Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

Erlend Økland

Landslide Susceptibility Mapping Using Machine Learning

A comparative study of automated machine learning and random forest for landslide susceptibility mapping in Vestland county, Norway

Master's thesis in Geography
Supervisor: Jan Ketil Rød
May 2024

NTNU
Norwegian University of
Science and Technology

Erlend Økland

Landslide Susceptibility Mapping Using Machine Learning

A comparative study of automated machine learning and random forest for landslide susceptibility mapping in Vestland county, Norway

Master's thesis in Geography
Supervisor: Jan Ketil Rød
May 2024

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

NTNU
Norwegian University of
Science and Technology

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

Erlend Økland

Machine Learning for Landslide Susceptibility Mapping

A comparative study of automated machine learning and random forest for landslide susceptibility mapping in Vestland county, Norway

Master's thesis in Master's Thesis in Geography
Supervisor: Jan Ketil Rød
May 2024

NTNU
Norwegian University of
Science and Technology

Erlend Økland

Machine Learning for Landslide Susceptibility Mapping

A comparative study of automated machine learning and random forest for landslide susceptibility mapping in Vestland county, Norway

Master’s thesis in Master's Thesis in Geography
Supervisor: Jan Ketil Rød
May 2024

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

NTNU
Norwegian University of
Science and Technology

Abstract

In Norway, landslides in soil represent a significant and escalating threat to infrastructure and residential areas, with projections indicating an increase in severity due to climate change. This study evaluated the applicability of machine learning algorithms for landslide susceptibility mapping (LSM) for the Vestland county in Norway. A dataset of historic landslide registries, together with event inventories, were used as the ground truth data for the classification task. A selection of fifteen conditioning factors were employed, encompassing morphological, geological, land cover-related, hydrological and anthropogenic factors.

Two different machine learning (ML) approaches was explored; The established Random Forest (RF) algorithm and an automated machine learning algorithm (Auto-Sklearn). The machine learning process involved feature selection to mitigate multicollinearity and enhance model performance, focusing on removing redundant and irrelevant predictors. The two models were trained on a training area of 6,478 km² to classify the data samples as landslide or non-landslide, treating the ML task as a binary classification and expressing the results in the form of a probability in order to produce susceptibility maps. The models were also validated through the test data and on an external validation area of 1,798 km² to assess their predictive accuracy and generalizability on unseen data.

Both models demonstrated high accuracy (88%) on test data, with the RF model achieving an AUC score of 0.96 and Auto-Sklearn's multilayer perceptron reaching 0.95. The multilayer perceptron model also showed fewer false positives and predicted more actual landslides than the RF model. When tested on external validation data to assess generalizability, the RF model's accuracy fell to 76% with an AUC of 0.87, while the multilayer perceptron model maintained higher accuracy (81%) and an AUC of 0.9. This suggests that automated machine learning can effectively optimize algorithms for specific datasets and outperforms traditional models, indicating robustness and the capability of the models to predict landslides across varying geographical regions.

The RF model identified distance to roads, Topographic Ruggedness Index, surficial deposit class 130 (bare rock/thin turf cover), average annual precipitation, and Normalized Difference Vegetation Index as key variables influencing landslide susceptibility, highlighting a spatial bias with an overrepresentation of landslide registries near roads. Despite this, removing road-associated variables only slightly impacted model effectiveness. A model focusing solely on slope, planform curvature, and water contributing area underperformed, underscoring the

necessity of incorporating a diverse array of factors for robust prediction. This variable importance may not be generalizable to other regions with distinct characteristics.

The landslide susceptibility map developed in this study demonstrates potential for accurately identifying high-risk areas within trained regions, offering a detailed tool for municipal planning over existing national maps. However, its utility is limited by the quality of the underlying landslide inventory, emphasizing the need for improved data accuracy for enhanced prediction accuracy.

The entire collection of codes that supports the results reported in this thesis is available in the online GitHub repository. This repository contains all the code mentioned throughout this study and is essential for replicating the reported results. You can access the repository at: https://github.com/ErlendOkland/LSM-Repository.git

Sammendrag

Løsmasseskred representerer en betydelig og økende trussel mot infrastruktur og bebyggelse i Norge, med prognoser som indikerer en økning i alvorlighetsgrad grunnet klimaendringer. Denne studien har vurdert anvendeligheten av maskinlæringsalgoritmer for kartlegging av aktsomhetsområder for løsmasseskred for Vestland fylke i Norge. Et datasett med historiske skredregistreringer, samt registreringer fra individuelle skredhendelser, ble brukt som grunndata for klassifiseringen. Et utvalg av femten betingende faktorer ble benyttet, som omfatter morfologiske, geologiske, arealressurs-relaterte, hydrologiske og antropogene faktorer.

To forskjellige maskinlæringsmetoder ble utforsket; den etablerte Random Forest (RF)-algoritmen og en automatisert maskinlæringsmodell (Auto-Sklearn). Maskinlæringsprosessen inkluderte variabelvalg for å redusere multikollinearitet og forbedre modellens ytelse, med fokus på å fjerne overflødige og irrelevante prediktorer. De to modellene ble trent på et treningsområde på 6 478 km² for å klassifisere prøvene enten som skredutsatte eller ikke-skredutsatte. Dette ble definert som en binær klassifisering, hvor resultatene ble uttrykt som sannsynlighetsgrad og visualisert gjennom et aktsomhetskart. Modellene ble videre validert gjennom testdata og på et eksternt valideringsområde på 1 798 km² for å vurdere deres prediktive nøyaktighet og generaliserbarhet på usette data.

Begge modellene viste høy nøyaktighet (88%) på testdata. RF-modellen oppnådde en AUC-verdi på 0,96 og Auto-Sklearns flerlags perceptron nevralt nettverk hadde en AUC-verdi på 0,95. Flerlags perceptron-modellen viste også færre falske positive og predikerte flere faktiske skred enn RF-modellen. De to modellene ble testet på eksternt valideringsdata for å vurdere generaliserbarhet; RF-modellens nøyaktighet falt til 76% med en AUC-verdi på 0,87, mens flerlags perceptron-modellen opprettholdt høyere nøyaktighet (81%) og en AUC-verdi på 0,9. Analysene i denne studien viser at automatisert maskinlæring effektivt kan optimalisere en algoritme for det spesifikke datasettet brukt i denne studien, og overgikk den tradisjonelle RF-algoritmen. Algoritmen viser sterk robusthet og evne til å predikere skred utenfor området den ble trent på.

RF-modellen identifiserte avstand til vei, topografisk ruhet, løsmassetype 130 (Bart fjell / fjell med tynt torvdekke), gjennomsnittlig årlig nedbør og vegetasjonsindeksen som de viktigste forklaringsvariablene. Dette funnet demonstrerer romlig bias i databasen for registrerte skredhendelser, hvor skredregistreringene er overrepresentert langs veinettverket. Fjerning av vei-assosierte forklaringsvariabler påvirket imidlertid ikke modellens prediksjonsnøyaktighet negativt i stor grad. Videre ble en modell som fokuserte utelukkende på forklaringsvariablene skråning, plankurvatur og akkumulert overflatestrømming testet; modellen underpresterte betydelig, noe som understreker nødvendigheten av å inkludere et bredt utvalg av relevante faktorer for robust prediksjon. Viktigheten av disse variablene må betraktes i sin kontekstuelle sammenheng og kan ikke nødvendigvis generaliseres til andre regioner.

Det resulterende aktsomhetskartet viser potensial for å nøyaktig identifisere områder utsatt for skred, spesielt innenfor samme området som modellen ble trent på. Metoden som ble utviklet i dette studiet viser størst potensial for prediksjon på lokalt og regionalt nivå, og kan oppnå større nøyaktighet i forhold til det det etablerte aktsomhetskartet av NGU. Nytteverdien til aktsomhetskartet begrenses imidlertid av kvaliteten på det underliggende skreddataen som er tilgjengelig, noe som understreker behovet for bedre datakvalitet for å forbedre prediksjonen.

Hele samlingen av koder som støtter resultatene rapportert i denne studien er tilgjengelig i en online GitHub-kolleksjon. Denne kolleksjonen inneholder all kode som er nevnt gjennom oppgaven og er essensiell for å kunne reprodusere de rapporterte resultatene: https://github.com/ErlendOkland/LSM-Repository.git

Preface

The best way to predict the future is to create it – Peter Drucker

This thesis marks my completion of a Master's degree in Geography at the Norwegian University of Science and Technology (NTNU).

First, I would like to extend my deepest gratitude to my supervisor, Jan Ketil Rød, for his unwavering support and guidance throughout the completion of this thesis.

I'm also immensely thankful to everyone at the reading room. A special thanks to Esten for the academic help and inspiration, and to all of you, thanks for the quirky conversations and fun times we shared. These moments of laughter and light-heartedness were just as vital to my journey as the studying itself.

Erlend Økland

Trondheim, May 2024

Table of Contents

Abstract V

Sammendrag VII

Preface IX

Table of Contents XI

List of Tables XIV

List of codes XV

List of Figures XV

List of Abbreviations XVIII

Chapter 1 1

Introduction 1

1.1. Background 1

1.2. The importance of mapping landslide susceptibility 3

1.2.1. Systematic mapping of landslides in Norway 4

1.3. Main Objectives and Research Questions 6

1.4. Outline 8

Chapter 2 9

Theoretical Background 9

2.1. Landslides 9

2.1.1. Classification of landslides 10

2.2. Landslide Susceptibility Mapping 12

2.3. Statistically based landslide susceptibility mapping 14

2.3.1 Logistic regression for LSM 14

2.3.2 Weight-of-evidence for LSM 15

2.3.3 Multi criteria decision analysis for LSM 15

2.3.4 Machine Learning for LSM 16

2.4. Machine Learning 17
2.4.1 Supervised classification 18
2.4.2 Binary Classification 20
2.4.3 Automated Machine Learning 22
Chapter 3 23
Data Preparation and Data Processing 23
3.1. Landslide Inventory Map 23
3.2. Conditioning Factors for LSM 25
3.2.1 Morphological Variables 26
3.2.2 Geological Variables 29
3.2.3 Land Cover related Variables 30
3.2.4 Hydrological Variables 32
3.2.5 Other Variables 33
3.3. Data processing 36
3.3.1 Mapping units 36
3.3.1 Sampling ratio and training/testing ratio 38
Chapter 4 40
Methodology 40
4.1. Presentation of the analyzed areas 41
4.2. Data Preparation 43
4.2.1 Creating point data for landslide / non-landslide 43
4.2.2 Creating influencing factors 47
4.3. Data Sampling 66
4.4. Feature Selection 67
4.5. Machine Learning 72
4.6. External Validation 75
4.7. Landslide susceptibility map 80

Chapter 5 ..... **82**

Results and Discussion 82

5.1. Machine learning for Landslide Susceptibility Mapping 82

5.2. Landslide Susceptibility Mapping using Random Forest 83

5.2.1. External Validation 87

5.2.2. Susceptibility Map 89

5.3. Landslide Susceptibility Mapping using Auto-Sklearn 92

5.3.1 External validation: 93

5.3. Comparing machine learning and NGU's method for landslide susceptibility mapping. 95

5.4. Exploring the importance of the influencing factors for prediction accuracy 97

5.4.1 Assessing the Performance of the random forest model with Targeted Topographic Factors 100

5.4.2 Exploring spatial bias in the landslide inventory 102

Chapter 6 ..... **103**

Conclusion and future work 104

6.1 Future Work 106

References 107

List of Tables

Table 1: Proposed modifications to the Varnes classification system. Red boxes indicate the landslide types proposed by NVE as synonyms to the Norwegian terms for landslide types under the common term “landslides in soil”. 12
Table 2: Confusion matrix for binary classification. 20
Table 3: Table showing the criteria defining the quality grade for historic landslide mapping in the NSDB database. Source: (Devoli et al., 2020). 25
Table 4: Overview of the influencing factors used in this study. 35
Table 5: Filtering of the landslide registries based on AOI, data, landslide type, initiation and locational accuracy. 44
Table 6: A subset of the resulting attribute table to the landslide / non-landslide shapefile after the cell value for the corresponding influencing factor is extracted. 66
Table 7: Subset of the final attribute table after feature selection. The variables: tpi, roughness, annual temperature and forest loss have been removed from the original training dataset. 72
Table 8: Overall accuracy, precision, recall and F1-score for the Random Forest model on the test set. 84
Table 9: Overall accuracy, precision, recall and F1-score for the Random Forest model on the external validation data set. 88
Table 10: Overall accuracy, precision, recall and F1-score for the Multilayer perceptron artificial neural network model on the test set. 92
Table 11: Overall accuracy, precision, recall and F1-score for the Multilayer perceptron artificial neural network model on the external validation data. 94
Table 12: Coverage of susceptibility zones and percentage of landslide records intersecting with NGU LSM and RF LSM, illustrating mapping accuracy of both methods. 96
Table 13: Overall accuracy, precision, recall and F1-score for the RF model only trained on the influencing factors used by the NGU approach: slope, planform curvature and water contributing area. 100
Table 14: Overall accuracy, precision, recall and F1-score for the RF model where the influencing factors “d_roads” and “landuse” are removed from training. 103

List of codes

Code 1: Code for generating non-landslide point data. 46
Code 2: Code for converting slope in degrees into slope in radians 51
Code 3: Code for calculating SPI 51
Code 4: gdal command to generate topographic position index based on a DEM..... 52
Code 5: gdal command to generate terrain ruggedness index based on a DEM 53
Code 6: ArcPy script to calculate TWI based on the equation given by (Saleem et al., 2019) 55
Code 7: ArcPy script to delineate landform types based on combinations of TPI at small and large scale and slope. Classes represents: 1: Canyons, 2: Valleys, 3: Headwaters, 4: U-valleys, 5: Plains, 6: Open slopes, 7: Upper slopes, 8: Ridges, 9: Hills, 10: High ridges 58
Code 8: Code for extracting the corresponding cell value for the influencing factors using the Extract Multi Values to Points tool..... 67
Code 9: The auto-sklearn classifier fitted to the training data. The classifier was run for two hours on all 8 of the computer's CPU cores. 75
Code 10: Code to predict landslide probability scores on the raster stack, based on the trained Random Forest model, called "model_rf" in the script. 80

List of Figures

Figure 1: The graph shows the total damage cost and cost per landslide event between 1980 and 2023. The numbers are adjusted for inflation. An exponential graph is plotted based on the cost per landslide event and total damage cost which gives an indication of future costs. Source: https://nask.finansnorge.no/ 3
Figure 2: Simplified scheme for the index-based combination of the three topographic properties for the identification of release areas. A cell is considered a starting cell when the index for all three input datasets is 1 (Fischer et al., 2014). 5
Figure 3: Flowchart showing the workflow of a typical supervised classification task. Source: (Mandal & Bhattacharya, 2019) 19
Figure 4: Curvatures where both profile and planforms curvatures are combined. The columns show planform curvature and rows show the profile curve. Orange square show what Sidle & Ochiai (2006) describes as the most unstable landforms. Source: (Esri, n.d-f). 28

Figure 5: A snippet from the study area showing generated slope units (A) and grid cells (B). The slope units were created using the Alvioli et al. (2016) method, with a minimum slope unit size of 10 000 m2. The grid cells have a size of 5 x 5 meters, the same size as the resolution of the influencing factors but are enlarged in the figure for better visualization. ... 38
Figure 6: Methodological framework for the thesis. ... 41
Figure 7: Figure showing how Vestland county was subdivided into two appropriate areas, one for training / prediction and one for external validation. For both areas, only what is defined as FKB-A and B areas was selected to be the final AOIs. ... 42
Figure 8: A snippet from the AOI showing landslide inventory points, the prepared fishnet and randomly selected grid cells. ... 47
Figure 9: The two thematic maps representing the sine and the cosine of the aspect. Aspect raster is computed from a digital elevation model (each cell representing altitude) and the Northnerness and Easterness is computed from the aspect raster. Inset maps from one area is shown for better visualization. ... 49
Figure 10: Profile Curvature raster ... 50
Figure 11: Stream Power Index Raster. Positive values indicate areas with higher potential for water flow erosion. Negative values represent areas with lower potential for water flow erosion. ... 51
Figure 12: Topographic Position Index raster. Positive values indicate that the central point is located higher than its average surroundings, while negative values indicate a position lower than the average. ... 52
Figure 13: Terrain Ruggedness Index raster. High values indicate a significant elevation difference between the central pixel and its surrounding pixels. Values closer to zero indicate areas of minimal elevation difference. ... 54
Figure 14: Topographic Wetness Index raster. Positive values represent areas of high-water accumulation and values closer to zero represents areas with low accumulation. ... 55
Figure 15: Slope raster measured in degrees. The output represents the rate of change of elevation for each DEM cell. ... 56
Figure 16: TPI raster data at 300- and 2000-meter scale. High values represent areas higher than the average terrain surrounding them, calculated using two different annulus sizes. TPI is combined with slope raster to delineate landform types as discrete raster. ... 59
Figure 17: Distance accumulation raster for road and river network. Each non-source cell represents the distance in meters to the closest road and river polyline. ... 60
Figure 18: Surficial deposits, land use and soil condition raster data. ... 62

Figure 19: Map snippet showing the NDVI. Cell values range from negative (water or non-vegetated areas), values between 0 and 0.3 indicate barren areas and values between 0.3 and 0.6 typically indicate sparse vegetation cover 63
Figure 20: Map snippet showing deforested areas between 2000 and 2023 64
Figure 21: Mean annual temperature and precipitation raster data 65
Figure 22: Correlation matrix with the Pearson correlation coefficient. The heatmap's colors range from blue (negative correlation) to red (positive correlation), with annotations providing the correlation values 69
Figure 23: New correlation matrix without highly correlated features. 70
Figure 24: Plot of the feature importance for the training data. The variable “forestloss” is removed as it has very little significance for the model. 72
Figure 25: Area for external validation, including the 250 landslide points and the 250 randomly generated non-landslide points. 77
Figure 26: Conditioning factors for the external validation area. 79
Figure 27: Confusion matrix for the random forest classifier on the test set. 84
Figure 28: Plot of the ROC curve for the RF model and the corresponding AUC value of 0.96. The diagonal line represents a score of 0.5 (random guess) and the green line represents a score of 1 (perfect prediction) 86
Figure 29: Confusion matrix for the RF model with classification thresholds of 0.3, 0.5, 0.7 and 0.9. 87
Figure 30: Confusion matrix for the random forest classifier on the external validation dataset. 89
Figure 31: Plot of the ROC curve for the external validation dataset with the corresponding AUC value of 0,87. The diagonal line represents a score of 0,5 (random guess) and the green line represents a score of 1 (perfect prediction). 89
Figure 32: LSM of the training AOI using the random forest classifier. Cell values represent the probability score by the classifier; Between 0 and 1. 91
Figure 34: Confusion matrix for the multilayer perceptron model on the test set. 93
Figure 34: Plot of the ROC curve for the multilayer perceptron model and the corresponding AUC value of 0.95. The diagonal line represents a score of 0.5 (random guess) and the green line represents a score of 1 (perfect prediction). 93
Figure 35: Confusion matrix for the multilayer perceptron model on the external validation data. 94

Figure 36: Plot of the ROC curve for the multilayer perceptron model and the corresponding AUC value of 0.9. The diagonal line represents a score of 0.5 (random guess) and the green line represents a score of 1 (perfect prediction). 95
Figure 37: Comparison between the NGU LSM and the RF LSM. The zones from the RF model include all areas with a probability score of 0.6 or greater. Landslide registries are added to better illustrate the relationship between the predictions and the 'ground truth'. 97
Figure 38: The feature importance of all the features used by the RF model. 99
Figure 39: Plot of the ROC curve for the RF model, trained only on the influencing factors slope, planar curvature and water contributing area..... 101
Figure 40: A snippet of the AOI, showing a comparison between LSM derived from the RF model trained on the targeted topographic factors used by NGU, and the comprehensive method employed primarily in this thesis. Both LSM include all areas with a probability score greater or equal to 0.6..... 102
Figure 41: Plot of the ROC curve for the RF model where the influencing factors "d_roads" and "landuse" are removed from training. 103

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
AOI	Area of Interest
AUC	Area Under the Curve
AutoML	Automated Machine Learning
CN	Climate Normal
CSV	Comma-separated values
DEM	Digital Elevation Model
FKB	The Joint Geospatial Database (*felles kartdatabase*)
FPR	False Positive Rate
GEE	Google Earth Engine
GIS	Geographical Information Systems
GLAD	The Global Land Analysis and Discovery

GFW	Global Forest Watch
GRASS	Geographic Resources Analysis Support System
InSAR	Interferometric synthetic aperture radar
JBV	Norwegian Railroad Administration (*Jernbaneverket*)
LiDAR	Laser imaging, detection and ranging
LR	Linear Regression
LSM	Landslide Susceptibility Mapping
MCDA	Multi-Criteria Decision Analysis
MCE	Multi-Criteria Evaluation
MET	The Norwegian Meteorological Institute (*Meteorologisk institutt*)
ML	Machine Learning
MLP	Multilayer Perceptron
NDVI	Normalized Difference Vegetation Index
NetCDF	Network Common Data Form
NIBIO	Norwegian Institute of Bioeconomy Research (*Norsk institutt for bioøkonomi*)
NIR	Near Infrared
NGI	Norwegian Geotechnical Institute (*Norges Geotekninske Institutt)*
NGU	Geological Survey of Norway (*Norges geologiske undersøkelse*)
NMA	Norwegian Mapping Authority (*Statens kartverk)*
NVE	Norwegian Water Resources and Energy Directorate (*Norges vassdrags- og energidirektorat*)
NSDB	Norwegian landslide database (den nasjonale skredhendingsdatabasen)
NTNU	Norwegian University of Science and Technology (*Norges teknisk-naturvitenskapelige universitet*)
RF	Random Forest
ROC	Receiver Operating Curve
RS	Remote sensing
SLSM	Statistical Landslide Susceptibility Mapping
SPI	Stream Power Index
SQL	Structured Query Language
SVM	Support Vector Machine
SVV	Norwegian Public Roads Administration (*Statens Vegvesen*)
TIFF	Tagged Image File Format
TPI	Topographic Position Index

TRI	Terrain Ruggedness Index
TWI	Topographic Wetness Index
VIF	Variation Inflation Factor
WSL	Windows Subsystem for Linux
XGBoost	Extreme Gradient Boosting

Chapter 1

Introduction

The first part of this thesis will contextualize the utilization of machine learning approaches in conjuncture with geographical data to describe the likelihood of landslides occurring in an area based on local terrain conditions. In addition, I will elaborate on the motivation behind the approach and present the research objectives this thesis aims to achieve. Finally, the structure of the remaining chapters is presented.

