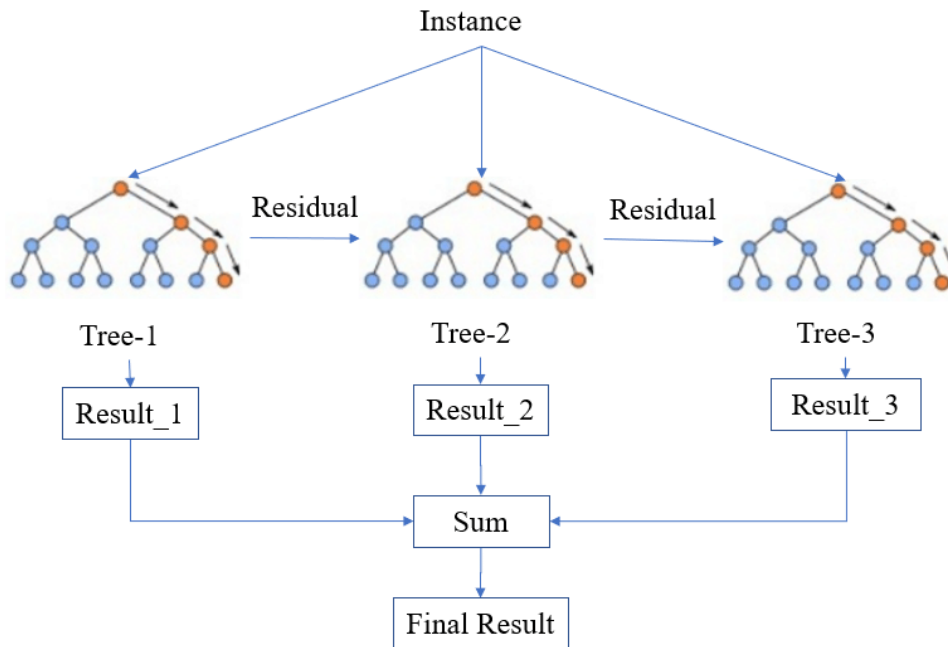


Figure 2.3: XGBoost Algorithm Structure

- **Hyperparameters**

Similar to RF, XGBoost performance efficiency depends on its hyperparameters. Table 2.3 lists the important hyperparameters and definitions according to Hussain et al. (2021), who used hyperparameters default values, and Wang et al. (2021), who utilized the BO method. A detailed explanation regarding the tuning methods will be written in the hyperparameter optimization techniques section.

Table 2.3: XGBoost Hyperparameters Listings

Hyperparameters	Definition
Eta	Booster parameter. This is the learning rate of the algorithm, and it is the step size shrinkage used in the update to prevent overfitting. It makes the model more conservative by shrinking the weights on each step.
Max_depth	The maximum depth of each tree in the model
Min_child_weight	It defines the minimum sum of weights of all observations required in a child. It is also used to control over-fitting.
Subsample	It denotes the fraction of observations to be randomly sampled for each tree. Lower values make the algorithm more conservative and prevent overfitting, but too small of values might lead to underfitting.
Gamma	Booster parameter. The node is split only when the resulting split positively reduces the loss function. Gamma specifies the minimum loss reduction required to make a split.
Colsample_bytree	The subsample ratio of columns when constructing each tree. Subsampling occurs once for every tree constructed.
nrounds	The number of trees

- **Strength and Limitations**

XGBoost has been proven to have great generalization performance and predictive ability since it possesses high AUC score in several LSM studies of multiple areas such as in Shigatse, Tibet (Zhang et al., 2019), Rangamati, Bangladesh (Rabby et al., 2020), Atartuk Dam, Turkey (Can et al., 2021) and Ayancik, Turkey (Sahin, 2020).

In the study of debris flow in Shigatse, Tibet (Zhang et al., 2019), the XGBoost with its AUC score of 98.8% had been outperforming 1D-CNN and RF, which both had the AUC of 97.6%. While XGBoost focuses on the residuals produced by the previous iteration, which can reduce both variation and bias, the RF algorithm concentrates on the final voting results of all DTs, which can minimize variance.

Furthermore, the landslide susceptibility mapping at Turkey's Ataturk Dam demonstrated that XGBoost had a strong prediction ability. Although, Can et al. (2021) conducted the LSM at the Ataturk Dam by incorporating lower spatial resolution of European Digital Elevation Model (EU-DEM), the XGBoost model resulted in an excellent AUC score of 96.0%. The high AUC value had been due to high prediction power of the XGBoost model.

Regardless of its high predictive ability, XGBoost was also shown to be prone to the occurrence of overestimation, where the highly susceptible area in the landslide susceptibility map is overestimated. XGBoost model classified more pixels as very highly susceptible; it increased the true positive rate while decreasing the true negative rate (Rabby et al., 2020; Zhang et al., 2019). Thus, this reduces its reliability in developing a final landslide susceptibility map despite its powerful prediction performance.

2.5.5.3 Stacking

- **Structure**

Due to their ease of programming and parameter selection, homogeneous classifiers are used in most ensemble approaches (Wang et al., 2019). The average error performance of all the classifiers in the ensemble is used to create the ensemble output. However, the performance of homogeneous ensembles is limited by the need for more diversity in classifier predictions resulting from similar types of errors. Research on heterogeneous ensembles was encouraged by a need for more diversity in forecasts.

One of the effective approaches in machine learning classification and regression problems is stacking. A common heterogeneous ensemble-learning technique called stacking or hybrid ensembles combines various base models with the aid of a metaclassifier or meta-learner to produce more accurate predictions (Shorfuzzaman, 2023; Wolpert, 1992). It can be considered a two-layer model, with the input layer comprising all the models and the prediction layer producing the output. First-level base learners use several different classification algorithms (Xiong et al., 2018). The meta-learner improves the stacking model performance in terms of its classification accuracy (Chaudhary et al. 2017; Gupta et al. 2020). The stacking of machine learning models has a few levels, as illustrated in Figure 2.4 (Divina et al., 2018) :

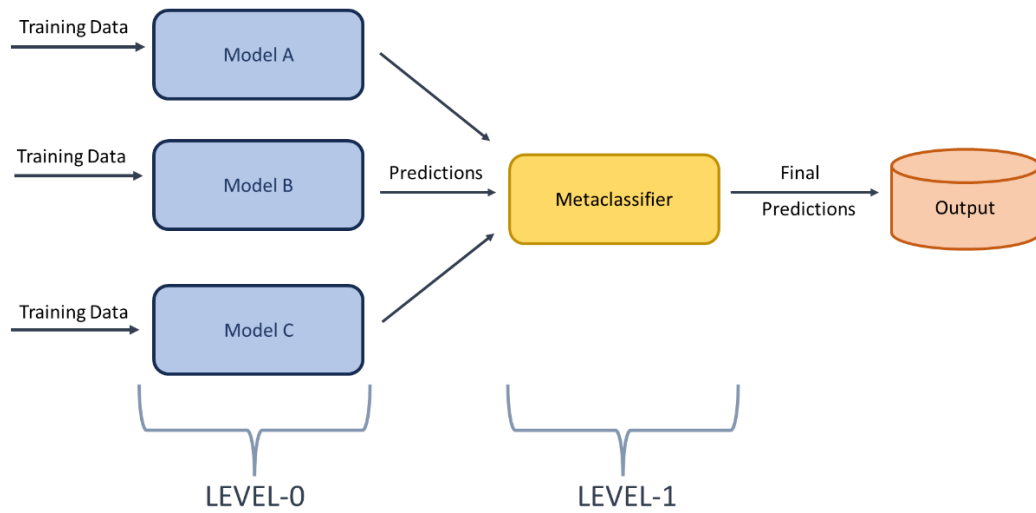


Figure 2.4: Stacking algorithm

- **Strength and Weakness**

The great performance of the stacked model has been proven in the latest study of LSM, such as the one conducted in Yanshan County, China, by Fang et al. (2020). To predict the susceptibility of landslides, the study examines four distinct heterogeneous ensemble-learning strategies, including stacking, blending, simple averaging (SA), and weighted averaging (WA). To produce accurate outcomes and avoid model selection problems, these methods combine several advanced classifiers, including LR, SVM, CNN, and RNN. The inventory comprised 380 historical landslide locations and 16 LCFs.

The results indicate that ensemble-learning algorithms outperform single classifiers in terms of prediction accuracy. The stacking and blending ensembles achieved higher prediction performance than other ensembles because they use meta-classifiers to correct the errors of the specific aspects that occur during the learning process of the base classifiers to improve their robustness and accuracy. Unlike stacking, blending model performance might be affected by information leakage during the K-fold cross-validation process of stacking.

Next, in the study conducted by Zeng et al. (2023) in Dazhou, Chongqing the current work focuses on two homogenous ensemble models, such as random forest (RF) and extreme gradient boosting (XGBoost), as well as three distinct fundamental classifiers, namely decision tree (DT), support vector machine (SVM), and multi-layer perceptron neural network model (MLPNN). To improve the spatial probability of landslide occurrence, the hierarchical building of the deep ensemble relied on the leading ensemble methods, which are bagging, boosting, and stacking ensemble techniques. The outcomes demonstrate that stacked models do not always perform better; for example, the DT-SVM-MLPNN-XGBoost model with bagging did not perform as well as the XGBoost model alone.

Similarly, the integration of DT-SVM-MLPNN performance is lower than the single RF and XGBoost. Nevertheless, the heterogeneous RF-XGBoost model, which is based on a stacking ensemble, showed the best performance with an AUC score of 95.5% in terms of its ability to forecast the areas affected by landslides. The overall result shows that stacking multiple base classifiers does not guarantee better performance compared to a single classifier. The selection of base classifiers and stacking techniques are important to ensure a great-performing model.

This finding obtained from the Dazhou, Chongqing (Zeng et al., 2023) is similar to the finding obtained from the LSM study in Kyushu Island, Japan (Dou et al., 2019), which incorporates the comparative analysis between stacking of SVM and single SVM. The final result showed that the single SVM model performed better than the stacked SVM, which only possessed an AUC score of 74%. Both studies in Dazhou, Chongqing, and Kyushu, Japan suggested that the selection of basic classifiers is important to ensure the best performance of the stacked models.

2.5.6 Summary of strengths and weaknesses of ML models

Based on the comparison of strength and weakness of the ML models in Table 2.4, it can be concluded that RF and XGBoost are the most suitable models for this research due to their high predictive accuracy, ability to handle non-linear data, and efficiency with tabular datasets. While other models like LR, SVM, and deep learning techniques have specific strengths, they are limited by issues such as lower accuracy, inefficiency with large datasets, or high computational demands. RF and XGBoost provide the best balance of performance and practicality for landslide susceptibility mapping.

Table 2.4: Summary of strengths and weaknesses of ML models

Model	Strengths	Weaknesses
Logistic Regression (LR)	• Good for addressing non-linear relationships between independent and dependent variables (Kavzoglu et al., 2019). • Simple and interpretable model.	• Unsuitable for complex non-linear problems (Zhou et al., 2018). • Limitations in generalization and simplification of landslide factors, affecting prediction performance (Wubalem & Meten, 2020).
Support Vector Machine (SVM)	• Excels with smaller samples, nonlinearity, and high-dimensional variables (Zhao et al., 2021). • Resistant to overfitting due to structural risk minimization.	• Unsuitable for large datasets and regional landslide susceptibility mapping (Zhao et al., 2021).
Deep Learning (e.g., CNN, RNN)	• High prediction accuracy for landslide susceptibility mapping (Zhang et al., 2019). • RNN can process sequential data effectively (Chen et al., 2015). • Suitable for large-scale studies involving extensive datasets and regions (Ngo et al., 2021).	• Limitations in handling tabular data (Zhang et al., 2019). • High computational requirements.
Random Forest (RF)	• High predictive performance (Sevgen et al., 2019). • Handles nonlinear relationships effectively.	• Can become computationally intensive with large datasets.
XGBoost	• Outperforms other models like CNN and RF in terms of prediction performance (Zhang et al., 2019). • Efficient for tabular data.	• Complex model that can be difficult to interpret.

2.6 Hyperparameter Optimization Techniques

In machine learning, hyperparameters are parameters that need to be set in advance during the learning process of a machine learning model, and algorithms are rarely hyperparameter-free. For all algorithms, some hyperparameters significantly impact the predictive power, and some hyperparameters significantly impact the predictive or model's accuracy. Therefore, hyperparameter optimization is particularly important. However, hyperparameter optimization is a combinatorial optimization problem. It is not optimized by gradient descent like normal parameters. The computation used to evaluate a set of hyperparameter configurations is very difficult because adjusting individual hyperparameters requires retraining to evaluate their effects.

Nowadays, hyperparameter optimization processes are performed using advanced methods that aim to find the best hyperparameters in shorter periods using a policy-based, informed search. Moreover, no initial setup or additional manual operations are required other than those conducted during the process. Some common strategies for optimizing hyperparameters, include Grid search, random search, and Bayesian optimization (Chang et al., 2019).

2.6.1 Manual search

Manual hyperparameter tuning involves experimenting with different sets of hyperparameters manually using the trial-and-error method. The results of each trial are tracked and used as feedback to obtain a combination of hyperparameters that yield the highest model performance. This method is also subjected to manually selecting values for hyperparameters based on prior knowledge or experience the prior knowledge of the researchers. For example, Khaliq et al. (2022) have manually set several hyperparameters

for RF, such as $\text{nodesize}=5$, $\text{ntree}=250$, and $\text{mtry}=4$ for the LSM study in Hattian Bala, Pakistan. Similarly, for XGBoost, Hussain et al. (2021) had manually set $\text{nrounds}=200$. Both Khaliq et al. (2022) and Hussain et al. (2021) presented studies setting the hyperparameters for RF and XGBoost based on the researcher's best assumption, and the findings had resulted in good performance of RF and XGBoost.

Besides, Chen et al. (2020) emphasized that the number of variables used (mtry) in each decision tree is also equal to the square root (classification model) or one-third (regression model) of the number of variables. Thus, Wang et al. (2020) had been setting their mtry for RF equal to 7 since it is one-third of the 22 LCFs incorporated in the study. After solving the mtry 's value, it is brought into the RF model for training. While the out-of-bag (OOB) error is stable, the minimum value of the abscissa is ntree . As the number of trees increases, the generalization error always becomes steady. Using this manual search, the RF had an AUC of 98.8%.

Setting the hyperparameters' values manually still can produce good performance of RF and XGBoost. However, giving every set of possible hyperparameters value a trial and error might be time-consuming. Moreover, in achieving the most accurate assumptions for optimal hyperparameter values requires extensive knowledge and experience in working with the models.

2.6.2 Grid Search

Grid search represents the most fundamental hyperparameter tuning method. This technique involves the creation, evaluation, and selection of a model for each possible combination of the provided hyperparameter values, aiming to identify the architecture that yields optimal results. While grid search can be universally applied across machine learning

to determine the best parameters for a given model, it encounters challenges related to dimensionality, particularly when the number of evaluated hyperparameters increases exponentially (Merghadi et al., 2020; Sameen et al., 2020).

Xia et al. (2022) had been using grid search method in finding the optimal values for the RF hyperparameters in the LSM study of Three Gorges Reservoir area. The finding shows the performance of RF on the training set was perfect, yielding an accuracy of 1, an F1 score of 1, and an AUC of 1, but it failed to achieve corresponding performance on the testing set (accuracy=0.908, F1 score=0.908, and AUC=0.9085), which revealed the occurrence of overfitting.

2.6.3 Bayesian Optimization

Bayesian optimization (BO) is a sequential model-based optimization (SMBO) methodology that enables the use of the outcomes of one iteration to enhance the sampling procedure for the subsequent trial (Pelikan, 2005). 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). The 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 is the Gaussian Process Upper Confidence Bound (GP-UCB), one of the most used acquisition functions.

The LSM study performed by Sun et al. (2021) has implemented BO incorporating LR and RF, in a frequent landslide-prone area in Fengjie County, China. The LR model's AUC value of the test dataset is improved by 4% based on the Bayesian algorithm. In comparison, the AUC value of the test dataset in the RF model is enhanced by 10%, showing that both models' hyperparameter optimization based on the Bayesian algorithm has significantly improved the models' accuracy.

Similarly, Wang et al. (2021) had optimized RF and XGBoost using BO for the LSM study in Loess Plateau of China. The final finding had shown that both RF and XGBoost prediction performance improved after optimization. BO had also reduced the overfitting rate in RF by 4%. Thus, BO efficiently improves prediction performance and reduces overfitting in RF and XGB.

2.6.4 Summary of the strength and weakness of various optimization methods.

Based on the strength and weakness of various optimization methods in Table 2.5, it can be concluded Bayesian Optimization (BO) stands out as the most effective hyperparameter optimization method for this study due to its ability to efficiently explore the parameter space and improve model performance. Unlike Manual Search, which relies heavily on trial-and-error and prior experience, BO systematically identifies optimal parameters, reducing the need for extensive manual input. Similarly, it surpasses Grid Search by avoiding the computational expense and overfitting risks associated with exhaustive evaluations of all parameter combinations. BO's demonstrated success in enhancing prediction accuracy and reducing overfitting in machine learning models makes it the ideal choice for optimizing hyperparameters in this research.

Table 2.5: Summary of strengths and weaknesses of various hyperparameters optimization methods

Method	Description	Strengths	Weaknesses
Manual Search	Manually selecting hyperparameter values based on prior knowledge or experience (Khaliq et al., 2022).	• Simple to implement (Khaliq et al., 2022). • Can produce good performance if prior knowledge is sufficient (Khaliq et al., 2022).	• Time-consuming due to trial-and-error process (Khaliq et al., 2022). • Requires extensive knowledge and experience (Khaliq et al., 2022).
Grid Search	Systematically evaluates all possible combinations of hyperparameters (Xia et al., 2022).	• Universally applicable across machine learning models (Merghadi et al., 2020; Sameen et al., 2020).	• Computationally expensive with increasing hyperparameter dimensions (Xia et al., 2022). • Prone to overfitting (Xia et al., 2022).
Bayesian Optimization	Sequential model-based optimization using Gaussian processes to guide hyperparameter selection (Pelikan, 2005; Wu et al., 2019).	• Efficient in exploring parameter space (Wu et al., 2019). • Reduces overfitting and improves model accuracy (Sun et al., 2021; Wang et al., 2021).	• Computational complexity increases with high-dimensional spaces (Wu et al., 2019).

2.7 Selection of Landslide Susceptibility Prediction Models

There are a few criteria needed to be considered in selecting an appropriate approach to be used, which are:

- The chosen approach must successfully deal with the non-linearity in the relationship between the landslides conditioning factors and the occurrences of landslides (Mezaal et al., 2017).
- The method of choice must handle geographic heterogeneity, consume numerous input parameters, and manage many output parameters such as classification and regression (Taalab et al., 2018).
- Selecting multiple machine learning models has an advantage over single models in that the hybrid-ensemble approach tends to focus more on the strengths of each model and minimize the weaknesses (Chen & Li, 2020).
- The selection depends on the study's scale and size of the study where regional and national study scales which cover larger areas will require more advanced ML models (Khosravi et al., 2020; Ngo et al., 2021).

Multiple factors have various degrees of influence on landslides. As a result, landslide possesses the traits of imbalance, nonlinearity, and randomness (Wang et al., 2020). The nonlinear relationship between the landslide conditioning factors and the landslide susceptibility can be addressed by the RF and XGBoost models (Sahin, 2022). Additionally, RF and XGBoost have been proven to have great prediction performance in regional scale studies across various environmental factors in different study areas (Arabameri et al., 2021). Thus, these two models are suitable for regional LSM studies.

However, the great accuracy came at the expense of overfitting, which made the model perform poorly in real-world scenarios (Hong et al., 2017; Park & Kim, 2019). On the contrary, despite its strengths, XGBoost has exhibited a susceptibility to overestimation. This tendency results in the production of a landslide susceptibility map that is less reliable,

even in instances where XGBoost outperforms Random Forest (RF) in terms of prediction accuracy (Agrawal & Dixit, 2023; Rabby et al., 2020). The challenges of overfitting encountered by RF and overestimation observed in XGBoost underscore the need for improvement to enhance their overall performance.

2.8 Challenges in developing Landslide Susceptibility Maps

- **Overfitting**

RF had successfully solved the overfitting problems in decision trees, but the study of Park and Kim (2019) showed that RF is still experiencing overfitting problems. When exposed to new datasets, a model that fits exceptionally well during training may not necessarily generalise well. Overfitting is the term for this issue, in which the model's accuracy as it was trained fails to correspond to its absolute accuracy on data that is not seen (Ghasemian et al., 2019).

The causes of overfitting can be numerous, such as complex models. Using an overly complex model for a simple task can lead to overfitting. For example, insufficient data. If there is insufficient data, the model might find patterns that do not exist (Rong et al., 2021). Lastly, the presence of noisy data poses a significant challenge. In instances where the training data incorporates errors or random fluctuations, an overfitted model may erroneously interpret these elements as meaningful patterns (Lee et al., 2018).

RF models encounter a bias-variance dilemma in prediction scenarios, whereby a rise in model complexity may lead to improved accuracy and a high model variance. Conversely, Cruz-Ramírez et al. (2014) found that lowering model complexity improves model generalisation, which is high bias but lowers accuracy. Therefore, the RF model's complexity

must be decreased to minimize overfitting. For some reason, it is challenging to maintain the models' simplicity while still ensuring that the predictions are accurate. This is because machine learning often performs better as it becomes more complicated. Creating an optimum model with good accuracy and reduced complexity is crucial, and an advanced hyperparameter optimization method, like BO, can do this (Kadam & Jadhav, 2020; Sun et al., 2020).

- **Overestimation**

Although XGBoost is less likely to undergo overfitting, it is susceptible to overestimation during the LSM development process. Even in cases where the model performs well in predictions, overestimation is still likely to happen, as proven in the studies of Agrawal and Dixit (2023), Rabby et al. (2020), and Zhu et al. (2019). Overestimation is where XGBoost model overexpanded landslide susceptibility and could not distinguish between high and low-susceptibility zones (Zhu et al., 2018). This often being caused by spatial autocorrelation where the overestimation in XGBoost model increased the actual positive rate while decreasing the valid negative rate by classifying more pixels as highly susceptible (Rabby et al., 2020; Zhu et al., 2019). Consequently, it will develop less reliable LSM, which might lead to time and cost inefficiencies for the local government and town planner, regardless of its powerful prediction performance. Unfortunately, no attempt has been made to improve the overestimation issue in XGBoost.

Nevertheless, the overestimation problem has been experienced in other models, such as in the study conducted in Chittagong Hill, Bangladesh (Rabby & Li, 2020). The study provided regional maps of landslide susceptibility for Bangladesh's Chittagong Hilly Areas, which incorporated the stacking of frequency ratio (FR) with logistic regression (FR_LR)

and the analytical hierarchy process (AHP) (FR_AHP). Thirteen landslide conditioning factors and 730 landslide locations were employed in a landslide inventory.

The prediction rate curves indicate that FR_LR has the greatest AUC at 79.46%, with FR_AHP coming in second at 77.15% and finally the single models of FR at 72.5%. In addition, not only does the FR have the lowest performance, but it also experiences overestimation issues due to its relatively low true-negative rate. Since most landslides occur in high or very high susceptibility zones, a practically feasible and consistent susceptibility map would identify the least amount of land in these zones (Huete et al., 2002; Rabby et al., 2020). The landslide susceptibility map's practical relevance is diminished by overestimating the sensitivity to landslides.

The integration of FR_LR lessened the overestimation problems. It produced a consistent and practical map while also improving accuracy. This is because it eliminated the overestimation of the southeast region of the study area's landslide vulnerability. Thus, this study has proven and given the idea that the integration of two different learners can overcome the overestimation problem. Integrating two learners does not always perform better than the single model. This is because the compatibility of the chosen base classifiers for the stacking model is very important.

