CHAPTER 1

INTRODUCTION

1.0 Introduction

Malaysia is protected from catastrophic natural disaster such earthquakes, tsunami, volcano and tornado but this country not immune to geological disaster such as floods, landslide, sinkholes that has always led to loss of lives, extensive damage to infrastructures and environment. Geographically, Malaysia located at the equator which receives high rainfall distribution and humidity throughout the year which mainly cause to the floods and landslides, especially during the Northeast Monsoon that occurs from November to March every year (Rahman & Mapjabil, 2017). Many landslides hazard happened in this season including the major landslide tragedy at Father's Organic Farm, Batang Kali, Selangor on 16 December 2022, and the collapse of the Highland Tower on 11 December 1993 was proved that monsoon (Harvendhar Singh et al. 2022).

Landslides described as the large-scale movement of soil because of high water content in the soil. It occurs in a variety of materials (soil, debris, rock, organics) and move at varying rates (mm/year to tens of m/second) and in a variety of patterns (topple, fall, flow, slide, spread) (Geertsema et al., 2020). Landslide risk assessment and Landslides susceptibility mapping (LSM) are required to detect the tendencies of landslides occurring and the potential risks occurred in the high vulnerability zones depending on the topographic and climatic condition of the area (Brabb, 1984; Reichenbach et al., 2018). The potential hazards consist of the injuries, lives lost, and property that might be affected to a system such as community or society in any given time span (Guo et al., 2020a, Guo et al., 2020b). Landslide hazard assessment techniques have been mostly developed using previously identified landslide data and GIS can be applied in the development of landslide spatial index for assessing quantitative potential landslides. Susceptibility analysis seeks to identify the locations that the landslides will happen and factor that led the hazards (Guzzetti et al., 2005). The analysis and prediction conducted in areas identified as having a high risk of landslides will help the

authorities take early actions to protect the safety of the residents and reduced the landslides impacts.

Machine learning techniques combined with ArcGIS software is the best tools for landslides susceptibility analysing and prediction that can improved the precision of geological hazards information and awareness to public. Machine learning (ML) and its derivative, like deep learning have become integral in many fields and disciplines, including geotechnical engineering where volume of observations and the availability of data sources such as remote sensing satellite data has promoted application of the ML in the study of landslides. ML algorithms are the best technique at identifying intricate patterns within the data and provide great potential for modeling issues related to landslides then enhance the accuracy of prediction compared to conventional approaches (Tehrani et al., 2022)

1.1 Problem Background

Although the conventional techniques that currently used is relevance, but it can provide only limited information about the location and time of the future landslides because of the complexity of these processes and the requirement to perform a numerous number of computations in real time. These methods work with laws of geology, geotechnics, hydrology, and meteorology; however, these methods are not always efficient in terms of explaining all the factors behind mass movements of soil (Tehrani et al., 2022).

Over the last few decades, research has primarily concentrated on model development for evaluating the risk of landslides through the development of knowledge based expert systems (Althuwaynee et al., 2012; Kayastha et al., 2013). The results heavily depend upon the experience of experts bringing in substantial aspects of subjectivity. Therefore, the expert systems are only valid in their distinctive survey area and are greatly relied on manual labour and the expertise of engineers and (He et al., 2024) geologists. Several research papers have been written on LSM by using various types of MLs and DLs to improve the accuracy of LSM. However, a very important factor often overlooked in this regard is selection of appropriate sample locations where landslides have never occurred, an aspect that greatly determines the precision of LSM (Rabby et al., 2023).

Agboola et al., 2024 applies geospatial analysis in conjunction with machine learning tools like Random Forest, Support Vector Machines, and Gradient Boosting to perform landslide susceptibility in consideration of various spatial data, geographical information systems (GIS) as a tool for data compilation and integration, optimization algorithms and validation criteria to enhance susceptibility mapping and classification for risk assessment. However, there are some constraints to using this method including quality and accessibility of input data since poor data quality or missing spatial data that reduce accuracy of the model.

ArcGIS is a powerful GIS tool for creating landslide susceptibility maps. However, it has limitation related to its machine learning functionalities because it can be restrictive to explore the state-of-the-art machine learning algorithms and its customization necessary towards the implementation of suitable landslide susceptibility mapping. Furthermore, it also has problem in handling big data sets used in other fields such as machine learning which results to poor performance of the analyses (Moziihrii Ado et al., 2022). Marjanović et al., 2011 used three types of ML models including Support Vector Machines (SVM), Decision Trees (DT), and Logistic Regression (LR) models in the evaluation of landslide susceptibility. These algorithms were chosen as they are appropriate for use with non-linear data and are widely applied in landslide research. But the study has some limitations of data availability, problems with generalization of models, which require the presence of specialized knowledge.

1.2 Problem Statement

In the rapid development of technology, more effective analysis and prediction are needed to identify the areas with a high risk of landslides to make careful preparations and planning for effective mitigation measurements. By introducing machine learning in creating analysis and prediction models for landslide susceptibility assessment and mapping, this tool can enhance the precision of landslides prediction and provide the accurate decision-making for mitigations planning and land-use management with cost effective and saving time. The problem question for this study are:

(a) How various geological and geospatial data can be used to produce the landslides susceptibility assessments and analysis?

(b) How machine learning and GIS can be applied in landslides susceptibility assessment and analysis?

(c) How machine learning and GIS can improve the decision-making process on mitigation and land use management**.**

1.4 Research Goal

This research aims to develop a machine learning model for landslide susceptibility analysis and prediction with visualization of the high-risk area of landslide occurrences.

1.5 Research Objectives

a) To create a robust dataset for analysis and prediction of landslide using topographical geological, meteorological and land-use data.

b) To improve the landslide spatial analysis and prediction accuracy using the machine learning models.

c) To identify and visualize the landslide prone areas using the landslide susceptibility maps from ArcGIS software for mitigation and land-use planning management.

1.6 Scope of study

This research focuses on the landslide susceptibility analysis in Kuala Lumpur and Selangor, Malaysia. The machine learning models are applied to improve the spatial analysis and prediction performance of landslide susceptibility and correlate it using ArcGIS software to get the visualization of landslide prone areas.

1.7 Significance of study

This study is important for disaster risk management which enables a prediction of areas prone to landslides. Machine learning helps geologists analyse the big and huge data and find out the factors which play a significant role for landslide

occurrence and can also develop models with high degree of accuracy. These models assist in partitioning areas by the vulnerability thus facilitating planning and resource deployment for mitigation measures of disasters where it can save peoples live and minimized the losses. The analysis enhances understanding and management of spatial resources within cities, thus enhancing the sustainable development of cities. Moreover, it promotes and builds the capacity for multidisciplinary engagement and commits to improving community disaster capacities, responses and protection of the environment and decisions.