1.1. Background

Hydrometeorological phenomena, including mass movements, are one of the most frequent and destructive natural hazards that cause a large number of deaths and damage to infrastructure around the world (Renza et al., 2021). In Norway, landslides in soil pose a considerable threat to infrastructure and dwellings, particularly along rivers, streams, and in old alluvial fans (Devoli et al., 2020). To address these challenges, the Norwegian Natural Perils Pool was established on January 1, 1980. Governed by the Natural Perils Insurance Act and the Pool's rules, it functions as a compulsory insurance mechanism linked to fire insurance. Based on data from the Norwegian Natural Perils Pool, landslides that occurred between 2013 and 2022 resulted in damages totaling 1 784 million NOK (Finans Norge, 2022).

At present, the effects of climate change increase the frequencies of specific extreme weather and climate events, such as floods and rainstorms, accelerating the occurrence of landslides around the world (Stott, 2016). Norway is expected to experience a warmer, wetter and wilder climate in the future, with an increased occurrence of extreme precipitation events and increased magnitude of such events.

To mitigate such influences, landslide susceptibility mapping (LSM) is proposed as a reference for planners and policymakers to better meet the Norwegian authority’s goal of preparing society to adapt to climate changes. Landslide susceptibility is considered as the

likelihood of a landslide occurring in an area based on local terrain and environmental conditions (Brabb, 1985).

Since the mid-1970s, the field of landslide susceptibility has seen a substantial body of research. This research primarily focuses on the functional correlation between landslide occurrences and geo-environmental factors predisposing to such events. Various statistical methods, scales, and mapping units have been employed in these studies (Guzzetti, 2006). Among these, two early investigations stand out for their innovative approach and findings; Neuland's 1976 study pioneered the use of statistical analysis in landslide research. It examined the relationship between various factors - morphometric, geo-mechanical, lithological, and structural - and the stability of 250 slopes in south-west Germany. This research led to the development of a bivariate discriminant analysis-based model for predicting landslide stability and instability.

Carrara's landmark 1983 study synthesized extensive research on the geological and geomorphological determinants of landslides in Calabria, Italy. Carrara employed discriminant and multiple regression analyses to predict landslide susceptibility, using an extensive dataset of landslides, as well as geological and geomorphological information. Notably, (Carrara et al., 1978; Carrara et al., 1977) developed specialized software for automated thematic cartography to manage this spatially distributed data. These tools represented early iterations of grid-based Geographical Information Systems (GIS), a nascent technology at that time (Goodchild, 2010).

Following these pioneering attempts in the mid-70s and 80s, scholars worldwide published hundreds of papers considering different environmental conditions and landslide distribution in different regions, using various statistical models (Reichenbach et al., 2018). Statistically-based landslide susceptibility mapping (SLSM) is now one of the most widely used method for landslide susceptibility assessment (Zhao et al., 2023). SLSM is an effective way of predicting and identifying areas prone to landslides on the regional scale by building the functional relationship between landslide environmental variables and existing landslide inventories (Alcántara-Ayala et al., 2022).

More recently, artificial intelligence represented by machine learning (ML), as well as the increased availability of observational data, like remote sensing data, has introduced new methods of function expression and data analysis for SLSM. The capability of ML in handling large datasets and discovering complex patterns hidden in the data makes it suitable for

landslide susceptibility mapping (Tehrani et al., 2022). Henceforth, the focus on SLSM research has shifted to the construction and analysis of ML models, striving to develop better and more suitable prediction models (Zhao et al., 2023).

1.2. The importance of mapping landslide susceptibility

Norway is a country dominated by valleys with steep slopes, making it essential to take landslide activity into significant consideration when planning and designing new infrastructure. In the last 500 years, 3 500 landslide occurrences have been registered, totaling over 4 000 dead (Furseth, 2006). Between 1980 and 2023, The Norwegian Natural Perils Pool registered 15 410 landslide related damages with a total cost of 3 871 million NOK. Furthermore, data from the NASK database indicates a rising trend in the cost associated with each landslide-related incident [1]. The increase could be attributed to either a higher frequency of landslide events, an escalation in the cost per event, or a combination of both factors. Hence, the data suggests that costs have increased and will likely increase exponentially over

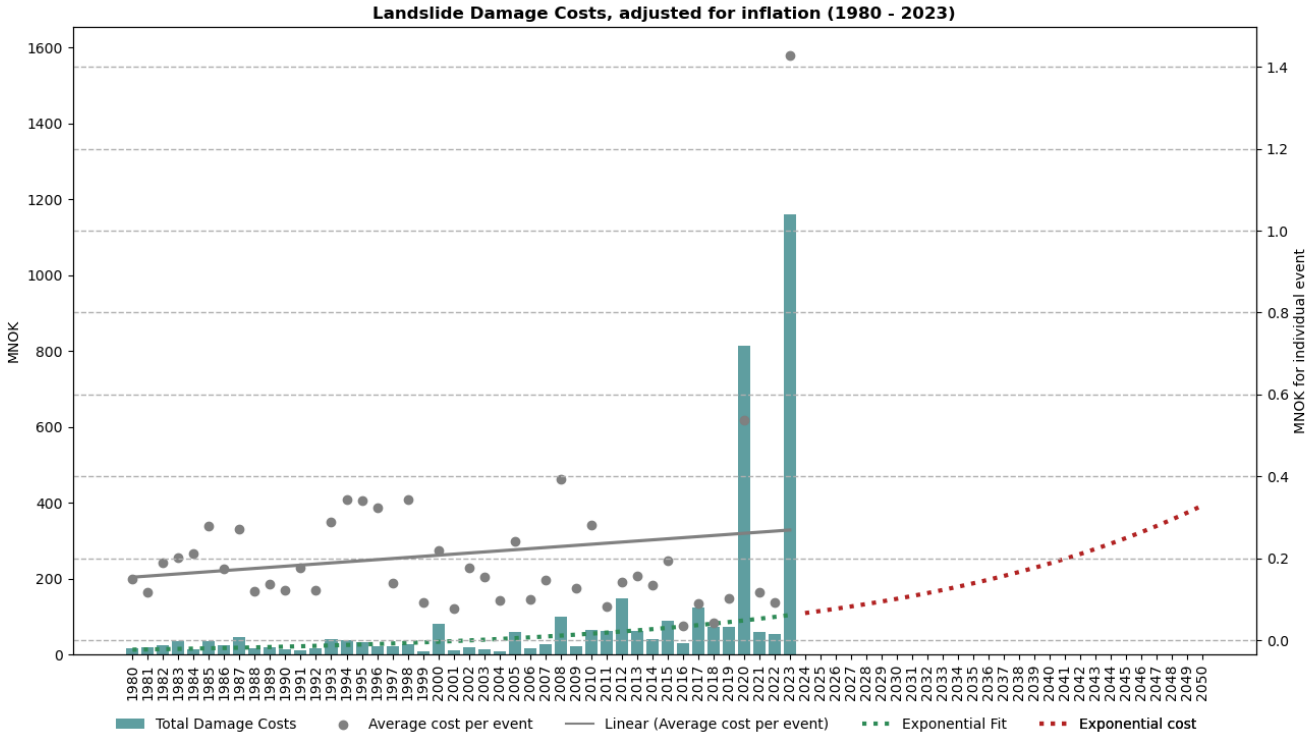


Figure 1: The graph shows the total damage cost and cost per landslide event between 1980 and 2023. The numbers are adjusted for inflation. An exponential graph is plotted based on the cost per landslide event and total damage cost which gives an indication of future costs. Source: https://nask.finansnorge.no/

[1] The NASK database contains statistics on number of reported damages related to natural disasters and total insurance costs. The statistics are all taken from the Norwegian Natural Damage Pool. https://nask.finansnorge.no/default.aspx

the years as a result of both more frequent and more expensive landslide related events (Bruland et al., 2023) (Figure 1).

To prepare for this trend and reduce the risk associated with landslides, numerous measures and significant investments are required. For these to gain societal acceptance and be feasible within given financial constraints, it is necessary to find the most socially optimal solutions. These solutions should offer the best cost/benefit ratio while also being sustainable in terms of the environment and other societal interests. This requires holistic thinking and difficult prioritizations (Bruland et al., 2023).

1.2.1. Systematic mapping of landslides in Norway

Systematic Mapping of landslides under state management has been ongoing since around 1980. Since 2009, the Norwegian Water Resources and Energy Directorate (NVE) has managed the state's landslide mapping, allocating funds to the Geological Survey of Norway (NGU). The Ministry of Petroleum and Energy defined the direction for the state`s work in managing risk associated with landslides (Meld.St.15, 2012). Landslide susceptibility maps and hazard maps were given particular importance as it provides knowledge about which areas are vulnerable and what consequences landslides can entail. The white paper makes a distinction between susceptibility and hazard maps, where susceptibility maps is defined as a map that shows areas prone to landslides, while hazard maps divide zones based on a given probability of a landslide event occurring (Meld.St.15, 2012). In the literature, confusion exists between the two terms landslide “susceptibility” and landslide “hazard”; They are often used as synonyms despite the two words expressing different concepts (Fell et al., 2008). In their review article on statistically-based landslide susceptibility models, (Reichenbach et al., 2018, p.61) makes the distinction between the two terms by considering landslide susceptibility as: “*the likelihood of a landslide occurring in an area on the basis of the local terrain and environmental conditions*” (Brabb, 1985), and landslide hazard as “*the probability that a landslide of a given magnitude will occur in a given period and in a given area*” (Reichenbach et al., 2018, p.61).

The Geological Survey of Norway developed the nationwide landslide susceptibility map on behalf of NVE. Created in 2014, the map encompasses types of landslides such as debris flow, debris slide and small to medium-sized debris floods. The map does not cover quick clay slides, shallow planar soil slides or large debris floods in gentle river courses (NVE, 2021).

The nationwide map highlights areas where caution must be exercised for the above-mentioned landslides. It delineates potential release and runout areas by first identifying the release points of the landslide, then the runout zone of the release points. For the identification of the potential release areas, an index-based approach based on empirical threshold values of the three topographic properties slope angle, planform curvature and water-contributing area were used. An area is only defined as a potential release point for a landslide when the threshold is exceeded on all three properties (see figure 2) (Fischer et al., 2014).

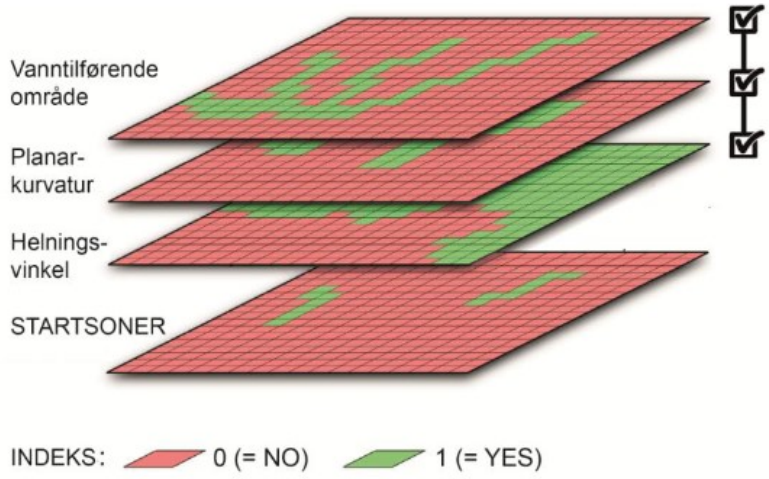


Figure 2: Simplified scheme for the index-based combination of the three topographic properties for the identification of release areas. A cell is considered a starting cell when the index for all three input datasets is 1 (Fischer et al., 2014).

The index-based method used for the susceptibility mapping by NGU is partly based on the work by (Blahut et al., 2010). This approach utilizes GIS-based models and susceptibility maps, integrating spatial and temporal probabilities to assess debris flow hazards. However, it has several disadvantages, including the tendency to oversimplify by not accounting for local geological and climatic variations (Blahut et al., 2010).

The established NGU method for landslide susceptibility mapping (LSM) in Norway employs expert knowledge, field observations and regional adaptations, making it a robust method that shows relevant susceptible areas without being too conservative (Fischer et al., 2014). However, its reliance on manual calibration, limited automation, and its simplicity might result in a model that does not capture susceptible areas on a smaller scale. While not claiming to supersede the traditional nation-wide method, this thesis explores an alternative approach by employing machine learning techniques, specifically Random Forest and

Automated Machine Learning (AutoML). These methods are chosen for their ability to handle large datasets, integrating numerous influencing factors and extensive records of landslides to potentially enhance prediction accuracy on a regional scale. This approach aims to leverage the strengths of machine learning to address some of the limitations of the NGU method.

While working on this master's thesis, the Norwegian government is in the process of preparing a new white paper on floods and landslides following several serious flood and landslide incidents in 2020 and 2023, specifically the quick-clay slide in Gjerdrum December 30, 2020 and the extreme weather event (Hans) of august 2023. The Norwegian Ministry of Petroleum and Energy has received input from a multitude of institutions to help shape the country's future plan regarding floods and landslides. The Norwegian University of Science and Technology (NTNU) makes it clear that in order to understand the magnitude that landslides pose, mapping is fundamental. Mapping is essential for understanding the natural processes that trigger landslide events and where such events happen. The adaptation of statistical models, emerging from the international literature, are beginning to be effective and can predict where landslide events may occur (Bruland et al., 2023).

1.3. Main Objectives and Research Questions

The main objective of this thesis is to explore the effectiveness of state-of-the-art machine learning methods in predicting landslides in soil within a regional context in Norway. This culminates in answering the following research questions:

(i) How effectively can machine learning, trained on historical landslide registries and incorporating morphological, geological, land cover, and hydrological data, predict landslides in soil within Vestland county?
(ii) How does the generalizability of a machine learning model, trained on morphological, geological, land cover, and hydrological data from one Vestland region, impact its landslide prediction accuracy in another Vestland region?
(iii) Does automated machine learning outperform the established Random Forest algorithm when it comes to the predictive performance on training data and external validation data?
(iv) How well can the Random Forest algorithm statistically evaluate which local terrain conditions contribute to landslide in soil formation?

Research Question (i) aims to thoroughly assess the effectiveness of comprehensive landslide susceptibility mapping for predicting landslides in soil within a specific region of Norway. Motivated by the methodology used by NGU to create the current nationwide susceptibility map—which necessitates dividing Norway into distinct zones because of its varied topography, geomorphology, geology, climate, and surficial deposit coverage—research question (ii) seeks to understand how machine learning can handle these disparities. Furthermore, state-of-the-art AutoML techniques will be employed to determine whether a model that automatically identifies the best algorithm and optimizes its hyperparameters can outperform the traditional Random Forest algorithm. The final research question explores another application of machine learning: using the Random Forest (RF) algorithm to assess the importance of geoenvironmental factors in the classification task and thereby elucidating their role in landslide formation.

1.4. Outline

The remainder of this thesis is organized into the following chapters: Theoretical background, Data preparation & data processing, Methodology, Results & discussion, and Conclusion.

Chapter 2 – Theoretical Background, describes the conceptual foundations necessary to understand the process of predicting landslides with machine learning. This include what kind of landslides it is possible predict, what kind of prediction method to use and what kind of metrics that should be used to measure the predictive performance of the LSM. This chapter also presents earlier studies that have used machine learning for LSM.

Chapter 3 – Data Preparation and Data Processing, presents how the landslide inventory was prepared, the justification for selecting the specific influencing factors, as well as the data processing phase. Data processing includes the choice of mapping units, sampling ratio and training/testing ratio. These decisions are based on previous approaches in the field.

Chapter 4 – Methodology, outlines the complete workflow of the study. It begins by defining the study area for both training and external validation. Landslide and non-landslide point data are created and mapped with influencing factors serving as independent variables for the binary classification task. Feature selection is employed to address cases of multicollinearity and redundancy. The dataset is then used to train both a Random Forest classifier and an AutoML classifier, which are also tested on an independent dataset. Finally, these models are utilized to generate a landslide susceptibility map for the region.

Chapter 5 – Result and Discussion, presents the results obtained from the analysis. Various performance metrics were utilized to evaluate the performance of the Random Forest model and the AutoML model. Furthermore, the results are discussed and compared with the established susceptibility map provided by NGU. Lastly, the importance of the influencing factors is explored and discussed.

Chapter 6 – Conclusion, concludes on the main findings and presents ideas for future work.

Chapter 2

Theoretical Background

2.1. Landslides

Over 90% of landscapes that are not currently glaciated consists of hillslopes, the remainder consisting of river channels and their floodplains (Holden, 2017). We can therefore say that all the landforms on earth have slopes as a kind of least common multiple. The processes where parent material is broken down and transported to the streams are vital in understanding how the catchment functions as a geomorphological system (Holden, 2017).

Slope processes can be systemized into two broad types: (i) weathering processes and (ii) transport processes. In a "transport limited" transport process, the rate of transport is constrained by the capacity of the process that moves the material. This occurs when there is an abundant supply of material, but the process responsible for moving it can only transport a limited amount over a short distance, thereby limiting the overall transport rate. Other processes are limited, not by the capacity to transport, but the supply of suitable material to transport, defined as supply limited (Holden, 2017). Mass movements fall under the category of supply-limited slope processes.

In mass movements, a block of rock or soil moves as a single unit. The movement of the block is primarily governed by forces acting on it as a whole. The individual rock or soil fragments within the block maintain close contact, resulting in their collective movement. Mass movements vary in their rates; those driven by large water flows are typically faster than movements in drier conditions. Furthermore, these rapid movements often transport material over greater distances and are usually limited by the available supply of material (Holden, 2017). In rapid mass movements, there is an important distinction between slides and flows. For slides, the moving mass moves as a block, while in flows, different parts of the mass move over each other with differential movements or shear.

Characterized by a range of processes, landslides involve the downward and outward displacement of slope-forming materials, including natural rocks, soil, artificial fill, or their combinations. The movement of these masses can manifest in several forms—falling, toppling, sliding, spreading, and flowing, or a combination thereof. While gravity serves as

the primary driving force behind landslides, its effect is often compounded by the presence of water (Sidle & Ochiai, 2006). In their 2006 book, Landslides – processes, Prediction and Land Use, Sidle & Ochiai (2006) included debris flows as they represent a special type of failure that can initiate either on hillslopes or within channels when augmented by an accumulation of pore water.

2.1.1. Classification of landslides

The aim of this study is to conduct a landslide susceptibility mapping of a region in Vestland county in Norway. A landslide susceptibility mapping depends on a landslide inventory – a comprehensive database detailing the occurrence and characteristics of past landslides in a specific area (Liu et al., 2023). The accuracy of the landslide inventory plays a crucial role in Landslide Susceptibility Mapping (LSM); It directly influences the quality and effectiveness of the samples extracted, which are essential for training and testing machine learning models (Liu et al., 2023). As the landslide inventory is derived from the NVE database for landslide occurrences, it is important to establish connections between English and Norwegian expressions and classify the different landslide types accordingly.

