



**Faculty of Engineering**

**Critical Appraisal of Slope Failure Contributing Parameters for  
Slope Risk Assessment System of Western Sarawak via Multi Statistical  
Approaches with Artificial Neural Network**

**Nur Hisyam bin Ramli**

**Master of Engineering  
2025**

Critical Appraisalment of Slope Failure Contributing Parameters for Slope  
Risk Assessment System of Western Sarawak via Multi Statistical  
Approaches with Artificial Neural Network

Nur Hisyam bin Ramli

A thesis submitted

In fulfillment of the requirements for the degree of Master of Engineering

(Civil Engineering)

Faculty of Engineering  
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: Nur Hisyam bin Ramli

Matric No.: 22020251

Faculty of Engineering

Universiti Malaysia Sarawak

Date: 04 / 03 / 2025

## **ACKNOWLEDGEMENT**

This research has received funding from the Ministry of Higher Education Malaysia, for funding this research through the Fundamental Research Grant Scheme under the grant number FRGS/1/2022/TK06/UNIMAS/01/1, and support from University Malaysia Sarawak, which were pivotal in completing this research. The support in terms of financial, and information provided by my supervisor, Professor Ir Dr Siti Noor Linda bt Taib, and my co supervisor, Professor Gs. Dr Tarmiji bin Masron has been detrimental to the formation of the study. Furthermore, mental support from family and friends has ensured that my psychological well-being was kept stable, ensuring a smooth execution of the project. Besides that, support received from lecturers in the form of a casual conversation, or a research-focused conversation has led to a fruitful publication idea. Finally, I would like to thank everyone that has been around for the duration of the research. Each one of you have both direct and indirect impact on the research. Wishing everyone a happy, healthy, and long-life. Till we meet again.

## **ABSTRACT**

Hazards related to slope failures often causes significant disruptions in a multitude of aspects to the victims. Thus, mitigating its risk is an utmost importance as the consequences are dire. Conventionally, the risks of slope failure are evaluated through a Slope Assessment System for Malaysia. However, the primary limitation of the system is its unsatisfactory level of accuracy. Furthermore, the system has been known to underwent performance degradation when used in areas outside of its data retrieval region such as Western Sarawak that is located at the Western-most tip of Sarawak consisting of Kuching (both landward and seaward), Serian, and Bau. Being home to the state capital city of Kuching, several higher learning institution, and a large-scale industrial park, the region is currently experiencing rapid development. Albeit development is an indicator of a prospering economy, it can increase the risk of exposure to slope failures as settlements being driven out of safe low-lying areas to regions with uncertain levels of slope failure susceptibility. Thus, this study was focused on critically appraising slope failure contributing parameters for Slope Risk Assessment System of Western Sarawak via Multi Statistical approaches with Artificial Neural Network. Unlike the traditional Slope Assessment System, approaching the development via an Artificial Neural Network ensures that the system was driven purely through the training data, making it free from personal biases. An Artificial Neural Network is a Machine Learning approach that is designed to mimic the human brain based on its decision-making process. Two Artificial Neural Networks models were developed in this study for different purposes, where one was used to develop a Landslide Susceptibility Map for the region, and the other was used to develop a new Slope Assessment System specifically for Western Sarawak. The Landslide Susceptibility Map model was developed with the input variables of aspect, curvature, elevation, soil type, lithology type, Land Use and Land Cover, slope

angle, rainfall intensity, and Topographic Wetness Index. The Assessment System model on the other hand was developed with the same variable excluding Land Use and Land Cover, and rainfall intensity. This allows the model to be used without having to wait for the dynamic variable's presence. The evaluation metrics for both models have shown that the development process was a success. The Landslide Susceptibility Model yielded a Root Mean Squared Error of 0.0057 with the hyperparameter of the model being eight neurons in a single hidden layer, a backpropagation learning algorithm, a learning rate of 0.001, and a maximum step of 1E+8. The predictive performance of said model has yielded a recall of 0.9, and a prediction success rate Area Under the Curve score of 0.99. As for the Slope Assessment System model, the same Root Mean Squared Error rate has been achieved with the hyperparameter of four neurons in a single hidden layer, a learning rate of 0.001, a backpropagation learning algorithm, and a maximum steps of 1E+8. The predictive performance of said model yielded a recall of 1.0, and a prediction success rate Area Under the Curve score of 1.0. The Landslide Susceptibility Map that has been developed from its Artificial Neural Network model has shown great correlation with on-site conditions where slope failure points have been identified to occur. However, its primary weakness was spatiotemporal issues arising from the raster files, where it could accurately predict slope failure risks in areas that has undergone major terrain changes for remediation. This issue was solved by implementing the new Slope Assessment System. The new Slope Assessment System was determined to have a recall score of 1, and an Area Under the Curve score of 0.95. Thus, the new system was deemed to have satisfactory levels of accuracy and can be used in Western Sarawak for slope failure susceptibility prediction.

**Keywords:** Data-Driven, machine learning, neural networks, geographical information system

***Penilaian Kritikal Parameter Penyumbang Kegagalan Cerun untuk Sistem Penilaian Risiko Cerun Sarawak Barat melalui Pendekatan Berbilang Statistik dengan Rangkaian Saraf Buatan***

**ABSTRAK**

*Bahaya berkaitan dengan kegagalan cerun sering menyebabkan gangguan yang ketara dalam pelbagai aspek kepada mangsa. Oleh itu, mengurangkan risikonya adalah sangat penting kerana akibatnya adalah serius. Secara konvensional, risiko kegagalan cerun dinilai melalui Sistem Penilaian Cerun untuk Malaysia. Namun, kekurangan utama sistem ini adalah tahap ketepatannya yang tidak memuaskan. Tambahan pula, sistem ini diketahui mengalami penurunan prestasi apabila digunakan di kawasan di luar kawasan pengambilan datanya. Salah satu kawasan yang dikecualikan untuk pengambilan data sistem adalah Sarawak Barat. Sarawak Barat terletak di hujung paling barat Sarawak yang terdiri daripada Kuching (baik di daratan mahupun di lautan), Serian, dan Bau. Menjadi rumah kepada ibu negeri Kuching, beberapa institusi pengajian tinggi, dan sebuah taman perindustrian berskala besar, kawasan ini kini mengalami pembangunan pesat. Walaupun pembangunan adalah petunjuk ekonomi yang makmur, ia boleh meningkatkan risiko pendedahan kepada kegagalan cerun kerana penempatan dipindahkan dari kawasan rendah yang selamat ke kawasan dengan tahap kerentanan kegagalan cerun yang tidak pasti. Oleh itu, kajian ini tertumpu kepada menilai secara kritis parameter-parameter yang menyumbang kepada kegagalan cerun untuk Sistem Penilaian Risiko Cerun Sarawak Barat melalui pendekatan Statistik Pelbagai dengan Rangkaian Neural Tiruan. Tidak seperti Sistem Penilaian Cerun tradisional, pendekatan pembangunan melalui Rangkaian Neural Tiruan memastikan bahawa sistem tersebut didorong sepenuhnya melalui data latihan, menjadikannya bebas daripada bias peribadi. Rangkaian Neural Tiruan adalah pendekatan Pembelajaran Mesin yang direka untuk meniru otak manusia berdasarkan proses membuat*

keputusan. Dua model Rangkaian Neural Tiruan telah dibangun dalam kajian ini untuk tujuan yang berbeza, di mana satu digunakan untuk membangunkan Peta Kecenderungan Tanah Runtuh bagi kawasan tersebut, dan yang satu lagi digunakan untuk membangunkan Sistem Penilaian Cerun yang baru khusus untuk Sarawak Barat. Model Peta Kecenderungan Tanah Runtuh dibangun dengan pembolehubah input seperti aspek, keluk, ketinggian, jenis tanah, jenis litologi, Penggunaan Tanah dan Penutupan Tanah, sudut cerun, intensiti hujan, dan Indeks Kelembapan Topografi. Model Sistem Penilaian sebaliknya dibangun dengan menggunakan pembolehubah yang sama kecuali Penggunaan Tanah dan Penutup Tanah, serta intensiti hujan. Ini membolehkan model digunakan tanpa perlu menunggu kehadiran pembolehubah dinamik. Metrik penilaian untuk kedua-dua model telah menunjukkan bahawa proses pembangunan adalah berjaya. Model Kerentanan Tanah Runtuh menghasilkan Ralat Kuasa Dua Purata (Root Mean Squared Error) sebanyak 0.0057 dengan hiperpemalar model terdiri daripada lapan neuron dalam satu lapisan tersembunyi, algoritma pembelajaran backpropagation, kadar pembelajaran sebanyak 0.001, dan langkah maksimum sebanyak  $1E+8$ . Prestasi ramalan model tersebut telah menghasilkan kadar ingatan sebanyak 0.9, dan kadar kejayaan ramalan dengan skor Area Under the Curve sebanyak 0.99. Bagi model Sistem Penilaian Cerun, kadar Root Mean Squared Error yang sama telah dicapai dengan hiperparameter empat neuron dalam satu lapisan tersembunyi, kadar pembelajaran 0.001, algoritma pembelajaran backpropagation, dan langkah maksimum  $1E+8$ . Prestasi ramalan model tersebut menghasilkan kadar ingatan sebanyak 1.0, dan kadar kejayaan ramalan dengan skor Area Under the Curve sebanyak 1.0. Peta Kecenderungan Tanah Runtuh yang telah dibangun daripada model Rangkaian Neural Tiruan telah menunjukkan korelasi yang baik dengan keadaan di lokasi di mana titik kegagalan cerun telah dikenal pasti berlaku.

*Namun, kelemahan utamanya adalah isu spatiotemporal yang timbul daripada fail raster, di mana ia dapat meramalkan risiko kegagalan cerun dengan tepat di kawasan yang telah mengalami perubahan terrain besar untuk pemulihan. Masalah ini diselesaikan dengan melaksanakan Sistem Penilaian Cerun yang baru. Sistem Penilaian Cerun yang baru didapati mempunyai skor ingatan sebanyak 1, dan skor Kawasan Di Bawah Lengkung sebanyak 0.95. Oleh itu, sistem baharu ini dianggap mempunyai tahap ketepatan yang memuaskan dan boleh digunakan di Sarawak Barat untuk ramalan kecenderungan kegagalan cerun.*

**Kata kunci:** *Berdasarkan data, pembelajaran mesin, rangkaian saraf, sistem maklumat geografi*

## TABLE OF CONTENTS

|                                  | <b>Page</b> |
|----------------------------------|-------------|
| <b>DECLARATION</b>               | i           |
| <b>ACKNOWLEDGEMENT</b>           | ii          |
| <b>ABSTRACT</b>                  | iii         |
| <b><i>ABSTRAK</i></b>            | v           |
| <b>TABLE OF CONTENTS</b>         | viii        |
| <b>LIST OF TABLES</b>            | xiii        |
| <b>LIST OF FIGURES</b>           | xv          |
| <b>LIST OF ABBREVIATIONS</b>     | xvii        |
| <b>LIST OF SYMBOLS AND UNITS</b> | xix         |
| <b>CHAPTER 1 INTRODUCTION</b>    | 1           |
| 1.1 Study Background             | 1           |
| 1.2 Study Area: Western Sarawak  | 4           |
| 1.3 Problem Statement            | 5           |
| 1.4 Aim of Research              | 6           |
| 1.5 Objectives                   | 6           |
| 1.6 Scope of Research            | 7           |
| 1.7 Thesis Structure             | 7           |

|                                                                     |           |
|---------------------------------------------------------------------|-----------|
| <b>CHAPTER 2 LITERATURE REVIEW</b>                                  | <b>9</b>  |
| 2.1 Introduction                                                    | 9         |
| 2.2 Slope Failure                                                   | 9         |
| 2.2.1 Impact of Slope Failure                                       | 10        |
| 2.2.2 Types of Slope Failure                                        | 13        |
| 2.2.3 Factors Leading to Slope Failures                             | 19        |
| 2.3 Slope Assessment Systems                                        | 32        |
| 2.3.1 Slope Assessment Systems in Malaysia                          | 33        |
| 2.3.2 Malaysia's SAS Accuracy                                       | 36        |
| 2.3.3 Slope Assessment System in Southeast Asia                     | 37        |
| 2.4 Geographical Information System                                 | 38        |
| 2.5 Artificial Neural Network                                       | 40        |
| 2.5.1 Innerworkings of an ANN                                       | 42        |
| 2.5.2 Multicollinearity                                             | 49        |
| 2.5.3 Evaluation Metrics                                            | 51        |
| 2.6 Slope Failures Input Data Comparison                            | 52        |
| 2.7 Landslide Susceptibility Mapping using GIS and Machine Learning | 54        |
| 2.8 Machine Learning and Traditional Factor of Safety               | 60        |
| 2.9 Summary of Review                                               | 61        |
| <b>CHAPTER 3 METHODOLOGY</b>                                        | <b>63</b> |

|       |                                                                |    |
|-------|----------------------------------------------------------------|----|
| 3.1   | Overview                                                       | 63 |
| 3.2   | Phase I Data Pooling and Data Pre-Processing                   | 64 |
| 3.2.1 | Data Collection                                                | 65 |
| 3.2.2 | Data Extraction                                                | 77 |
| 3.2.3 | Multicollinearity Assessment                                   | 78 |
| 3.2.4 | Training Data and Testing Data                                 | 78 |
| 3.2.5 | Remarks                                                        | 83 |
| 3.3   | Phase II Development of ANN Models for Western Sarawak         | 83 |
| 3.3.1 | ANN Model Development                                          | 84 |
| 3.3.2 | AUC, Recall, and RMSE                                          | 87 |
| 3.3.3 | Remarks                                                        | 87 |
| 3.4   | Phase III LSM Development and Verification for Western Sarawak | 88 |
| 3.4.1 | LSM Development                                                | 88 |
| 3.4.2 | Verification                                                   | 90 |
| 3.4.3 | Remarks                                                        | 90 |
| 3.5   | Phase IV Development and Verification for Western Sarawak SRAS | 90 |
| 3.5.1 | Data Analysis and Extraction                                   | 91 |
| 3.5.2 | SRAS Development                                               | 92 |
| 3.5.3 | Accuracy Assessment                                            | 93 |
| 3.5.4 | Remarks                                                        | 94 |

|                                                 |                                                                                                                                       |     |
|-------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.6                                             | Validity of Remote Sensing data and Comparison of Performance for Machine Learning models and WoE and FOA in Slope Failure Prediction | 94  |
| <b>CHAPTER 4 RESULTS AND DISCUSSION</b>         |                                                                                                                                       | 95  |
| 4.1                                             | Overview                                                                                                                              | 95  |
| 4.2                                             | Multicollinearity Assessment – Phase I                                                                                                | 95  |
| 4.3                                             | ANN Models Development and Evaluation – Phase                                                                                         | 102 |
| 4.3.1                                           | LSM ANN Model Evaluation                                                                                                              | 102 |
| 4.3.2                                           | SRAS ANN Model Evaluation                                                                                                             | 105 |
| 4.4                                             | LSM of Western Sarawak – Phase III                                                                                                    | 107 |
| 4.5                                             | Development and Verification for Western Sarawak SRAS – Phase IV                                                                      | 114 |
| 4.5.1                                           | Data Analysis for SRAS’s ANN Model Variable Importance                                                                                | 115 |
| 4.5.2                                           | SRAS’s Development                                                                                                                    | 116 |
| 4.5.3                                           | SRAS’s Evaluation                                                                                                                     | 121 |
| 4.6                                             | Validity of Remote Sensing data in Slope Failure Prediction                                                                           | 123 |
| 4.7                                             | Comparison of Slope Failure Prediction Techniques                                                                                     | 124 |
| 4.8                                             | Remarks                                                                                                                               | 125 |
| <b>CHAPTER 5 CONCLUSION AND RECOMMENDATIONS</b> |                                                                                                                                       | 127 |
| 5.1                                             | Introduction                                                                                                                          | 127 |
| 5.2                                             | Contribution of This Research                                                                                                         | 128 |
| 5.3                                             | Achievement of Objective                                                                                                              | 128 |
| 5.4                                             | Recommendation for Future Study                                                                                                       | 131 |

|                   |     |
|-------------------|-----|
| <b>REFERENCES</b> | 132 |
| <b>APPENDICES</b> | 158 |

## LIST OF TABLES

|                                                                                         | <b>Page</b> |
|-----------------------------------------------------------------------------------------|-------------|
| Table 2.1: Notable Slope Failure Cases in Malaysia                                      | 11          |
| Table 2.2: Historical Slope Failures and Forest Area Destroyed                          | 12          |
| Table 2.3: Classification of slope failure                                              | 13          |
| Table 2.4: Debris Flow Cases in Malaysia                                                | 15          |
| Table 2.5: Historical Slope Failure Triggered by Earthquake in Sabah                    | 20          |
| Table 2.6: Slope Failure Factors used by the SAS                                        | 35          |
| Table 2.7: SAS Accuracy in Predicting Slope Failure in Cut-Slopes of Granitic Formation | 36          |
| Table 2.8: SAS Accuracy in Predicting Slope Failure in Cut-Slopes of Meta Sediments     | 37          |
| Table 2.9: Comparison of Slope Failure Input                                            | 53          |
| Table 2.10: Previous Machine Learning LSM Study                                         | 54          |
| Table 3.1: Min – Max of Numerical Variables                                             | 79          |
| Table 3.2: Aspect Data Categorisation                                                   | 80          |
| Table 3.3: Curvature Data Categorisation                                                | 81          |
| Table 3.4: Lithology Categorisation                                                     | 81          |
| Table 3.4 Continued                                                                     | 82          |
| Table 3.5: LULC Categorisation                                                          | 82          |
| Table 3.6: Soil Type Categorisation                                                     | 83          |
| Table 3.7: Formula for Respective ANN Model                                             | 85          |
| Table 4.1: PCA Matrix                                                                   | 96          |
| Table 4.2: Slope Failures based on Aspect Orientation                                   | 98          |
| Table 4.3: Slope Failures based on Curvature                                            | 99          |
| Table 4.4: Slope Failures based on Lithology types                                      | 100         |

|             |                                                                       |     |
|-------------|-----------------------------------------------------------------------|-----|
| Table 4.5:  | Slope Failures based on LULC                                          | 101 |
| Table 4.6:  | Slope Failures based on soil types                                    | 101 |
| Table 4.7:  | LSM's ANN Model RMSE                                                  | 103 |
| Table 4.8:  | LSM ANN Model Performance Matrix                                      | 104 |
| Table 4.9:  | SRAS's ANN Model RMSE                                                 | 106 |
| Table 4.10: | SRAS ANN Model Performance Matrix                                     | 106 |
| Table 4.11: | Mean Slope Failure Susceptibility based on Zones                      | 109 |
| Table 4.12: | Area Coverage Percentage based on Slope Failure Susceptibility Levels | 109 |
| Table 4.13: | Tangga Mawang failure point properties                                | 111 |
| Table 4.14: | Gayu failure point properties                                         | 112 |
| Table 4.15: | Nirvana Cemetery failure point properties                             | 113 |
| Table 4.16: | SRAS's ANN Model Variable Importance                                  | 115 |
| Table 4.17: | Assigned Weightages and Biases in the Hidden Layer                    | 118 |
| Table 4.18: | Example's Input Variable                                              | 119 |
| Table 4.19: | Data Pre-processing Outcome                                           | 119 |
| Table 4.20: | Active Weights and Biases                                             | 120 |
| Table 4.21: | Assigned Weightages and Biases in the Output Layer                    | 121 |
| Table 4.22: | SRAS ANN Performance Matrix                                           | 121 |
| Table 4.23: | Performance Metrics Comparison                                        | 123 |
| Table 4.24: | Models' confusion matrix                                              | 125 |

## LIST OF FIGURES

|                                                                         | <b>Page</b> |
|-------------------------------------------------------------------------|-------------|
| Figure 1.1: Map of Study Area                                           | 5           |
| Figure 2.1: Rockfall Illustration                                       | 14          |
| Figure 2.2: Rotational Slide and Translational Slide Illustration       | 16          |
| Figure 2.3: Topple Illustration                                         | 17          |
| Figure 2.4: Lateral Spreads Illustration                                | 18          |
| Figure 2.5: Terrain Curvature Illustration                              | 28          |
| Figure 2.6: AI and its' Subset                                          | 41          |
| Figure 2.7: ANN's Architecture                                          | 42          |
| Figure 2.8: How an ANN works                                            | 43          |
| Figure 2.9: ANN Model of the Study in Comparison to Conventional Method | 61          |
| Figure 3.1: Flowchart of the Research Methodology                       | 64          |
| Figure 3.2: Flowchart of Phase I                                        | 65          |
| Figure 3.3: Historical Failed Slopes and Non-Failed Slopes              | 67          |
| Figure 3.4: Aspect in Western Sarawak                                   | 68          |
| Figure 3.5: Curvature in Western Sarawak                                | 69          |
| Figure 3.6: Elevation in Western Sarawak                                | 70          |
| Figure 3.7: Soil type in Western Sarawak                                | 71          |
| Figure 3.8: Lithogy in Western Sarawak                                  | 73          |
| Figure 3.9: LULC in Western Sarawak                                     | 74          |
| Figure 3.10: Rainfall Intensity in Western Sarawak                      | 75          |
| Figure 3.11: Slope Angle in Western Sarawak                             | 76          |
| Figure 3.12: TWI in Western Sarawak                                     | 77          |
| Figure 3.13: Flowchart of Phase II                                      | 84          |

|                                                                                             |     |
|---------------------------------------------------------------------------------------------|-----|
| Figure 3.14: Flowchart of Phase III                                                         | 88  |
| Figure 3.15: Flowchart of Phase IV                                                          | 91  |
| Figure 4.1: Scatterplot of Elevation vs Slope Angle                                         | 97  |
| Figure 4.2: AUCs of LSM's ANN Model (a) in Training Phase, and (b) Testing Phase            | 104 |
| Figure 4.3: SRAS's ANN Model AUC (a) Training Success Rate, and (b) Prediction Success Rate | 107 |
| Figure 4.4: LSM of Western Sarawak                                                          | 108 |
| Figure 4.5: Visited Points for LSM Verification                                             | 110 |
| Figure 4.6: Tangga Mawang Slope Failure                                                     | 112 |
| Figure 4.7: Gayu Slope Failure                                                              | 113 |
| Figure 4.8: Nirvana Cemetery conditions                                                     | 114 |
| Figure 4.9: SRAS Application Procedure                                                      | 116 |
| Figure 4.10: Handling Input Variables                                                       | 117 |
| Figure 4.11: SRAS's AUC                                                                     | 122 |

## **LIST OF ABBREVIATIONS**

|       |                                           |
|-------|-------------------------------------------|
| US    | United States of America                  |
| SAS   | Slope Assessment System                   |
| PWD   | Public Works Department                   |
| SRAS  | Slope Risk Assessment System              |
| ANN   | Artificial Neural Network                 |
| LSM   | Landslide Susceptibility Map              |
| GIS   | Geographical Information System           |
| GDP   | Gross Domestic Product                    |
| TWI   | Topographical Wetness Index               |
| LULC  | Land Use and Land Cover                   |
| NDVI  | Normalised Difference Vegetation Index    |
| SMS   | Slope Maintenance System                  |
| SPRS  | Slope Priority Ranking System             |
| SIMS  | Slope Information Management System       |
| SMART | Slope Management and Risk Tracking System |
| ASEAN | Association of Southeast Asian Nations    |
| WoE   | Weight of Evidence                        |
| FOA   | Factor Overlay Analysis                   |
| min   | Minimum                                   |
| max   | Maximum                                   |
| MSE   | Mean Squared Error                        |
| RMSE  | Root Mean Squared Error                   |
| NASA  | National Space Agency                     |

|           |                                                                   |
|-----------|-------------------------------------------------------------------|
| VIF       | Variance Inflation Factor                                         |
| PCA       | Pearson's Correlation Analysis                                    |
| AUC       | Area Under the Curve                                              |
| ROC       | Receiver Operating Characteristics                                |
| TPR       | True Positive Rate                                                |
| FPR       | False Positive Rate                                               |
| RF        | Random Forest                                                     |
| CNN       | Convolutional Neural Network                                      |
| Fuzzy AHP | Fuzzy Analytical Hierarchy Process                                |
| LR        | Linear Regression                                                 |
| SVM       | Support Vector Machine                                            |
| EBF       | Evidential belief model                                           |
| JMG       | Department of Minerals and Geosciences Malaysia                   |
| GDAL      | Geospatial Data Abstraction Library                               |
| QGIS      | Quantum Geographical Information System                           |
| SAGA      | System for Automated Geoscientific Analyses                       |
| FAO       | Food and Agriculture Organisation                                 |
| UNESCO    | United Nations Educational, Scientific, and Cultural Organisation |
| AO        | Orthic Acrisols                                                   |
| BF        | Ferralic Cambisols                                                |
| GD        | Dystric Gleysols                                                  |
| OD        | Dystric Histosols                                                 |
| QA        | Albic Arenosols                                                   |
| ESRI      | Environmental Systems Research Institute Incorporation            |

## LIST OF SYMBOLS AND UNITS

|          |                         |
|----------|-------------------------|
| %        | Percentage              |
| m        | Metres                  |
| mm       | Millimetres             |
| PWD      | Public Works Department |
| km       | Kilometres              |
| °        | Degrees                 |
| rad      | Radians                 |
| pentad   | Five days               |
| <i>e</i> | Exponent                |

## **CHAPTER 1**

### **INTRODUCTION**

#### **1.1 Study Background**

Slope failures or landslides are one of the most common forms of natural disasters as it has been observed to occur everywhere around the world (Moragues et al., 2024). Furthermore, the frequency of slope failure cases has been observed to be on the rise at a global scale, which was attributed to global climate change triggered by the accelerating trend of ice loss and glacier retreat (Zemp et al., 2015). Slope failures are characterised by the outward movement of the slope forming materials, followed by a downward slide towards a gentler slope as it gradually loses its potential energy whilst being pulled by gravity (Rosly et al., 2022). Physically, there are two opposing forces on any given slope: the shear stress that promotes the movement of the slope forming materials, and the shear strength which negates the shear stress, where if the shear stress exceeds the strength, slope failure would occur (Althuwaynee & Pradhan, 2018).

As with all natural disasters, slope failures have been regarded as a significant contributor to a wide range of crises, including environmental and societal concerns, loss of life, property destruction, psychological trauma, disputes over land ownership, and degradation of land (Rahman & Mapjabil, 2017). Additionally, the effects of slope failure are observed globally, including in developed nations such as the United States of America (US), where the country experiences annual losses surpassing 2 billion US dollars due to slope failures (Akter et al., 2019). The impacts of slope failures are more severe in developing nations due to their limited preparedness for natural disasters, and a potentially

reduced effectiveness of mitigation efforts as a result of budget constraints (Vakhshoori et al., 2019).

It is of utmost importance to ascertain the susceptibility of slope failure in a particular area to mitigate the potential damages caused by such occurrences, despite the inherent complexity of this task. Nevertheless, it has been established that there are commonalities among different instances of slope failure, as a recurring pattern of causal factors for slope failure has been identified (Tengtrairat et al., 2021). Therefore, it is crucial to consider historical data on slope failures in order to identify areas that are prone to such occurrences (Zêzere et al., 2017). Through the analysis of historical data on slope failures, researchers can explore the various factors contributing to the extent of these incidents. Therefore, it is possible to create a Slope Assessment System (SAS) based on the provided data. However, an expert view-based SAS is not universally applicable for determining the susceptibility of slope failure, as its precision relies heavily on the properties of the specific area being studied matching those in the data source area for slope failures. The concept articulated by Tobler's first law of geography asserts that all things are interconnected, but objects in close proximity have a stronger relationship than those that are farther apart (Fan et al., 2023).

Malaysia, located in Southeast Asia, is considered a developing nation. The country is comprised of two primary regions, namely Peninsular Malaysia and parts of Borneo, which are Sabah and Sarawak. Situated along the equator, the nation encounters a tropical climate characterised by two annual rainy seasons, namely the Southwest Monsoon and the Northeast Monsoon (Azis, 2018). Despite the prevalence of extensive flat terrain in the country, the occurrence of slope failures is still evident. Slope failures contribute to 8.3% of the country's total annual natural disasters, with approximately one hundred cases occurring

each year, ranging in severity (Akter et al., 2019). Moreover, it has been established that all documented slope failures occur during periods of heavy rainfall, with precipitation identified as the principal causative factor of slope failure in the nation (Rosly et al., 2022). The Public Works Department (PWD) in Malaysia has developed multiple SAS using data collected from Peninsular Malaysia and Sabah for the purpose of assessing the likelihood of slope failure. The accuracy of the SAS was found to vary, with results showing inconsistency across different regions (H. Singh et al., 2008).

As of the year 2019, there have been 21,000 identified slope failure prone areas in Malaysia, where 2,000 were in the state of Sarawak (Majid et al., 2020). It is projected that Sarawak will experience an increase in slope failure occurrences due to the state receiving the highest amount of rainfall, with expectations of this increasing further because of climate change. Additionally, Sarawak has the highest surface temperature, contributing to the rise in annual rainfall from 3000 millimetres (mm) to 4000 millimetres over the past thirty years (Isia et al., 2023). Furthermore, the state of Sarawak is experiencing significant growth and advancement, particularly in the region of Western Sarawak, which is also the location of the state's capital city. The correlation between economic growth and urban development is frequently linked to instances of slope failure, particularly in Malaysia. The country has identified construction errors in susceptible areas as a leading factor in the occurrence of slope failures (Rahman & Mapjabil, 2017). Consequently, Western Sarawak was chosen as the area of interest in this study.

This study intends to conduct a critical appraisal of slope failure contributing parameters for Slope Risk Assessment System (SRAS) of Western Sarawak via multi statistical approaches with Artificial Neural Network (ANN). Where the SRAS is intended

to be used as a substitute for the SAS in Western Sarawak. Through the advancement in computer science, and Machine Learning, several researchers have applied Machine Learning algorithm to predict the slope failure susceptibility of their respective area of interest with high success rate (Khajehzadeh et al., 2021; Pham et al., 2019; Yakub et al., 2022). Furthermore, the implementation of remote sensing data in their research has allowed them to plot the Landslide Susceptibility Map (LSM) for their area of interest. An LSM is a graphical representation for the slope failure susceptibility. Alongside ANN, Geographical Information System (GIS) was heavily used to manage spatial data in this study. Although at the state level the study area is known as the Southern region, at the national level, the study area is known as the Western region, as it located on the Western most region of East Malaysia (Breitfeld, 2021; Mohd-Azlan et al., 2022).

## **1.2 Study Area: Western Sarawak**

Western Sarawak is in the Western-most tip of Borneo, with three main administrative divisions, which are Kuching, Samarahan, and Serian. However, the study area of this research was confined to the districts of Bau, Kuching, and the division of Serian as seen in Figure 1.1 (Sarawak Government, n.d.). The exclusion of several areas was implemented as there were little information on historical slope failures cases were obtained from the areas. The study area was determined to have an area of approximately 4778.58 km<sup>2</sup>, and as it is located on equatorial line, tropical climate is present. The weather conditions in study area are at its driest during the late Southwest monsoon from May to August, whilst the wettest during the Northeast Monsoon from December to February (Dambul & Jones, 2007). The state of Sarawak has experienced a Gross Domestic Product (GDP) growth of 4.7% in 2019, with a nominal GDP of 114 billion in Ringgit Malaysia. Thus, Western Sarawak, where the capital city of the state is in, experienced development and urbanisation

at a steady rate (Liaw, 2020). However, progress does not come without a trade-off, the depletion of flat areas that are known to be slope failure safe areas will push the development to locations with steep slopes increasing the exposure to slope failure risk (Alamgir et al., 2020).

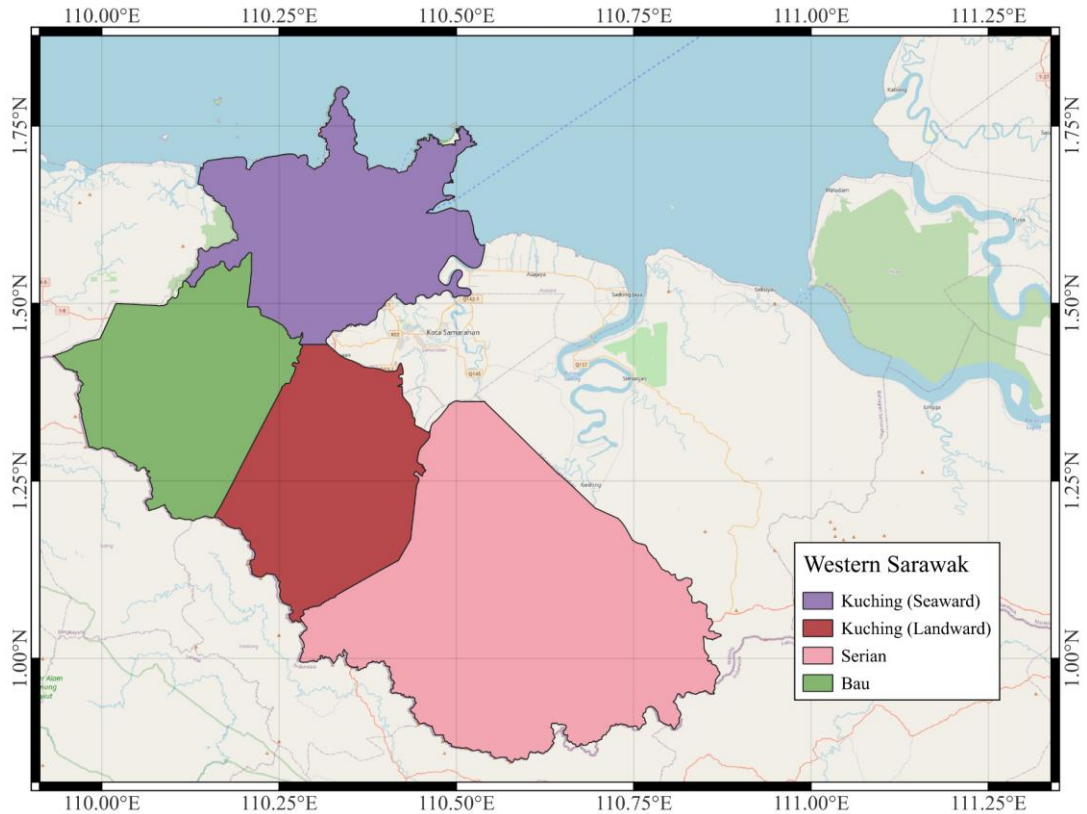


**Figure 1.1:** Map of Study Area

### 1.3 Problem Statement

Numerous previous studies have successfully deployed Machine Learning models to predict the slope failure susceptibility of their respective area of interest such as Data-Driven Landslide Spatial Prediction and Deformation Monitoring: A Case Study of Shiyan City, China, Landslide Susceptibility Mapping using support vector machine for Meghalaya, India, and Landslide Susceptibility Model Using Artificial Neural Network (ANN)

Approach in Langat River Basin, Selangor, Malaysia (Ado & Amitab, 2023; Selamat et al., 2022b; Sheng et al., 2023). However, the temporal issues that persists when using spatial data has not been addressed as many of the previous study solely focused on the deployment of the Machine Learning model as the predictor. Furthermore, in the study area, there has not been a research conducted to deploy Machine Learning models for slope failure predictions. Besides that, the existing techniques known as Slope Assessment Systems in assessing slope failure susceptibility in Malaysia has been deemed to be insufficient as it lacks the minimum accuracy and has been determined to be inconsistent when used outside of its data-retrieval region such as Western Sarawak (Huat & Jamaludin, 2005). Thus, this study was focused on the critical appraisal of slope failure contributing parameters for slope risk assessment system of Western Sarawak via multi statistical approaches with ANN. The outcome in the form of a new Slope Assessment System known as SRAS, which solves the spatiotemporal and model availability dependence of the previous study, and the issues with the existing techniques of slope failure assessment, which will be used in conjunction with the LSM. It is the pioneer for slope failure prediction with Machine Learning models in Western Sarawak.

#### **1.4 Aim of Research**

This research aims to conduct a critical appraisal of slope failure contributing parameters for Western Sarawak via multi statistical approaches with ANN to develop a new SRAS, and an LSM for the region.

#### **1.5 Objectives**

This study was conducted to fulfil the following objectives.

- i. To analyse and critically appraise the contributing factors of slope failures such as geological, hydrological, meteorological, and physical factors leading to slope failures in Western Sarawak – Phase I.
- ii. To assess the suitability of ANN in terms of its accuracy in the development of a framework analysis for SRAS, and an LSM – Phase II
- iii. To develop a an LSM and a SRAS for Western Sarawak from the ANN model - Phase III, and Phase IV

## **1.6 Scope of Research**

This research is limited to the development of ANN models for the formulation of a SRAS, and plotting the LSM, with data obtained from areas which have historical slope failures data in Western Sarawak. The input variables for the ANN model were limited to accessible data.

## **1.7 Thesis Structure**

This thesis consists of five chapters with their own sub-chapters. The first chapter is introduction. This chapter provides a summary of the thesis, including basic information on the Machine Learning method used, the cruciality of this study, and information on the area of study.

Followed by the second chapter, which is literature review. This chapter shows previous study on the topic or research that are involved in the topic. Notably, the method of Machine Learning, type of data lab required, landslides parameters, and previous SAS systems in used for landslide susceptibility investigation of Malaysia.

The third chapter is methodology, where it shows the processes that were involved to collect the spatial data, development of the ANN model, and the conversion of the ANN model into a new SRAS for Western Sarawak.

The next chapter is results and discussion, where the section is focused on the analysis and interpretation of the output obtained from the trained and tested ANN model, as well as the SRAS. Justification for the results that were obtained is presented and discussed in detail here.

The final chapter is the conclusion and recommendations that summarised the project. This section presents the project's strengths and weaknesses, as well as recommendations for improvements in the future.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

This chapter examines and reviews the essential aspect of slope failure susceptibility assessment. Prior works associated with the topics were summarised as it provides the foundation of this research. This chapter is partitioned into four sections, where the first section focuses on slope failure, the second section deals with SAS, the third section discusses on GIS, and the fourth section examines ANN.

#### **2.2 Slope Failure**

Slope failure or generally known as landslide is a form of natural disaster that occur in a plethora of materials and possibly exhibits various types of movement at different rates, that changes the landscape as it moves the slope forming material (Ghasemian et al., 2022). By definition, slope failure is the downward and outward movement of a slope forming material which occur when the external forces exerted upon the slope is larger than the internal forces that is responsible for the slope's stability (Yilmaz et al., 2022). Slope failures mostly take place in mountainous regions with steep slopes, where it poses a serious threat to both humans and the surrounding, thus, reliable mitigation efforts are necessary to avoid or reduce the potential geological hazard (Hakim et al., 2022).

Slope failure occurrences are inherently complex as it does not occur solely through a single variable nor does it exhibits the same type of movement consistently (Vidal-Paez et al., 2020). A plethora of variables influences its occurrences ranging from geological factors to human factors, at a different extent respectively (Li et al., 2021). The complexity and

intricate relationship amongst the variables to induce slope failure has enticed numerous researchers to develop a system for assessing the slope failure susceptibility (Fan et al., 2023). Moreover, the influence of each variable differs from one location to another, limiting the usage of a developed system, however the basic framework can be transferred to a new area (Peng et al., 2022).

### **2.2.1 Impact of Slope Failure**

Slope failures involve the movement of slope forming materials making it very deadly, where annually it has been determined to cause death in the thousands globally. Whenever it occur at a large scale, the impact tends to be long lasting (Xiao et al., 2022). The devastating aftermath of a slope failure greatly effects anyone and anything that were present in its wake (Utomo et al., 2020).

#### **2.2.1.1 Socioeconomic Impact**

Slope failure has a high possibility to block or destroy access of communities to and from services for goods, work, and healthcare on every occasion it occurs nearby or on a road, as the debris tends to be deposited on the road whenever it is situated on the gentler slope (Winter et al., 2016). Repairing the damages caused by slope failures is known as the direct impact on the economy, on the other hand, indirect socioeconomical impact of slope failures are as follows (Schuster & Turner, 1996):

- i. Depreciation in values for real estates in slope failure prone areas.
- ii. Lower amounts of tax revenues gathered from depreciated property caused by slope failure.
- iii. Decreasing productivity of human resources as well as animals because of injury, death, or psychological trauma.

Since 1973, slope failures have costed Malaysia more than a billion US dollar from loss of life as well as private and public properties damages, some of the most costly slope failure cases in the country can be seen in Table 2.1 (Diana et al., 2021).

**Table 2.1:** Notable Slope Failure Cases in Malaysia

| Case                                | Year | Damages                                                        |
|-------------------------------------|------|----------------------------------------------------------------|
| Highland Tower, Selangor            | 1993 | 48 deaths, 2 injured, and one high rise building collapsed     |
| Kuala Lumpur to Karak Highway       | 1995 | 20 deaths, 22 wounded, and ten cars damaged                    |
| Pos Dipang, Perak                   | 1996 | 44 deaths                                                      |
| Keningau, Sabah                     | 1996 | 300 deaths through mudslide and floods during the Greg's Storm |
| North to South Expressway, Selangor | 1999 | Road was inaccessible for a day leaving thousands stranded.    |
| Jalan Seri Pechala 1, Kuala Lumpur  | 2004 | A total of 24 houses emptied                                   |

#### 2.2.1.2 Topographical Impact

Slope failure significantly alters the landscape in areas that it has taken place, as it displaces failed slope forming material from one location to another, especially major ones with large amount of mass movement (Sassa & Canuti, 2008). As it modifies the terrain, it also causes the destruction of forest in the area, which is a very common phenomenon around the world. However, the effect has been determined to be more severe in tropical regions due to heavy rainfall and earthquakes where some can be observed in Table 2.2 (Geertsema et al., 2009).

**Table 2.2:** Historical Slope Failures and Forest Area Destroyed

| Location | Year | Forest Area Destroyed (km <sup>2</sup> ) |
|----------|------|------------------------------------------|
| Chile    | 1960 | 250                                      |
| Panama   | 1976 | 54                                       |
| Colombia | 1994 | 250                                      |
| Ecuador  | 1987 | 54                                       |

#### 2.2.1.3 Hydrological Impact

Slope failures greatly degrades water quality in streams and irrigation facilities in its vicinity as debris from the failure fell into the water bodies (Schuster & Turner, 1996). Furthermore, slope failure does not only decreases the quality of water on the surface with excessive amount of sediments that are essential for the longevity of the fish and wildlife habitat, it also changes the behaviour of the groundwater flow (Geertsema et al., 2009). Moreover, the presence of excessive sediments in the water greatly decreases water quality in other areas that are connected to the affected areas, by reducing algae production which occur as the sediments blocks transmission of light rays, damaging the gills of fish that finally leads to its death, and high amount of deceased fish in a water body will lead to degradation of water quality (Schuster & Highland, 2017).

#### 2.2.1.4 Remarks

The aftermath of slope failures leads to numerous crisis such as socioeconomic issues, landscape alteration, and degradation of water qualities. However, these are only several examples of the negative impacts of slope failures, as slope failure could trigger a chain reaction of crisis in its' wake.

## 2.2.2 Types of Slope Failure

Slope failure is a term used to describe a wide array of phenomenon which involves mass wasting be it rock, soil, artificial fill or a combination of said materials on a slope where the mass moves outward and downward towards a gentler slope (Highland, 2004). As there are multiple movements exhibited by slope failures in different types of materials, Varnes in 1996 has taken the initiative to categorise slope failures based on said criteria where the type of movements are falls, topples, rotational or translational slides, lateral spreads, flows, and complex, and the type of materials are bedrock, and engineering soils: predominantly fine or predominantly coarse (Varnes & Cruden, 1996). The classification of slope failure that was developed Varnes are still used till today where it has gone through several revisions to provide a better understanding on each type of slope failure where the latest classification can be seen in Table 2.3 (Rahman & Mapjabil, 2017).