- **Proposed Solution for the Issues of Overfitting and Overestimation**

Thus, the study of Rabby and Li (2020) gave the idea of where stacking the XGBoost with other ML models might minimize the overestimation problem. Since RF is prone to overfitting but resistance to overestimation while XGBoost is resistance to overfitting and prone to overestimation, both RF and XGBoost might be compatible to be stacked together. This is because the strength of RF can cover the overestimation issue in XGBoost and vice

versa. The strength of XGBoost can help with the overfitting issue in RF. In addition, implementing advanced hyperparameter optimization methods such as the Bayesian Optimization method will help improve the overfitting issues, as stated by Kadam and Jadhav (2020) and Sun et al. (2020).

Thus, in this study it was a proposed solution to develop hybrid ensemble RF-XGBoost with Bayesian hyperparameter optimization (BO). In this study, the hybrid ensemble RF-XGBoost followed by Bayesian optimization refers to a stacking of two ensemble machine learning models, RF and XGBoost, and further developed using the Bayesian hyperparameter optimization technique. The term ‘hybrid’ reflects both the stacking of two models to benefit from their complementary capabilities as well as the combination of these models with an optimization technique so that the models’ hyperparameters are optimized. The goal of this method is to improve the accuracy of prediction and at the same time overfitting is decreased and a more robust landslide susceptibility model is developed.

2.8.1 Summary on the challenges in developing LSM and the proposed solution

To conclude, this thesis focuses on addressing the issues of overfitting and overestimation in RF and XGBoost, with an emphasis on implementing the proposed solutions, such as hyperparameter optimization, and model stacking, to enhance model generalization and reliability as summarized in Table 2.6.

Table 2.6: Summary of strengths and weaknesses of various hyperparameters optimization methods

Challenges	Causal Factor	Proposed Solution	Citations
Overfitting	• Bias-Variance dilemma • Model's complexity	• Reduce model complexity by implementing Bayesian hyperparameter optimization method (BO).	Ghasemian et al., 2019; Park & Kim, 2019; Kadam & Jadhav, 2020; Sun et al., 2020
Overestimation	• Spatial autocorrelation	• Stacking XGBoost with RF to overcome overestimation.	Rabby & Li, 2020; Agrawal & Dixit, 2023; Rabby et al., 2020; Zhu et al., 2018, 2019

2.9 Performance Evaluation Methods

2.9.1 Area Under the Curve of Receiver Operating Characteristics

Validating the model's performance is crucial and required in landslide susceptibility mapping (Wang et al., 2020). The most popular technique for accurately assessing a landslide's vulnerability is the receiver operating characteristic (ROC) curve. Bui et al. (2020) state that the ROC curve measures continuous data specificity and sensitivity variables. The AUC-ROC shows the model's accuracy. The model's prediction accuracy is higher when the AUC value is closer to 1. The AUC can be grouped into success rate and prediction rate AUC, which usually being used to observe the rate of overfitting occurrence in the model.

- **Success rate and Prediction rate AUC**

The success rate curve shows how well-suited the developed model is for assessing landslide susceptibility (Pham et al., 2017). A comparison between the number of landslides utilized in the training dataset and the LSMs is used to generate the success rate curve (Pradhan et al., 2010). Prediction rate curves help evaluate the probability that landslides will occur. As such, the model can be utilized to validate a model's predictive ability (Brenning, 2005). Predictive rate curve is generated using employs the testing dataset (Fabbri et al., 2002).

AUC is a graph with varying index values that typically range from 0.5, or 50%, to a maximum value of 1, or 100%. Using minimum AUC parameters with a minimum limit of 60% is recommended. Based on the AUC, 90% indicate an excellent model classification, 80%–90% indicate a good model classification, 70%–80% indicate a medium or decent

model classification, and $< 60\%$ indicate a poor one. The influence of the landslide event increases with a parameter's AUC value (Pradhan, 2013).

- **DeLong Test**

DeLong's test is a non-parametric statistical method used to compare the area under the receiver operating characteristic (ROC) curves (AUC) of two or more correlated models (DeLong et al., 1988). It is particularly suited for cases where the same dataset is used to generate the ROC curves, accounting for the inherent correlation between the models. The test evaluates whether the observed difference in AUCs is statistically significant, providing a robust measure of the models' discriminative performance.

In practice, DeLong's test is widely used in machine learning and statistical research to assess the predictive abilities of models, especially when comparing their ability to distinguish between positive and negative outcomes. The test outputs a p-value, which determines the statistical significance of the AUC difference. If the p-value is lesser than 0.05, the difference in AUCs is considered statistically significant, indicating that one model has superior discriminative performance. Conversely, if the p-value is greater than or equal to 0.05, the difference is not statistically significant, suggesting that the models perform comparably in terms of AUC.

- **Overfitting**

Overfitting is when the model becomes so good at training datasets, making the model perform very well with training data but poorly with test or validation data (Ying, 2019). Thus, if the model's AUC score of success rate is extremely better than the prediction rate it is confirmed that the model is experiencing overfitting (Hong et al., 2016; Park & Kim, 2019; Silalahi et al., 2019; Wubalem & Meten, 2020)

2.9.2 Confusion matrix and Statistics

At the same time, numerous studies such as Daviran et al. (2023), Nnanwuba et al. (2022), Sun et al. (2021), Wang et al. (2021), and Wang et al. (2023) had utilized the confusion matrix accuracy, precision, sensitivity, specificity, and kappa to evaluate the model predictive performance. Figure 2.5 shows the confusion matrix table, meanwhile the accuracy, precision, sensitivity, specificity, and kappa are calculated as shown in Equation 2.4 to Equation 2.5.

	Actual Positive	Actual Negative
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (TN)	True negative (TN)

Figure 2.5: Confusion matrix

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{Equation 2.4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Equation 2.5}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Equation 2.6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad \text{Equation 2.7}$$

By referring to the confusion matrix (Figure 2.5) and the Equation 2.4 to Equation 2.7, the stated FP and FN are the numbers of landslides that were mistakenly identified, whereas TP (and TN are the numbers of correctly classified landslides. These values range from 0 to 1 or 0%-100% for accuracy, precision, sensitivity, and specificity. As these values increase, the model's performance improves.

$$\text{kappa, } k = \frac{p_o - p_e}{1 - p_e} \quad \text{Equation 2.8}$$

Where based on Equation 2.8:

P_o = Observed Agreement

P_e = Expected Agreement

Meanwhile, the Kappa index ranges from 0 to 1 and can be categorized into: almost perfect (0.8–1.0); substantial agreement (0.6–0.8); moderate agreement (0.4–0.6); fair agreement (0.2–0.4); slight agreement (0.0–0.2); and poor (<0.0) (Landis & Koch, 1977).

2.10 Conclusion

In summary, this chapter covers the reviews on the LCFs followed by the landslide inventories, machine learnings used in LSM studies, hyperparameter optimization challenges in developing LSM, and ends with the model's performance evaluation methods.

While choosing the LCFs, it is very important to check the significant factors of the landslide events of the study area. The LCFs selection can further be carried out by utilizing the multicollinearity test and feature selection method, such as Boruta algorithm to avoid collinearity issues which will destabilize the prediction model. Next, for the landslide inventories, it is crucial to ensure that both landslides and non-landslides datasets are balanced, which can be achieved using the LaGRiSU.

In addition, a comprehensive review of the evolution of machine learning approaches in landslide susceptibility mapping (LSM) revealed that machine learning approach generally offer improved accuracy and reliability over traditional techniques. After analyzing various machine learning models applied in LSM studies, it was concluded that ensemble methods, such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost), consistently demonstrate superior predictive performance.

Nevertheless, there is no such perfect machine learning model as there will always be challenges in developing the LSM. Previous researchers have widely used RF and XGBoost for the past few years, and regardless of their great performance in predicting landslides, there are still drawbacks. Most of the studies conducted using these two models focused on prediction performance and comparative studies. However, issues such as overestimation experienced by the XGBoost and overfitting in RF have been overlooked.

Thus, in this current study, the overestimation and overfitting issues as stated in section 2.8, will be addressed and planned to be improved through the stacking of RF-XGBoost and further enhanced by implementing BO for the hyperparameter tuning of the two models.

CHAPTER 3

MODELLING AND IMPLEMENTATION

3.1 Overview

In this chapter, a detailed methodology that is proposed for this research project is laid out. Firstly, the chapter explains the Penang Island whose landslide conditioning factors and data have been acquired. The selected data, comprising mainly of the landslide locations topographical, hydrological, environmental, and anthropogenic factors used in this project, is stated and explained along with its significance. It also describes how the data has been pre-processed and arranged to make it suitable for operation. The second portion of this chapter describes the proposed implementation methodology for landslide susceptibility prediction. This includes the algorithms, the hyperparameter optimization, and the models' performance evaluation. The final part includes developing landslide susceptibility maps from the generated models.

3.2 Research Layout

The methodology was started with data collection, then the development of spatial databases for landslide conditioning factors, the development of an inventory map for landslides, followed by the evaluation of the significance of landslide conditioning factors (LCFs), the modelling and validation of landslide susceptibility models, and finally the development of the final landslide susceptibility map (Figure 3.1). The landslide susceptibility maps used in this study were produced from various data sources. This study fully utilized open-source tools such as RStudio and GIS, including ArcMap 10.4 and QGIS 3.18. The information utilized in this study was gathered from public sources.

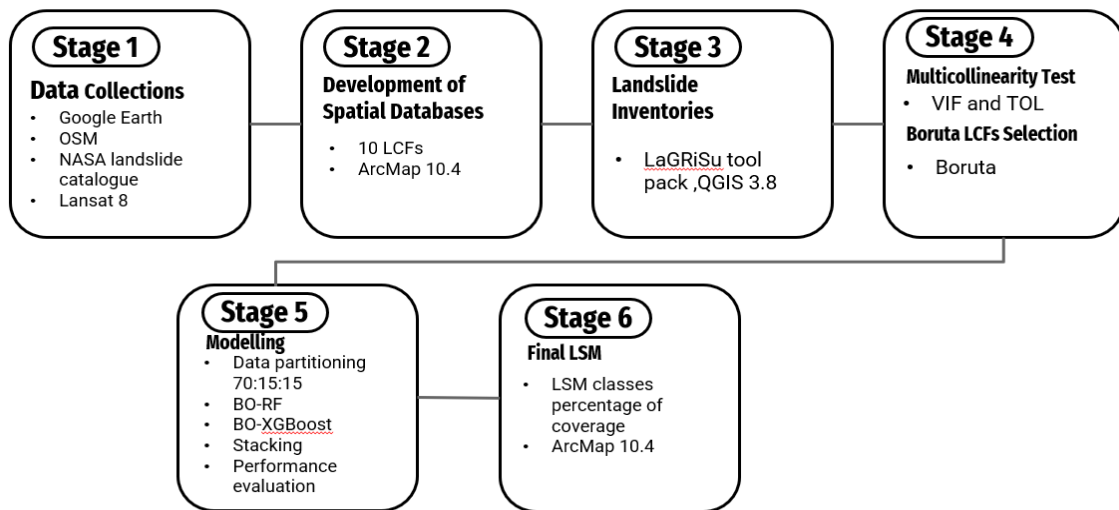


Figure 3.1: Research Methodology Flowchart

3.3 Study Area

The Penang Island is situated northwest of Peninsular Malaysia in the Strait of Malacca. A 301.248 km² region roughly describes the size of the island (Figure 3.2). Between latitudes 5°15'N and 5°30'N and longitudes 100°10'E and 100°20'E is about where the study area is located. The terrain of the island is notably hilly, with steep slopes covering around half of the territory at elevations as low as 18 metres.

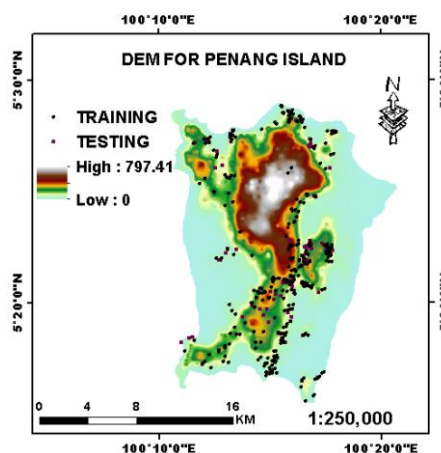


Figure 3.2: Penang Island

The Timeline of Landslides in Malaysia in 2022 had reported several landslide cases happening in the Penang Island as on the 19th of October 2022, and it was reported that the recent bad weather has caused small landslides and fallen branches at various spots on Penang Hill. In addition, on the 21st of June 2021, it was also reported that in Tanjung Bungah, Penang, an excessive rainfall run-off caused a flash flood of muddy water in front of the Surin Condominium in Tanjung Bungah. This resulted from cutting into a steep slope during the extension of Solok Tanjung Bungah along a 250ft contour. Then on the 21st of November 2018, a minor landslide took place along Jalan Paya Terubong between Green Garden and the Jalan Tun Sardon junction.

3.4 Landslide Conditioning Factors Spatial Databases

This study determined ten landslide causative factors to be applied in developing landslide susceptibility models and forecasting the distribution of landslide susceptibility. The topographical features include slope angle, slope length, plan and profile curvatures, and DEM. The hydrological factors comprise the SPI, the TWI, and the distance from the stream. The NDVI and distance from roads round out the environmental component lists.

3.4.1 Topographical Factors

DEM is a vital tool for understanding topography, which is also a crucial input for topography in precise modelling and LSM prediction. The quality of the DEM directly affects the generated factors such as slope angle, slope length, and curvatures (Green, 1988). In this study the DEMs were generated from the elevations of Penang Island in few stages as stated below:

- The elevation points were extracted from Google Earth and stored in the "kml" file format (Figure 3.3).
- Then, by visiting the website (www.gpsvisualizer.com), the "kml file" was converted into a "GPX file". It was ensured that the elevation information was added to the data while the data conversion took place. The elevation "GPX" file was exported into ArcMap 10.4 via the "GPX to Features" conversion tools as a shapefile. It was ensured that each elevation point's elevation value appeared in the attribute table (Figure 3.4).
- The spatial analyst "Interpolation Tool," which is the Inverse Distance Weighted (IDW), was used to create the DEM with an output cell of 30 m x 30 m. The coordinate system of the layer was set to the CGS_WGS_1984_47N.
- Finally, by using the DEM layer as the input data, the Spatial Analyst Surface Tool created slope angles, plan, and profile curvatures of the study area.

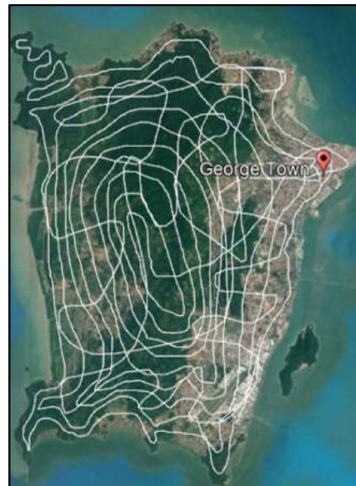


Figure 3.3: Extraction of elevation points from Google Earth Pro



Figure 3.4: Exported Elevations shapefile and the attribute table with the elevation values

Penang Island's DEM falls in the range from 0 m to 797.41 (Figure 3.11a), and the slope angle was in the range of 0° to 60.44° (Figure 3.11b). Meanwhile, the plan (Figure 3.11c) and profile curvature have the ranges of (-5.08 to 4.86) and (-7.59 to 8.29), respectively (Figure 3.11d). Finally, the slope length ranges from 0 to 124.02 (Figure 3.11e)

3.4.2 Hydrological Factors

Hydrological factors, such as the distance from a stream, the TWI, and the SPI, are water-related factors that affect the likelihood of erosion and landslides. The distance from stream, TWI, and SPI was developed by firstly producing the fill layer, then flow direction and flow accumulation layers using the Hydrology tool (Figure 3.5).

- **Distance from stream**

According to Chowdhury (2023), the development of the distance from stream map includes the steps as follows:

- I. The development of the flow direction layer, which was produced using fill as the input raster. The flow direction function takes a surface as an input and

creates a flow direction raster from each pixel to its steepest downslope neighbour.

- II. Next, by using the flow direction as the input, flow accumulation was produced. The flow accumulation tools determined the cumulative flow as the total weight of overall cells flowing into respective downslope cells in the output raster.
- III. Flow accumulation value was conditionally divided to extract the higher-order streams.
- IV. Then, the stream link, the stream order, and the stream network layers (Figure 3.6) were extracted by using the "Stream Link", "Stream Order", and "Stream to Feature" tools from the Hydrology toolset. Stream links were the stream channel sections connecting two adjacent junctions, an outlet, or a drainage divide; meanwhile, the stream order acts as the stream identifiers based on tributaries.
- V. The "dissolve" tool was used to dissolve the stream network layer, which was then used as an input layer for the Euclidean distance tool, which produced the final distance to the stream layer.

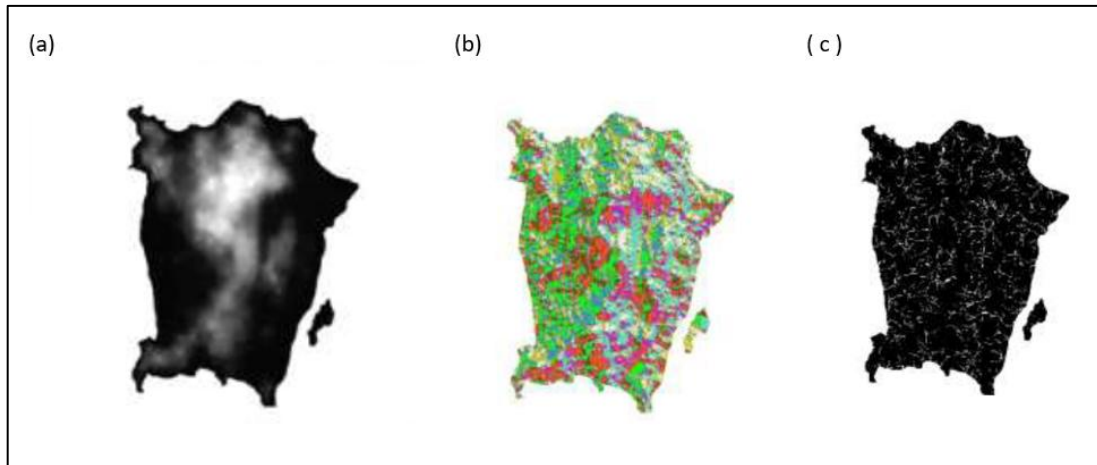


Figure 3.5: (a) Fill layer (b) Flow direction layer (c) Flow accumulation layer

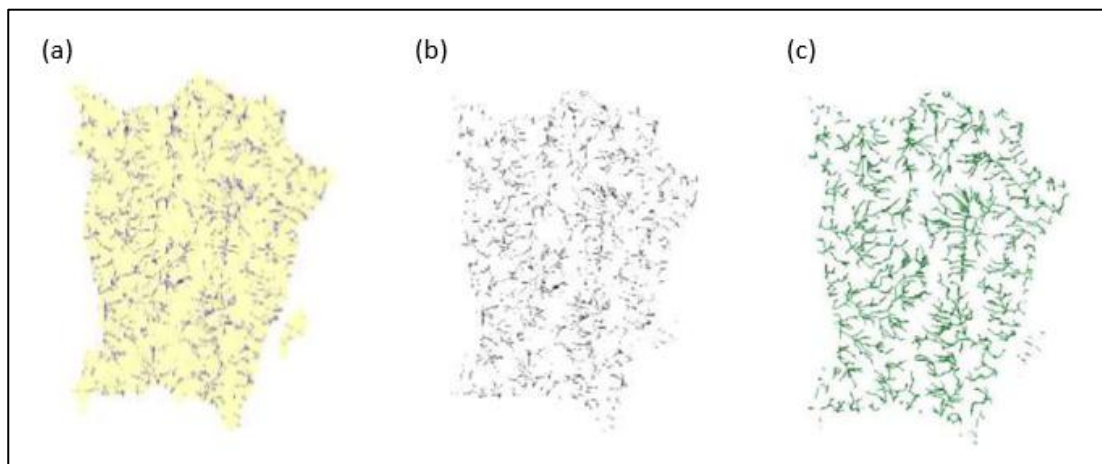


Figure 3.6: (a) Stream link (b) Stream order (c) Stream network

The distance from stream ranges for Penang Island was 0 m to 785.18 m (Figure 3.11g). Water can cause saturation sediments, decreasing the integrity of the slope and causing it to slope movement or mass movement (Tien Bui et al., 2016).

- **Topographic Wetness Index**

TWI exhibits a topographic effect at the site of the saturated area size of runoff generation, so it can be said that lower TWI values and higher slope length and SPI values represent higher landslide susceptibility (Moore et al., 1991; Oh et al., 2018). The development of TWI then involves several calculations using raster calculator tool in

ArcMap of “ Radian of Slope”, “Tanslope”, “Scaled flow accumulation” and the final calculation of TWI as shown in Figure 3.7 and Equation 3.1 to 3.4 (Beven & Kirkby, 1979; Chowdhury, 2023). The TWI range for each study area was 2.97 to 25.95 (Figure 3.11j).

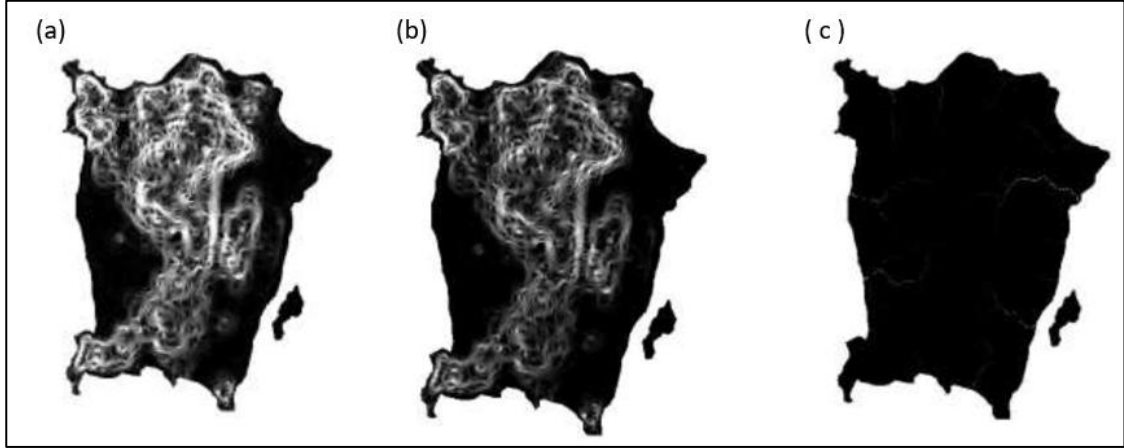


Figure 3.7: (a) Radian slope layer (b) Tanslope layer (c) Scaled flow accumulation layer

$$\text{Radian of Slope} = \frac{(\text{Slope angle} * 1.570796)}{90} \quad \text{Equation 3.1}$$

$$\text{Tanslope} = \text{con}(\text{"Radian of slope"} > 0, \text{Tan}(\text{"Radian of Slope"}), 0.001) \quad \text{Equation 3.2}$$

$$\text{Scaled flow accumulation} = (\text{"flow accumulation"} + 1) * 30 \quad \text{Equation 3.3}$$

$$\text{TWI} = \text{Ln}(\text{"Scaled flow accumulation"} / \text{Tanslope}) \quad \text{Equation 3.4}$$

- **Stream Power Index**

SPI, a measure of the flow erosion power of the stream at the given point of the topographical surface, which can control the initiation of landslides (Jaafari et al., 2014; Rahmati et al., 2019). For SPI, the calculation is as shown in the equation below, resulting

in SPI in the ranges of 0 to 117894 (Figure 3.11i). The calculations were calculated using the raster calculator and the equations below for SPI (Equation 3.5) (Ahmad et al., 2019) development.

$$SPI = \text{"Flow Accumulation"} * 30 - \tan(\text{"slope degree"} * 0.017453) \quad \text{Equation 3.5}$$

3.4.3 Environmental and Anthropogenic Factors

Environmental and anthropogenic factors include the distance from roads and NDVI. Meanwhile, Open Street Map was used to import road data into ArcGIS. The road data was in polyline type (Figure 3.8). The distance from the road was then calculated by using Euclidean Distance tool. The distance from the road to Penang Island falls from 0 m to 1561.15 m (Figure 3.11f).