Landslides can be distinguished by the rate of the slope movement. Rates of movement range from less than 15 cm per year to more than 150 cm per second according to Cruden and Varnes (1992, as cited in Xie, 2014). Another method of classifying landslides involves characterizing the movement within the displaced mass, focusing on the kinematics of landslides. For the classification used in this study, the "The Varnes Classification System" (Varnes, 1978) is utilized, derived from the updated article written by (Hungr et al., 2014). The widely used classification scheme developed by Varnes (1978) distinguishes five kinematically distinct types of landslide movements: (a) falls; (b) topples; (c) slides; (d) spreads; (e) flows), plus a combination of these movements. The classification is coupled with the type of material present in the movement (bedrock, coarse soils and predominately fine soils) as well as the speed of movement (Sidle & Ochiai, 2006).

Hungr et al. (2014) proposed modifications to the Varnes classification system to more accurately represent contemporary advancements in the understanding of landslide phenomena, including the materials and mechanisms involved. A landslide is a physical system developed through several stages. As reviewed by Skempton and Hutchinson (1969), the evolution of mass movements comprises pre-failure deformations, failure itself and post-

failure displacement. In addition to this, many landslides exhibit many different movement episodes, often separated by periods of standstill. A landslide is a complex process that might develop through different stages, each individual stage being described differently. Hungr et al. (2014) points out the fact that all landslides are complex and we should be able to apply simple traditional terms to whole scenario. Otherwise, a classification of such an event would need to be torn into fragments. Hungr et al. (2014) proposes the simple term assigned to a given landslide type should reflect the particular focus of the researcher.

The concern of this research is to map potential areas susceptible to landslides, specifically landslides in soil. Hungr et al. (2014) categorizes landslides by geological material types classified by origin, where soil and rock are the two main classes. Similarly, NVE cites international terminology in the report on landslide hazard mapping in Norway (Cruden & Varnes, 1996, as cited in (Øydvin. et al., 2011). The report refers to a need to adopt a more general and simplified terminology, adapted to Norwegian natural conditions as well as the language. Landslides are grouped by geological material: (a) rock; (b) loose material; (c) snow. Hence, “loose material” is defined as all types of loose material residing on top of the bedrock. NVE uses the term “Løsmasseskred”, translated as “Landslides in soils” for all landslides in loose materials. Furthermore, landslides in soil are divided into debris slide, debris flow, debris flood and quick-clay landslides. The terms debris slides, debris flows and debris floods are often used interchangeably, and will in this study be defined as “Landslides in soil” and used to define the landslide types for the landslide inventory, as shown in table 1.

Table 1: Proposed modifications to the Varnes classification system. Red boxes indicate the landslide types proposed by NVE as synonyms to the Norwegian terms for landslide types under the common term "landslides in soil".

Type of movement	Rock	Soil
Fall	1. *Rock/ice* fall[a]	2. *Boulder/debris/silt* fall[a]
Topple	3. Rock block topple[a]	5. *Gravel/sand/silt* topple[a]
	4. Rock flexural topple	
Slide	6. Rock rotational slide	11. *Clay/silt* rotational slide
	7. Rock planar slide[a]	12. *Clay/silt* planar slide
	8. Rock wedge slide[a]	13. *Gravel/sand/debris* slide[a]
	9. Rock compound slide	14. *Clay/silt* compound slide
	10. Rock irregular slide[a]	
Spread	15. Rock slope spread	16. *Sand/silt* liquefaction spread[a]
		17. Sensitive clay spread[a]
Flow	18. *Rock/ice* avalanche[a]	19. *Sand/silt/debris* dry flow
		20. *Sand/silt/debris* flowslide[a]
		21. Sensitive clay flowslide[a]
		22. Debris flow[a]
		23. Mud flow[a]
		24. Debris flood
		25. Debris avalanche[a]
		26. Earthflow
		27. Peat flow
Slope deformation	28. Mountain slope deformation	30. Soil slope deformation
	29. Rock slope deformation	31. Soil creep
		32. Solifluction

2.2. Landslide Susceptibility Mapping

Landslide Susceptibility Mapping is the process of predicting "where" landslides are likely to occur. Several methods and approaches have been proposed and tested to ascertain landslide susceptibility, including analysis of landslide inventories, heuristic terrain zoning, geomorphological mapping, numerical modeling and statistically- based classification methods (Reichenbach et al., 2018). All the proposed methods are based upon a select few, widely accepted set of assumptions (Xie, 2014):

(1) Slope failures leave discernible morphological features; most of them can be recognized, classified, and mapped.

(2) A landslide is controlled by mechanical laws that can be determined empirically, statistically or in deterministic way. Conditions that cause landslides directly or indirectly linked to slope failure, can be mapped and used to predict landslide occurrence in the future.

(3) The past and present are keys to the future. This principle implies that slope failures in the future will likely occur under the conditions which led to past and present slope instability.
(4) Landslide occurrence in space and time, can be inferred from heuristic investigation, computed through the analysis of environmental information, or inferred from physical models. Therefore, a territory can be zoned into hazard classes ranked according to different probabilities.

Ideally, the assessment and mapping of landslide hazards should be based on all these assumptions. Not adhering to these principles can limit the usefulness of any hazard evaluation, regardless of method or the objective of the study (Xie, 2014).

Landslide susceptibility methods can be broadly categorized as either quantitative or qualitative, which can be further classified into knowledge-driven, data-driven and deterministic physical model methods (Yong et al., 2022). Knowledge-driven methods leverage expert insights (Sujatha & Rajamanickam, 2011). Yet, this approach is inherently subjective, and expert opinions might not always be accurate. On the other hand, data-driven approaches rely predominantly on statistical theories (Chen et al., 2020; Guo et al., 2023). These methods streamline the influencing and controlling factors of geological disasters but omit detailed examination of the mechanisms triggering these disasters. The accuracy of these models heavily depends on the quality of the data, with any errors in the input leading to significant discrepancies in the outcomes.

Deterministic physical models (Gao and Wang, 2019), which utilize mechanical equations to describe physical processes, are typically reserved for situations with homogenous landslide types and physical characteristics. They are most effective for in-depth studies of smaller regions. Although this technique can pinpoint key factors based on physical principles, it demands precise parameter information and is generally constrained to smaller-scale studies (Yong et al., 2022).

2.3. Statistically based landslide susceptibility mapping

Among the three previously descibed methods, the data-driven method is the most commonly used method in landslide susceptibility mapping (Yong et al., 2022). The data-driven approach can be roughly divided into binary and multivariate statistical methods.

Binary statistical methods evaluate individual factors, such as slope or lithology, in isolation. This approach overlays each factor onto a landslide inventory map to assign a weight score based on how closely the factor correlates with landslide events. Prominent models within this methodology include the information quantity model and the frequency ratio model (Umar et al., 2014). The binary approach does not account for the interplay among factors.

To bridge this gap, multivariate statistical methods, synonymous with machine learning in this context, concurrently analyze multiple factors (Yong et al., 2022). The model calculates how different factors relate to landslides, allowing for the assessment of each factor's significance. This helps identify key contributors to landslide risk, aiding in the management and prevention of such disasters. Techniques such as logistic regression, artificial neural networks, and random forests are commonly used (Reichenbach et al., 2018).

Reichenbach et al. (2018) emphasizes how models used in the literature employs one of the following six main groups of classification methods:

(i) Classical statistics (e.g., logistic regression)
(ii) Index-based (e.g., weight-of-evidence)
(iii) Multi criteria decision analysis
(iv) Machine learning (e.g., fuzzy logic, support vector machines, random forest)
(v) Neural networks
(vi) Other statistics

For reference, only (ii) Index-based, can be defined as a binary statistical method, while the other five would be considered multivariate statistical approaches.

2.3.1 Logistic regression for LSM

Logistic regression can be seen as an extension of multiple linear regression in situation where the dependent variable, in this case the presence or absence of landslides, is not a continuous parameter. The outcome of the logistic regression is a probability, ranging from 0 to 1 (Nefeslioglu et al., 2008). The general purpose of a logistic regression method in to

determine the best fitting model to describe the relationship between the dependent variable and a set of independent variables.

Kavzoglu et al. (2014) used a logistic regression model to assess the shallow landslide susceptibility of the Trabzon province in Turkey. The logistic regression method was compared to a machine learning approach and a multi criteria decision method. The logistic regression model was used as a benchmark method because of its popularity and ease of use. The logistic regression method resulted in a classification with an accuracy of 62 percent compared to an accuracy of 75 and 77 percent for the machine learning approach and a multi criteria decision method, respectively. Compared to the logistic regression model, the multi criteria decision method and the machine learning method was shown to have several advantages, including handling complex and non-linear data (Kavzoglu et al., 2014).

2.3.2 Weight-of-evidence for LSM

The weight-of-evidence method is an index-based approach used to combine datasets to predict the location of a target variable. Neuhäuser and Terhorst (2007) described how overlaying landslide locations with conditioning factors allows for the statistical evaluation of their significance in past landslide occurrences. Conversely, the Weight of Evidence model calculates the influence of each predictive factor by determining positive (X+) and negative (X-) weights, based on the presence or absence of landslides within the study area. A positive weight indicates that the predictive variable is present at the landslide location and the magnitude of the weight is an indication of the correlation between the predictive variable and the landslides. Concurrently, a negative weight indicates the absence of the predictive variable and shows the magnitude of the negative correlation (Regmi et al., 2014).

Regmi et al. (2014) investigated the application of the weights-of-evidence method and compared it to other index-based approaches for landslide susceptibility mapping of a road section in central Nepal. Validation results showed that all three methods show equal performance. The weights-of-evidence method had a success rate of 75%.

2.3.3 Multi criteria decision analysis for LSM

A multi-criteria decision analysis (MCDA) approach is a type of decision analysis in which decision alternatives are evaluated in terms of explicitly stated criteria (Stewart, 1992). Kamp

et al. (2008) used a multi-criterion evaluation to determine the significance of event-controlling parameters in triggering of the landslides. The study used a multi-criteria Evaluation (MCE) technique. The word “decision” in this context refers to a choice between a set of factors (conditioning factors that explain landslides). These factors are evaluated and weighted to generate criterions. The MCE combines these criteria to construct a single composite that can be used for decision making. Kamp et al. (2008) used a weighting system developed by collecting expert opinion obtained from field work, implying some element of subjectivity.

2.3.4 Machine Learning for LSM

Compared to alternative methods, machine learning techniques excel at mitigating overfitting, capturing the complex non-linear interactions among variables affecting landslide susceptibility and automatically generating the best features and context to achieve high prediction accuracy (Reichenbach et al., 2018). Machine learning methods are also well suited to handle large amount of data, making it especially suitable for classification tasks associated with predicting landslide occurrences. Machine learning can produce repeatable and highly accurate results through continuous learning (Ado et al., 2022). Huang et al. (2020) compared the performance of heuristic, general statistical and machine learning models for landslide susceptibility assessment. The study inferred that machine learning models have higher predictive performance than general statistical and heuristic models. General statistical models were shown to be limited by its linear analysis and heuristic models limited by the subjective weighting process (Huang et al., 2020).

Many machine learning-based models are used in the field by different authors; Ado et al. (2022) points to the random forest (RF), Support Vector Machine (SVM), Linear Logistic Regression (LR) and Artificial Neural Network (ANN) as the most popular models. Liu et al. (2023) looked at data from the Web of Science and found 572 publications between 2009 and 2022 that utilized machine learning with landslide susceptibility mapping. They found that RF, ANN and SVM where the three most prevalent methods applied in the field. Most recent research have compared the performance of different models; Huang and Zhao (2018) compared the performance of SVM with analytic hierarchy process (AHP), logistic regression (LR), ANN and RF. They concluded that the quality and quantity of the data was the most important factor.

Ado et al. (2022) also discussed the recent advancement of hybrid and ensemble methods. Hybrid methods combine multiple machine learning models with optimization and feature selection techniques. Ensemble methods combines several weak machine learning models and uses a voting system to derive the result of the model. Based on results from many studies, Ado et al. (2022) concluded that there is a noticeable improvement in the prediction accuracy with hybrid and ensemble methods. This is because hybrid methods can overcome limitations of conventional machine learning (Ado et al., 2022). Similarly, Liu et al. (2023) compared multiple methods to predict landslide susceptibility in the landslide-prone county of Yunyang in China. The ensemble technique Extreme Gradient Boosting (XGBoost) and random forest (RF) proved much better performance than the others, especially on the training, but also on the testing phase. Liu et al. (2023) concluded that the tree-based methods (XGBoost and RF) showed better performance, especially for larger areas where the training suffers from limited data points.

Ageenko et al. (2022) conducted the first landslide susceptibility mapping in Denmark; by testing three established machine learning algorithms – RF, SVM and LR. Training data, testing data and external validation data were used to predict landslides. The RF model showed superior accuracy, especially on the external validation data. The RF model was shown to have the robustness and potential to be used for landslide susceptibility mapping of Denmark. In their comprehensive review, Liu et al. (2023) examined machine learning-based approaches for Landslide Susceptibility Mapping. They found that traditional machine learning techniques can create reliable susceptibility maps, provided there is access to high-quality data samples. Furthermore, the study highlighted that employing sophisticated ensemble models typically enhances the accuracy of these predictions.

2.4. Machine Learning

Machine learning falls under the umbrella of artificial intelligence (AI) and can be described as computational techniques using past experiences to enhance performance or to generate precise predictions (Mohri et al., 2018). Arthur Lee Samuel, a professor and pioneer of machine learning, defined it as the field of study that allows computers to learn without being explicitly programmed (Aurélien, 2019).

Common for the definitions is that machine learning consists of designing a prediction algorithm. This algorithm attempts to discover patterns in data, referred to as “experience”.

Mohri et al. (2018) emphasizes that in this context, “experience” pertains to previously gathered information accessible to the algorithm, usually existing as electronic data prepared for examination. This experience, the so called “previously gathered information” can be structured in different ways, which affects the machine learning task.

Machine learning systems can be classified based on the amount, and type of supervision they get during the training process. If all data points used in the training have their corresponding ground truth values, the machine learning is defined as supervised learning. In the context of landslide susceptibility, supervised learning occurs when a training set includes labels indicating whether a landslide is present or absent—essentially providing the desired solution to the analysis. Conversely, unsupervised learning uses data points that lack ground truth values, enabling the model to learn “without a teacher.” (Aurélien, 2019).

A typical supervised learning task is classification. Supervised classification is used when the target value belongs to a finite set of classes. Another common task is to predict a target numerical value, such as the price of a house based on a set of features called predictors. This method is called supervised regression. Some regression algorithms can be used for classification and vice versa. Logistic regression is an effective algorithm that can be used for classification tasks. This is because it can output a value that corresponds to the probability of belonging to a given class, for example 50% chance of a landslide occurring (Aurélien, 2019).

As discussed in chapter 2.3.4, machine learning has been extensively used in the field of landslide susceptibility. This is treated as a supervised classification task for two main reasons: (1) it involves supervised learning, as the method relies on ground truth data to determine the presence or absence of landslides, and (2) it is a classification task, as the model predicts whether the outcome falls into one of two categories: landslide present or landslide absent.

2.4.1 Supervised classification

In supervised classification methods, the objective is to identify or predict predefined classes and label new objects as a member of a specific class (Dinov, 2023). This classification algorithm learns from a set of training data to categorize new data points and predicts the class of new entries based on their attributes, which are critical parameters for creating a precise model. Classification tasks can be binary, involving two outcomes such as predicting landslide/ non-landslide, or multi-label, where outcomes exceed two. Additionally, samples

can also be mapped to more than one label (Mandal & Bhattacharya, 2019). Constructing a classification model involves multiple steps; first (1), a data set need to be collected and preprocessed. Second, (2) the classifier model needs to be initialized. Third (3), the dataset is split into training data and testing data and the classifier is fed with the training data. Fourth, (4) the trained classifier is used to predict the class of new observations. Last, (5) the model's performance is assessed by evaluating its error rate on a separate test dataset (See figure 3) (Mandal & Bhattacharya, 2019).

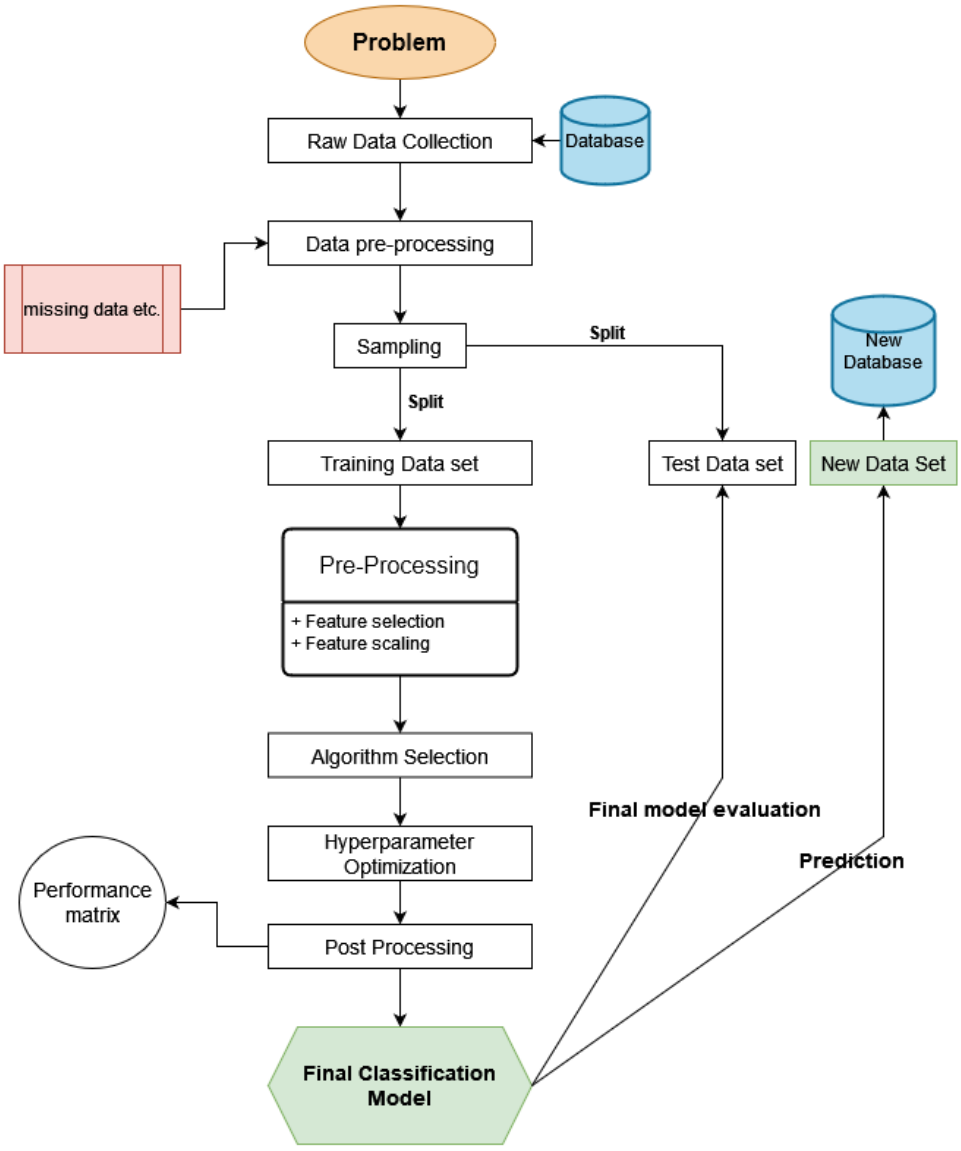


Figure 3: Flowchart showing the workflow of a typical supervised classification task. Source: (Mandal & Bhattacharya, 2019)

2.4.2 Binary Classification

Binary classification is a fundamental concept within machine learning, where the goal is to categorize entities into one of two distinct groups based on a set of features. We can think of these outcomes as positive (1) or negative (0); this exemplification works especially well in the context of LSM where we want to predict an area prone to landslides (1) and areas not prone (0), based on various geographical and environmental features.

Many metrics can be used to measure the performance of a binary classifier. The accuracy of a model is a metric that measures how often the model correctly predicts the outcome. A higher value usually means better model performance (Liu et al., 2023). When using cross validation – a method for assessing how the result of a statistical analysis will generalize to an independent dataset – the accuracy score can often be misleading (Aurélien, 2019). Using an extreme example illustrates an import point: if a dataset is skewed with only 10% of the data representing class 1 (landslide occurring) and 90% being class 0 (landslide not occurring), an accuracy score of 90% can be obtained merely by always guessing the outcome to be 0. In other words, the accuracy score treats all classes as equally important and quantifies all correct predictions. In the field of information retrieval, precision and recall are preferred as metrics for model performance (Aurélien, 2019). A confusion matrix is a way to define the performance of a classification algorithm. Given a binary classification of a dataset, there are four basic combinations of actual data and assigned category: (1) true positive (correct positive assignment), (2) true negative (correct negative assignment), (3) false positive (incorrect positive assignment) and (4) false negative (incorrect negative assignment). These four possibilities can be arranged as a table where rows correspond to the actual class and columns corresponds to predicted class (see table 2).

Table 2: Confusion matrix for binary classification.