**Table 2.3:** Classification of slope failure

| Type of Movement | Type of Material                            |                      |                    |
|------------------|---------------------------------------------|----------------------|--------------------|
|                  | Bedrock                                     | Engineering Soils    |                    |
|                  |                                             | Predominantly Coarse | Predominantly Fine |
| Falls            | Rock fall                                   | Debris fall          | Earth fall         |
| Topples          | Rock topples                                | Debris topple        | Earth topples      |
| Slides           | Rockslide                                   | Debris slide         | Earth slide        |
| Lateral Spreads  | Rock spread                                 | Debris spread        | Earth spread       |
| Flows            | Rock flow                                   | Debris flow          | Earth flow         |
| Complex          | Combination of two or more type of movement |                      |                    |

### 2.2.2.1 Fall

A fall slope failure as seen in Figure 2.1, happens when the slope forming materials fail and moves mostly through free fall, leaping, bounding, or rolling, with minimal or no

interaction amongst one moving mass to the other, as there is a lack of resistance from anything, fall slope failure moves very rapidly until all of the potential energies are transferred (Varnes & Cruden, 1996). Fall slope failure can be in the form of rockfall when it occurs on bedrock, debris fall when it occurs on engineering soils that are predominantly coarse, or earth fall when the engineering soils are predominantly fine (Rahman & Mapjabil, 2017). Fall especially rockfalls are commonly observed to occur in mountainous areas with snow or ones that receives very large amounts of rainfall that infiltrate the fracture in bedrocks which in turns causes the materials to separate and moves independently (Ravanel et al., 2023). Malaysia has also experienced rockfalls in the past such as the documented rockfall in Kinta Valley, where the area is known to be very susceptible to rockfall as early as the case in 1919, where it has caused a total of 12 deaths, and the most devastating case as of 2022, was the October 1973 incident, that claimed 40 lives as a rock debris from rockfall demolished a longhouse in the area (Ghani et al., 2022).



**Figure 2.1:** Rockfall Illustration

#### 2.2.2.2 Flows

Flows are slope failures that behave similarly to a viscous fluid, the moving mass consists of many small units. However, flows does not only occur in the presence of water,

as dry flows has also been observed to happen (Varnes & Cruden, 1996). There are five categories of flows differs from one to the other in major aspects, which are debris flow that is triggered by intense surface water flow that moves the slope forming material with less than 50% fines, debris avalanche is a debris flow that moves more rapidly, earthflow that occurs in fine grained material under saturated condition where it liquefies and forms an hourglass shapes as it move. Mudflow happens when the slope forming materials is made of more than 50% fine particles and is wet enough to flow rapidly. Meanwhile, creep which is a slow and steady downward movement of the slope forming materials be it soil or rock caused by a large enough shear stress to cause permanent deformation, albeit not enough to cause shear failure (Highland, 2004). Malaysia being a tropical country that receives large amounts of precipitation especially in the wet seasons, has experienced several flow slope failures, particularly debris flow as seen in Table 2.4, recorded as early as 1995 with varying degrees of severity.

**Table 2.4:** Debris Flow Cases in Malaysia

| No. | Location                                                        | Year | Fatalities |
|-----|-----------------------------------------------------------------|------|------------|
| 1.  | KM 38.6 Kuala Lumpur to Karak Highway, Genting Sempah, Selangor | 1995 | 20         |
| 2.  | Keningau, Sabah                                                 | 1996 | 300        |
| 3.  | Ruan Changkul, Simunjan, Sarawak                                | 2002 | 16         |
| 4.  | Gunung Jerai, Yan, Kedah                                        | 2021 | 5          |
| 5.  | Gunung Inas, Sungai Kupang, Baling, Kedah                       | 2022 | 3          |

#### 2.2.2.3 Slides

The term slide is usually associated with the general term used to describe mass movement on a slope which is ‘landslide’, however, the specific use case of the term is only restricted to mass movement be it rotational or translational, that occurs due to a distinct

zone of weakness on a slope that separates the sliding material from the stable material (Varnes & Cruden, 1996). On a rotational slide, the surface of rupture curved concavely upwards as the sliding materials rotates on an axis, while translational slides moves roughly linearly along a planar surface with rotation at a minimum or none (Highland, 2004). Malaysia has also experienced slide slope failures in the past such as the shallow rotational slide case in January 1999 that was triggered by heavy rain which buried several houses or huts in a squatter settlement killing 13 people, and in January of 2001, where another shallow rotational slide occurred in Simunjan, Sarawak on a vegetable farm that buried several houses at the bottom of the slope, killing 16 people (Rahman & Mapjabil, 2017). Rotational and translational slides are illustrated in Figure 2.2.

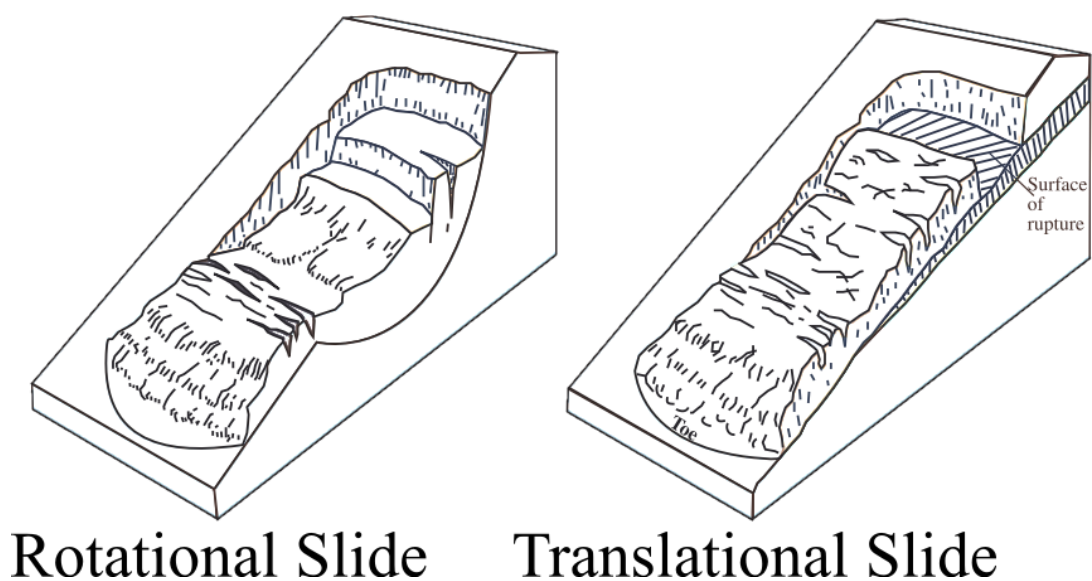


**Figure 2.2:** Rotational Slide and Translational Slide Illustration

#### 2.2.2.4 Topple

Topple occurs when the slope forming material fails due to gravity, forces exerted by adjacent units, or by the presence of fluids in cracks, and moves via rotational about a pivotal point in a single unit or multiple units as seen in Figure 2.3 (Highland, 2004). Topple

may occur on bedrock materials resulting in a rock topple, predominantly coarse soils resulting in a debris topple, or predominantly fine soils resulting in an earth topple (Rahman & Mapjabil, 2017).

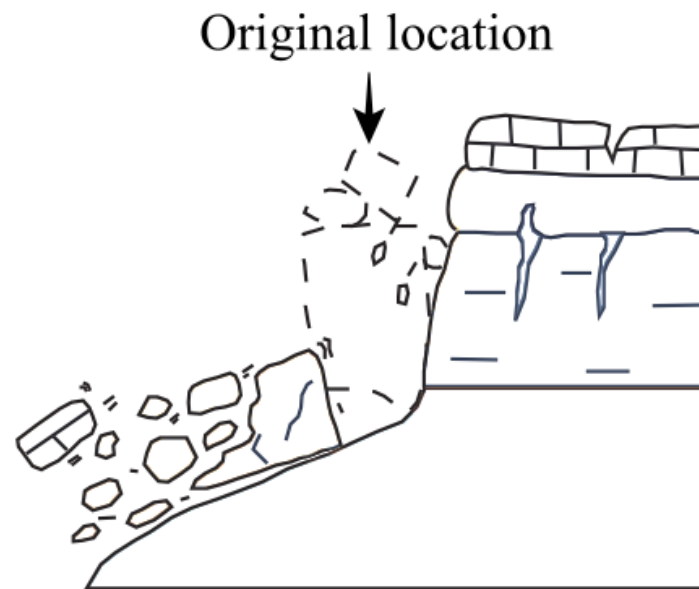


**Figure 2.3:** Topple Illustration

#### 2.2.2.5 Lateral Spread

Unlike any other type of slope failures, lateral spread occurs on very gentle slope or flat terrain as seen in Figure 2.4 where the main type of movement that causes it is a lateral extension with shear or tensile fractures triggered by rapid ground motion be it due to earthquake or artificially induced (Varnes & Cruden, 1996). Lateral spread may occur in bedrock material which forms rock spread, predominantly coarse soils which forms debris spread, or predominantly fine soils which form earth spread, and the failure has been observed to start suddenly in a small area which then spreads in a very rapid manner (Highland, 2004). Thus far, there is a lack of reported lateral spreads cases in Malaysia,

potentially due to the main triggering factor which is seismic activity being relatively low in the country (Rosly et al., 2022).

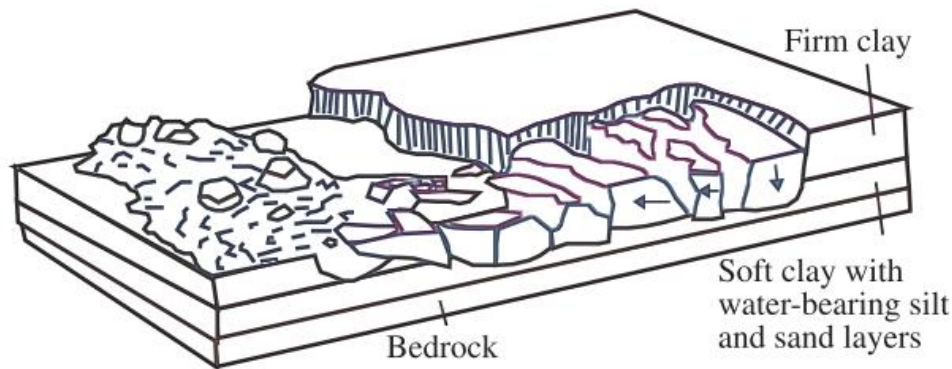


**Figure 2.4:** Lateral Spreads Illustration

#### 2.2.2.6 Complex

Slope failures may occur through a combination of two or more types of slope failure and this is known as a complex slope failure (Highland, 2004). There is a potential where variation in the type of movement exhibited by the slope failure and the materials that it has occurred in from time to time, and from one location to another, thus a rigid categorisation of that particular slope failure is not suitable (Varnes & Cruden, 1996).

#### 2.2.2.7 Remarks

There are several types of slope failures categorised by the type of movement and the type of slope-forming materials. However, it has been determined that the mechanics of each failure type is the same, as the slope-forming materials loses its strength due to the underlying conditioning variables and fails when a sufficient triggering variable is present.

### **2.2.3 Factors Leading to Slope Failures**

There are many types of slope failures, however the causes for all slope failures tend to be same albeit differs in the amount of influence, as the main component behind the mechanism is the loss of shear strength (Das et al., 2011). There are many factors that causes slope failures to occur such as geological factors, morphological factors, and human factors (Highland, 2004). Furthermore, the factors can be further categorised based on how it influences slope failure to occur: be it a triggering factors that causes slope failure to occur directly upon its presence, or conditioning factors that reduces the shear strength of slope forming material leaving it more susceptible to failure whenever the triggering factors are present (Althuwaynee & Pradhan, 2018). The factors can further be categorised based on how the factor changes from time to time, where dynamic factor such as rainfall, and human factors changes drastically according to time, whilst static factor tends to stay the same or changes slowly through time such as elevation and slope angle (Tengtrairat et al., 2021).

#### **2.2.3.1 Triggering Factors**

The following are examples of triggering factors:

##### **i. Earthquakes**

Earthquakes refers to shaking or trembling of the earth's surface that transmits seismic energy to the surrounding area, and it is known to be one of the most destructive forms of natural disaster as it causes a domino effect which triggers other natural disaster such as tsunamis and slope failures (S. Chen et al., 2020). Earthquakes trigger slope failures by applying large external forces that shake the slope forming material leading to slope failures, and the impact of earthquakes on slope failures occurrences have been determined to be long-lasting as the seismic waves loosen up any slope forming materials making it very

unstable and susceptible to slope failures such as debris flow in the presence of rain (Zhao et al., 2022). Furthermore, earthquake triggered slope failure has been determined to be the most destructive form of slope failures, and the losses caused by it are larger than losses from the triggering earthquake, thus hindering the development of society and the economy of the affected area (Fan et al., 2023). In slope failure susceptibility studies, earthquake is often time represented through peak ground acceleration, a well-known engineering parameter that is commonly used to seismic structural analysis and risk assessment (Gandomi et al., 2016). Malaysia albeit located outside active seismic zone, its proximity to active earthquake sources such as the Manila Trench has exposed the country to the disaster (Aziz et al., 2023). Amongst all states in Malaysia, Sabah has experienced the most slope failure cases triggered by earthquake as seen in Table 2.5, which originated from the nearby region of Southern Philippines, straits of Macassar, Sulu Sea, and Celebes Sea (Rosly et al., 2022).

**Table 2.5:** Historical Slope Failure Triggered by Earthquake in Sabah

| No. | Location                            | Year | Case                                                           |
|-----|-------------------------------------|------|----------------------------------------------------------------|
| 1.  | Ranau and Kundasang                 | 2015 | 5.9 Richter scale earthquake triggered two mudslide incidents. |
| 2.  | Ranau, Kundasang, and Kota Kinabalu | 2018 | 5.2 Richter scale earthquake triggered rockfall.               |

The influence of earthquake on slope failures in Western Sarawak was confirmed to be at a minimum where the nearest fault from the region were the Kelawit fault, and the Bukit Mersing fault. Both fault lines are located 500 km away from Western Sarawak, with negligible ground motions originating from the faults (Halim Ida et al., 2022).

## ii. Rainfall

Rainfall triggers slope failures by infiltrating the slope forming materials causing the water table to rise and the processes causes the slope forming material to be saturated resulting in a weaker slope which then slides if it has become completely unstable (Najar et al., 2023). In comparison to earthquake, rainfall has triggered more slope failures around the world albeit it only triggers slope failure when the minimum critical rainfall threshold has been achieved, which differs based on a plethora of properties present in the specific area (Ma et al., 2020). Rainfall is a form of dynamic slope failure triggering factor as it does not occur all the time, and greatly changes the hydrological condition of an area when it happens (Huat et al., 2006). Furthermore, most shallow slides slope failures are triggered by heavy rainfall, and it is hypothesized that rainfall causes a rapid increase in porewater cases due to the varying saturated hydraulic conductivity, which in turn causes the slide to occur (Gomes & Vieira, 2016). Rainfall is commonly represented through its' intensity, which measures the amount of rainfall received by an area per unit time, and the data can be obtained through remote sensing, rainfall station data, or an incorporation of both (Sa'adi et al., 2019). In Malaysia, rainfall has been identified to be the main slope failure triggering factor, as all of the recorded cases took place during the rainy season (Rosly et al., 2022). Moreover, debris flow slope failure has been determined to be increasing in frequency in Malaysia as intense rainfall causes water surge in the upstream areas (Jaapar et al., 2023). The impact of rainfall combined with intense temperatures in the tropics can cause severe weathering that could reach depths in the hundreds of meters (Huat & Jamaludin, 2005). In Western Sarawak, cumulative daily rainfall impose additional load to the slope, and when a sudden increase in the rainfall intensity occur, the slope would inherently fail (Tay & Selaman, 2011).

#### 2.2.3.2 Conditioning Factors

In the presence of a triggering factor, given that the forces exerted by the trigger is sufficient to destabilise the slope forming materials as the conditioning factors that are present on the slope are susceptible, a slope failure would occur (Fan et al., 2023). The following are examples of conditioning factors:

##### i. Lithology

Lithology is a geological factor in slope failure susceptibility studies and it is known to be the parent material of soil, where different lithology refers to the gross physical characteristics of the parent materials which includes the mineral composition, colour, and grain size (Gray et al., 2016). Lithology has a huge impact on slope stability as each lithology possesses a different degree of susceptibility towards failure where ‘softer’ lithology are often times more prone to failure than ‘harder’ ones, for example clay units are more susceptible towards failure in comparison to calcareous units (Anis et al., 2019). Furthermore, information of lithology provides details on the degree of strength and permeability of soils and rocks in the area (Nhu et al., 2020). In conclusion, the types of lithology exhibited variations in the physical and mechanical properties of the units, such as strength, degree of weathering, durability, density, and permeability (Youssef & Pourghasemi, 2021). In Western Sarawak, exists eight distinct types of lithologies, coming from different period, with varying levels of susceptibility towards slope failure. The lithology types can be seen below, arranged based on slope failure susceptibility, starting from the most prone lithology.

a. Clay, Silt, and Peat (Pleistocene and Recent)

Clay is very prone to slope failure especially in the presence of intense rainfall, due to their low shear strength and high expansive capacities (Yalcin, 2007). Silt is even more susceptible to slope failure as it is easily eroded (Fonseca et al., 2017). Peat on the other hand, only fails in areas nearby water courses, and peat slope failure is mostly concentrated in Scotland (Mills & Rushton, 2024).

b. Arenaceous and Argillaceous Rocks with Coal and Calcareous beds (Neogene)

These lithologies have varying levels of susceptibility. However, the beds consisting of coal and calcareous rocks are weak, and very prone to weathering, which increases the likelihood of slope failure in the lithologies (Miller & Degg, 2012).

c. Argillaceous Rocks with Some Arenaceous, and Calcareous beds (Palaeogene)

Argillaceous rocks are considerably susceptible to failure, with high clay contents and rich in organic matter (Faizin & Nur, 2018). In combination with the Arenaceous, and Calcareous beds, this type of lithology also has beds that are prone to weathering, and highly permeable to rainfall, which greatly increases the likelihood of slope failure (Mezughi et al., 2011).

d. Argillaceous Rocks with Some Arenaceous, and Calcareous beds (Jurassic to Cretaceous)

This lithology type is identical to the previous however coming from an older period. Thus, potentially due to the older and potentially more compacted materials, it is possible for the lithology to be less susceptible to slope failure.

- e. Sandstone and associated Lava, Pyroclastics, and Argillaceous Rocks (Triassic)

In comparison to pure Argillaceous rocks, Sandstone can be more stable, and with the presence of lava and Pyroclastics, the strength of the lithology should be improved (Trisnawati et al., 2022).

- f. Phyllite, Metagraywacke, and Limestone, with some Schists and Chert (Upper Palaeozoic)

This type of lithology is considerably stronger than the previous ones as it is made of metamorphic rocks albeit limestone being susceptible to Karst process (Le et al., 2021).

- g. Andesitic, Basaltic Lava, and Pyroclastics (Triassic to Quaternary)

Volcanic rocks that make up this lithology type is generally safe from slope failure as they are strong, and more resistant to weathering. However, they can be susceptible to slope failure if fracture is present (Misbahudin, 2020).

- h. Granodiorite, Diorite, Granite, and their fine-grained equivalents (Triassic to Late Tertiary)

Amongst all lithology types that are present in Western Sarawak, this lithology is the least susceptible towards slope failure, due to their high strength, and resistance to weathering. The only instances where slope failure occur in this lithology is when large live loads are constantly applied, and even in this case, the slope failure is minor (Chigira et al., 2011).

- ii. Surface Runoff

Water disposition from rainfall on the terrain significantly impacts the slope failure susceptibility as the water will infiltrate the surface layer increasing the water content thus

weakening the slope (Selamat et al., 2022a). As water accumulates, the slope forming materials become saturated, thus increasing porewater pressure whilst reducing the shear strength (Peng et al., 2022). Topographic Wetness Index (TWI) equation as seen in Equation 2.1, is commonly used to determine the degree of the surface runoff as it is a physical indicator of flood inundation zones and provides information on the severity of runoff at each location in the area of interest (Rudra & Sarkar, 2023). Moreover, in slope failure susceptibility studies, the deployment of TWI provides details on the geomorphological parameters to describe the local relief (Selamat et al., 2023).

$$TWI = \ln (\alpha / \tan \beta) \quad \text{Equation 2.1}$$

Where  $\alpha$  is the cumulative upslope area drainage through a point (per unit contour length), and  $\beta$  is the angle of the slope at the point.

### iii. Elevation

Elevation represents altitudes from a datum of an area and it has been regarded as one of the most crucial slope failure conditioning factors, as most slope failures occurred in areas with high elevation (Anis et al., 2019). Moreover, the higher altitude increases the ruggedness of the terrain, and the varying elevation corresponds to different rainfall patterns and vegetation types (Sur et al., 2022). Elevation also affects the degree of weathering as the temperature inversely correlates with elevation, where lower elevation receives a higher temperature, and vice versa (Selamat et al., 2022b). Moreover, elevation controls the deposition of soils, and soil thickness where at a lower elevation, unconsolidated materials are deposited, while more competent materials are commonly located in the higher elevation (Mekonnen et al., 2022).

#### iv. Aspect

Aspect refers to direction of a given slope face and is identified through the compass orientation where the aspect values starts from 0° North circling clockwise 360° back to the start, which has been determined to have significant influence on the properties of the slope forming materials (Mekonnen et al., 2022). The slope aspect determines the amount of sunlight exposure, the effects of rainfall, moisture, and wind conditions in the study area, which translates to degree of vegetation coverage, surface weathering rate, and surface evaporation that influences the occurrence of slope failure (Li et al., 2021). Furthermore, several research works have concluded some aspect orientation of slope is more likely to have slope failure cases in comparison to other aspect, depending on the study area (Anis et al., 2019).

#### v. Slope Angle

Slope angle is the angle formed between the run and the climb of an angular deviation from an idealised flat surface as seen in Equation 2.2. Slope angle is highly regarded as the most influential slope failure contributing factor where the steeper the slope, the higher the likelihood of slope failure, as it has been determined that the driving force of mass movement is directly proportional to slope angle steepness, where a steeper slope has a higher driving force of mass movement (Youssef & Pourghasemi, 2021). Furthermore, a steeper slope increases the velocity of surface runoff which increases the probability of sediments transfer that normally causes slides and debris flow slope failure to occur (Nhu et al., 2020). Some studies that were carried out in Malaysia determined that slope angle is the most influential factor among other slope failure conditioning factors due to the increase of shear stress with the increase in slope angle (Zulkafli et al., 2023). Past slope failure case in Ruan Changkul,

Sarawak occurred with slope angle value from 25° to 40°, indicating that at an angle of 25°, the slope was already deemed to be unstable as the catastrophe destroyed an eight-door longhouse and two houses nearby (Hashim & Among, 2003).

$$\theta = \arctan\left(\frac{y}{x}\right) \quad \text{Equation 2.2}$$

Where  $\theta$  is the slope angle,  $y$  is the vertical distance, and  $x$  is the horizontal distance that coincides with the vertical distance.

vi. Curvature

Curvature is one of the many terrain factor in slope failure studies as seen in Figure 2.5 where it shows the shape of the terrain, a positive curvature imply an upwards convex gradient, a zero value indicates a flat terrain, and a negative curvature value indicates a concave gradient (Rudra & Sarkar, 2023). The impact of curvature on slope failures varies from one location to another, where in some cases the type of curvature does not have any sort of impact on slope failures (Pham et al., 2019). However, it is important to consider curvature in slope failure susceptibility studies to determine its influence, convex curvature are known to accelerate rainfall water flow whilst concave and flat curvature are known to deaccelerate the flow and accumulate the water content, where in both cases, influences the slope failure (I. Ullah et al., 2022).

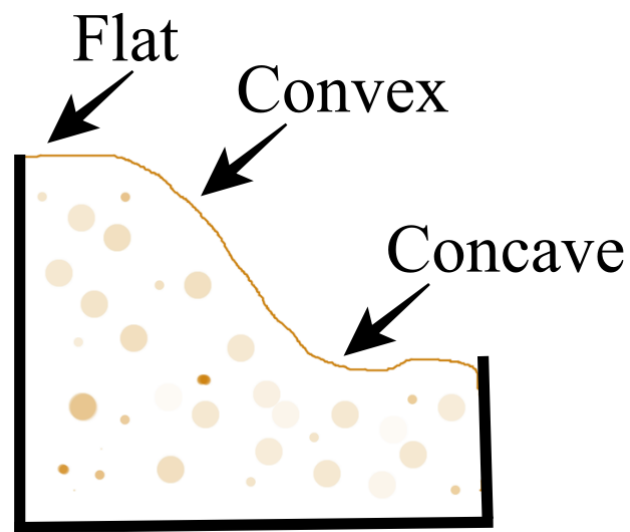


**Figure 2.5:** Terrain Curvature Illustration

vii. Land Use and Land Cover

Human activity has a significant impact on the occurrence of slope failure, where in Malaysia most of the slope failures occurred in areas where construction was done in slope failure prone areas (Majid et al., 2020). To show areas which have been disturbed by human activities, Land Use and Land Cover (LULC) maps are commonly referred, to track extension of urban areas and infrastructure construction as these activities have the tendency to disturb the slope stability (Youssef & Pourghasemi, 2021). LULC also tracks other surface properties such as vegetation cover, and forest changes (Baig et al., 2022).

viii. Vegetation Density

Vegetation can have both positive and negative impacts on the susceptibility of slope failures, depending on the location. Generally, high vegetation can decrease the chance of slope failures as the roots help in slope stabilisation by reducing the amount of water reaching the ground surface, however on rare occasions, highly packed vegetation increases

the amount of water retained on the slope (Achour & Pourghasemi, 2020; Sur et al., 2022). Normalised Difference Vegetation Index (NDVI) is used in slope failure susceptibility mapping to establish the relation between vegetation cover and slope failure occurrences (Achour & Pourghasemi, 2020). NDVI values range from -1 to 1, where each group value within the range corresponds to a certain vegetation condition, that shows whether the area of interest has a high value of vegetation or otherwise (Nnanwuba et al., 2022; Zulkafli et al., 2023).

#### ix. Soil Type

Soils are the product of lithologies that underwent soil forming processes such as weathering (Gray et al., 2016). As soils are typically deposited on the surfaces, it is the first to be exposed to rainfall, and different types of soil have been determined to have different hydraulic conductivities, where soils with high hydraulic conductivity increases the likelihood of slope failures (Gomes & Vieira, 2016). There are five different soil types in Western Sarawak as seen in the following sections arranged from most to least susceptible to slope failure.

##### a. Dystric Histosols (OD)

This soil type consists of mostly organic materials, and form under poor drainage conditions. The term Dystric refers to the acidity of the soil, as Dystric Histosols pH levels are mostly below 5.5. This soil type is not structurally sound, being soft and compressible, and its ability to retain water (Bončina et al., 2023).

b. Dystric Gleysols (GD)

Dystric Gleysols is formed in a similar manner to Dystric Histosols, albeit different in the composition. This soil type consists of mostly mineral particles with some organic matter, making it less susceptible to slope failure in comparison to Dystric Histosols (Tuffour et al., 2019).

c. Orthic Acrisols (AO)

Orthic Acrisols is composed of clay, with a lower water retention capacity in comparison to the previous soil types. Its susceptibility to slope failure is mostly caused by the shrink-swell behaviour of the clay content (Aldiansyah & Risna, 2023).

d. Ferralic Cambisols (BF)

Ferralic Cambisols consists of a variety of parent materials, including volcanic ash, alluvial deposits, and weathered rocks, which has high content of iron and aluminium oxides. The iron and aluminium oxides help in stabilising the soil, and being well-structured, the soil have good drainage which reduces the risk of slope failure. The risk of slope failure in this soil type is mainly caused by the weathering conditions of the soil (Wakwoya et al., 2023).

e. Albic Arenosols (QA)

Albeit prone to erosion due to consisting mostly of sand which has loose and granular structures, this soil type is the least prone to slope failure as it provides a good drainage that reduces the risk of water accumulation (Desyatkin et al., 2021).

x. Distance from Rivers, and Distance from Roads

As a slope's increases in proximity to rivers, the likelihood of failures increases, as the rate of erosion in areas close to a river is higher in comparison to areas that are further, as shear stress is higher in said areas (Mihai et al., 2020). Distance from roads on the other hand is an anthropogenic variable with human settlements tend to be near the road. In addition, cut-fill constructions are mostly carried out nearby a road, it is also a contributor towards slope failure as design error, and construction error on cut-fill may have occurred (Chuma et al., 2023).

2.2.3.3 Slope Failure Factors in Western Sarawak

In Western Sarawak, it was identified that the sole slope failure triggering factor was rainfall, which is prevalent in the wet season of Northeast monsoon for the region (Sa'adi et al., 2019). Seismically triggered slope failures has not been recorded in the region, thus eliminating earthquake as a triggering factor (Halim Ida et al., 2022). The topographical conditioning factors of aspect, curvature, elevation, and slope angle are considered as universal factors as it is present throughout a multitude of research regarding slope failure predictions. Human influence in Western Sarawak is representable through LULC as it tracks the type of human activity in the form of Land Use and Land Cover types (Rudra & Sarkar, 2023). Another variable that is commonly used to track human activity is NDVI, which shows the vegetation density of an area, where dense area are commonly undisturbed areas (Suwanno et al., 2023). However, in Western Sarawak, NDVI was deemed to be unsuitable for representing human activity, as it may provide a false positive confirmation, as areas disturbed for palm oil plantation or other agriculture related activities are detected as high density in NDVI (Cobos-Mora et al., 2023)

#### 2.2.3.4 Remarks

The occurrence of slope failures is influenced by two types of variables, which are triggering factors, and conditioning factors. The weightage of the variables varies depending on the location. As an example, even in a single type of variable such as aspect, exist the probability of a failure would occur more in a certain class, rather than the other. Furthermore, in a slope failure susceptibility study, multivariable conditions are an important viewpoint to provide clarity towards the decision-making process. In Western Sarawak, rainfall is the primary triggering factor for slope failure as the region is safe from any hazardous seismic activity. Conditioning factor of elevation and slope angle has been identified to be the most crucial factor for slope failure assessment throughout the world including Western Sarawak.

### 2.3 Slope Assessment Systems

Traditional SASs are slope failure susceptibility ranking systems, which are very crucial in order to avoid or minimise the damages caused by slope failure, and the systems were mainly developed through quantitative approaches making it an expert view-based method (Albar & Mohd-Nordin, 2022). The accuracy of an SAS is highly dependent on the data used by the experts in the development process, and most of the data originated from a specific or multiple regions that are in near proximity to each other, making the use case of the SAS restricted to the area where the development data was obtained, thus, around the globe, each country has their own SAS (Taherynia et al., 2014). Furthermore, the traditional SAS accuracy deteriorates over time as it was developed with empirical thresholds and physical based model resulting in the need for an update from time to time due to the weightage of each slope failure factors to the occurrence also changes from time to time (Mondini et al., 2023).

### **2.3.1 Slope Assessment Systems in Malaysia**

Malaysia has experienced numerous slope failure cases on a yearly basis, thus the PWD has developed several SASs over the years to update the earlier SAS, where each SAS was used in large-scale to medium-scale assessments of slope. However, the accuracy of the SAS was determined to be unsatisfactory for evaluating slope failure susceptibility in certain use cases, due to various reasons such as the usage of a hazard scoring system derived from other country's SAS, insufficient database, and the use of an overgeneralised approach (Salmanfarsi et al., 2019).

#### **2.3.1.1 Types of SASs in Malaysia**

The following were the SASs developed for Malaysia by the PWD in a chronological manner:

##### **i. Slope Maintenance System**

The first SAS in Malaysia was developed in 1996, which was the Slope Maintenance System (SMS) as part of the East-West Highway long-term preventive measures where it was developed through a statistical method using discriminant analyses based on slope type which are embankment or fill, and cut or natural slope (H. Singh et al., 2008). There were several parameters considered in the development of the SAS, which were slope morphology, lithology type, groundwater, existing mitigation structures, vegetation or slope cover, and erosion (Rahim et al., 2023). The weightings for each one of the parameters were determined through the Factor Overlay Analysis (FOA) method proposed by Anbalagan in 1995, where the maximum weightage of two was assigned to the most hazardous sub-parameter, whereas the weightage for other sub-parameters was determined through the ratio

of cases with the sub parameter to the total cases multiplied by maximum weightage (Equation 2.3) (H. Singh et al., 2008).

$$W_{FOA} = f * W_{max} \quad \text{Equation 2.3}$$

Where  $W_{FOA}$  is the weightage for the sub-variable,  $f$  is the frequency of landslide for the sub-variable, and  $W_{max}$  is the maximum weightage.

ii. Slope Priority Ranking System

Slope Priority Ranking System (SPRS) was developed by PWD in 1999 for a quick assessment of all slopes in Malaysia in order to determine which slope requires the most attention for budget allocation priority, where in comparison to SMS, the approach to develop the SAS was rather simple as the hazard scoring was associated with ratings of 0, 1 or 2 for each sub-parameter in each main parameter (H. Singh et al., 2008). The slope failure factors that were used to develop the SAS were slope morphology, weathering profile, structural geology, groundwater, vegetation or slope cover, and erosion (Rahim et al., 2023).

iii. Slope Information Management System

In 2002, PWD has developed the Slope Information Management System (SIMS), however, the SAS was not developed solely by PWD, instead it was the results of a collaboration between PWD and the Japanese International Cooperation Agency (Huat & Jamaludin, 2005). As an alternative, the hazard scoring system used in the SAS was adopted from a pre-existing hazard scoring system used in Japan (H. Singh et al., 2008). The slope failure factors used develop the SAS were slope morphology, structural geology, groundwater, presence of an existing mitigation structures, vegetation or slope cover, and erosion (Rahim et al., 2023).

iv. Slope Management and Risk Tracking Systems

The Slope Management and Risk Tracking Systems (SMART) was developed by PWD in 2004 with data obtained from the Tamparuli – Sandakan Road in Sabah. The hazard score for each parameter ranges from 0 to 1 derived from the integration of results from statistical method which used stepwise discriminant function where the outcome was converted into probability, deterministic method through Monte-Carlo analysis, and expert knowledge when it was required (Huat & Jamaludin, 2005). SMART was developed using several slope failure factors which were slope morphology, weathering profile, groundwater, presence of existing mitigation structures, vegetation or slope cover, and erosion (Rahim et al., 2023).

2.3.1.2 Remarks

In conclusion, several SASs have been developed by PWD throughout the years with data obtained from several regions in Malaysia with several slope failure factors used to determine the slope failure susceptibility as seen in Table 2.6. The variables considered in each one of the SASs provides crucial information when considering the selection of variables for slope failure studies.

**Table 2.6:** Slope Failure Factors used by the SAS

| <b>Slope Failure Factors</b> | <b>SMS</b> | <b>SPRS</b> | <b>SIMS</b> | <b>SMART</b> |
|------------------------------|------------|-------------|-------------|--------------|
| Slope Morphology             | Yes        | Yes         | Yes         | Yes          |
| Lithology                    | Yes        | No          | No          | No           |
| Weathering                   | No         | Yes         | No          | Yes          |
| Structural Geology           | No         | Yes         | Yes         | No           |
| Groundwater                  | Yes        | Yes         | Yes         | Yes          |
| Mitigation Structures        | Yes        | No          | Yes         | Yes          |
| Vegetation or slope cover    | Yes        | Yes         | Yes         | Yes          |

**Table 2.6** Continued

|         |     |     |     |    |
|---------|-----|-----|-----|----|
| Erosion | Yes | Yes | Yes | No |
|---------|-----|-----|-----|----|

### 2.3.2 Malaysia's SAS Accuracy

A study was conducted in 2008 to evaluate the accuracy of SAS used in Malaysia in predicting the slope failures for rock cut slopes, and metasedimentary rocks, where it has been concluded that none of the SAS has achieved satisfactory results except for SMART in use case of metasedimentary rocks (Salmanfarsi et al., 2020). A total of 44 slope failure cases were used to evaluate the SAS accuracy in cut slopes of rock (granitic formation), where SMS accurately predicted 17 cases, SPRS accurately predicted 23 cases, SIMS only predicted a single case accurately, and SMART predicted 27 cases accurately, resulting in a accuracy score of 39%, 52%, 2%, and 61% respectively as seen in Table 2.7, whereas in the use case of predicting 29 slope failures of cut-slopes in metasedimentary rocks, SMS accurately predicted 13 cases, SPRS accurately predicted 17 cases, SIMS accurately predicted five cases, and SMART accurately predicted 26 cases, resulting in accuracy score of 45%, 59%, 17%, and 90% respectively as seen in Table 2.8 (H. Singh et al., 2008).

**Table 2.7:** SAS Accuracy in Predicting Slope Failure in Cut-Slopes of Granitic Formation

| SAS                | SMS | SPRS | SIMS | SMART |
|--------------------|-----|------|------|-------|
| Failed Slopes      | 44  | 44   | 44   | 44    |
| Correct Assessment | 17  | 23   | 1    | 27    |
| Accuracy           | 39% | 52%  | 2%   | 61%   |

**Table 2.8:** SAS Accuracy in Predicting Slope Failure in Cut-Slopes of Meta Sediments

| SAS                | SMS | SPRS | SIMS | SMART |
|--------------------|-----|------|------|-------|
| Failed Slopes      | 29  | 29   | 29   | 29    |
| Correct Assessment | 13  | 17   | 5    | 26    |
| Accuracy           | 45% | 59%  | 17%  | 90%   |

In conclusion, the accuracy of a SAS is very important as incorrect predictions increases the likelihood of exposure to the dangers of slope failures, and it is very important to have a region specific hazard scoring system, using historical slope failures data obtained from the specific region to develop the SAS for the region with sufficient amount of data, and avoid using oversimplified approaches (Jamaludin et al., 2006).

### **2.3.3 Slope Assessment System in Southeast Asia**

In Southeast Asia, the Association of Southeast Asian Nations (ASEAN) have recommended the Weight of Evidence (WoE) approach to determine the susceptibility of an area of interest preferably a river basin towards slope failures due to the adaptability, and performance (Rama Krishna et al., 2021). Like the method used in developing SMS, and SMART in Malaysia, WoE development is limited to categorical variables (Naceur et al., 2022). The method used to figure out the weightage that each variable has on the occurrence of landslide can be seen in Equation 2.4.

$$W_{WOE} = \ln \left( \frac{\% L_{sv}}{\sum \% NL_{sv}} \right) \quad \text{Equation 2.4}$$

Where  $W_{WOE}$  is the weight of the sub-variable,  $\ln$  is the natural log,  $\% L_{sv}$  is the percentage of landslide event in the sub-variable, and  $\sum \% NL_{sv}$  is the percentage of non-landslide event in said sub-variable.

WoE is highly dependent on the variables to be categorised to determine the weightage of the variables that are considered during the development of the SAS. The simple approach requires historical slope failure cases to be present in each class of the variables. However, exists a possibility where prior slope failure cases do not exist in a certain range of the variables, thus leading to a larger class range, which causes overgeneralisation to occur.

## 2.4 Geographical Information System

Geographical Information System (GIS) is a software that is used to deal with georeferenced spatial data. Spatial data provides information on features that are available on the earth's surface. Spatial data nowadays can be obtained via remote sensing where satellites are used. Remote sensing data are becoming increasingly accessible, as well as accurate, where it has aided countless researches (Adnan et al., 2023). Prior to remote sensing, direct field surveys were the only source of spatial data, which was not cost-effective as it consumes a large amount of time, and could only cover a small area of interest (Chuma et al., 2023). The application of GIS in slope failure susceptibility assessment is very important to find relationships amongst spatial data of historical slope failure cases and the input variables. In any data extraction method for spatial data, the extracted data are spatiotemporal as any recent changes in real life are not tracked in the spatial data (Bambang

et al., 2023). Issues in the form of recent changes affect the LSMs contemporality as it is not shown in the LSMs.

There are two major types of GIS file format, which are vector and raster (Baig et al., 2022). Spatial data in the form of polygons, points, or lines are known as vector data, which is very suitable for storing data regarding information with well-defined borders that is not influenced by numerical value changes such as, the borders of neighbouring districts, road networks, and historical slope failures points (Senanayake et al., 2022). In certain cases, vector data are required to be in the raster format, and the conversion can be done through a process known as rasterization, which converts vector data into raster data, based on the assigned resolution, due to vector data not being represented by resolution (Zulkifli et al., 2023). Moreover, a single vector file can store numerous information of several type such as date, and integer in the attribute table (QGIS Development Team, 2023). Vector data can be created through drawing the polygon, line, points, or importing the csv file with the required information such as latitudes and longitudes (Tengtrairat et al., 2021).

Raster data files are a form of image data, thus, raster data are stored in the form of pixels, where a higher resolution raster provides a more detailed information on the data it represents (Ighile et al., 2022). Raster data files are very suitable for data with many variation in value over a large area, where representing the data in the form of vector would be tedious as a large number of borders are required (Babitha et al., 2022). Slope failure contributing factors such as elevation, slope angle, and curvature are very suitable to be stored as raster files as creating borders with specific ranges of values would decrease the amount of information that can be represented (Ado & Amitab, 2023). Furthermore, a single raster file can be used to derive several other raster file especially elevation raster, which could produce slope angle raster, curvature raster, and TWI, through terrain analysis in the GIS's software

or the raster calculator tool (Senanayake & Pradhan, 2022). A single raster file can comprises multiple raster through raster stacking, which save the raster file in their own respective raster layers known as band (Tehrani et al., 2022). A raster file can also be converted to a vector file through vectorisation, the pixel values affect the outcome of the process, with a higher resolution raster would produce a better data gradient in comparison to a lower resolution raster (QGIS Development Team, 2023).

The use case of the spatial data format is highly dependent on the intended application. In the case of slope failure related studies, to obtain the input values of the triggering, and causative variables, the data must be in the raster format, whereas the output of slope failure and slope failure cases must be in the vector format. As for mapping purposes, especially through a function, the resolution of each input raster is very influential to the end-product, as a high-resolution input, would yield a high-resolution output map. Thus, whenever possible, the highest resolution raster files should be used for the mapping purposes, given that the computing power available is adequate. The application of GIS is crucial in natural hazards assessment, as it has been determined that natural hazards occurrences often time correlated spatially with its causative variables. The usage of GIS has certainly provided a platform for managing, converting, and creating the spatial data.

## **2.5 Artificial Neural Network**

Artificial intelligence is a system that is programmed to sense, reason, act, and adapt to the data on which it was trained. Machine Learning is a subset of artificial intelligence, where its predictive performance increases when exposed to a large amount of data. In the case of an ANN, a large enough dataset would require the neural network to be more complex, and this would convert the Machine Learning model, to a deep learning model as

seen Figure 2.6 (Khajehzadeh et al., 2022). Machine Learning covers a broad definition of several Machine Learning algorithms such as ANN, support vector machines, fuzzy logic, and random forest, where these algorithm has been widely utilised to solve science and engineering problems for the past decades (Lary et al., 2016).

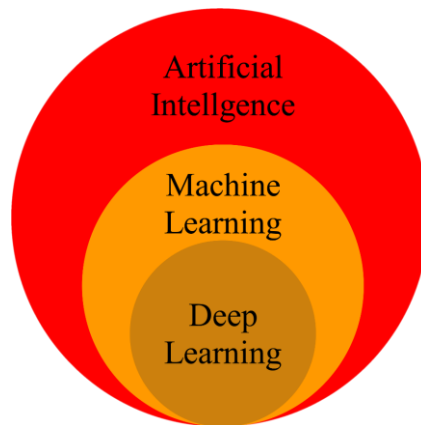


**Figure 2.6:** AI and its' Subset

Everything under the banner of artificial intelligence is very similar as it attempts to mimic in a way that a computer could understand, the human brain to solve non-linear problems with large amounts of data (Shahin, 2016).