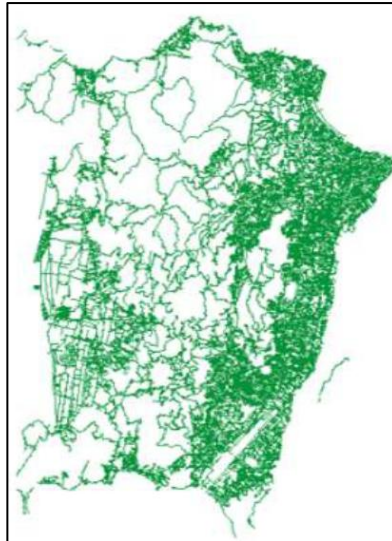


Figure 3.8: Open Street Map polyline road data

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 a better capture, the cloud cover was

ensured to be less than 20% (Figure 3.9). Only bands 4 and 5 of the seven bands on the satellite images were imported into ArcMap (Figure 3.10). Equation 3.6 was used in the calculation of the NDVI layer. (Hashim et al., 2019). The NDVI of study area falls in the range of -0.17 to 0.59 (Figure 3.11h).

$$\text{NDVI} = (\text{BAND 5} - \text{BAND 4}) / (\text{BAND 5} + \text{BAND 4}) \quad \text{Equation 3.6}$$

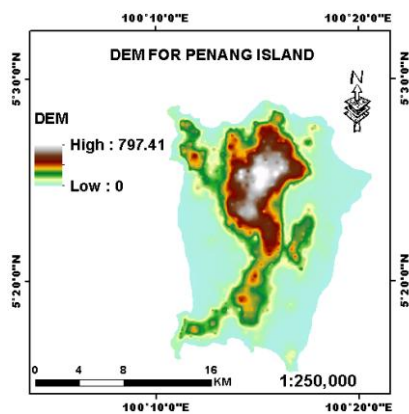


Figure 3.9: Landsat 8 Image from the USGS Earth Explorer

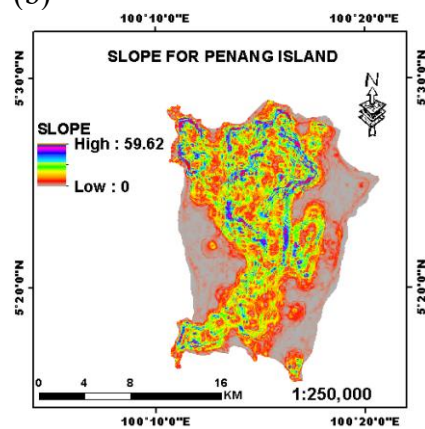


Figure 3.10: Penang Island Landsat 8 image of Band 4 and 5 in ArcMap

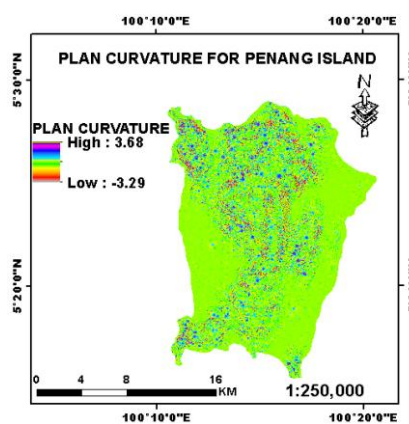
(a)



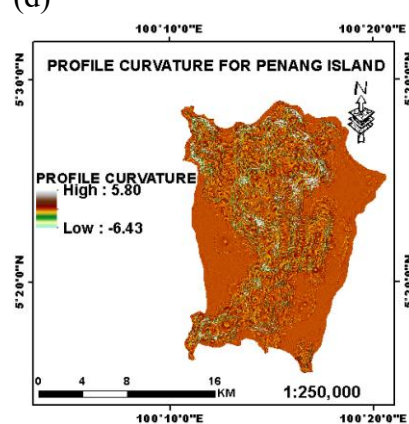
(b)



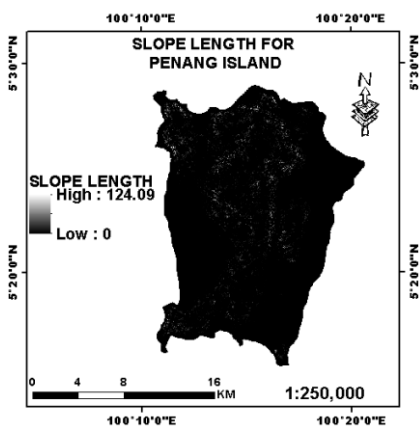
(c)



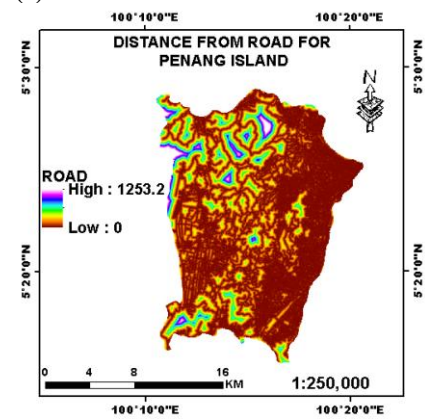
(d)



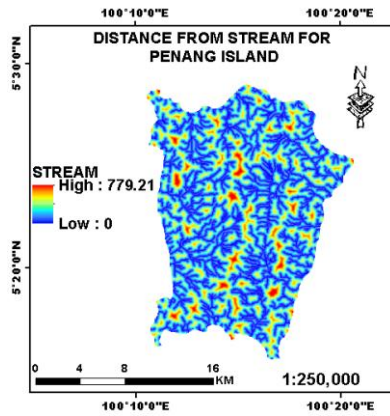
(e)



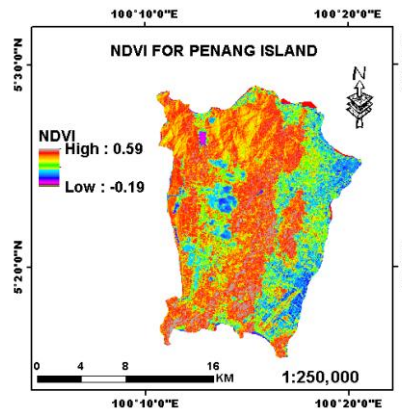
(f)



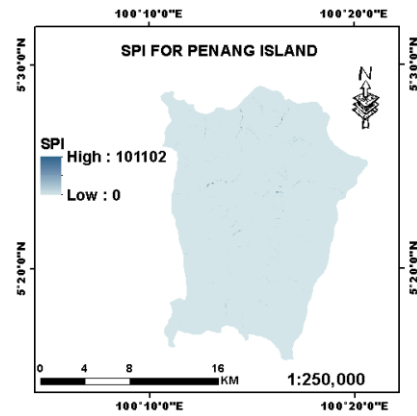
(g)



(h)



(i)



(j)

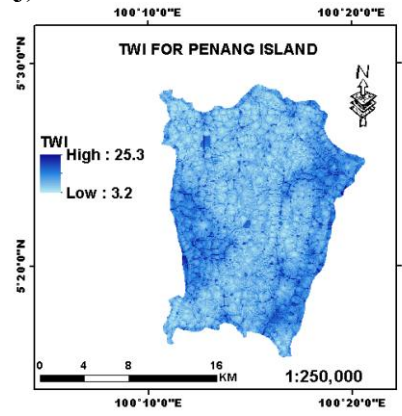


Figure 3.11: LCFs (a)DEM (b) Slope angle (c) Plan Curvature (d) Profile Curvature (e) Slope Length (f) Distance from Road (g) Distance from stream (h) NDVI (i) SPI (j)TWI

3.5 Landslide Inventories Map

A hazard inventory keeps track of a geological hazard body's location and time of occurrence, as well as the several sorts of hazards that can be identified in a given area (Pašek, 1975). The historical geohazard literature published between 2006 and 2021 (Elmahdy et al., 2016; Han et al., 2021; Lee & Pradhan, 2006) as well as the open-source NASA Landslide Catalogue website (<https://gpm.nasa.gov/landslides/index.html>, accessed on 30th March 2021) were used to gather the landslide data for both study areas covered in this paper. The landslide inventory map also featured the most recent incidents reported by news outlets, including The Stars and The Channel NewsAsia. Penang Island has a total of 443 landslide spots.

In the present investigation, 443 landslide representations of previous landslides, a LaGriSu tool pack (Al-Thwaynee et al., 2023) in QGIS 3.18, and a digital elevation model (DEM) layer were used to easily extract binary features with values of 1, which were the landslide points and 0 non-landslide points. In the process of grid unit extraction, the slope threshold that did not experience any landslides, known as a safe area, was first determined by extracting the distribution of landslides on the slope (Figure 3.12).

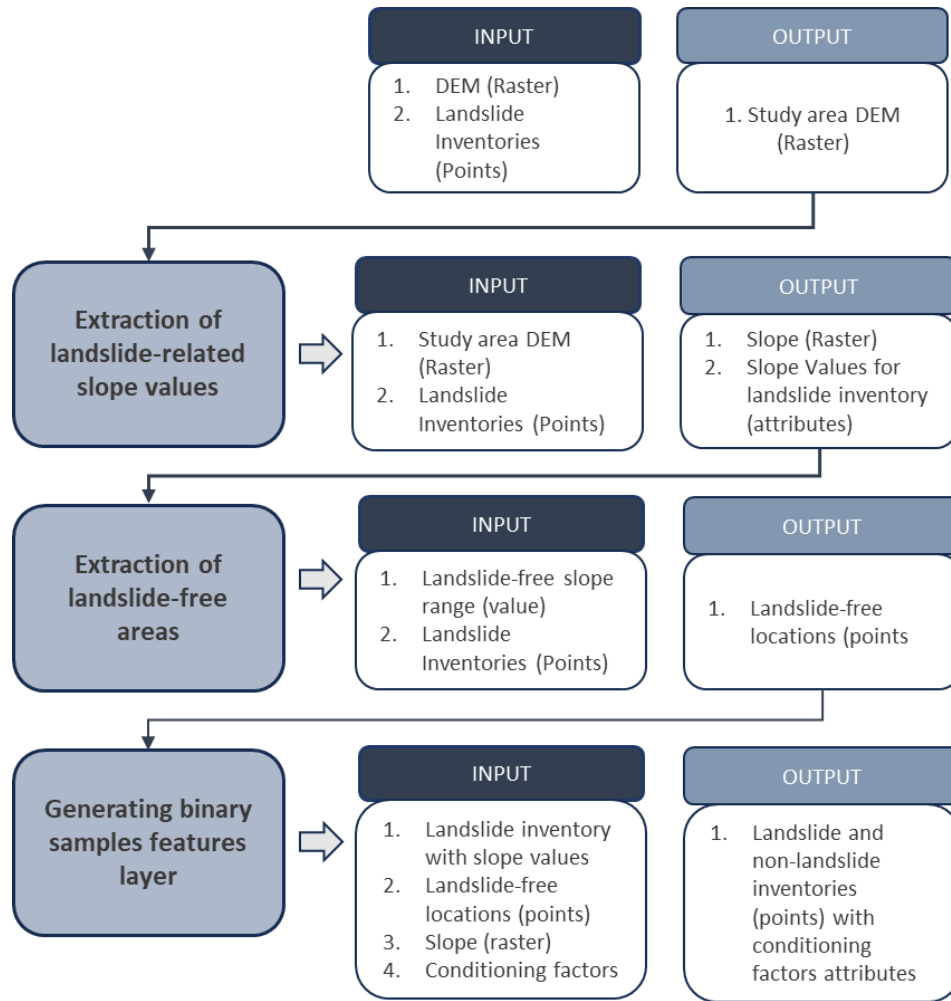


Figure 3.12: Grid Unit Extraction via LaGriSu Tool packs process

Next, the non-landslide points were extracted randomly within the slope landslide-free zone (Huang et al., 2017). The number of points with no landslides was equal to those with landslides to prevent the problem of class imbalance (Chen et al., 2017; Park & Kim, 2019; Sahin, 2020). Then, the parameters of the 443 points of the landslide-free inventories, the landslides within the slope value, and the slope raster were integrated to determine the training and testing zones. The total points of both landslides and non-landslides after the extraction were 886. Finally, the values for modelling landslide susceptibility were extracted for all 10 conditioning components and attached to the extracted datasets.

3.6 Modelling

The modelling process was carried out in the RStudio Software of version 4.3.0. The multicollinearity test, feature selection using the Boruta model, and frequency ratio (FR) between variables and landslide occurrence were carried out using the whole dataset. The landslides datasets and all the 10 spatial databases were imported into R. Datasets for both the study areas were split into 70% of training, 15% of testing datasets, and 15% of validation datasets randomly. There were no exact criteria or rule of thumb on the ratio selection for the dataset splitting (Achour & Reza, 2020; Hussain et al., 2021; Saha et al., 2021; Sahin, 2022). The unwelcomed attributes in the data were removed, and the data was also scaled. After the data pre-processing, proceed to the RF and XGBoost modelling with the bayesian hyperparameter optimization and finally perform the model stacking and development of the final landslide susceptibility map.

3.6.1 Multicollinearity Test

The variance inflation factor (VIF) and tolerance (TOL) approaches were utilized in this study to test for multicollinearity among the variables. Similarly, prior studies frequently used the VIF and TOL methods (Ado et al., 2022; Amiri et al., 2019; Chen et al., 2018; Gao et al., 2020; Rabby & Li, 2020; Saha & Saha, 2022). The model has a significant correlation issue when $VIF > 10$ and $TOL < 0.1$ (Chen et al. 2018; Xia et al. 2022).

The "olsrr" package and "ols_vif_tol()" function in RStudio were used to implement the multicollinearity test for VIF and TOL among the landslide conditioning variables. These multicollinearity tests can identify redundant components that should be removed (O'brien, 2007; Pourghasemi et al., 2013). The pseudocode for the implementation of multicollinearity test is shown in Table 3.1.

Table 3.1: Pseudocode for variance inflation factor (VIF) and tolerance (TOL)

Algorithm 1: Multicollinearity Test

INPUT Landslide datasets

OUTPUT multicollinearity test result

1) Load Packages

olsrr

2) Compute linear regression

`lm ()` function

(3) Compute VIF and TOL

`ols_vif_tol()` function

END

3.6.2 Landslide Conditioning Factors selection with Boruta

In this study, the Boruta algorithm was used to prioritize the selected factors triggering landslides. The Boruta is a productive factor optimization process (FOP) (Amiri et al., 2019). According to Ahmadpour et al. (2021), the Boruta utilized shadow features copies of the original features, and produced decision trees based on the objects to which the shadow features were randomly assigned. The significance of an attribute was determined by the accuracy loss of the classification model caused by attributes assigned to objects at random.

Next, the accuracy loss's mean and standard deviation were calculated. The Z-scores were next produced by dividing the accuracy loss's mean and standard deviation. The Z-score was regarded as an essential metric in the Boruta algorithm (Amiri et al., 2019). To determine the importance of original attributes, the set of the importance of shadow attributes

was utilized as a reference (Habibi et al., 2023). The significance of original features then contrasted with the maximum significance of shadow features (Kursa & Rudnicki, 2010).

The Boruta model was implemented using the “Boruta” package in R (Kursa & Rudnicki, 2010). The Boruta pseudocode was shown in Table 3.2.

Table 3.2: Pseudocode for Boruta model

Algorithm 2: Boruta
INPUT Landslide datasets
OUTPUT Boruta feature selection result
1) Load Packages
<i>boruta</i>
2) Compute Boruta model
boruta () function
(3) Plotting output
plot()function
END

Following the feature prioritization with Boruta, sensitivity analysis was also conducted to assess the robustness of the models and evaluate the impact of different landslide conditioning factors on model performance. Sensitivity analysis helps identify which factors have the most significant influence on the accuracy of the models, providing further insight into the reliability and performance of the model in predicting landslide susceptibility. This analysis helps to ensure that the models remain robust and that their performance is not overly sensitive to specific features.

3.6.3 Frequency Ratio

The susceptibility of landslide in any region can be estimated based on the contribution of causative factors and the relationship between them. After confirming the chosen conditioning factors, a more in-depth analysis of the relationship between the parameters chosen and the occurrence of landslides was conducted. This was achieved by conducting the frequency ratio (FR) analysis as performed in previous studies by Acharya and Lee (2019) and Rabby and Li (2020).

FR was defined as the ratio of the landslide are in the total study area and was also referred to as the ratio of the probabilities of a landslide occurrence to a non-occurrence for a given attribute (Lee & Sambath, 2006; Pradhan & Lee, 2010). It was based on the observed relationship between the spatial distribution and causative factors of landslides and can be expressed mathematically in Equation 3.7:

$$FR = \frac{A/A'}{B/B'} \quad \text{Equation 3.7}$$

For each factor, A is the number of pixels that have a landslide; B is the number of pixels in the factor's class region; A' is the total number of landslides in the research area; B' represents the total number of pixels in the study region, and FR denotes a class's frequency ratio for the factor. The FR for each class of every chosen LCFs were computed in Microsoft Excel.

3.6.4 Comparative Analysis

In this study, a comparative analysis was conducted between five machine learning algorithms, Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Logistic Regression (LR). The primary purpose of this comparison is to justify the selection of RF and XGBoost for the proposed methodology. Each algorithm was evaluated based on its performance metrics, such as accuracy, precision, sensitivity and specificity, using the same dataset and preprocessing pipeline. The pseudocode to train each model is presented in Table 3.3.

This initial analysis focused on evaluating algorithmic performance on raw tabular data without parameter tuning. This approach ensures a fair comparison of the models' baseline capabilities and avoids introducing optimization biases. For RF and XGBoost, default hyperparameters were used at this stage, as the emphasis was on understanding their inherent strengths and weaknesses when applied to the dataset. Fine-tuning of these parameters is planned for future phases of the research to maximize model performance and generalization

Convolutional Neural Networks (CNNs), while highly effective for image and spatially structured data, were not included in this study due to the nature of the dataset and the objectives of the research. The dataset primarily consists of tabular data without spatial or hierarchical features, making CNNs less suited for this analysis. Moreover, CNNs typically require significantly larger datasets and higher computational resources to achieve optimal performance, which was not feasible or necessary for this study. The focus was on algorithms better aligned with the dataset characteristics and research goals, such as Random Forest and XGBoost, which excel in tabular data processing with minimal preprocessing and are more interpretable for practical applications.

Table 3.3: Pseudocode for KNN, LR, SVM, RF and XGBoost

Algorithm 3: RF, XGBoost, LR, KNN and SVM

INPUT

Datasets

Training datasets=0.7

Validation datasets=0.15

OUTPUT Prediction.

1) Load Packages

caret

random forest

xgboost

kernlab

readr

tidymodels

themis

plot3D

2) Input Datasets

Training= 0.7

3) Define train Control

Method = 'cv' and number = 10.

7) Train models:

Train KNN model using train() with method = "knn"

Train RF model using train() with method = "rf"

Train SVM model using train() with method = "svmRadial"

Train XGBoost model using train() with method = "xgbTree"

Train Logistic Regression model using train() with method = "glm" and family = "binomial".

8) Input Datasets

Validation datasets= 0.15

5) Compute prediction and confusion matrix

predict () function

confusion matrix () function

END

3.6.5 Hyperparameter Optimization

3.6.5.1 Selection of hyperparameter optimization technique

The two most popular types of hyperparameter optimization techniques are automated search techniques and manual search techniques. Hyperparameter settings are manually tested by manual search. It relies on the basic intuition and expertise of knowledgeable users who can recognise the crucial parameters that have the most significant influence on the outcomes and then utilize the visualisation tools to establish the relationship between those particular parameters and the result (Bardenet et al., 2013).

More professional background knowledge and real-world experience are required to use manual search. It is challenging for non-expert users to apply. Adjusting hyperparameters is a non-replicable procedure. In addition, managing a large number of hyperparameters and a wide range of values becomes challenging since humans struggle to handle high-dimensional data and frequently misinterpret or overlook patterns and relationships in hyperparameters (Sandha et al., 2020).

There are automatic search techniques, like grid search, that have been developed to address the shortcomings of manual search. Grid search operates on the principle of comprehensive searching. Grid search uses any possible combination of hyperparameter values on the training set to train a machine-learning model. On a cross-validation set, it assesses performance using a predetermined measure. Ultimately, the hyperparameters that yield optimal performance are those produced via grid search. This approach has the advantage of automatic tuning and can potentially yield the global optimal value of the optimisation objective function; however, it is constrained by the curse of

dimensionality since algorithmic efficiency decreases rapidly with an increasing hyperparameter range and number of parameters to be tuned (Bergstra & Bengio, 2012).

On the other hand, Bergstra and Bengio (2012) demonstrates that grid search is less unreliable when training some complex models. Therefore, a challenge that has not yet been fully resolved in machine learning is enabling the automatic tuning method to attain high efficiency and precision. An optimization problem known as "hyperparameter tuning" has an unidentified or "black box" goal function. Bayesian Optimization (BO) can effectively solve this problem (Betrò, 1991). Bayesian optimization is a very effective optimization algorithm in solving this kind of optimization problem (Jones et al., 1998).

3.6.5.2 Bayesian Hyperparameter Optimization

I. Bayesian Optimization Process

The Bayesian optimization (BO) process works as following according to Wu et al. (2019) and Wang et al. (2021):

- Step 1: The BO process was begun with the sample initialization where limited set of sample points were selected randomly.
- Step 2: The model was initialized where those points were utilized to calculate a surrogate function, Gaussian process (GP).
- Step 3: Acquisition function was used to obtain the next point. The model was then updated by reevaluating the surrogate function.
- Step 4: In this validation point stage, the surrogate function was verified whether the function was stable or not.

The Bayesian algorithm possessed strong robustness and utilized prior knowledge in the Gaussian process (GP). The algorithm only required input and output data. It fitted the posterior distribution of the objective function by increasing the number of samples, thereby realizing the hyperparameter optimization of the model and obtaining the optimal solution.

II. Gaussian Process

Gaussian process (GP) was used as surrogate models in Bayesian optimization because GP can easily calculate the predicted distribution of the target (Garrido-Merchán & Hernández-Lobato, 2020). GP defined a distribution over functions, where any finite set of function values followed a multivariate Gaussian distribution. In the context of Bayesian optimization, a GP was used to model the unknown objective function, and it provided a posterior distribution over the function values given the observed data (Snoek et al., 2012). GP particularly useful in domains with limited prior knowledge because they rely on the assumption that similar inputs lead to similar outputs. Thus making GP well-suited for detecting patterns and trends in the data when not much prior information was available (Abbas et al., 2023).

Based on Abbas et al. (2023), Wang et al. (2021) and Sun et al. (2020), when GP was used as the basic model, assuming that the objective function $f(x)$ being optimized was sampled randomly from GP, it was expressed as $f(x) \sim \text{GP}(m(x), k(x, x'))$, where $k(x, x')$ represents the covariance function, and $m(x)$ represents the mean function. The intrinsic characteristics of the objective function $f(x)$, such as smoothness, additive noise, and amplitude were specified by the covariance function $k(x, x')$, and the output result was the covariance of $f(x)$ and $f(x')$. In the general Gaussian model, the probability of each feature

was calculated and then accumulated. In the multivariate Gaussian probability model in the GP, the covariance matrix was constructed, and the probability values of all feature vectors were used. The GP model was mathematically expressed in Equation 3.8 to Equation 3.10:

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\text{cov}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \text{cov}^{-1}(x - \mu)\right) \quad \text{Equation 3.8}$$

Where μ (mean) and cov (covariance) are as follows:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{Equation 3.9}$$

$$\text{cov} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad \text{Equation 3.10}$$

III. Function Acquisition

Based on the Bayes rule, a trade-off between active learning and the optimum objective function was made up until the present iteration (Wu et al., 2019). Furthermore, to maximize accuracy for a classification problem, the BO method decided whether to focus on the region that contains the best point that produced the maximum objective function until the current iteration, known as exploitation or examined a different region with the highest uncertainty, known as exploration. Each iteration experienced this.

In the present research, the acquisition function for the Bayesian hyperparameter optimization in the RF model was implemented using the Gaussian Process Upper Confidence Bound (GP-UCB). GP-UCB was a combination of mean and variance functions. It promoted a high level of variance and means (Dai et al., 2019). Additionally, it was able

to a smaller trade-off coefficient and maintain the regret bound. As a result, the algorithm can better strike a balance between local and global investigations (Berk et al., 2020).