Assigned **Actual**	**Test outcome positive**	**Test outcome negative**
Condition positive	True positive	False negative
Condition negative	False positive	True negative

Eight basic ratios exist from the confusion matrix which comes in four complementary pairs, each pair summing to 1. By selecting one ratio from each of the pairs, one can concisely describe a model's performance across different dimensions, providing a balanced view of its

predictive capabilities and areas for improvement. This approach offers a nuanced understanding of model accuracy beyond a single metric, allowing for a more comprehensive evaluation of its performance in binary classification tasks (Aurélien, 2019).

The accuracy score of the positive prediction, called *precision* is a measure of the number of true positive divided by the amount of true positive plus false positives (Aurélien, 2019).

$$precision = \frac{TP}{TP + FP}$$

Precision alone cannot be used as a comprehensive metric because it does not account for the entirety of relevant instances, potentially ignoring a significant number of true positives beyond the single instance it correctly identifies. Precision is therefore typically used alongside another metric called *recall*. This is a measure of the true positive rate and is the ratio of positive instances that are correctly detected by the classifier, described with the following formula:

$$recall = \frac{TP}{TP - FN}$$

The precision and recall metrics can be combined into a single metric, called the F1 score. The F1 computes a "harmonic mean" of the two but puts more to weight to low values. The F1 score therefore only rises when both precision and recall have high values (Aurélien, 2019).

$$F1 = \frac{TP}{TP + \frac{FN + FP}{2}}$$

As the F1 score favors classification that have similar recall and precision, a problem arises; in practice, achieving high precision and high recall is challenging due to the inverse relationship between precision and recall in many classification tasks. Increasing precision typically means being more conservative about labeling an instance as positive, which can lead to missing out on true positives, thus lowering recall. On the other hand, increasing recall means striving to identify as many true positives as possible, which can lead to mistakenly labeling more negatives as positives, thus lowering precision (Aurélien, 2019).

The receiver operating characteristic (ROC) curve is another tool used to assess the usefulness of a classifier. The ROC curve plots the true positive rate (Recall) against the false positive

rate (FPR). The FPR is the ratio of negative instances incorrectly classified as positive. The ROC curve gives a clear visualization of a model's ability to differentiate between two classes at any point along a scale from 0 to 1, which represents the full range of possible cut-off points a model can use to decide whether an instance belongs to one class or the other (Aurélien, 2019).

Several algorithms can perform binary classification tasks, including logistic regression, decisions trees, random forest, support vector machine and gradient boosting algorithms like XGBoost. Random Forests and XGBoost are among the most effective algorithms for binary classification. Random Forests, introduced by Leo Breiman and Adele Cutler in 2001, leverage an ensemble of decision trees to improve prediction accuracy and control overfitting (Breiman, 2001). By averaging the predictions of numerous trees, each trained on a subset of the data, Random Forests provide robustness and reduce the risk of errors from individual trees. This method has been widely adopted for its simplicity and effectiveness across various domains.

2.4.3 Automated Machine Learning

Automated machine learning (AutoML) addresses the fundamental difficulties when applying machine learning to solve real life problems. Every effective machine learning task needs to decide which machine learning algorithm to use on a given dataset, decide how to preprocess its features and how to set up the hyperparameters (Feuer et. al., 2015). As of today, there is no such thing as a perfect machine learning model; no algorithm exist that will guarantee to select the best model for a particular task. AutoML approaches this problem by providing methods and processes to make machine learning available for non-ML experts and to improve efficiency of machine learning.

Auto-Sklearn is one such AutoML framework. Built around the scikit-learn machine learning library for python, auto-sklearn automatically searches for the right learning algorithm for a new machine learning dataset and optimizes its hyperparameters (Feurer et al., 2022). In addition to the performance assessment of different machine learning models, the auto-sklearn includes a stacked ensemble of the tested models. The ensemble utilizes multiple algorithms for the prediction, possibly resulting in improved performance compared to only using one algorithm (Feurer et al., 2022).

Chapter 3

Data Preparation and Data Processing

As the foundation of creating effective machine learning models, the process of data preparation is the most important part. Within the field of landslide susceptibility mapping (LSM), this task emerges as particularly intricate, encompassing both the acquisition of precise landslide inventories as well as the careful selection of relevant conditioning factors (Liu et al., 2023).

This chapter will present the data on landslide locations (landslide inventory), data used for generating the predictive variables, the selection process of the predictive variables, as well as a detailed look into the data processing. Data processing includes the choice of mapping units, sampling ratio and training/testing ratio. In addition to highlighting these key aspects of the study, the decisions will be contextualized by comparing them with previous methods in the field.

3.1. Landslide Inventory Map

The landslide inventory refers to a database containing records of the location, extent, and often the types of landslides in a certain area (Yong et al., 2022). This dataset is the basis for landslides susceptibility mapping, acting as the ground truth needed for a supervised machine learning task. Therefore, the completeness of the landslide inventory is directly related to the accuracy and usefulness of the final susceptibility map. There are multiple methods for obtaining a landslide inventory. Among the most commonly used approaches documented in the literature are field investigations, satellite imagery analysis, Google Earth interpretation, and aerial photograph analysis (Liu et al., 2023).

Field investigation is generally regarded as the most precise method to create a landslide inventory map. The process involves studying an affected area and thoroughly investigating and measuring each potential slope failure. Field surveys for documenting landslides are time-consuming and sometimes impractical due to inaccessibility and high costs. Additionally,

human error can affect the accuracy of data collection (Liu et al., 2023). These factors highlight the need for alternative, more efficient methods.

A remote sensing (RS) approach has the advantage of full coverage of remote sensing images and a strong interpretation capability for small and medium-sized landslides. The remote sensing image survey method can stem from multiple sources, including unarmed aerial vehicles, photogrammetry, aerial photos, satellite images, interferometric synthetic aperture radar (InSAR) and laser imaging, detection and ranging (LiDAR) (Yong et al., 2022). More recently, machine learning has been used to identify landslides. Kreuzer and Damm (2020) developed an automated method using machine learning to filter and classify digital data for landslide inventories, successfully reducing irrelevant data by 91% through a keyword alert system. However, such methods are very complex and the degree of applicability is questionable. These methods are thus usually not introduced in practice as a replacement for traditional and professionally recorded inventory maps (Smith et al., 2021).

In this study, a database containing landslide locations were prepared based on the Norwegian landslide database (NSDB). The Norwegian Water Resources and Energy Directorate (NVE) is responsible for state management tasks regarding preventative measures related to flood and landslide events. NVE are responsible for maintaining a reliable and updated registry system for landslides and floods. Historical landslide events are registered in the NSDB database and consist of over 70 000 separate landslide registrations from multiple sources, including The Norwegian Public Roads Administration (SVV), The Norwegian National Rail Administration (JBV), Norwegian Geotechnical Institute (NGI), The Geological Survey of Norway (NGU), NVE and unaffiliated people through different websites for natural hazard observations[2]. However, the credibility and accuracy of these registrations vary significantly, with many lacking precise information regarding location, initiation time, landslide type, as well as spatial bias in the mapping. According to Fischer et al. (2014), many of the registries mapped by SVV and JBV are mapped on, or very close to the road and railroad network. This variability in data quality directly affects hazard and risk assessment accuracy.

In their 2020 report, NVE presented a recommendation for quality control of registered historical landslide events. The quality of landslide data was categorized into a grading

[2] Unaffiliated people can register landslide events through websites like Regobs: https://www.regobs.no/?SelectedNumberOfDays=3&&NWLat=71.7739410364347&NWLon=-15.996093750000009&SELat=56.43820369358167&SELon=67.6318359375

system ranging from A to D, based on the accuracy of their location, time, and landslide type (Devoli et al., 2020). Given the critical importance of high-quality landslide inventory for LSM, only records that satisfied the spatial accuracy and landslide type certainty of grade A were utilized (see table 3). Because LSM is not capable, nor intended to predict the temporal occurrence of a landslide, the criteria regarding temporal accuracy in the NSDB database could be ignored. As a result, a collection of landslide events with a high degree of spatial accuracy (≤ 10 meter) and a certainty of the landslide type were collected using the software ArcGIS Pro 3.2.0 with SQL expressions on the raw dataset.

Table 3: Table showing the criteria defining the quality grade for historic landslide mapping in the NSDB database. Source: (Devoli et al., 2020).

A	B	C	D
Location: Have to be confirmed from available documentation	**Location:** Have to be confirmed from available documentation	**Location:** Have to be confirmed from available documentation	**Location:** Unsure (no estimate for the uncertainty of the location is defined)
Spreading area: • Landslide point and landslide initiation point have a spatial accuracy +/- 10 meters or more • Landslide initiation area and runout area as a polygon	**Spreading area:** • Landslide point and landslide initiation point have a spatial accuracy +/- 50 meters or more	**Spreading area:** • Landslide point have spatial accuracy > 50. (No registered accuracy is not accepted).	Both landslide events that are quality controlled and which is not quality controlled can have this quality grade.
Date and time: Correct date, time +/- 3 hours or better	**Date and time:** Date +/- 1 day, unknown time of day	**Date and time:** Date unsure (temporal accuracy > 2 days).	
Landslide type: Correct (easy to interpret from available information **No duplicates:** (Inside quality grade A or B)	**Landslide type:** Correct (easy to interpret from available information **No duplicates:** (Inside quality grade A or B)	**Landslide type:** Unsure, or unspecified	

3.2. Conditioning Factors for LSM

In addition to landslide data (the "dependent" variable, also called target variable), landslide susceptibility mapping requires information on geoenvironmental factors, called "conditioning factors". The conditioning factors in LSM act as the explanatory variables in a supervised machine learning task. The conditioning factors are used to explain the target variable by analyzing the relationship between the conditioning factors and the target variable. As one of the keys to create effective prediction models, selecting the conditioning factors is

another important step in landslide susceptibility mapping. According to Reichenbach et al. (2018), 596 such conditioning factors exist. It is unrealistic for this study to take them all into consideration; A select group of variables will be used.

The selection of conditioning factors for this study was guided by prior research, particularly the work of Reichenbach et al. (2018). This foundational study identified topographical, hydrological, geological, land cover, and anthropogenic factors as significant contributors to landslide susceptibility. The choice also reflects the availability of data for the region and the feasibility to produce such data. A more detailed discussion on this topic will be provided in Chapter 4.

Reichenbach et al. (2018) classified variables into five classes:

(i) Morphological
(ii) Geological
(iii) Land cover
(iv) Hydrological
(v) Other variables

3.2.1 Morphological Variables

In LSM, morphological variables are obtained by processing terrain elevation data, often in the form of a digital elevation model (DEM) (Reichenbach et al., 2018). A DEM is a raster representation of a three-dimensional modeling of the Earth's topography (Saleem et al., 2019). For this study, a DEM was downloaded based on elevation data from the project "*Nasjonal detaljert høgdemodell*" – a detailed mapping of elevation in every square meter of Norway, carried out between 2016 and 2022. The data was downloaded in the GeoTIFF format with a resolution of 5 meters[3].

DEM derivatives have proven particularly effective for predicting the spatial distribution of landslides. Although results from LSM vary based on differences in local triggering factor and quality of available data, the importance of DEM derivatives for LSM is unquestionable. Fabbri et al. (2003) showed how a predictive model that used DEM derivatives, including elevation, aspect and slope, performed much better than a model using geological,

[3] The DEM was resampled to a spatial resolution of 5 meters to better capture the surface characteristics contributing to landslide in soil formation. The DEM was downloaded from hoydedata.no: https://hoydedata.no/LaserInnsyn2/

depositional and land use data, and slightly better than a model that combined the variables. According to Reichenbach et al. (2018), simple and direct measures of the Earth's surface is the most used conditioning factors. This is in part because they can be generated using a DEM and because there exist theoretical reasonings to justify their use. The following paragraphs describe the morphological variables used in this study.

The raw product of a DEM is an elevation representation of the surface. *Elevation* is often used in LSM as it functions as a proxy for variability in slope gradient, lithology, weathering, precipitation, ground motion, soil thickness and land use (Sidle & Ochiai, 2006). Reichenbach et al. (2018) report that numerous studies have established significant statistical correlations between elevation and landslide frequency. However, elevation alone does not provide a direct physical explanation for landslides. Despite this, elevation remains a popular variable, frequently utilized as a direct measure of terrain morphology (Reichenbach et al., 2018).

The gradient of a *slope* is considered as an essential conditioning factor for determining slope instability (Ado et al., 2022). The slope gradient represents the quantitative measure of the steepness of the terrain and has a direct control on the balance of the retaining and destabilizing forces that acts on a slope. Sidle & Ochiai (2006) points out that although slope gradient is important related to landslide initiation, the wide range in the lower limit of slope gradients known to trigger different types of landslides shows that other geomorphic, as well as hydrological and geological factors are often equally important determinants of slope stability.

In addition to the gradient of the slope, the shape of the slope exerts a strong influence on the slope stability in steep terrain by concentrating or dispersing surface and primarily subsurface water in the landscape (Sidle & Ochiai, 2006). Part of a surface can be concave, convex or planar (straight). Profile *curvature*, aligning with the slope's direction, influences flow acceleration or deceleration on the surface. An upward convex shape will generally decelerate the flow, while upward concavity accelerates the flow across the surface. Planform curvature align perpendicular with the slope direction and affects flow convergence or divergence across a surface. According to Sidle & Ochiai (2006), laterally and upward convex landforms are most stable in steep terrain, followed by planar landforms and concave landforms (least stable) (see figure 4). For laterally convex slopes, water is dispersed while concave slopes tend to concentrate water into small areas of the slope, generating rapid pore water pressure.

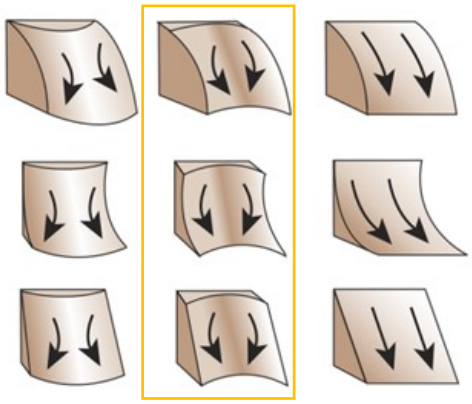


Figure 4: Curvatures where both profile and planforms curvatures are combined. The columns show planform curvature and rows show the profile curve. Orange square show what Sidle & Ochiai (2006) describes as the most unstable landforms. Source: (Esri, n.d-f).

Neither slope nor elevation alone exerts a dominant control on slope stability, although these variables interact with other factors to influence landslide susceptibility. *Slope aspect* is a representation of the direction of the downhill slope, expressed as the compass direction the surface faces. According to Sidle & Ochiai (2006), slope aspect affects hydrological processes via evapotranspiration, thus affecting weathering processes, vegetation and root development. Moreover, on the Northern Hemisphere, north-facing slopes receive higher and less variable moisture compared to south-facing slopes, resulting in greater weathering and thus increase in landslide susceptibility (Sidle & Ochiai, 2006). However, the use of slope aspect as an influencing factor is less justified compared to other metrics, and may be controlled by local conditions. While aspect characteristics have been shown to influence landslide susceptibility in some areas, other areas don't (Sidle & Ochiai, 2006).

Other more complex morphological variables that captures the overall morphology of the entire slope or sub-catchments are also shown to be good predictors for slope instability. The expression of the *roughness* of the surface is such metric; computed by the standard deviation of elevation or slope, high roughness is expected to be larger in areas susceptible to landslides (Reichenbach et al., 2018). In their 2008 study, García-Rodríguez et al. (2008) used logistic regression to evaluate probability of landslides for El Salvador. The results of the study illustrate the importance of terrain roughness for LMS.

Another variable that is extensively used to express terrain roughness is the *Terrain Ruggedness Index* (TRI). TRI can be extracted from a DEM and serves as a measurement of the land surface condition, characterizing terrain by its smoothness or ruggedness. It captures the local variations in curvatures and gradients (Saleem et al., 2019). The usability of surface

roughness to delineate surface morphological units has been invigorated because of high resolution topographical data. It is wildly used in LSM, for example by Ageenko et al. (2022) to map landslide susceptibility in Denmark. In the case study, using a random forest classification model, TRI was shown to be the most important influencing factor, contributing more to the model's decision making than any other factor.

A DEM can also be utilized to distinguish topographic features in the terrain, such as hilltops, valley bottoms, exposed ridges, flat plains and upper and lower slopes. The *Topographic position index* (TPI) was introduced by Weiss (2001) and has been applied to a number of landslide susceptibility studies. TPI is computed as the difference between the cell elevation and the mean elevation of neighboring cells. The TPI factor is considered a geomorphological landslide conditioning parameter because landslide initiation events usually take place on ridges (Saleem et al., 2019). TPI have been extensively used in previous research. Ageenko et al. (2022) used TPI in conjunction with other factors to predict landslides in Denmark.

TPI is expressed as a continuous raster where positive cell values represent a cell that lays higher in the terrain than its surrounding cells and a negative value represents sunken features in the terrain (Weiss, 2001). The range of TPI values depends not only on elevation differences, but also the predetermined radius of which the mean elevation of neighboring cells is calculated. Weiss (2001) used this fact by using the parameters from two neighborhood sizes (300 and 2000 meters) and combining them in order to identify complex landscape features. More recently, De Reu et al. (2013) also employed TPI to distinguish features from deeply incised streams to open flat areas.

3.2.2 Geological Variables

Geological variables are extensively used in LSM. According to Reichenbach et al. (2018), the most common geological information is the type of rock present in the area. The justification for employing this variable remains ambiguous since geological maps predominantly depict the bedrock, detailing chronostratigraphic units and formations. These aspects may not directly correlate with the mechanical characteristics of the materials implicated in landslide initiation (Reichenbach et al., 2018). Studies have also used data on rock structure and bedding attitude for LSM. Sidle & Ochiai (2006) summarized the most important geological factors for landslide initiation as weathering of regolith, bedrock structure and tectonics.

As this study deals with landslides in soils / loose materials, data on *surficial deposits* were used as an influencing factor. Landslides have been associated with different geological materials around the world. For the generation of a nation-wide susceptibility map of Norway, Fischer et al. (2014) classified the surficial deposits in Norway based on how sensitive the respective surficial deposits generally are to the triggering of landslides. However, as this study uses a machine learning approach for LSM, such weighting was not used. Nevertheless, the dataset provided by NGU for surficial deposits cover, was employed in this study[4].

Furthermore, Fischer et al. (2014) found it impractical to incorporate bedrock geology into the LSM model due to the significant variations in bedrock geology across Norway. Consequently, data on bedrock geology were not utilized in this study

Soil conditions are relevant to the field of landslide susceptibility for several reasons, including water infiltration, soil sheer strength and stability of slopes (Sidle & Ochiai, 2006). As the influencing factor describing soil condition, data from the Norwegian Institute of Bioeconomy Research (NIBIO) were used[5]. Encompassed under the "Geovekst"-cooperation, NIBIOs detailed data on the soil thickness, soil type and soil cover prevalence are used (NIBIO, 2020). The classification system categorizes the land based on soil depth, organic soil layer thickness, and human impact. Organic layers (1) thicker than 30 cm in peat and forests, and 20 cm in agricultural land, differentiate areas for cultivation, peatlands, and bogs. Soil-covered areas (2) have more than 50% of the land with soil depth over 30 cm, while shallow-soiled lands (3) have more than 50% of the area with less than 30 cm depth but are not classified as bare rock. Bare rock areas (4) are predominantly exposed rock, block fields (5) consist mainly of stone blocks from various origins, and constructed grounds (6) are significantly altered by human activities, indicating low biological productivity.

3.2.3 Land Cover related Variables

Land use activities and disturbances related to land cover affect the magnitude, frequency and type of landslide (Sidle & Ochiai, 2006). The major widespread land use activities that influence mass wasting events include timber harvesting, forest conversion, grazing, recreation, as well as concentrated human activities like building of roads, urban development

[4] The dataset can be downloaded from the National surficial deposits database: https://geo.ngu.no/kart/losmasse_mobil/
[5] More information on the dataset can be found from the NIBIO website: https://www.nibio.no/tema/jord/arealressurser/arealressurskart-ar5/grunnforhold

(Sidle & Ochiai, 2006). In the context of LSM, the majority of studies have used combinations of vegetation, land cover and land use data derived from existing maps or data processed by satellite imagery (Reichenbach et al., 2018).

Data on *land cover* information is widely used is LSM, its theoretical rationale being that land cover conditions slope stability. However, varied local conditions makes it difficult to understand the functional link between land use and landslides. Promper et al. (2014) found that development of new building areas in Austria resulted in landslide risk hotspots. Carrera et al. (1991) showed that a land cover of forests was a good predictor for slope stability. Yet, previous studies from the same area in Italy showed contradicting results. Nevertheless, land cover data is a commonly used in the field and will also be used as an influencing factor in this study. Encompassed under the "Geovekst"-cooperation, NIBIOs dataset on land use types shows the main types of land use classes, including built-up areas, agricultural land, forests, barren land, peatland, snow/ice cover and freshwater[6]. This dataset is used as a representation on the land cover for this study.