ANN is a Machine Learning model that is designed to mimic the interconnected network of neurons in the human brain with an input layer, one or several hidden layers, and an output layer as seen in Figure 2.7 (Khajehzadeh et al., 2021). An ANN is very similar to a human brain to the point where it operates is designed based on the information transfer process, where the neurons receive signals through synapses and if the signals are strong enough, the neurons is activated, and the information is passed to another neuron via the axon (Gershenson, 2003). Moreover, if the data used to develop the ANN model is large enough, the number of neurons in the hidden layer, and the number of hidden layers may be

adjusted to further accommodate the increasing complexity of the problem, where if the model and data is complex enough, it has evolved into deep learning (Hakim et al., 2022).

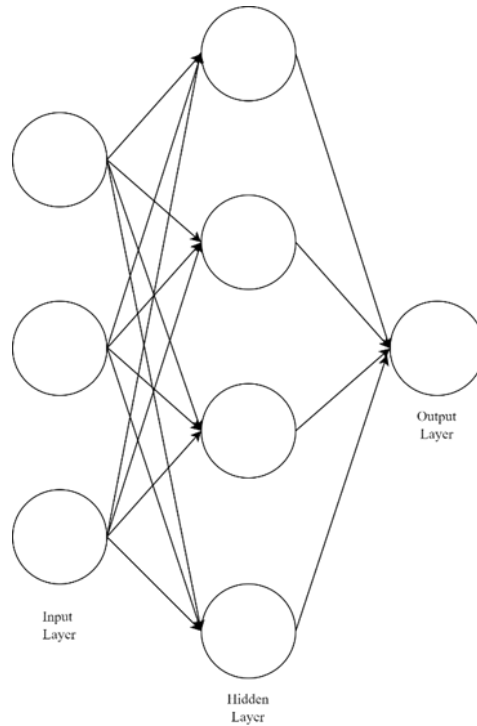


**Figure 2.7:** ANN's Architecture

### 2.5.1 Innerworkings of an ANN

As an ANN model is designed to mimic the human brain, there are many processes involved to calculate the outcome, which involves a lot of components such as the formula, bias, number of hidden layer neurons as well as the number of the layers itself, learning algorithm, error function, and activation function, as seen in Figure 2.8 (Fristch et al., 2022). The inner-working of an ANN model starts from data-preprocessing process in the development phase. The input of an ANN model is divided into training data and testing data. Training data is used to develop the predictive property of an ANN model, whilst the

testing data is used to evaluate the performance of the model. From the input, the data is passed into the neuron(s) in the hidden layer, which was then passed to the output layer. The passing process is reiterated until the lowest possible error has been achieved.

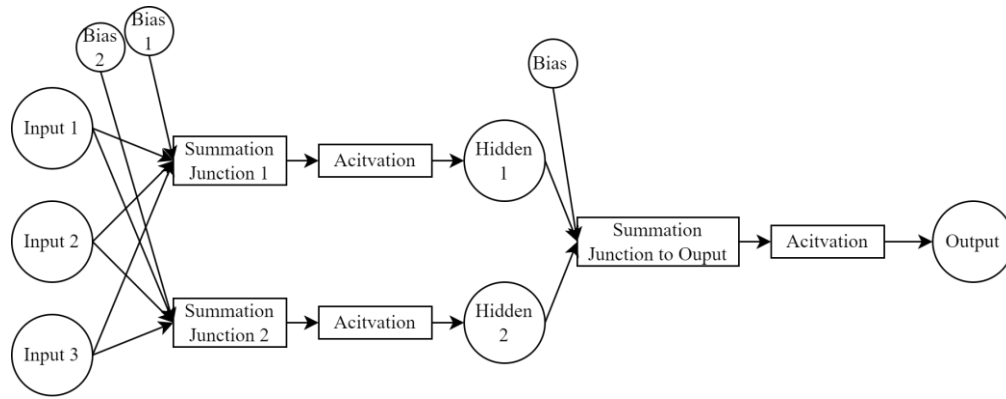


**Figure 2.8:** How an ANN works

#### 2.5.1.1 Training and Testing

Data is essential in every type of Machine Learning models, as it is used to train the model to develop its' predictive properties as well as to evaluate the accuracy of the model prediction (Hassanzadeh et al., 2018). Both of the training data and testing data originated from a single primary dataset, thus both having the same type of variables present, where the only difference between the training data and the testing data is the amount of data or rows present (Das et al., 2011). There are no absolute rules to divide the primary dataset into the training data and testing data, however, as a general rule of thumb, the number of instances in the training data should be more than that of the testing data, as the training data is vital in the learning process of the ANN model (Khan et al., 2016). A good dataset for Machine Learning model must represent all possible combinations of input parameters to each type of expected output, and some Machine Learning models such as Support Vector Machine requires data balancing to avoid an oversimplified prediction (Gupta & Shukla, 2020).

#### 2.5.1.2 Data Preprocessing

Data preprocessing is an important step in any Machine Learning model development. It is done prior to the splitting of the main dataset into the training data and testing data to create a uniform data format that can be understood by computers (Utomo et al., 2020). Data quality and data preprocessing are known as one of the major phases in the development of a Machine Learning model where it constitutes at least 80% of the total development (Ma et al., 2020). There are two main types of data that requires preprocessing, which are numerical data and categorical data, and the method used to process each type of data is different (Swamynathan, 2017).

##### i. Numerical Data

Numerical data preprocessing is done through scaling where each of the variables is scaled to be in the same range as the raw data especially in slope failure studies ranges widely from one to the other, such as slope angle and elevation. An unscaled numerical dataset increases the chance of the Machine Learning model's learning process to be lengthy or unstable that could cause the learning process to fail or exhibit a high error score (Sharma, 2022). A popular method to conduct data preprocessing on numerical data is data scaling through Min-Max scaling or minimum (min) – maximum (max) scaling which scales the values to be in the range of 0 to 1 based on the maximum values and minimum values of each parameters (Horton et al., 2015). The Min-Max scaling equation can be seen in Equation 2.5.

$$X_{scaled} = \frac{X - \min(X)}{\max(X) - \min(x)} \quad \text{Equation 2.5}$$

Where,  $X$  is the unscaled value of variable,  $X_{scaled}$  is the scaled value of variable,  $\min(X)$  is the lowest value of variable, and  $\max(X)$  is the highest value of variable.

## ii. Categorical Data

Unlike numerical data, categorical data is not represented by numbers, instead it is represented by categories as it is qualitative instead of quantitative in nature, such as the type of soil, compass orientation, and the type of surface curvature (Wissmann & Toutenburg, 2007). Thus, categorical data need to be categorised before using it to develop the Machine Learning model, where a popular method to conduct data pre-processing of categorical data is through one-hot encoding, where it works by categorising the data into its' unique field, whenever the data is present within the category, it is marked as 1, if otherwise it is represented by 0, marking the absence of the data in the category field (Guerzhoy & Hinton, 2017). In conclusion, one-hot encoding is widely used in models to ensure the Machine Learning models understand that said data is not quantified as a numerical data as it is represented by its' unique category that is set by the programmers (Ul Haq et al., 2019).

### 2.5.1.3 Formula

Formula is the basis of an ANN model, it is a symbolic description of the process that will be carried out the ANN model, where the target variable and input variables are present (Fristch et al., 2022). The formula is assigned at the start of the ANN model, where the input and output values are present from the training data to provide a learning material for the ANN model. When the input and output values are assigned, the ANN model is known as a supervised Machine Learning model (Shapiro et al., 1999).

#### 2.5.1.4 Learning Algorithm

An ANN learns from the training data by weights and biases adjustment through a learning process known as feed-forward, and if the output of the feed-forward process is found to be unsatisfactory, the output are fed back to the network, improving the algorithm accuracy, this process is known as backpropagation, a well-known ANN learning algorithm (Shapiro et al., 1999). Each time the training data is re-inserted into the feed forward network, the output error is minimised through a gradient descent method, that calculates the loss function of the all weights that are present in the network (Swamynathan, 2017). In conclusion, backpropagation works by configuring random connection weight at the start of the training phase. The weights are adjusted iteratively until the lowest value of error has been achieved, where the sample actual value and the ANN model predicted value differs at its lowest (Sheng et al., 2023).

#### 2.5.1.5 Error Function

Mean square error (MSE) and root mean square error (RMSE) is a popular error function used to evaluate the performance of an ANN model where MSE estimates the generalisation error of the ANN model, and RMSE measures the forecasting errors of model, and can be seen in Equation 2.6, and Equation 2.7 (Nguyen et al., 2019). MSE and RMSE provide an indication on how accurate the ANN model prediction is, thus, lower MSE and RMSE values are favourable, and if the accuracy is not satisfactory, the hyperparameter of the ANN model can be tuned or the main dataset can be re-evaluated (Khan et al., 2016).

$$MSE = \frac{\sum_{i=1}^N (X_{obs} - X_{est})^2}{n} \quad \text{Equation 2.6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_{obs} - X_{est})^2}{n}} \quad \text{Equation 2.7}$$

Where,  $X_{obs}$  denotes the original values of the output, and  $X_{est}$  denotes the predicted values of the output.

#### 2.5.1.6 Summation Junction and Activation Function

In every ANN model from one layer to the other, there is a summation junction that sums up the pre-processed data with regards to the weightage and biases, which is then activated through an activation function (Lary et al., 2016).

##### i. Weightage

Weightage is a numerical representation of the importance of each input variable to the output in a summation junction, and it is determined by the ANN model based on the training data (Selamat et al., 2022a). In an ANN model with a backpropagation algorithm the weightage is adjusted iteratively until the error value is at its lowest (Abidoye et al., 2018). Each assigned weightage will act as a coefficient to the output before passing the information to the subsequent layer(s) (Senanayake & Pradhan, 2022).

##### ii. Biases

Biases are another numerical value that is assigned by the ANN model, however, in comparison to the weightage, it is not a coefficient for each variable, instead it is a numerical value that is added to summed weightage and their respective variable values (Abidoye et

al., 2018). The primary function of the ANN's network parameter such as biases and weightage are to reduce the likelihood of overfitting, where the ANN just memorise the training data, thus having a lower accuracy when it comes to new data, as per evaluated by testing data (Das et al., 2011).

#### 2.5.1.7 Activation Function

Whenever the transferred information exceeds a certain threshold, it is activated through an activation function (Gershenson, 2003). In cases of slope failure susceptibility, the activation functions specifically the sigmoid activation function is responsible for the susceptibility score values to fall in the range of 0 to 1 (Bradley, 1997). The sigmoid activation function is a popular activation function used to develop an ANN model due to its ability to handle complex relationships amongst the input data to the output, and the smooth continuous outcome (Youssef & Pourghasemi, 2021). The sigmoid activation function can be seen in Equation 2.8 (Shapiro et al., 1999).

$$f(x) = \frac{1}{1 + e^{-x}} \quad \text{Equation 2.8}$$

Where,  $f(x)$  is activated passed information, and  $x$  is passed information.

#### 2.5.1.8 Remarks

An ANN model is a complex Machine Learning model. A simple form of an ANN model with a single neuron in a single hidden layer is known as a linear regression (Yao et al., 2022). The use case of an ANN for slope failure susceptibility studies is not a novelty, especially for LSM development. However, past research did not consider the possibility that the LSM could obsolete, given that certain important remote sensing data are seldomly updated such as the elevation data from the National Space Agency (NASA), which was last

updated in 2015 (Laurencelle et al., 2015). Thus, the inner workings of an ANN model could provide a framework for a new SAS that could be used at any given time to evaluate the slope failure susceptibility of an area of interest.

### **2.5.2 Multicollinearity**

Multicollinearity is a phenomenon where a linear relationship can be seen amongst the input variables in both training data and testing data, which is not beneficial for the ANN model, as it detects variables with high collinearity to be the same variable, making the impact of variables with collinearity the same, thus, introducing data redundancy to the ANN model (Bravo-López et al., 2023). A multicollinearity assessment is done on the main dataset prior to the splitting process that produces the training data, and the testing data to choose the appropriate variables that should be kept in the dataset (Selamat et al., 2022b). If and only if, significant collinearity is observed to be present amongst the variables, one or several input variables must be removed depending on the multicollinearity score to avoid any further problems down the line such as degradation of the ANN model's accuracy (Diebold, 2019). There are several methods used to find multicollinearity amongst input variables for Machine Learning purposes.

#### **2.5.2.1 Variance Inflation Factor**

Variance Inflation Factor (VIF) is a method used to determine the presence of multicollinearity in Machine Learning model's data where if the VIF score is more than 2.5, bias is introduced to the model, VIF score of less than five indicates a moderate multicollinearity, and the VIF score exceeds ten, multicollinearity is present (Adnan et al., 2023). In data science, the multicollinearity assessment is done through Equation 2.9 in various packages for a programming language (Swamynathan, 2017).

$$VIF = \frac{1}{1 - R^2} \quad \text{Equation 2.9}$$

Where R is the coefficient of determination of variable.

#### 2.5.2.2 Pearson's Correlation Analysis

Pearson's Correlation Analysis (PCA) multicollinearity assessment coefficients fall in the range of -1 to +1, where whenever the multicollinearity coefficients gets closer to the positive and negative limits, a strong correlation is observed (Djukem et al., 2020). A PCA coefficient of 0 to 0.2 indicates a very weak correlation, 0.2 to 0.4 indicates a weak correlation 0.4 to 0.6 indicates a moderate correlation, 0.6 to 0.8 indicates a strong correlation, and 0.8 to 1.0 indicates a very strong correlation, and generally variables with PCA coefficients of more than 0.8, undergoes removal of one of the variables (Moore et al., 2011). The equation for PCA can be seen in Equation 2.10.

$$PCA = \frac{N\sum xy - \sum x \sum y}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} \quad \text{Equation 2.10}$$

Where N is the number of pairs of variables, X is variable X, and Y is variable Y.

#### 2.5.2.3 Remarks

Multicollinearity is something that should be avoid in any use case of a Machine Learning model. Both PCA, and VIF are valid methods for evaluating multicollinearity amongst the input variables. Although a multicollinearity check amongst the input variables is the main purpose of conducting a multicollinearity assessment, it can also be used to evaluate relationship between the input variables to the output variables. Furthermore, when including the output variables in the multicollinearity assessment, it provides an overview of which input variables should be removed when multicollinearity amongst the input

variables exists. This is done by removing the input variables variable with the lower correlation to the output variable.

### **2.5.3 Evaluation Metrics**

An ANN model accuracy for the training and testing phase is commonly evaluated through the Area Under the Curve (AUC) method as seen in Equation 2.11. AUC is a component of Receiver Operating Characteristic (ROC) Curve, where it is the area underneath the curve. ROC curve is plotted with True Positive Rate (TPR) and False Positive Rate (FPR) value as the axis with TPR as the vertical axis, and FPR as the horizontal axis (Lee et al., 2020). Upon plotting the ROC curve, the AUC can be determined. AUC provides a method to determine the accuracy of an ANN where it has values in the range of 0 to 1, with 1 being 100% accurate while 0 being 0% accurate (Nnanwuba et al., 2022). Whilst AUC provides the information on the general accuracy of an ANN model, recall or TPR is also integral as it measures the performance of the model in predicting relevant instance as is it a measure of ratio for correctly predicted positive values as seen in Equation 2.12 (Shahabi et al., 2022). Recall is important in slope failure susceptibility prediction models as generally the number of failed slopes is smaller than non-failed slopes, thus, a higher recall indicates that the model is able to predict slope failure successfully (A. Singh et al., 2024).

$$AUC = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{Equation 2.11}$$

$$TPR = \frac{TP}{TP + FN} \quad \text{Equation 2.12}$$

Where, TP is the amount of true positive prediction, TN is the amount of true negative prediction, FP is the amount of false positive prediction, and FN is the amount of false negative prediction.

While RMSE provides an overview of the precision of the Machine Learning model, accuracy assessment is conducted to evaluate the overall accuracy. Generally, in slope failure susceptibility assessment, an AUC score of more than 0.8 is accepted the minimum threshold for a Machine Learning model to be considered as accurate. Recall is also an integral metric especially in a study where the number of positive instances is smaller than the number of negative instance such as the number of failed slopes is smaller than non-failed slopes. It is important to note that the accuracy of the Machine Learning model for any purpose is highly dependent on the accuracy of the data that it has been trained with.

## 2.6 Slope Failures Input Data Comparison

Slope failures input data in this case is the data used to develop the SAS, and the Machine Learning model for slope failure susceptibility analysis purposes. This was done to provide more context on what factor should be investigated for the development of a new ANN model. The SAS input variables were compared to Machine Learning based approach of Random Forest (RF), ANN, Convolutional Neural Network (CNN), and Fuzzy Analytical Hierarchy Process (Fuzzy AHP) as seen in Table 2.9. It has been determined that 8 input variables are the most common number of variables to be considered in developing the

system, and slope angle is the most prevalent input variables with it being present in every system.

**Table 2.9:** Comparison of Slope Failure Input

| Variable           | System |   |   |   |   |   |   |   | % Present |
|--------------------|--------|---|---|---|---|---|---|---|-----------|
|                    | A      | B | C | D | E | F | G | H |           |
| Aspect             | 1      | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 62.5      |
| Curvature          | 0      | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 50.0      |
| Cut Slope Age      | 1      | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12.5      |
| Elevation          | 0      | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 50.0      |
| Erosion            | 1      | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 37.5      |
| LULC               | 0      | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 37.5      |
| Lithology          | 0      | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 37.5      |
| Natural Water Path | 1      | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 62.5      |
| Ponding            | 0      | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 12.5      |
| Rainfall           | 0      | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 37.5      |
| Seepage            | 0      | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 37.5      |
| Slope Angle        | 1      | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 100.0     |
| Slope Cover        | 1      | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 62.5      |
| Slope Height       | 1      | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 50.0      |
| Soil Series        | 1      | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 50.0      |
| Surface Drains     | 1      | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 37.5      |
| TWI                | 0      | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 25.0      |
| No.                | 8      | 8 | 7 | 7 | 9 | 6 | 8 | 6 |           |

Note. A is SMS (1996), B is SPRS (1999), C is SIMS (2002), D is SMART (2004), E is ANN by (Selamat et al., 2022a)., F is CNN by (K. Ullah et al., 2022), G is RF by (Nnanwuba

et al., 2022), and H is Fuzzy AHP by (Sadiki et al., 2023). No. is the number of variables. 1 represents the variable is present in said model, and 0 represents the variable is absent in said model. % present is percentage of the variable used in every model.

## 2.7 Landslide Susceptibility Mapping using GIS and Machine Learning

Knowing the susceptibility of area towards slope failures is crucial especially in the presence of an extensive amount of slope failures triggering factors to efficiently plan the emergency response, land planning, and disaster mitigation (Chen et al., 2020). An LSM is a graphical representation of slope failure susceptibility, which provides details on the susceptibility rate as well as the coordinates based on the reference system (Bravo-López et al., 2023). The LSM of an area can be developed through a multitude of approaches be it using the local SAS, or Machine Learning algorithms, which is gaining traction due to its' accuracy and deployment speed (Kalantar et al., 2019). The susceptibility rate ranking is highly dependent on the model used to determine the susceptibility, where generally a higher score correlates with a higher risk of susceptibility (Moresi et al., 2020). A detailed review of Machine Learning model use for LSM can be seen in Table 2.10.

**Table 2.10:** Previous Machine Learning LSM Study

| Author(s)               | Description                                                                                                                                                                                                         | Gap                                                        |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| (Mekonnen et al., 2022) | The decision-making process was done through Analytical Hierarchy Process, and experts were referred to for the weightage of each slope failure factors. Resulting LSM was determined to have an accuracy of 84.8%. | A mix of expert-view, and Machine Learning method for LSM. |

**Table 2.10** Continued

|                            |                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (Badola et al., 2023)      | Rapid development increases the likelihood of slope failure. Due to the large dataset, the application of traditional SAS for LSM development is not suitable. The Mamdani's fuzzy inference system has achieved an accuracy of 77.7% for the LSM with the incorporation of remote sensing data.              | No new SAS was derived from the Machine Learning model.                                                                                                                        |
| (Md.Sharafat et al., 2024) | A slope failure inventory of 255 points was used in this study where the resulting LSM developed from LR, and RF have been determined to have accuracy exceeding 90%.                                                                                                                                         | ANN was not deployed, and no new SAS was derived from the Machine Learning models.                                                                                             |
| (Moragues et al., 2024)    | The study focused on the dynamic change of a terrain due to glacier melts, that causes the slope failure susceptibility to increase. Remote sensing data was incorporated, and the slope failure susceptibility was determined through a Frequency Ratio Model, resulting in an LSM with an accuracy of 80.4% | The Machine Learning model required the data to be in a categorical form, where the weight of factor was determined through the percentage of landslide pixels in said factor. |

**Table 2.10** Continued

|                               |                                                                                                                                                                                                                                                                                                                          |                                                                          |
|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|
| (Yilmaz et al., 2022)         | Typical Machine Learning model have been determined to have a poor generalisation capability as evident by the average satisfactory score for landslide susceptibility assessment. It has been determined that a deep learning model outperform a Machine Learning model with an increase of accuracy for the LSM by 6%. | No new SAS was developed. The Machine Learning model compared to was RF. |
| (Vidal-Paez et al., 2020)     | Development of an LSM through fuzzy logic. No mention of accuracy. Remote sensing data was applied.                                                                                                                                                                                                                      | No new SAS was developed and the LSMs were accepted as the end-product.  |
| (Youssef & Pourghasemi, 2021) | Comparison of Machine Learning model in the development of an LSM. All Machine Learning models were highly accurate with accuracy ranging from 89% to 95.1%. The Machine Learning models consisted of support vector machine (SVM), RF, Multivariate Adaptative Regression Spline, and ANN.                              | No new SAS was developed and the LSMs were accepted as the end-product.  |

**Table 2.10** Continued

|                           |                                                                                                                                                                                                                                                                              |                                                                         |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| (Patil & Panhalkar, 2023) | Development of an LSM through a Landslide Numerical Risk Factor (LNRF) model, and remote sensing data. The model accuracy was determined to be 77%.                                                                                                                          | No new SAS was developed and the LSMs were accepted as the end-product. |
| (W. Chen et al., 2018)    | Comparison of Machine Learning models' accuracy in the development of an LSM. The Machine Learning models consisted of Bayes' net, radical basis function classifier, logistic model tree, and RF. The random forest outperforms the other models with an accuracy of 0.752. | No new SAS was developed and the LSMs were accepted as the end-product. |
| (Fan et al., 2023)        | Development of an LSM through LR and a hybrid Machine Learning model with accuracy of 70.7%. The study was focused on earthquake influences on slope failure in the study area.                                                                                              | No new SAS was developed and the LSMs were accepted as the end-product. |

**Table 2.10** Continued

|                              |                                                                                                                                                                                                         |                                                                                                                                                                  |
|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (Naceur et al., 2022)        | Development of an LSM through WoE, Radial Basis Function Network, and SVM, where the models each resulted in an AUC of 83.72%, 93.68%, 94.37% respectively.                                             | No new SAS was developed and the LSMs were accepted as the end-product. However, WoE as suggested by ASEAN was deemed to have the lowest accuracy in this study. |
| (Achour & Pourghasemi, 2020) | This study has proven Machine Learning models are suitable, and viable to be used in the development of an LSM, where the AUC of the RF model used in this study was 97.2%.                             | No new SAS was developed and the LSMs were accepted as the end-product.                                                                                          |
| (Anis et al., 2019)          | In this study, the evidential belief model (EBF) was compared to WoE for developing an LSM in North-Western Tunisia. The WoE model outperforms the EBF mode, with AUC of 86.5%, and 84.8% respectively. | No new SAS was developed and the LSMs were accepted as the end-product.                                                                                          |

**Table 2.10** Continued

|                         |                                                                                                                                                                                                                                                                               |                                                                         |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| (Zulkifli et al., 2023) | LSM development through fuzzy logic in Majene, West Sulawesi. The study did not provide details on the accuracy. Furthermore, it did not mention the sensitivity of each slope failure variables present based on the dataset used in the Machine Learning model development. | No new SAS was developed and the LSMs were accepted as the end-product. |
| (Selamat et al., 2022a) | Development of an LSM with ANN through remote sensing data in Langat River Basin, Selangor. Through the ANN model variable importance, curvature was determined to be most influential variable.                                                                              | No new SAS was developed and the LSMs were accepted as the end-product. |

Flood is another natural disaster that is influenced by a plethora of factors and causes tremendous amounts of damages to the society and economy, making the determination of flood susceptibility a crucial step to avoid the potential harm (Rudra & Sarkar, 2023). Unlike slope failures, flood does not have a well-established qualitative method to determine the susceptibility, thus most cotemporary research opted to use Machine Learning based method instead to develop the flood susceptibility map (Ighile et al., 2022). GIS provides researchers with a method to investigate the spatial agreement between spatial data and flood events by observing the patterns of floods to develop the flood inventory, selecting the spatial data that

causes floods to occur, prior to the development of the flood susceptibility map (Adnan et al., 2023).

Erosion susceptibility is inherently complex to predict, however with the combination of GIS and Machine Learning algorithms, researchers have successfully mapped out the soil erosion susceptibility or their respective area of interest (Chuma et al., 2023). Through GIS, the relationship between the soil erosion causative factors such as rainfall, slope angle, and curvature to soil erosion was determined (Senanayake et al., 2022).

## **2.8 Machine Learning and Traditional Factor of Safety**

Traditional approaches in evaluating slopes stability are still applied to this day, such that of the Finite Element Method, Fellenius Method, or Bishop Method, with the deployment being unsuitable for immediate action as lab studies are required, and costly (Rabie, 2014). In a recent study that was done to compare the accuracy of Machine Learning model specifically an ANN model to the conventional method, it was discovered that the ANN model of the study was comparable to the Finite Element Method's factor of safety as seen in Figure 2.9 (Alqawasmeh, 2023).

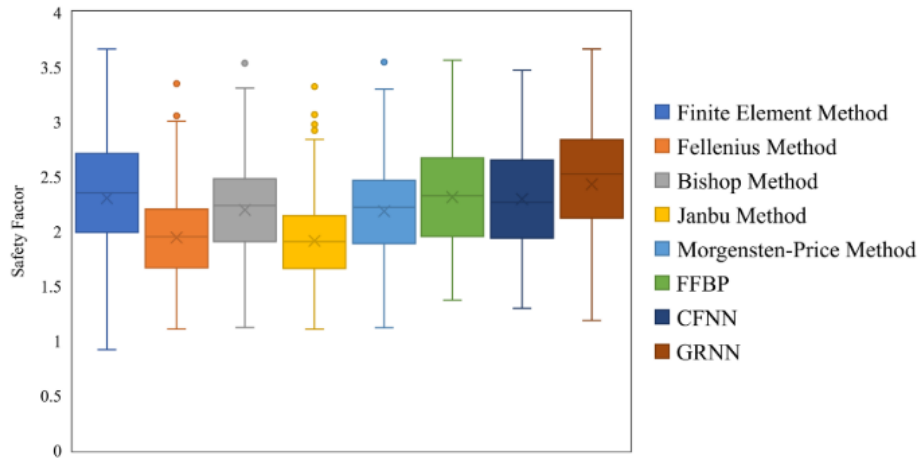


**Figure 2.9:** ANN Model of the Study in Comparison to Conventional Method

## 2.9 Summary of Review

Slope failure is certainly one of the prevalent forms of natural disaster. The impact of slope failure can last for years depending on the severity. There are numerous types of slope failures, however, the factors leading to its occurrence are generally the same, albeit at a different rate. To mitigate the potential risk posed by slope failures in Malaysia, several SASs have been developed by PWD. The inconsistency of the SASs has deterred its application to region outside of the source data location. Thus, this study provides a contemporary approach to the problem through a Machine Learning model with GIS. Throughout the literature review, several gaps were observed from the previous studies which were the dependency of the pre-existing SAS on expert's opinions which may introduce personal biases in the system, the dependency on computing device access for deployment of ANN models in predicting slope failure, the spatiotemporal issues of the LSMs, and the techniques suggested by ASEAN has not been compared to each other. Although an LSM can be considered as a viable preliminary assessment in slope failure susceptibility analysis, due to the temporal properties of certain input variables such as LULC, and rainfall intensity, in addition with the terrain morphology that changes from time

to time that might not be reflected in a digital elevation model given the update time, an SAS is crucial to complete the LSM. Unlike LSM, an SAS accuracy is not restricted to any changes on site. Hence, the novelty of this research lies in the formation of the new SAS made from an ANN model which solves the issue of spatiotemporal and dependency on computing devices, which is complementary to the LSM.

## **CHAPTER 3**

### **METHODOLOGY**

#### **3.1 Overview**

This chapter describes the materials and methods that have been applied in in this research to develop the ANN models and subsequently develop the LSM and SRAS through the respective ANN models. This study was carried out in four primary phases as seen in Figure 3.1, which were Phase I data pooling and data pre-processing, Phase II development of ANN models for Western Sarawak, Phase III LSM development and verification for Western Sarawak, and Phase IV development and verification for Western Sarawak SRAS. Traditional soil physical properties were not used as variables in this study due to the unavailability of said data, where to obtain the information is costly and time consuming as the study area was large. A separate research was conducted to validate the usability of remote sensing data in comparison to traditional soil physical properties variables and comparing the ANN model performance to ASEAN recommended methods at river basin level.

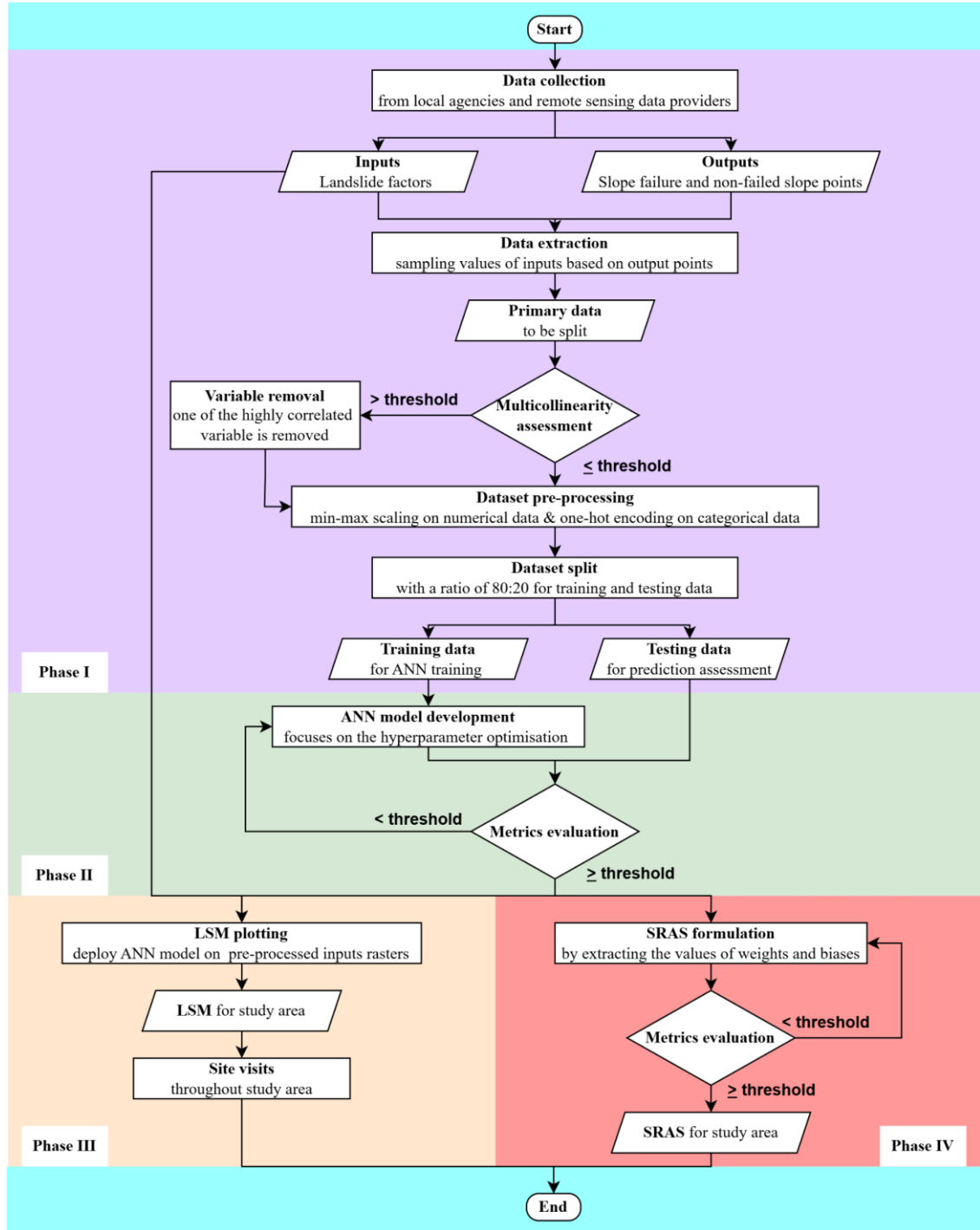


**Figure 3.1:** Flowchart of the Research Methodology

### 3.2 Phase I Data Pooling and Data Pre-Processing

Phase I focuses on collecting the required data and pre-processing the data for the ANN model development and evaluation. There were two categories of data required to form

a single primary data that will later be split into the training data and testing data which were the inputs in the form of landslide factors, and outputs in the form of historical slope failure and non-failed slopes points. Multicollinearity assessment was conducted on the primary data to determine whether highly correlated inputs are available in the dataset. Highly correlated inputs will cause issues during the ANN model training phase as two variables are detected as a single variable, which will cause input redundancy. If there were no multicollinearity issues or the issue has been solved, the primary data were then to be split into the training data and testing data with a ratio of 80:20. The procedures conducted in Phase I is shown in Figure 3.2.

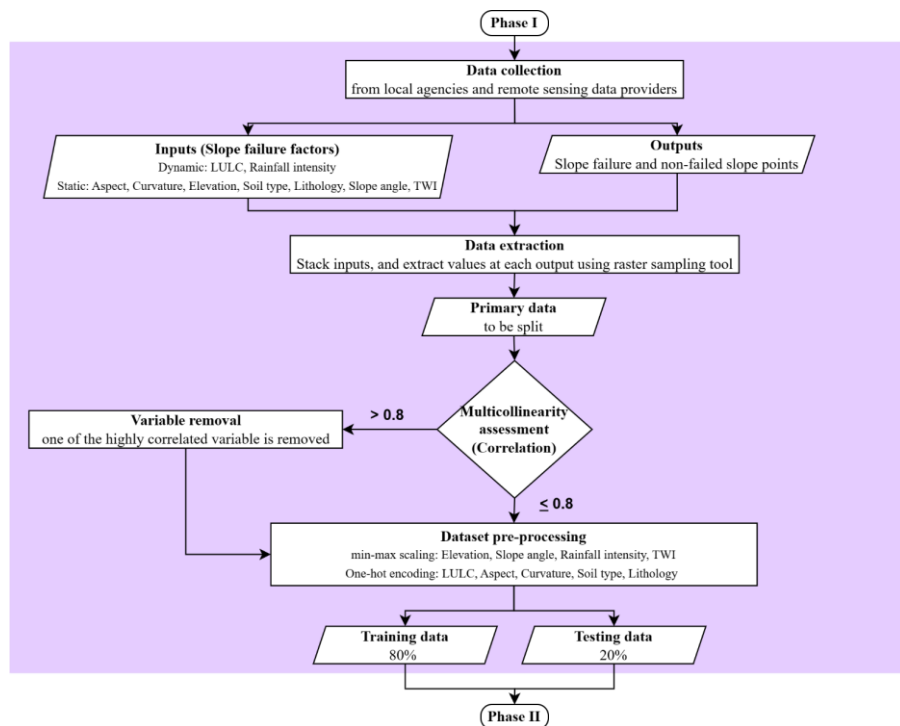


**Figure 3.2:** Flowchart of Phase I

### 3.2.1 Data Collection

Data is the basis of any Machine Learning models, and the ANN models developed in this study requires both inputs and outputs the approach was conducted through a

supervised learning model. The required data were collected from various remote sensing data providers and the local authorities. As spatial data are spatiotemporal, it was very important for the inputs to represent the at-time properties at each outputs. However, future spatiotemporal issues were to handle by the SRAS.

#### 3.2.1.1 Outputs - Slope Failure Inventory

Slope failure inventory refers to past slope failure and non-failed slope points location, which were assigned to a value of 1 and 0 respectively (Kaamin et al., 2016). The past slope failure points were provided by the Department of Minerals and Geosciences Malaysia (JMG) in addition to the failed slopes points discovered during site visits which can be referred in Appendix A. All slope failure points utilised in this study were obtained from 2018 to 2023. The non-failed slope points on the other hand were randomly selected based on the SAS developed by Jamaludin and collaborators in 2006, where it was identified that slopes with angle lower than have the lowest likelihood of failure (Jamaludin et al., 2006). In total, there were 49 failed slopes, and 1167 non-failed slopes points were obtained and utilised in this study as seen in Figure 3.3. Although at an imbalance, in the case of developing an ANN model, the ratio of failed slopes and non-failed slopes were considered to be acceptable, as there is a lack of recorded slope failure cases, especially in regions that are far from human settlements (Gupta & Shukla, 2020).

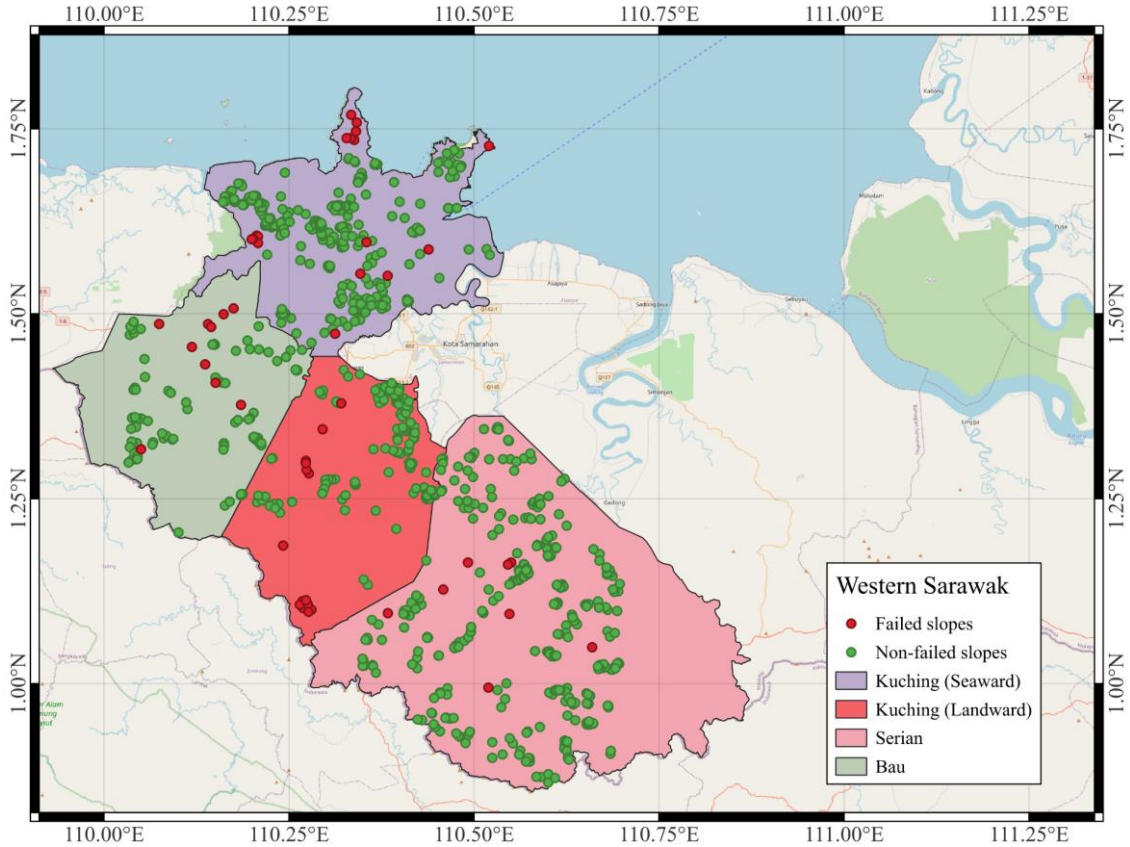


**Figure 3.3:** Historical Failed Slopes and Non-Failed Slopes

#### 3.2.1.2 Slope Failure Factors

The slope failure factors in this study can be categorised into dynamic and static variables (Akter et al., 2019). The dynamic variables were LULC, and rainfall intensity, whereas the static variables were aspect, curvature, elevation, lithology, slope angle, soil type, and TWI. To develop the ANN model for the SRAS, only static variables were utilised as this avoid any issues of spatiotemporal changes, whereas the LSM's ANN model have used both dynamic and static variables (Abidoye et al., 2018). To ensure that the dynamic variables accurately represent the on-site conditions for the duration of the output extractions, the variable extraction time were kept at the year at 2021, during the pandemic as human activities were limited at the time, thus restraining the dynamic changes in the variables (Meyer et al., 2020).

i. Aspect

The aspect raster file in this study was obtained as a derivative product of the elevation raster by applying the aspect analysis function from the Geospatial Data Abstraction Library (GDAL) in Quantum Geographical Information System (QGIS) (Lemenkova & Debeir, 2023). The distribution of aspect for Western Sarawak with a class range of  $45^\circ$  can be seen in Figure 3.4. Most of the past slope failures were identified to be located in areas with aspect range of  $90^\circ$  to  $135^\circ$  with 25% of total cases, followed by  $315^\circ$  to  $360^\circ$  with 16.67% of total cases,  $0^\circ$  to  $45^\circ$  and  $45^\circ$  to  $90^\circ$  both have the same number of cases with 14.58% of total cases,  $135^\circ$  to  $180^\circ$  with 10.42% of total cases, and  $180^\circ$  to  $225^\circ$ ,  $225^\circ$  to  $270^\circ$ ,  $270^\circ$  to  $315^\circ$  each contributed to 6.25% of total cases.