3.6.6 Random Forest

3.6.6.1 Structure

To explain the spatial relationship between landslide events, a Random Forest (RF) model was used to build several decision trees. It operated by creating many decision trees during the training phase and producing classes that represent the individual trees' modes of categorization or regression (Breiman, 2001). The RF significantly improved bootstrap aggregating, also called bagging, which was suggested in the literature as an effective method for solving regression and classification issues (Naghbi et al., 2015; Trigila et al., 2015). The average of the outcomes was used to estimate the dependent variable for the regression technique.

For the classification algorithm as used in this study, a decision tree was constructed to output the class. The RF does not make any assumptions regarding the relationship between the response and explanatory variables. This method was suitable for analyzing nonlinearities and hierarchical interactions in large datasets. Thus, it is possible to forecast new data classes more accurately using an RF method (Kim et al., 2018).

The algorithm process of RF was performed in two major steps based on Breiman (2001) :

- Step 1: A training set of several RF bootstrap samples was constructed. The classification rule is also being created. During the bootstrapping phase, some unused observations during tree construction will be “left out” of the training set. The test set known as the "out-of-bag" (OOB) samples was produced by

these "left out" observations. The OOB evaluated predicted prediction accuracy and misclassification error. The OOB error rate was an unbiased estimate of the generalization error rate.

- Step 2: As the tree grew, a particular number of variables, indicated by the letter "mtry," were randomly selected at each node from all the input variables, using the best split possible. Selecting features at random for every node reduces the forest error rate because there is less correlation between any two pairs of trees in the forest. RF includes two strong concepts in machine learning algorithms: bagging and random feature selection. The RStudio used default settings to set the RF model's hyperparameters. Those hyperparameters will be optimized with BO to attain the model's optimal performance.

3.6.6.2 The out-of-bag (oob) error estimate

In RF there was a need to calculate or estimate the minimum number of trees to reduce the Out-Of-Bag error (OOB) (Sahin, 2020). The input data, which was used to generate the tree model, was classified using a set of individual trees to estimate of the test error in OOB to get an unbiased estimate of the test set error. It was estimated internally, during the run, as follows (Breiman, 2001):

- Several samples, typically two-thirds of the original sample set, were extracted from the original sample set using the bootstrap resampling approach in the RF model to create a new training sample set.

- From the new training sample set, several decision trees were built to form the forest, and a corresponding decision tree was generated for each independent training sample.
- The remaining samples, which made up one-third of the initial sample set, were referred to as the test sample set or Out of Bag (OOB) samples. These samples are used to confirm the model's performance.

3.6.6.3 Hyperparameters

"Mtry" was the number of sampled predictors at each stage, and "min_n" was the minimal number of occurrences in the node to be split further. Furthermore, the number of trees that should be constructed was determined by the RF model's "ntree" (Schonlau & Zou, 2020). When dealing with many predictors, some predictors will be correlated. In such a scenario, a small value of "mtry" was used (Chen et al., 2018).

A more conservative model was resulted from larger samples (min_n) and possessed possibility for the occurrence of underfitting. It was noted that the performance of RF was closely related to the strength of each tree and the inter-tree correlation. According to Chen et al. (2018), "mtry" subset was produced by random sampling. The best split on the predictor subset was used to split the node. As "mtry" increases, the strength of each tree was improved, but the inter-tree correlation increases, and the RF error rate increase.

As "mtry" decreases, both inter-tree correlation and the strength of individual trees decreases. So, the optimal value of "mtry" were required to ensure that the trees as uncorrelated as possible.

In addition, since the final classification result of algorithm was produced by all decision trees, the number of decision trees, “ntree” in the RF will also affect the accuracy of the model. The performance of single tree in the forest was low, if the “ntree” was too small, and this led the whole RF model to have bad performance (Wu et al., 2019). In this study, the “min_n” “ntree” and “mtry” were tuned by using BO as shown in the pseudocode in Table 3.4.

Table 3.4: Pseudocode for Random Forest optimized model with Bayesian optimization

Algorithm 4: Random Forest and Bayesian Optimization

INPUT

Datasets

Training datasets=0.7

Validation datasets=0.15

Hyperparameters

mtry

min_n

ntree

OUTPUT Random Forest Optimized model.

1) Load Packages

caret

random forest

readr

tidymodels

themis

plot3D

2) Input Datasets

Training= 0.7

3) Hyperparameters

set mtry=tuned()

set min_n=tuned()

set ntree=tuned()

4) Compute Bayesian Optimization

Tune_bayes function

5) Optimized Random Forest hyperparameters

6) Compute Cross-Validation

set k= 10 folds

7) Input Datasets

Validation datasets= 0.15

5) Compute prediction and confusion matrix

predict () function

confusion matrix () function

END

3.6.7 Extreme Gradient Boosting

3.6.7.1 Structure

A boosting algorithm's primary principle was to combine the outputs of weak learners to improve performance sequentially. XGBoost integrated various classification and regression trees (CART) with the gradient boosting technique. The regularized objective function for enhanced generalization, gradient tree boosting for additive training, and shrinkage and column subsampling for avoiding overfitting are the three key components of XGBoost (Chen & Guestrin, 2016).

The Equations (3.11) to (3.14) summarized by (Chen & Guestrin, 2016) depict the algorithmic process. The XGBoost algorithm aims to minimize the regularized objective function given in Eq (3.11).

$$L(\Phi) = \sum_i \mathbf{1}(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad \text{Equation 3.11}$$

The first term in Equation 3.11 is a loss function, which measures the difference between target y_i and the prediction $\hat{y}_i$, and the second one is a penalty term, as explained in Equation 3.12 for controlling model complexity to avoid overfitting.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad \text{Equation 3.12}$$

where;

T: the number of leaves in the tree;

w: the score of each leaf;

γ, λ : the regularization degrees.

An iterative approach is applied in the method to minimize the objective function. The objective function to be minimized at step t is given in Equation 3.13.

$$L^{(t)} = \sum_{i=1}^n \mathbf{1} \left(y_i \hat{y}^{(t-1)} + f_t(x_i) \right) t \Omega(f_t) \quad \text{Equation 3.13}$$

To speed up the optimization process, second order Taylor expansion is applied to the objective. After removing the constant terms, a simplified objective function at step t is given in Equation 3.14.

$$\tilde{L}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad \text{Equation 3.14}$$

The training process of XGBoost aims at minimizing of the aforementioned objective function. Notably, while the level of model complexities grows as the learning of trees progresses, the regularization term prevents the over fitting problems by regulating the number of leaf nodes in the tree (Arabameri et al., 2021). In addition to the regularization, shrinkage, and feature subsampling are used to prevent overfitting in the XGBoost .

3.6.7.2 Hyperparameters

The XGBoost model contains several common booster hyperparameters including eta, max_depth, min_child_weight, subsample, gamma, colsample_bytree , nrounds and general parameter such as booster. From the previous studies of Can et al. (2021), Wang et al. (2021) and Hussain et al. (2022), it was observed that there was no fixed number of XGBoost hyperparameters used in each research and the assigned values for the hyperparameters used were also different in each study. All the booster parameters descriptions, default values, permissible ranges and the hyperparameters values in several

previous studies were tabulated in Table 3.5. The default values and permissible ranges of values of each of the booster hyperparameters were stated by Chen and Guestrin (2016) and Friedman et al. (2000).

Table 3.5: XGBoost Hyperparameters

Hyperparameters	Description	Default values and permissible hyperparameters. ranges		Hyperparameters Values used in previous studies		
		Values	Search interval	(Can et al., 2021)	(Wang et al., 2021)	(Hussain et al., 2022)
Eta	Shrinkage factor	0.3	[0,1]	0.3	0.14	N/A
Max_depth	Maximum depth of a tree	6	[0,∞]	10	3	6
Min_child_weight	Minimum weight to create a new node Maximum	1	[0,∞]	5	N/A	N/A
Subsample	Subsample ratio of the observations for each tree	1	(0,1]	0.3	N/A	1
Gamma	Regularization term	0	[0,∞]	3	0.1	N/A
Colsample_bytree	Subsample ratio of features for each tree	1	(0,1]	0.5	N/A	0.75
nrounds	Number of gradient boosted trees	100	(0,∞]	100	668	200

The pseudocode of optimized XGBoost model was shown in Table 3.5. The booster parameter used in this study was the default “gbtree”. In this study, all the booster hyperparameters had been considered and the values used to define some bounds for the optimizer to search within were selected based on the stated permissible range for each hyperparameter as stated in Table 3.6.

Table 3.6: Pseudocode for XGBoost Optimized model with Bayesian Optimization

Algorithm 5: XGBoost Bayesian Optimization

INPUT

Dataset

Training dataset=0.7

Validation dataset=0.15

Hyperparameters

eta

max_depth

min_child_weight

subsample

gamma

colsample_bytree

nrounds

OUTPUT XGBoost optimized model

1) Load Packages

xgboost

ParBayesianOptimization

mlbench

dplyr

2) Input Datasets

Training= 0.7

3) Hyperparameters

set eta = c(0.001,0.2)

set max_depth= c(1,10)

set min_child_weight = c(1,50)

set subsample= c(0.1,1)

set gamma=c(0.0,0.1)

set colsample_bytree=c(0.75,1)

4) Compute Bayesian Optimization

Bayes_opt ()

-
- 5) Output
 - Optimized XGBoost hyperparameters
 - Best rounds values
 - 6) Cross-Validation
 - set k=10 folds
 - 7) Input
 - Validation dataset=0.15
 - 8) Compute prediction and confusion matrix
 - Predict () function
 - Confusion matrix () function

END

3.6.8 Stacking

3.6.8.1 Meta-classifier

To improve the base-classifier predictions, an ensemble model can include a meta-classifier or meta-learning layer as an extra classification layer. Classifier ensembles, in which a predetermined number of trained classifiers independently predict the class label of the test sample and the decisions are combined, are the precursors of meta-learning (Sagi & Rokach, 2018). The defining characteristic of this ensemble lies in the diversity of its base classifiers, which constitute its components. Owing to this diversity, the predictions generated by the base classifiers exhibit a range of variations. This diversity can be harnessed through integration by a meticulously designed meta-learning classification stage to achieve high accuracy in predictions (Susan et al., 2021).

The prevailing trend in recent studies adopting a stacking approach, particularly within the domain of LSM, involves the utilization of the Generalized Linear Model (glm) (Gyamerah et al., 2023; Kablan et al., 2023; Kwon et al., 2019). There is no specific

guideline for selecting the kind of meta-learning model. Somehow, the meta-learner must be less complex and interpretable. This is because the stacking framework's primary layer resembles an extremely complex nonlinear feature converter. The preceding instances have new representations after this change. The second layer does not require complicated classifiers (Song et al., 2020).

3.6.8.2 Implementation of Stacking

In this research, there were two stages to the architecture of stacking the RF and XGBoost models. The stacking phases were consisting of two phases, as illustrated in Figure 3.13 and as listed below (Divina et al., 2018):

- Level 0 Models – At this level, RF and XGBoost were trained individually as base models.
- Level 0 Prediction - The predictions that were the resultant of the individual RF and XGBoost models were known as the level 0 prediction. This was based on the training data.
- Level 1 Model - This was the model that combined the prediction of all the level 0 or base models. It was also known as a meta-classifier. The meta-classifier used in this research was the glm.
- Level 1 Prediction - The glm meta-classifier was trained on multiple predictions given by the RF and XGBoost base models, and then it optimally integrated the predictions of the base models on the testing data.

Predictions were then produced using validation datasets fed into the base models along with the expected outputs, and these input and output pairs were utilized as the training dataset for the meta-model.

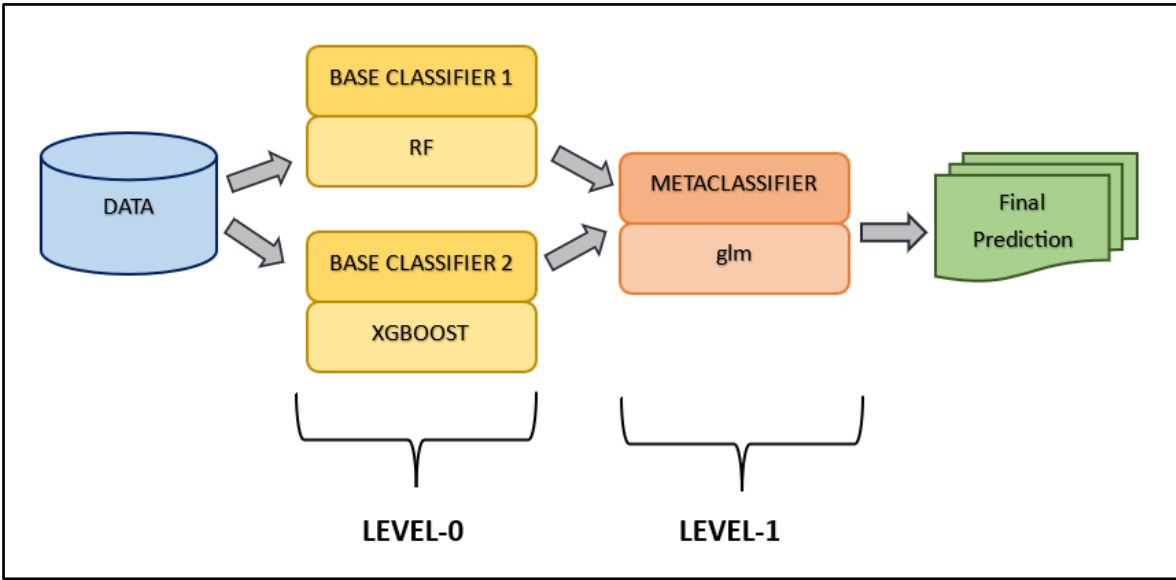


Figure 3.13: Stacked model of current research

Following that, the confusion matrix, statistics, and AUC were used to assess the performance of the stacked model. The pseudocode for the stacking model was shown in Table 3.7.

Table 3.7: Pseudocode for the stacking model

Algorithm 6: Stacked model

INPUT Data

Random Forest prediction

XGBoost prediction

metaclassifier

Testing dataset=0.15

OUTPUT Stacked model

1) Create Data frame

data.frame () function

Random Forest prediction

XGBoost prediction

2) Compute Stacking

Train () function

Set metaclassifier= “glm”

(3) Input

Testing dataset=0.15

(4) Compute prediction and confusion matrix

Predict ()function

Confusion matrix () function

END

3.7 Final Landslide Susceptibility Map

The final landslide susceptibility maps were prepared in the stage after the base and stacked models were developed and evaluated. All the landslide conditioning databases' raster layers were loaded, and it was confirmed that all the layers' extents were identical. It was necessary to resample if the layer's extent was different. The various raster layers were stacked, and the longitude and latitude information were added to a data frame. The LSM of XGBoost and RF, as well as stacked models of Penang Island, were developed after

performing the prediction at the grid location and extracting the predicted level class. The landslide and non-landslide classes were included in the level class. Each study region's DEM raster layer served as the final coordinate reference system for the LSMs.

Imported into ArcMap and reclassified using the "reclasses" spatial analyst were the raster layers from the stacked, XGBoost, and RF models of Penang Island. The natural breaks (Jenks) were used to divide the classification's values into two classes, namely "susceptible" and "non susceptible" (Chen et al., 2013; Sahana & Sajjad, 2017). The entire area of the susceptible and non-susceptible regions was then determined and tabulated. The percentage of the coverage area is then calculated in ArcMap 10.4 calculator by using the formula below in Equation 3.15:

$$\text{Percentage of coverage area} = \frac{\text{No.Pixels} \times \text{raster cell size}}{\text{Total study area}} \times 100\% \quad \text{Equation 3.15}$$

Default raster cell size = 30m x 30 m,

Total study area = 301.248 km²

The resulting susceptibility map categorizes the study region into two distinct susceptibility classes: high and low. The percentage of coverage area for each class is calculated using the RF, XGBoost and stacked models to assess any potential overestimation in the model outputs. The high-susceptibility areas represent regions at a greater risk of landslides, while the low-susceptibility areas are less prone to such events.

3.8 Conclusion

In conclusion of this chapter, the progression commenced with the development of landslide inventories for the study area, followed by the establishment of spatial databases

for LCFs, followed by the LCFs selection, prediction modelling, and the final stage is the production of the LSM. The steps in developing LCFs spatial databases using ArcMap software had been explained in detailed with all the specific tool such as the spatial analyst tool involved. Besides, the modelling stage started off with hyperparameter tuning using BO and the optimal hyperparameter then being used to build the RF and XGBoost model. Then, the two optimized single model were stacked together using glm as the metaclassifier. The performance of single and stacked models was evaluated. The final LSM for all the models are developed. The modelling took place in RStudio software and all the involved packages for every algorithm had been listed in the methodologies.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Overview

This chapter provides a detailed account of each experimental stage for the study and prediction of landslide susceptibility, its results, and the final mapping for this research. The experiments' findings were presented in tables and graphs for simple understanding. Additionally, the explanation of the results provides descriptions of the procedure, the results, and their importance. The multicollinearity test, the Boruta model for feature selection, the correlational analysis between landslide conditioning factors and landslides, the landslide susceptibility models, and the final maps are the outputs of the experiments discussed in this chapter. This study focuses on analysing the performance of the implemented models with hyperparameter optimization and developing a hybrid ensemble of the base learner models.

4.2 Feature Selections

The goal of screening for conditioning variables is to eliminate conditioning factors that are unnecessary or redundant. Furthermore, sufficient dominant conditioning components can reduce the model's input data resources and runtime. VIF and TOL were used to conduct the multicollinearity test. The Boruta feature selection method was then used to identify the dominant conditioning factors from all the conditioning factors. The FR was used to determine the correlation between the factors and the occurrence of landslides once the most important factors had been selected.

4.2.1 Multicollinearity Test

O'brien (2007) claimed that multicollinearity may overstate estimates of variable variance. The ability to comprehend the results and estimate the parameters of model approaches is negatively impacted by high levels of correlation among independent variables. The VIF and TOL are reliable methods that aid in discovering correlated variables. According to Dou et al. (2015), this indicator is the most often used measurement of the multicollinearity level in the regression model. They are especially preferred for identifying the risk elements in landslide evaluation. If the TOL value is less than 0.1 or the VIF value is more significant than 10, a multi-collinearity problem in the predisposing factor is likely present (Dormann et al., 2013).

Table 4.1 shows 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 found to have the highest VIF value (1.78) and the lowest tolerance (TOL) value (0.56). On top of that, the stream power index (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 showed that there are no significant collinearity issues between the 10 landslide factors. Hence, these 10 factors were selected for the next feature selection stage via the Boruta Model.

Table 4.1: 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

4.2.2 Landslide Conditioning Factors Selection Via Boruta

The contributions of the 10 contributing factors to the occurrence of disasters in the research area were determined using the Boruta model. All 10 conditioning factors are confirmed by the Boruta model to be significant in the occurrences of landslides in Penang Island, as shown in Figure 4.1 and Table 4.2. Similar to the landslide susceptibility mapping results in the Abha Basin (Youssef & Pourghasemi, 2021) and Muzzafarabad district (Hussain et al., 2022), the slope and DEM contribute more to the incidence of risks than the other contributing elements. Other factors such as the distance from the road, TWI, SPI, distance from stream, slope length, plan, and profile curvatures were also significant in several previous studies. These factors were also dominant conditioning factors in a large number of previous studies (Chen & Li, 2020; Nhu et al., 2020; Pham et al., 2022; Sun et al., 2020).

Table 4.2: Boruta Model's result of importance of variables

Landslide Conditioning Factors	Mean Importance	Median Importance	Minimum Importance	Maximum Importance	Norm Hits	Decision
Slope	47.67	47.59	44.91	50.72	1.00	Confirmed
DEM	38.54	38.57	36.59	40.08	1.00	Confirmed
TWI	26.91	26.69	25.52	28.55	1.00	Confirmed
SPI	17.30	17.07	15.47	19.49	1.00	Confirmed
Plan Curvature	14.39	14.50	13.56	15.58	1.00	Confirmed
NDVI	13.73	13.57	11.56	15.60	1.00	Confirmed
Slope Length	12.81	12.83	12.30	13.73	1.00	Confirmed
Profile Curvature	11.95	12.16	10.69	13.36	1.00	Confirmed
Distance from Road	10.69	10.64	8.77	12.17	1.00	Confirmed
Distance from Stream	7.38	7.34	5.34	9.19	1.00	Confirmed

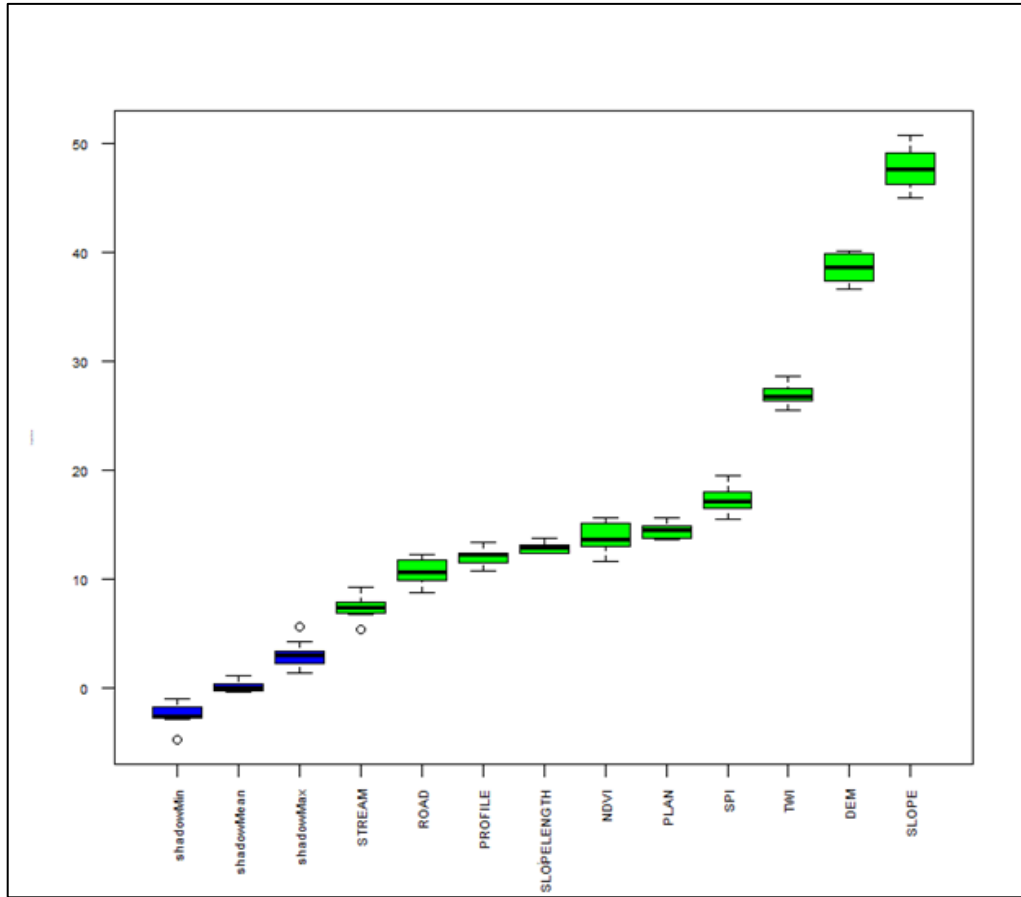


Figure 4.1: Boruta model's chart plot on the importance of variables

4.2.3 Frequency Ratio

As indicated in Figure 4.2, the frequency ratio (FR) model analyses the relationship between landslide occurrence and its influence factors. The FR value can be used to show in detail the contribution of every class range of each LCFs that had been confirmed by the Boruta to the occurrence of landslide. For the Penang Island the FR revealed that the elevation influential effect toward landslide reduced as elevation increased, with the landslide occurring most frequently at elevations between 78 and 200 meters above sea level. This is because as height increases, precipitation increases, temperatures drop, and weathering accelerates, encouraging landslide development (Gruber & Haeberli, 2007).