Measurements of the physical characteristics of the vegetation cover is also extensively used in LSM. Vegetation cover contributes to slope stability, primarily through the water uptake capabilities of plants, which result in vegetated slopes being, on average, 12.84% drier and exhibiting matric suctions that are threefold higher compared to fallow slopes (Liu et al., 2023). Therefore, the *normalized difference vegetation index* (NDVI) is considered as one of the main land cover factors, despite lack of a clear relationship between vegetation cover and slope instability (Reichenbach et al., 2018). NDVI is used to quantify vegetation density by observing the rate of absorption and reflection of visible and near infrared sunlight. NDVI was obtained using Google Earth Engine and sentinel 2 satellite data from 2022. Methodology will be discussed further in chapter 4.

The *loss of forest cover* has also been shown to have an effect on slope stability. According to Sidle & Ochiai (2006), deforestation in steep terrain can reduce site stability by causing root strength deterioration and increasing soil water due to lower evapotranspiration. In their 2020 report, Nordrum et al. (2020) looked at stabilizing effects of vegetation cover in Norway. Including the effects mentioned by Sidle & Ochiai (2006), interception of water was shown to be over 50% of the annual precipitation in dense forests and 30% in more open areas with

[6] The dataset was downloaded from Geonorge.no. Access is restricted to only the partners of the Geovekst corporation. https://kartkatalog.geonorge.no/metadata/arealressurskart-fkb-ar5/243751e8-5803-4627-898c-d0ddabe82056

scarce vegetation cover. In areas of timber harvest, the root strength deteriorates dramatically some years after the deforestation and then increases as a function of new growth (Nordrum et al., 2020). This presents a problem: the temporal discrepancy between the available forest loss data, which summarize net changes over extended periods, and the specific timing of landslides in localized areas. This discrepancy makes it difficult to assess the immediate impact of forest loss on slope stability for individual landslide events, obscuring the direct temporal relationship between deforestation and landslide occurrence in any specific location. The machine learning methodology employed in this study will include steps to check for irrelevant features to remove possible variables that acts as "noise" in the prediction. Consequently, data on forest loss will be used. The "Hansen global forest change" dataset was used in this study. By using Landsat images between 2000 and 2022, forest loss was quantified based on vegetation cover 5 meter or taller (Hansen et al., 2013). The methodology is described in more detail in chapter 4.

3.2.4 Hydrological Variables

Hillslope hydrology is another crucial factor to consider as it exerts a major control on landslide initiation. According to Sidle & Ochicai (2006), hillslope hydrology effects landslides through precipitation, water recharge into soils (potential for overland flow), movement of water within the regolith and evapotranspiration and interception of water. All these factors help control the amount of groundwater in different parts of a hillslope, thus controlling the potential for slope failure. As an influencing factor for LSM, Liu et al. (2023) points to data on aridity, the topographic wetness index (TWI), the index of moisture and data that describes long-term erosion effects in the terrain as the most important hydrological factors used in LSM.

The *Topographic Wetness Index* (TWI) is a commonly used variable to predict the amount of soil moisture at the catchment scale and measures where water traverses and how much accumulates in different spots (Saleem et al., 2019). The TWI was originally developed by Beven & Kirkby in 1979 and is computed using the local upslope area draining through a cell and the local slope. Many researchers have incorporated TWI as a factor in LSM. Notably, Ageenko et al. (2022) identified TWI as one of the most critical variables in their landslide susceptibility study in Denmark. Hong et al. (2018) used TWI alongside other commonly used influencing factors to predict landslides in China. The TWI was shown to have one of the highest predictive weights associated with landslide in the region.

The *Stream Power Index* (SPI) is another commonly used conditioning factor that can be derived from a DEM. The SPI also uses the local upslope drainage area and slope angle for the computation, but unlike TWI, SPI utilizes the amount of water accumulation and the slope to predict the potential for the erosive power of flowing water on the terrain (Saleem et al., 2019). The SPI parameter is also widely used in LSM (Reichenbach et al., 2018).

Because of the erosive effects of rivers, the *distance from rivers* is shown to have a great influence on landslide formation (Liu et al., 2023). Distance from rivers can be expressed as the distance allocation for each cell to the provided river source, which typically is some vector representation of the river system for the area. For this study, a representation of the river network was downloaded from the ELVIS river network database, provided by NVE[7]. A number of researchers have used distance to rivers as an influencing factor. According to Reichenbach et al. (2018), the "distance from river" variable has strong a theoretical rationale; the distance captures hydrological conditions that negatively affect the slope stability towards a river or stream due to the concentration of groundwater flow and the destabilizing effect of river incision.

3.2.5 Other Variables

In addition to the abovementioned variables that can be easily categorized into classes, other factors fall outside this division. Based on the classification by Reichenbach et al. (2018) and similar classification done by Liu et al. (2023), natural factors and human-related factors is accommodated in this section.

Spatial patterns of precipitation are closely associated with landslide initiation (Sidle & Ochiai, 2006). This is particularly relevant for landslide initiation in Norway where the most common triggering mechanism are related to a buildup of pore water pressure in loose materials, often due to precipitation and/or snowmelt in the spring and heavy rainfall in the autumn (NVE, 2014). Furthermore, total rainfall and short-term intensity of rainfall are rainfall attributes that strongly affect the buildup of pore water pressure in unstable hillslopes. Consequently*, annual mean rainfall* is usually used in LSM (Liu et al., 2023). For this study,

[7] The ELVIS database represents all catchment related elements, notably rivers and smaller stream as vector polylines. Downloaded from the NVE Atlas: https://atlas.nve.no/Html5Viewer/index.html?viewer=nveatlas#

data on mean annual rainfall between 1991 and 2020 from The Norwegian Meteorological Institute (MET) were used[8].

Temperature has also been shown to have an effect on landslide formation because the resistance of soils on hillslopes weakens at cooler temperatures. Liu et al. (2023) recommends using annual temperature as an influencing factor, a practice exemplified by Ageenko et al. 2022 and Yao et al. (2023). For this study, data on *mean annual temperature* between 1991 and 2020 from The Norwegian Meteorological Institute (MET) were used[9].

Anthropogenic influences are another major factor in landslide initiation. As one of the major transformations of nature caused by humans, road networks directly destabilize hillslopes (Sidle & Ochiai, 2006). Roads affect slope stability by altering natural pathways where water runs and where water accumulates, undercutting unstable slopes and by overloading and oversteepening the fill slope (the inclined slope extending from the outside edge of the road shoulder). Furthermore, studies have demonstrated that areas with unimproved roads, particularly forest roads, experience an increase in landslide erosion by approximately two orders of magnitude compared to undisturbed forest land, and by one order of magnitude relative to deforested areas (Sidle & Ochiai, 2006). Nevertheless, using the *distance from roads* as an influencing factor for LSM is shown to have questionable validity (Reichenbach et al., 2018). Also, the presence of a road on a hillslope has different effects on the slope above and belove the road; distance accumulation does not take this into account (Reichenbach et al., 2018). Still, distance from roads is commonly used and will be used in this study as well. For this study, a representation of the road network was downloaded from the FKB-Veg road network dataset, provided by The Joint Geospatial Database (FKB)[10].

[8] Data on mean annual precipitation between 1990 and 2020 can be downloaded from the thredds server: https://thredds.met.no/thredds/catalog/KSS/Gridded_climate_normals_1991-2020/precipitation/catalog.html
[9] Data on mean annual temperature between 1990 and 2020 can be downloaded from the thredds server: https://thredds.met.no/thredds/catalog/KSS/Gridded_climate_normals_1991-2020/temperature/catalog.html
[10] FKB-Veg contains detailed information about all public and private road network for Norway. The dataset can be downloaded from Geonorge.no: https://kartkatalog.geonorge.no/metadata/fkb-veg/4920b452-75cc-45f2-964c-3378204c3517?search=veg

Table 4: Overview of the influencing factors used in this study.

Category	Variables	Type	Spatial Resolution	Source
Geomorphological	Elevation, from DEM	Continuous	5 m	NMA
	Slope, from DEM	Continuous	5 m	-
	Profile curvature, from DEM	Continuous	5 m	-
	Aspect, from DEM	Continuous	5 m	-
	Roughness, from DEM	Continuous	5 m	-
	TPI, from DEM	Continuous	5 m	-
	TRI, from DEM	Continuous	5 m	-
	Landforms, from DEM	Categorical	5 m	-
Hydrological	TWI, from DEM	Continuous	5 m	-
	SPI, from DEM	Continuous	5 m	-
	Distance to rivers	Continuous	5 m	NVE
Geological	Surficial deposits (Quaternary deposits)	Categorical	1:50 000	NGU
	Soil condition	Categorical	1:1000	NIBIO
Land Cover	Land use, from FKB AR5	Categorical	1:1000	NIBIO
	NDVI	Continuous	30 m	GEE
	Forestless	Categorical	30 m	GEE
Other (Anthropogenic, climatic)	Average annual temperature	Continuous	30 m	MET
	Average annual precipitation	Continuous	30 m	MET
	Distance to roads	Continuous	5 m	NMA

3.3. Data processing

Constructing the landslide inventory database and choosing the influencing factors for LSM can be described as data preparation. After the data is prepared, the next step involves processing the data.

3.3.1 Mapping units

Using landslide inventories and influencing factors to evaluate the likelihood of a landslide occurring in an area requires the selection of a suitable mapping unit (Guzzetti, 2006). In LSM, choosing a mapping unit is crucial because the model's output segments the land by susceptibility levels, and when utilizing multiple predictive raster factors, these values must be assigned within a defined framework that partitions the map accordingly. Based on the concept of distinct and clearly definable mapping units, Guzzetti (2006) summed all methods in the literature as either: (i) grid cells, (ii) terrain units, (iii) unique condition units, (iv) slope units, (v) geo-hydrological units, (vi) topographic units, and (vii) political or administrative units.

According to Reichenbach et al. (2018), grid cells (i.e. “pixels”) are the most common type of mapping unit (see Figure 5b). Grid cells divides the land into regular drawing cells of the same size which become the mapping unit of reference. Grid cells are favored for raster-based GIS applications in landslide susceptibility mapping. They allow each cell to hold values for the various influencing factors. Despite the challenge of accurately representing continuous landscapes and features within discrete cells, advancements in computing power have significantly mitigated this issue, enabling the use of smaller grid cells that more accurately reflect terrain characteristics (Guzzetti, 2006).

An alternative to grid-cells is slope units; Slope units partition the land into hydrological regions bounded by drainage and divide lines (Guzzetti, 2006) as shown in Figure 5a. Slope units help solve some of the drawbacks when using grid-cells. Since landslides occur primarily on hillslopes, slope units are in theory well suited for LSM. Slope units can correspond to an individual slope, many adjacent slopes or a small catchment. In contrast to grid-cells, conditioning factors obtained for individual slope units do not reflect the value for one cell. Instead, the value refers to the entire terrain subdivision, providing more meaningful results. As an example, descriptive statistics (e.g. mean) of continuous conditioning factors (e.g. slope) in a slope unit are better predictors of landslide presence compared to the same

indices computed for a single cell in a DEM. However, slope units are difficult to prepare, the process is prone to errors and the size of the slope unit directly influence what kinds of landslides will be mapped, necessitating tailoring the size to the known landslide in the study area. Slope units have therefore rarely been used in LSM (Reichenbach et al., 2018).

Alvioli et al. (2016) proposed a tool for automatic delineation of slope units. The software is written in python for the GRASS GIS software and requires a digital elevation model (DEM) and a set of user-defined input parameters. The tool works by dividing the landscape into large half basins based on hydrological and terrain aspects, then iteratively subdividing these into smaller, more defined slope units that reflect the natural terrain divisions. This process is fine-tuned through user-defined parameters, where the most important ones are (1) the minimum area of a slope unit and (2), how much variation in terrain aspect is allowed within one slope unit (Alvioli et al., 2016). For this study, slope units were made for the study area using a minimum slope area of 10 000 m2 and a low variation in terrain aspect (0.1). Figure 5 shows a comparison between grid cells and slope units[11].

For this master study, grid cells were chosen as the preferred mapping unit. In addition to being the most used method, grid cells have the advantage of being simple to process at different resolutions. GIS also make it easy to use cell-based (raster) data, for instance terrain elevation data that can be used to compute further derivatives (Guzzetti, 2006). It should be stated that the grid cell approach also has some clear disadvantages. Firstly, there is no physical relationship between dividing the map based on grid-cells and landslides. Second, the majority of models that adopt grid-cells as the mapping unit uses cells having the same spatial resolution as the DEM. This can be problematic because the size (resolution) of the grid cells determine what kinds of landslides characteristics can be captured. Reichenbach et al. (2018) emphasizes how small grid cells (<5m) allows for the capture of small, shallow landslides but might have little or no correlative significance for larger, deep-seated landslides whose signature is better captured by larger grid cells. Lastly, the use of grid-cells can often result in zonations that is difficult to use in the real world because cells within a small geographical area might have very different susceptibility scores.

Slope units were also considered. However, despite the automated approach proposed by Alvioli et al. (2016), the delineation still necessitates fine-tuning. It is also much more time-

[11] A description if the Slope Unit delineation software can be found here: https://geomorphology.irpi.cnr.it/tools/slope-units

consuming to calculate statistical values for entire slope unit polygons compared to extracting single cell values from raster data. However, it should be noted that the use of slope units has been demonstrated to yield better assessment results than using grid cells, as exemplified by Ba et al. (2018). Comparing results between grid cells and slope units was considered beyond the scope of this thesis.

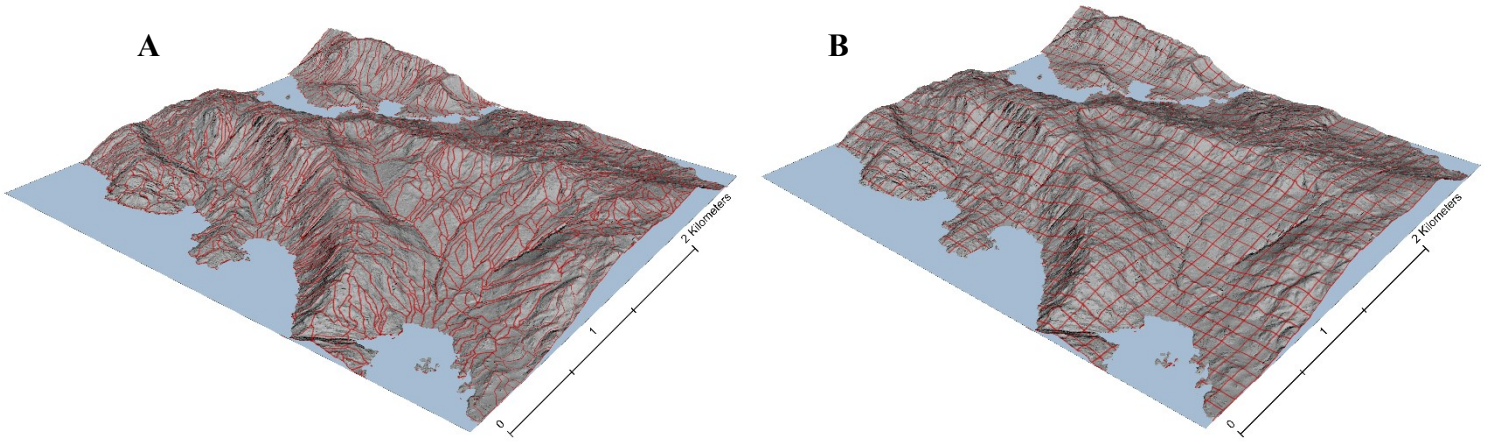


Figure 5: A snippet from the study area showing generated slope units (A) and grid cells (B). The slope units were created using the Alvioli et al. (2016) method, with a minimum slope unit size of 10 000 m2. The grid cells have a size of 5 x 5 meters, the same size as the resolution of the influencing factors but are enlarged in the figure for better visualization.

3.3.1 Sampling ratio and training/testing ratio

The sampling ratio (ratio of landslide samples and non-landslide samples) plays an important role on the outcome of the modelling. According to Liu et al. (2023), a sampling ratio of 1:1 (same amount of landslide and non-landslide samples) is commonly adopted by researchers. However, research has shown that different datasets and models usually have an optimal sampling ratio that can improve the accuracy of the LSM. Yang et al. (2023) explored this aspect of LSM by testing different machine learning algorithms on different sampling ratios. They tested sampling ratios of 1:1, 1:2, 1:4, 1:6, 1:8 and the result showed that a slight change in the ratio caused a visible change in the distribution of different susceptibility areas. For the random forest model, optimizing the sampling ratio resulted in more areas described as highly susceptible and improved the overall accuracy of the model by 1,3% (Yang et al., 2023). Implementing the method developed by Yang et al. (2023) is time-consuming and similar methods are seldom used by other researchers. Because of the space constraints associated with this study, a sampling ratio of 1:1 was used for this study.

Another important decision arises after the data is fully processed and prepared; the data need to be randomly split into two separate datasets, one for training the data and one for testing the accuracy of the trained classifier. This ratio can affect the performance of the model. Sahin et al. (2020) analyzed the effect of different training/testing ratios, more specifically the ratios 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20 and 90:10. Using the random forest algorithm, Sahin et al. (2020) used different evaluation methods to conclude that splitting the landslide inventory data into 70% for training and 30% for testing showed the best results. This 70/30 ratio is commonly used. However, randomly splitting the samples into a training set and a testing set cannot guarantee that both the training and testing sets contain a proportionate representation of landslide and non-landslide samples (Liu et al., 2023). Ageenko et al. (2022) improved this 70/30 ratio by ensuring a comparably equal number of landslide and non-landslide samples in both the training and testing datasets through stratified splitting based on the presence or absence of landslides. Similar method was utilized in this study; splitting training and testing in 70/30 ratio using stratification to prevent bias in the model training and testing.

Chapter 4

Methodology

Figure 6 presents a general overview of the methodological framework for this thesis. First, the area of interest (AOI) is defined for both training and external validation. Data was collected from multiple sources. Factors influencing landslide formation were created, some are DEM derivatives, others are from other sources. Point data representing landslide locations and non-landslide locations were then created. Landslide data were extracted from the NVE landslide database and refined to ensure high spatial accuracy. Non-landslide data were randomly generated in the study area to facilitate the binary classifier. The influencing factors were then mapped onto the landslide/non-landslide point data to act as the independent variables for the binary classification. Feature selection was then used to refine the dataset and resolve issues of multicollinearity. Variance Inflation Factor (VIF) and Pearson correlation coefficient (PCC) were used to check for multicollinearity and remove several sets of correlated variables.

The preparatory phase sets the stage for the application of the machine learning classifier. The dataset containing landslide and non-landslide points with their related independent variables is used to train the RF and the AutoML model. A subsection of the data is used for prediction. An independent dataset outside the AOI is used to test the predictive power of the two models, ensuring the model's generalizability and reliability. Finally, the trained RF model is applied on the entire stack of influencing feature raster data to create a landslide susceptibility map where the output for each pixel is a probability score between 0 (very little chance of landslide occurring) and 1 (very high chance of landslide occurring).

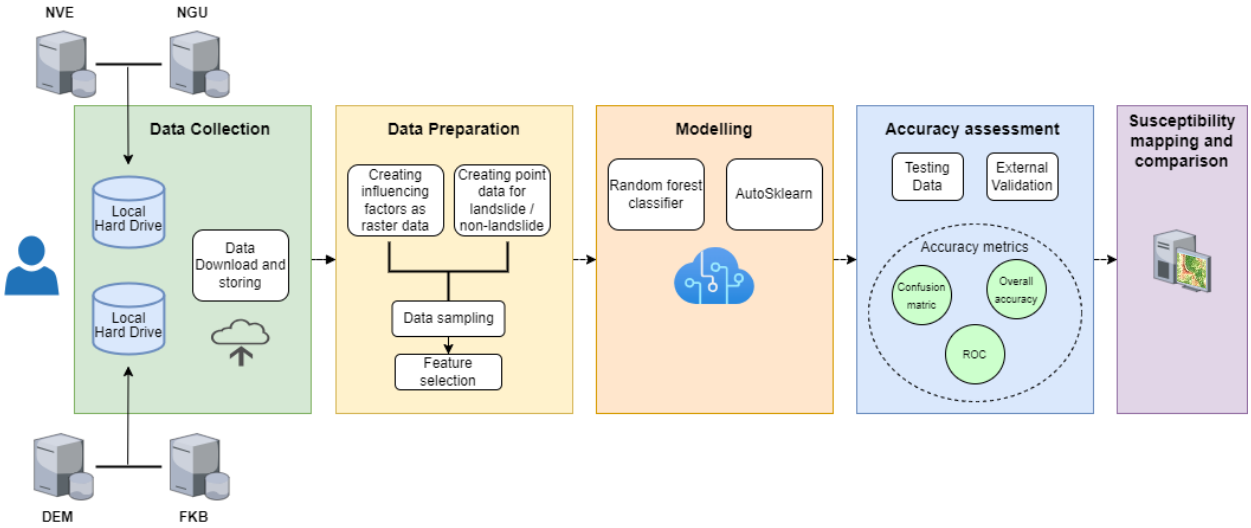


Figure 6: Methodological framework for the thesis.