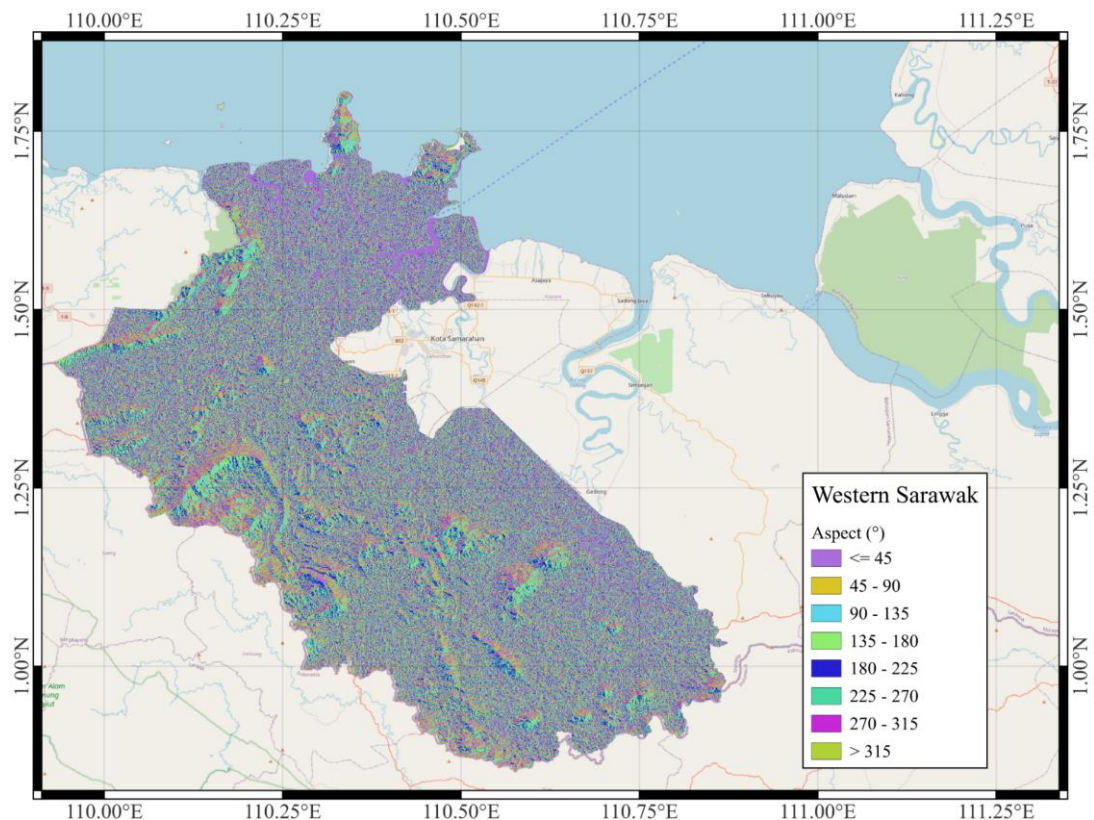


**Figure 3.4:** Aspect in Western Sarawak

ii. Curvature

The curvature raster of the study area was obtained as a derivative product of the elevation raster obtained through System for Automated Geoscientific Analyses (SAGA) with their package of terrain analysis – morphometry (Conrad et al., 2015). The curvature values in Western Sarawak ranges from -0.114 rad to 0.1498 rad as seen in Figure 3.5. The curvature goes from concave to convex, as the curvature values goes from the negative to positive (Babitha et al., 2022). In Western Sarawak, both convex, and concave curvature have the highest likelihood of slope failures in comparison to flat areas with 50% and 45.83% as compared to flat areas with 4.17%.

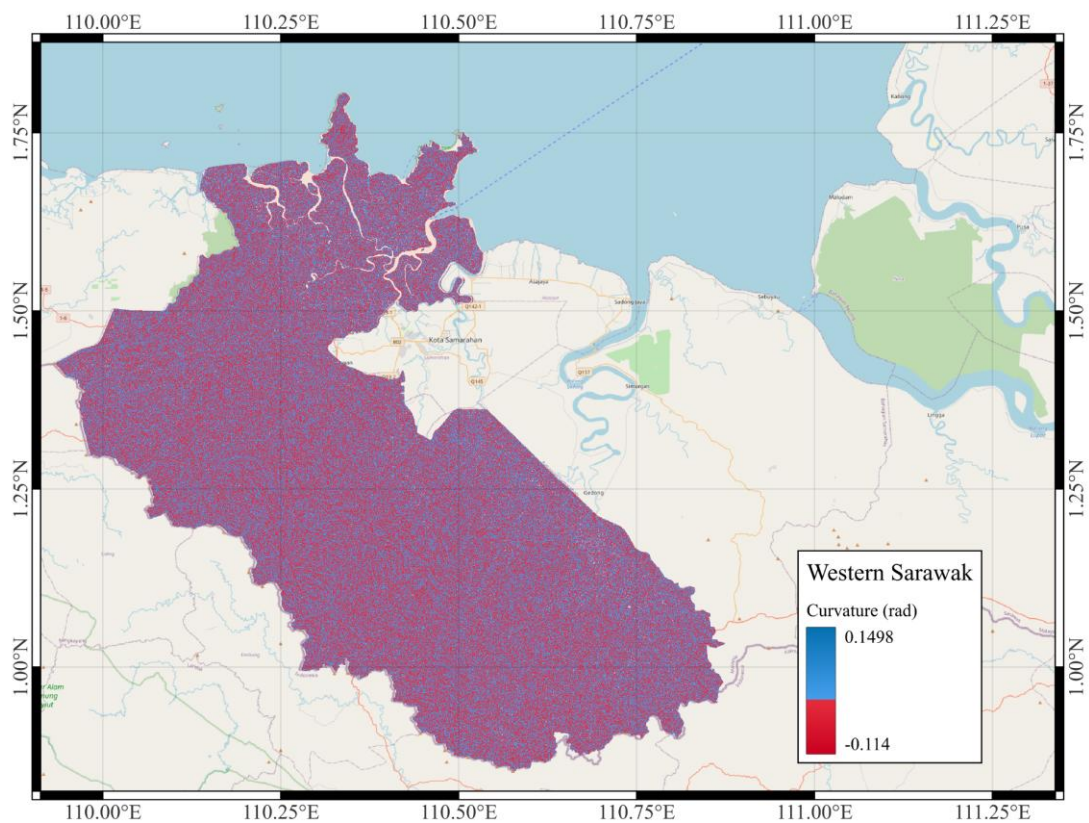


**Figure 3.5:** Curvature in Western Sarawak

### iii. Elevation

The elevation raster in this study was obtained from NASA through the Alaska State Facility at a resolution of 12.5m per pixel, where the raster was fetched by the Advanced Land Observing Satellite Phase Array Type L-band Synthetic Aperture Radar and latest freely available elevation raster provided by NASA (Laurencelle et al., 2015). Several other terrain related raster files were extracted from the elevation raster in this study, which were aspect, curvature, slope angle, and TWI (Naceur et al., 2022). Elevation in the study ranges from 1266m to -37m as seen in Figure 3.6, where the negative elevation indicates locations that are lower than the datum such as mean sea level (Chuma et al., 2023).

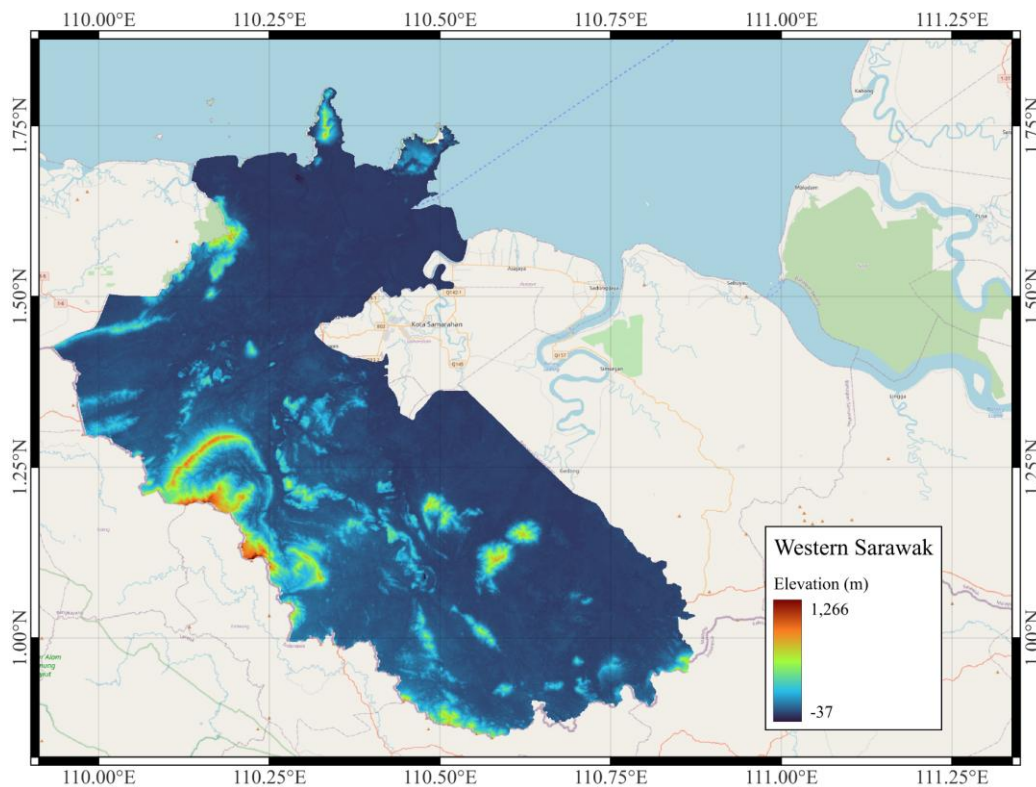


**Figure 3.6:** Elevation in Western Sarawak

#### iv. Soil Type

The soil type raster in this study was obtained from the Food and Agriculture Organisation (FAO) of the United Nations Educational, Scientific, and Cultural Organisation (UNESCO) which was a part of the organisation's soil map of the world mapping in 1979 (FAO-UNESCO, 1979). The original data was in the shapefile format, and due to the requirements of sampling and LSM plotting, it was rasterised using the built in rasterise tool in QGIS (QGIS Development Team, 2023). There were five distinct soil types in Western Sarawak as seen in Figure 3.7. It was determined that most of the slope failure cases occurred in soil type AO with 70.83%, GD with 16.67%, QA with 12.50%, and no cases in the others.

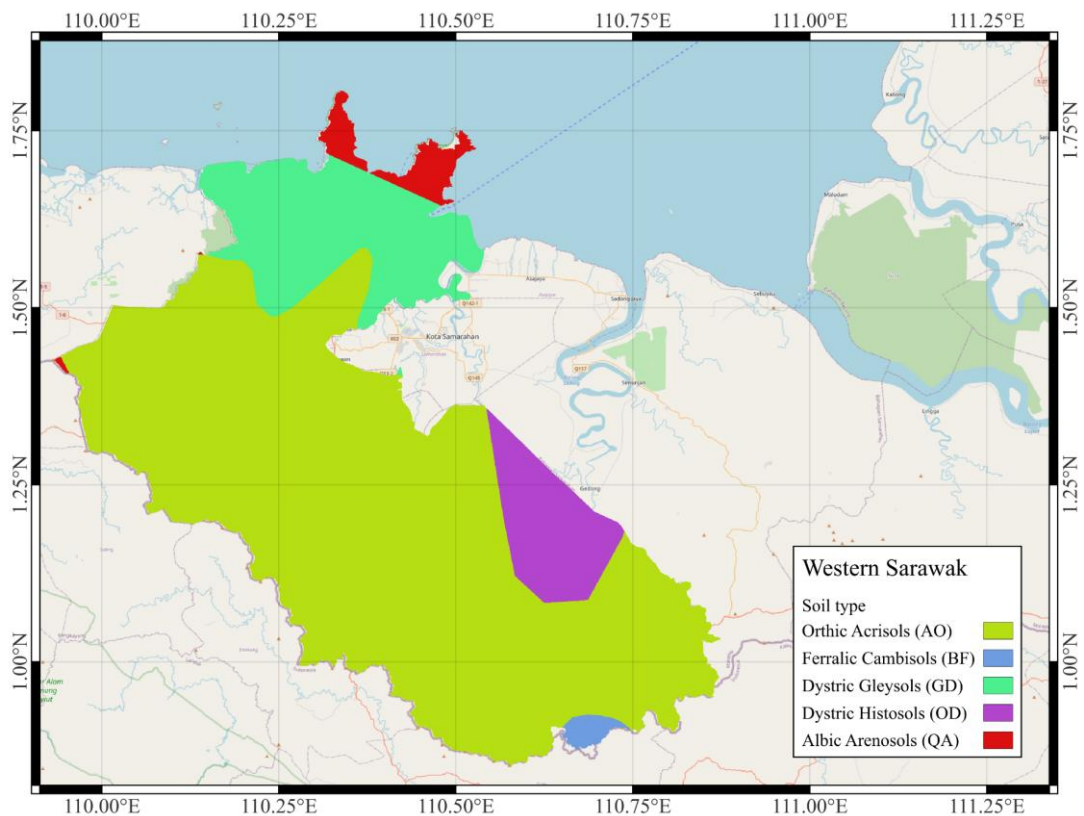
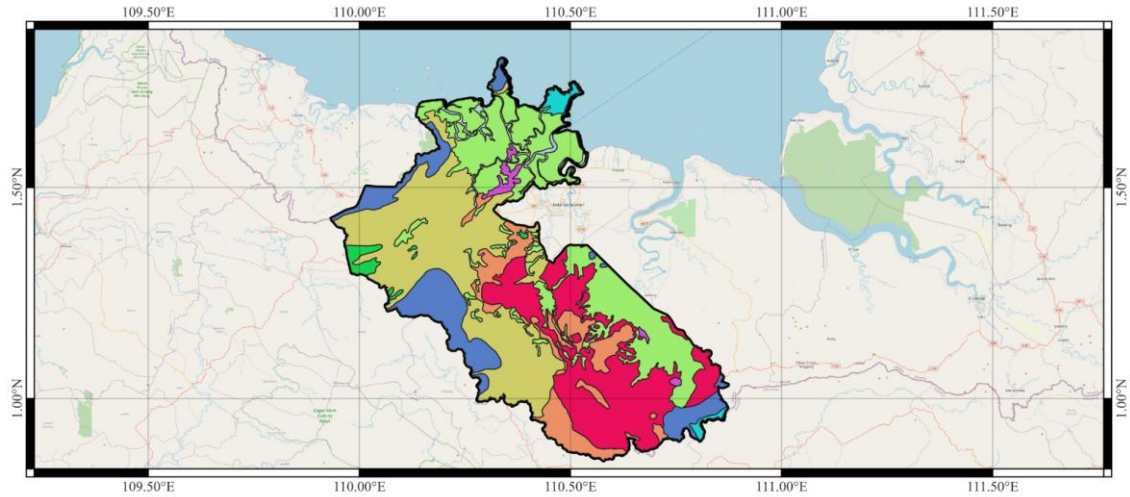


**Figure 3.7:** Soil type in Western Sarawak

v. Lithology

In this study the lithology raster was obtained from JMG. Initially, it was in the vector format, as the ANN model requires the input to be in the form of raster, it was converted through QGIS rasterization function (QGIS Development Team, 2023). Furthermore, in the raster format the value was converted from text to numerical values, to ensure the ANN model fully understand the data (Culhane, 2013). The lithology type for the study area consisted of clay, silt, and peat from the Pleistocene to recent represented by 1, Arenaceous, and Argillaceous rocks, coal and Calcareous beds from the Neogene period represented by 2, Argillaceous rocks with some Arenaceous, and Calcareous beds from the Palaeogene period represented by 3, Argillaceous rocks with some Arenaceous and Calcareous from the Jurassic to the Cretaceous period represented by 4, Sandstone, associated lava, pyroclastic, and Argillaceous rocks from the Triassic period represented by 5, Phyllite, Metagraywacke, and limestone with some schists and chert from the Upper Palaeozoic period represented by 6, Granodiorite, Diorite, Granite, and their fine-grained equivalents from the Triassic to the Late Tertiary period represented by 7, and Andesitic, Basaltic lava, and Pyroclastic from the Triassic to the Quaternary represented by 11. The resolution of the raster file was 12.5m per pixel. The lithology map can be seen in Figure 3.8. Slope failure cases in Western Sarawak were most prevalent in Arenaceous, and Argillaceous rocks, coal and Calcareous beds from the Neogene with 41.67%, followed by Argillaceous rocks with some Arenaceous and Calcareous from the Jurassic to the Cretaceous with 31.25%, and Andesitic, Basaltic lava, and Pyroclastic from the Triassic to the Quaternary with 10.42%.



### Legend

#### Lithology

- Andesitic and basaltic lava and pyroclastics. TRIASSIC - QUATERNARY
- Arenaceous and argillaceous rocks, coal and calcareous beds. NEOGENE
- Argillaceous rocks; some arenaceous and calcareous beds. PALAEOGENE
- Argillaceous rocks; some arenaceous and calcareous rocks. JURASSIC - CRETACEOUS
- Clay, silt, sand and peat. PLEISTOCENE AND RECENT
- Granodiorite, diorite, granite and their fine-grained equivalents. TRIASSIC - LATE TERTIARY
- Phyllite, metagraywacke and limestone: some schists and chert.- UPPER PALAEOZOIC
- Sandstone and associated lava and pyroclastics. Argillaceous rocks. TRIASSIC

**Figure 3.8:** Lithogy in Western Sarawak

#### vi. LULC

The LULC raster in this study was retrieved from the Environmental Systems Research Institute Incorporation (ESRI), where the maps were derived from the European Space Agency Sentinel-2 satellite imagery at a resolution of 10m per pixel, developed through billions of human-labelled pixels used to train a deep learning model for land classification which was determined to have a minimum accuracy of 75% (Karra et al., 2021). The LULC that were present in the study area consisted of water represented by 1, trees represented by 2, grass by 3, flooded vegetation represented by 4, crops represented by 5, scrub or shrub represented by 6, built area represented by 7, bare ground represented by 8, and clouds represented by 10, where the values were classed by the data providers (Shahabi et al., 2022). The LULC map can be seen in Figure 3.9. The LULC of trees has the

most slope failure with 83.33% from total cases, followed by built areas with 14.58%, grass with 2.08%, and no cases in others. This has indicated there is a shift from trees to built areas for slope failures in Western Sarawak. Thus, LULC was suitable to represent human influence on slope failures in the study area (Rudra & Sarkar, 2023). The LULC data was fetched from ESRI via Google's Earth Engine (see Appendix F).

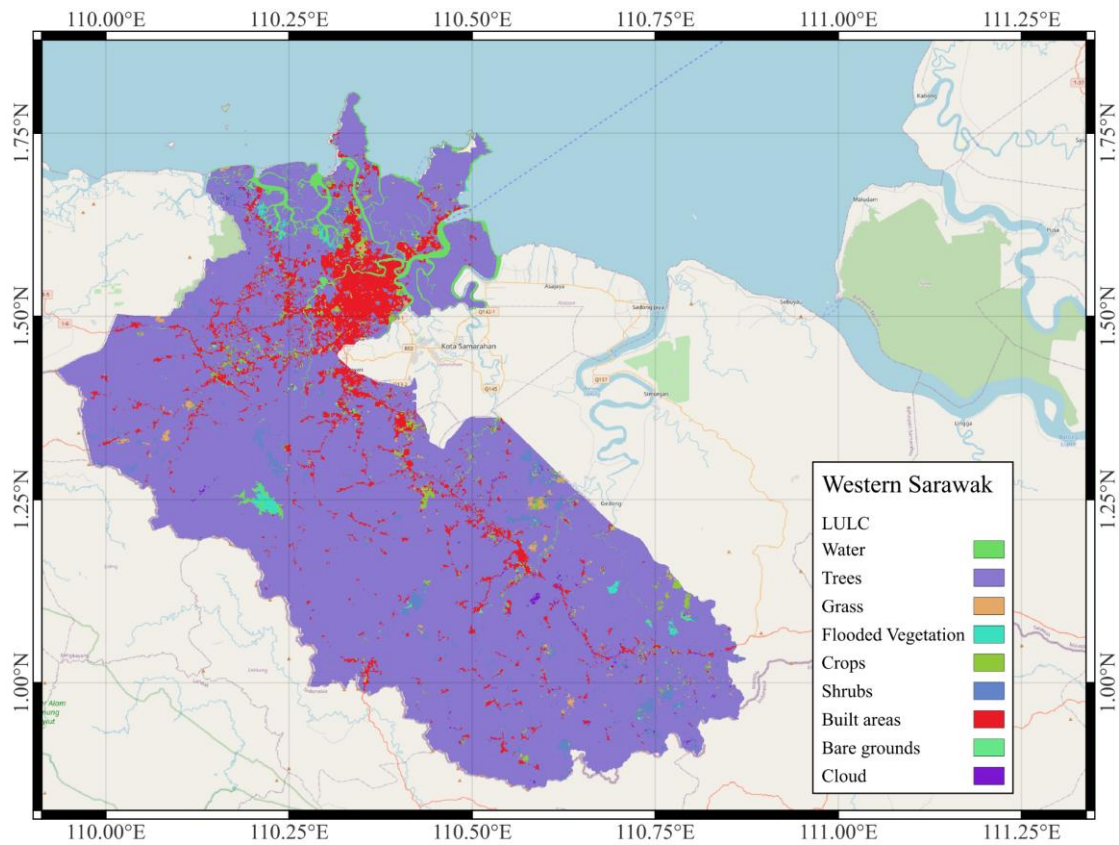


**Figure 3.9:** LULC in Western Sarawak

vii. Rainfall Intensity

The rainfall intensity raster was obtained from the Climate Hazards Group that extracted the data from their Climate Hazards Group InfraRed Precipitation with Station data programme that incorporates in-situ station data as well as a thirty plus years quasi-global rainfall dataset resulting the rainfall estimates (Funk et al., 2015). The rainfall intensity in

the study area was observed be in the range of 67.821mm per five days (pentad) to 115.362mm, where the rainfall intensity data was extracted from the rainy season of Western Sarawak where all cases have been reported, thus it was an average for the season (Rosly et al., 2022). Although the resolution of the raster file was low in comparison to other raster files in this study, it was acceptable as the data was rainfall, and rainfall typically covers a wide area (Shahabi et al., 2022). The rainfall intensity map can be seen in Figure 3.10. The rainfall intensity data was fetched from ESRI via Google's Earth Engine (see Appendix F).

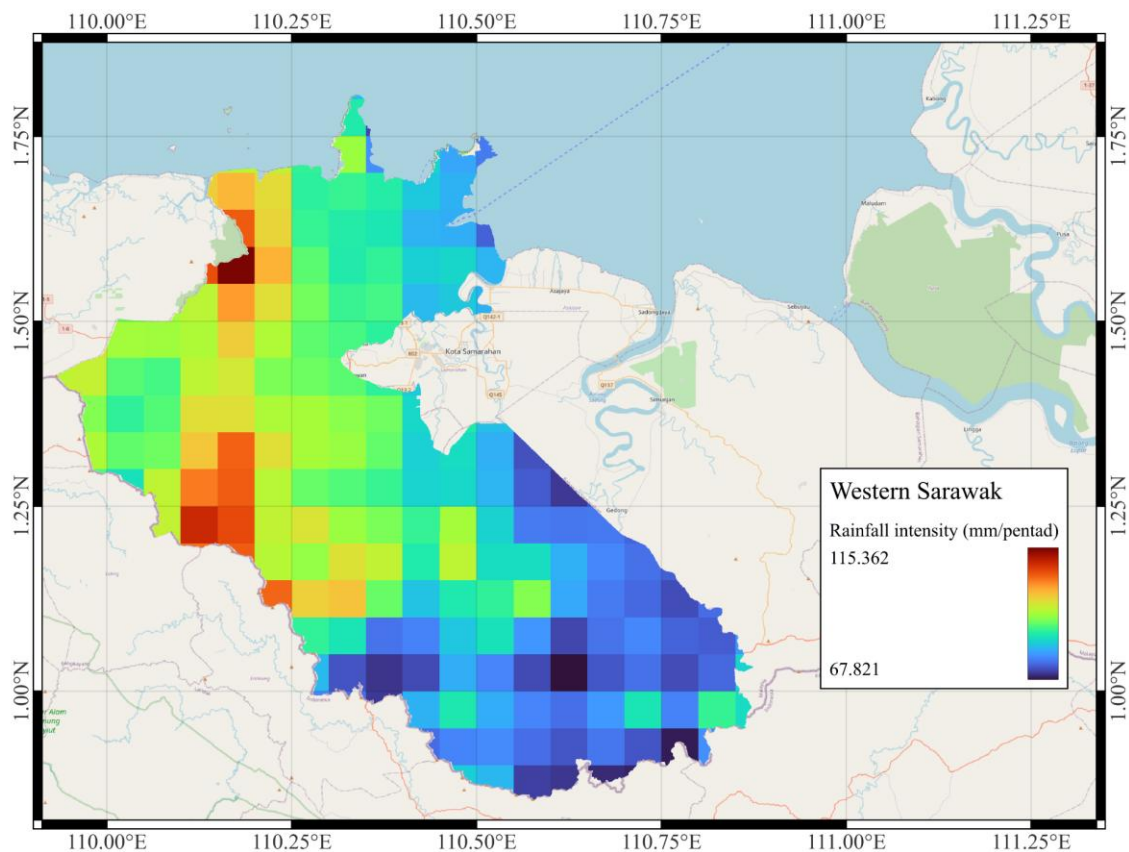


**Figure 3.10:** Rainfall Intensity in Western Sarawak

#### viii. Slope Angle

The slope angle raster in this study was obtained as a derivative product of the elevation raster through QGIS GDAL slope function, and it is known to be the most

important raster and data in terms of slope failure studies (Martin & Chai, 2022). As the slope angle raster was obtained from the elevation raster, with values ranging from  $0^{\circ}$  to  $76.1^{\circ}$  as seen in Figure 3.11 (Laurencelle et al., 2015).

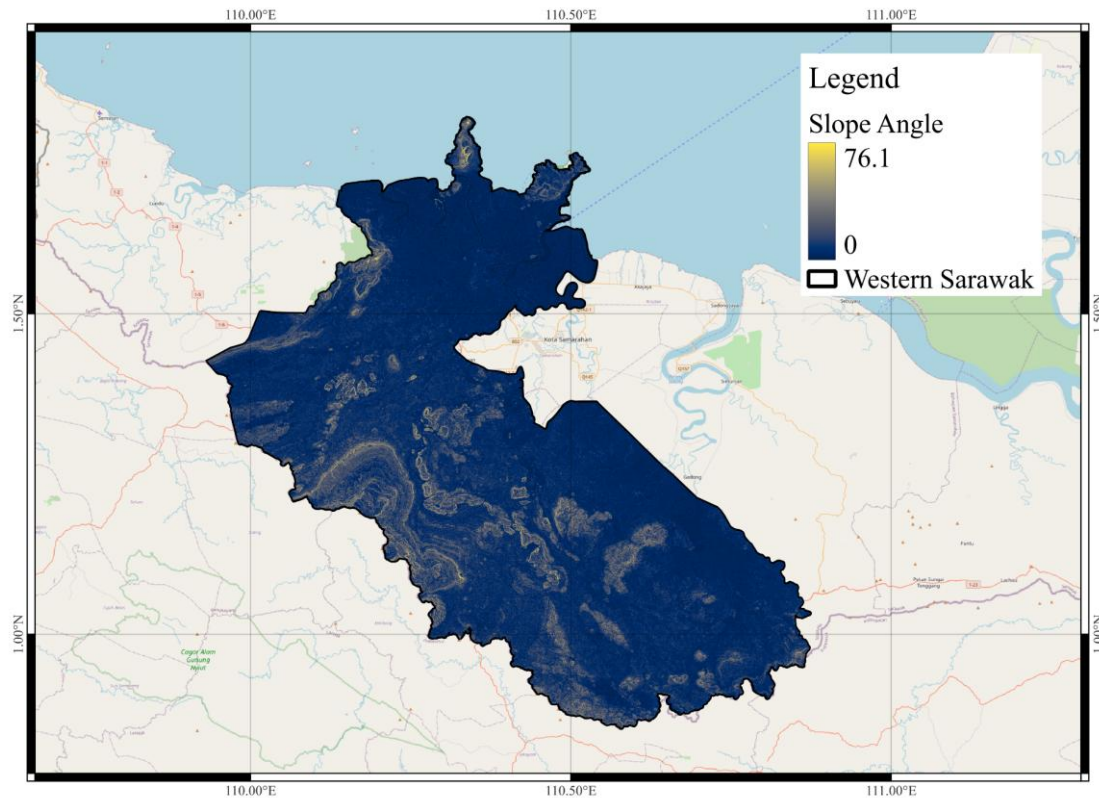


**Figure 3.11:** Slope Angle in Western Sarawak

#### ix. TWI

The TWI raster was obtained as a derivative product of the elevation raster obtained through the principle of Equation 2.1 by using the QGIS raster calculator (Naceur et al., 2022). The TWI for the study area was observed to be in the range of 1.41 to 18.43 as seen in Figure 3.12 (QGIS Development Team, 2023).

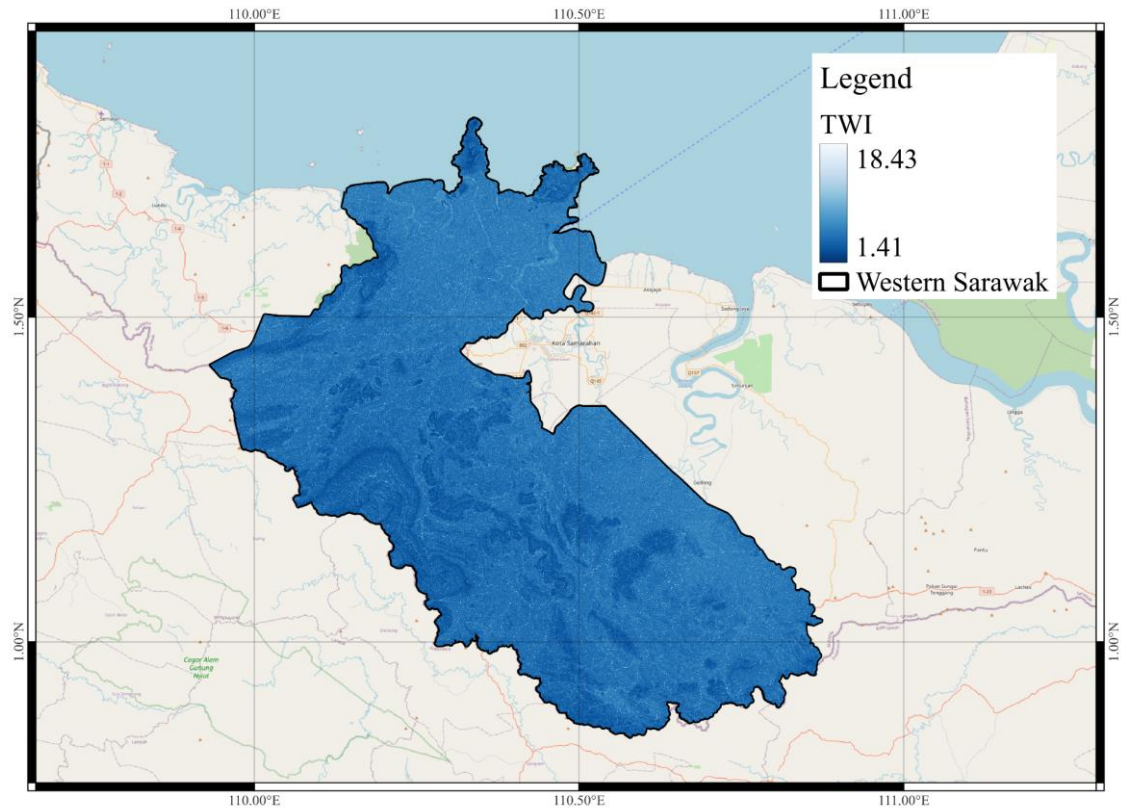


**Figure 3.12:** TWI in Western Sarawak

### 3.2.2 Data Extraction

To extract the data of slope failure factors at each slope failure inventory location efficiently, the slope failure factors raster files were stacked using the One Click Raster Stacking plugin in QGIS developed by Andreas Rabe where the plugin was developed as an extension of the GDAL raster stack tool, and in order to stack all of the raster files, a uniform coordinate reference system was required where in this study, the selected coordinate reference system was the European Petroleum Survey Group 3376, a projected coordinate reference system used specifically for East Malaysia be it onshore or offshore, where the unit was in m with the datum being the Geodetic Datum of Malaysia 2000 (JUPEM, 2009; Wang et al., 2021). The raster stacking process stacks the raster files in such a way that each respective raster files were represented by their individual band in a single raster file, and

the raster sampling tool was used to extract the values of each slope failure factors from the stacked raster at each slope failure inventory, where the raster values were represented by pixels (QGIS Development Team, 2023).

### **3.2.3 Multicollinearity Assessment**

In this study, the PCA method was used to determine collinearity amongst the input variables which were the slope failure factors, and high collinearity is not a favourable condition for the Machine Learning model, if there were variables that exhibited high collinearity, one of the variables were removed from the dataset (Djukem et al., 2020). Furthermore, it was also conducted to determine the preliminary view on the relationship of numerical input variables to the occurrence of slope failures in the study area (Sheng et al., 2023). A PCA score or coefficient lower than 0.8 was determined to be very favourable in the development of a training and testing dataset for Machine Learning models, however, in slope failure studies, with a small data pool, a PCA score lower than 0.9 is acceptable, where in this research the PCA score threshold was maxed at 0.8, if the score was larger than the threshold, removal of factors would be performed (Schober & Schwarte, 2018). Microsoft Excel's data analysis correlation plugin was used to conduct the PCA multicollinearity assessment.

### **3.2.4 Training Data and Testing Data**

Upon completion of the data extraction in QGIS, the data was exported into a comma separated values format known as the main dataset, which was then split into training data and testing data with a ratio of 80:20 after the PCA exhibits satisfactory results (Selamat et al., 2023). There are no explicit rule on the splitting ratio for the training data and testing data, however, researchers agree that the training data must be more than the testing data as

it will be used to develop the ANN model predictive capabilities in a supervised Machine Learning development (Horton et al., 2015). The testing data on the other was set aside to avoid “peeking” and it was only used after the training phase was completed (Tengtrairat et al., 2021). Before the training data and testing data can be used to develop and evaluate the ANN model, it must first undergo data preprocessing for the numerical data and the categorical data as to avoid any overpowering of a single variables due to the values, such as elevation values are large as it can goes up to the thousands, whilst slope angle values can only reach up to ninety (Swamynathan, 2017).

#### 3.2.4.1 Data Preprocessing

In this study, the numerical variables underwent data preprocessing through a data scaling method known as Min-Max scaling or min-max scaling as seen in Equation 2.3, whilst the categorical variables were categorised through one-hot encoding (Sharma, 2022).

##### i. Numerical Variables

There were several numerical variables in the training data and testing data, which were elevation, rainfall intensity, slope angle, and TWI, and the Min-Max scaling was conducted to scale variables value to be in the range of 0 to 1 (Sheng et al., 2023). The min and max values for each numerical variables can be seen in Table 3.1.

**Table 3.1:** Min – Max of Numerical Variables

| Variable           | Min              | Max               |
|--------------------|------------------|-------------------|
| Elevation          | -37 m            | 1266 m            |
| Rainfall Intensity | 67.821 mm/pentad | 115.362 mm/pentad |
| Slope Angle        | 0 °              | 76.1 °            |
| TWI                | 1.41             | 18.43             |

## ii. Categorical Variables

The categorical variables underwent categorisation through one-hot encoding, where if the data belonged to the pre-specified class, it was marked with an integer of 1 representing that it was present in the class, and if it was absent, it was marked with an integer of 0 (Ul Haq et al., 2019). There were several categorical data in the training data as well as the testing data which were aspect, curvature, lithology type, LULC, and soil type (Naceur et al., 2022). The following were the pre-specified classes for the classification of the categorical data.

### a. Aspect

The aspect data were categorised based on the compass notation with a class range of 45° starting from 0° North (Anis et al., 2019). The following Table 3.2 shows the classes for the aspect data where each class were assigned with a code from a to h for a simpler data handling process in the development stage.

**Table 3.2:** Aspect Data Categorisation

| Code | Range     | Compass Notation   |
|------|-----------|--------------------|
| a    | 0°-45°    | North to Northeast |
| b    | 45°-90°   | Northeast to East  |
| c    | 90°-135°  | East to Southeast  |
| d    | 135°-180° | Southeast to South |
| e    | 180°-225° | South to Southwest |
| f    | 225°-270° | Southwest to West  |
| g    | 270°-315° | West to Northwest  |
| h    | 315°-360° | Northwest to North |

### b. Curvature

The curvature values were categorised based on the shape whether the surface protrudes upwards indicating a convex curvature where in the raster file it was marked as

positive values, flat terrain with raster values of null or zero, and concave curvature where the surface concaves downwards with raster values in the negative as seen in Table 3.3 (Li et al., 2021).

**Table 3.3:** Curvature Data Categorisation

| Code    | Range               | Description              |
|---------|---------------------|--------------------------|
| Convex  | Positive values     | Surface protrude upwards |
| Flat    | Null or zero values | Flat terrain             |
| Concave | Negative values     | Surface concaves         |

c. Lithology Type

The lithology type data for the study area was obtained from JMG, where the organisation have predetermined numerical values to represent each type of lithology, and the lithology type was classified as such, with an addition of the L alphabet in front of the integer to denote the data as lithology (Najar et al., 2023). The classification of the lithology type can be seen in Table 3.4.

**Table 3.4:** Lithology Categorisation

| Code | Lithology                                                                         |
|------|-----------------------------------------------------------------------------------|
| L1   | Clay, silt, sand, and peat. (PLEISTOCENE AND RECENT)                              |
| L2   | Arenaceous and argillaceous rocks, coal, and calcareous beds. (NEOGENE)           |
| L3   | Argillaceous rocks; some arenaceous and calcareous beds. (PALAEOGENE)             |
| L4   | Argillaceous rocks; some arenaceous and calcareous rocks (JURASSIC - CRETACEOUS)  |
| L5   | Sandstone and associated lava and pyroclastic. Argillaceous rocks. (TRIASSIC)     |
| L6   | Phyllite, metagraywacke and limestone: some schists and chert. (UPPER PALAEOZOIC) |

**Table 3.5** Continued

|     |                                                                                                |
|-----|------------------------------------------------------------------------------------------------|
| L7  | Granodiorite, diorite, granite, and their fine-grained equivalents. (TRIASSIC - LATE TERTIARY) |
| L11 | Andesitic and basaltic lava and pyroclastic. (TRIASSIC - QUATERNARY)                           |

d. LULC

The LULC data was divided based on the land cover class determined by ESRI, where in the study area there are 9 LULC types present, which were water with class value of 1, trees with class value of 2, grass with class value of 3, flooded vegetation with class value of 4, crops with class value of 5, scrub or shrub with class value of 6, built area with class value of 7, bare ground with class value of 8, and clouds with class value of 10, and each class value was accompanied by the alphabet 'U' to denote that the data was LULC data as seen in Table 3.6 (Karra et al., 2021).

**Table 3.6:** LULC Categorisation

| Class Value | Code | LULC               |
|-------------|------|--------------------|
| 1           | U1   | Water              |
| 2           | U2   | Trees              |
| 3           | U3   | Grass              |
| 4           | U4   | Flooded Vegetation |
| 5           | U5   | Crops              |
| 6           | U6   | Scrub or Shrubs    |
| 7           | U7   | Built Areas        |
| 8           | U8   | Bare Ground        |
| 10          | U10  | Cloud              |

e. Soil Type

The soil type was categorised based on the predetermined class set by FAO, and each FAO assigned numerical code was accompanied by the alphabet ‘S’ as an indicator for soil followed by the assigned integer value as seen Table 3.7 (FAO-UNESCO, 1979).

**Table 3.7:** Soil Type Categorisation

| Code | FAO Code | Soil type          |
|------|----------|--------------------|
| S1   | AO       | Orthic Acrisols    |
| S2   | BF       | Ferralic Cambisols |
| S3   | GD       | Dystric Gleysols   |
| S4   | OD       | Dystric Histosols  |
| S6   | QA       | Albic Arenosols    |

**3.2.5 Remarks**

In conclusion, Phase I was a very important phase in this study as the overall accuracy of the subsequent ANN model was highly dependent on the data in this phase. In this phase, two sets of training data and testing data were prepared, which were used to develop the ANN model for the LSM and SRAS, with only static data for the SRAS.

**3.3 Phase II Development of ANN Models for Western Sarawak**

Developing the ANN models for the LSM and SRAS were the primary focus of Phase II as seen in Figure 3.13. The primary difference between the SRAS ANN model and the LSM ANN model were the remove of dynamic variables of rainfall intensity and LULC in the training data for the SRAS ANN model. The same steps were applied for the SRAS ANN model testing data. The hyperparameters of the ANN models were optimised based on the evaluation metrics that must be fulfilled, which were an AUC score at least 0.8, a recall

score at least 0.8, and RMSE of not more than 0.1. Post development of each respective ANN models, the LSM was developed in Phase III, and the SRAS in Phase IV.

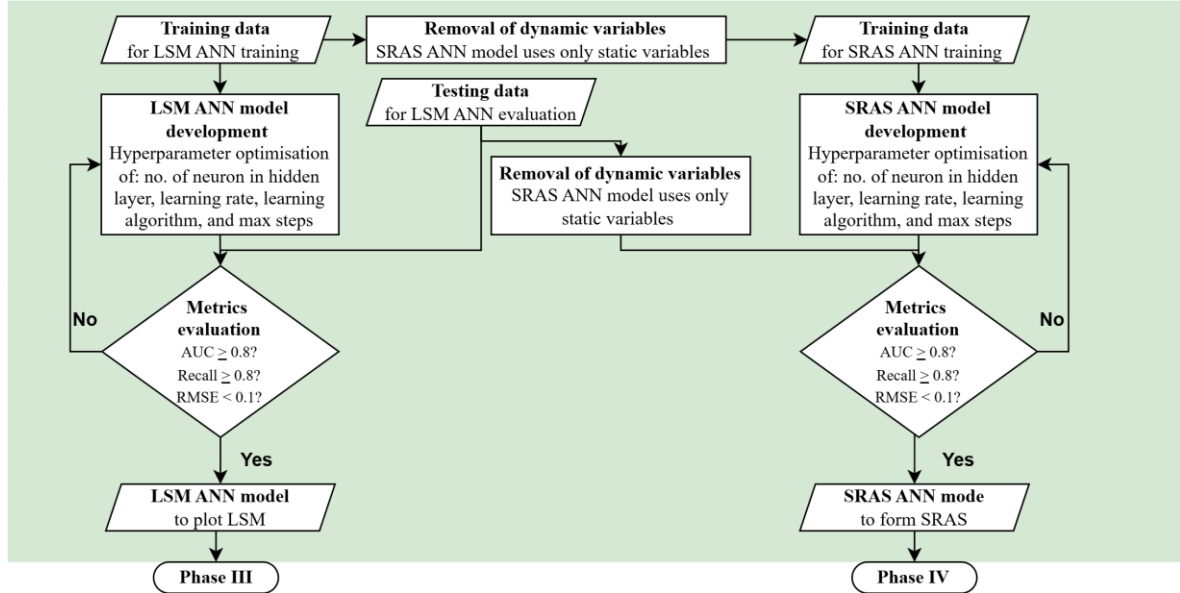


**Figure 3.13:** Flowchart of Phase II

### 3.3.1 ANN Model Development

R is a freely available, and open-source general purpose statistical software package used in numerous fields of research including GIS related ones, making the software a great choice for this research, and the programming session to develop the ANN model was done through the official integrated development environment known as RStudio (Horton et al., 2015). To develop the ANN models, a package specifically designed to develop an ANN model in R known as “neuralnet” was used, as other package such as “nnet” was found to have certain limitations such as restriction in the complexity of the neural network (Adnan et al., 2023). The main components that were tweaked to increase the accuracy of the ANN models were the hyperparameters as listed below (Fristch et al., 2022):

i. Formula

Formula in an ANN model is known to be a symbolic description of the ANN model, where given a set of input, will result in an output as seen in Equation 3.1.

$$Output = \sum Input \quad \text{Equation 3.1}$$

Where, Output is the ‘string’ of the output, and Input is the ‘string’ of the input.

As the ANN model for the SRAS was developed with fewer variables, the formula was shorter in comparison the LSM’s ANN model as seen in Table 3.8.