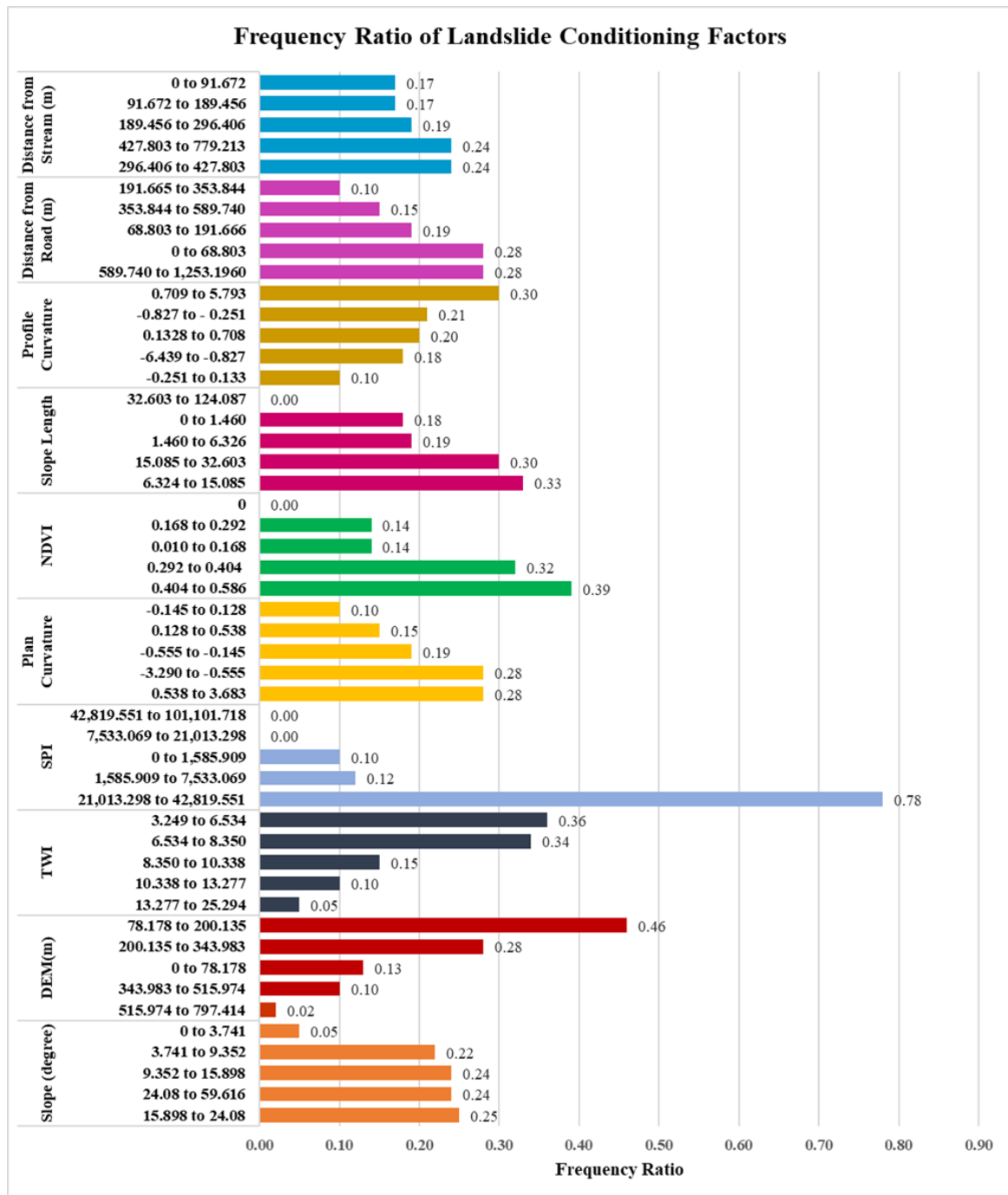


Figure 4.2: Frequency Ratio of landslide conditioning factors

Curvatures on Penang Island were substantially associated with the occurrence of landslides at the profile curvature point, where 0.708 to 5.793 landslides occurred. Additionally, there was a strong relation between the plan curvature and the occurrence of landslides in the ranges of 0.328 to -0.555 and 0.538 to 3.683. The degree of slope erosion is influenced by both the profile curvature and the plan curvature, which both regulate the speed of flowing water (Aghdam et al., 2016).

Landslides are more likely to occur on slopes greater than 10° angle. The previous finding indicated that landslides were more frequent on moderately steep slopes (15°–25°), suggesting that this range could serve as a threshold (Nakileza & Nedala, 2020). In the western highlands of Uganda, Nseka et al. (2019) found additional landslides on slopes between 25 and 35°. According to Bizimana & Sonmez (2015), the places most susceptible to landslides are those with slope angles of more than 14° on convex slopes and more than 41° on concave slopes. Zhuang et al. (2015) found that the higher slope range of 15° to 40° had the highest occurrence of landslides (51.81%). Due to the terrain being soft and covered in thick colluvium or residual soils, which require higher water levels to cause slope collapses, the prevalence of landslides is low at lower grades. Similarly, there are not many landslides at very high slope gradients since the terrain is so steep and there is not much colluvium to mobilize it. Steeper slopes are more susceptible because the slope gradient controls the shear pressures on the hillside (Silalahi et al., 2019).

The SPI value had the highest FR value with an incidence of landslides in the range of 21013.3 to 42819.6. TWI was significantly connected with landslides only in the range of 3.2 to 6.5 for Penang Island. The FR data demonstrate that the association between this variable and landslides declines with increasing distance from the road for both research

locations. This aligned with what Lei et al. (2020) found in their study. Although the FR value for the range from 68.8m to 589.7m is lower, it was discovered in Penang Island that the correlation value is high in the range of 589.74m to 1,253.2m. This is supported by research showing that most landslides happen within a 1000 m radius of a road (Wang et al., 2022). Concurrently, the construction of roadways alters the terrain's initial stability (Wang et al., 2020).

In contrast, as the distance from the stream grows, so does the relationship of factors with the occurrence of landslides on Penang Island. In Penang Island, the maximum distance from the stream is 427.8 m, which is still seen as being too close to the stream. The stability of the slope is impacted by the strength of the erosion at the slope's base, which increases with proximity to the river (Yalcin, 2008). Finally, FR discovers that for Penang Island, the NDVI value has the largest positive association with the occurrence of a landslide, between 0.4 and 0.5.

4.3 Modelling

4.3.1 Performance comparison between XGBoost, RF, LR, KNN and SVM

To justify the selection of RF and XGBoost in this study, both RF and XGBoost performances had been compared with other three commonly used machine learning models namely the K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Logistic Regression (LR). Therefore, after training all the five ML models, the cross-validation accuracy and the accuracy obtained from the confusion matrix on the prediction of trained models on validation datasets were tabulated on Table 4.3.

From Table 4.3, Among the machine learning models evaluated, XGBoost demonstrated the best overall performance, achieving a cross-validation accuracy of 90.45% and a prediction accuracy of 94.71%, indicating its strong ability to generalize and effectively capture the underlying patterns in the data. Random Forest also performed well, with a cross-validation accuracy of 89.17% and a prediction accuracy of 90.15%, making it a reliable alternative to XGBoost. LR and KNN showed consistent and moderate performance, with prediction accuracies of 88.11% and 88.51%, respectively, suggesting they are suitable for scenarios where model simplicity or interpretability is prioritized. In contrast, SVM exhibited significant overfitting, as evidenced by its high cross-validation accuracy (89.53%) but substantially lower prediction accuracy (75.76%), highlighting the need for further tuning or adjustments. Overall, XGBoost was identified as the most robust and accurate model for the landslide prediction task.

RF and XGBoost often outperform LR, SVM, and KNN due to several key factors. First, both RF and XGBoost excel at capturing non-linear relationships between features and the target variable. Unlike LR, which assumes a linear relationship, RF and XGBoost are tree-based algorithms that split the data hierarchically based on feature importance. This allows them to model complex, non-linear patterns in the data without the need for additional feature engineering. On the other hand, LR may struggle to capture non-linear relationships unless extra features (e.g., polynomial terms) are manually created. While SVM with a non-linear kernel like the Radial Basis Function can also handle non-linearity, it often requires careful parameter tuning to achieve good performance, which may be challenging with a smaller dataset. KNN, being distance-based, can suffer when there are irrelevant features or non-linearities, as it treats all features equally without considering their importance.

In addition to handling non-linearities, RF and XGBoost are adept at evaluating feature importance and identifying interactions between features. This ability is particularly valuable in datasets where the relationship between features and the target is not straightforward. In contrast, LR and SVM do not automatically model feature interactions, which can be a limitation when dealing with complex datasets. While KNN does not evaluate feature importance either, it can be sensitive to irrelevant or redundant features, which impacts its overall performance.

The moderate size of the data (620 training data) also plays a role in the performance differences. RF and XGBoost are robust algorithms that can handle datasets of this size well, even without parameter tuning. LR, however, may underfit if the dataset does not provide enough data points to capture the relationships between features and the target variable effectively. Similarly, SVMs require careful selection of hyperparameters and kernel types, which can be difficult to tune for small datasets, often leading to suboptimal performance. KNN, relying on local distances between data points, can become unreliable when the dataset is small and the feature space is high-dimensional.

Lastly, RF and XGBoost provide scalability and interpretability. They naturally scale to higher-dimensional data and allow for a better understanding of which features are most influential in making predictions, making them ideal for tabular datasets. In contrast, KNN and SVM often require significant preprocessing, such as feature scaling and normalization, and may be more computationally intensive when handling moderately sized datasets.

In conclusion, RF and XGBoost outperform LR, SVM, and KNN in datasets of this size because they handle non-linearities, feature interactions, and small-to-moderate data

sizes more effectively. Their robustness make them well-suited for real-world applications with tabular data.

Table 4.3: Cross-validation and prediction accuracies of all MLmodels

ML models	Cross-validation accuracies	Prediction accuracies
XGBoost	0.90451	0.9471
RF	0.8917	0.9015
KNN	0.8805	0.8851
SVM	0.8953	0.7576
LR	0.8903	0.8811

4.3.2 Sensitivity Analysis

Across all models in Table 4.4, slope angle stands out as one of the most significant landslide conditioning factors, as it considerably determines the extent to which gravitational force affects a certain region. It is determined that the steeper the slope angle, or the more perpendicular it is to the horizontal plane, the more unsteady a slope becomes. This aspect allows for accurate susceptibility mapping to be conducted.

The models' predictive ability was improved because of TWI and SPI as they incorporated the effects of water flow accumulation and slope steepness on slope failure in different environments. These hydrological factors influence the level of saturation of soils and the possibility of slope failure. Along with the aforementioned factors, distance from roads and distance from streams in Model 9 fostered prediction accuracy as they demonstrated the human and natural factors affecting slope stability. Areas around roads tend to be construction disturbed while streams are known to contribute to soil erosion and waterlogging.

As much as NDVI and slope length and curvature are known to be critical in landslide susceptibility modelling, they impact differently. NDVI is more of a vegetation density indicator which is useful in stabilizing slopes thereby reducing landslides but is a poor indicator of erosion. Nevertheless, its contribution to models was not constant since cross validation accuracy was increased while prediction accuracy decreased in sparsely vegetated areas or urbanized regions. Slope length, which affects the volume of water accumulated and the energy from runoff, raised predictive accuracy in some examples.

The performance of both RF and XGBoost did not change with respect to the models used, with XGBoost exhibiting greater cross-validation accuracy. The prediction accuracy for both candidates was inconsistent which points to a greater discrimination of conditioning factors necessary for the models. The steady improvement in the accuracy of the predictions when a progression of models is from model 1 to model 9 is an indication that the inclusion of wider and contradictory parameters helps shield the models against potential weaknesses. The consistent performance of both the RF and XGBoost models in model 9 when used in landslide susceptibility mapping indicates the efficacy of this model.

Table 4.4: Sensitivity Analysis for RF and XGBoost models

Model	Landslide Conditioning Factors	Cross-validation accuracies		Prediction accuracies	
		RF	XGBoost	RF	XGBoost
Model_1	DEM, Slope angle	0.8914	0.9033	0.8593	0.8593
Model_2	DEM, Slope angle, TWI	0.9054	0.9169	0.833	0.833
Model_3	DEM, Slope angle, TWI, SPI	0.891	0.9154	0.9213	0.9134
Model_4	DEM, Slope angle, TWI, SPI, Plan curvature	0.8843	0.8956	0.8512	0.8347
Model_5	DEM, Slope angle, TWI, SPI, Plan curvature, NDVI	0.8996	0.9137	0.8707	0.819

Model_6	DEM, Slope angle, TWI, SPI, Plan curvature, NDVI, Slope length	0.9015	0.9174	0.875	0.8
Model_7	DEM, Slope angle, TWI, SPI, Plan curvature, NDVI, Slope length, Profile curvature	0.8925	0.9019	0.878	0.9024
Model_8	DEM, Slope angle, TWI, SPI, Plan curvature, NDVI, Slope length, Profile curvature, Distance from road	0.8949	0.9127	0.9016	0.8934
Model_9	DEM, Slope angle, TWI, SPI, Plan curvature, NDVI, Slope length, Profile curvature, Distance from road, Distance from stream	0.8995	0.9141	0.9167	0.9167

4.3.3 Random Forest with Bayesian Optimization

The minimum number of occurrences in the node to split further (min_n), the number of sampled predictors at each stage (mtry) and the number of trees built (ntree) are the three important RF hyperparameters which had the most significant impacts on the outcomes, and were chosen for the current study hyperparameter optimization.

The hyperparameter values for each iteration were obtained by optimizing the ntree, mtry and min_n hyperparameters using the Bayesian optimization (BO) approach. This BO for the RF produced a maximum area under curve (AUC) value of 0.969097, with the corresponding hyperparameter values of ntrees=352, mtry=2 and min_n=6, as shown in Table 4.5.

The RF model is then trained using the optimal mtry, ntree, and min_n values. Additionally, the Out-of-Bag (OOB) error for the RF model with classification type, 352

ntrees, and two variables tried at each split (mtry) was 9.26%, resulting in a trained model accuracy of 90.74%. This demonstrates that the RF model has good accuracy.

Table 4.5: Hyperparameters Bayesian Optimization of Random Forest

Mtry	min_n	ntrees	.metric	mean	n	std_err
2	6	352	roc_auc	0.969057	10	0.007477
1	6	338	roc_auc	0.968074	10	0.008223
2	1	348	roc_auc	0.967953	10	0.007667
1	8	330	roc_auc	0.967633	10	0.00817
1	9	350	roc_auc	0.967522	10	0.008196
3	4	336	roc_auc	0.96746	10	0.007698
1	7	389	roc_auc	0.967418	10	0.008304
1	6	314	roc_auc	0.967085	10	0.008428
6	7	319	roc_auc	0.966756	10	0.007187
1	3	338	roc_auc	0.966655	10	0.008193
4	2	391	roc_auc	0.966255	10	0.007379
3	5	240	roc_auc	0.965325	10	0.00796
10	1	165	roc_auc	0.963284	10	0.00768
9	4	468	roc_auc	0.962288	10	0.008137
7	9	104	roc_auc	0.961528	10	0.008282

Table 4.6 displays the confusion matrix and statistical findings of the trained model tested with the validation datasets. The confusion matrix shows that the RF model built possessed excellent performance with an accuracy of 89.06%. Besides, the kappa value posed by the RF model was 0.6 to 0.8, indicating that the model is substantially reliable; hence, it is substantially agreed, as stated by Landis & Koch (1977). Both the sensitivity and specificity of the RF model in this current study fall in the range of 0.8 to 0.9, which indicates that the models are highly sensitive and specific. The RF model classifies 91.53% of the non-landslides area and 86.96% of the landslides area correctly within the validation datasets. In

addition, the precision result indicated that the RF model accurately predicted the area of landslide zones 92.31% of the time.

Table 4.6: Validation Dataset Confusion Matrix and Statistics for RF

	Confusion Matrix		Accuracy (%)	Kappa	Precision (%)	Sensitivity (%)	Specificity (%)
	Landslide	Non-Landslide					
Landslide	60	5	89.06	0.781	92.31	86.96	91.53
Non Landslide	9	54					

The difference between the success rate, represented by the AUC value of the training dataset, and the predictive rate, represented by the AUC value of the validation dataset, was just 3.40% (Table 4.7). The difference between the success and prediction AUC score which is lesser than 5% had demonstrated that overfitting occurrence are still in small percentage.

Table 4.7: Success rate and Prediction rate of RF

Model Performance Evaluator	Success Rate	Prediction Rate
Area Under Curve (AUC) %	99.3	95.9

4.3.4 Extreme Gradient Boosting with Bayesian Optimization

Seven hyperparameters, including learning rate of algorithm (eta), the maximum depth of each tree (max_depth), the minimum sum of weights of all observations required in a child (min_child_weight), the percentage of rows used for each tree construction (subsample), the minimum loss reduction (gamma), the subsample ratio of columns (colsample_bytree), and the number of trees (nround) are part of the XGBoost model. The

tree-based model that the booster employed was set to “gbtree”. The remaining hyperparameters were optimized using the Bayesian approach, and the hyperparameter value for each iteration was obtained. In Table 4.8, the optimum values were shown.

Table 4.8: Hyperparamters optimal values for XGBoost

Hyperparameters	Optimal Values
eta	0.2
Max_depth	10
Min_child_weight	1
Subsample	1
Gamma	0.1
Colsample_bytree	0.9040207
nrounds	69

The XGBoost model was then constructed using the optimized hyperparameters, and K-fold cross-validation was utilized. The cross-validated resampling accuracy and kappa results were 0.8925 and 0.7850, respectively.

Table 4.9 provide the confusion matrix and statistics result of the XGBoost model tested with validation datasets. The confusion matrix demonstrates that the XGBoost model, which had an accuracy of 91.41%, performed very well. Additionally, the XGBoost model’s kappa value falls in the range between 0.8 to 1.00, demonstrating the model's great reliability and near-perfect agreement. In the validation datasets, the XGBoost model accurately classifies 89.86% of the landslide area and 93.22% of the non-landslide area. Furthermore, the precision result showed that 93.94% of the time, the XGBoost model predicted the location of landslide zones correctly.

Table 4.9: Validation Datasets Confusion Matrix and Statistics of XGBoost

	Confusion Matrix		Accuracy (%)	Kappa	Precision (%)	Sensitivity (%)	Specificity (%)
	Landslide	Non-Landslide					
Landslide	62	4	91.41	0.8277	93.94	89.86	93.22
Non-Landslide	7	55					

Table 4.10 shows the AUC of success and predictive rate of the XGBoost model. The success rate's AUC curve was 2.9% greater than the forecast rate. The small percentage difference between the success and predictive rate indicated only small overfitting occurrence in the model. This discovery has effectively demonstrated how Bayesian hyperparameter optimization reduces the likelihood of overfitting in XGBoost.

Table 4.10: Success Rate and Prediction Rate for XGBoost

Model Performance Evaluator	Success Rate	Prediction Rate
Area Under Curve (AUC) %	100.00	97.1

4.3.5 Stacked Extreme Gradient Boosting and Random Forest

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 confusion matrix and statistics of the stacked model with testing datasets were presented in Table 4.11.

The confusion matrix shows that the RF-XGBoost stacked model performed well, with an accuracy of 86.93%. Besides, the kappa value posed by the stacked model was 0.6 to 0.8, indicating that the model is substantially reliable and had achieved substantial

agreement. The stacked model is highly sensitive and specific as the model had managed to identify 90% of the landslides area and 84.34% of the non-landslides area correctly within the testing datasets. In addition, the precision result indicated that the stacked model accurately predicted the location of landslide zones 82.89% of the time.

Table 4.11: Testing Datasets Confusion matrix and Statistics of Stacked model

	Confusion Matrix		Accuracy (%)	Kappa	Precision (%)	Sensitivity (%)	Specificity (%)
	Landslide	Non-Landslide					
Landslide	63	13	86.93	0.7384	82.89	90.00	84.34
Non-Landslide	7	70					

Meanwhile the prediction rate AUC was 1.17%, slightly higher than the success rate prediction as shown in Table 4.12. The percentage difference was very small and insignificant thus the overfitting was successfully reduced via stacking.

Table 4.12: Success Rate and Prediction Rate for Stacked model

Model Performance Evaluator	Success Rate	Prediction Rate
Area Under Curve (AUC) %	96.8	95.6

4.3.6 Model Comparisons

The results of the accuracy tests, AUC, kappa value, sensitivity, and specificity were utilized to assess the overall performance of the models. The XGBoost model outperformed RF, and the stacked model of RF and XGBoost when performance metric data were analyzed based on the validation and testing datasets. Overall accuracy findings revealed that the

XGBoost model consistently predicted more samples than other models, with an accuracy rate of 91.41%, followed by the RF (89.06%) and the stacked model (86.93%).

Additionally, the kappa index was used to measure the reliability of the LSMs. The stacking model had a kappa value of 0.7384, followed by the RF model (0.781), and then the XGBoost model (0.8277). A "nearly perfect" agreement (0.81 to 1.00) is shown by the XGBoost model. On the other hand, because it falls between 0.61 and 0.80, the kappa value of the stacked and RF models showed "substantial" agreement (Landis & Koch, 1977). According to Kappa's value, the three models have generally come to a solid agreement.

Meanwhile, concerning the success rate, as measured by the AUC, XGBoost attained a perfect score of 100%, followed by RF with 99.3%, and the stacked model with 96.8%, as illustrated in Figure 4.3.

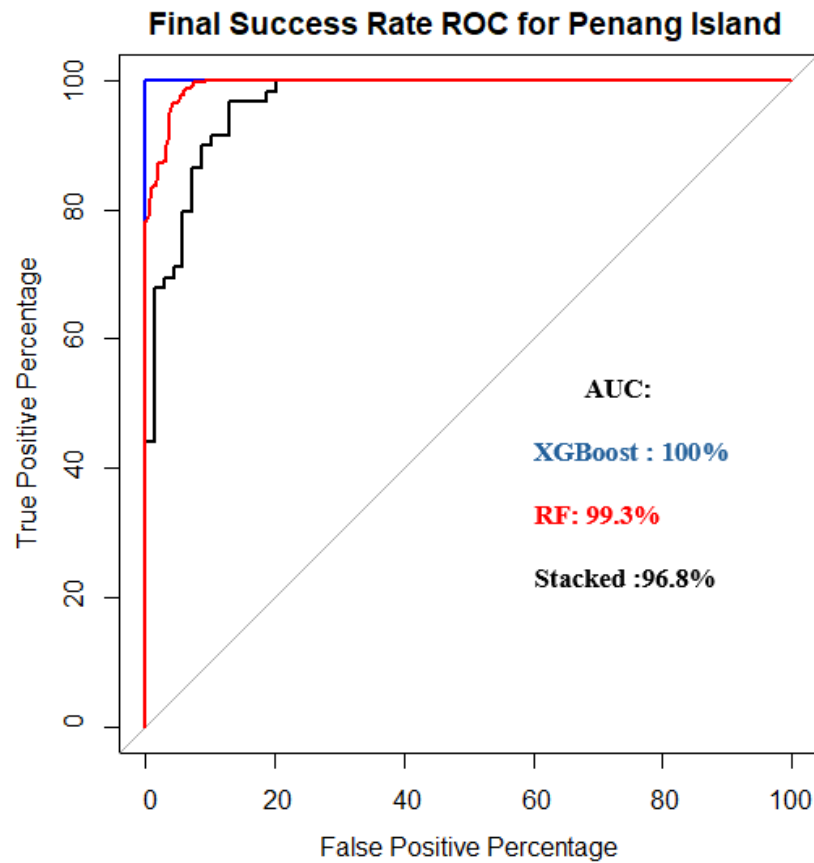


Figure 4.3: Success Rate AUC result

Additionally, the XGBoost (97.1%) also demonstrated better prediction ability in terms of the prediction rate of AUC results, followed by the RF (95.9%) and the stacked model (95.6%) (Figure 4.4). While the differences in AUC scores appear small, statistical comparisons were conducted to provide a more objective evaluation of the models.