4.1. Presentation of the analyzed areas

The susceptibility mapping was carried out for a region in the middle of Vestland county in the westernmost part of Norway. An area was chosen for the training and predicting of the machine learning model, another smaller area was chosen for external validation. The chosen area of interest (AOI) used for the training and prediction, shown in Figure 7, covers approximately 6 478 km2 and is located between Haugesund in the south, The Sognefjord in the north and the county border in the east. The chosen AOI used for external validation covers approximately 1 798 km2 and is located just North of the AOI for training, north of The Sognefjord between Florø in the west and Jostedalsbreen glacier in the east.

The choosing of the AOI represents a balance between the desire to include many historic landslide registries and the need to maintain the computational demands and data storage requirements of the influencing factor raster data. Furthermore, Vestland emerges as a particularly relevant study location for landslides in soil; Vestland is dominated by steep hillslopes with loose deposits, making it particularly prone to landslides. Vestland also receive most precipitation in Norway. According to Sidle (2006), total rainfall and short-term rainfall intensity strongly influence landslide initiation, both of which have been shown to have increased by between 15 to 20% for certain regions in Vestland (Øydvin. et al., 2011).

For clarity throughout this chapter, the term “AOI” will refer to both the area of interest utilized for training/prediction as well as the area used for external validation. Selecting the appropriate AOI is crucial for the effectiveness of the machine learning model in applying

predictive variables and identifying the landslide inventory. Given the vastness of Vestland county, using the entire area as the AOI would be impractical due to high computational and data storage demands. A balanced approach was employed by initially considering the entire Vestland county, which was then segmented into areas for training and external validation. By utilizing The Joint Geospatial Database (FKB) standards, the focus was narrowed to FKB-A and FKB-B areas. These subcategories represent urban and developing areas with a high degree of utilization, which are of particular interest due to their higher population density and infrastructure significance—factors that amplify the risk and impact of landslides (Geovekst, 2016). In contrast, FKB-C and FKB-D areas represent rural, cultivated, mountainous, forested land, areas with a low land value (see figure 7).

This applied strategy effectively narrows the study's scope to areas where landslide vulnerability is highest while also optimizing computational resources and ensuring data quality. As a result, the AOI is approximately 40% the size of the original area but keeps 75% of the landslide registries.

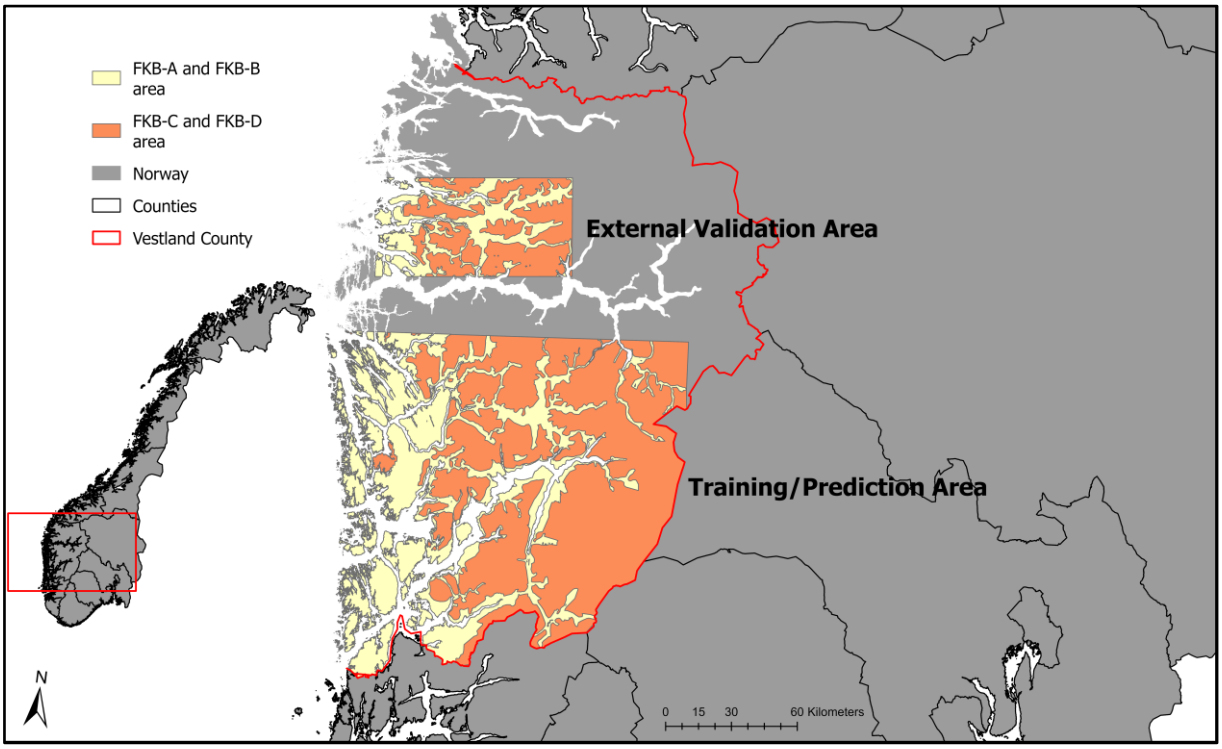


Figure 7: Figure showing how Vestland county was subdivided into two appropriate areas, one for training / prediction and one for external validation. For both areas, only what is defined as FKB-A and B areas was selected to be the final AOIs.

4.2. Data Preparation

The data preparation phase includes cleaning, transforming and organizing the raw data into a format that the machine learning algorithm can understand and use to predict landside susceptible locations. For this study, two primary types of data are used; raster data for influencing factors and vector point data for the landslide inventory.

4.2.1 Creating point data for landslide / non-landslide

As stated in chapter 3, the raw data for landslide inventory is downloaded from the NSDB database. For the AOI, 21 520 landslide registries are found. The landslide registries are divided into two datasets; one dataset encompass registered landslide initiation/runout points and the other is a more general database of landslide registries. These registries encompass various landslide types, with spatial accuracy ranging from exact locations to a margin of error up to 10,000 meters. The database also includes historic records dating back to the year 1100. Considering the dynamic nature of development and land-use changes, this study chose to include only landslide records after and including the year 1980. While other research, such as Rød et al. (2012), has considered records dating back to 1650, this approach was deemed unnecessary for the current analysis. Between 1650 and 1980, only 16 landslide events meeting the selection criteria were recorded, reinforcing the decision to focus on more recent data. For the general landslide dataset, an SQL query was used to extract only the records that satisfy landslide type 140, 141, 142, 143, 144, 145 and 160[12]. The selected landslide classes represent different types of landslides in soil, similar to the classification method described in chapter 2.1 (NVE, 2018). In addition, the records also needed to satisfy a spatial accuracy less than or including 10 meters. The dataset for landslide initiation and runout points was further filtered to include only landslide initiation points. This adjustment aligns with the study's objective to explore areas susceptible to landslides, rather than the pathways landslides travel downslope. The resulting SQL query narrowed down to dataset to 743 records for the training data and 197 for the external validation area. This dramatic filtering of 96 % is deemed necessary, especially regarding the landslide type and the accuracy of the mapping.

[12] Following the classification scheme provided by NVE; landslide type **140** = landslide in soil (unspecified). **141** and **143** = landslide in clayley soil. **142** = debris flow/debris flood. **144** and **145** = debris flow/debris slide/debris avalanche. **160** = shallow planar soil slide. Source: (NVE, 2018)

Table 5: Filtering of the landslide registries based on AOI, data, landslide type, initiation and locational accuracy.

Training Data:

Operation type	Filter	**Landslide registries**	**Landslide initiation/runout points**
		Number of events	Number of events
Select by Location	AOI	13 083	3955
Select by Attribute	Year ≥ 1980	7410	3955
Select by Attribute	Landslide type (initiation points for initiation/runout dataset)	1243	653
Select by Attribute	Locational accuracy ≥ 10 m	652	90

External Validation Data:

Operation type	Filter	**Landslide registries**	**Landslide initiation/runout points**
		Number of events	Number of events
Select by Location	AOI	3795	687
Select by Attribute	Year ≥ 1980	2010	687
Select by Attribute	Landslide type (initiation points for initiation/runout dataset)	356	66
Select by Attribute	Locational accuracy ≥ 10 m	182	15

According to Reichenbach et al. (2018), event inventories (landslides caused by a single trigger, e.g., an extreme rainfall event) are the second most used inventory data in the literature. Event inventories are good at evaluating the predictive capability of the model. However, these records need be used with caution as the geographical distribution and abundance depend on the pattern of the trigger, pattern of the geo-environmental variables and how detailed the mapping has been carried out (Reichenbach et al., 2018). Keeping this in mind, a dataset with 53 landslide registries mapped by Åberg (2021) in the Sunnfjord district was included for the AOI used for external validation. In addition, 124 starting points mapped

on the detailed NGU quaternary map by Pullarello (2021) was used for the AOI used for training. To address overlapping landslide registries and mitigate road-related bias in the landslide inventory, registries that were both located within a 10-meter distance of each other and intersecting with roads were excluded. Consequently, the refined landslide inventory for the training and external validation datasets contained 699 and 233 records, respectively.

Generating points to represent non-landslide occurrences, or areas where landslides are absent, is an essential step for enabling ML models to distinguish between areas that are susceptible to landslides and those that are not. Modeled after the methodology presented in the study by Ageenko et al. (2022), a random sample of non-landslide points, matching the number of landslide points, was created. By maintaining an equal distribution of landslide and non-landslide data points, the model can interpret predicted probabilities without adjustment, thereby circumventing the problem of class imbalance and model bias. Such a balanced dataset ensures the model is not skewed towards overrepresenting either condition, which is important to identify areas at risk of landslides and preventing the misclassification of potentially safe zones as hazardous (Ageenko et al., 2022).

To generate the non-landslide points, a python script was developed. The script uses GeoPandas, a python library built upon the pandas library to allow handling of geospatial data and tabular data, by introducing geometric types (e.g. point data). First, the landslide inventory data is read to a geodataframe, where both the location of the points and additional information is stored. To avoid overlapping between landslide and non-landslide points, and avoid non-landslide points being generated in water bodies, a fishnet of rectangular cells, measuring 40 by 40 meters, was generated in ArcGIS Pro for the AOI. Water bodies were excluded by using an erase function to remove parts overlapping lakes, fjords and rivers. A 120-meter buffer is applied to each landslide point so ensure that no grid cell within this distance is selected. A Similar approach was done by Ageenko et al. (2022). As seen in Code 1, line 12, 13 and 15, the script identifies all the grid cells from the fishnet that do not intersect with the buffered landslide zones and randomly selects a number of grid cells equal to the amount of landslide records. A visual representation can be seen in figure 8.

Next, one point is generated randomly within each of the selected grid cells, ensuring that the non-landslide points are dispersed throughout the AOI, not confined to specific locations (Code 1, line 17-26). The geodataframe is then saved as a new shapefile.

To prepare the datasets for further analysis, a new field is made for the original landslide inventory shapefile and the newly created non-landslide shapefile. This column represents a binary system (presence/absence of landslide), which serves as the basis for the machine learning algorithm's predictions, called the target variable. The column, “Landslides”, marked as ‘0’ for non-landslide points and ‘1’ for landslide points, respectively (Code 1, line 30-34). Finally, the two shapefiles are combined into a single file.

```
1  # Generating point data for non-landslide locations
2  import geopandas as gpd
3  import pandas as pd
4  import random
5  from shapely.geometry import Point
6
7  landslide_points = gpd.read_file('<path_to_input_landslide_point_data>')
8  fishnet_grid = gpd.read_file('<path_to_fishnet_data>')
9
10 buffered_zones = landslide_points.buffer(120)
11
12 exclusion_zone = buffered_zones.unary_union
13 eligible_cells = fishnet_grid[~fishnet_grid.intersects(exclusion_zone)]
14
15 selected_cells = eligible_cells.sample(n=len(landslide_points))
16
17 def generate_point_in_cell(cell):
18     minx, miny, maxx, maxy = cell.bounds
19     while True:
20         point = Point(random.uniform(minx, maxx), random.uniform(miny, maxy))
21         if point.within(cell):
22             return point
23
24 non_landslide_points = selected_cells.geometry.apply(generate_point_in_cell)
25
26 non_landslide_gdf = gpd.GeoDataFrame(geometry=non_landslide_points)
27 non_landslide_gdf.to_file('<path_to_output_non_landslide_data>', driver='ESRI Shapefile')
28 print('Non-landslide points generated and saved as Shapefile.')
29
30 non_landslides = gpd.read_file('<path_to_output_non_landslide_data>')
31 non_landslides['Landslides'] = 0
32
33 landslides = gpd.read_file('<path_to_input_landslide_point_data>')
34 landslides['Landslide'] = 1
35
36 common_crs = 'epsg:25833'
37 non_landslides = non_landslides.to_crs(common_crs)
38 landslides = landslides.to_crs(common_crs)
39
40 combined_dataset = pd.concat([non_landslides, landslides], ignore_index=True)
41 combined_dataset.to_file('<path_to_combined_landslide_non_landslide_data>', driver='ESRI
   Shapefile')
```

Code 1: Code for generating non-landslide point data.

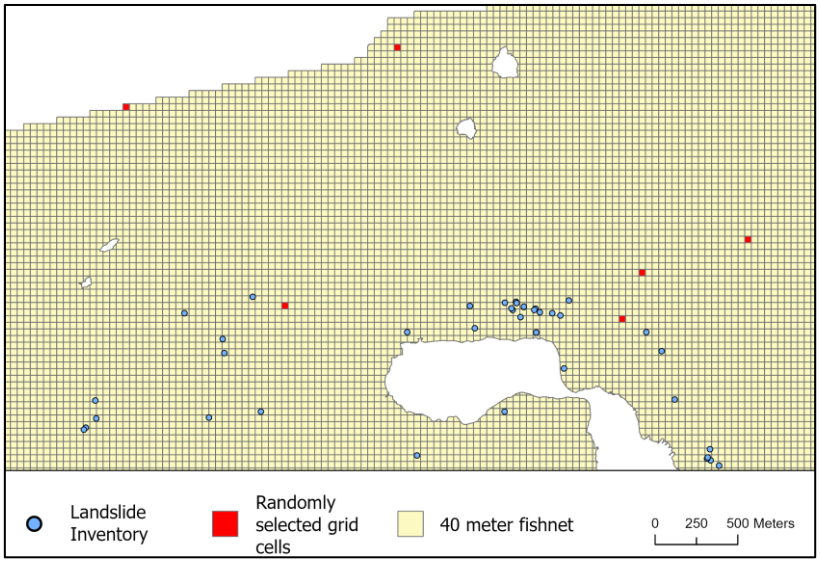


Figure 8: A snippet from the AOI showing landslide inventory points, the prepared fishnet and randomly selected grid cells.

4.2.2 Creating influencing factors

In this chapter, the focus is turned towards the generation of the influencing factor data, which serve as the predictor variables in the analysis. Following the establishment of the target variable through the script described previously, the derivation of the influencing factors is now described. The rationale behind selecting these specific influencing factors was detailed in Chapter 3.2.

In the derivation of influencing factors for the landslide prediction model, Python 3.9 with the ArcPy library is utilized for its integration with ArcGIS's spatial data analysis capabilities, facilitating the manipulation of geographic information for predictor variables[13]. The methodology is further enhanced with a Python 3.11 scripts that make use of libraries such as Rasterio, GDAL, Rioxarray, and RichDEM [14]. RichDEM is notable for its efficient processing of DEM derivatives, which are vital since they constitute a noteworthy portion of the influencing factors. This combination of ArcPy for geographic data management alongside ArcGIS Pro with the computational efficiency of specialized libraries streamlines the process. For all influencing factors, a cell size of 5 meters was chosen. This is partly because of

[13] More information about the ArcGIS Pro Python reference can be found here: https://pro.arcgis.com/en/pro-app/latest/arcpy/main/arcgis-pro-arcpy-reference.htm
[14] Source to the utilized python libraries can be found here: Rasterio: https://rasterio.readthedocs.io/en/stable/. GDAL: https://gdal.org/index.html. Rioxarray: https://corteva.github.io/rioxarray/stable/. RichDEM: https://richdem.readthedocs.io/en/latest/

computational restrictions, but mainly because 5 meters resolution, especially for DEM derivatives, is recommended because at this scale, terrain variabilities that contribute to landslide formation is best captured (L. Rubensdotter, personal communication, January 31, 2024). Similarly, Arnone et al. (2016) compared different resolutions for the influencing factors and concluded that the best predictive capabilities were found using medium resolution (between 5 and 20 meter), showing that an increase of resolution doesn't always imply an improvement in the susceptibility analysis (Arnone et al., 2016).

4.2.2.1 Aspect

Aspect raster was generated using the *Surface Parameters* tool in ArcGIS Pro. The tool provides a newer implementation of aspect compared to the older *Aspect* tool. The Surface Parameters tool is recommended as the algorithm used provides a more natural fit to the terrain. The output identifies the compass direction the slope faces, expressed in positive degrees from 0 to 360, measured clockwise from north (Esri, n.d-a). Aspect is therefore a circular parameter. This circular nature can pose a challenge for machine learning models because these models don't inherently understand the continuity between the end points of the aspect values. For example, an aspect of 359 degrees is very close to 0 degrees in terms of direction, but numerically, they appear far apart. This can mislead a model into thinking two nearly identical orientations are vastly different, potentially skewing the analysis or predictions. Therefore, aspect was decomposed into its sine and cosine component. In this way it is possible to represent direction of the slope as a two-dimensional system; one raster represents the sine of aspect where north-south slopes have positive and negative values between 1 and -1, respectively. The other raster represents the cosine of aspect where east-west slopes have positive and negative values between 1 and -1, respectively (Ageenko et al., 2022).

The script to compute the sine and cosine of the aspect raster is based on the work by Ageenko et al. (2022). However, due to the large size of raster files, Dask was used to divide the computation into smaller section, called chunks, each of which small enough to fit into memory [15]. The "dask.array.sin" and "dask.array.cos" functions were used to compute the sine and cosine of the original aspect raster, resulting in two raster files shown in figure 9.

[15] More information on the Dask library: https://docs.dask.org/en/stable/array.html

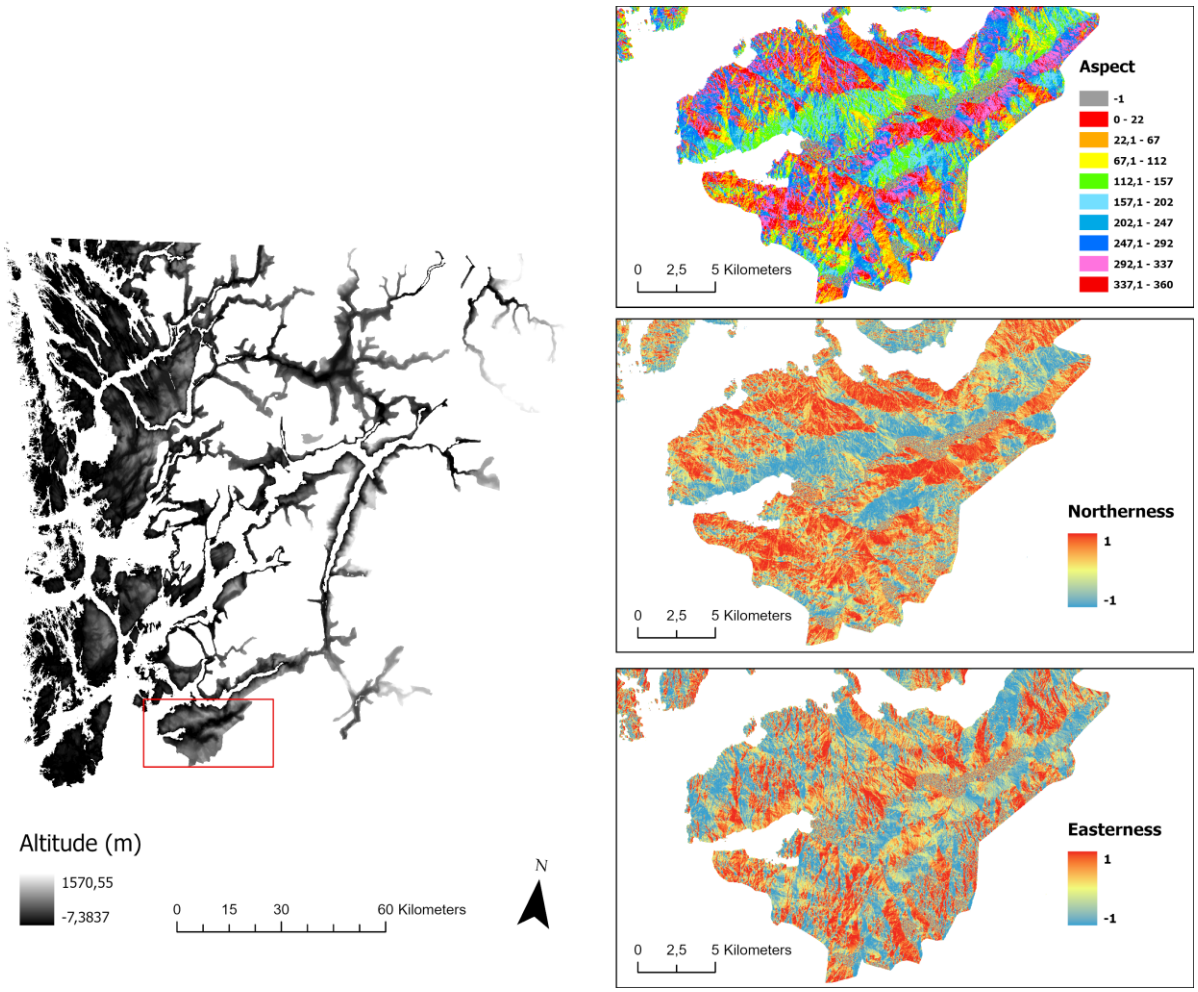


Figure 9: The two thematic maps representing the sine and the cosine of the aspect. Aspect raster is computed from a digital elevation model (each cell representing altitude) and the Northnerness and Easterness is computed from the aspect raster. Inset maps from one area of the AOI is shown for better visualization.