**Table 3.8:** Formula for Respective ANN Model

| ANN Model | Formula                                                                                                                |
|-----------|------------------------------------------------------------------------------------------------------------------------|
| SRAS      | Slope Failure ~ Elevation + Slope Angle + TWI + Aspect + Curvature + Lithology + Soil Type                             |
| LSM       | Slope Failure ~ Elevation + Rainfall Intensity + LULC + Slope Angle + TWI + Aspect + Curvature + Lithology + Soil Type |

ii. Data

Data in the hyperparameter refers to the data frame that contains the variables specified in the formula, as it will be used to train the ANN model, the data was the training data.

iii. Hidden

Hidden in the package refers to the number of hidden layers that the ANN model requires, which was represented by an integer or if it were to have several columns, it will be formatted with a comma to separate the hidden layers. In this study, the ANN model for

the LSM was developed with eight neurons, whilst the ANN model for the SRAS was developed with four.

iv. Learning Rate

Learning rates refer to size of step taken in each iteration of the weight adjustment in a neural network. Both ANN models were developed with a learning rate of 0.001, where the values was small enough to ensure that the process converges smoothly, but not too small to hinder the learning process (Baig et al., 2022).

v. Algorithm

Algorithm refers to the type of learning algorithm that will be used the ANN model for the weight adjustment (Tehrani et al., 2022). In this study, backpropagation was selected as the learning algorithm for both LSM's ANN model, and SRAS's ANN model.

vi. Error Function

Error function refers to the calculation method used to determine the error in the training process of the ANN model, by looking into the difference between the actual and predicted values (Youssef & Pourghasemi, 2021). The error function of RMSE was chosen for both ANN model.

vii. Linear Output

Linear output refers to the expected outcome of the ANN model, as both ANN model was intended to have outcome that ranges from 0 to 1, instead of logical outcome of 0 or 1, it was set as false, as the ANN model was intended to have value ranging from 0 to 1 for the slope failure susceptibility score.

viii. Stepmax

Stepmax or maximum steps refers to the number of iterations that will take place in the ANN model training phase. Both ANN models had a stepmax of  $1e+10$ .

ix. Activation Function

Activation function refers to the function that will be used to smooth out the results in each summation junction of the ANN model, where in this study, both ANN model was developed with the sigmoid activation function (Ma et al., 2020).

### **3.3.2 AUC, Recall, and RMSE**

AUC, recall, and RMSE were important evaluation metrics in this study. The ANN model training success rate and prediction success rate using testing data were evaluated through said metrics (Khajehzadeh et al., 2022). Knowing both training and testing performance has helped in determining the overall generalisation capabilities of the ANN models. In this study, a minimum AUC and Recall of 0.8, and a maximum RMSE of 0.1 were adopted. This was to ensure the ANN model produces the most accurate prediction in general, with focus on the minority output of slope failures, and minimising the model's prediction error in units, all based on AUC, recall, and RMSE respectively (Senanayake et al., 2022).

### **3.3.3 Remarks**

Phase II was a very important phase in this research as the prediction function for the LSM, and the SRAS was developed in this phase, where if and only if the RMSE score was lower than or at most 0.1 with an AUC and recall score higher than or at least 0.8, the subsequent phase can be commenced.

### 3.4 Phase III LSM Development and Verification for Western Sarawak

The LSM was developed in Phase III as it required the ANN model to be conducted. Phase III and Phase IV can be done simultaneously, however, as the LSM development was a part of the code for the ANN model, it was given the priority for time saving reasons. Phase III starts with the development of the LSM, and followed by on-site verification, which was done by conducting site visits to several location in each district available within the study area. The flowchart can be seen in Figure 3.14.

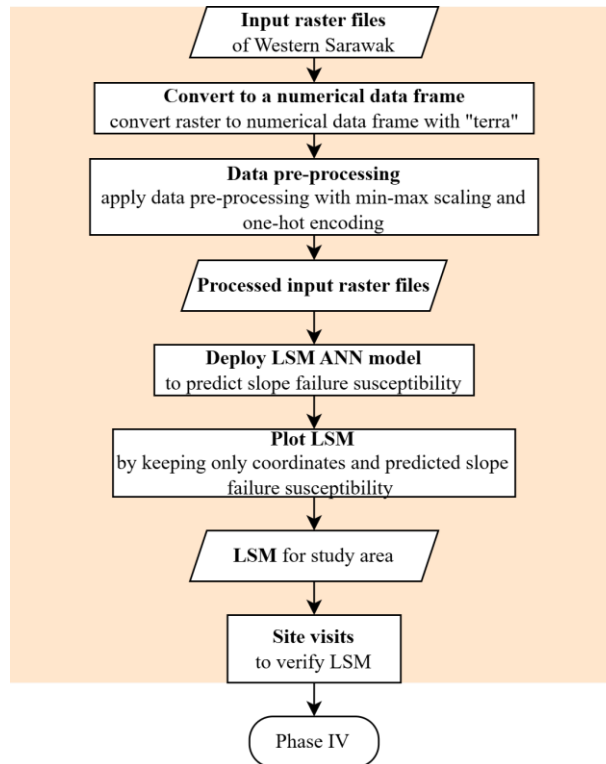


**Figure 3.14:** Flowchart of Phase III

#### 3.4.1 LSM Development

To develop the LSM, the following procedures were conducted:

i. Raster Stacking

A package known as ‘terra’ was installed in the ANN model code script for the LSM, which was a package that enables R to handle spatial data in the raster format (Bivand et al., 2024). All of the raster files of the variable present in the LSM’s ANN model were imported into RStudio with the same coordinate reference system, and the extent of each raster file was checked, and if the extent was determined to be not uniform, the raster file was resized through resampling by the size of the smallest raster (Shahabi et al., 2022). Once the raster file was uniformed, the raster stacking commenced.

ii. Converting Stacked Raster Files to a Data frame

The ANN model could not understand the data if it were not converted into a numerical data frame, as it was not trained with the raster format (Loukrakpam & Oinam, 2021). Thus, the stacked raster was converted into a numerical data frame, where each coordinate in the raster had the respective slope failure factor values, similar to the raster sampling method, however, in this stage, it covered the whole study area without the presence of slope failure inventory in the final data frame (Loukrakpam & Oinam, 2021).

iii. Data Preprocessing

The same data preprocessing technique applied to the training data and testing data was again applied to the data frame, as the ANN model required variables that were present in the formula used in the development phase. The numerical variables underwent Min-Max scaling, and the categorical variables were one hot encoded with the categories that were specified in the ANN model’s data preprocessing phase (Khajehzadeh et al., 2022).

iv. Run ANN Model

Similar to the usage of the testing data, the processed data frame underwent slope failure susceptibility prediction through the ANN model predictive properties (Luo et al., 2021).

v. Conversion to LSM

After the ANN model had successfully predicted the slope failure susceptibility at each coordinate point in the data frame, the input variables were removed, leaving only the coordinates, and slope failure susceptibility to plot the LSM (Bravo-López et al., 2023).

### **3.4.2 Verification**

The verification was the final stage in Phase III. It was conducted to determine whether the slope failure susceptibility score in the LSM was justified, by applying a qualitative approach of visiting several areas in each district present in the study area (Huat & Jamaludin, 2005). There were three districts visited, which were Bau, Kuching (divided into landwards and seawards region), and Serian.

### **3.4.3 Remarks**

In conclusion, the main goal of this phase is to represent the ANN model prediction on slope failure susceptibility of Western Sarawak in a graphical manner, which was the LSM.

## **3.5 Phase IV Development and Verification for Western Sarawak SRAS**

Phase IV was the final phase in this study where it intended to develop a new SAS system based on the ANN model inner workings known as SRAS as seen in Figure 3.15. The same process in the ANN modelling mentioned in Phase II for SRAS was applied to

predict the slope failure susceptibility. To use the SRAS, data preprocessing is required, and the calculation was based on the assigned weightages, biases, and activation function (Shapiro et al., 1999).

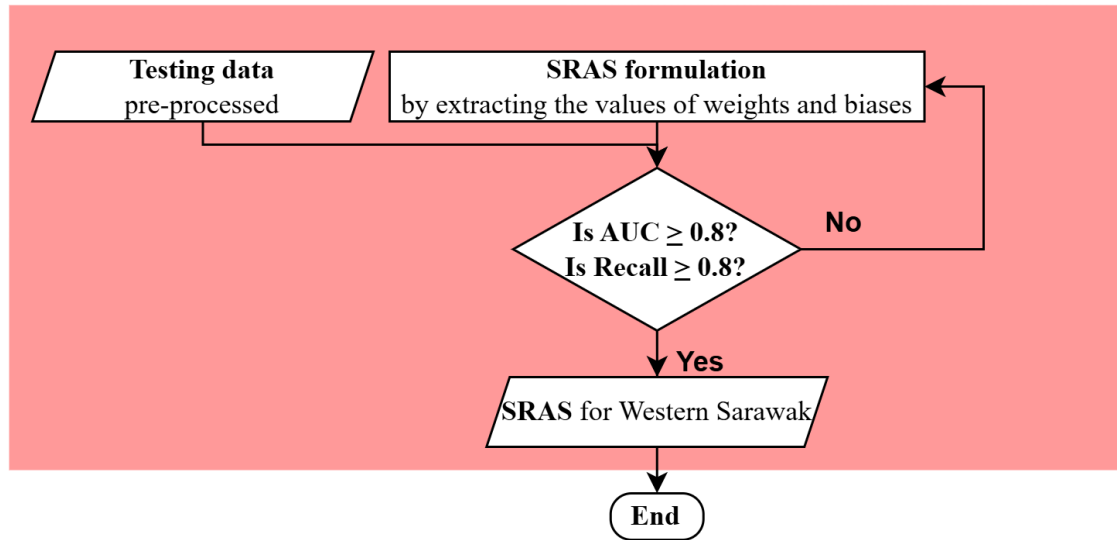


**Figure 3.15:** Flowchart of Phase IV

### 3.5.1 Data Analysis and Extraction

The data analyses and extractions were processed and used to extract the weightage of each variable, and biases for SRAS's ANN model as it would be used to develop the SRAS. This process was done by exporting the relevant information in a numerical data frame to be tested. The Olden algorithm was used as a simple method to provide an overview of the ANN model behaviour before extracting the actual weightages from nodes to nodes (Marcus & Beck, 2022).

### 3.5.2 SRAS Development

#### 3.5.2.1 Data Preprocessing

As the used to train the ANN model underwent preprocessing, the same parameters for the data scaling of numerical variables, and one-hot encoding class were applied (Abidoye et al., 2018).

##### i. Numerical Variable Min-Max scaling

The simplified data preprocessing equation for elevation can be seen in Equation 3.2, slope angle simplified data preprocessing equation can be seen in Equation 3.3, and simplified data preprocessing equation for TWI can be seen in Equation 3.4.

$$El_{norm} = \frac{El + 37}{1229} \quad \text{Equation 3.2}$$

Where  $El_{norm}$  refers to the normalised elevation, and  $El$  refers to the elevation value in m.

$$\Theta_{norm} = \frac{\Theta}{76.1} \quad \text{Equation 3.3}$$

Where  $\Theta_{norm}$  refers to normalised slope angle, and  $\Theta$  refers to slope angle value in °.

$$TWI_{norm} = \frac{TWI - 1.41}{17.02} \quad \text{Equation 3.4}$$

Where  $TWI_{norm}$  refers to normalised TWI, and  $TWI$  refers to the TWI value.

Minimum and maximum values for elevation, slope angle, and TWI which were incorporated into Equation 3.2, Equation 3.3, and Equation 3.4, can be referred from Table 3.1.

ii. Categorical Data

The categorisation of categorical data follows the same one-hot encoded categorisation for the ANN model.

3.5.2.2 SRAS Calculation

The calculation for the SRAS in this study was designed from the calculation method applied the ANN model to obtain the final output. For the SRAS ANN model, there were five output that was calculated, each one representing a neuron (excluding the neuron in the input layer). The simplified equation at each neuron for the four neurons in the hidden layer, and the output neuron is as seen in Equation 3.5. The equation was derived from the weightage of variables, biases, and the sigmoid activation function of the ANN model (Lee et al., 2020). Unlike the neuron in the hidden layer where the slope failure factors were the input, the output neuron inputs were the value obtained at the four neuron in the hidden layer (Shapiro et al., 1999).

$$f(x) = \frac{1}{1 + e^{-(B + \sum(X_n * W_n))}} \quad \text{Equation 3.5}$$

Where  $f(x)$  refers to smoothed output at each summation junction,  $B$  refers to bias assigned by the ANN at each summation junction,  $X_n$  refers to pre-processed variable value, and  $W_n$  refers to weightage assigned by the ANN for each input variable at each summation junction.

3.5.3 Accuracy Assessment

The accuracy assessment of the SRAS was compared to the predicted values of the ANN model based on the testing data through AUC, and recall (Hakim et al., 2022). This was done to determine the degree of success for the conversion of the ANN model into

equations. RMSE was not involved as it was used to determine the prediction error based on the training data for the ANN model training phase.

#### **3.5.4 Remarks**

Phase IV was conducted to determine whether an ANN model can be transformed into a reliable SRAS if it were used as a replacement of the pre-existing SAS in Malaysia. If this phase was conducted successfully, the research was determined to be finished.

### **3.6 Validity of Remote Sensing data and Comparison of Performance for Machine Learning models and WoE and FOA in Slope Failure Prediction**

To validate the usability of remote sensing data in comparison to traditional soil physical properties variables, a separate study was conducted. The prediction performance of ANN model was compared to the ASEAN recommended and SAS development approach for slope failure prediction at a river basin level, which were WoE and FOA respectively.

A study was conducted in collaboration with Professor Hyuck-Jin Park from Sejong University. The previous study by Park et al., 2017 where the performance of Machine Learning models namely Fuzzy and probabilistic analysis developed with variables of soil type, slope angle, elevation, soil friction angle, soil cohesion, unit weight hydraulic conductivity, and soil thickness as variables were compared to an ANN model developed with remote sensing data of aspect, curvature, elevation, hydraulic conductivity, LULC, NDVI, and slope angle. The performance comparison were done through recall and AUC.

Another study was conducted to compare ANN to WoE which was recommended by ASEAN and FOA, which was the method used to develop the previous SAS to predict slope failure susceptibility at the Sarawak River basin level.

## CHAPTER 4

### RESULTS AND DISCUSSION

#### 4.1 Overview

This chapter has focused on discussing the results from each analysis that has been done throughout the study. The analysis consists of multicollinearity assessment from Phase I, LSM ANN model and SRAS ANN model metrics evaluation alongside the hyperparameter from Phase II, the resulting LSM of Western Sarawak as predicted by the ANN model from Phase III, the new formulated SRAS and its metrics evaluation from Phase IV, and the extras that consisted of remote sensing data comparison to soil physical properties as input variables for Machine Learning models in slope failure prediction, and comparison of ANN model performance to WoE and FOA for slope failure prediction at a river basin scale.

#### 4.2 Multicollinearity Assessment – Phase I

Multicollinearity amongst input variables is not a favourable condition in any Machine Learning model development process. High correlation amongst two input variables happens when the input variables form a linear relationship between each other. This will cause problems to the model as it overfits the impact of two input variables. Through PCA, the highly correlated input variables exhibits correlation coefficient at least or more than 0.8 (Kalantar et al., 2019). If and only if the correlation coefficients amongst two input variables exceed 0.8, one of it were to be removed. Furthermore, the multicollinearity assessment can provide a preliminary overview of the influence that each input variable has on the output without the consideration of the other input variables. The multicollinearity assessment through PCA in this is visualised through a matrix as seen in Table 4.1.

**Table 4.1:** PCA Matrix

|                  | <i>aspect</i> | <i>curvature</i> | <i>elevation</i> | <i>lithology</i> | <i>lulc</i> | <i>rain</i> | <i>slope</i> | <i>soil</i> | <i>twi</i> | <i>fail</i> |
|------------------|---------------|------------------|------------------|------------------|-------------|-------------|--------------|-------------|------------|-------------|
| <b>aspect</b>    | 1.00          |                  |                  |                  |             |             |              |             |            |             |
| <b>curvature</b> | -0.05         | 1.00             |                  |                  |             |             |              |             |            |             |
| <b>elevation</b> | 0.10          | 0.02             | 1.00             |                  |             |             |              |             |            |             |
| <b>lithology</b> | 0.11          | -0.07            | 0.05             | 1.00             |             |             |              |             |            |             |
| <b>lulc</b>      | 0.02          | 0.00             | -0.15            | -0.19            | 1.00        |             |              |             |            |             |
| <b>rain</b>      | 0.08          | -0.03            | 0.05             | 0.17             | 0.33        | 1.00        |              |             |            |             |
| <b>slope</b>     | 0.35          | 0.03             | 0.70             | 0.06             | 0.08        | 0.16        | 1.00         |             |            |             |
| <b>soil</b>      | -0.13         | 0.04             | -0.04            | -0.27            | 0.12        | 0.15        | -0.11        | 1.00        |            |             |
| <b>twi</b>       | -0.14         | -0.15            | -0.24            | -0.01            | 0.02        | 0.02        | -0.33        | 0.09        | 1.00       |             |
| <b>fail</b>      | 0.15          | 0.00             | 0.59             | 0.05             | 0.07        | 0.16        | 0.72         | 0.17        | 0.26       | 1.00        |

Note: rain is rainfall intensity, slope is slope angle, soil is soil type, and fail is slope failure

Solely based on the PCA matrix, it was identified that slope angle and elevation correlated highly with slope failure occurrence with a correlation coefficient of 0.72 and 0.59 respectively. On the other hand, TWI presents the most negatively correlated input variable to slope failure with a coefficient of -0.26. This was justified as slope angle increases, the capability of the slope to hold surface runoff decreases, reducing the risk of slope failure. Positively correlated variables show that the tendency of slope failure increases with the input variables, and vice versa (Iban & Sekertekin, 2022). However, this was just an overview of data, as the decision making process of the ANN model was dependent on the hyperparameters (Loukrakpam & Oinam, 2021).

Amongst the input variables, it was identified that the highest positively correlated input variables were slope angle and elevation, with a PCA correlation coefficient of 0.70. This indicates that slope angle increases with elevation based on the primary data for Western Sarawak. As the highest positive correlation coefficient did not exceed 0.8, none of the two input variables were removed. Analysing the information provided by the respective

raster files of slope angle and elevation in Western Sarawak through the “terra” package in R has revealed that there is indeed a positive correlation between the two variables overall as seen in Figure 4.1. It was important to note that the coefficient of determination ( $R^2$ ) was obtained by evaluating the overall raster values and not on the primary data where only areas with slope failure cases and areas with slope angle at a maximum of  $5^\circ$  were sampled.

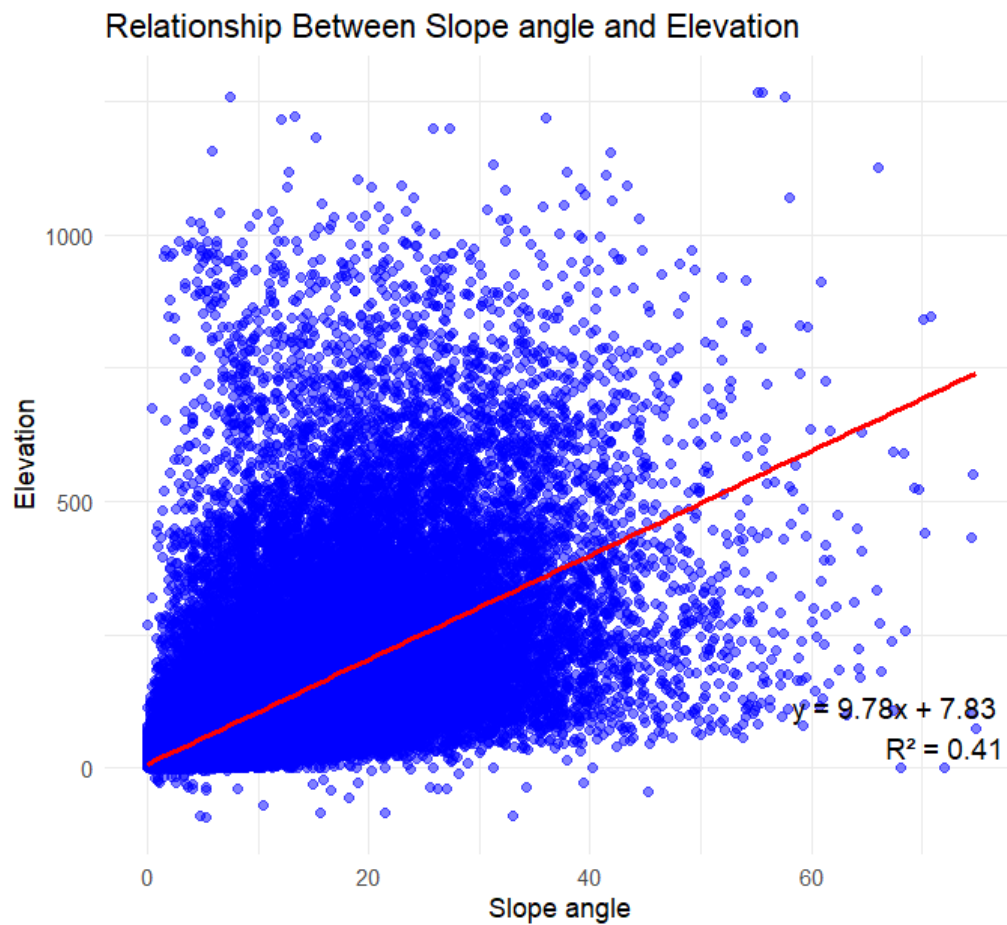


**Figure 4.1:** Scatterplot of Elevation vs Slope Angle

Although the other seemingly does not have high influence on the occurrences of slope failures, it was crucial for the variables to be included as it provides the ANN model with a clearer description on the required combination of input variables required for slope failure to occur (Saikh & Mondal, 2023). The input variables which have shown the low correlation to the output consisted mostly of categorical variable which were important to the ANN model decision making process.

A previous study in Malaysia for Langat River Basin did not evaluate the correlation of the categorical input variables to the occurrence of slope failure, however it was still included as the ANN model performed well with said variables (Selamat et al., 2022b). This was proven by observing the frequency of slope failures in the categorical variables. Based on the aspect orientation, it was determined that slope failures mostly occurred in aspect of 90° to 135° with 25% of the total slope failures, followed by aspect of 315° to 360° with 16.67% as seen in Table 4.2.

**Table 4.2:** Slope Failures based on Aspect Orientation

| Aspect      | Percentage |
|-------------|------------|
| 0° - 45°    | 14.58      |
| 45° - 90°   | 14.58      |
| 90° - 135°  | 25.00      |
| 135° - 180° | 10.42      |
| 180° - 225° | 6.25       |
| 225° - 270° | 6.25       |
| 270° - 315° | 6.25       |
| 315° - 360° | 16.67      |

In terms of curvature, flat areas had the lowest likelihood of slope failure in Western Sarawak, with 4.17%, followed by concave curvature at 45.83%, and convex curvature with half of the total slope failures as seen in Table 4.3.

**Table 4.3:** Slope Failures based on Curvature

| Curvature | Percentage |
|-----------|------------|
| Concave   | 45.83      |
| Flat      | 4.17       |
| Convex    | 50.00      |

Slope failure cases in Western Sarawak with regards to lithology types in the form of percentages can be seen in Table 4.4. Granodiorite, diorite, granite, and their fine-grained equivalents from Triassic to Late Tertiary period is the strongest lithology type present in Western Sarawak (Chigira et al., 2011). Its resistance to slope failure was proven with the lowest percentage slope failure occurrence in the region with only 2.08%. Albeit being the weakest lithology type, clay, silt, sand, and peat from the period Pleistocene and recent only accounted to 4.17% of the total recorded cases of slope failure in Western Sarawak, the second lowest percentage alongside Phyllite, metagraywacke and limestone with some schists and chert from the Upper Palaeozoic period. Here, it was important to note that most of clay, silt, sand, and peat were deposited in low-lying areas, that are generally free from slope failure cases. Furthermore, peat slope failure is more prevalent in countries that experience winter, as it is moved by the thawing process of ice (Mills & Rushton, 2024). In terms of lithology type, most slope failures in Western Sarawak had occurred in Argillaceous rocks, some arenaceous and calcareous beds (Palaeogene) with 41.67%, followed by Argillaceous rocks; some arenaceous and calcareous rocks (Jurassic - Cretaceous) at 31.25%. The percentage of slope failures in the lithology types seems to be justified based on the period of composition, as the Jurassic – Cretaceous was the earlier, giving enough time to compact the lithology, resulting in a greater resistance to slope failure (Trisnawati et al., 2022). Judging solely through the strength of each lithology type present in the study area could not fully justify the susceptibility towards slope failure based on the percentage

of slope failure cases. This have shown that it was important to consider the slope failure occurrences were influenced by a variety of variables.

**Table 4.4:** Slope Failures based on Lithology types

| <b>Lithology</b>                                                                               | <b>Percentage</b> |
|------------------------------------------------------------------------------------------------|-------------------|
| Andesitic and basaltic lava and pyroclastic. (TRIASSIC - QUATERNARY)                           | 10.42             |
| Argillaceous rocks; some arenaceous and calcareous beds. (PALAEOGENE)                          | 41.67             |
| Argillaceous rocks; some arenaceous and calcareous rocks (JURASSIC - CRETACEOUS)               | 31.25             |
| Clay, silt, sand, and peat. (PLEISTOCENE AND RECENT)                                           | 4.17              |
| Granodiorite, diorite, granite, and their fine-grained equivalents. (TRIASSIC - LATE TERTIARY) | 2.08              |
| Phyllite, metagraywacke and limestone: some schists and chert. (UPPER PALAEOZOIC)              | 4.17              |
| Sandstone and associated lava and pyroclastic. Argillaceous rocks. (TRIASSIC)                  | 6.25              |

Historical slope failure occurrences based on the LULC as seen in Table 4.5 have revealed that most of the slope failures had occurred in areas with LULC of trees with 83.33%, followed by built areas with 14.58%, and grass with 2.08%. Other LULCs in Western Sarawak did not experienced slope failure at the time of this study. The type of LULCs with slope failure cases revealed that most of it were possibly to had occurred in undisturbed slopes. Thus, as of the time of this study, it can be concluded that most of the recorded slope failure cases to have occur naturally. However, the amount of built areas exposure to slope failures is expected to increase if no mitigation efforts are implemented (Dai et al., 2021).

**Table 4.5:** Slope Failures based on LULC

| <b>LULC</b> | <b>Percentage</b> |
|-------------|-------------------|
| Built Areas | 14.58             |
| Grass       | 2.08              |
| Trees       | 83.33             |
| Other       | 0.00              |

Only three soil types have been confirmed to have experienced slope failures in the past at the time of study, which were Orthic Acrisols, Dystric Gleysols, and Albic Arenosols, with 70.83%, 16.67%, and 12.50% respectively as seen in Table 4.6. Albeit Dystric Histosols, and Dystric Gleysols being the two weakest soil types in Western Sarawak, no prior slope failure cases have been reported in its area, potentially due to deposition of the soil types in low-lying areas, and Albic Arenosols albeit being the strongest soil type, has experienced slope failures in the past.

**Table 4.6:** Slope Failures based on soil types

| <b>Soil Type</b> | <b>Percentage</b> |
|------------------|-------------------|
| Orthic Acrisols  | 70.83             |
| Dystric Gleysols | 16.67             |
| Albic Arenosols  | 12.50             |
| Others           | 0.00              |

In conclusion, the multicollinearity assessment was crucial for the sake of the ANN model learning phase. Judging by the percentage of slope failures that had occurred in the past is not to develop the system, as slope failures occurrences are influenced by a plethora of variables, where some may require certain conditions of other variables for slope failures to occur.

### **4.3 ANN Models Development and Evaluation – Phase**

There was two ANN models developed in this study for the purpose of developing an LSM of Western Sarawak, and the SRAS. The fundamental difference of the ANN models was the exclusion of dynamic slope failures variables for the SRAS ANN model. This was done to ensure that the SRAS can be applied directly on-site when required, without the need to wait for the dynamic variable to occur. The performance of the ANN models was evaluated through AUC, recall, and RMSE.

#### **4.3.1 LSM ANN Model Evaluation**

The LSM ANN model was developed with the slope failure factors of aspect, curvature, elevation, soil type, lithology type, LULC, slope angle, rainfall intensity, and TWI. RMSE was the initial measurements for the ANN model hyperparameter tuning process, as it determines the potential error in the form of deviation from the actual predicted values. The goal was to minimise the RMSE while obtaining an AUC and recall a minimum of 0.8. The threshold for RMSE for the training data was set at 0.1, with lower values being preferred (Hodson, 2022). The ANN model's hyperparameter was tuned for five time with the number of neuron the hidden layer, learning rate, learning algorithm, and stepmax being adjusted as seen in Table 4.7. The fifth iteration of the hyperparameters has yielded the lowest RMSE amongst the trials with 0.0057. Thus, the final hyperparameters configuration of eight neurons in a single hidden layer, a learning rate of 0.001, a backpropagation learning algorithm, and a stepmax of 1.00E+08, were selected as the final hyperparameter for the LSM's ANN model. Here it was deemed that eight neurons were suitable for the complexity of the problem, the learning rate of 0.001 was a suitable size for the weight adjustments process. The resilient backpropagation did not performed as well as the traditional backpropagation learning algorithm likely due to the fact that resilient backpropagation

adjusts the weights based solely on the direction of the gradient adjustment, ignoring the magnitude of the variables (Fristch et al., 2022). The traditional backpropagation learning algorithm allows for a more thorough understanding of importance for each slope failure factors.

**Table 4.7:** LSM's ANN Model RMSE

| Trial | Hyperparameter        |               |                           |          |        |
|-------|-----------------------|---------------|---------------------------|----------|--------|
|       | Hidden Layer's Neuron | Learning Rate | Algorithm                 | Stepmax  | RMSE   |
| 1     | 8                     | 0.01          | Backpropagation           | 1.00E+10 | 0.0116 |
| 2     | 8                     | 0.01          | Resilient Backpropagation | 1.00E+10 | 0.093  |
| 3     | 6                     | 0.001         | Backpropagation           | 1.00E+10 | 0.0279 |
| 4     | 18                    | 0.001         | Backpropagation           | 1.00E+08 | 0.0285 |
| 5     | 8                     | 0.001         | Backpropagation           | 1.00E+08 | 0.0057 |

The training phase of the ANN model was evaluated to determine whether the training process was a success or otherwise, and the testing phase was evaluated to determine the performance of the ANN model based on its capabilities to predict slope failure based on the never-before-seen testing data. The resulting performance matrix for each phase can be seen in Table 4.8. Due to the “true” instances being the minority, and it was crucial for the ANN model to be able to accurately predict slope failure, recall was chosen as one of the evaluation metric. For the training phase, the ANN model was able to classify all instances accordingly, with all 38 slope failure cases predicted as slope failures, resulting in a recall score of 1. On the other hand, when deploying the ANN model on the testing data, it was

able to classify all non-slope failures accordingly, whereas only one of out of ten slope failures was classified as non-slope failure, resulting in a recall of 0.9.

**Table 4.8:** LSM ANN Model Performance Matrix

| Actual\Prediction | Training |      | Testing |      |
|-------------------|----------|------|---------|------|
|                   | False    | True | False   | True |
| 0                 | 928      | 0    | 230     | 0    |
| 1                 | 0        | 38   | 1       | 9    |

As recall was a measure that focuses on the minority of slope failure cases, AUC provides an overall view of the ANN model performance as seen in Figure 4.2. In both training and prediction, the ANN model has yielded high AUC scores of 1, and 0.99.

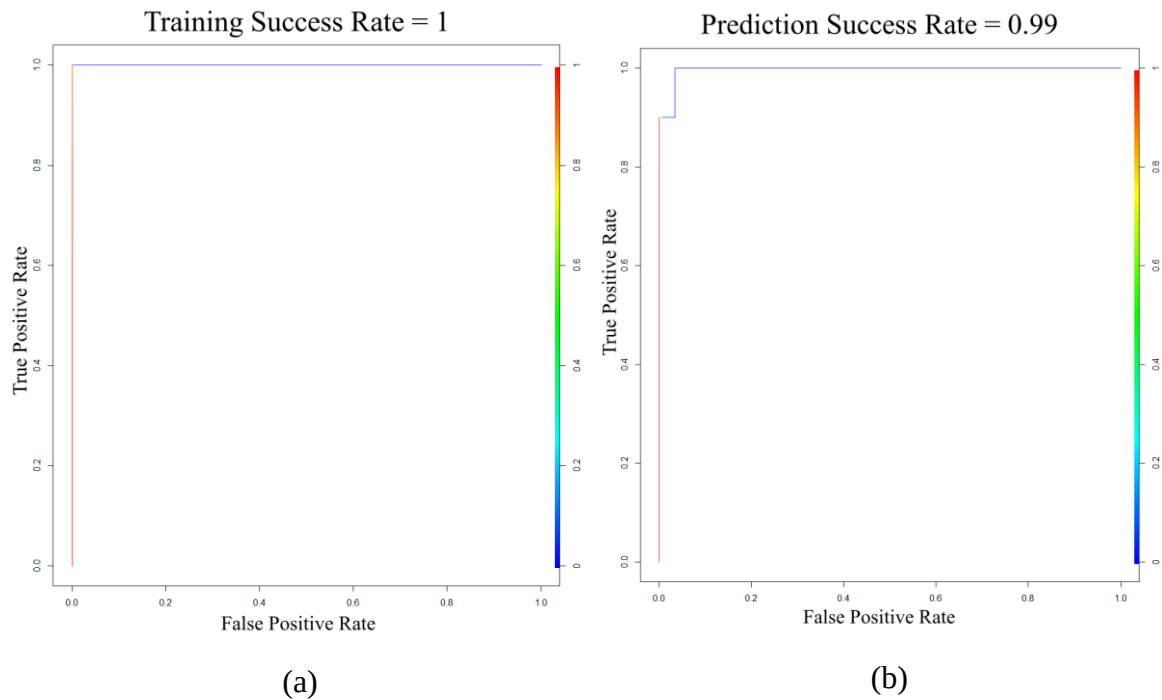


**Figure 4.2:** AUCs of LSM's ANN Model (a) in Training Phase, and (b) Testing Phase

The high AUC score of 1 for the ANN model in the training phase as shown in Figure 4.2(a) has shown that the ANN model training was a success. As for the prediction success rate AUC of 0.99 as seen in Figure 4.2(b), it indicates that although the prediction was not perfect, it was able to predict slope failures accordingly with high confidence. Thus, based on the RMSE of 0.0057, recall of 0.9 for the testing phase, and an AUC of 0.99 for the prediction success rate, the LSM ANN model was deemed highly accurate, and suitable for deployment to develop the LSM for Western Sarawak.

#### **4.3.2 SRAS ANN Model Evaluation**

Unlike the LSM's ANN model, the SRAS's ANN model was developed with fewer slope failure factors to facilitate its usage directly on the site. This limitation has caused the SRAS's ANN model to be developed with fewer variables, which were aspect, curvature, elevation, lithology, slope angle, soil type, and TWI. As rainfall intensity was removed from the developmental data for the SRAS's ANN model, it can be seen a measure of susceptibility prior to heavy rainfall. Thus, it would help to prepare all areas that are highly susceptible in the event of future heavy rainfall. The hyperparameter optimisation of the SRAS's ANN model was straightforward due to the availability of the LSM's ANN model. A single hyperparameter optimisation trial was done by initialising the final layer of the LSM's ANN model seed resulting the SRAS ANN model achieving a similar RMSE of 0.0057 as seen in Table 4.9. Due to the less complex nature of the data used in developing the SRAS's ANN model the number of neuron in the single hidden layer was adjusted to achieve the RMSE. The final hyperparameter of the SRAS's ANN model were four neuron in a single hidden layer, a backpropagation learning algorithm, a learning rate of 0.001, and a maximum step of 1.00E+8.

**Table 4.9:** SRAS's ANN Model RMSE

| Trial | Hyperparameter        |               |                 |          |        |
|-------|-----------------------|---------------|-----------------|----------|--------|
|       | Hidden Layer's Neuron | Learning Rate | Algorithm       | Stepmax  | RMSE   |
| 1     | 4                     | 0.001         | Backpropagation | 1.00E+08 | 0.0057 |

The prediction results for the training data and testing data can be seen in Table 4.10. It was determined that in both phase, the ANN model was able to accurate predict all instances, resulting in a recall of 1 for both training and testing phase.

**Table 4.10:** SRAS ANN Model Performance Matrix

| Actual\Prediction | Training |      | Testing |      |
|-------------------|----------|------|---------|------|
|                   | False    | True | False   | True |
| 0                 | 928      | 0    | 230     | 0    |
| 1                 | 0        | 38   | 0       | 10   |

The highly accurate prediction has also contributed to the perfect AUC score for both training success rate and prediction success rate as seen in Figure 4.3(a), and Figure 4.3(b) respectively. The increase in performance for the SRAS's ANN model was attributed to the removal of dynamic variables namely the rainfall intensity and LULC. However, this does not indicate that the LSM's ANN model is any lesser in comparison to the SRAS's ANN model. The LSM's ANN model provide a more comprehensive view on the landslide susceptibility due to the higher number of variables involved. Instead, the SRAS's ANN model completes the LSM's ANN model as the LSM later on is limited to the spatiotemporal properties of the raster files, which was to be overcome by the SRAS (Bambang et al., 2023).

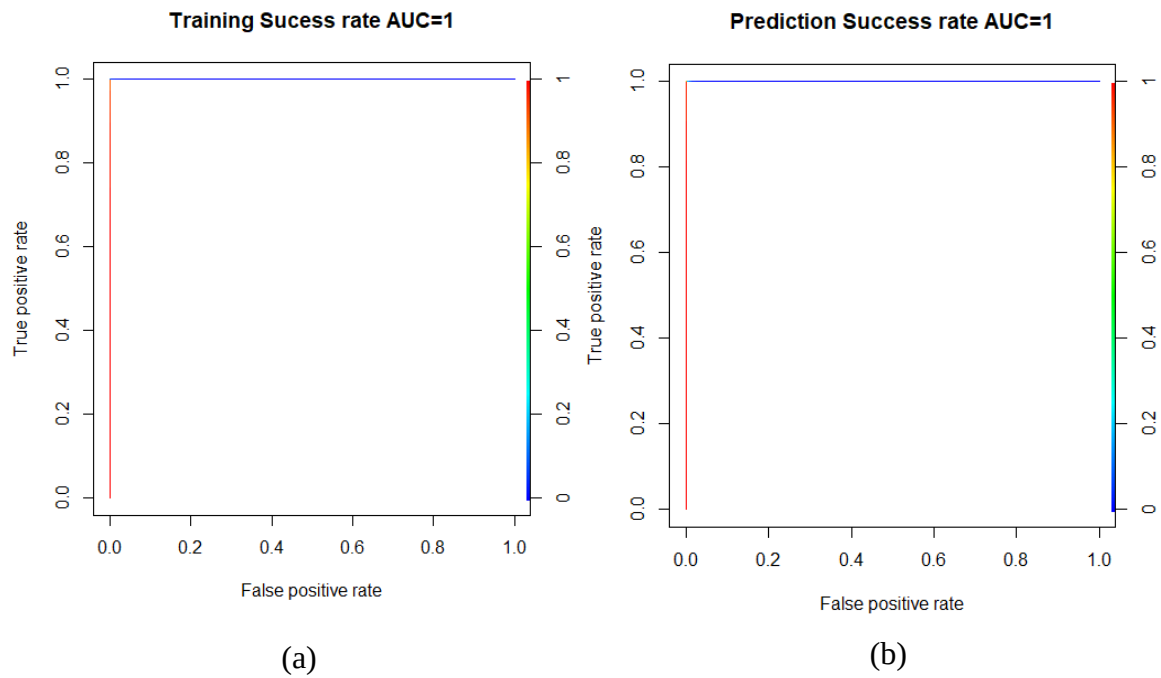


**Figure 4.3:** SRAS’s ANN Model AUC (a) Training Success Rate, and (b) Prediction Success Rate

In conclusion, the SRAS’s ANN model was deemed to be highly accurate and suitable to be extracted for the SRAS development.

#### 4.4 LSM of Western Sarawak – Phase III

To develop the LSM, the each raster files for the input variables of slope failure factors were transformed into a numerical data frame through the “terra” package in R (Bivand et al., 2024). The numerical data frame then underwent the same data-preprocessing steps as the training data and testing data. Post processing, the ANN model was deployed on the numerical data frame to predict the slope failure susceptibility at each points in Western Sarawak. Keeping only the latitude, longitude, and the slope failure susceptibility level, the LSM was plotted, where it can be seen in Figure 4.4. The susceptibility value ranges from 0 to 1, from least to most severe. The slope failure susceptibility of Western Sarawak was

determined to have a mean of 0.131, indicating that most of the areas in the region are not susceptible to slope failure (QGIS Development Team, 2023).

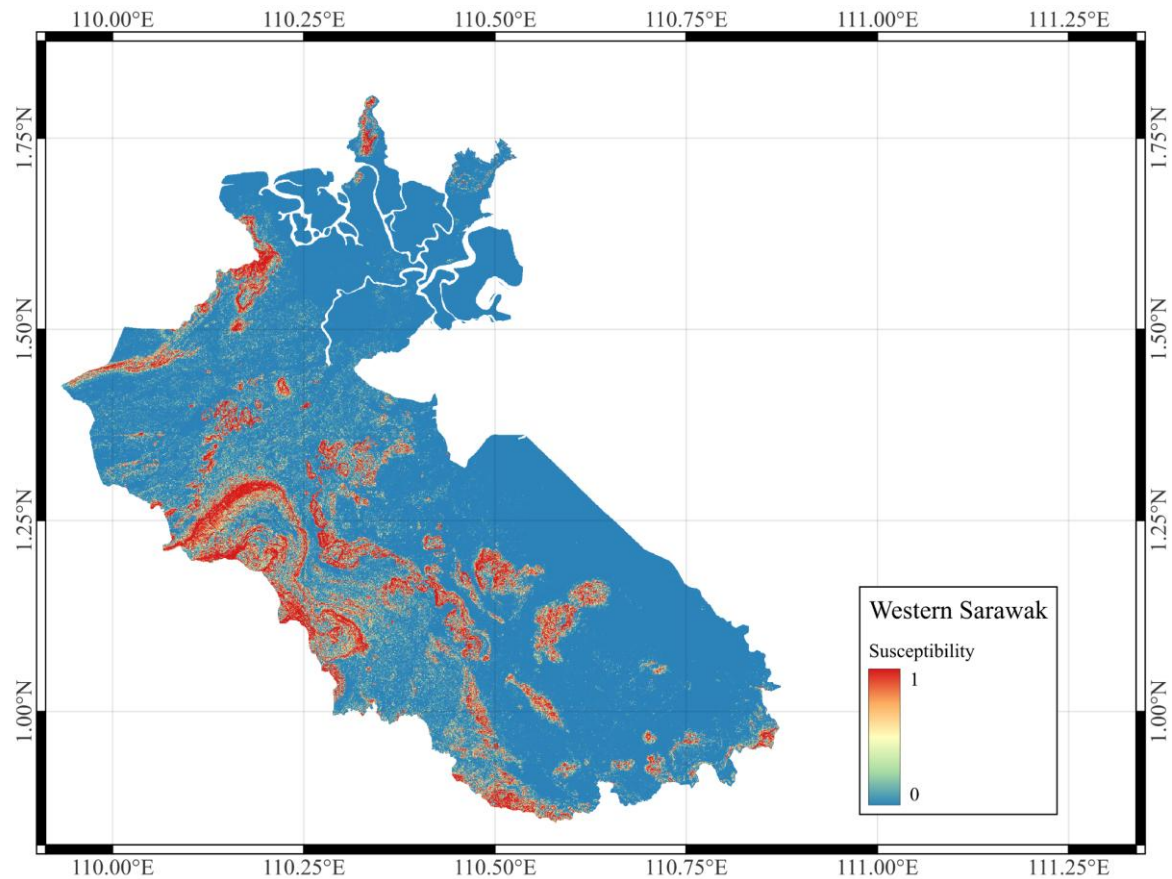


**Figure 4.4:** LSM of Western Sarawak

Evaluating the slope failure susceptibility in a zonal manner, it was determined that Kuching (landward) has the highest mean slope failure susceptibility level with 0.235, followed by Bau, Serian, and Kuching (seaward) with a mean slope failure susceptibility of 0.183, 0.098, and 0.048 respectively. Amongst the sub-districts, Kuching (seaward) is the most populated area, and home to the capital city of the state. Thus, it can be concluded that most of the general population in Western Sarawak are located outside areas with high risks of slope failure.