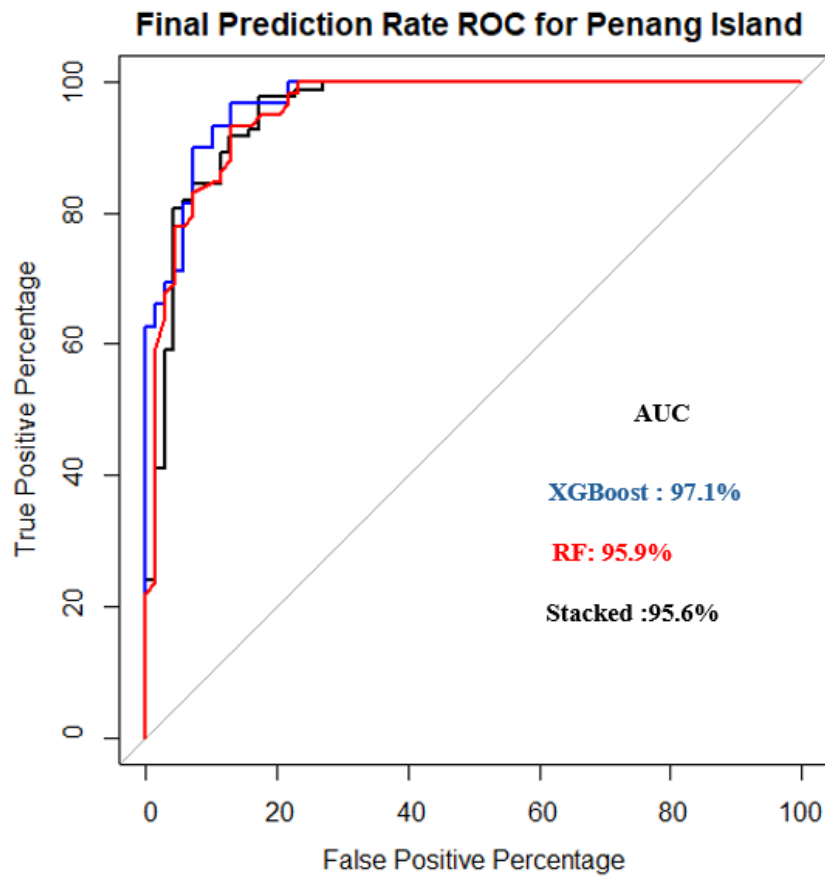


Figure 4.4: Prediction Rate AUC results

Table 4.13: DeLong Test results on models AUC

Paired Models	p-value
XGBoost and RF	0.2309
XGBoost and stacked	0.4717
RF and stacked	0.7392

Based on Figure 4.4, XGBoost achieved the highest AUC (97.1%), followed by RF (96.3%) and the stacked model (95.6%), the differences in AUC between the models were not statistically significant according to DeLong's test (Table 4.13) as the p-values obtained

are >0.05 . This indicates that all three models perform comparably well in terms of overall predictive ability.

Since the AUC values alone are not significantly different, other factors should be considered when determining the best model. XGBoost, with its higher AUC, suggests it might be the most reliable model for distinguishing positive and negative cases. Additionally, XGBoost's advanced regularization techniques help reduce overfitting, making it suitable for datasets with potential noise. On the other hand, while RF's AUC is slightly lower than XGBoost's, it remains a robust and interpretable model that is easier to implement and tune, with consistent performance across various studies. The stacked model, despite having a slightly lower AUC, addresses overestimation tendencies observed in XGBoost by combining the strengths of both base models. This makes the stacked model a valuable choice in applications where balanced predictions is important.

Overall, all the models achieved AUC scores above 90%, they can be classified as having excellent predictive capability (Pradhan, 2013). The high AUC values highlight the strong reliability of the models in distinguishing between positive and negative cases in landslide susceptibility mapping. Therefore, the high accuracies and AUC values across all three models indicate robust prediction performance.

Comparative analysis of gradient boosting research by Cao et al. (2020) showed that the XGBoost method outperformed RF and support vector machine (SVM) algorithms with the highest AUC score. In addition, the XGBoost model has been demonstrated to slightly outperform RF in the LSM study on the Karakorum Highway, Pakistan (Hussain et al., 2021), with only a 1.22% difference in AUC score. Similarly, in this study, the prediction AUC difference between the XGBoost and RF scores was 1.20%.

In the study by Can et al. (2021) in Ataturk Dam, Turkey, the XGBoost had a high AUC score of 96%. The result was slightly lower than what was reported in the current study. Despite using a DEM with a lower spatial resolution, the anticipated AUC value in this case was 97.1%. The high accuracy values can be attributed to the landslide conditioning factors, geographic databases' low noise level, and the XGBoost method's excellent predictive power. Due to its high regularization, the XGBoost model performs better than other machine learning models (Sahin, 2020). Therefore, there RF and XGBoost were proven to be the two effective prediction models for landslide susceptibility.

Meanwhile, the XGBoost-RF stacked model predictive AUC is 1.5% and 0.3% lower than the XGBoost and RF individual model respectively. Although the performance of the stacked model is powerful and slightly lower than the base models but still it did not outperform the single base model. This finding was similar to the study conducted by Nhu et al. (2020), also where the poor performance of Alternating Decision Tree (ADTree) with the AUC score of 59% had been significantly improve to 94% after being stacked with the Ada-Boost (AB). Nevertheless, the performance of the single AB still outperformed the stacked model of AB-ADTree. Thus, from this study and the one conducted by Nhu et al. (2020) it can be concluded that heterogenous ensembling can produce a prediction model with high accuracy although there is no guarantee it can outperformed the base model.

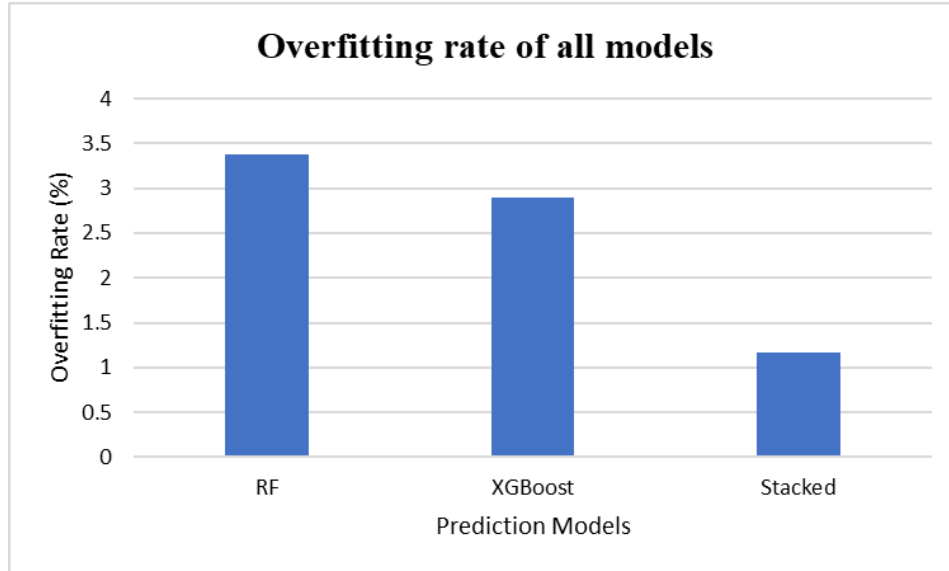


Figure 4.5: Overfitting Rate of all models

The optimized XGBoost and RF models also exhibit minimal overfitting. The LSM of Rangamatti Hill, Bangladesh Rabby et al. (2020) used grid search optimization, which reduced the overfitting percentage to 4.64% and 4.52% for XGBoost and RF models, respectively. The overfitting problem in this study's XGBoost and RF models has been further minimized by BO, with the resulting overfitting percentages being 2.9% and 3.4%, respectively. The stacking of the two optimized models also produced a 1.2% overfitting rate, which is the lowest (Figure 4.5). As a result, for the XGBoost and RF, BO was a successful advanced hyperparameter optimization strategy.

4.4 Final Landslide Susceptibility Maps

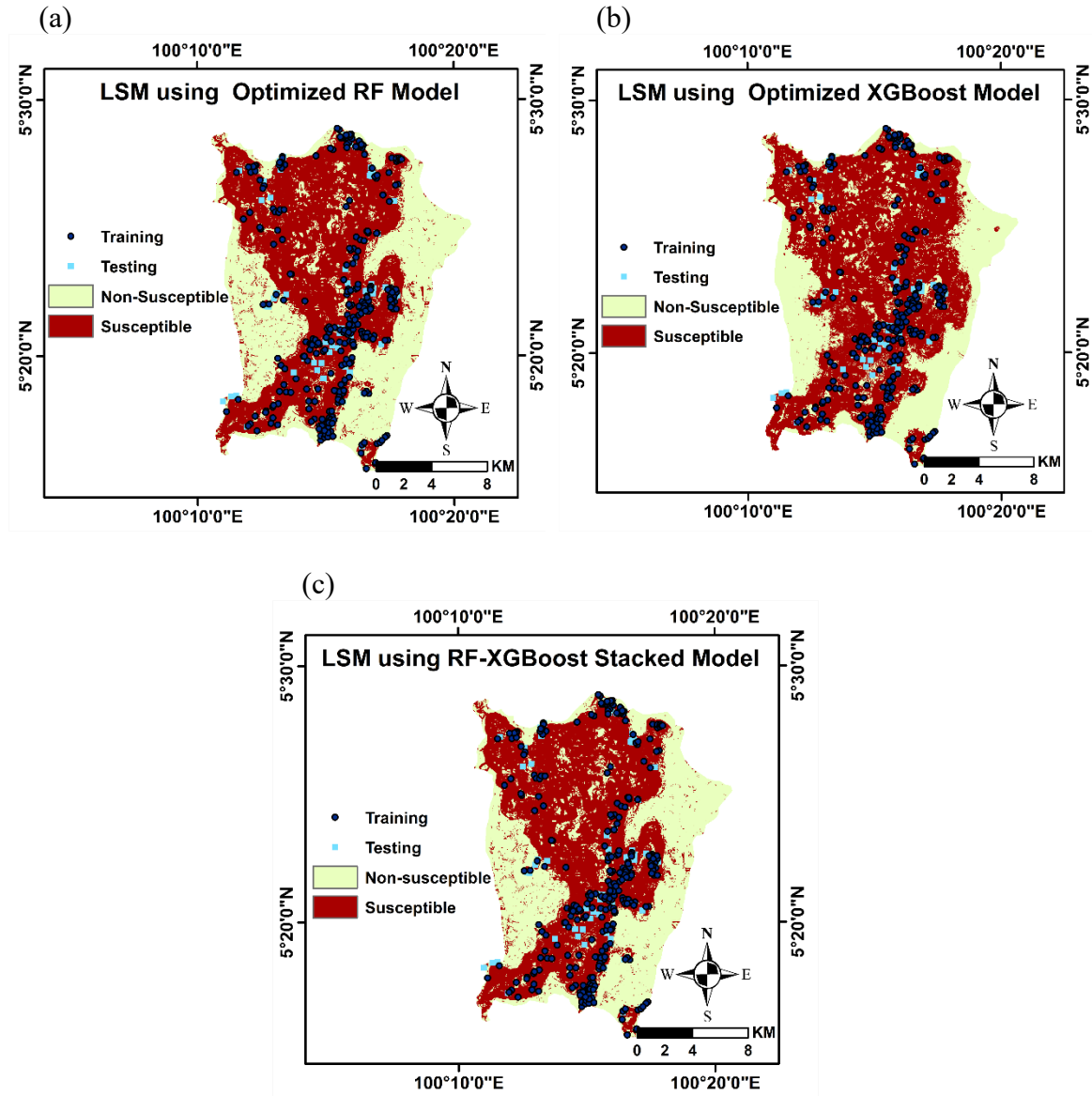


Figure 4.6: Penang Island LSMs via (a) RF (b) XGBoost (c) Stacking

After the reclassification process, to determine the statistical reliability of the results, the susceptible level was considered landslide zones, and the non-susceptible level was considered non-landslide zones. The Jenks natural breakpoint approach Jaafari et al. (2014) was used to divide landslide susceptibility into two classes Chen et al. (2013) based on the results of the three algorithms, which are the non-susceptible and susceptible.

This tool was used to create landslide susceptibility zoning maps for each algorithm. The categorization ratings for the three models varied from 0 to 1, representing non-susceptible and susceptible zones. The final LSM produced by RF, XGBoost, and stacked models are shown in Figure 4.6. Even though XGBoost outperformed RF and stacked models, XGBoost and RF generated number of pixels for susceptible and non-susceptible zones with only slight difference. Table 4.12 shows the percentage coverage area of the landslides and non-landslides zones area based on respective model predictions.

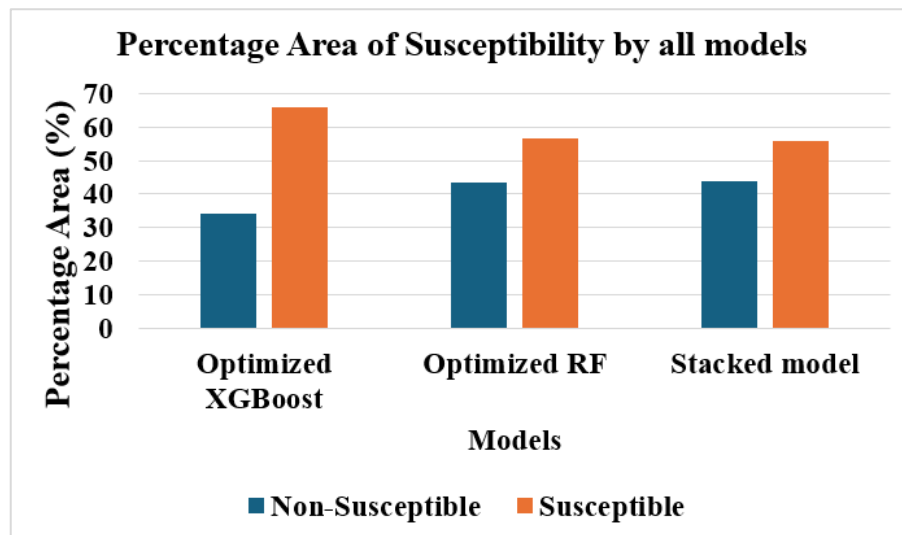


Figure 4.7: Percentage of susceptibility areas covered by LSM generated by all models

Based on Figure 4.7, the stacking of optimized RF and XGBoost models successfully addressed the overestimation issues observed in XGBoost, as evidenced by a percentage difference of less than 1% in the classification of susceptible and non-susceptible areas between the RF and stacked models. This improvement was achieved by maintaining a

balance between true positives and false positives in both models, thereby preventing overestimation.

This finding is supported by Wang et al. (2020), who demonstrated in their study that stacking two models could effectively mitigate overestimation issues. It highlights that the overestimation problem can be minimized through the stacking strategy, particularly when using optimized models. While the optimized XGBoost model outperformed other models in this study, it tended to allocate a disproportionately high percentage of steep slope areas and densely vegetated zones as susceptible, indicating over-prediction in these regions. Notably, the stacking strategy, which combines two models, slightly reduced the overestimation tendency exhibited by XGBoost, resulting in a more balanced and reliable prediction.

According to Abedini and Tulabi (2018), the least amount of territory should typically be assigned to the zone with the highest susceptibility. If a susceptibility model assigns an excessive area to highly susceptible zones, it risks having good accuracy but poor practical relevance, as emphasized by Reichenbach et al. (2018). In this study, while XGBoost demonstrated strong predictive performance, its over-prediction in highly susceptible areas reduced its practical applicability. By contrast, the stacking approach addressed this issue, leading to a more realistic allocation of susceptibility zones and enhancing the model's practical relevance.

4.4.1 Practical Implementation of LSM

The findings of this study hold significant value for real-world applications, particularly in disaster management, urban planning, and environmental conservation. The landslide susceptibility maps produced using the stacking strategy of RF and XGBoost models offer

accurate and balanced predictions, which can serve as vital tools for decision-makers. Urban planners can utilize these maps to identify high-risk areas and implement targeted mitigation strategies, such as slope stabilization, drainage improvement, and the regulation of construction activities in susceptible zones. Additionally, these maps can guide the safe development of infrastructure, ensuring minimal disruption to natural landscapes and reducing the risk of damage caused by landslides.

For local authorities, these findings underscore the importance of integrating machine learning-based susceptibility assessments into regional development planning. The use of a stacking approach can help policymakers allocate resources more effectively by prioritizing interventions in areas identified as highly susceptible. For example, community relocation programs, early warning systems, and the establishment of no-development zones can be planned based on these susceptibility maps.

- **Policy Implications and Recommendations:**

1. **Integration into Regional Planning:** Local governments and planning agencies should incorporate machine learning-based susceptibility maps into land-use planning processes. The adoption of such tools would allow for the identification of potential hazard zones and ensure that new developments comply with safety standards.
2. **Targeted Mitigation Measures:** Policies should be developed to focus on areas classified as high susceptibility, such as steep slopes and densely vegetated regions identified in this study. Measures could include enforcing construction restrictions, improving slope drainage systems, and implementing vegetation management programs to stabilize slopes.

3. **Capacity Building and Awareness:** Local authorities should invest in training programs to enhance technical expertise in using and interpreting machine learning models for landslide susceptibility mapping. In addition, public awareness campaigns could educate communities about landslide risks and encourage proactive measures to minimize their impact.
4. **Monitoring and Early Warning Systems:** Based on the insights from susceptibility maps, authorities should establish real-time monitoring systems in high-risk areas. Early warning systems could be integrated with local disaster response frameworks to reduce casualties and property damage during landslide events.

4.5 Conclusion

The results and discussion chapter had presented all the final findings of the current research study. The VIF and TOL multicollinearity test had confirmed that all the 10 suggested LCFs, DEM, slope angle, slope length, plan curvature, profile curvature, TWI, SPI, NDVI, distance from road, and stream had no multicollinearity issues. In addition, the Boruta model had also confirmed that all the 10 LCFs significantly contribute to the landslide occurrence in Penang Island.

The BO had proven to enhance the performance of the RF and XGBoost models. Two of the single optimized models and stack model (RF-XGBoost) had shown such a great prediction performance that all the models had an AUC score of more than 90%. Somehow, the stacked model of RF-XGBoost had been outperformed by the optimized single RF and XGBoost models. Besides, in the landslide susceptibility map (LSM) had experienced the optimized overestimation issues. The stacked model had produced where the susceptible area is 2% slightly more than the map produced from the single RF and XGboost models.

Somehow, the percentage is considered small and insignificant. Thus, it was proven that overfitting issue in RF and overestimation problem in XGBoost could be overcome by developing a hybrid ensemble model via the stacking of RF-XGBoost and by applying the BO.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Findings and Conclusions

This study used 886 datasets of the landslides and non-landslide points and the ten landslide conditioning factors to predict landslide susceptibility. The datasets were then split randomly into the ratio of 70:15:15 for the training, validating, and testing datasets. The training datasets were used to train the base models. In contrast, the validating datasets were used to generate a confusion matrix and predict the base models. The testing datasets generated a confusion matrix and predicted the stacked model. The results of the conducted experiments provide insight into the prediction efficiency and accuracy of the proposed model with the help of performance measures. The findings of this research are as follows:

- The multi-collinearity test indicates that there was no multi-collinearity among the ten variables. All the variables had a value of TOL of more than 0.1 and a variance inflation factor VIF value of less than 10.
- In addition, based on the Boruta finding, all ten landslide conditioning factors were necessary, and the slope angle was the most important.
- The validation process was done by comparing the produced LSM with the actual landslide occurrences using the AUROC curve, confusion matrix, and statistical methods. Results indicated that the model's performance is high, with prediction rate AUC values of 97.1%, 95.9%, and 95.6% for XGBoost,

RF, and stacked models, respectively. The RF, XGBoost, and stacked models of RF-XGBoost are confirmed to be suitable for use in regional-scale LSM.

- All the developed models had excellent performance where the AUC scores of the models are above 90%. Since all models performed excellently and no significant differences were found in AUC, XGBoost can be considered the best model due to its highest AUC and advanced predictive capabilities. However, the final choice should depend on the specific application. If interpretability or ease of implementation is crucial, RF may be the preferred model. If addressing overestimation and ensuring balanced predictions are priorities, the stacked model might be the better option. Ultimately, all models are strong performers, but XGBoost stands out slightly as the top choice for applications where maximum discrimination accuracy is prioritized.
- The BO method was proven to be an efficient optimization method. This is because the RF and XGBoost model performed the best at their optimum state. Besides, both BO and stacking had successfully minimized the occurrence of overfitting and overestimation.
- From the overall findings, all four objectives of this research study were achieved, and the problems stated were solved.

5.2 Limitations and Recommendations

The accuracy of landslide susceptibility mapping is influenced by the quality and resolution of input data, as public data sources may contain inaccuracies or outdated information. Future studies should focus on developing high-resolution spatial databases using advanced tools like LiDAR and satellite imagery to improve data reliability.

Additionally, the random splitting of datasets in the study may have overlooked spatial autocorrelation, which could affect model performance in new regions. To address this, future research should adopt spatially stratified sampling methods to ensure datasets are representative of real-world conditions. Incorporating dynamic factors, such as rainfall intensity and human activities, is also essential to capture temporal variability and improve prediction accuracy. Expanding the scope of the study to include diverse regions or larger scales will help validate the model's transferability. Enhancing landslide occurrence records through field surveys and crowdsourced data collection is crucial to reduce biases and capture smaller or less reported events. Finally, exploring additional optimization techniques, such as genetic algorithms or particle swarm optimization, could improve model efficiency and robustness, complementing the success of Bayesian Optimization in the study.

Besides, several assumptions were made during the modelling process that could influence the outcomes. For instance, it was assumed that all conditioning factors, such as slope angle and vegetation index, remain constant over time, despite the possibility of dynamic changes due to human activities or natural events. Therefore, it was recommended that future research should focus on incorporating dynamic changes in the conditioning factors to better reflect real-world scenarios. For instance, monitoring land-use changes, vegetation cover fluctuations, and slope modifications due to human activities or natural processes over time could improve the models' accuracy and relevance. Temporal datasets, such as multi-temporal satellite imagery or historical climate data, could be integrated to account for these dynamic variations.

5.3 Summary

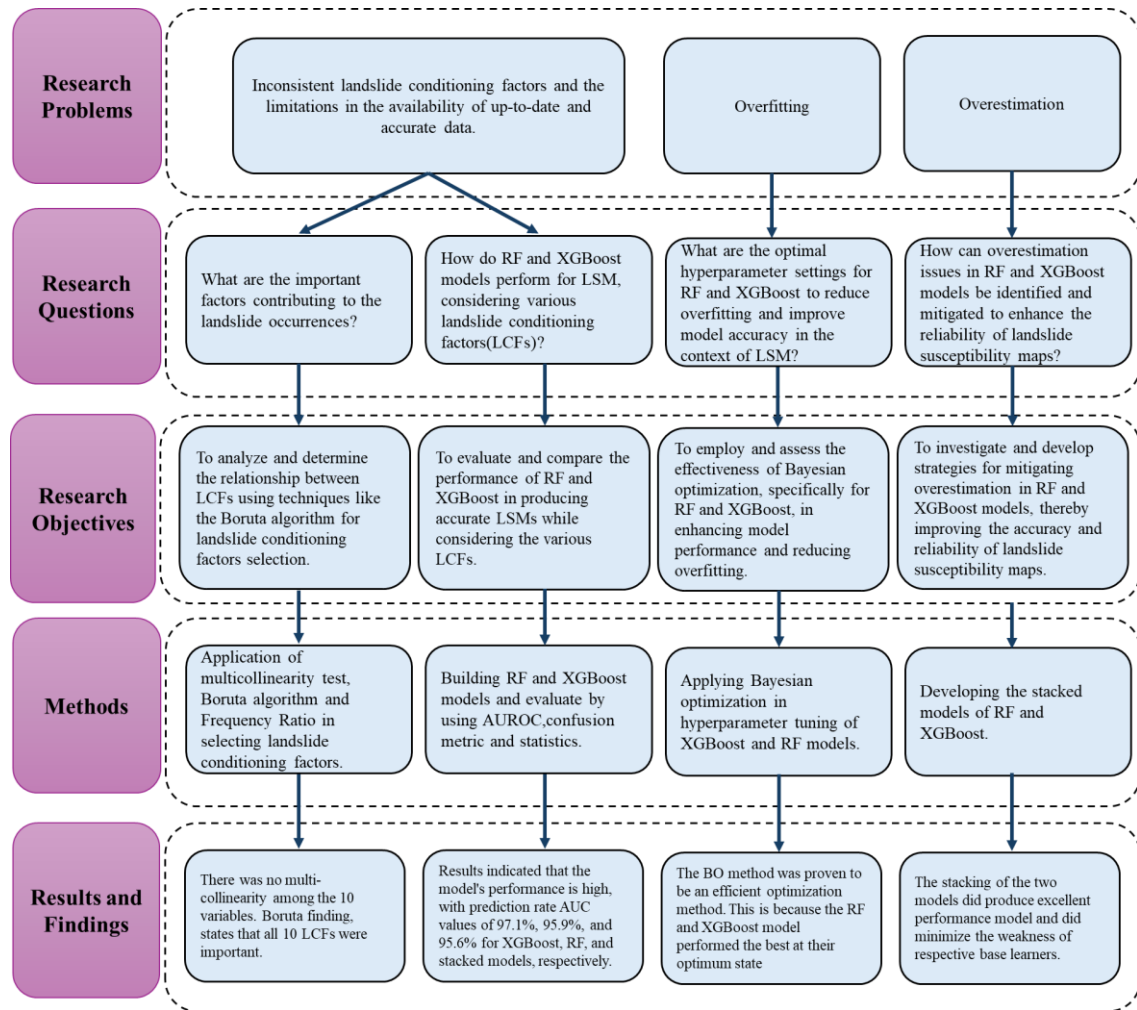


Figure 5.1: Summary of the research problem, questions, objectives, methods, results, and findings.