4.2.2.2 Profile Curvature

Profile curvature was generated using the *Surface Parameters* tool in ArcGIS Pro. The output measures the terrain's curvature along its slope line. Positive profile curvature indicates convex terrain that accelerates surface water flow and erosion. Conversely, negative profile curvature shows concave terrain that decelerates water flow. A zero value signifies flat terrain with no significant impact on water flow velocity (Esri, n.d-a). Figure 10 displays a small snippet of the AOI showing the profile curvature raster, arranged for easier visual interpretation.

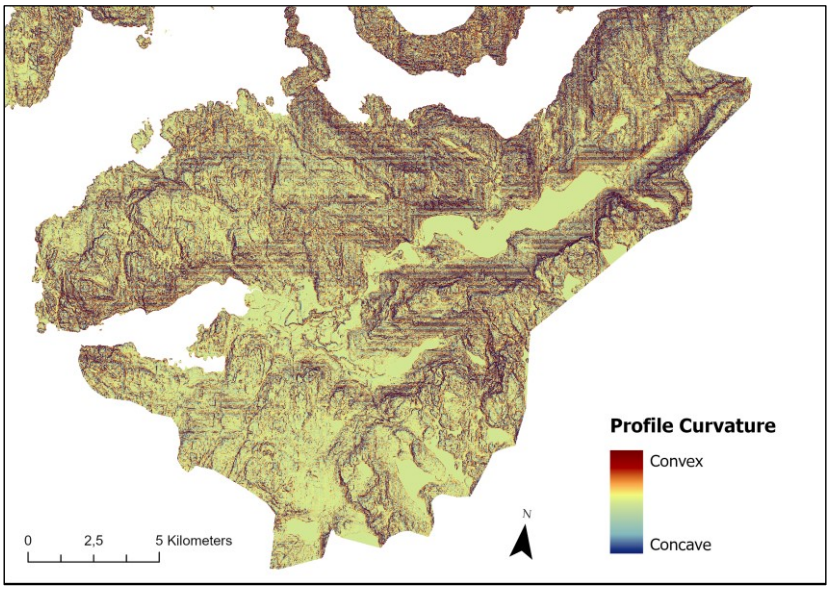


Figure 10: Profile Curvature raster

4.2.2.2 Stream Power Index

Introduced by Moore et al. in 1988, SPI uses the upstream contributing area and the slope to estimate the erosive power of flowing water. SPI can be computed as follows;

$$SPI = a * tan\beta$$

Where a is the local upstream accumulated flow area and β is the slope angel (Saleem et al., 2019). The local upstream area is a measure of the area contributing surface runoff to a specific point on a landscape. It's essentially an indicator of how much water would flow through a point, considering only the water that arrives due to gravity pulling it downhill from the immediate landscape. This metric can be generated using ArcGIS Pro with the *Flow Accumulation* tool. The tool creates a raster of accumulated flow into each cell by using a flow direction raster where the flow direction from each cell to its downslope neighbor is defined. The accumulated flow is thus based on the number of cells whose flow direction is interlinked in the downslope flowing of surface water. Cells with high flow accumulation are areas of concentrated flow and cells with zero accumulation represent local topographic high points (Esri, n.d-b). To calculate the slop angel β, the tool *Surface Parameters* tool in ArcGIS Pro was used. The resulting slope raster in degrees was then converted to slope radians using the python script, seen in Code 2:

```
import arcpy

slope_degrees = "<input_slope_raster_in_degrees>"

slope_rad = "<output_slope_raster_in_radians>"
slope_in_radians = Raster(slope_degrees)
rrconvert = (slope_in_radians * 3.141593 / 180)
rrconvert.save(slope_rad)
```

Code 2: Code for converting slope in degrees into slope in radians

For the calculation of SPI, representing slope in radians is preferred compared to slope in degrees (Ageenko et al., 2022). The script used to calculate the SPI utilizes the Gdal library for geospatial data manipulation and NumPy library for the numerical computation. It calculates SPI with the python function, seen in code 3:

```
def calculate_spi(flow_acc, slope):
    return np.log((flow_acc*5*5 + 0.001) * np.tan(slope + 0.001))
```

Code 3: Code for calculating SPI

The function uses the flow accumulation raster and slope in radians to calculate SPI, based on the method employed by Ageenko et al. (2022). A small constant is introduced to avoid zero values. Output SPI values are stored in a new GeoTIFF file, maintaining the original spatial properties, as shown in figure 11.

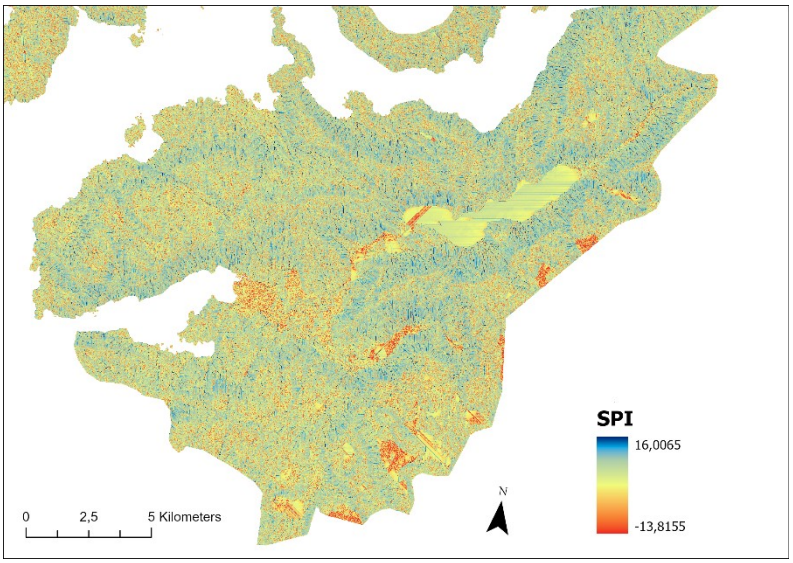


Figure 11: Stream Power Index Raster. Positive values indicate areas with higher potential for water flow erosion. Negative values represent areas with lower potential for water flow erosion.

4.2.2.3 Topographic Position Index

Introduced by Weiss. (2001), The Topographic Position Index (TPI) is defined as the difference between a central cell and the mean of its surrounding cells. TPI is a quantitative measure generated using GDALs DEM Processing module, which is part of the GDAL library specialized for use on digital elevation models [16]. As seen in Code 4, GdalDEM has a built-in command that uses a DEM as input to calculate topographic position of each cell by comparing the elevation of the cell to the mean elevation of its surrounding cells within a specified neighborhood. This measure is calculated for each cell in the input DEM, resulting in a TPI raster where each cell's value reflects its relative position (see figure 12): positive values indicate that the cell is higher than its surroundings, suggesting prominent features like ridges or hilltops, while negative values indicate cells lower than their surroundings, typical of valleys or depressions. A TPI value close to zero suggests flat or uniformly sloping terrain (De Reu et al., 2013).

```
1  import gdal
2
3  dem = gdal.Open("<input_digital_elevation_model_raster>")
4  tpi = gdal.DEMProcessing("<output_tpi_raster", dem, 'TPI', computeEdges=True)
5
```

Code 4: gdal command to generate topographic position index based on a DEM

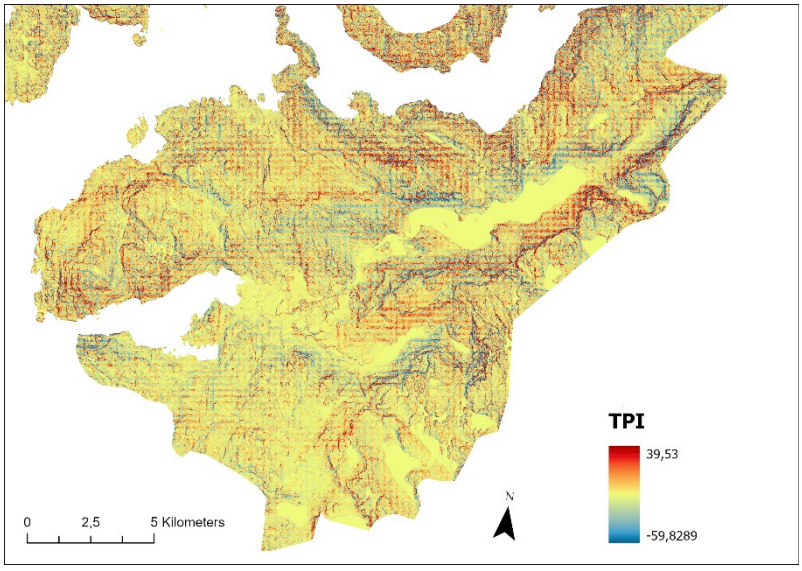


Figure 12: Topographic Position Index raster. Positive values indicate that the central point is located higher than its average surroundings, while negative values indicate a position lower than the average.

[16] Manual for the gdal library can be found here: https://gdal.org/programs/gdaldem.html

4.2.2.4 Terrain Ruggedness Index

Terrain Ruggedness Index (TRI) quantifies the ruggedness of the terrain by measuring the difference between a central pixel and its surrounding cells. TRI can be seen as a measure to assist in describing the terrain as smooth or rugged by depicting the local variance of curvature and gradients (Ageenko et al., 2022). TRI is calculated using the built-in function in GdalDEM, as shown in code block 5. The calculation uses the algorithm developed by Riley et al. (1999) that calculates the sum change in elevation between a grid cell and its eight neighbor grid cells.

As shown in figure 13, the resulting raster data reflect the elevation differences between cells; high values indicate a significant elevation difference between the central pixel and its surrounding pixels. Values closer to zero indicate areas of minimal elevation difference between a central pixel and its surrounding pixels, meaning a level terrain (Riley et al., 1999). The index describes topographic heterogeneity at small scale. If the elevation difference between the central pixel and each of its neighbours is minimal or consistent, the terrain within that specific 3x3 grid will be considered level or smooth by the TRI, resulting in a low ruggedness value. This can happen on a steep slope if the slope is uniform and the elevation changes between adjacent pixels are consistent across the 3x3 grid being analyzed. Furthermore, TRI is calculated based on the absolute values of the differences in elevation and will not distinguish between a small depression and a small hill surrounded by flat terrain; it does not differentiate between positive and negative elevation differences (Riley et al., 1999).

```
1  import gdal
2
3  dem = gdal.Open("<input_digital_elevation_model_raster>")
4  tri = gdal.DEMProcessing("<output_tri_raster>", dem, 'TRI', computeEdges=True)
```

Code: 5: gdal command to generate terrain ruggedness index based on a DEM

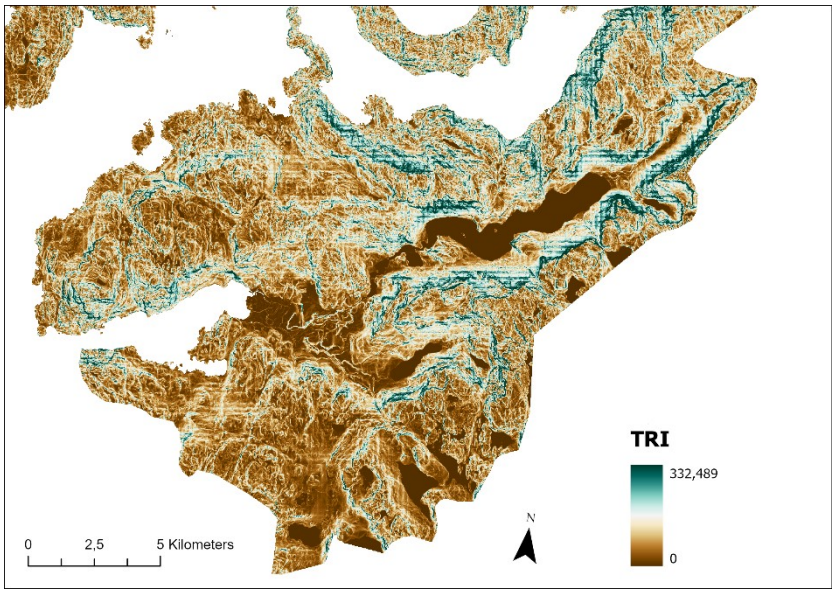


Figure 13: Terrain Ruggedness Index raster. High values indicate a significant elevation difference between the central pixel and its surrounding pixels. Values closer to zero indicate areas of minimal elevation difference.

4.2.2.5 Topographic Wetness Index

Introduced by Beven & Kirkby. (1979), the Topographic Wetness Index uses the local upstream contributing area and the local slope to estimate amount of soil moisture in the terrain. SPI can be computed as follows;

$$TWI = ln\left(\frac{a}{\tan(\beta)}\right)$$

Where *a* is the local upstream accumulated flow area and β is the local slope (Saleem et al., 2019). As shown in Code block 6, an ArcPy script was used to compute the Topographic Wetness Index (TWI) based on the beforementioned formula where *a* is represented with a flow accumulation raster file and β represents a slope raster in radians. The creation of both files is discussed in further detail in section 4.2.2.2. The slope raster is slightly adjusted by adding a small constant to avoid division by zero errors. These raster values are then fed into the *Raster Calculator tool* of ArcGIS Pro through an expression that applies the logarithm and tangent functions as described in the TWI formula. The output is a new raster file providing a spatial distribution of wetness potential across the terrain. High values indicate areas with greater potential for water to accumulate, typically in depressions or flat areas conducive to moisture retention. Conversely, lower values generally represent ridges or steep slopes where water is less likely to accumulate and more likely to run off (see figure 14).

```python
import arcpy

flow_acc = "<input_flow_accumulation_raster>"
slope_radians = "<input_slope_raster_in_radians>"

expression = f"Ln(('{flow_acc}'+ 1) / Tan('{slope_radians}' + 0.0001))"
output_twi_raster = "<output_twi_raster>"
# Perform the calculation
arcpy.gp.RasterCalculator_sa(expression, output_twi_raster)
```

Code 6: ArcPy script to calculate TWI based on the equation given by (Saleem et al., 2019)

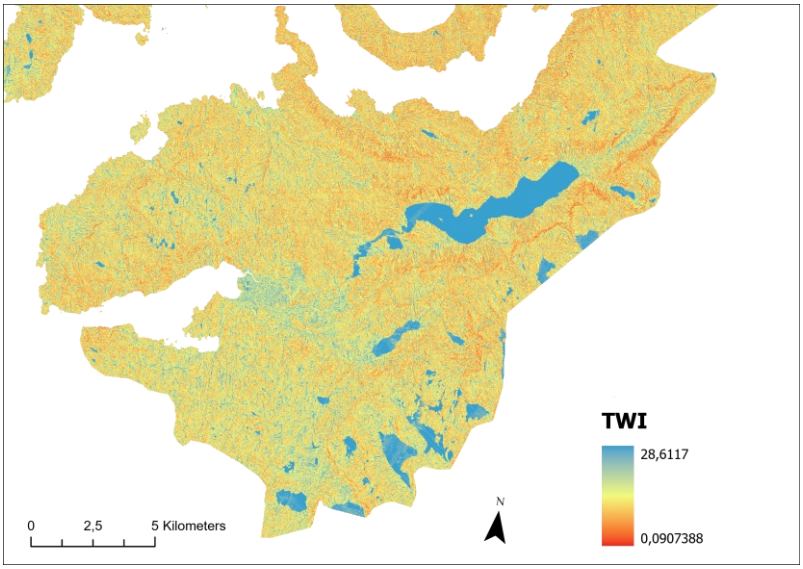


Figure 14: Topographic Wetness Index raster. Positive values represent areas of high-water accumulation and values closer to zero represents areas with low accumulation.

4.2.2.6 Slope

Slope raster was generated using the *Surface Parameters* tool in ArcGIS Pro. The output measures the terrain's curvature along its slope line. The *Surface Parameters* tool is recommended to use instead of the older *slope* geoprocessing tools in ArcGIS Pro as it provides a newer implementation of slope (Esri, n.d-a); While the older Slope tool calculates slope by fitting a flat plane to a small, fixed grid of land cells, which can sometimes oversimplify or misrepresent natural land variations, the *Surface Parameters* tool fits a curved surface to a potentially larger area. This method captures the true shape and variations of the landscape in a more realistic way, especially on terrain where features like hills, valleys, or man-made structures like roads and streams are present (Esri, n.d-a).

The algorithm identifies the gradient from each cell of the DEM raster, as seen in figure 15. When slope is expressed in degrees, a value of 0 degrees indicates completely flat terrain, while a value of 90 degrees signifies a vertical cliff.

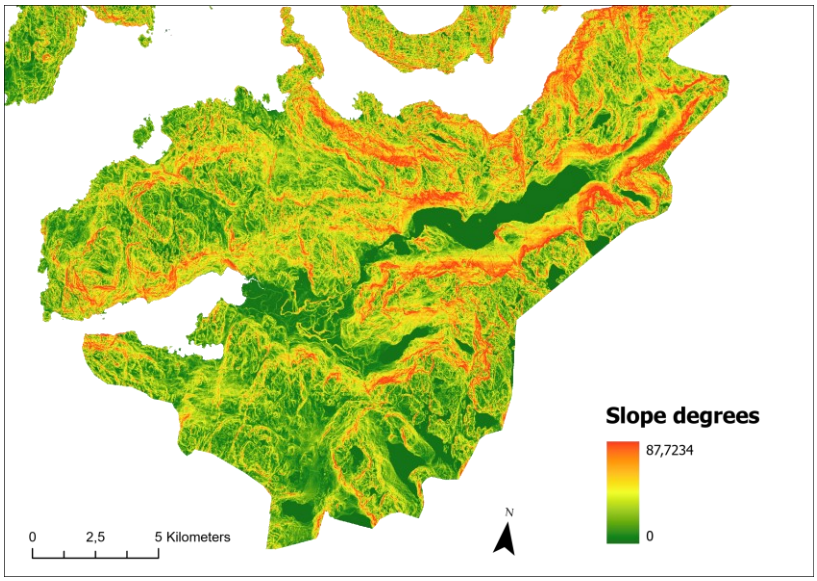


Figure 15: Slope raster measured in degrees. The output represents the rate of change of elevation for each DEM cell.

4.2.2.7 Landform

The aforementioned topographic position index (TPI) is used to measure topographic slope positions but has also been showed to work to classify landform types (De Reu et al., 2013). TPI was introduced by Weiss. (2001) and compares the elevation of each cell in a DEM to the mean elevation of a specified neighborhood around the cell (Weiss, 2001). By combining TPI at small and large scale allows a variety of nested landforms to be distinguished. As recommended by Weiss. (2001), a scale factor of 300 and 2000 was used in this study to represent small scale and large-scale landforms. When using a cell size of 5 meters the following formulas are used to compute TPI at 300- and 2000-meters scales:

Tpi300 = int((dem – focalmean(dem, annulus, 30, 60)) + 0.5)

Tpi2000 = int((dem – focalmean(dem, annulus, 370, 400)) + 0.5)

To accomplish the calculation, the *Focal Statistics* tool and the *Minus* tool in ArcGIS Pro was utilized. *Focal Statistics* calculates the values within a specific neighborhood around each input cell (Esri, n.d-c). As recommended by Weiss. (2001), an annulus neighborhood type was selected, essentially representing a donut-shaped area around the cell, defined by an inner and

an outer radius. Within this area, the mean elevation was calculated for each cell in the DEM. According to the formula, the new focal statistics raster data where then subtracted from the original DEM using the *Minus* tool. In this way the resulting raster data shows the difference between the cell's elevation and the average elevation of its surrounding area. This difference highlights whether the cell is situated higher, lower, or at the same level as its surroundings.