**Table 4.11:** Mean Slope Failure Susceptibility based on Zones

| Area               | Mean Susceptibility Level |
|--------------------|---------------------------|
| Kuching (Landward) | 0.235                     |
| Bau                | 0.183                     |
| Serian             | 0.098                     |
| Kuching (Seaward)  | 0.048                     |

By removing the built areas from the LULC vector in QGIS, the percentage of areas left for development was able to be determined. This was done by converting the raster files into a numerical data frame through the “terra” package (Bivand et al., 2024). The values were then counted by using R basic function that crawls through the data (Horton et al., 2015). It was revealed that outside of the current built areas in Western Sarawak, there is an estimated 82.51% area coverage that are known to be safe with susceptibility levels from 0.0 to 0.2 as seen in Table 4.12. On the other hand, the current area coverage of highly susceptible areas is 9.06%. However, these odds are subject to change, as climate change patterns shifts, and development are being pushed further landward (Raj et al., 2023).

**Table 4.12:** Area Coverage Percentage based on Slope Failure Susceptibility Levels

| Susceptibility Levels | Area Coverage (%) |
|-----------------------|-------------------|
| 0.0 to 0.2            | 82.51             |
| 0.2 to 0.4            | 3.48              |
| 0.4 to 0.6            | 2.42              |
| 0.6 to 0.8            | 2.54              |
| 0.8 to 1.0            | 9.06              |

Several site visits were conducted to further validate the accuracy of the LSM. Most of the high-risk areas in the LSM are located around the mountainous regions, thus the focus of the site visits was to determine the validity of the LSM in areas that are not in the mountainous region or around it as seen in Figure 4.5. More information can be found in Appendix B, C, and D.

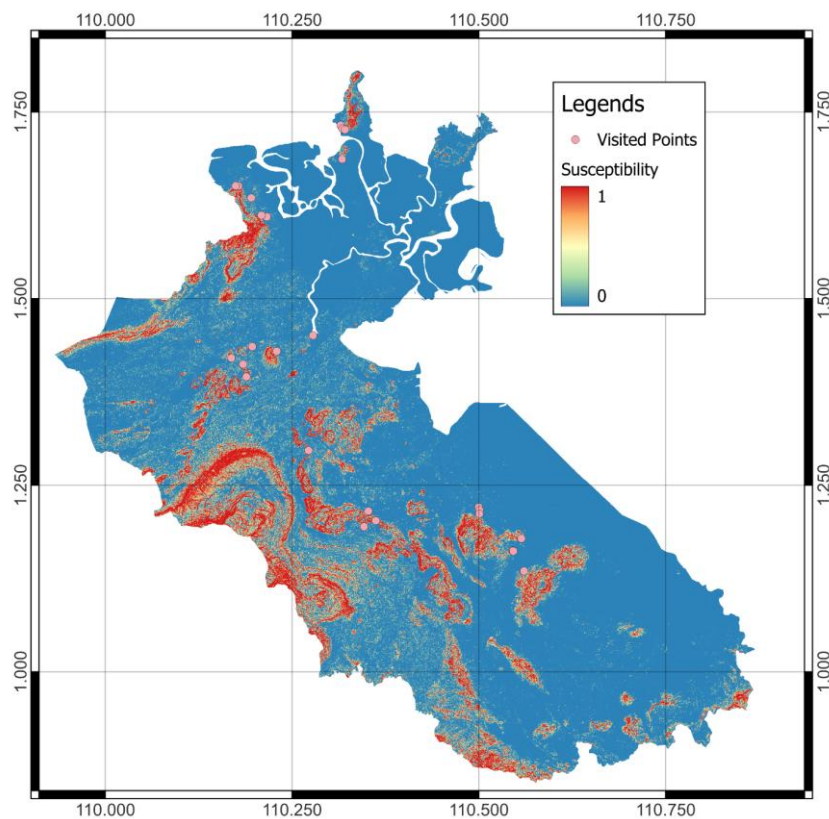


**Figure 4.5:** Visited Points for LSM Verification

Throughout the site visits, several recent slope failures were identified in areas that were near human settlements. One of the slope failures occurred in Kampung Tangga Mawang, Serian, albeit at a smaller scale. The properties of the variables for the failed points can be seen in Table 4.13. Although the values of the variables seemed to be average, the

ANN model was able to predict the slope failure accordingly as the failure point referenced through the LSM revealed that it had a susceptibility level of 0.81.

**Table 4.13:** Tangga Mawang failure point properties

| Variable       | Values  |
|----------------|---------|
| Aspect         | 355.01° |
| Curvature      | -0.012  |
| Elevation      | 66 m    |
| Lithology type | L11     |
| Soil type      | S1      |
| Slope angle    | 19.351° |
| TWI            | 3.787   |

The slope failure in Tangga Mawang can be seen in Figure 4.6, where a few days prior, the area received heavy rainfall, and the site visit was conducted during the rainy season.

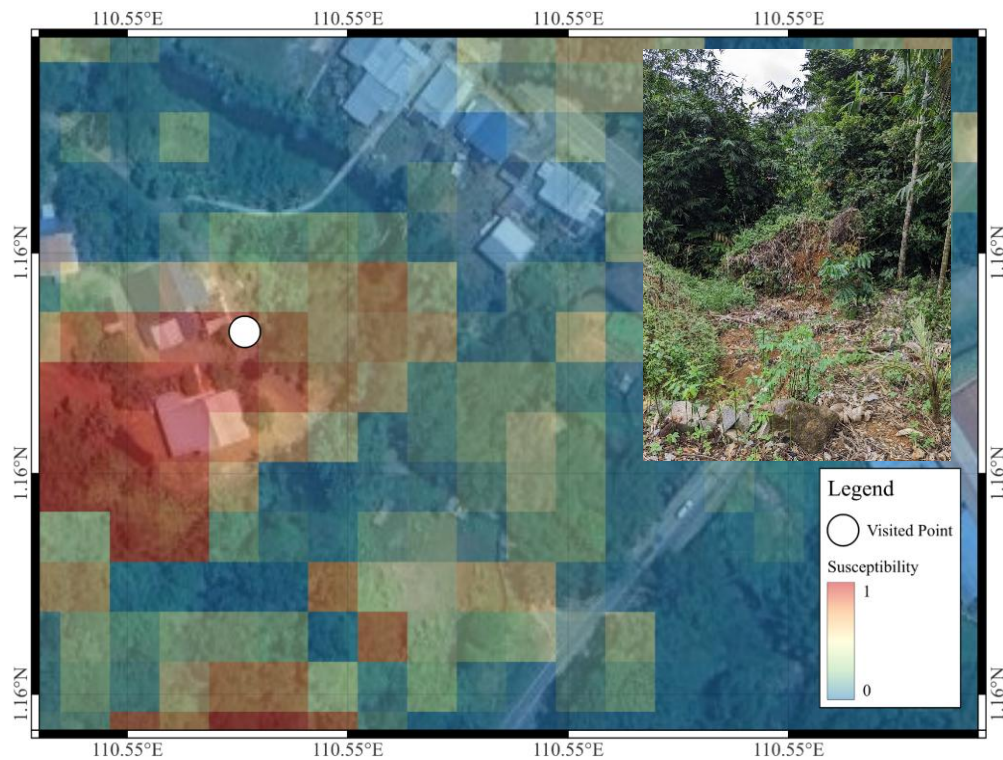


**Figure 4.6:** Tangga Mawang Slope Failure

Another slope failure that was located in Kampung Gayu, Kuching (landward). The failure point properties Gayu is very similar to Tangga Mawang's failure point properties, except for the aspect as seen in Table 4.14. The susceptibility level of the point was determined to be at 0.95 based on the LSM.

**Table 4.14:** Gayu failure point properties

| Variable       | Values  |
|----------------|---------|
| Aspect         | 15.75°  |
| Curvature      | -0.001  |
| Elevation      | 94 m    |
| Lithology type | L11     |
| Soil type      | S1      |
| Slope angle    | 22.053° |
| TWI            | 3.766   |

The slope failure at Gayu is illustrated in Figure 4.7. Similar to Tangga Mawang, the site visit to Gayu was conducted during the rainy season. However, the difference between the two areas is that Gayu location is at the feet of a mountain known as Muan Tabi. This has proven that the susceptible areas in Western Sarawak are not only located at the top part of the mountainous region. Thus, proving the importance of a multivariable approach in predicting the slope failure susceptibility for Western Sarawak.

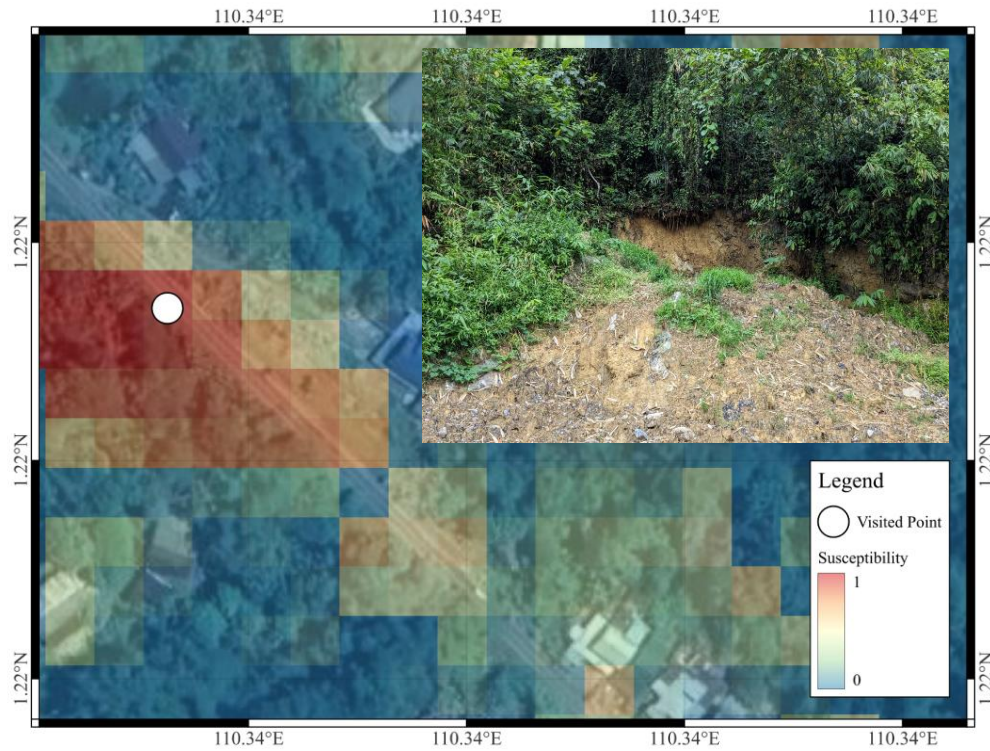


**Figure 4.7:** Gayu Slope Failure

However, not all high risks areas based on the LSM is justifiable. This was observed at Nirvana Cemetery, Bau, where based on the LSM, the susceptibility at the point of interest was 0.81. The properties of the point of interest can be seen in Table 4.15. Upon site visit, the area was relatively flat, as slope remediation effort had been conducted to make way for the construction of the cemetery.

**Table 4.15:** Nirvana Cemetery failure point properties

| Variable       | Values  |
|----------------|---------|
| Aspect         | 128.66° |
| Curvature      | -0.001  |
| Elevation      | 64 m    |
| Lithology type | L4      |
| Soil type      | S1      |
| Slope angle    | 16.053° |
| TWI            | 4.103   |

The current condition and the susceptibility level for the study area can be seen in Figure 4.8. However, this does not indicate that the ANN model was at fault, instead it was due to the spatiotemporal issue of the raster files. Specifically, it was linked to the elevation raster file as NASA had developed it in 2005, making it unable to track the current situation on site (Shahabi et al., 2022). Thus, the terrain change was not available in the raster file that had been used to plot the LSM. The spatiotemporal issue was to be overcome by using the SRAS to determine the slope failure susceptibility level on site.

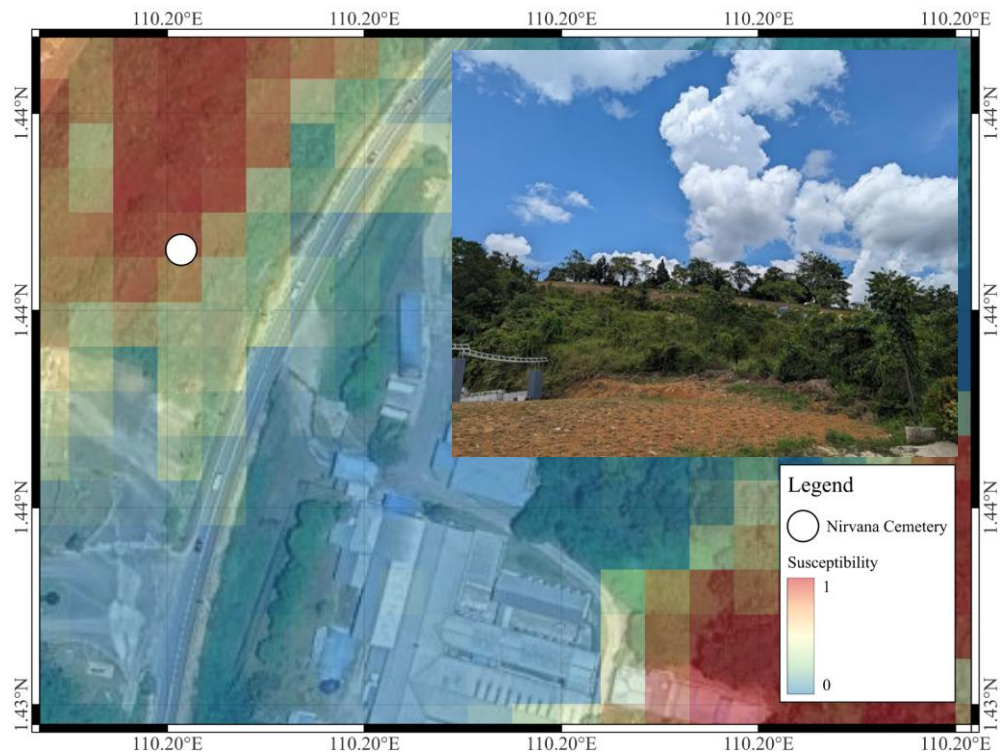


**Figure 4.8:** Nirvana Cemetery conditions

#### 4.5 Development and Verification for Western Sarawak SRAS – Phase IV

To develop the SRAS, the weightage from each connected neurons in the ANN model was extracted alongside the bias. Prior to developing the SRAS, a data analysis phase was conducted through the Olden algorithm to determine the general influence that each

variable has on the outcome of slope failure (Olden et al., 2004). This was done to provide an overview of the decision-making process done by the ANN model.

#### 4.5.1 Data Analysis for SRAS's ANN Model Variable Importance

The relative importance of each variable was determined through the “Olden” algorithm, through the “olden” function from the “NeuralNetTools” package in R (Marcus & Beck, 2022). This function has provided information on the relative importance of each slope failure factors to the outcome of slope failure occurrences which can be seen in Table 4.16. As seen in said table, the variables are represented in the pre-processed form, based on the ANN model training process. Amongst the slope failure factors, slope angle was the most influential variable, with a weight of 0.196, followed by TWI with 0.082, and elevation with 0.062. However, the relative importance does not reflect the final decision made on the weight by the ANN model.

**Table 4.16:** SRAS's ANN Model Variable Importance

| Variable    | Weight | Variable | Weight | Variable | Weight |
|-------------|--------|----------|--------|----------|--------|
| Elevation   | 0.062  | g        | 0.022  | L5       | 0.043  |
| Slope Angle | 0.196  | h        | 0.019  | L6       | 0.028  |
| TWI         | 0.082  | Convex   | 0.019  | L7       | 0.036  |
| a           | 0.017  | Flat     | 0.035  | L11      | 0.023  |
| b           | 0.019  | Concave  | 0.036  | S1       | 0.044  |
| c           | 0.024  | L1       | 0.017  | S2       | 0.013  |
| d           | 0.045  | L2       | 0.023  | S3       | 0.044  |
| e           | 0.028  | L3       | 0.040  | S4       | 0.024  |
| f           | 0.019  | L4       | 0.022  | S6       | 0.018  |

#### 4.5.2 SRAS's Development

The SRAS was developed based on the inner-workings of the ANN model, where weightages, biases, and activation function were crucial elements of the SRAS (Shapiro et al., 1999). To apply the SRAS as seen in Figure 4.9, the required slope failure factors must first undergo data pre-processing in the same manner as the training data and testing data. The pre-processed variables were then used as the input for each neuron in the hidden layer. Finally, the output from each neuron was used as the input to obtain the final score of slope failure susceptibility.

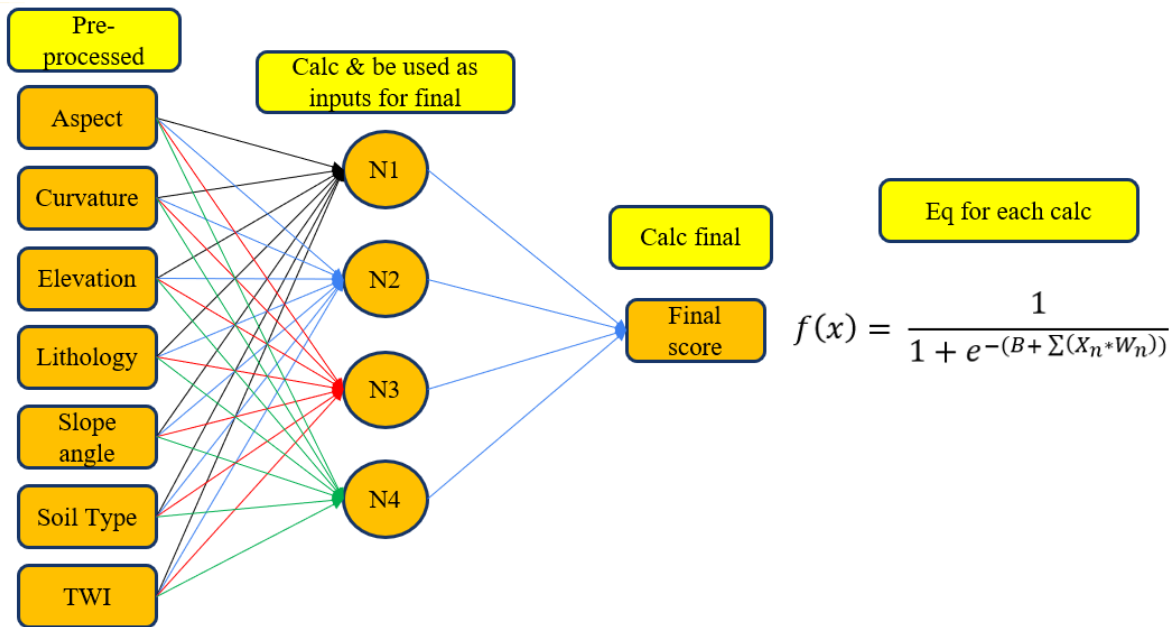


**Figure 4.9:** SRAS Application Procedure

For the SRAS, the input variables consisted of numerical variables which were by elevation, slope angle, and TWI, and categorical variables which were aspect, lithology, curvature, soil type as seen in Figure 4.10. The categorical variables were categorised through one-hot encoding in the same manner as the training and testing data for the ANN model.

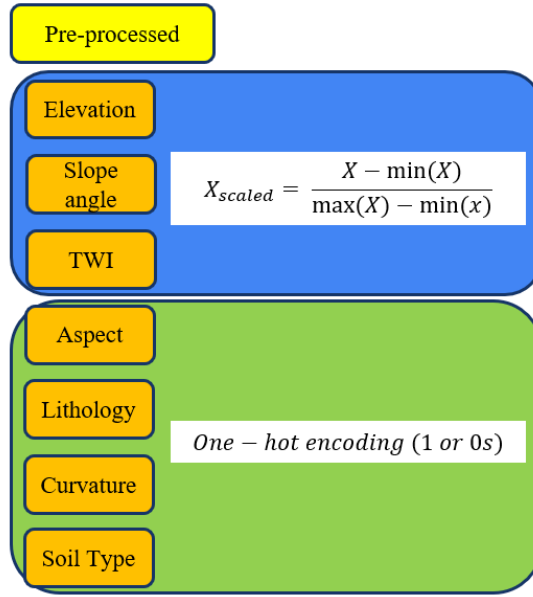


**Figure 4.10:** Handling Input Variables

On the other hand, the numerical input variables were processed through min-max scaling. The slope failure factors of elevation, slope angle, and TWI were processed through the simplified version of the min-max scaling equation for each variables as seen in Equation 4.1, Equation 4.2, and Equation 4.3 respectively. To calculate the output for each neuron in the SRAS, Equation 4.4 was used.

$$Processed(Elevation) = \frac{Elevation + 37\ m}{1229\ m} \quad \text{Equation 4.1}$$

$$Processed(Slope\ Angle) = \frac{Slope\ Angle}{76.1^\circ} \quad \text{Equation 4.2}$$

$$Processed(TWI) = \frac{TWI - 1.41}{17.02} \quad \text{Equation 4.3}$$

$$Output = \frac{1}{1 + e^{-(Bias + \sum(Processed\ Variable * Weight))}} \quad \text{Equation 4.4}$$

To determine outputs at each neurons, weightages and biases associated with each respective neurons are required. The value of weightages and biases for each variable in the hidden layer can be seen in Table 4.17.

**Table 4.17:** Assigned Weightages and Biases in the Hidden Layer

| Neuron 1     |               | Neuron 2     |               | Neuron 3     |               | Neuron 4     |               |
|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|
| Bias (B)     | -0.148        | Bias (B)     | 0.878         | Bias (B)     | 1.552         | Bias (B)     | -1.612        |
| Variable (X) | Weightage (W) | Variable (X) | Weightage (W) | Variable (X) | Weightage (W) | Variable (X) | Weightage (W) |
| Elevation    | 1.436         | Elevation    | 1.239         | Elevation    | -2.262        | Elevation    | 1.073         |
| Slope        | -0.091        | Slope        | 6.182         | Slope        | -6.765        | Slope        | 7.311         |
| TWI          | 1.291         | TWI          | -3.150        | TWI          | 2.883         | TWI          | -2.225        |
| a            | 0.558         | a            | -0.683        | a            | -0.233        | a            | -0.903        |
| b            | -0.111        | b            | -0.940        | b            | -0.381        | b            | -1.679        |
| c            | -1.052        | c            | -0.796        | c            | -0.234        | c            | -0.644        |
| d            | -1.628        | d            | -0.651        | d            | -0.180        | d            | -0.745        |
| e            | 1.114         | e            | -0.715        | e            | 0.463         | e            | -0.472        |
| f            | 0.256         | f            | -0.606        | f            | -0.706        | f            | -0.905        |
| g            | -1.267        | g            | -0.501        | g            | -0.597        | g            | -0.724        |
| h            | -0.938        | h            | -1.326        | h            | -0.567        | h            | -0.888        |
| Convex       | -0.516        | Convex       | -1.087        | Convex       | 0.619         | Convex       | -0.179        |
| Flat         | -0.958        | Flat         | -1.574        | Flat         | 0.988         | Flat         | 0.773         |
| Concave      | 0.667         | Concave      | -0.961        | Concave      | 0.605         | Concave      | -0.120        |
| L1           | -0.729        | L1           | 0.849         | L1           | 2.147         | L1           | 0.705         |
| L2           | -0.180        | L2           | -0.839        | L2           | 0.861         | L2           | 0.803         |
| L3           | -0.838        | L3           | -0.527        | L3           | 0.353         | L3           | 1.307         |
| L4           | -1.227        | L4           | 0.343         | L4           | 0.050         | L4           | 0.951         |
| L5           | -0.357        | L5           | 0.177         | L5           | 0.230         | L5           | 0.812         |
| L6           | -0.299        | L6           | 0.286         | L6           | -0.571        | L6           | 1.335         |
| L7           | 0.702         | L7           | -0.353        | L7           | -0.129        | L7           | -0.010        |
| L11          | -0.771        | L11          | -0.018        | L11          | -0.281        | L11          | 0.702         |
| S1           | 0.513         | S1           | -0.226        | S1           | -0.783        | S1           | 0.432         |
| S2           | -1.497        | S2           | -0.662        | S2           | -1.245        | S2           | -0.605        |
| S3           | -2.258        | S3           | -0.380        | S3           | -0.669        | S3           | 0.678         |
| S4           | 1.398         | S4           | -1.294        | S4           | -0.269        | S4           | -0.437        |
| S6           | -0.782        | S6           | 0.002         | S6           | -0.032        | S6           | 0.060         |

Each neurons in the hidden layer must be calculated for the output before proceeding to the final output neuron. Table 4.18 shows the unprocessed values for the input variables of a randomly selected high-risk area from the LSM, which will be used an example. Table 4.19 shows the pre-processed value or class for respective categorical and numerical input variables in the example point.

**Table 4.18:** Example's Input Variable

| No. | Input Variable | Values                                                                | Data Type   |
|-----|----------------|-----------------------------------------------------------------------|-------------|
| 1.  | Aspect         | 285.61°                                                               | Categorical |
| 2.  | Curvature      | 0.001                                                                 | Categorical |
| 3.  | Elevation      | 492m                                                                  | Numerical   |
| 4.  | Lithology Type | Argillaceous rocks; some Arenaceous and Calcareous beds (Palaeogeone) | Categorical |
| 5.  | Slope Angle    | 73°                                                                   | Numerical   |
| 6.  | Soil Type      | Albic Arenosols                                                       | Categorical |
| 7.  | TWI            | 1.42                                                                  | Numerical   |

**Table 4.19:** Data Pre-processing Outcome

| No. | Variable       | Processed |
|-----|----------------|-----------|
| 1.  | Elevation      | 0.406     |
| 2.  | Slope Angle    | 0.959     |
| 3.  | TWI            | 0.0001    |
| 4.  | Aspect         | g         |
| 5.  | Curvature      | Convex    |
| 6.  | Lithology Type | L3        |
| 7.  | Soil Type      | S6        |

Post data pre-processing, categorical variables classes that were not present based on the example points are not used. This provides a simpler understanding of the required weights, thus, reducing the confusion when applying the SRAS. Table 4.20 shows the only required fields from the total weights and biases of the hidden layer.

**Table 4.20:** Active Weights and Biases

| Neuron 1     |               | Neuron 2     |               | Neuron 3     |               | Neuron 4     |               |
|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|
| Bias (B)     | -0.148        | Bias (B)     | 0.878         | Bias (B)     | 1.552         | Bias (B)     | -1.612        |
| Variable (X) | Weightage (W) | Variable (X) | Weightage (W) | Variable (X) | Weightage (W) | Variable (X) | Weightage (W) |
| Elevation    | 1.436         | Elevation    | 1.239         | Elevation    | -2.262        | Elevation    | 1.073         |
| Slope        | -0.091        | Slope        | 6.182         | Slope        | -6.765        | Slope        | 7.311         |
| TWI          | 1.291         | TWI          | -3.150        | TWI          | 2.883         | TWI          | -2.225        |
| g            | -1.267        | g            | -0.501        | g            | -0.597        | g            | -0.724        |
| Convex       | -0.516        | Convex       | -1.087        | Convex       | 0.619         | Convex       | -0.179        |
| L3           | -0.838        | L3           | -0.527        | L3           | 0.353         | L3           | 1.307         |
| L7           | 0.702         | L7           | -0.353        | L7           | -0.129        | L7           | -0.010        |
| L11          | -0.771        | L11          | -0.018        | L11          | -0.281        | L11          | 0.702         |
| S6           | -0.782        | S6           | 0.002         | S6           | -0.032        | S6           | 0.060         |

The pre-processed values can now be used in conjunction with the weights and biases of each neuron to calculate the output value for each neuron as seen in the following calculations. It was determined through the calculations that neuron 1, neuron 2, neuron 3, and neuron 4 has the output value of 0.045, 0.994, 0.04, and 0.998 respectively.

$$Neuron1 = \frac{1}{1 + e^{-(-0.148 + \sum(1.436*0.406 + -0.091*0.959 + 1.291*0.0001 + -1.267*1 + -0.516*1 + -0.838*1 + -0.782*1))}} = 0.045$$

$$Neuron2 = \frac{1}{1 + e^{-(0.878 + \sum(1.239*0.406 + 6.182*0.959 + -3.150*0.0001 + -0.501*1 + -1.087*1 + -0.527*1 + 0.002*1))}} = 0.994$$

$$Neuron3 = \frac{1}{1 + e^{-(1.552 + \sum(-2.262*0.406 + -6.765*0.959 + 2.883*0.0001 + -0.597*1 + 0.619*1 + 0.353*1 + -0.032*1))}} = 0.004$$

$$Neuron4 = \frac{1}{1 + e^{-(-1.612 + \sum(1.073*0.406 + 7.311*0.959 + -2.225*0.0001 + -0.724*1 + -0.179*1 + 1.307*1 + 0.060*1))}} = 0.998$$

After the output of each neurons in the hidden layer has been calculated, the output were then used as the input for the final output neuron, which was the slope failure susceptibility score. The same calculation principle is applied with Equation 4.4 and each neurons in the hidden layer were assigned to a weight, and a bias for the final neuron as seen in Table 4.21.

**Table 4.21:** Assigned Weightages and Biases in the Output Layer

| Output       |               |
|--------------|---------------|
| Bias         | -2.792        |
| Variable (X) | Weightage (W) |
| Neuron 1     | -1.057        |
| Neuron 2     | 7.097         |
| Neuron 3     | -7.928        |
| Neuron 4     | 8.023         |

The final calculation is shown in the following for the output of slope failure susceptibility score. It was determined that for the high risk randomly sampled point, the output was 1, and indication of a high slope failure susceptibility. Thus, it was concluded that the SRAS was implementable.

$$Output = \frac{1}{1 + e^{-(-2.792 + \sum(-1.057*0.045 + 7.097*0.994 + -7.928*0.004 + 8.023*0.998))}} = 1$$

#### 4.5.3 SRAS's Evaluation

To evaluate the SRAS's, it was compared to its own ANN model for the AUC assessment. In terms of recall, the SRAS has achieved a score of 1, with all slope failure accurately classified for the testing data as seen in Table 4.22.

**Table 4.22:** SRAS ANN Performance Matrix

| Actual\Prediction | Testing |      |
|-------------------|---------|------|
|                   | False   | True |
| 0                 | 220     | 10   |
| 1                 | 0       | 10   |

There were ten false positive prediction made by the SRAS, however, this was acceptable as the focus was on the recall, due to the criticality of determining slope failure is more important than wrongfully predicted non-failure (Dai et al., 2021). In terms of overall performance as compared to the ANN model through AUC. The resulting AUC was 0.95 as seen in Error! Reference source not found., indicating that potentially, only 5% of the total instances were wrongfully predicted. As the AUC score was above the threshold of 0.8, the SRAS was deemed to be highly successful (Ado & Amitab, 2023). Thus, it was concluded that the SRAS was suitable for slope failure susceptibility prediction for the study area of Western Sarawak with a recall of 1, and AUC of 0.95.

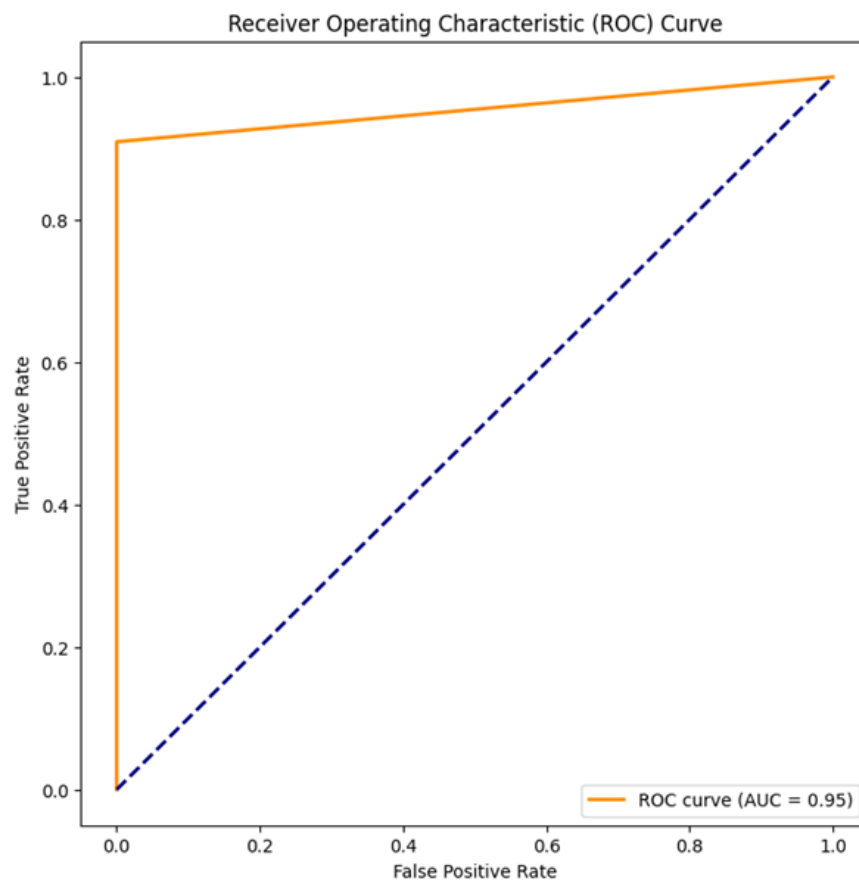


**Figure 4.11:** SRAS's AUC

#### 4.6 Validity of Remote Sensing data in Slope Failure Prediction

A separate study was conducted to validate the usability of remote sensing data as compared to soil physical properties when it comes to slope failure prediction through a Machine Learning based approach. The study area was Gangwon, Inje, South Korea, and the study was based on the a previous study done by Park et al., 2017. The remote sensing data obtained from various providers were aspect, curvature, elevation, saturated hydraulic conductivity, LULC, NDVI, and slope angle. The previous study on the other hand have used soil type, slope angle, elevation, soil friction angle, soil cohesion, unit weight, hydraulic conductivity, and soil thickness as the input variables. The resulting performance metrics of the models developed with the different variables type can be seen in Table 4.23. The model developed with the remote sensing data was able to outperform the model developed with soil physical properties as the input variables, with a recall of 0.885 as compared to the remote sensing model, and an AUC of 0.736 as compared to the soil physical properties model AUC of 0.736. This has validated the usability of remote sensing data when it comes to developing a Machine Learning model for slope failure susceptibility prediction (Al-Ruzouq et al., 2024).

**Table 4.23:** Performance Metrics Comparison

| Variable Type            | Recall | AUC   |
|--------------------------|--------|-------|
| Remote Sensing           | 0.885  | 0.736 |
| Soil Physical Properties | 0.955  | 0.995 |

#### **4.7 Comparison of Slope Failure Prediction Techniques**

A test conducted at the river basin level to evaluate the accuracy and precision of Weight of Evidence, ANN, and Factor Overlay Analysis (proposed by Anbalagan, 1995), using the input variables of aspect, curvature, distance from river, distance from road, elevation, soil type, and slope angle, based on the Guidelines On Integrating Climate Change Projection into Landslide Risk Assessment and Mapping of ASEAN (Rama Krishna et al., 2021). The primary difference between Weight of Evidence, and Factor Overlay Analysis to an ANN model was the capability of the ANN model to not only handle categorical variables, but also numerical variables (Zêzere et al., 2017).

Unlike Weight of Evidence, and Factor Overlay Analysis, the weightage assigned by the ANN model was determined by the lowest amount of error possible based on the maximum steps, and for each neuron in the hidden layer, there is an assigned weightage for the sub-variable (Loukrakpam & Oinam, 2021). Alongside the weightage, the ANN model comes with a bias assigned for each neuron, that is a constant used to adjust the overall fit of the model based on the training data (Fristch et al., 2022).

The accuracy assessment was done by evaluating the predictive performance of the ANN model, Factor Overlay Analysis model, and Weight of Evidence model using the ANN model's testing data, with a total of 18 instances: 3 landslide occurrences; and 15 non-landslide occurrences. The results of the prediction with a threshold of 0.5 can be seen in Table 4.24. It was determined that the ANN model was the most accurate model with an AUC of 1, followed by Factor Overlay Analysis model with an AUC of 0.78, and Weight of Evidence model with an AUC of 0.50 (Edamo et al., 2022). In terms of precision, the ANN model has scored 1, followed by the Factor Overlay Analysis model with a score of 0.40,

and Weight of Evidence with a precision score of 0.25. Thus, it was concluded that the ANN model was the most accurate model out of the three models with the score of 1 for AUC, and precision (Shreevastav et al., 2022).

**Table 4.24:** Models' confusion matrix

|        | ANN      |      | Factor Overlay Analysis |      | Weight of Evidence |      |
|--------|----------|------|-------------------------|------|--------------------|------|
| Actual | Non-fail | Fail | Non-fail                | Fail | Non-fail           | Fail |
| 0      | 15       | 0    | 12                      | 3    | 6                  | 9    |
| 1      | 0        | 3    | 1                       | 2    | 0                  | 3    |

#### 4.8 Remarks

In the primary dataset which was used to develop the training data and testing data for the LSM and SRAS ANN model, there were no significant multicollinearity amongst the input variable detected with the highest correlation amongst the input variable were between slope angle and elevation with a coefficient of 0.7. The LSM ANN model was developed with aspect, curvature, elevation, soil type, lithology type, LULC, slope angle, rainfall intensity, and TWI. The ANN model has achieved a recall of 0.9, and AUC score of 0.99 based on the testing data. The developed LSM has shown significant correlation to on-site conditions based on the slope failure cases that has been identified. However, the limitation of the LSM was the spatiotemporal properties of the raster files, where a remediated area was detected as a high-risk area. The SRAS ANN model was developed with fewer input variables, as dynamic variables were removed to facilitate the usability of the SRAS on-site. The SRAS ANN model was developed with aspect, curvature, elevation, soil type, lithology type, slope angle, and TWI. In terms of performance based on the testing data, the SRAS

ANN model has achieved a recall of 1, and an AUC score of 1. The SRAS was developed based on the weights and biases assigned by its ANN model, and the calculation was based on the inner workings of the ANN model. The SRAS has achieved a recall of 1, and an AUC of 0.95. Thus, it was concluded that the SRAS was suitable for use in predicting slope failure susceptibility in Western Sarawak. The usability of remote sensing data as input variables for the ANN models were validated by conducting a research to compare it with soil physical properties variables. It was deemed suitable as the remote sensing data model outperform the soil physical properties model as it scored 0.955 and 0.995 for recall and AUC respectively in comparison to the soil physical properties model which scored 0.885, and 0.736 for the metrics. The validity of an ANN model as a method to predict slope failure susceptibility was also compared with Weight of Evidence (a method suggested for ASEAN), and Factor Overlay Analysis (a method used to developed SMS and SMART). The ANN model outperformed the other models with an AUC of 1, followed by Factor Overlay Analysis with an AUC of 0.78, and Weight of Evidence with an AUC of 0.5.

## **CHAPTER 5**

### **CONCLUSION AND RECOMMENDATIONS**

#### **5.1 Introduction**

The lack of a region-specific SAS and LSM for Western Sarawak has been one of the biggest challenges faced by geotechnical engineers and policymakers in the region. A highly accurate SAS and LSM are crucial as a mitigation effort to avoid the potential risks of exposure to the hazards of slope failures in Western Sarawak. In this study, the two ANN models were developed for plotting the LSM, and development of the SRAS respectively. The LSM ANN model was developed with input variables of aspect, curvature, elevation, soil type, lithology type, LULC, slope angle, rainfall intensity, and TWI. The SRAS ANN model on the other hand was developed with fewer input variables with the dynamic variable of LULC and rainfall intensity removed to enable its application on-site. Evaluating the ANN models through recall and AUC has revealed that the models were highly accurate with a recall of 0.9 and an AUC of 0.99 for the LSM ANN model, and a recall of 1 and an AUC of 1 for SRAS ANN model. The LSM was determined to be significantly correlated with the on-site conditions that has not been disturbed significantly. In areas where significant terrain changes, the LSM was found to be lacking due to the spatiotemporal properties of the raster files. The SRAS is complementary to the LSM as it can be used to evaluate slope failure susceptibility on-site, thus, eliminating spatiotemporal issues. The SRAS was evaluated with recall and AUC, where it has achieved a score of 1 and 0.95 for respective metrics. Thus, it was concluded that the LSM and SRAS that has been developed were highly suitable for evaluating slope failure susceptibility in Western Sarawak.

The lack of SAS and LSM for Western Sarawak has been one of the biggest challenges faced by geotechnical engineers, and policy makers in the region. A suitable SAS, and a highly accurate LSM is very important to avoid slope failure occurrences as well as to determine areas that require mitigation efforts. An SAS is very important to provide continuous assessment of slopes in Western Sarawak, as remote sensing data such as elevation are updated occasionally. This thesis has focused on the critical appraisal of slope failure contributing parameters for Western Sarawak via multi statistical approaches where the Artificial Neural Network was utilised to produce an LSM, and the SRAS. This final chapter gives a summary of the key components of the research and provides recommendations for future work in the area.

## **5.2 Contribution of This Research**

The main contribution of this research are as follows:

- i. ANN models for slope failure susceptibility prediction of Western Sarawak.
- ii. The first LSM for the region generated through an ANN model.
- iii. An SRAS specifically for the region of Western Sarawak.

## **5.3 Achievement of Objective**

There were three primary objectives that were achieved in this study as seen in the following:

- i. To analyse and critically appraise the contributing factors of slope failures such as geological, hydrological, meteorological, and physical factors leading to slope failures in Western Sarawak – Phase I.