The overall view of the research stages is shown in Figure 5.1. The study starts by stating the problem and defining the objectives of the research study, which is developing a landslide susceptibility map for Penang Island using the RF and XGBoost hybrid ensemble with BO. The problems were then established by reviewing and interpreting ample literature involving landslide occurrences, LCFs, susceptibility prediction, and mapping processes. Firstly, LCFs had been determined. Then, the feature selection and multicollinearity test techniques were elaborated on in detail. The selection of highly important LCFs ensured the models' excellent prediction performance.

In addition, the landslide susceptibility prediction techniques were then reviewed in detail. It started from commonly used MLs such as LR, SVM, and KNN to more advanced machine learning such as RF, XGBoost, CNN, RNN, and hybrid machine learning. Besides, several optimisation methods, such as manual search, random search, and BO can be used to optimize the performance of the MLs. It has been concluded, citing several references, that hybrid ensemble machine learning has proven to be far more effective in terms of prediction processes than single learners. The suitability of using RF and XGBoost techniques has been established, mainly because the former can address the non-linearity issue efficiently while being a robust model in landslide susceptibility prediction.

Hybrid ensemble machine learning means that two or more machine learners are stacked together with the implementation of hyperparameter optimisation methods. This is mainly because the optimisation of machine learning will ensure the optimum performance of the model, and integrating multiple models will reduce the weakness in the single learner. Thus, in this current study, the integration of RF and XGBoost were stacked, and each model was optimized using the BO method.

The method of the research process was well-designed and explained in detail. This research project has fully utilized open-access data and software such as RStudio, ArcMap, and QGIS. Every package required in RStudio for developing prediction models, analysis, and mapping was all stated. The RF, XGBoost, and stacked models' performances were evaluated and validated using a confusion matrix, statistics, and the AUC of the receiver operating characteristic curve. The results of each model were then discussed and compared.

REFERENCES

- Abbas, F., Zhang, F., Abbas, F., Ismail, M., Iqbal, J., Hussain, D., Khan, G., Alrefaei, A. F., & Albeshr, M. F. (2023). Landslide susceptibility mapping: Analysis of different feature selection techniques with artificial neural network tuned by bayesian and metaheuristic algorithms. *Remote Sensing*, 15(17), 4330. <https://doi.org/10.3390/rs15174330>
- Abd Majid, N., & Rainis, R. (2019). Aplikasi sistem maklumat geografi (GIS) dan analisis diskriminan dalam pemodelan kejadian kegagalan cerun di Pulau Pinang, Malaysia. *Sains Malaysiana*, 48(7), 1367–1381. <https://doi.org/10.17576/jsm-2019-4807-06>
- 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. *Environmental Earth Sciences*, 77, 1–13. <https://doi.org/10.1007/s12665-018-7524-1>
- Acharya, T. D., & Lee, D. H. (2019). Landslide Susceptibility Mapping using relative frequency and predictor rate along Araniko Highway landslide susceptibility. *KSCE Journal of Civil Engineering*, 23, 763-776. <https://doi.org/10.1007/s12205-018-0156-x>
- 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>

- Ada, M., & San, B. T. (2018). Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. *Natural Hazards*, 90(1), 237–263. <https://doi.org/10.1007/s11069-017-3043-8>
- Adnan, M. S. G., Rahman, M. S., Ahmed, N., Ahmed, B., Rabbi, M. F., & Rahman, R. M. (2020). Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Remote Sensing*, 12(20), 1–23. <https://doi.org/10.3390/rs12203347>
- Ado, M., Amitab, K., Maji, A. K., Jasińska, E., Gono, R., Leonowicz, Z., & Jasiński, M. (2022). Landslide Susceptibility Mapping Using Machine Learning: A Literature Survey. *Remote Sensing*, 14(13), 3029. <https://doi.org/10.3390/rs14133029>
- 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, 1-20. <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. *Bulletin of Engineering Geology and the Environment*, 82(5), 170. <https://doi.org/10.1007/s10064-023-03188-2>
- Ahmad, I., Dar, M. A., Teka, A. H., Gebre, T., Gadissa, E., & Tolosa, A. T. (2019). Application of hydrological indices for erosion hazard mapping using spatial analyst tool. *Environmental Monitoring and Assessment*, 191(8), 482.

<https://doi.org/10.1007/s10661-019-7614-x>

- 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. *Sustainability*, 13(18), 10110. <https://doi.org/10.3390/su131810110>
- Akinci, H., & Zeybek, M. (2021). Comparing classical statistic and machine learning models in landslide susceptibility mapping in Ardanuc (Artvin), Turkey. *Natural Hazards*, 108(2), 1515–1543. <https://doi.org/10.1007/s11069-021-04743-4>
- Al-Najjar, H. A. H., Kalantar, B., Pradhan, B., & Saeidi, V. (2019). Conditioning factor determination for mapping and prediction of landslide susceptibility using machine learning algorithms. *Earth Resources and Environmental Remote Sensing/GIS Applications X*, 11156, 97–107. <https://doi.org/10.1117/12.2532687>
- 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). *Korean Society of Engineering Geology*, 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. <https://doi.org/10.1016/j.envsoft.2023.105657>
- Alin, A. (2010). Multicollinearity. *WIREs Computational Statistics*, 2(3), 370–374.

<https://doi.org/10.1002/wics.84>

Alqadhi, S., Mallick, J., Talukdar, S., Bindajam, A. A., Van Hong, N., & Saha, T. K. (2022). Selecting optimal conditioning parameters for landslide susceptibility: An experimental research on Aqabat Al-Sulbat, Saudi Arabia. *Environmental Science and Pollution Research*, 29, 3743–3762. <https://doi.org/10.1007/s11356-021-15886-z>

Althuwaynee, O. F., Pradhan, B., & Lee, S. (2012). Application of an evidential belief function model in landslide susceptibility mapping. *Computers and Geosciences*, 44, 120–135. <https://doi.org/10.1016/j.cageo.2012.03.003>

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>

Arabameri, A., Chandra Pal, S., Costache, R., Saha, A., Rezaie, F., Seyed Danesh, A., Pradhan, B., Lee, S., & Hoang, N. D. (2021). Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms. *Geomatics, Natural Hazards and Risk*, 12(1), 469–498. <https://doi.org/10.1080/19475705.2021.1880977>

Arabameri, A., Chandra Pal, S., Rezaie, F., Chakraborty, R., Saha, A., Blaschke, T., Di Napoli, M., Ghorbanzadeh, O., & Thi Ngo, P. T. (2022). Decision tree based ensemble machine learning approaches for landslide susceptibility mapping. *Geocarto International*, 37(16), 4594–4627.

<https://doi.org/10.1080/10106049.2021.1892210>

ASEAN (2021). *On Integrating Climate Change Projection into Landslide Risk Assesment & Mapping*. ASEAN Secretariat.

Atkinson, P. M., & Massari, R. (1998). Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Computers & Geosciences*, 24(4), 373–385. [https://doi.org/10.1016/S0098-3004\(97\)00117-9](https://doi.org/10.1016/S0098-3004(97)00117-9)

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*, 11(10), 2076. <https://doi.org/10.3390/w11102076>

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>

Ayenew, T., & Barbieri, G. (2005). Inventory of landslides and susceptibility mapping in the Dessie area, northern Ethiopia. *Engineering Geology*, 77, 1–15. <https://doi.org/10.1016/j.enggeo.2004.07.002>

Bardenet, R., Brendel, M., Kégl, B., & Sebag, M. (2013). Collaborative hyperparameter tuning. In *International Conference on Machine Learning*, 199–207.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(2), 281–305.

<http://jmlr.org/papers/v13/bergstra12a.html>

- Berk, J., Gupta, S., Rana, S., & Venkatesh, S. (2020). Randomised gaussian process upper confidence bound for bayesian optimisation. *arXiv preprint arXiv:2006.04296*.
<https://doi.org/10.48550/arXiv.2006.04296>
- Betrò, B. (1991). Bayesian methods in global optimization. *Journal of Global Optimization*, 1, 1–14. <https://doi.org/10.1007/BF00120661>
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, 24(1), 43–69.
<https://doi.org/10.1080/02626667909491834>
- Bizimana, H., & Sonmez, O. (2015). Landslide occurrences in the hilly areas of Rwanda, their causes and protection measures. *Disaster Science and Engineering*, 1(1), 1–7.
- 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>
- Brabb, E. E. (1985). Innovative approaches to landslide hazard and risk mapping. *International Landslide Symposium Proceedings, Toronto, Canada*, 1, 17–22.
<https://www.cabidigitallibrary.org/doi/full/10.5555/19861906693>

- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Brenning, A. (2005). Spatial prediction models for landslide hazards: Review, comparison and evaluation. *Natural Hazards and Earth System Sciences*, 5(6), 853–862. <https://doi.org/10.5194/nhess-5-853-2005>
- Budha, A. K. S. D. P. (2018). GIS Based Landslide Susceptibility Mapping along the Road Section from Bandeu to Barahabise, Sindhupal Chowk District of Nepal. *International Journal of Science and Research (IJSR)*, 7(11), 465–471.
<https://doi.org/10.21275/ART20192474>
- Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B., Revhaug, I., Tien Bui, D., Tuan, T. A., Klempe, H., Pradhan, B., & Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13(2), 361–378. <https://doi.org/10.1007/s10346-015-0557-6>
- Bui, D., Tsangaratos, P., Tien, N., Liem, N., & Trinh, P. (2020). Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *Catena*, 188, 104426. <https://doi.org/10.1016/j.catena.2019.104426>
- 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. *Applied Sciences*, 11(11), 4993.

<https://doi.org/10.3390/app11114993>

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>

Çellek, S. (2022). Effect of the slope angle and its classification on landslides. *Himalayan Geology*, 43(1), 85–95. <https://doi.org/10.5194/nhess-2020-87>

Chang, K.-T., Merghadi, A., Yunus, A. P., Pham, B. T., & Dou, J. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Scientific Reports*, 9(1), 12296. <https://doi.org/10.1038/s41598-019-48773-2>

Chaudhary, N., Abu-Odeh, A., Karaman, I., & Arróyave, R. (2017). A data-driven machine learning approach to predicting stacking faulting energy in austenitic steels. *Journal of Materials Science*, 52(18), 11048–11076. <https://doi.org/10.1007/s10853-017-1252-x>

Chen, H., Zeng, Z., & Tang, H. (2015). Landslide Deformation Prediction Based on Recurrent Neural Network. *Neural Processing Letters*, 41(2), 169–178. <https://doi.org/10.1007/s11063-013-9318-5>

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->

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD, San Francisco, CA, USA*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, T., Zhu, L., Niu, R., Trinder, C. J., Peng, L., & Lei, T. (2020). Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models. *Journal of Mountain Science*, 17(3), 670–685. <https://doi.org/10.1007/s11629-019-5839-3>
- 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., & Li, Y. (2020). GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *Catena*, 195, 104777. <https://doi.org/10.1016/j.catena.2020.104777>
- Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., Zhu, A. X., Pei, X., & Duan, Z. (2018). Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Science of the Total Environment*, 626, 1121–1135. <https://doi.org/10.1016/j.scitotenv.2018.01.124>

- Chen, W., Xie, X., Peng, J., Wang, J., Duan, Z., & Hong, H. (2017). GIS-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models. *Geomatics, Natural Hazards and Risk*, 8(2), 950–973. <https://doi.org/10.1080/19475705.2017.1289250>
- 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>
- Chen, W., Zhang, S., Li, R., & Shahabi, H. (2018). Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve bayes tree for landslide susceptibility modeling. *Science of the Total Environment*, 644, 1006–1018. <https://doi.org/10.1016/j.scitotenv.2018.06.389>
- Chen, W., Zhao, X., Shahabi, H., Shirzadi, A., Khosravi, K., Chai, H., Zhang, S., Zhang, L., Ma, J., Chen, Y., Wang, X., Bin Ahmad, B., & Li, R. (2019). Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto International*, 34(11), 1177–1201. <https://doi.org/10.1080/10106049.2019.1588393>
- Cheng, G., Guo, L., Zhao, T., Han, J., Li, H., & Fang, J. (2013). Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA. *International Journal of Remote Sensing*, 34(1), 45–59. <https://doi.org/10.1080/01431161.2012.705443>

- Chimidi, G., Raghuvanshi, T. K., & Suryabhagavan, K. V. (2017). Landslide hazard evaluation and zonation in and around Gimbi town, western Ethiopia—a GIS-based statistical approach. *Applied Geomatics*, 9(4), 219–236. <https://doi.org/10.1007/s12518-017-0195-x>
- Chowdhury, M. S. (2023). Modelling hydrological factors from DEM using GIS. *MethodsX*, 10, 102062. <https://doi.org/10.1016/j.mex.2023.102062>
- Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., & Smith, J. T. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, 73(2), 209–263. <https://doi.org/10.1007/s10064-013-0538-8>
- Cruz-Ramírez, N., Acosta-Mesa, H. G., Mezura-Montes, E., Guerra-Hernández, A., Hoyos-Rivera, G. de J., Barrientos-Martínez, R. E., Gutiérrez-Fragoso, K., Nava-Fernández, L. A., González-Gaspar, P., Novoa-del-Toro, E. M., Aguilera-Rueda, V. J., & Ameca-Alducin, M. Y. (2014). How Good Is Crude MDL for Solving the Bias-Variance Dilemma? An empirical investigation based on Bayesian networks. *PloS one*, 9(3), e92866. <https://doi.org/10.1371/journal.pone.0092866>
- Dai, F. C., Lee, C. F., Li, J., & Xu, Z. W. (2001). Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmental Geology*, 40(3), 381–391. <https://doi.org/10.1007/s002540000163>
- Dai, Z., Yu, H., Low, B. K. H., & Jaillet, P. (2019). Bayesian optimization meets Bayesian

- optimal stopping. *International Conference on Machine Learning*, 97, 1496–1506. <https://proceedings.mlr.press/v97/dai19a.html>
- Dao, D. Van, Jaafari, A., Bayat, M., Mafi-Gholami, D., Qi, C., Moayedi, H., Phong, T. Van, Ly, H.-B., Le, T.-T., Trinh, P. T., Luu, C., Quoc, N. K., Thanh, B. N., & Pham, B. T. (2020). A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena*, 188, 104451. <https://doi.org/10.1016/j.catena.2019.104451>
- Das, S., Bora, P. K., & Das, R. (2022). Estimation of slope length gradient (LS) factor for the sub-watershed areas of Juri River in Tripura. *Modeling Earth Systems and Environment*, 8(1), 1171–1177. <https://doi.org/10.1007/s40808-021-01153-0>
- Daviran, M., Shamekhi, M., Ghezelbash, R., & Maghsoudi, A. (2023). Landslide susceptibility prediction using artificial neural networks, SVMs and random forest: Hyperparameters tuning by genetic optimization algorithm. *International Journal of Environmental Science and Technology*, 20(1), 259–276. <https://doi.org/10.1007/s13762-022-04491-3>
- DeLong, E.R, DeLong , D.M., & Clarke-Pearson, D. L. (1998) . Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), 837-845.
- 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>

Divina, F., Gilson, A., Gómez-Vela, F., Garcia Torres, M., & Torres, J. (2018). Stacking Ensemble learning for short-term electricity consumption forecasting. *Energies*, 11, 949. <https://doi.org/10.3390/en11040949>

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>

Dou, J., Tien Bui, D., P. Yunus, A., Jia, K., Song, X., Revhaug, I., Xia, H., & Zhu, Z. (2015). Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. *PloS one*, 10(7), e0133262. <https://doi.org/10.1371/journal.pone.0133262>

Dou, J., Yunus, A. P., Bui, D. T., Merghadi, A., Sahana, M., Zhu, Z., Chen, C. W., Han, Z., & Pham, B. T. (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>

Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C. W., Khosravi, K., Yang, Y., & Pham, B. T. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide

- susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of the Total Environment*, 662, 332–346. <https://doi.org/10.1016/j.scitotenv.2019.01.221>
- Dou, J., Yunus, A. P., Tien Bui, D., Sahana, M., Chen, C.-W., Zhu, Z., Wang, W., & Pham, B. T. (2019). Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. *Remote Sensing*, 11(6), 638. <https://doi.org/10.3390/rs11060638>
- Dou, J., Yunus, A. P., Xu, Y., Zhu, Z., Chen, C.-W., Sahana, M., Khosravi, K., Yang, Y., & Pham, B. T. (2019). Torrential rainfall-triggered shallow landslide characteristics and susceptibility assessment using ensemble data-driven models in the Dongjiang Reservoir Watershed, China. *Natural Hazards*, 97(2), 579–609. <https://doi.org/10.1007/s11069-019-03659-4>
- Duan, J., Asteris, P., Nguyen, H., Bui, X.-N., & Moayedi, H. (2021). A Novel Artificial Intelligence Technique to Predict Compressive Strength of Recycled Aggregate Concrete Using ICA-XGBoost Model. *Engineering With Computers*, 37, 3329–3346. <https://doi.org/10.1007/s00366-020-01003-0>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- 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), 345–359. <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>
- Fang, Z., Wang, Y., Peng, L., & Hong, H. (2020). A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *International Journal of Geographical Information Science*, 35(2), 321–347. <https://doi.org/10.1080/13658816.2020.1808897>
- 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/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2), 337–407. <http://dx.doi.org/10.1214/aos/1016218223>
- Gao, H., Fam, P. S., Tay, L. T., & Low, H. C. (2020). Three oversampling methods applied in a comparative landslide spatial research in Penang Island, Malaysia. *SN Applied Sciences*, 2(9). <https://doi.org/10.1007/s42452-020-03307-8>

- Garrido-Merchán, E. C., & Hernández-Lobato, D. (2020). Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes. *Neurocomputing*, 380, 20–35. <https://doi.org/10.1016/j.neucom.2019.11.004>
- Ghasemian, A., Hosseinmardi, H., & Clauset, A. (2019). Evaluating overfit and underfit in models of network community structure. *IEEE Transactions on Knowledge and Data Engineering*, 32(9), 1722–1735. <https://doi.org/10.1109/TKDE.2019.2911585>
- Gruber, S., & Haeberli, W. (2007). Permafrost in steep bedrock slopes and its temperature-related destabilization following climate change. *Journal of Geophysical Research: Earth Surface*, 112(2). <https://doi.org/10.1029/2006JF000547>
- Guerriero, L., Confuorto, P., Calcaterra, D., Guadagno, F., Revellino, P., & Di Martire, D. (2019). PS-driven inventory of town-damaging landslides in the Benevento, Avellino and Salerno Provinces, southern Italy. *Journal of Maps*, 15, 619–625. <https://doi.org/10.1080/17445647.2019.1651770>
- 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>
- Gupta, A., Khan, R. U., Singh, V. K., Tanveer, M., Kumar, D., Chakraborti, A., & Pachori, R. B. (2020). A novel approach for classification of mental tasks using multiview ensemble learning (MEL). *Neurocomputing*, 417, 558–584.

<https://doi.org/10.1016/j.neucom.2020.07.050>

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, 31(1), 181–216. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)

Gyamerah, S. A., Mensah, A. A., Asare, C., & Dzupire, N. (2023). Improving mortality forecasting using a hybrid of Lee–Carter and stacking ensemble model. *Bulletin of the National Research Centre*, 47(1), 158. <https://doi.org/10.1186/s42269-023-01138-2>

Habibi, A., Delavar, M. R., Sadeghian, M. S., Nazari, B., & Pirasteh, S. (2023). A hybrid of ensemble machine learning models with RFE and Boruta wrapper-based algorithms for flash flood susceptibility assessment. *International Journal of Applied Earth Observation and Geoinformation*, 122, 103401. <https://doi.org/10.1016/j.jag.2023.103401>

Hakim, W. L., Rezaie, F., Nur, A. S., Panahi, M., Khosravi, K., Lee, C. W., & Lee, S. (2022). Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in Icheon, South Korea. *Journal of Environmental Management*, 305, 114367. <https://doi.org/10.1016/j.jenvman.2021.114367>

Hammad Khaliq, A., Basharat, M., Talha Riaz, M., Tayyib Riaz, M., Wani, S., Al-Ansari, N., Ba Le, L., & Thi Thuy Linh, N. (2022). Spatiotemporal landslide susceptibility mapping using machine learning models: A case study from district Hattian Bala,

- NW Himalaya, Pakistan. *Ain Shams Engineering Journal*, 14(3), 101907.
<https://doi.org/10.1016/j.asej.2022.101907>
- Hamza, T., & Raghuvanshi, T. K. (2017). GIS based landslide hazard evaluation and zonation – A case from Jeldu District, Central Ethiopia. *Journal of King Saud University - Science*, 29(2), 151–165. <https://doi.org/10.1016/j.jksus.2016.05.002>
- 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>
- Hashim, B. M., Sultan, M. A., Attyia, M. N., Al Maliki, A. A., & Al-Ansari, N. (2019). Change detection and impact of climate changes to Iraqi southern marshes using Landsat 2 MSS, Landsat 8 Oli and sentinel 2 MSI data and GIS applications. *Applied Sciences*, 9(10), 2016. <https://doi.org/10.3390/app9102016>
- Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17(2), 24016. <https://doi.org/10.1088/1748-9326/ac4d4f>
- He, S., Pan, P., Dai, L., Wang, H., & Liu, J. (2012). Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. *Geomorphology*, 171, 30–41.
<https://doi.org/10.1016/j.geomorph.2012.04.024>
- Hearn, G. J., & Hart, A. B. (2019). Landslide susceptibility mapping: A practitioner's view. *Bulletin of Engineering Geology and the Environment*, 78(8), 5811–5826.
<https://doi.org/10.1007/s10064-019-01506-1>

- Hong, H., Pourghasemi, H. R., & Pourtaghi, Z. S. (2016). Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology*, 259, 105–118. <https://doi.org/10.1016/j.geomorph.2016.02.012>
- Hong, H., Tsangaratos, P., Ilia, I., Chen, W., & Xu, C. (2017). Comparing the Performance of a Logistic Regression and a Random Forest Model in Landslide Susceptibility Assessments. the Case of Wuyaun Area, China. In *Advancing Culture of Living with Landslides: Volume 2 Advances in Landslide Science*, (pp.1043-1050). Springer International Publishing. https://doi.org/10.1007/978-3-319-53498-5_118
- Hu, X., Zhang, H., Mei, H., Xiao, D., Li, Y., & Li, M. (2020). Landslide susceptibility mapping using the stacking ensemble machine learning method in lushui, southwest China. *Applied Sciences*, 10(11), 4016. <https://doi.org/10.3390/app10114016>
- Huang, F., Cao, Z., Guo, J., Jiang, S. H., Li, S., & Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena*, 191, 104580. <https://doi.org/10.1016/j.catena.2020.104580>
- Huang, F., Yin, K., Huang, J., Gui, L., & Wang, P. (2017). Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Engineering Geology*, 223, 11–22. <https://doi.org/10.1016/j.enggeo.2017.04.013>
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002).

- Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1), 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- 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), 4129. <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>
- Jabatan Perancang Bandar dan Desa Negeri Pulau Pinang. (2019). *Rancangan Struktur Negeri Pulau Pinang 2030*. 1–146. [https://jpbdd.penang.gov.my/images/faris/pdf/RSN dan RKK/RANCANGAN STRUKTUR NEGERI PULAU PINANG 2030 BM.pdf](https://jpbdd.penang.gov.my/images/faris/pdf/RSN_dan_RKK/RANCANGAN_STRUKTUR_NEGERI_PULAU_PINANG_2030_BM.pdf)
- Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4), 455–492. <https://doi.org/10.1023/A:1008306431147>

- Kablan, R., Miller, H. A., Suliman, S., & Frieboes, H. B. (2023). Evaluation of stacked ensemble model performance to predict clinical outcomes: A COVID-19 study. *International Journal of Medical Informatics*, 175, 105090. <https://doi.org/10.1016/j.ijmedinf.2023.105090>
- 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>
- Kadirhodjaev, A., Kadavi, P. R., Lee, C. W., & Lee, S. (2018). Analysis of the relationships between topographic factors and landslide occurrence and their application to landslide susceptibility mapping: A case study of Mingchukur, Uzbekistan. *Geosciences Journal*, 22(6), 1053–1067. <https://doi.org/10.1007/s12303-018-0052-x>
- Kanungo, D., Arora, M., Sarkar, S., & Gupta, R. P. (2006). A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology*, 85, 347–366. <https://doi.org/10.1016/j.enggeo.2006.03.004>
- Kavzoglu, T., Colkesen, I., & Sahin, E. (2019). Machine Learning Techniques in landslide susceptibility mapping: A Survey and a Case Study. In *Advances in Natural and Technological Hazards Research*, 283–301. https://doi.org/10.1007/978-3-319-77377-3_13
- Kementerian Kerja Raya, J. K. R. (2009). *National Slope Master Plan : Kajian Pelan*

Induk Langkah-Langkah Pembaikan Cerun di Malaysia (Kajian Pelan Induk Cerun Negara) 2009-2023.

- 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/10.1016/j.jhydrol.2020.125552>
- Kim, J. C., Lee, S. S., Jung, H. S., & Lee, S. S. (2018). Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto International*, 33(9), 1000–1015. <https://doi.org/10.1080/10106049.2017.1323964>
- Korzh, O., Cook, G., Andersen, T., & Serra, E. (2017). Stacking approach for CNN transfer learning ensemble for remote sensing imagery. *2017 Intelligent Systems Conference (IntelliSys)*, 599–608. <https://doi.org/10.1109/IntelliSys.2017.8324356>
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software*, 36(11), 1–13. <https://doi.org/10.18637/jss.v036.i11>
- Kwon, H., Park, J., & Lee, Y. (2019). Stacking Ensemble Technique for Classifying Breast Cancer. *Healthcare informatics research*, 25(4), 283–288. <https://doi.org/10.4258/hir.2019.25.4.283>
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*,

363–374. <https://doi.org/10.2307/2529786>

Lee, J.-H., Sameen, M. I., Pradhan, B., & Park, H.-J. (2018). Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology*, 303, 284–298. <https://doi.org/10.1016/j.geomorph.2017.12.007>

Lee, S. (2019). Current and Future Status of GIS-based landslide susceptibility mapping: A Literature Review. *Korean Journal of Remote Sensing*, 35(1), 179–193. <https://doi.org/10.7780/kjrs.2019.35.1.12>

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>

Lee, S., & Sambath, T. (2006). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology*, 50, 847–855. <https://doi.org/10.1007/s00254-006-0256-7>

Lei, X., Chen, W., & Pham, B. T. (2020). Performance evaluation of GIS-based artificial intelligence approaches for landslide susceptibility modeling and spatial patterns analysis. *ISPRS International Journal of Geo-Information*, 9(7), 1–20. <https://doi.org/10.3390/ijgi9070443>

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136.

<https://doi.org/10.1057/palgrave.jors.2601545>

- 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, H., Xu, Y., Zhou, J., Wang, X., Yamagishi, H., & Dou, J. (2020). Preliminary analyses of a catastrophic landslide occurred on July 23, 2019, in Guizhou Province, China. *Landslides*, 17(3), 719–724. <https://doi.org/10.1007/s10346-019-01334-0>
- Li, W., Fang, Z., & Wang, Y. (2022). Stacking ensemble of deep learning methods for landslide susceptibility mapping in the Three Gorges Reservoir area, China. *Stochastic Environmental Research and Risk Assessment*, 36(8), 2207–2228. <https://doi.org/10.1007/s00477-021-02032-x>
- Liu, J., Wang, J., Xiong, J., Cheng, W., Sun, H., Yong, Z., & Wang, N. (2021). Hybrid Models Incorporating Bivariate Statistics and Machine Learning Methods for Flash Flood Susceptibility Assessment Based on Remote Sensing Datasets. *Remote Sensing*, 13(23), 4945. <https://doi.org/10.3390/rs13234945>
- Masrur Ahmed, A. A., Deo, R. C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., & Yang, L. (2021). Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. *Journal of Hydrology*, 599, 126350. <https://doi.org/10.1016/j.jhydrol.2021.126350>

- Matthew, W. (2011). Bias of the random forest out-of-bag (OOB) error for certain input parameters. *Open Journal of Statistics*, 2011. <http://dx.doi.org/10.4236/ojs.2011.13024>
- Mc, M., MORACI, N., Rosone, M., & Farulla, C. (2016). Experimental study of a naturally weathered stiff clay. *Canadian Geotechnical Journal*, 53(12), 2047-2057. <https://doi.org/10.1139/cgj-2016-0175>
- 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/10.1016/j.earscirev.2020.103225>
- Mezaal, M. R., Pradhan, B., Sameen, M. I., Shafri, H. Z. M., & Yusoff, Z. M. (2017). Optimized neural architecture for automatic landslide detection from high-resolution airborne laser scanning data. *Applied Sciences*, 7(7), 730. <https://doi.org/10.3390/app7070730>
- Mohandoss, D. P., Shi, Y., & Suo, K. (2021). Outlier prediction using random forest classifier. *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, 27–33.

<https://doi.org/10.1109/CCWC51732.2021.9376077>

- Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1), 3–30. <https://doi.org/10.1002/hyp.3360050103>
- Mumtaz, R., Amin, A., Khan, M., Asif, M., Anwar, Z., & Bashir, M. J. (2023). Impact of Green Energy Transportation Systems on Urban Air Quality: A Predictive Analysis Using Spatiotemporal Deep Learning Techniques. *Energies*, 16, 6087. <https://doi.org/10.3390/en16166087>
- Nachappa, T. G., Piralilou, S. T., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., Blaschke, T., Gudiyangada Nachappa, T., Tavakkoli Piralilou, S., 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>
- Naghibi, S. A., Pourghasemi, H. R., & Dixon, B. (2015). GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environmental Monitoring and Assessment*, 188(1), 44. <https://doi.org/10.1007/s10661-015-5049-6>
- Nakileza, B. R., & Nedala, S. (2020). Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda. *Geoenvironmental Disasters*, 7, 1-13. <https://doi.org/10.1186/s40677-020-00160-0>

- 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>
- Nandi, A., & Shakoor, A. (2010). A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology*, 110(1-2), 11–20. <https://doi.org/10.1016/j.enggeo.2009.10.001>
- 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., & Egwunwu, E. M. (2022). Prediction of spatial likelihood of shallow landslide using GIS-based machine learning in Awgu, Southeast/Nigeria. *Sustainability*, 14(19), 1–20. <https://doi.org/10.3390/su141912000>
- Nseka, D., Kakembo, V., Bamutaze, Y., & Mugagga, F. (2019). Analysis of topographic parameters underpinning landslide occurrence in Kigezi highlands of southwestern Uganda. *Natural Hazards*, 99(2), 973–989. <https://doi.org/10.1007/s11069-019-03787-x>
- O'brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation

- Factors. *Quality & Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Oh, H.-J., Kadavi, P. R., Lee, C.-W., & Lee, S. (2018). Evaluation of landslide susceptibility mapping by evidential belief function, logistic regression and support vector machine models. *Geomatics, Natural Hazards and Risk*, 9(1), 1053–1070. <https://doi.org/10.1080/19475705.2018.1481147>
- Paoletti, M. E., Haut, J. M., Plaza, J., & Plaza, A. (2018). A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 120–147. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2017.11.021>
- Pardeshi, S. D., Autade, S. E., & Pardeshi, S. S. (2013). Landslide hazard assessment recent trends and techniques. *SpringerPlus*, 2(1), 523. <https://doi.org/10.1186/2193-1801-2-523>
- 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), 942. <https://doi.org/10.3390/app9050942>
- Parmar, A., Katariya, R., & Patel, V. (2019). A review on random forest: An ensemble classifier. *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, 758–763. https://doi.org/10.1007/978-3-030-03146-6_86
- Pašek, J. (1975). Landslides inventory. *Bulletin of the International Association of Engineering Geology - Bulletin de l'Association Internationale de Géologie de*

l'Ingénieur, 12(1), 73–74. <https://doi.org/10.1007/BF02635432>

Pelikan, M. (2005). Bayesian optimization algorithm. *Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms*, 31–48. https://doi.org/10.1007/978-3-540-32373-0_3

Pham, B. T., Phong, T. Van, Nguyen-Thoi, T., Parial, K., K. Singh, S., Ly, H. B., Nguyen, K. T., Ho, L. S., Le, H. Van, & Prakash, I. (2022). Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto International*, 37(3), 735–757. <https://doi.org/10.1080/10106049.2020.1737972>

Pham, B. T., Pradhan, B., Bui, D. T., Prakash, I., & Dholakia, M. B. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environmental Modelling & Software*, 84, 240–250. <https://doi.org/10.1016/j.envsoft.2016.07.005>

Pham, B. T., Tien Bui, D., Pourghasemi, H. R., Indra, P., & Dholakia, M. B. (2017). Landslide susceptibility assessment 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(1–2), 255–273. <https://doi.org/10.1007/s00704-015-1702-9>

Pham, B. T., Tien Bui, D., Prakash, I., & Dholakia, M. B. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena*,

149, 52–63. <https://doi.org/10.1016/j.catena.2016.09.007>

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., Gayen, A., Park, S., Lee, C.-W., & Lee, S. (2018). Assessment of landslide-prone areas and their zonation using logistic regression, logitboost, and naïvebayes machine-learning algorithms. *Sustainability*, 10(10), 3697. <https://doi.org/10.3390/su10103697>

Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T., & Cerda, A. (2020). Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-60191-3>

Pourghasemi, H. R., Moradi, H. R., & Fatemi Aghda, S. M. (2013). Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural Hazards*, 69(1), 749–779. <https://doi.org/10.1007/s11069-013-0728-5>

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. (2010). Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software*, 25(6), 747–759. <https://doi.org/10.1016/j.envsoft.2009.10.016>
- 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/10.1016/j.compenvurbsys.2009.12.004>
- Prasad, P., Loveson, V. J., Das, B., & Kotha, M. (2022). Novel ensemble machine learning models in flood susceptibility mapping. *Geocarto International*, 37(16), 4571–4593. <https://doi.org/10.1080/10106049.2021.1892209>
- Prasad, P., Loveson, V. J., Das, S., & Chandra, P. (2021). Artificial intelligence approaches for spatial prediction of landslides in mountainous regions of western India. *Environmental Earth Sciences*, 80(21), 720. <https://doi.org/10.1007/s12665-021-10033-w>
- 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), 483. <https://doi.org/10.3390/geosciences10120483>

- Raghuvanshi, T. K., Ibrahim, J., & Ayalew, D. (2014). Slope stability susceptibility evaluation parameter (SSEP) rating scheme – An approach for landslide hazard zonation. *Journal of African Earth Sciences*, 99, 595–612. <https://doi.org/10.1016/j.jafrearsci.2014.05.004>
- Rahmati, O., Kalantari, Z., Samadi, M., Uuemaa, E., Moghaddam, D. D., Nalivan, O. A., Destouni, G., & Tien Bui, D. (2019). GIS-Based Site Selection for Check Dams in Watersheds: Considering Geomorphometric and Topo-Hydrological Factors. *Sustainability* 11(20), 5639 . <https://doi.org/10.3390/su11205639>
- Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H. R., & Feizizadeh, B. (2017). Evaluating the influence of geo-environmental factors on gully erosion in a semi-arid region of Iran: An integrated framework. *The Science of the Total Environment*, 579, 913–927. <https://doi.org/10.1016/j.scitotenv.2016.10.176>
- Rana, S., Li, C., Gupta, S., Nguyen, V., & Venkatesh, S. (2017). High dimensional Bayesian optimization with elastic Gaussian process. *International Conference on Machine Learning*, 2883–2891. <https://proceedings.mlr.press/v70/rana17a.html>
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Renard, K. G., Foster, G. R., Yoder, D. C., & McCool, D. K. (1994). RUSLE revisited: Status, questions, answers, and the future. *Journal of Soil and Water Conservation*, 49(3), 213–220.

- Rong, G., Li, K., Su, Y., Tong, Z., Liu, X., Zhang, J., Zhang, Y., & Li, T. (2021). Comparison of tree-structured parzen estimator optimization in three typical neural network models for landslide susceptibility assessment. *Remote Sensing*, 13(22), 4694. <https://doi.org/10.3390/rs13224694>
- Saaty, T. (2008). Decision making with the Analytic Hierarchy Process. *International Journal of services sciences*, 1(1), 83–98. <https://doi.org/10.1504/IJSSCI.2008.017590>
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1249. <https://doi.org/10.1002/widm.1249>
- 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>
- Saha, S., Roy, J., Pradhan, B., & Hembram, T. K. (2021). Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at East Sikkim Himalayan, India. *Advances in Space Research*, 68(7), 2819–2840. <https://doi.org/10.1016/j.asr.2021.05.018>
- Sahana, M., & Sajjad, H. (2017). Evaluating effectiveness of frequency ratio, fuzzy logic and logistic regression models in assessing landslide susceptibility: a case from Rudraprayag district, India. *Journal of Mountain Science*, 14(11), 2150–2167. <https://doi.org/10.1007/s11629-017-4404-1>
- 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., Bui, D. T., Alamri, A. M., Al-Zuhairi, M., Pradhan, B., Bui, D. T., Alamri, A. M., Sameen, M. I., Pradhan, B., Bui, D. T., & Alamri, A. M. (2020). Systematic sample subdividing strategy for training landslide susceptibility models. *Catena*, 187, 104358. <https://doi.org/10.1016/j.catena.2019.104358>

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>

Sandha, S. S., Aggarwal, M., Fedorov, I., & Srivastava, M. (2020). Mango: A Python library for parallel hyperparameter tuning. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3987–3991. <https://doi.org/10.1109/ICASSP40776.2020.9054609>

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>

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>

Shahabi, H., Khezri, S., Ahmad, B. Bin, & Hashim, M. (2014). Retracted: Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. *Catena*, 115, 55–70. <https://doi.org/10.1016/j.catena.2013.11.014>

Shirzadi, A., Soliamani, K., Habibnejhad, M., Kavian, A., Chapi, K., Shahabi, H., Chen, W., Khosravi, K., Pham, B. T., Pradhan, B., Ahmad, A., Bin Ahmad, B., Bui, D. T., Thai Pham, B., Pradhan, B., Ahmad, A., Bin Ahmad, B., & Tien Bui, D. (2018). Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors*, 18(11), 3777. <https://doi.org/10.3390/s18113777>

Shorfuzzaman, M. (2023). IoT-enabled stacked ensemble of deep neural networks for the diagnosis of COVID-19 using chest CT scans. *Computing*, 105(4), 887–908. <https://doi.org/10.1007/s00607-021-00971-5>

Sidaway, J. (2022). Working Principles of Digital Elevation Model and its Applications. *Journal of Remote Sensing and GIS*, 11(6), 237. <https://doi.org/10.35248/2469-4134.22.11.237>.

Sidle, R. C., & Ochiai, H. (2013). *Landslides : processes , prediction , and land use*. American Geophysical Union. <https://doi.org/10.1029/WM018>

Silalahi, F. E. S., Pamela, Arifianti, Y., & Hidayat, F. (2019). Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia. *Geoscience Letters*, 6(1), 10. <https://doi.org/10.1186/s40562-019-0140-4>

- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25.
- Song, J., Wang, Y., Fang, Z., Peng, L., & Hong, H. (2020). Potential of ensemble learning to improve tree-based classifiers for landslide susceptibility mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 4642–4662. <https://doi.org/10.1109/JSTARS.2020.3014143>
- 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. <http://dx.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>
- 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. *Engineering Geology*, 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. In *Inventive Communication and Computational Technologies: Proceedings of ICICCT 2020*, 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>

Thai Pham, B., Tien Bui, D., & Prakash, I. (2018). Landslide susceptibility modelling using different advanced decision trees methods. *Civil Engineering and Environmental Systems*, 35(1–4), 139–157.
<https://doi.org/10.1080/10286608.2019.1568418>

Thi Ngo, P. T., Panahi, M., Khosravi, K., Ghorbanzadeh, O., Kariminejad, N., Cerda, A., & Lee, S. (2021). Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geoscience Frontiers*, 12(2), 505–519.
<https://doi.org/10.1016/j.gsf.2020.06.013>

Tien Bui, D., Pradhan, B., Lofman, O., & Revhaug, I. (2012). Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naïve bayes models. *Mathematical Problems in Engineering*, 2012, 974638.
<https://doi.org/10.1155/2012/974638>

Tien Bui, D., Shahabi, H., Shirzadi, A., Chapi, K., Hoang, N.-D., Pham, B. T., Bui, Q.-T., Tran, C.-T., Panahi, M., Bin Ahmad, B., & Saro, L. (2018). A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides. *Remote Sensing* 10(10), 1538. <https://doi.org/10.3390/rs10101538>

Timeline of landslides in Malaysia in 2022. (2022, December 16). Channel News Asia.

<https://www.channelnewsasia.com/asia/malaysia-landslides-2022-timeline-3150361>

Tofani, V., Ventisette, C., Moretti, S., & Casagli, N. (2014). Integration of Remote Sensing Techniques for Intensity Zonation within a Landslide Area: A Case Study in the Northern Apennines, Italy. *Remote Sensing*, 6, 907–924. <https://doi.org/10.3390/rs6020907>

Trigila, A., Iadanza, C., Esposito, C., & Scarascia-Mugnozza, G. (2015). Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology*, 249, 119–136. <https://doi.org/10.1016/j.geomorph.2015.06.001>

Wu, W., & Sidle, R. (1996). A distributed slope stability model for steep forested basin. *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts*, 33, 178A-178A. [https://doi.org/10.1016/0148-9062\(96\)85201-6](https://doi.org/10.1016/0148-9062(96)85201-6)

Wang, G., Chen, X., & Chen, W. (2020). Spatial Prediction of Landslide Susceptibility based on GIS and Discriminant Functions. *ISPRS International Journal of Geo-Information*, 9(3), 144. <https://doi.org/10.3390/ijgi9030144>

Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, 38(1), 223–230. <https://doi.org/10.1016/j.eswa.2010.06.048>

Wang, G., Lei, X., Chen, W., Shahabi, H., & Shirzadi, A. (2020). Hybrid computational intelligence methods for landslide susceptibility mapping. *SS symmetry Landslide Susceptibility Mapping*. 12(3), 325. <https://doi.org/10.3390/sym12030325>

- Wang, Q., Li, W., Xing, M., Wu, Y., Pei, Y., Yang, D., & Bai, H. (2016). Landslide susceptibility mapping at Gongliu county, China using artificial neural network and weight of evidence models. *Geosciences Journal*, 20(5), 705–718. <https://doi.org/10.1007/s12303-016-0003-3>
- 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, 712240. <https://doi.org/10.3389/feart.2021.712240>
- Wang, Y., Feng, L., Li, S., Ren, F., & Du, Q. (2020). A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. *Catena*, 188, 104425. <https://doi.org/10.1016/j.catena.2019.104425>
- 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>
- Wang, Y., Wang, D., Geng, N., Wang, Y., Yin, Y., & Jin, Y. (2019). Stacking-based ensemble learning of decision trees for interpretable prostate cancer detection. *Applied Soft Computing*, 77, 188–204. <https://doi.org/10.1016/j.asoc.2019.01.015>
- Wang, Z., Chen, Z., Ma, K., & Zhang, Z. (2023). Machine-Learning-Based Hybrid Modeling for Geological Hazard Susceptibility Assessment in Wudou District, Bailong River Basin, China. *GeoHazards*, 4(2), 157–182.

<https://doi.org/10.3390/geohazards4020010>

Wang, Z., Ma, C., Qiu, Y., Xiong, H., & Li, M. (2022). Refined zoning of landslide susceptibility: A case study in Enshi County, Hubei, China. *International Journal of Environmental Research and Public Health*, 19(15), 9412. <https://doi.org/10.3390/ijerph19159412>

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)

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. *Journal of Electronic Science and Technology*, 17(1), 26–40. <https://doi.org/10.11989/JEST.1674-862X.80904120>

Wu, X., Niu, R., Ren, F., & Peng, L. (2013). Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges, China. *Environmental Earth Sciences*, 70(3), 1307–1318. <https://doi.org/10.1007/s12665-013-2217-2>

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>

Xia, D., Tang, H., Sun, S., Tang, C., & Zhang, B. (2022). Landslide susceptibility mapping based on the Germinal Center Optimization algorithm and support vector classification. *Remote Sensing*, 14(11), 2707. <https://doi.org/10.3390/rs14112707>

- 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>
- Xiong, Y., Wang, Q., Yang, J., Zhu, X., & Wei, D.-Q. (2018). PredT4SE-Stack: Prediction of Bacterial Type IV secreted effectors from protein sequences using a stacked ensemble method . *Frontiers in Microbiology*, 9, 422705. <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02571>
- 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>
- Yao, X., Tham, L. G., & Dai, F. C. (2008). Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China. *Geomorphology*, 101(4), 572–582. <https://doi.org/10.1016/j.geomorph.2008.02.011>
- 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>

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, 1168, 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>

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>

Zeng, T., Wu, L., Peduto, D., Glade, T., Hayakawa, Y. S., & Yin, K. (2023). Ensemble learning framework for landslide susceptibility mapping: Different basic classifier and ensemble strategy. *Geoscience Frontiers*, 14(6), 101645. <https://doi.org/10.1016/j.gsf.2023.101645>

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/10.1016/j.cageo.2021.104966>

- Zhang, K., Wu, X., Niu, R., Yang, K., & Zhao, L. (2017). The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. *Environmental Earth Sciences*, 76, 1-20. <https://doi.org/10.1007/s12665-017-6731-5>
- 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), 2801. <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, 589630. <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>
- Zhu, A.-X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T., & Hong, H. (2019). A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. *Catena*, 183, 104188.

<https://doi.org/10.1016/j.catena.2019.104188>

Zhu, A.-X., Miao, Y., Yang, L., Bai, S., Liu, J., & Hong, H. (2018). Comparison of the presence-only method and presence-absence method in landslide susceptibility mapping. *Catena*, 171, 222–233. <https://doi.org/10.1016/j.catena.2018.07.012>

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36. <https://doi.org/10.1109/MGRS.2017.2762307>

Zhuang, J., Peng, J., Iqbal, J., Liu, T., Liu, N., Li, Y., & Ma, P. (2015). Identification of landslide spatial distribution and susceptibility assessment in relation to topography in the Xi'an Region, Shaanxi Province, China. *Frontiers of Earth Science*, 9(3), 449–462. <https://doi.org/10.1007/s11707-014-0474-3>