

**A MACHINE LEARNING APPROACH FOR  
LANDSLIDE SUSCEPTIBILITY MODELING FOR  
RATHNAPURA DISTRICT, SRI LANKA**

Madirawalage Anjali Shanika Perera

(189339R)

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa  
Sri Lanka

May 2021

**A MACHINE LEARNING APPROACH FOR  
LANDSLIDE SUSCEPTIBILITY MODELING FOR  
RATHNAPURA DISTRICT, SRI LANKA**

Madirawalage Anjali Shanika Perera

(189339R)

Dissertation submitted in partial fulfillment of the requirements for the degree Master  
of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa  
Sri Lanka

May 2021

## DECLARATION OF THE CANDIDATE & SUPERVISOR

M.A.S Perera (189339R) hereby declare that this report was carried out by me under the supervision of Dr. Uthayasanker Thayasivam, in the Department of Computer Science and Engineering, University of Moratuwa. It has not been submitted to any other institution or study program by me for any other purpose.

***UOM Verified Signature***

14/11/2021

.....  
M.A.S Perera

.....  
Date

I, Dr. Uthayasanker Thayasivam, certify that the above declaration made by the candidate is true & this report is forwarded for the purpose of evaluation.

***UOM Verified Signature***

14/11/2021

.....  
Dr. Uthayasanker Thayasivam

.....  
Date

Senior Lecturer,

Department of Computer Science and Engineering,

University of Moratuwa,

Katubadda.

## **ACKNOWLEDGMENT**

I would like to express my heartfelt gratitude to my supervisor, Senior Lecture Dr. Uthyasanker Thayasivam, for providing me with excellent guidance, advice, and time from the beginning of this study, as well as for providing me with exceptional experiences during the work, and for his unwavering encouragement and support in various ways.

I would also like to thank, MSc Coordinator, Dr. Charith Chitraranjan for sparing his valuable time and guide me on this work. And I express my gratitude to the entire senior and junior lecturer panel, as well as the faculty and staff of the University of Moratuwa's Faculty of Engineering, for their guidance, supervision, and critical contribution to this study.

Finally, I would like to express my gratitude to my family and friends for their help in making this study a success.

## **Abstract**

In certain areas of the world, landslides are the most common and recurrent natural hazard, resulting in substantial human deaths and property damage. Landslides are extremely common in Sri Lanka, with landslides affecting approximately 30.7% of the country's land area. As the demand for human growth has increased, landslides have become a major problem in Sri Lanka's mountainous regions. As a result, detecting landslide potential associated with terrain data and remote sensing data is crucial for ensuring the long-term viability of projects while minimizing the risk of landslide disasters.

The aim of this study is to develop a susceptibility map for Sri Lanka using a novel data science approach. This study has been used the Rathnapura district in Sri Lanka as the study area. In this study, five ensemble machine learning algorithms: Random Forest, Bagged Decision Tree, AdaBoost, XGBoost, and Gradient Boost were used for landslide prediction and landslide susceptibility map modeling. Using the K-Means clustering algorithm, the class probability values from ensemble-based machine learning algorithms were used to reclassify the study area into susceptibility levels: Extreme Low (EL), Low (L), Moderate (M), High (H), Very High (VH), and Extreme High (EH). In addition, landslide susceptibility maps were generated using the Frequency Ratio technique. The Landslide Susceptibility Index (LSI) was generated using the Frequency Ratio values. The study area was then categorized into six landslide susceptibility classes based on the LSI value: Extreme Low, Low, Moderate, High, Very High, and Very High.

The F-Score, Accuracy, Precision, and Recall values were used to evaluate the landslide prediction results, while the Landslide Density value was used to evaluate the LSMs. Finally, a web application was developed to visualize landslide susceptibility maps, landslide locations, and landslide conditioning factor maps.

## TABLE OF CONTENTS

DECLARATION OF THE CANDIDATE & SUPERVISOR .....	i
ACKNOWLEDGMENT .....	ii
Abstract .....	iii
LIST OF FIGURES .....	viii
LIST OF TABLES .....	ix
LIST OF APPENDICES .....	x
CHAPTER 1: INTRODUCTION .....	1
1.1    Landslide .....	1
1.2    Landslide Susceptibility Mapping .....	2
1.3    Problem Definition .....	3
1.4    Research Objectives .....	3
CHAPTER 2: LITERATURE REVIEW .....	5
2.1    Introduction .....	5
2.2    Landslide Disaster Risk Reduction in Sri Lanka.....	5
2.3    Landslide prediction systems in Sri Lanka.....	7
2.4    Landslide Susceptibility Mapping Techniques .....	7
2.4.1 Landslide Susceptibility Mapping.....	8
2.4.2 Landslide Susceptibility Mapping in Sri Lanka .....	11
2.5    Summary .....	13
CHAPTER 3: STUDY AREA .....	18
3.1    Study Area: Rathnapura district .....	19
CHAPTER 4: EXPERIMENTS .....	20
4.1    Experiments on Dataset.....	20
4.2    Causative factors of landslide.....	20