There were two approaches in analysing and appraising the contribution of each slope failure factors that has been used as the input variables in this study, which were the multicollinearity assessment, and Olden's algorithm. The main goal of conducting a multicollinearity assessment was to determine whether significant correlation exists or otherwise amongst the input variables of slope failure factors. Significant correlation is determined when the correlation coefficient amongst a pair of input variables exceeds the threshold of 0.8. No significant correlation amongst the input variables were detected. Moreover, the multicollinearity assessment also reveals the correlation between the input variables to the out of slope failure. It was determined that for numerical variables, slope failure was significantly correlated with slope failure in Western Sarawak, with a corresponding correlation coefficient of 0.88, followed by elevation with 0.59. TWI negatively correlated with slope failure in the region, with a correlation coefficient of -0.37. This indicates that as the terrain wetness decreases, the likelihood of slope failure increases. This was due to the inability of the less wet surface to handle heavy rainfall which causes slope failure susceptibility to increase. However, multicollinearity assessment only analyses and appraise the importance of variables based purely on the primary dataset. To understand the importance of variables based on the decision-making process of the ANN model at a relative importance level, the Olden's algorithm was deployed. Understanding the relative importance for the SRAS ANN model has helped in comprehending the SRAS. Based on the algorithm, terrain variable of slope angle, elevation, and TWI has the highest influence on slope failure susceptibility in the region with a relative importance of 0.196, 0.082, and 0.062 respectively. In terms of aspect, class d was the most influential aspect orientation with a relative importance of 0.045. Concave curvature being the most susceptible curvature to slope failure with 0.036, followed closely by flat curvature with 0.035. Lithology class L5

was the most susceptible lithology with a relative importance of 0.043. Finally soil type S3 and S1 was the most susceptible soil type with a relative importance of 0.044 respectively.

ii. To assess the suitability of ANN in terms of its accuracy in the development of a framework analysis for SRAS, and an LSM – Phase II

The developed ANN models were evaluated through recall, and AUC for the predictive performance. For the LSM ANN model, it has yielded a recall of 0.9, and an AUC score of 0.99, whereas the SRAS ANN model has achieved a recall and AUC score of 1 for respective metrics. In terms of comparison to other approaches in slope failure susceptibility assessment, at a basin level, an ANN model performance was compared to Weight of Evidence, and Factor Overlay Analysis, where it has outperformed the other approaches. The ANN model has achieved an AUC of 1, followed by Factor Overlay Analysis with an AUC of 0.78, and Weight of Evidence with an AUC of 0.50. Thus, an ANN model was deemed to be a suitable approach for a framework analysis for a SRAS, and an LSM.

iii. To develop an LSM and a SRAS for Western Sarawak from the ANN model - Phase III, and Phase IV

The LSM's slope failure susceptibility level was predicted by its ANN model. To develop the LSM, the slope failure factors raster files were converted into a numerical data frame that underwent data-preprocessing in the same manner as the training data and testing data before deploying the ANN model. The LSM was highly accurate based on its ANN model evaluation metrics where some slope failure points were discovered to be in the high risks area marked by the LSM. However, the limitation of the LSM was spatiotemporal, where recent significant terrain changes were not reflected in the LSM. Thus, certain low-risk areas that had underwent remediation efforts were still marked as high-risk areas in the

LSM. The SRAS was a complementary part of the LSM. Whenever discrepancies are determined on-site where the conditions do not correlate with the LSM, the SRAS can be deployed to determine the slope failure susceptibility without access to its ANN model. The SRAS was compared to its ANN model where it has achieved an AUC of 0.95. Thus, both the LSM and SRAS development were deemed to be highly successful.

#### **5.4 Recommendation for Future Study**

This research could serve as a source of information for future research in the slope failure susceptibility prediction with Machine Learning field. The followings are future research areas that can be explored:

- i. This study was conducted with varying resolution of raster files that has been resampled to ensure the resolution were kept uniform. Future study should investigate the impact of uniform raster files resolution from source to the ANN models. A uniform raster file resolution, especially high-resolution raster files, would yield xa more accurate LSM in spatial terms.
- ii. Activation function has played an important role in formulating the SRAS. In this study, the sigmoidal activation function was used. Future study should investigate the suitable framework for formulating a new SAS through a different activation function. A different activation should yield a different output, yet the correlation between the output and the activation function used should be a novel approach.
- iii. Spatiotemporal was the primary limitation of the LSM. Future study should consider developing an LSM based on real time variables that are available through remote sensing data providers. This would remove the spatiotemporal issues from the LSM.

## REFERENCES

- Abidoye, L. K., Mahdi, F. M., Idris, M. O., Alabi, O. O., & Wahab, A. A. (2018). ANN-derived equation and ITS application in the prediction of dielectric properties of pure and impure CO<sub>2</sub>. *Journal of Cleaner Production*, 175, 123–132. <https://doi.org/10.1016/j.jclepro.2017.12.013>
- Achour, Y., & Pourghasemi, H. R. (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>
- Adnan, M. S. G., Siam, Z. S., Kabir, I., Kabir, Z., Ahmed, M. R., Hassan, Q. K., Rahman, R. M., & Dewan, A. (2023). A novel framework for addressing uncertainties in Machine Learning-based geospatial approaches for flood prediction. *Journal of Environmental Management*, 326(PB), 116813. <https://doi.org/10.1016/j.jenvman.2022.116813>
- Ado, M., & Amitab, K. (2023). Landslide Susceptibility Mapping using support vector machine for Meghalaya, India. *2023 4th International Conference on Computing and Communication Systems*, I3CS 2023. <https://doi.org/10.1109/I3CS58314.2023.10127361>
- Akter, A., Noor, M. J. M. M., Goto, M., Khanam, S., Parvez, A., & Rasheduzzaman, M. (2019). Landslide Disaster in Malaysia: An Overview. *International Journal of Innovative Research and Development*, 8(6), 292–302. <https://doi.org/10.24940/ijird/2019/v8/i6/jun19058>
- Al-Ruzouq, R., Shanableh, A., Jena, R., Gibril, M. B. A., Hammouri, N. A., & Lamghari, F. (2024). Flood susceptibility mapping using a novel integration of multi-temporal

- sentinel-1 data and eXtreme deep learning model. *Geoscience Frontiers*, 15(3), 1–18.  
<https://doi.org/10.1016/j.gsf.2024.101780>
- Alamgir, M., Campbell, M. J., Sloan, S., Engert, J., Word, J., & Laurance, W. F. (2020). Emerging challenges for sustainable development and forest conservation in Sarawak, Borneo. *PLoS ONE*, 15(3). <https://doi.org/10.1371/journal.pone.0229614>
- Albar, A., & Mohd-Nordin, M. M. (2022). Slope Mass Rating (SMR) classification for rock slope stability and geohazard vulnerability assessment. *International Journal of Integrated Engineering*, 14(5), 57–65. <https://doi.org/10.30880/ijie.2022.14.05.006>
- Aldiansyah, S., & Risna, R. (2023). Assessing potential habitat suitability of vulnerable endemic species: a case study of *Diospyros celebica* Bakh and *Rhyticeros cassidix*. *Forum Geografic*, 22(2), 159–169. <https://doi.org/10.5775/fg.2023.2.3596>
- Alqawasmeh, H. (2023). Comparative Assessment of Various Artificial Neural Network Techniques for Estimating the Safety Factor of Road Embankments. *International Journal of GEOMATE*, 24(103), 42–51. <https://doi.org/10.21660/2023.103.3532>
- Althuwaynee, O. F., & Pradhan, B. (2018). Landslides: Causes, Mapping and Monitoring – Examples from Malaysia. In *Natural Hazards: Earthquakes, Volcanoes, and Landslides* (pp. 363–396). CRC Press Taylor & Francis Group.  
<http://taylorandfrancis.com>
- Anis, Z., Wissem, G., Vali, V., Smida, H., & Mohamed Essghaier, G. (2019). GIS-based Landslide Susceptibility Mapping using bivariate statistical methods in North-western Tunisia. *Open Geosciences*, 11(1), 708–726. <https://doi.org/10.1515/geo-2019-0056>
- Azis, M. (2018). Malaysia: Country profile. In *Asian Disaster Reduction Center (ADRC)* (Issue November).  
[https://www.adrc.asia/countryreport/MYS/FY2018/Malaysia\\_CR2018A.pdf](https://www.adrc.asia/countryreport/MYS/FY2018/Malaysia_CR2018A.pdf)

- Aziz, A. F., Mardi, N. H., Abdul Malek, M., Teh, S. Y., Wil, M. A., Shuja, A. H., Ahmed, A. N., Kumar, P., Sherif, M., & Elshafie, A. (2023). Development of inundation maps along East Coast of Peninsular Malaysia due to predicted seaquake from Manila Trench. *Applied Water Science*, 13(2). <https://doi.org/10.1007/s13201-022-01860-8>
- Babitha, Vincent, S., Pathan, S., & Garcia Benitez, S. R. (2022). Landslides in Goa: A Weight of Evidence(WoE) approach for mapping. *Proceedings - 2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems, ICMACC 2022*, 154–160. <https://doi.org/10.1109/ICMACC54824.2022.10093638>
- Badola, S., Mishra, V. N., Parkash, S., & Pandey, M. (2023). Rule-based fuzzy inference system for Landslide Susceptibility Mapping along national highway 7 in Garhwal Himalayas, India. *Quaternary Science Advances*, 11(March), 100093. <https://doi.org/10.1016/j.qsa.2023.100093>
- Baig, M. F., Mustafa, M. R. U., Baig, I., Takaijudin, H. B., & Zeshan, M. T. (2022). Assessment of Land Use Land Cover Changes and Future Predictions Using CA-ANN Simulation for Selangor, Malaysia. *Water (Switzerland)*, 14(3). <https://doi.org/10.3390/w14030402>
- Bambang, I., Saravanan, S., Reddy, N. M., & Abijith, D. (2023). An investigation of the changing patterns of rainfall in the Indravathi subbasin utilizing the Mann-Kendall and Sen's slope methods. *IOP Conference Series: Earth and Environmental Science*, 1173(1), 2024. <https://doi.org/10.1088/1755-1315/1173/1/012036>
- Bivand, R., Hijmans, R. J., Pebesma, E., & Sumner, M. D. (2024). *Package "terra."* CRAN. <https://rspatial.org/>
- Bončina, A., Klopčič, M., Trifković, V., Ficko, A., & Simončič, P. (2023). Tree and stand

- growth differ among soil classes in semi-natural forests in central Europe. *Catena*, 222(December 2022). <https://doi.org/10.1016/j.catena.2022.106854>
- Bradley, A. E. (1997). The use of the Area Under the ROC Curve in the evaluation of Machine Learning algorithms. *Pattern Recognition*, 30(7), 1145–1159. [https://doi.org/https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/https://doi.org/10.1016/S0031-3203(96)00142-2)
- Bravo-López, E., Fernández Del Castillo, T., Sellers, C., & Delgado-García, J. (2023). Analysis of conditioning factors in Cuenca, Ecuador, for Landslide Susceptibility Maps generation employing Machine Learning methods. *Land*, 12(6), 1–28. <https://doi.org/10.3390/land12061135>
- Breitfeld, H. T. (2021). Provenance of Pleistocene Sediments in West Sarawak and Evidence for Pliocene Acid Magmatism in Central Borneo. *Berita Sedimentologi*, 47(1), 5–32. <https://doi.org/10.51835/bsed.2021.47.1.51>
- Chen, S., Miao, Z., Wu, L., & He, Y. (2020). Application of an incomplete landslide inventory and One Class Classifier to earthquake-induced Landslide Susceptibility Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1649–1660. <https://doi.org/10.1109/JSTARS.2020.2985088>
- 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>
- Chigira, M., Mohamad, Z., Sian, L. C., & Komoo, I. (2011). Landslides in weathered granitic rocks in Japan and Malaysia. *Bulletin of the Geological Society of Malaysia*, 57(57), 1–6. <https://doi.org/10.7186/bgsm57201101>
- Chuma, G. B., Mugumaarhahama, Y., Mond, J. M., Bagula, E. M., Ndeko, A. B., Lucungu,

- P. B., Karume, K., Mushagalusa, G. N., & Schmitz, S. (2023). Gully erosion susceptibility mapping using four Machine Learning methods in Luzinzi watershed, eastern Democratic Republic of Congo. *Physics and Chemistry of the Earth*, 129, 1–15. <https://doi.org/10.1016/j.pce.2022.103295>
- Cobos-Mora, S. L., Rodriguez-Galiano, V., & Lima, A. (2023). Analysis of landslide explicative factors and susceptibility mapping in an andean context: The case of Azuay province (Ecuador). *Heliyon*, 9(9), e20170. <https://doi.org/10.1016/j.heliyon.2023.e20170>
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Culhane, A. (2013). *Introduction to Programming and Statistical Modelling in R* (1st ed.). Harvard University. [https://doi.org/https://ds.dfci.harvard.edu/~aedin/courses/R/Bio503\\_winter13.pdf](https://doi.org/https://ds.dfci.harvard.edu/~aedin/courses/R/Bio503_winter13.pdf)
- Dai, L., Zhu, M., He, Z., He, Y., Zheng, Z., Zhou, G., Wang, C., Ren, J., Tang, H., Liu, Q., Huang, F., Li, Z., Li, M., Wang, Z., Li, M., & Jiang, L. (2021). Landslide risk classification based on ensemble Machine Learning. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2021-July, 3924–3927. <https://doi.org/10.1109/IGARSS47720.2021.9553034>
- Dambul, R., & Jones, P. (2007). Regional and temporal climatic classification for Borneo. *GEOGRAFIA Malaysia Journal of Society and Space*, 3, 84–105.
- Das, S. K., Samui, P., Khan, S. Z., & Sivakugan, N. (2011). Machine Learning techniques applied to prediction of residual strength of clay. *Central European Journal of*

- Geosciences*, 3(4), 449–461. <https://doi.org/10.2478/s13533-011-0043-1>
- Desyatkin, R. V., Okoneshnikova, M. V., Ivanova, A. Z., Desyatkin, A. R., & Filippov, N. V. (2021). Sandy soils of desert-like landscapes (tukulans) of Central Yakutia. *IOP Conference Series: Earth and Environmental Science*, 862(1). <https://doi.org/10.1088/1755-1315/862/1/012003>
- Diana, M. I. N., Muhamad, N., Taha, M. R., Osman, A., & Alam, M. M. (2021). Social vulnerability assessment for landslide hazards in Malaysia: A systematic review study. In *Land* (Vol. 10, Issue 3). MDPI AG. <https://doi.org/10.3390/land10030315>
- Diebold, F. X. (2019). Econometric data science: a predictive modelling approach. In *Econometrics*. pdf. University of Pennsylvania. <https://www.sas.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>
- Djukem, W. D. L., Braun, A., Wouatong, A. S. L., Guedjeo, C., Dohmen, K., Wotchoko, P., Fernandez-Steege, T. M., & Havenith, H. B. (2020). Effect of soil geomechanical properties and geo-environmental factors on landslide predisposition at mount oku, cameroon. *International Journal of Environmental Research and Public Health*, 17(18), 1–28. <https://doi.org/10.3390/ijerph17186795>
- Edamo, M. L., Ukumo, T. Y., Lohani, T. K., Ayana, M. T., Ayele, M. A., Mada, Z. M., & Abdi, D. M. (2022). A comparative assessment of multi-criteria decision-making analysis and Machine Learning methods for flood susceptibility mapping and socio-economic impacts on flood risk in Abela-Abaya floodplain of Ethiopia. *Environmental Challenges*, 9(October). <https://doi.org/10.1016/j.envc.2022.100629>
- Faizin, & Nur, B. A. (2018). Landslides susceptibility mapping at Gunung Ciremai National Park. *E3S Web of Conferences*, 31, 3–6. <https://doi.org/10.1051/e3sconf/20183112010>
- Fan, X., Liu, B., Luo, J., Pan, K., Han, S., & Zhou, Z. (2023). Comparison of earthquake-

- induced shallow landslide susceptibility assessment based on two-category LR and KDE-MLR. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-28096-z>
- FAO-UNESCO. (1979). Soil map of the world: South East Asia. In *Soil Map of The World* (1st ed., Vol. 9). <https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/>
- Fonseca, L. D. M., Lani, J. L., Filho, E. I. F., dos Santos, G. R., Ferreira, W. P. M., & Trancoso Santos, A. M. R. (2017). Variabilidade dos atributos físicos do solo em áreas suscetíveis ao deslizamento de terra. *Acta Scientiarum - Agronomy*, 39(1), 109–118. <https://doi.org/10.4025/actasciagron.v39i1.30561>
- Fristch, S., Guenther, F., Wright, M. N., & Mueller, S. M. (2022, February 7). *Package “neuralnet.”* CRAN. <https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*, 2, 1–21. <https://doi.org/10.1038/sdata.2015.66>
- Gandomi, M., Soltanpour, M., Zolfaghari, M. R., & Gandomi, A. H. (2016). Prediction of peak ground acceleration of Iran’s tectonic regions using a hybrid soft computing technique. *Geoscience Frontiers*, 7(1), 75–82. <https://doi.org/10.1016/j.gsf.2014.10.004>
- Geertsema, M., Highland, L., & Vaugeouis, L. (2009). Environmental Impact of Landslides. In S. Kyoji, P. Canuti, L. Highland, N. Casagli, S. Tsuchiya, & F. Wang (Eds.), *Landslides Disaster Risk Reduction* (pp. 589–607). Springer-Verlag.
- Gershenson, C. (2003). *Artificial Neural Networks for Beginners*. ResearchGate. <https://www.researchgate.net/publication/1956697>

- Ghani, M. F. A., Simon, N., Mohamed, T. R. T., & Roslee, R. (2022). 3D modelling of rockfall hazard at Gunung Lang, Ipoh. *IOP Conference Series: Earth and Environmental Science*, 1103(1). <https://doi.org/10.1088/1755-1315/1103/1/012028>
- Ghasemian, B., Shahabi, H., Shirzadi, A., Al-Ansari, N., Jaafari, A., Kress, V. R., Geertsema, M., Renoud, S., & Ahmad, A. (2022). A robust deep-learning model for Landslide Susceptibility Mapping: A case study of Kurdistan province, Iran. *Sensors*, 22(4). <https://doi.org/10.3390/s22041573>
- Gomes, M. C. V., & Vieira, B. C. (2016). Saturated hydraulic conductivity of soils in a shallow landslide area in the Serra do Mar, São Paulo, Brazil. *Zeitschrift Fur Geomorphologie*, 60(1), 53–65. <https://doi.org/10.1127/zfg/2016/0229>
- Gray, J. M., Bishop, T. F. A., & Wilford, J. R. (2016). Lithology and soil relationships for soil modelling and mapping. *Catena*, 147, 429–440. <https://doi.org/10.1016/j.catena.2016.07.045>
- Guerzhoy, M., & Hinton, G. (2017). *One-Hot encoding CSC411: Machine Learning and data mining*. University of Toronto. <http://www.cs.toronto.edu/~guerzhoy/411/>
- Gupta, S. K., & Shukla, D. P. (2020). Data imbalance in landslide susceptibility zonation: A case study of Mandakini river basin, Uttarakhand, India. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 5234–5237. <https://doi.org/10.1109/IGARSS39084.2020.9323230>
- 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. <https://doi.org/10.1016/j.jenvman.2021.114367>
- Halim Ida, S. A., Shuib, R., & Noh Muhammad, R. M. (2022). Site-suitability analysis on

- seismic stations using geographic information systems. *Disaster Advances*, 15(2), 1–14. <https://doi.org/10.25303/1502da001014>
- Hashim, K., & Among, L. H. (2003). Geological investigation on Ruan Changkul landslide. *Geological Society of Malaysia*, 46, 125–132.
- Hassanzadeh, H., Nguyen, A., Karimi, S., & Chu, K. (2018). Transferability of artificial neural networks for clinical document classification across hospitals: A case study on abnormality detection from radiology reports. *Journal of Biomedical Informatics*, 85, 68–79. <https://doi.org/10.1016/j.jbi.2018.07.017>
- Highland, L. (2004). Landslide types and processes. In *U.S Geological Survey (USGS)*. <https://doi.org/http://pubs.usgs.gov/fs/2004/3072/>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Horton, N. J., Kleinman Nicholas Horton, K. J., & Kleinman, K. (2015). *Using R and RStudio for data management, statistical analysis, and graphic* (2nd ed.). Taylor & Francis Group.
- Huat, B. B. K., Ali, F. H., & Rajoo, R. S. K. (2006). Stability analysis and stability chart for unsaturated residual soil slope. *American Journal of Environmental Sciences*, 2(4), 154–160. <https://doi.org/10.3844/ajessp.2006.154.160>
- Huat, B. B. K., & Jamaludin, S. (2005). Evaluation of slope assessment system in predicting landslides along roads underlain by granitic formation. *American Journal of Environmental Sciences*, 1(2), 90–96. <https://doi.org/10.3844/ajessp.2005.90.96>
- Iban, M. C., & Sekertekin, A. (2022). Machine Learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin

- provinces, Turkey. *Ecological Informatics*, 69(April), 101647.  
<https://doi.org/10.1016/j.ecoinf.2022.101647>
- Ighile, E. H., Shirakawa, H., & Tanikawa, H. (2022). A study on the application of GIS and Machine Learning to predict flood areas in Nigeria. *Sustainability (Switzerland)*, 14(9), 1–33. <https://doi.org/10.3390/su14095039>
- Isia, I., Hadibarata, T., Jusoh, M. N. H., Bhattacharjya, R. K., Shahedan, N. F., Bouaissi, A., Fitriyani, N. L., & Syafrudin, M. (2023). Drought analysis based on standardized precipitation evapotranspiration index and standardized precipitation index in Sarawak, Malaysia. *Sustainability (Switzerland)*, 15(1). <https://doi.org/10.3390/su15010734>
- Jaapar, A. R., Md Aripin, M. F. S., Komoo, I., Ali, C. A., Ramli, Z., Harun, A. R., Bachat, J., Mohamad, Z., Omar, R. C., Razak, K. A., & Lim, C. S. (2023). The emerging widespread debris flow disasters in tropical terrain of Peninsular Malaysia: Understanding the risk and policy intervention. *E3S Web of Conferences*, 415. <https://doi.org/10.1051/e3sconf/202341505008>
- Jamaludin, S., K Huat, B. B., & Omar, H. (2006). Evaluation and development Of cut–slope assessment systems for Peninsular Malaysia in predicting landslides in granitic formation. *Jurnal Teknologi*, 44(B), 31–46. <https://doi.org/10.11113/jt.v44.360>
- JUPEM. (2009). Garis panduan mengenai penukaran koordinat, transformasi Datum dan unjuran peta untuk tujuan ukur dan pemetaan. In *Pekeliling Ketua Pengarah Ukur dan Pemetaan* (Vol. 3).
- Kaamin, M., Kadir, A. A., Zarin, N. N. A., Ahmad, N. F. A., Razali, S. N. M., Bukari, S. M., & Ahmad, M. A. (2016). Hazard and risk slope inventory using Geographical Information System (GIS): Case study at federal route Simpang Pulai from section 35 to 45. *IOP Conference Series: Materials Science and Engineering*, 160(1), 1–7.

<https://doi.org/10.1088/1757-899X/160/1/012072>

Kalantar, B., Ueda, N., Lay, U. S., Al-Najjar, H. A. H., & Halin, A. A. (2019). Conditioning factors determination for Landslide Susceptibility Mapping using support vector Machine Learning. *IGARSS*, 9626–9629.

<https://doi.org/10.1109/IGARSS.2019.8898340>

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021). Global land use / land cover with Sentinel 2 and deep learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4704–4707.

<https://doi.org/10.1109/IGARSS47720.2021.9553499>

Khajehzadeh, M., Taha, M. R., Keawsawasvong, S., Mirzaei, H., & Jebeli, M. (2021). Three-dimensional slope stability predictions using artificial neural networks. *International Journal for Numerical and Analytical Methods in Geomechanics*, 45(13), 1988–2000.

<https://doi.org/10.1002/nag.3252>

Khajehzadeh, M., Taha, M. R., Keawsawasvong, S., Mirzaei, H., & Jebeli, M. (2022). An effective artificial intelligence approach for slope stability evaluation. *IEEE Access*, 10, 5660–5671. <https://doi.org/10.1109/ACCESS.2022.3141432>

Khan, S. Z., Suman, S., Pavani, M., & Das, S. K. (2016). Prediction of the residual strength of clay using functional networks. *Geoscience Frontiers*, 7(1), 67–74.

<https://doi.org/10.1016/j.gsf.2014.12.008>

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine Learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3–10.

<https://doi.org/10.1016/j.gsf.2015.07.003>

Laurencelle, J., Logan, T., & Gens, R. (2015). *ALOS PALSAR - Radiometric Terrain Correction*. JAXA/METI ALOS PALSAR. <https://doi.org/10.5067/J4JVCFDDPEW1>

- Le, T. T. T., Tran, T. V., Hoang, V. H., Bui, V. T., Bui, T. K. T., & Nguyen, H. P. (2021). Developing a Landslide Susceptibility Map using the analytic hierarchical process in ta van and hau thao communes, Sapa, Vietnam. *Journal of Disaster Research*, 16(4), 529–538. <https://doi.org/10.20965/JDR.2021.P0529>
- Lee, D. H., Kim, Y. T., & Lee, S. R. (2020). Shallow landslide susceptibility models based on artificial neural networks considering the factor selection method and various non-linear activation functions. *Remote Sensing*, 12(7). <https://doi.org/10.3390/rs12071194>
- Lemenkova, P., & Debeir, O. (2023). GDAL and PROJ Libraries Integrated with GRASS GIS for Terrain Modelling of the Georeferenced Raster Image. In *Technologies* (Vol. 11, Issue 2). <https://doi.org/10.3390/technologies11020046>
- 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(1), 1–14. <https://doi.org/10.1155/2021/4758062>
- Liaw, S. E. (2020). *Business and investment opportunities in Sarawak at MIDA Invest Series: Unfolding states' business potential*. Ministry of Industrial and Entrepreneur Development Sarawak. [https://www.mida.gov.my/wp-content/uploads/2020/07/20190131110151\\_190128-MIED-Slides-for-MIDA-INVEST-SERIES.pdf](https://www.mida.gov.my/wp-content/uploads/2020/07/20190131110151_190128-MIED-Slides-for-MIDA-INVEST-SERIES.pdf)
- Loukrakpam, C., & Oinam, B. (2021). Linking the past, present and future scenarios of soil erosion modeling in a river basin. *Global Journal of Environmental Science and Management*, 7(3), 457–472. <https://doi.org/10.22034/GJESM.2021.03.09>
- Luo, S., Sarabandi, K., Tong, L., & Pierce, L. E. (2021). Probability assessment of rainfall-induced landslides based on safety factors using soil moisture estimation from SAR

- images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7), 5579–5597.  
<https://doi.org/10.1109/TGRS.2020.3025996>
- Ma, Z., Mei, G., & Piccialli, F. (2020). Machine Learning for landslides prevention: a survey. *Neural Computing and Applications*. <https://doi.org/https://doi.org/10.1007/s00521-020-05529-8>
- Majid, N. A., Taha, M. R., & Selamat, S. N. (2020). Historical landslide events in Malaysia 1993-2019. *Indian Journal of Science and Technology*, 13(33), 3387.  
<https://doi.org/10.17485/IJST/v13i33.884>
- Marcus, M., & Beck, W. (2022). *Package ‘NeuralNetTools.’* <https://cran.r-project.org/package=NeuralNetTools/NeuralNetTools.pdf>
- Martin, D., & Chai, S. S. (2022). A study on performance comparisons between KNN, Random Forest and XGBoost in prediction of landslide susceptibility in Kota Kinabalu, Malaysia. *2022 IEEE 13th Control and System Graduate Research Colloquium, ICSGRC 2022 - Conference Proceedings*, 159–164.  
<https://doi.org/10.1109/ICSGRC55096.2022.9845146>
- Md.Sharafat, C., Md.Naimur, R., Md.Sujon, S., Md.Abu, S., Mahmud, K. H., & Hafsa, B. (2024). GIS-based Landslide Susceptibility Mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh. *Heliyon*, 10(e23424), 1–19. <https://doi.org/10.1016/j.heliyon.2023.e23424>
- Mekonnen, A. A., Raghuvanshi, T. K., Suryabhadgavan, K. V., & Kassawmar, T. (2022). GIS-based landslide susceptibility zonation and risk assessment in complex landscape: A case of Beshilo watershed, northern Ethiopia. *Environmental Challenges*, 8(August), 1–17. <https://doi.org/10.1016/j.envc.2022.100586>
- Meyer, J., Herring, M., McDowell, C., Lansing, J., Brower, C., Schuch, F., Smith, L., Tully,

- M., Martin, J., Caswell, S., Cortes, N., & Boolani, A. (2020). Joint prevalence of physical activity and sitting time during COVID-19 among US adults in April 2020. *Prev. Med. Rep.*, 20(101256), 101256. <https://doi.org/10.1016/j.pmedr.2020.101256>
- Mezughi, T. H., Akhir, J. M., Rafek, A. G., & Abdullah, I. (2011). Landslide susceptibility assessment using frequency ratio model applied to an area along the E-W highway (Gerik-Jeli). *American Journal of Environmental Sciences*, 7(1), 43–50. <https://doi.org/10.3844/ajessp.2011.43.50>
- Mihai, N., Mihai, C. M., Cosmin, C., Nicusor, N., Georgiana, V., & Valeriu, S.-L. (2020). River-landslide erosion interaction assessed through LiDAR and UAV SfM high-resolution DEMs, SAR and photogrammetry. *EGUSPHERE*. <https://doi.org/10.5194/EGUSPHERE-EGU2020-5451>
- Miller, S., & Degg, M. (2012). Landslide Susceptibility Mapping in North-East Wales. *Geomatics, Natural Hazards and Risk*, 3(2), 133–159. <https://doi.org/10.1080/19475705.2011.600778>
- Mills, A. J., & Rushton, D. (2024). *Risk-based approach to peatland restoration and peat instability*. <https://www.nature.scot/doc/naturescot-research-report-1259-risk-based-approach-peatland-restoration-and-peat-instability>
- Misbahudin, M. (2020). Landslide susceptibility analysis in Kabandungan District and Salak Geothermal Field, West Java. *Jurnal Geografi Lingkungan Tropik*, 4(2). <https://doi.org/10.7454/jglitrop.v4i2.75>
- Mohd-Azlan, J., Yong, J. Y., Mohd Hazzrol, N. N., Pengiran, P., Atong, A., & Abdul Aziz, S. (2022). Local hunting practices and perceptions regarding the distribution and ecological role of the Large Flying Fox (Chiroptera: Pteropodidae: Pteropus vampyrus) in western Sarawak, Malaysian Borneo. *Journal of Threatened Taxa*, 14(1), 20387–

20399. <https://doi.org/10.11609/jott.6977.14.1.20387-20399>
- Mondini, A. C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications*, 14(1). <https://doi.org/10.1038/s41467-023-38135-y>
- Moore, D., Notz, W. I., & Fligner, M. A. (2011). *The basic practice of statistics* (6th ed.). W.H Freeman and Company.
- Moragues, S., Lenzano, M. G., Jeanneret, P., Gil, V., & Lannutti, E. (2024). Landslide Susceptibility Mapping in the Northern part of Los Glaciares National Park, Southern Patagonia, Argentina using remote sensing, GIS and frequency ratio model. *Quaternary Science Advances*, 13(July 2023), 1–18. <https://doi.org/10.1016/j.qsa.2023.100146>
- Moresi, F. V., Maesano, M., Collalti, A., Sidle, R. C., Matteucci, G., & Mugnozza, G. S. (2020). Mapping landslide prediction through a GIS-based model: A case study in a catchment in southern Italy. *Geosciences (Switzerland)*, 10(8), 1–22. <https://doi.org/10.3390/geosciences10080309>
- Naceur, H. A., Abdo, H. G., Igmoullan, B., Namous, M., Almohamad, H., Al Dughairi, A. A., & Al-Mutiry, M. (2022). Performance assessment of the landslide susceptibility modelling using the support vector machine, radial basis function network, and weight of evidence models in the N'fis river basin, Morocco. *Geoscience Letters*, 9(1). <https://doi.org/10.1186/s40562-022-00249-4>
- Najar, I. A., Ahmadi, R., Khalik, Y. K. A., Taib, S. N. L., Sutan, N. B. M., & Ramli, N. H. Bin. (2023). Soil suffusion under the dual threat of rainfall and seismic vibration. *International Journal of Design and Nature and Ecodynamics*, 18(4), 849–860. <https://doi.org/10.18280/ijdne.180411>
- Nguyen, V. V., Pham, B. T., Vu, B. T., Prakash, I., Jha, S., Shahabi, H., Shirzadi, A., Ba, D.

- N., Kumar, R., Chatterjee, J. M., & Bui, D. T. (2019). Hybrid Machine Learning approaches for landslide susceptibility modeling. *Forests*, 10(2). <https://doi.org/10.3390/f10020157>
- Nhu, V. H., Mohammadi, A., Shahabi, H., Ahmad, B. Bin, Al-Ansari, N., Shirzadi, A., Geertsema, M., Kress, V. R., Karimzadeh, S., Kamran, K. V., Chen, W., & Nguyen, H. (2020). Landslide detection and susceptibility modeling on cameron highlands (Malaysia): A comparison between random forest, logistic regression and logistic model tree algorithms. *Forests*, 11(8). <https://doi.org/10.3390/F11080830>
- Nnanwuba, U. E., Qin, S., Adeyeye, O. A., Cosmas, N. C., Yao, J., Qiao, S., Jingbo, S., & Egwuonwu, E. M. (2022). Prediction of spatial likelihood of shallow landslide using GIS-based Machine Learning in Awgu, Southeast Nigeria. *Sustainability (Switzerland)*, 14(19). <https://doi.org/10.3390/su141912000>
- Olden, J., Joy, M. ., & Death, R. . (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178, 389–397. [https://doi.org/10.1016/S0304-3800\(04\)00156-5](https://doi.org/10.1016/S0304-3800(04)00156-5)
- Park, H. J., Jang, J. Y., & Lee, J. H. (2017). Physically based susceptibility assessment of rainfall-induced shallow landslides using a fuzzy point estimate method. *Remote Sensing*, 9(5). <https://doi.org/10.3390/rs9050487>
- Patil, A. S., & Panhalkar, S. S. (2023). Remote sensing and GIS-based Landslide Susceptibility Mapping using LNRF method in part of Western Ghats of India. *Quaternary Science Advances*, 11(March), 1–15. <https://doi.org/10.1016/j.qsa.2023.100095>
- Peng, T., Chen, Y., & Chen, W. (2022). Landslide susceptibility modeling using remote sensing data and random subspace-based functional tree classifier. *Remote Sensing*,

- 14(19). <https://doi.org/10.3390/rs14194803>
- Pham, B. T., Shirzadi, A., Shahabi, H., Omidvar, E., Singh, S. K., Sahana, M., Asl, D. T., Ahmad, B. Bin, Quoc, N. K., & Lee, S. (2019). Landslide susceptibility assessment by novel hybrid Machine Learning algorithms. *Sustainability (Switzerland)*, 11(16). <https://doi.org/10.3390/su11164386>
- QGIS Development Team. (2023). *QGIS desktop 3.22 user guide*. <https://docs.qgis.org/3.22/pdf/en/QGIS-3.22-DesktopUserGuide-en.pdf>
- Rabie, M. (2014). Comparison study between traditional and finite element methods for slopes under heavy rainfall. *HBRC Journal*, 10(2), 160–168. <https://doi.org/10.1016/j.hbrcj.2013.10.002>
- Rahim, A. F. A., Rafek, A. G. M., Serasa, A. S., Jaapar, A. R., Lai, G. T., Roslee, R., Ern, L. K., Huy, N. X., & Van Xuan, T. (2023). A review of rock slope stability assessment practice in malaysia. *Sains Malaysiana*, 52(2), 399–416. <https://doi.org/10.17576/jsm-2023-5202-07>
- Rahman, H. A., & Mapjabil, J. (2017). Landslides disaster in Malaysia: An overview. *Health & the Environment Journal*, 8(1), 58–71. <http://psasir.upm.edu.my/id/eprint/54691/>
- Raj, A. S., Henrietta, H. M., & Angelena, J. P. (2023). Rainfall data classification using Mann-Kendall test statistics associated with Neuro Fuzzy technique: A case study of Chennai district. *International Journal of Hybrid Intelligent Systems*, 19(1–2), 95–109. <https://doi.org/10.3233/HIS-230010>
- Rama Krishna, P. S. V., Shivakoti, B. R., Dewi, A., Arambepola, N. M. S. ., Dwi Kartiko, R., Jayasinghe, S., Basnayake, S., Towashiraporn, P., & Keaokiriya, P. (2021). *Guidelines on integrating climate change projection into landslide risk assessment & mapping*. Association of Southeast Asian Nations (ASEAN).

<https://asean.org/book/guideline-on-integrating-climate-change-projection-into-landslide-risk-assessments-and-mapping-at-the-river-basin-level/>

- Ravanel, L., Duvillard, P. A., Astrade, L., Faug, T., Deline, P., Berthet, J., Cathala, M., Magnin, F., Baratier, A., & Bodin, X. (2023). The Tacconnaz rockfall (Mont-Blanc Massif, European Alps) of November 2018: A complex and at-risk rockwall-glacier-torrent morphodynamic continuum. *Applied Sciences (Switzerland)*, 13(17). <https://doi.org/10.3390/app13179716>
- Rosly, M. H., Mohamad, H. M., Bolong, N., & Harith, N. S. H. (2022). An overview: Relationship of geological condition and rainfall with landslide events at East Malaysia. *Trends in Sciences*, 19(8). <https://doi.org/10.48048/tis.2022.3464>
- Rudra, R. R., & Sarkar, S. K. (2023). Artificial neural network for flood susceptibility mapping in Bangladesh. *Heliyon*, 9(6). <https://doi.org/10.1016/j.heliyon.2023.e16459>
- Sa'adi, Z., Shahid, S., Ismail, T., Chung, E. S., & Wang, X. J. (2019). Trends analysis of rainfall and rainfall extremes in Sarawak, Malaysia using modified Mann–Kendall test. *Meteorology and Atmospheric Physics*, 131(3), 263–277. <https://doi.org/10.1007/s00703-017-0564-3>
- Sadiki, M., Manaouch, M., Aghad, M., Batchi, M., & Al Karkouri, J. (2023). Identifying landslides prone areas using GIS-based fuzzy analytical hierarchy process model in Ziz Upper Watershed (Morocco). *Ecological Engineering and Environmental Technology*, 24(1), 67–83. <https://doi.org/10.12912/27197050/154916>
- Saikh, N. I., & Mondal, P. (2023). GIS-based Machine Learning algorithm for flood susceptibility analysis in the Pagla river basin, Eastern India. *Natural Hazards Research*, 3(3), 420–436. <https://doi.org/10.1016/j.nhres.2023.05.004>
- Salmanfarsi, A. F., Awang, H., & Ali, M. I. (2019). A review on the development of rock

- slope stability assessment in Malaysia. *IOP Conference Series: Earth and Environmental Science*, 365(1). <https://doi.org/10.1088/1755-1315/365/1/012063>
- Salmanfarsi, A. F., Awang, H., & Ali, M. I. (2020). Rock mass classification for rock slope stability assessment in Malaysia: A review. *IOP Conference Series: Materials Science and Engineering*, 712(1). <https://doi.org/10.1088/1757-899X/712/1/012035>
- Sarawak Government. (n.d.). *Sarawak Administrative Division and Districts*. Sarawak Government. Retrieved April 29, 2024, from [https://sarawak.gov.my/web/home/article\\_view/358/298/](https://sarawak.gov.my/web/home/article_view/358/298/)
- Sassa, K., & Canuti, P. (2008). *Landslides Disaster Risk Reduction* (S. Kyoji, P. Canuti, N. Casagli, L. Highland, F. Wang, & S. Tsuchiya (eds.); 1.). Springer-Verlag.
- Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Schuster, R. L., & Highland, L. M. (2017). *Impact of landslides and innovative landslide-mitigation measures on the natural environment*. <https://pubs.usgs.gov/publication/sir20175013>
- Schuster, R. L., & Turner, A. K. (1996). Socioeconomic significance of landslides. In *Transportation Research Board Special Report* (247th ed., pp. 12–35). National Academy Press.
- Selamat, S. N., Abd Majid, N., & Mohd Taib, A. (2023). A comparative assessment of sampling ratios using artificial neural network (ANN) for landslide predictive model in Langat River Basin, Selangor, Malaysia. *Sustainability (Switzerland)*, 15(861), 1–21. <https://doi.org/10.3390/su15010861>
- Selamat, S. N., Majid, N. A., Taha, M. R., & Osman, A. (2022a). Application of

- Geographical Information System (GIS) using artificial neural networks (ANN) for landslide study in Langat Basin, Selangor. *IOP Conference Series: Earth and Environmental Science*, 1064(1), 1–8. <https://doi.org/10.1088/1755-1315/1064/1/012052>
- Selamat, S. N., Majid, N. A., Taha, M. R., & Osman, A. (2022b). Landslide susceptibility model using artificial neural network (ANN) approach in Langat River Basin, Selangor, Malaysia. *Land*, 11(6). <https://doi.org/10.3390/land11060833>
- Senanayake, S., & Pradhan, B. (2022). Predicting soil erosion susceptibility associated with climate change scenarios in the Central Highlands of Sri Lanka. *Journal of Environmental Management*, 308. <https://doi.org/10.1016/j.jenvman.2022.114589>
- Senanayake, S., Pradhan, B., Alamri, A., & Park, H. J. (2022). A new application of deep neural network (LSTM) and RUSLE models in soil erosion prediction. *Science of the Total Environment*, 845. <https://doi.org/10.1016/j.scitotenv.2022.157220>
- Shahabi, H., Rahimzad, M., Ghorbanzadeh, O., Piralilou, S. T., Blaschke, T., Homayouni, S., & Ghamisi, P. (2022). Rapid mapping of landslides from Sentinel-2 data using unsupervised deep learning. *2022 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium, M2GARSS 2022 - Proceedings*, 17–20. <https://doi.org/10.1109/M2GARSS52314.2022.9840273>
- Shahin, M. A. (2016). State-of-the-art review of some artificial intelligence applications in pile foundations. *Geoscience Frontiers*, 7(1), 33–44. <https://doi.org/10.1016/j.gsf.2014.10.002>
- Shapiro, A. F., Scott Pflumm, J., & DeFilippo, T. A. (1999). The inner workings of neural networks and genetic algorithms. *Actuarial Research Clearing House*, 1, 415–426.
- Sharma, V. (2022). A study on data scaling methods for Machine Learning. *International*

- Journal for Global Academic & Scientific Research*, 1(1).  
<https://doi.org/10.55938/ijgasr.v1i1.4>
- Sheng, Y., Xu, G., Jin, B., Zhou, C., Li, Y., & Chen, W. (2023). Data-driven landslide spatial prediction and deformation monitoring: A case study of Shiyan City, China. *Remote Sensing*, 15(21). <https://doi.org/10.3390/rs15215256>
- Shreevastav, B. B., Tiwari, K. R., Mandal, R. A., & Singh, B. (2022). “Flood risk modeling in southern Bagmati corridor, Nepal” (A study from Sarlahi and Rautahat, Nepal). *Progress in Disaster Science*, 16(October), 1–10.  
<https://doi.org/10.1016/j.pdisas.2022.100260>
- Singh, A., Dhiman, N., Niraj, K. C., & Shukla, D. P. (2024). Ensembled transfer learning approach for error reduction in Landslide Susceptibility Mapping of the data scare region. *Sci. Rep.*, 14(1), 29060. <https://doi.org/10.1038/s41598-024-76541-4>
- Singh, H., Huat, B. B. K., & Jamaludin, S. (2008). Slope assessment systems: A review and evaluation of current techniques used for cut slopes in the mountainous terrain of West Malaysia. *EJGE*, 13(E), 1–23.
- Sur, U., Singh, P., Meena, S. R., & Singh, T. N. (2022). Predicting landslides susceptible zones in the Lesser Himalayas by ensemble of per pixel and object-based models. *Remote Sensing*, 14(8). <https://doi.org/10.3390/rs14081953>
- Suwanno, P., Yaibok, C., Pornbunyanon, T., Kanjanakul, C., Buathongkhue, C., Tsumita, N., & Fukuda, A. (2023). GIS-based identification and analysis of suitable evacuation areas and routes in flood-prone zones of Nakhon Si Thammarat municipality. *IATSS Research*, 47(3), 416–431. <https://doi.org/10.1016/j.iatssr.2023.08.004>
- Swamynathan, M. (2017). Mastering Machine Learning with Python in six steps. In T. Green (Ed.), *Mastering Machine Learning with Python in Six Steps* (1st ed.). Apress.

<https://doi.org/10.1007/978-1-4842-2866-1>

- Taherynia, M. H., Mohammadi, M., & Ajalloeian, R. (2014). Assessment of slope instability and risk analysis of road cut slopes in Lashotor Pass, Iran. *Journal of Geological Research*, 2014, 1–12. <https://doi.org/10.1155/2014/763598>
- Tay, J. E., & Selaman, O. S. (2011). A study on the rainfall and landslides along Sarawak Road using the antecedent rainfall analysis. *UNIMAS E-Journal of Civil Engineering*, 2(1). <https://publisher.unimas.my/ojs/index.php/JCEST/article/view/80/57>
- Tehrani, F. S., Calvello, M., Liu, Z., Zhang, L., & Lacasse, S. (2022). Machine Learning and landslide studies: recent advances and applications. *Natural Hazards*. <https://doi.org/10.1007/s11069-022-05423-7>
- Tengtrairat, N., Woo, W. L., Parathai, P., Aryupong, C., Jitsangiam, P., & Rinchumphu, D. (2021). Automated landslide-risk prediction using web GIS and Machine Learning models. *Sensors (Basel, Switzerland)*, 21(13). <https://doi.org/10.3390/s21134620>
- Trisnawati, D., Najib, Hidayatillah, A. S., & Najib. (2022). The relationship of lithology with landslide occurrences in Banyumanik and Tembalang districts, Semarang City. *IOP Conference Series: Earth and Environmental Science*, 1047(1). <https://doi.org/10.1088/1755-1315/1047/1/012026>
- Tuffour, H. O., Abubakari, A., Agbeshie, A. A., Khalid, A. A., Tetteh, E. N., Keshavarzi, A., Bonsu, M., Quansah, C., Oppong, J. C., & Danso, L. (2019). Pedotransfer functions for estimating saturated hydraulic conductivity of selected benchmark soils in Ghana. *Asian Soil Research Journal*, May, 1–11. <https://doi.org/10.9734/asrj/2019/v2i230046>
- Ul Haq, I., Gondal, I., Vamplew, P., & Brown, S. (2019). Categorical features transformation with compact one-hot encoder for fraud detection in distributed environment. *Communications in Computer and Information Science*, 996, 69–80.

[https://doi.org/10.1007/978-981-13-6661-1\\_6](https://doi.org/10.1007/978-981-13-6661-1_6)

- Ullah, I., Aslam, B., Shah, S. H. I. A., Tariq, A., Qin, S., Majeed, M., & Havenith, H. B. (2022). An integrated approach of Machine Learning, remote sensing, and GIS data for the Landslide Susceptibility Mapping. *Land*, 11(8). <https://doi.org/10.3390/land11081265>
- Ullah, K., Wang, Y., Fang, Z., Wang, L., & Rahman, M. (2022). Multi-hazard susceptibility mapping based on Convolutional Neural Networks. *Geoscience Frontiers*, 13(5), 101425. <https://doi.org/10.1016/j.gsf.2022.101425>
- Utomo, D., Hu, L. C., & Hsiung, P. A. (2020). Deep neural network-based data reconstruction for landslide detection. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3119–3122. <https://doi.org/10.1109/IGARSS39084.2020.9323124>
- Vakhshoori, V., Pourghasemi, H. R., Zare, M., & Blaschke, T. (2019). Landslide Susceptibility Mapping using GIS-based data mining algorithms. *Water (Switzerland)*, 11(11), 1–30. <https://doi.org/10.3390/w11112292>
- Varnes, D. J., & Cruden, D. M. (1996). Landslide types and processes. In *LANDSLIDES: INVESTIGATION AND MITIGATION* (Vol. 247). Transportation Research Board.
- Vidal-Paez, P., Clavero, J., Droguett, B., Perez-Martinez, W., Briceno-De-Urbaneja, I., & Oliva, P. (2020). Landslide susceptibility using remote sensing data GIS in a high Andean area of central Chile. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 6604–6607. <https://doi.org/10.1109/IGARSS39084.2020.9324317>
- Wakwoya, M. B., Woldeyohannis, W. H., & Yimamu, F. K. (2023). Characterization and classification of soils of upper Hoha sub-watershed in Assosa District, Western

- Ethiopia. *Heliyon*, 9(4), e14866. <https://doi.org/10.1016/j.heliyon.2023.e14866>
- Wang, H., Zhang, L., Yin, K., Luo, H., & Li, J. (2021). Landslide identification using Machine Learning. *Geoscience Frontiers*, 12(1), 351–364. <https://doi.org/10.1016/j.gsf.2020.02.012>
- Winter, M. G., Shearer, B., Palmer, D., Peeling, D., Harmer, C., & Sharpe, J. (2016). The economic impact of landslides and floods on the road network. *Procedia Engineering*, 143, 1425–1434. <https://doi.org/10.1016/j.proeng.2016.06.168>
- Wissmann, M., & Toutenburg, H. (2007). *Role of categorical variables in multicollinearity in linear regression model*. Ludwig Maximilian University of Munich. <http://www.stat.uni-muenchen.de>
- Xiao, B., Zhao, J., Li, D., Zhao, Z., Zhou, D., Xi, W., & Li, Y. (2022). Combined SBAS-InSAR and PSO-RF algorithm for evaluating the susceptibility prediction of landslide in complex mountainous area: A case study of Ludian County, China. *Sensors*, 22(20), 1–24. <https://doi.org/10.3390/s22208041>
- Yakub, I., Kueh, A. B. H., Pineda De La O, E. A., Rahman, M. R., Barawi, M. H., Abdullah, M. O., Amran, M., Fediuk, R., & Vatin, N. I. (2022). Employing an artificial neural network in correlating a hydrogen-selective catalytic reduction performance with crystallite sizes of a biomass-derived bimetallic catalyst. *Catalysts*, 12(7). <https://doi.org/10.3390/catal12070779>
- Yalcin, A. (2007). The effects of clay on landslides: A case study. *Applied Clay Science*, 38(1–2), 77–85. <https://doi.org/10.1016/j.clay.2007.01.007>
- Yao, J., Zhang, X., Luo, W., Liu, C., & Ren, L. (2022). Applications of stacking/blending ensemble learning approaches for evaluating flash flood susceptibility. *International Journal of Applied Earth Observation and Geoinformation*, 112(July), 102932.

<https://doi.org/10.1016/j.jag.2022.102932>

- Yilmaz, E. O., Teke, A., & Kavzoglu, T. (2022). Performance evaluation of depthwise separable CNN and random forest algorithms for landslide susceptibility prediction. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022-July, 5477–5480. <https://doi.org/10.1109/IGARSS46834.2022.9883954>
- 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>
- Zemp, M., Frey, H., Gärtnert-Roer, I., Nussbaumer, S. U., Hoelzle, M., & Paul, F. (2015). Historically unprecedented global glacier decline in the early 21st century. *Journal of Glaciology*, 61(228), 745–762. [https://doi.org/https://doi.org/10.3189/2015JoG15J017](https://doi.org/10.3189/2015JoG15J017)
- Zêzere, J. L., Pereira, S., Melo, R., Oliveira, S. C., & Garcia, R. A. C. (2017). Mapping landslide susceptibility using data-driven methods. *Science of the Total Environment*, 589, 250–267. <https://doi.org/10.1016/j.scitotenv.2017.02.188>
- Zhao, Z., Xu, T., Min, J., Zhang, Q., Li, W., & He, C. (2022). Multi-temporal remote sensing analysis of landslide evolution around the Yingxiu hydropower station after the 2008 Wenchuan earthquake, in Sichuan Province, Southwest China. *2022 9th International Forum on Electrical Engineering and Automation, IFEEA 2022*, 495–498. <https://doi.org/10.1109/IFEEA57288.2022.10038022>
- Zulkafli, S. A., Abd Majid, N., & Rainis, R. (2023). Spatial analysis on the variances of landslide factors using geographically weighted logistic regression in Penang Island, Malaysia. *Sustainability (Switzerland)*, 15(1). <https://doi.org/10.3390/su15010852>
- Zulkifli, Zainuddin, Z., Niswar, M., & Nirwana, H. (2023). Landslide susceptibility zoning

using fuzzy logic algorithm & GIS: Study case in Majene, West Sulawesi. 2023  
*International Conference on Computer, Control, Informatics and Its Applications*  
(IC3INA), 183–188. <https://doi.org/10.1109/IC3INA60834.2023.10285732>

## APPENDICES

### Appendix A: Slope Failure Points

| xcoord             | ycoord             | District | Source     |
|--------------------|--------------------|----------|------------|
| 110.5436409720156  | 1.1619361834323105 | Serian   | Site Visit |
| 110.34427318122995 | 1.2168827705073566 | Kuching  | Site Visit |
| 110.32776960559713 | 1.7372917094715865 | Kuching  | JMG        |
| 110.16173745616841 | 1.4993879980643867 | Bau      | JMG        |
| 110.0501716485025  | 1.3168627203857006 | Bau      | JMG        |
| 110.18517395126622 | 1.3767933772144136 | Bau      | JMG        |
| 110.14046819949564 | 1.4862206028171958 | Bau      | JMG        |
| 110.14506808151457 | 1.482295245731548  | Bau      | JMG        |
| 110.13677504324373 | 1.4315667616300953 | Bau      | JMG        |
| 110.07473693715863 | 1.4864558550213993 | Bau      | JMG        |
| 110.11880543684973 | 1.45522707278301   | Bau      | JMG        |
| 110.15080318465827 | 1.4067584009214973 | Bau      | JMG        |
| 110.17515440367254 | 1.5072213215256431 | Bau      | JMG        |
| 110.45854376092214 | 1.1272302423945648 | Serian   | JMG        |
| 110.49185526754901 | 1.163731922712594  | Serian   | JMG        |
| 110.55017693366928 | 1.1636846214576886 | Serian   | JMG        |
| 110.54587191379227 | 1.160959223423971  | Serian   | JMG        |
| 110.65975033691788 | 1.0492452145432971 | Serian   | JMG        |
| 110.5195238258275  | 0.9946350865410785 | Serian   | JMG        |
| 110.38362872760847 | 1.0952159684746745 | Serian   | JMG        |
| 110.5481900643521  | 1.0942642149270752 | Serian   | JMG        |
| 110.26827971876307 | 1.1114258155882466 | Kuching  | JMG        |
| 110.28079777114475 | 1.0998261650130812 | Kuching  | JMG        |
| 110.27689029689537 | 1.284185917053083  | Kuching  | JMG        |
| 110.27370998585262 | 1.28901629131979   | Kuching  | JMG        |
| 110.26911616219469 | 1.102214460637692  | Kuching  | JMG        |
| 110.27286811863115 | 1.3011659874920545 | Kuching  | JMG        |
| 110.29542196146484 | 1.3438628398364971 | Kuching  | JMG        |
| 110.27293910429461 | 1.2979849524804563 | Kuching  | JMG        |
| 110.2645574454333  | 1.106729588045076  | Kuching  | JMG        |

|                    |                    |         |     |
|--------------------|--------------------|---------|-----|
| 110.24226301242608 | 1.1866724604305963 | Kuching | JMG |
| 110.27532732698349 | 1.107534349092711  | Kuching | JMG |
| 110.27651920476292 | 1.0971241574442534 | Kuching | JMG |
| 110.27309941493105 | 1.1127751845392169 | Kuching | JMG |
| 110.4390031528494  | 1.5868826336789401 | Kuching | JMG |
| 110.3124737550508  | 1.4729116683448247 | Kuching | JMG |
| 110.3464491633072  | 1.5541652337311818 | Kuching | JMG |
| 110.35492723843883 | 1.5961853551112979 | Kuching | JMG |
| 110.38338509339292 | 1.5511529954085357 | Kuching | JMG |
| 110.32059000220038 | 1.3791426714198363 | Kuching | JMG |
| 110.20630992321412 | 1.6055579249975775 | Kuching | JMG |
| 110.20831673202035 | 1.604411097079457  | Kuching | JMG |
| 110.20840348692758 | 1.5955678041200334 | Kuching | JMG |
| 110.20170717062064 | 1.6014871973367066 | Kuching | JMG |
| 110.19977222352388 | 1.600665765091343  | Kuching | JMG |
| 110.33831400178346 | 1.7349105770831315 | Kuching | JMG |
| 110.33377811280087 | 1.7372941751461997 | Kuching | JMG |
| 110.34011174521946 | 1.7467764519704243 | Kuching | JMG |
| 110.33402324895965 | 1.7688564345958135 | Kuching | JMG |

Appendix B: Site Visit Reports

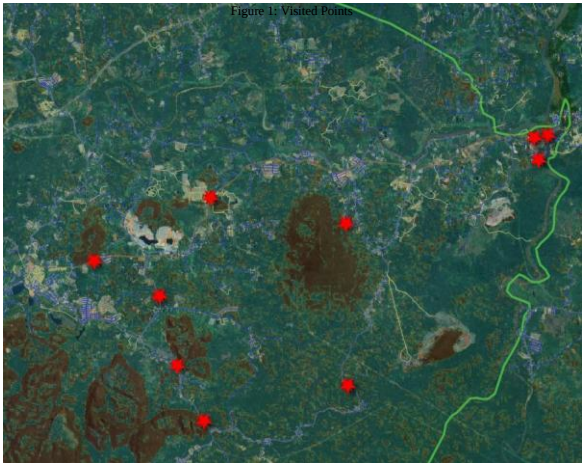
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Site Visit Report: Bau Date of Visit: 6/11/2023                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <p>BAU is home to the fairy cave and several other caves in the areas. The districts consist mostly of scattered towns. As the district is in the outskirts of Western Sarawak, development is still in progress, thus a large amount of heavy hauling trucks moves through the roads every day. The terrain of the district consists of a large amount of hill scattered throughout the district. A total of 10 locations were planned to be visited based on the accessibility as seen in Figure 1. However, due to the sudden change in weather conditions, only 6 were visited.</p> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <p>1. 1.45079051, 110.2782298</p> <p>The first point visited has a susceptibility rate of 0.344. The area is located next to a Kampung. Furthermore, even the area nearby the roads seems to have a high susceptibility rate. Looking around the susceptibility rate seems to be reasonable. Upon site visit, the road was observed to have quite a significant slope. Furthermore, at the red area nearby, it was observed that the area possibly has experienced landslides before, as a gabion wall was seen installed next to road. As the for the high susceptibility on the other side, the slope of the nearby hill seems to be the major indicator. The details of the first point can be seen in Figure 2.</p> |



*Figure 2: Point 1 Details*

2. 1.42965757, 110.2295993

The second point visited has a susceptibility rate of 0.07. However, it was very near to the red areas. Upon site visit, the road was quite flat relative to the terrain. As for the red areas, it was on the mountain nearby, which was Bung Muan. The mountain possibly has experienced rockfalls in the past as seen in the non-vegetated areas on the surface of the mountain.

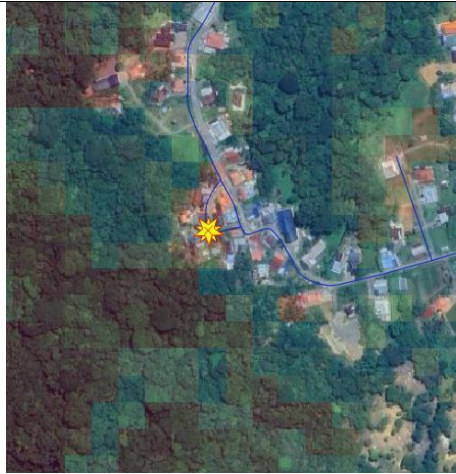


*Figure 3: Point 2 Details*

### 3. 1.435792262, 110.1968856

The third point visited has a susceptibility rate of 0.88, which was extremely high for a road. Upon site visit, the red areas are a cemetery, and it slope failure mitigation effort has been applied. As for the current conditions of the roads, it was relatively flat



IV. 1.420506888000000, 110.1689293

The fourth point has a susceptibility rate of 0.265. The large amount of read areas in the region is represented by the large hill nearby which has quite the steep slope, while the small bit of red on the bottom was a small hill.

Figure 5: Point 4 Details



4. 1.412135882, 110.1845509

The fifth point visited has achieved a susceptibility rate of 0.9. The road itself is considerably flat, however is located on a cliff like terrain. The red areas also include the hill nearby. Thankfully, a drain was installed nearby, possibly as a mitigation measure.





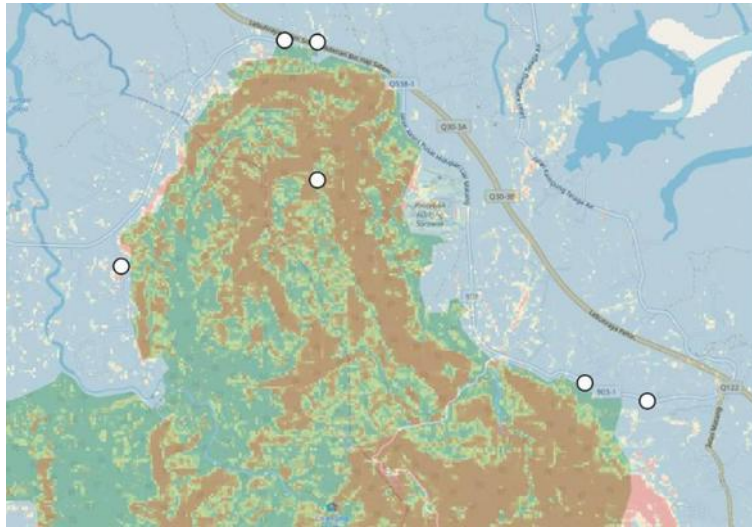
Figure 6: Point 5 Details

VI. 1.395412543000000, 110.188827399999994

The sixth point visited has achieved a susceptibility rate of 0.0005. The road was flat and located on a low elevation area. The red areas are on the mountain nearby, with the very steep slopes.



Figure 7: Point 6 Details



5. 1.609699, 110.2163

The location was nearby Pearl's Happy Farm with a susceptibility rate of 0.036. Past landslide's aftermath can be seen from the points with an estimated distance of 5km, at 120°-150°. The site land use was primarily for agricultural purposes, with high vegetation. The access road for the site situated at a considerably higher elevation than the valley next to it.



Figure 8: Point 1 Details

6. 1.611783, 110.2091

The location was nearby Meteor Recreational Park 1km away from the first site however having a slightly higher elevation, and accordingly both shares an almost similar susceptibility rate of 0.038. Past landslide's aftermath can be seen from the points with an estimated distance of 5km, at 120°-180°. The site land use was primarily for ecotourism purposes, with high vegetation. The access road for the site situated at a considerably higher elevation than the valley next to it.



*Figure 9: Point 2 Details*

7. 1.635073, 110.195442

The location was nearby Sekolah agama Matang, with a susceptibility level of 0.88. Past landslide's aftermath can barely be seen from the points with an estimated distance of more than 15km, at 90°-150°. The site land use was primarily for school, with high vegetation. The access road for the site situated at a considerably flat with the school next to it, however to the right, the road was located on a mild slope with bare grounds upon validation through satellite imagery.



*Figure 10: Point 3 Details*

8. 1.650597, 110.178336

The location was located at kubah's road to sungai rayu with 0.003 landslide susceptibility level, while the visible mountains were in the reds. Past landslide's aftermath can barely be seen from the points with an

estimated distance of more than 8km, at 90°-180°. The site land use was primarily for agricultural purposes, with high vegetation. The access road for the site situated at a considerably flat with the school next to it.



*Figure 11: Point 4 Details*

9. 1.651152,110.174547

The location was located at sungai rayu access road with 0.00005 chances of landslide, however the nearby mountain was predicted to have a high susceptibility of landslide on the peak, less on the lower cut region. No past landslide's aftermath was observed. The site land use was primarily for undisturbed jungle, with high vegetation. The access road for the site situated at a considerably flat as for the mountain range, it was at a considerably lower elevation as compared to the previous sites.



*Figure 12: Point 5 Details*

10. 1.625163, 110.15571

The site was nearby Matang Wildlife centre, the point of coordinate itself was predicted to have a high susceptibility level of 0.76. Recent landslide's aftermath can barely be seen clearly next to the road. The

site land use was primarily for quarry purposes, with medium vegetation for the undisturbed areas. The access road for the site was rolling.



*Figure 13: Point 6 Details*

#### Site Visit Report: Padawan (Kuching Landward)

Kuching landward has the highest density of mountain range per unit area in comparison to other parts of the study area, thus, it is considered as the highland areas of Western Sarawak, and eco-tourism activities has been determined to be on the rise in the area based on the amount of eco-tourism hotspots.

11. 110.34638258576747, 1.194514847554495

Point 1 had a susceptibility score of 0.65 based on the LSM which was considerably high as it was located next to the road. Visiting the location in real life, a minor landslide was observed. Furthermore, it was observed that the point location and the surrounding area was located at the peak of the road stretch.



*Figure 14: Point 1 Details*

12. 110.36214915970235, 1.2025950144146094

Point 2 was very alarming, as a large portion of the road was located on high-risk areas. As for the road, it seems that it was in an area which resembles a hill side, as in the sudden change of vegetation height, and the slope angle for the road itself was quite steep.



*Figure 15: Point 2 Details*

13. 110.35219837723643, 1.2154102084109177

Similar to point 2, a large portion of Point 3's road was located on high-risk, with a susceptibility score of 0.92. Visiting the site reveals that point 3 was located in between two hills, which was probably a single hill before the cut-and-fill for the road construction took place. Furthermore, a recent landslide was observed at Point 3.



*Figure 16: Point 3 Details*

14. 110.27221472353766, 1.296566871347397

Point 4 was observed to not be in direct contact with the high-risk areas, however there was a bend on the road that was on the reds. The large high-risk areas were determined to be a mountain next to the road, and the high-risk area on the road was where the road had a sudden change in elevation.

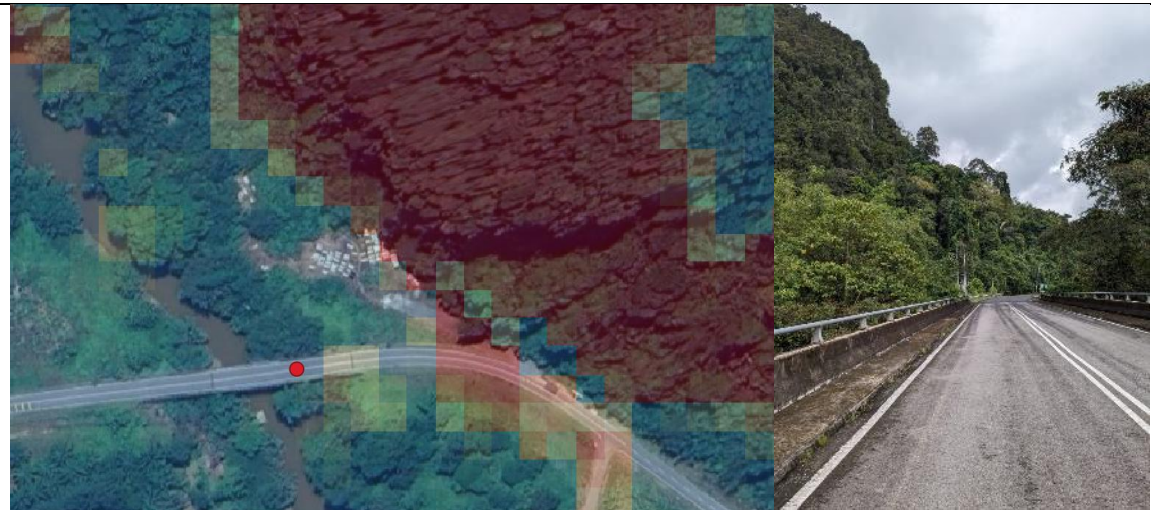


*Figure 20: Point 4 Details*

Site Visit Report: Serian Date of Visit: 30/1/2024.

The site visit to Serian was conducted on 30/1/2024. Several locations which were determined to be in the vicinity of high-risk areas, which were Kampung Bunga, Kampung Kakai, Kampung Tangga Mawang, and Kampung Slabi. The following were conditions in the visited areas:

15. Kampung Bunga

At point 1 the susceptibility score was 0.33. The road had a slight slope located on a short (below 10m) cliff on both sides of the road. The actual point was not visited as it was guarded by dogs. However, the road was located on a small hill and can be seen by comparing the rooftop of the nearby house. Going further down the road provides details on the large number of high areas in the map. It was determined that the high-risk areas correspond to the hills nearby.



*Figure 21: Point 1 Details, and site details*

Point 2 had the highest risk of landslides amongst the three points in Kampung Bunga, which was 0.61. Point 2 was in the backroads of Kampung Bunga, which is mainly used by the villagers to go their orchard. A retaining wall was seen at the point, indicating a landslide most likely had occurred in the specific point which can be seen. Point 2 was a valley like terrain, as a stream was seen nearby, and judging by the elevation profile, it is likely that the landslide that had occurred was a medium-scaled landslide.



*Figure 22: Point 2 Details*

Point 3 was the furthest accessible through car area in Kampung Bunga. From the LSM, point 3 was determined to have a susceptibility score of 0.33. The road at point 3 was considerably steep. Point 3 itself was a small albeit high hill, which provides a hiking trail entry to the large high-risk areas nearby which was determined to be Bung Sadong that can be viewed.



Figure 23: Point 3 Details

## 16. Kampung Kakai

Five points were visited throughout the stretch of Kampung Kakai's road.

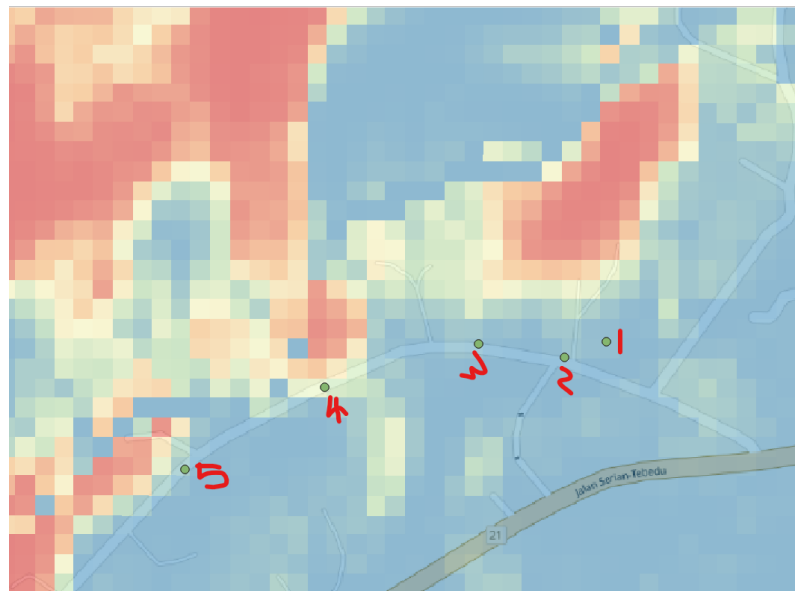


Figure 24: Kampung Kakai Points

Point 1 and Point 2 located nearby each other where each point was determined to have a susceptibility score of 0.03 to 0.06 respectively. The point was chosen as it provides a clear image of the high-risk area

nearby. The high-risk area was determined to be the hill nearby the village and point 1 and point 2 were flat grounds. Point 3 provides a different view of the hill from the roadside. Point 4 had a susceptibility score of 0.42. Point 4 seems to be located on the end for the slope of the hill. It was a highly elevated road with a slightly steep angle. Furthermore, a natural drain was observed at point 4. Point 5 is the starting point for the steeper side of the slope as it is located next to a hill



*Figure 25: Kampung Kakai Point 1 to 5*

## 17. Kampung Tangga Mawang

Kampung Tangga Mawang is a mountain village in Serian, where a few points were visited.

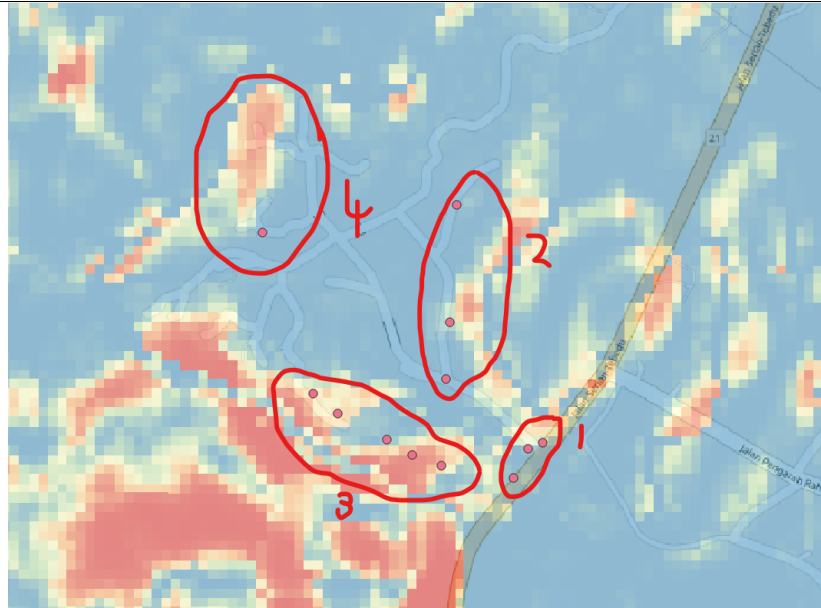


*Figure 26: Kampung Tangga Mawang Points*

Group 1 was located at the entrance of Kampung Tangga Mawang, and alongside the main road. The high susceptibility area to the entrance is explained by the hill nearby. However, the high-risk area on the road seemingly has been removed from the road to construct the road. The high-risk area along the main road which was a small hill nearby. Further down the road from group 1, the high-risk area was explained by the presence of a large hill. Group 2 is the interior of the village. The high-risk area was explained by the presence of the hill. Group 3 is the most exposed to landslide area amongst all the visited points. Several recent landslide cases were observed in the area. Moving further up shows that the group was located on a cliff next to a stream. Group 4 provides a view of the high-risk area nearby. It can be observed how the LSM correlates to the real-life situation as it moves from a gentle to a steeper slope.



*Figure 27: Kampung Tangga Mawang Group 1-3 Site Conditions*



*Figure 27: Kampung Tangga Mawang Group 4 Site Conditions*

## 18. Kampung Slabi

Kampung Slabi is another mountain village in Serian, and several points were visited.

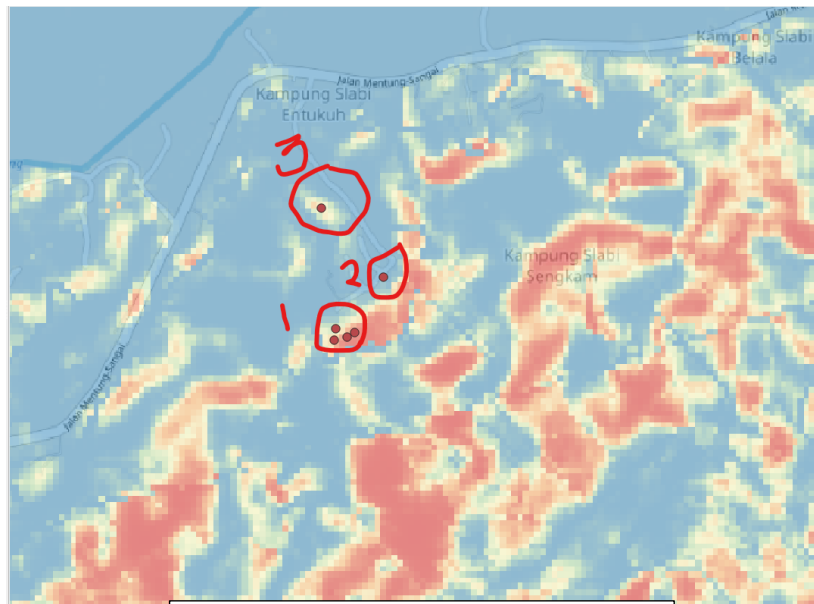


*Figure 28: Kampung Slabi Visited Points*

Group 1 was the highest point of the village, and a very highly elevated cliff with very steep slope. The area is very dangerous and prone to landslide as it is highly elevated with steep slopes.



*Figure 29: Group 1 Site Conditions*

Group 2 provides a view of the high-risk area nearby which was a hill. At group 3, the entrance of the village, a landslide warning sign. Furthermore, on point of group 3, is a small hill with a gentle slope.



*Figure 30: Group 2 and 3 Site Conditions*

Site Visit Report: Santubong Date of Visit: 26/10/2023

Santubong is home to the Santubong's National Park and located next to sea. The site visit was conducted to see the real-life conditions of the high susceptibility areas within the Landslide Susceptibility Map that has been developed. A total of 4 locations were visited based on the accessibility

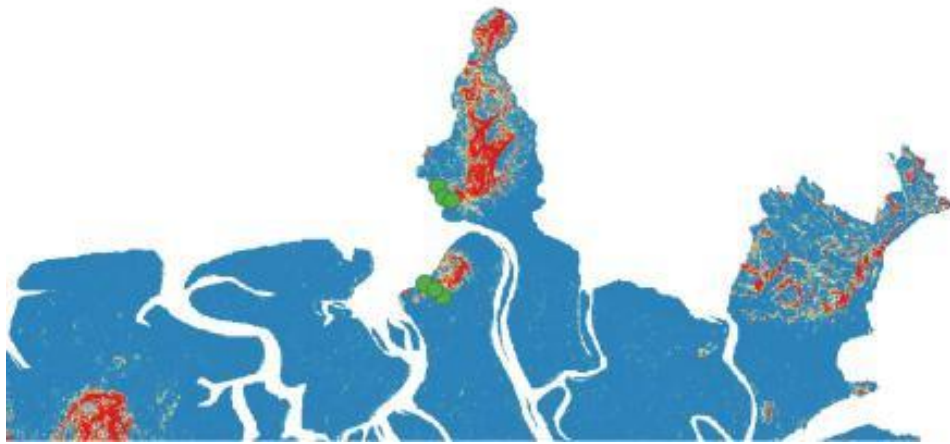


*Figure 31: Santubong Visited Points*

1. 1.732261523, 110.3151208

The first point visited has a susceptibility rate of 0.34, which is quite high for a road. Upon site visit, the road was observed to have quite a significant slope. Furthermore, at the red area nearby, it was observed that there was a recent minor landslide.

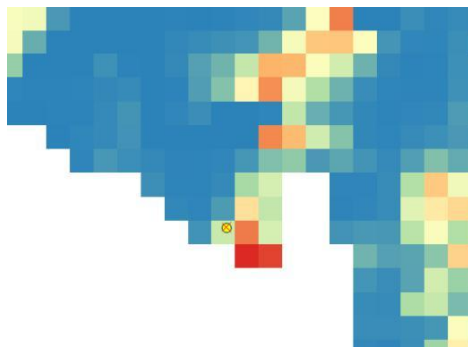


*Figure 32: Point 1 Site Conditions*

2. 1.727742206, 110.3176326

The second point visited has a susceptibility rate of 0.01. Upon site visit, the road was quite flat relative to the terrain. For the red areas, it was in the hills nearby.



*Figure 33: Point 2 Site Conditions*

3. 1.726453641, 110.3209381

The third point visited has a susceptibility rate of 0.92, which was extremely high for a road. Upon site visit, the road was observed to be quite steep, but not enough to justify the high susceptibility rate. However, upon walking around the susceptibility rate and why the exact was in the end of the high susceptibility areas was justified. It was located on a cliff.



*Figure 34: Point 3 Site Conditions*

4. 1.686862935, 110.3171844

The final point has a susceptibility rate of 0.003. Upon visit, the susceptibility rate was validated as the point was located on a flat area at a newly built residential area. However, for the high susceptibility areas, it was located at a hill nearby.



## Appendix D: ANN model Rstudio IDE

R is a programming language used for data science purposes, which includes Machine Learning.

The following are the base of the ANN model.

### 1. Data Prep

```
# change directory based on your own working directory  
setwd("C:/Users/user/Desktop/by district")  
  
.libPaths("C:/Users/user/Desktop/by district/library")
```

### 2. Importing Data

```
data_train <- read.csv("C:/Users/user/Desktop/FINALE RUN/ANN  
model/training.csv", header = T)
```

```
data_train <-data.frame(data_train)
```

3. Data preprocessing command

```
seq(0,361,45), right=FALSE, labels=c("a","b","c","d","e","f","g","h"))
```

```
breaks = c(-10, -0.00000000001, 0, 10), labels=c("concave","flat","convex"))
```

```
flags=data.frame(Reduce(cbind,lapply(levels(Lithologyr),function(x){(Lithology  
r == x)*1})))) names(flags) = levels(Lithologyr)
```

```
data_train = cbind(data_train, flags)
```

```
data_train <-(na.omit(data_train))
```

```
as.data.frame(table(data_train$slide))
```

```
maxs <- apply(data_train, 2, max)
```

4. Using the ANN model in prediction

```
head(scaled_test,6) ##select which column has the training/testing
```

```
compute.output_Tr=predict(nn.ce,scaled_train[,c(-6)])
```

## Appendix E: GEE IDE

GEE code editor uses JavaScript as the programming language. Being developed by Google, it has seamless integration with the Google's environment including Google Drive, where the extracted data is stored. Numerous remote sensing data agencies from around the world provides free access to their data in GEE. Some data have not been updated in a while such as elevation, whilst other data are updated from a few weeks to a few days. The data supplied in GEE consisted of an Image or an Image Collection, with

different methods to extract the data respectively.

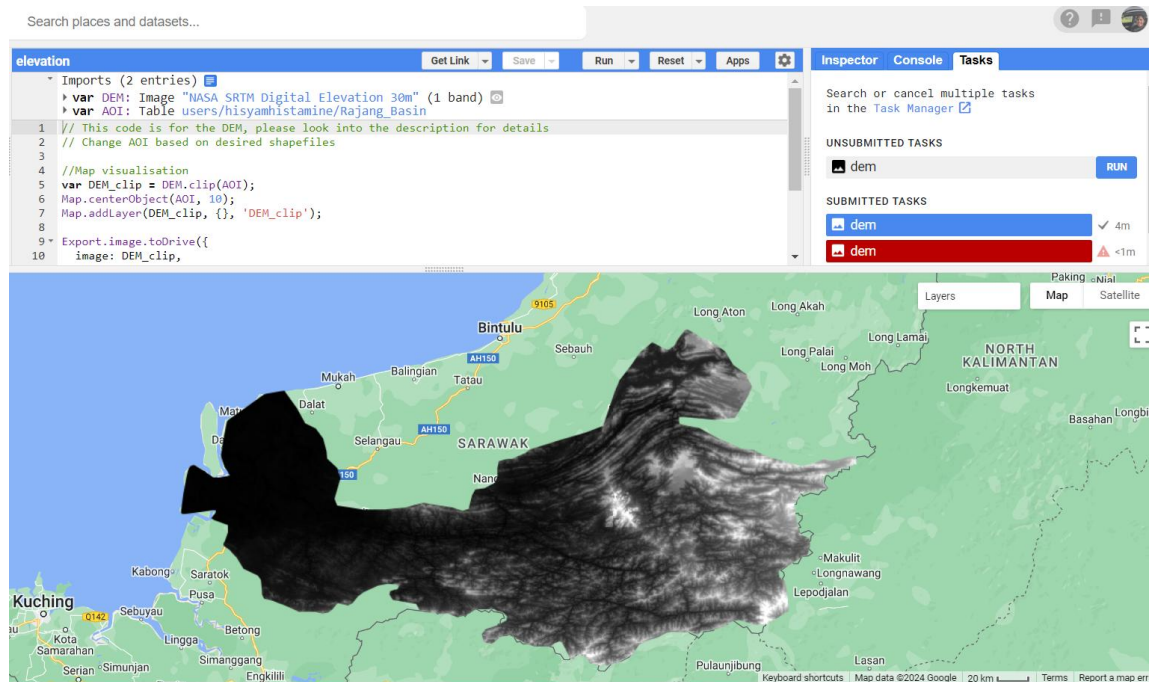


Image Process:

```
//Map visualisation
```

```
var DEM_clip = DEM.clip(AOI);
```

```
Map.centerObject(AOI, 10);
```

```
Map.addLayer(DEM_clip, {}, 'DEM_clip');
```

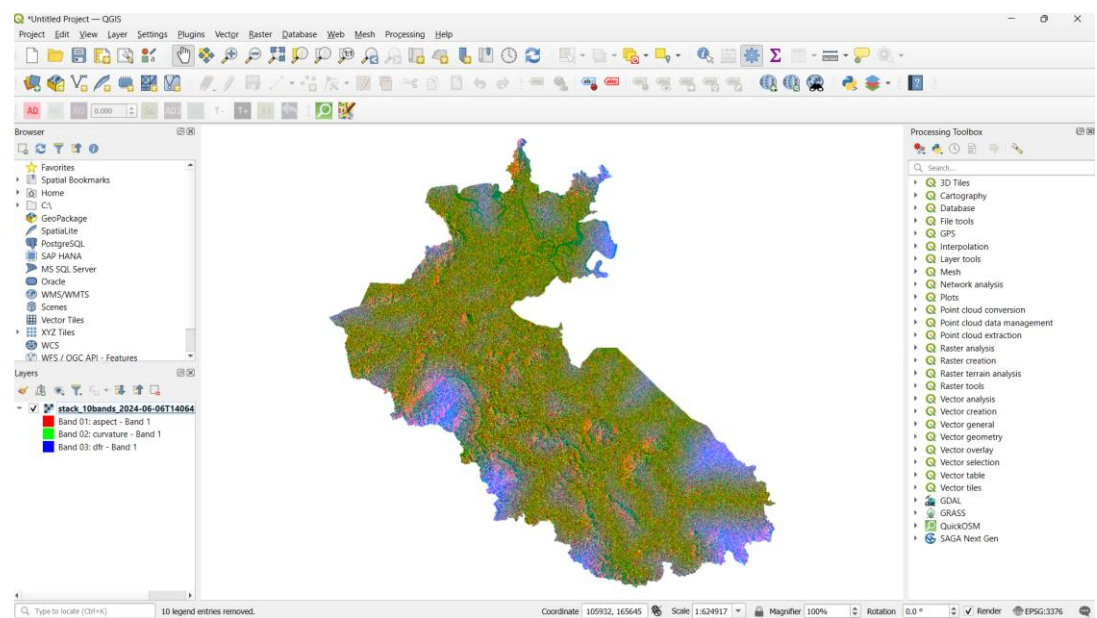
Image Collection Process:

```
var clippedImage = esri_lulc10.mosaic().clip(AOI);
```

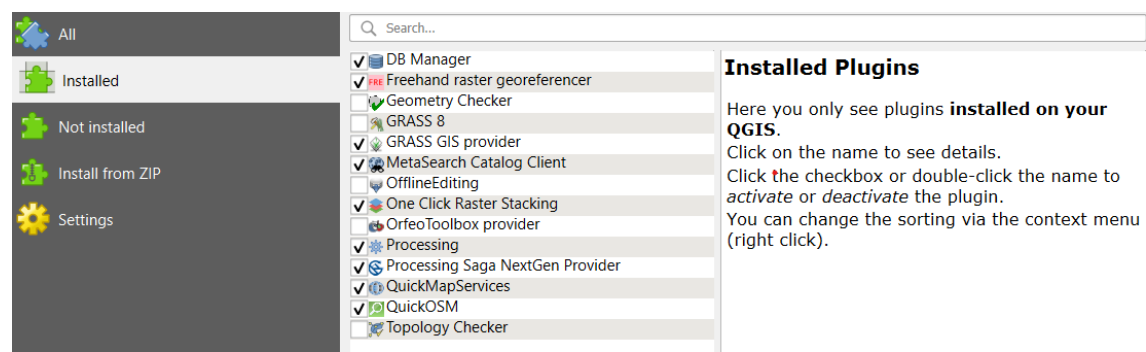
given that the image collection is in Imports.

Appendix F: QGIS

QGIS was the main GIS software used to handle the spatial data in this study. QGIS is a free open-source GIS software.



QGIS was the GIS software of choice as it was free, and a lot of plugins are available.



Freehand raster georeferencer, SAGA, QuickOSM, QuickMapServices, and One Click Raster Stacking were the main plugins used extensively in this study.