4.2.1	Slope .....	21
4.2.2	Land use .....	22
4.2.3	Water Catchment .....	22
4.2.4	Landform .....	22
4.2.5	Vegetation .....	22
4.2.6	Elevation.....	22
4.2.7	Distance to water streams.....	22
4.2.8	Overburden.....	23
4.2.9	Drainage Density .....	23
4.2.10	Annual Rainfall.....	23
4.3	Preparation of the training and validation datasets .....	24
4.4	Experiments on existing methodologies.....	25
CHAPTER 5: METHODOLOGY .....		26
5.1	Landslide Causative factor selection .....	29
5.1.1	Tree based feature important methods .....	29
5.1.2	Information Gain (IG) factor selection method.....	30
5.2	Re-Sampling Techniques for Class Imbalance Issue .....	31
5.2.1	Penalize Algorithms (Cost-sensitive Training) .....	32
5.2.2	Change the Algorithms or use Tree-based Algorithms .....	32
5.3	Landslide Prediction Modeling .....	32
5.3.1	Random Forest Algorithm.....	33
5.3.2	Bagged Decision Tree (BDT) Algorithm .....	34
5.3.3	Extreme Gradient Boosting (XGBoost) Algorithm.....	34
5.3.4	AdaBoost Algorithm .....	35
5.3.5	Gradient Boosted Tree (GBRT) Algorithm.....	35
5.4	Landslide Susceptibility Modeling.....	37

5.4.1	Landslide Susceptibility Modeling using a Semi-supervised machine learning model .....	37
5.4.1.1	K-Means Algorithm.....	38
5.4.2	Landslide Susceptibility Modeling using Frequency Ratio (FR) model .....	39
5.5	Model Validation Techniques .....	42
5.5.1	Confusion Matrix .....	43
5.5.2	Accuracy.....	44
5.5.3	Precision .....	44
5.5.4	Recall.....	44
5.5.5	F-Score .....	45
5.5.6	Specificity.....	45
5.5.7	Sensitivity.....	45
5.5.8	Landslide Density Index (LDI) .....	45
CHAPTER 6: RESULTS AND EVALUATION .....		47
6.1	Feature Selection Method Results.....	47
6.1.1	Tree-based feature selection method results.....	47
6.1.2	Information Gain (IG) feature selection method results.....	48
6.2	Landslide prediction results.....	49
6.3	Landslide Susceptibility Map (LSM) results.....	51
6.3.1	LSM by Machine Learning Model Results .....	51
CHAPTER 7: DATA VISUALIZATION WEB APPLICATION .....		55
7.1	Technologies used for web application implementation.....	56
7.1.1	Leaflet JavaScript Library .....	56
7.1.2	CesiumJS JavaScript Library with GeoJson files.....	56
CHAPTER 8: CONCLUSION AND FUTURE WORKS .....		58
REFERENCES.....		62

APPENDIX A: GIS FEATURE MAPS.....	69
APPENDIX B: THE CORRELATION BETWEEN THE LANDSLIDE LOCATION AND LANDSLIDE CONDITIONING FACTOR CLASSES .....	71
APPENDIX C: SCREENSHOTS OF THE APPLICATION .....	74

## LIST OF FIGURES

Figure 2.1: Landslide disaster risk reduction process in Sri Lanka .....	6
Figure 2.2: Digital elevation model and landslide occurrence in the study area .....	9
Figure 3.1: Spatial Distribution of Landslides [43] .....	18
Figure 3.2: Landslide Map in Rathnapura District .....	19
Figure 5.1: Methodological diagram for Landslide Susceptibility map construction	28
Figure 5.2: Landslide distribution .....	31
Figure 5.3: Relationship between landslide and causative factors (a) Rainfall (b) Distance To Water Streams (c) Drainage Density (d) Elevation (e) Catchment (f) Vegetation (g) Land Use (h) Overburden (i) Landform and (j) Slope maps .....	42
Figure 5.4: Confusion matrix terminology.....	43
Figure 6.1: Feature Importance of landslide conditioning factors .....	47
Figure 6.2: Information Gain of landslide conditioning factors .....	48
Figure 6.3: Performance comparison of Machine learning classifiers.....	50
Figure 6.4: Landslide susceptibility maps produced by semi-supervised machine learning model.....	51
Figure 6.5: Landslide susceptibility maps produced by ensemble FR method (a) without considering weights for each causative factor (b) by considering weights for each causative factor .....	51
Figure 6.6: Landslide proportion of each susceptibility level.....	53
Figure 6.7: Landslide density distribution of each susceptibility level.....	54

## LIST OF TABLES

Table 2.1: Comparison between Data Science based Landslide Susceptibility approaches.....	14
Table 4.1: Landslide conditioning factors with their corresponding sources .....	21
Table 4.2: Classification of causative factors .....	24
Table 5.1: ExtraTreeClassifier parameters.....	30
Table 5.2: Information gain feature selection classifier parameters .....	30
Table 5.3: Classifier Parameters .....	36
Table 6.1: Performance comparison of ensemble machine learning algorithms on landslide prediction .....	49
Table 6.2: Accuracy statistics for LSM models .....	52

**LIST OF APPENDICES**

APPENDIX A: GIS FEATURE MAPS..... 69

APPENDIX B: THE CORRELATION BETWEEN THE LANDSLIDE LOCATION  
AND LANDSLIDE CONDITIONING FACTOR CLASSES ..... 71

APPENDIX C: SCREENSHOTS OF THE APPLICATION ..... 74