However, to make TPI data comparable at different scales Weiss. (2001) recommends standardizing the TPI values. The *Raster Calculator* tool in ArcGIS Pro was used to standardize the two TPI raster data by centering the data around zero by subtracting the mean and dividing by the standard deviation to normalize the variance in the data (Weiss, 2001).

A Python script utilizing the ArcPy library was employed to assign specific classes to each landform. The script processes raster layers representing TPI at small and large scale, standardized to mean 0 and standard deviation 1, and slope in degrees. It defines conditions within a list, where each condition comprises a logical statement that evaluates whether the TPI values and slope for each cell fall within a threshold. The specific conditions follow the work by Weis. (2001). These conditions are designed to capture distinct topographic features, assigning values ranging from 1 to 10, each representing a unique landform type. For instance, the condition checking if the TPI values for both scales are greater than – or equal to one scaled standard deviation above the average, categorize high ridges, expressed as category 10. As seen in figure 16, the results from these evaluations are compiled into a single raster output, where each cell's value corresponds to its classified landform type, generated through conditional checks implemented using the *Con* function from ArcPy (see code 7)[17].

[17] A more detailed description of the Con function can be found from the Esri website: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/con-.htm

```python
import arcpy

tp300_stdi = "<input_tpi_300_meter_scale_standardized>"
tp2000_stdi = "<input_tpi_2000_meter_scale_standardized>"
slope = "<input_slope_raster_in_degrees"

conditions = [
    ((Raster(tp300_stdi) > -100) & (Raster(tp300_stdi) < 100) &
     (Raster(tp2000_stdi) > -100) & (Raster(tp2000_stdi) < 100)
     & (Raster(slope) <= 5), 5),

    ((Raster(tp300_stdi) > -100) & (Raster(tp300_stdi) < 100) &
      (Raster(tp2000_stdi) > -100) & (Raster(tp2000_stdi) < 100)
      & (Raster(slope) >= 6), 6),

    ((Raster(tp300_stdi) > -100) & (Raster(tp300_stdi) < 100)
     & (Raster(tp2000_stdi) >= 100), 7),

    ((Raster(tp300_stdi) > -100) & (Raster(tp300_stdi) < 100)
     & (Raster(tp2000_stdi) <= -100), 4),

    ((Raster(tp300_stdi) <= -100) & (Raster(tp2000_stdi) > -100)
     & (Raster(tp2000_stdi) < 100), 2),

    ((Raster(tp300_stdi) >= 100) & (Raster(tp2000_stdi) > -100)
     & (Raster(tp2000_stdi) < 100), 9),

    ((Raster(tp300_stdi) <= -100) & (Raster(tp2000_stdi) >= 100), 3),

    ((Raster(tp300_stdi) <= -100) & (Raster(tp2000_stdi) <= -100), 1),

    ((Raster(tp300_stdi) >= 100) & (Raster(tp2000_stdi) >= 100), 10),

    ((Raster(tp300_stdi) >= 100) & (Raster(tp2000_stdi) <= -100), 8)
    ]
# Initialize the result with a default value
result = Raster(tp300_stdi) * 0

for condition, value in conditions:
    result = Con(condition, value, result)
result.save("<output_landform_raster")
```

Code 7: ArcPy script to delineate landform types based on combinations of TPI at small and large scale and slope. Classes represents: 1: Canyons, 2: Valleys, 3: Headwaters, 4: U-valleys, 5: Plains, 6: Open slopes, 7: Upper slopes, 8: Ridges, 9: Hills, 10: High ridges

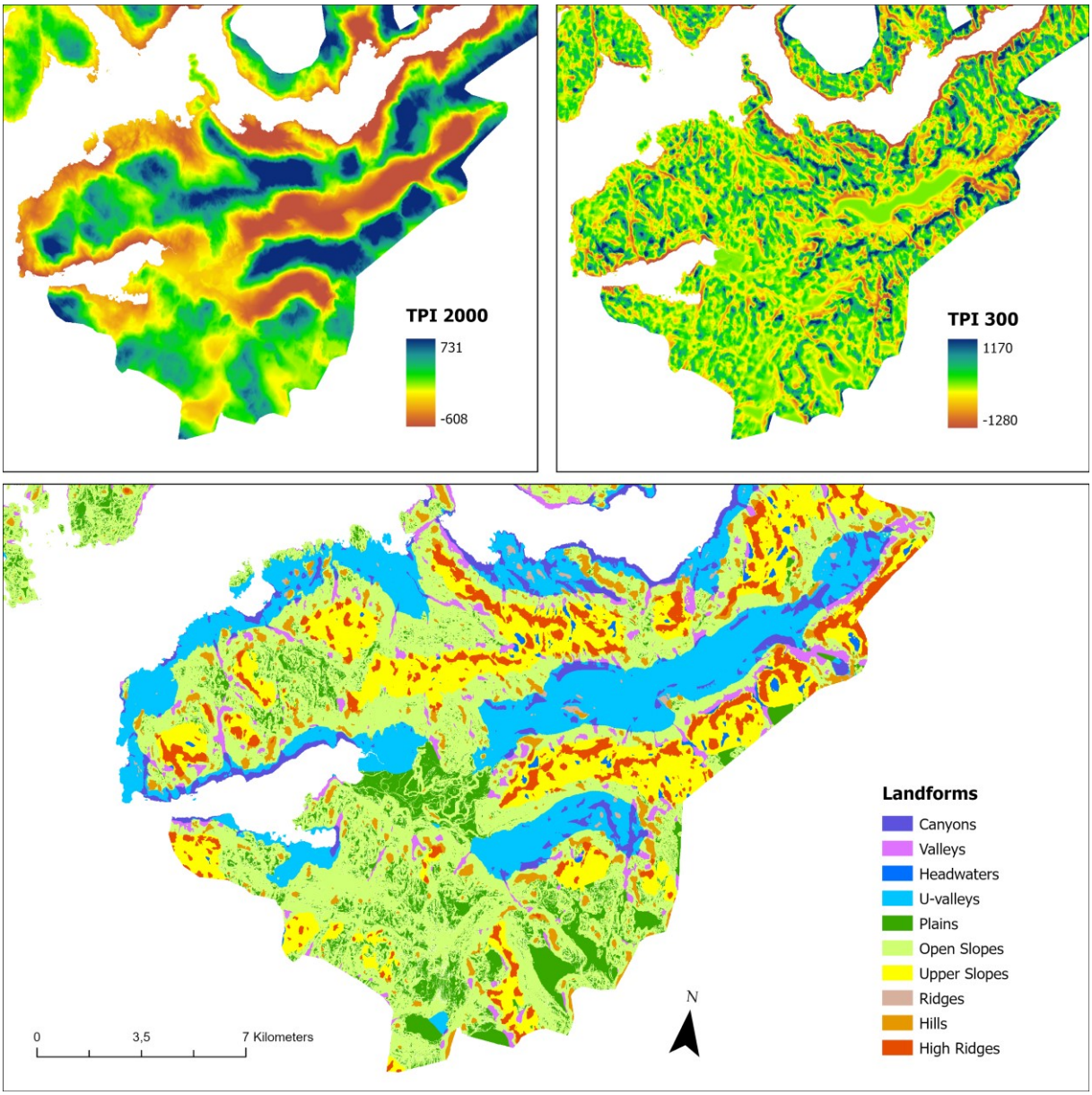


Figure 16: TPI raster data at 300- and 2000-meter scale. High values represent areas higher than the average terrain surrounding them, calculated using two different annulus sizes. TPI is combined with slope raster to delineate landform types as discrete raster.

4.2.2.8 Distance to roads and rivers

To calculate the distance to rivers and roads, the *Distance Accumulation* tool in ArcGIS Pro was used. This tool computes the accumulated distance for each cell to defined sources (Esri, n.d-d). To determine the distance for each cell in the AOI to the nearest road and river, the input source data must represent the road and river networks in the area. For roads, vector polygon data from the FKB-Roads dataset were downloaded and imported into ArcGIS Pro. For the river system, the nationwide Elvis-River network dataset developed by the Norwegian Water Resources and Energy Directorate (NVE) was downloaded and imported into ArcGIS

Pro. Both datasets were clipped to the AOI and then used as input source data. To account for the actual surface distance across the terrain, an input surface raster (DEM) was used to provide an elevation surface for the distance calculation. The two output raster datasets record the measured distance in meters from every non-source cell to the closest vector representation of roads and rivers, as seen in figure 17.

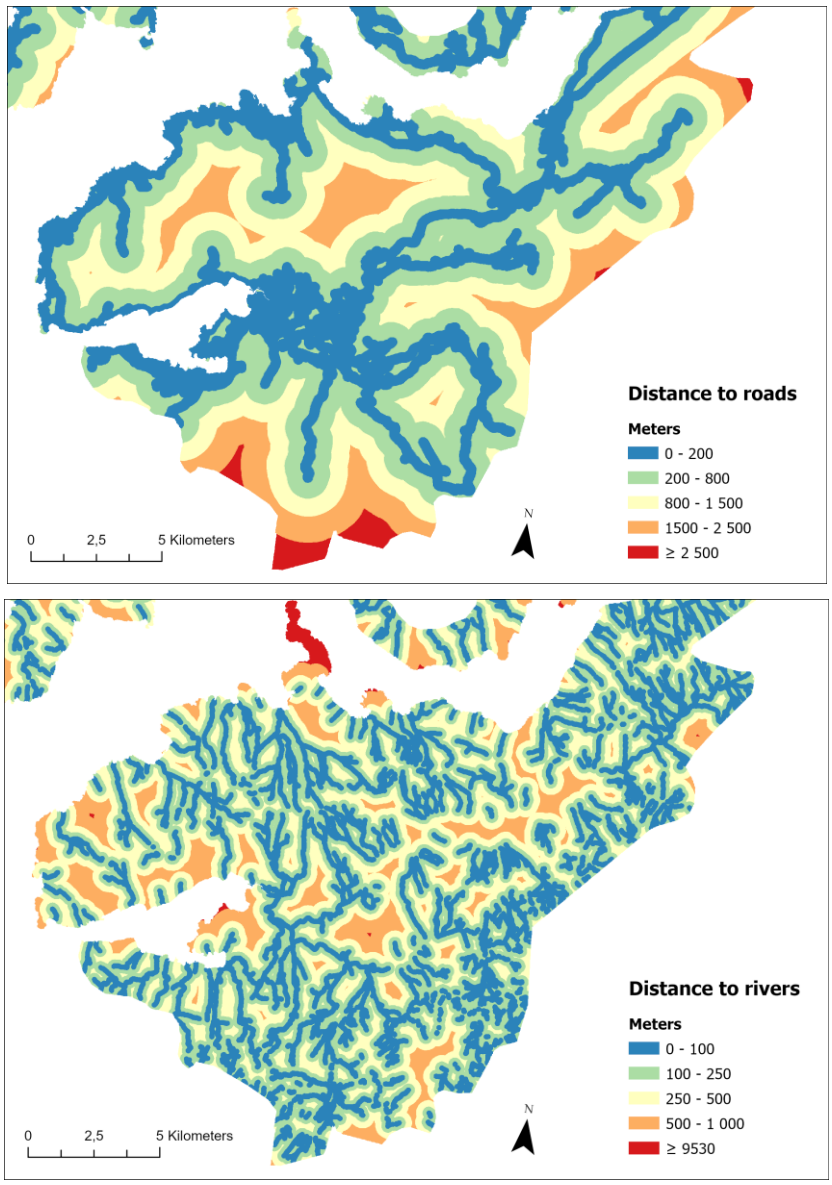


Figure 17: Distance accumulation raster for road and river network. Each non-source cell represents the distance in meters to the closest road and river polyline.

4.2.2.9 Surficial deposits, Land use and Soil Condition

Raster representation of the categorical data for quaternary surficial deposits, land use types and soil condition were created in ArcGIS Pro using the *Polygon to Raster* conversion tool. Data on quaternary surficial deposits was downloaded for Vestland county in the shapefile format [18]. The dataset primarily shows the distribution of unconsolidated sediment types covering the rock surface, mostly formed during and after the last ice age. The data represents the dominant soil type in the top meters of the surface. The data on the surficial deposit types are derived from the content of quaternary geological maps (NGU, n.d).

For the representation of land use and soil condition, the FKB-AR5 dataset was downloaded as Esri file geodatabase format [19]. FKB-AR5 is a nationwide detailed land use dataset where the land is categorized by land use type, forest quality, tree species, and soil conditions, represented as categorical values in four separate columns.

The FKB-AR5 dataset is managed within a single vector polygon file, represented as a multipart feature with discontinuous parts. This makes it possible to represent the four abovementioned classes even though they don't necessarily overlap. Firstly, the value field that represents soil conditions in the feature class was used to generate a raster representation using the *Polygon to Raster* tool. Secondly, the tool was employed again to create a raster for land use, this time using the value field that denotes the land use class. The same methodology was applied to create a raster representation of the surficial deposits dataset, utilizing the value field for soil type. The raster files were then clipped to the AOI. Figure 18 provides a snippet of this area, displaying the surficial deposits, land use, and soil conditions.

[18] More detailed description of the dataset can be found from the Geonorge website: https://kartkatalog.geonorge.no/metadata/loesmasser/3de4ddf6-d6b8-4398-8222-f5c47791a757
[19] More detailed description of the dataset can be found from the Geonorge website: https://kartkatalog.geonorge.no/metadata/arealressurskart-fkb-ar5/243751e8-5803-4627-898c-d0ddabe82056

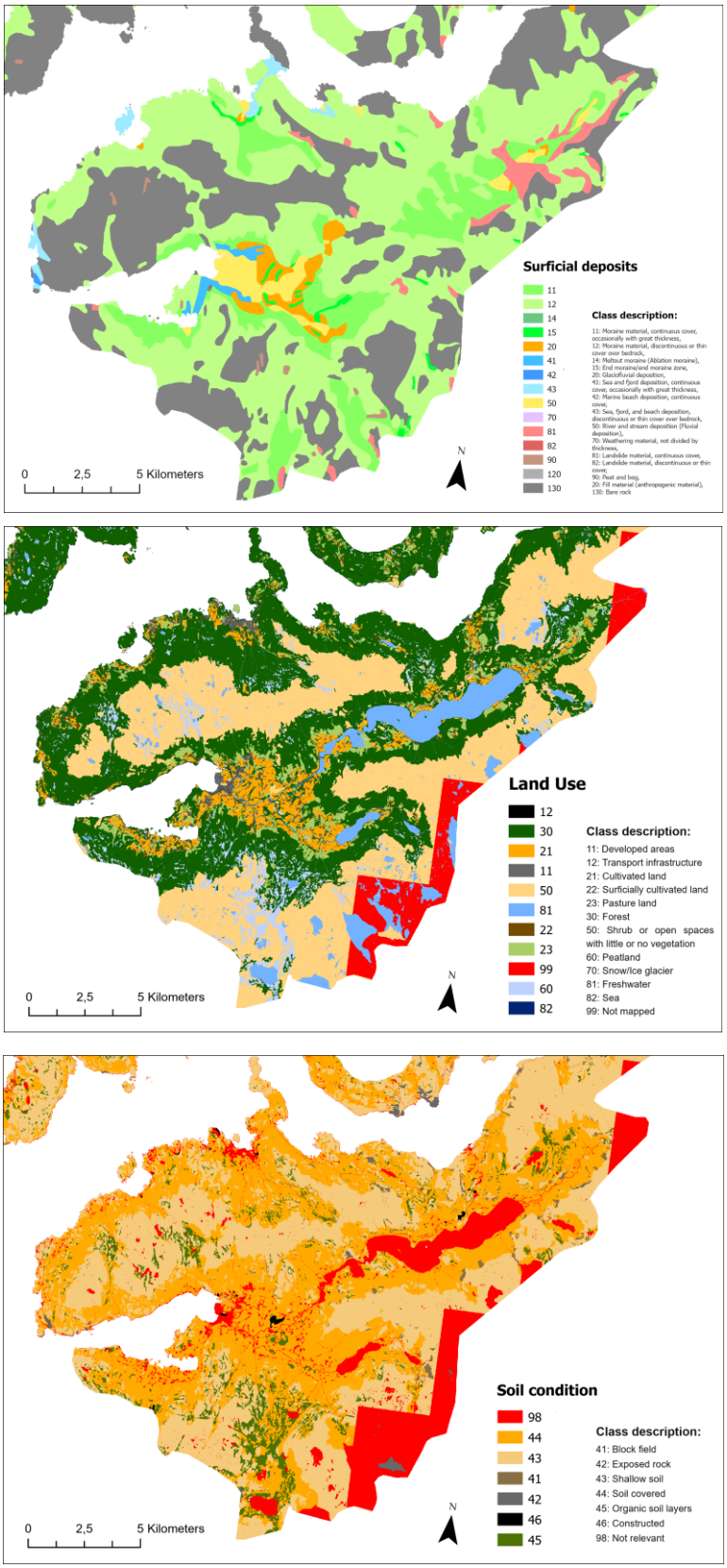


Figure 18: Surficial deposits, land use and soil condition raster data.

4.2.2.10 NDVI and Forest loss

The Normalized Difference Vegetation Index (NDVI) is a satellite-derived vegetation index calculated using the difference and sum of the near-infrared (NIR) and red (RED) bands of the electromagnetic spectrum (Hamel et al., 2009). Google Earth Engine (GEE) was utilized in this study to calculate the NDVI. GEE is a cloud-based platform with extensive data repositories, accessible through an internet-accessible application programming interface. This setup facilitates rapid prototyping and result visualization of satellite data (Qasimi et al., 2022).

To compute NDVI, Sentinel 2 satellite data was imported to the GEE script and filtered to only include imagery between January and December 2022 with less than 20% cloud cover. The command *median()* was used to reduce the image collection to one image by calculating the median of all values at each pixel across the stack of images. NDVI was then calculated using the *normalizedDifference()* command. The command computes the normalized difference between two bands, in this case band 4 (Red) and band 8 (NIR) with the following calculation:

$$(NIR - red)/\,(NIR + red)$$

The resulting raster file roughly reflects the photosynthetic activity occurring at each pixel. This is because vegetation reflects light in the near-infrared and absorbs light in the red part of the electromagnetic spectrum (Google, n.d). Finaly the resulting imagery was then exported as a TIFF file, as shown in figure 19.

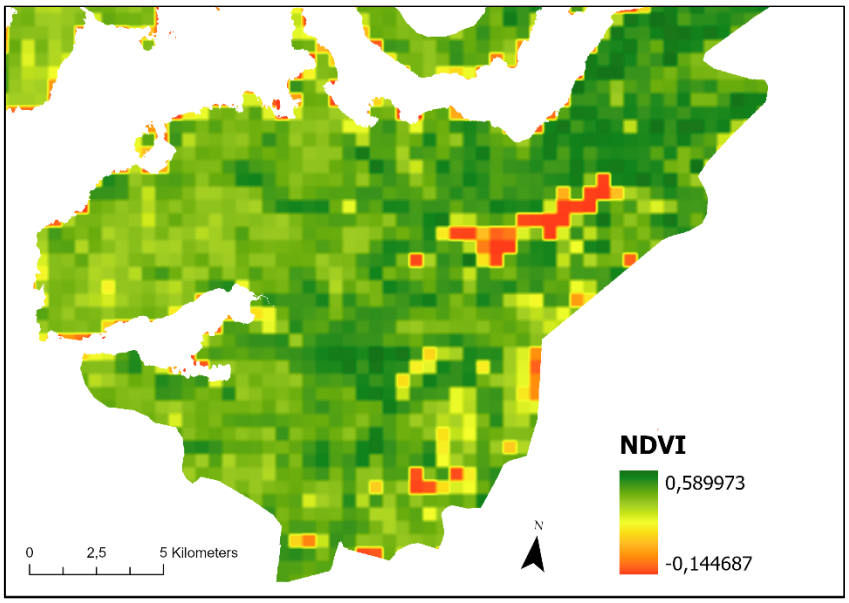


Figure 19: Map snippet showing the NDVI. Cell values range from negative (water or non-vegetated areas), values between 0 and 0.3 indicate barren areas and values between 0.3 and 0.6 typically indicate sparse vegetation cover

To quantify forest loss cover in Norway, the “Hansen global forest change” dataset was imported to GEE and filtered to display the band reflecting forest loss between 2000 and 2023, as shown in figure 20. The particular band quantifies forest loss during the period for each pixel, defined as a change from forest to a non-forest state, based on Landsat imagery (Hansen et al., 2013). Specifically, the composite imagery are median observations during the growing season in four spectral bands (Red, Near infrared, Shortwave infrared 1 and Shortwave infrared 2). The Global Land Analysis and Discovery (GLAD) laboratory at the University of Maryland, in partnership with Global Forest Watch (GFW), provides annually updated version of the dataset [20].

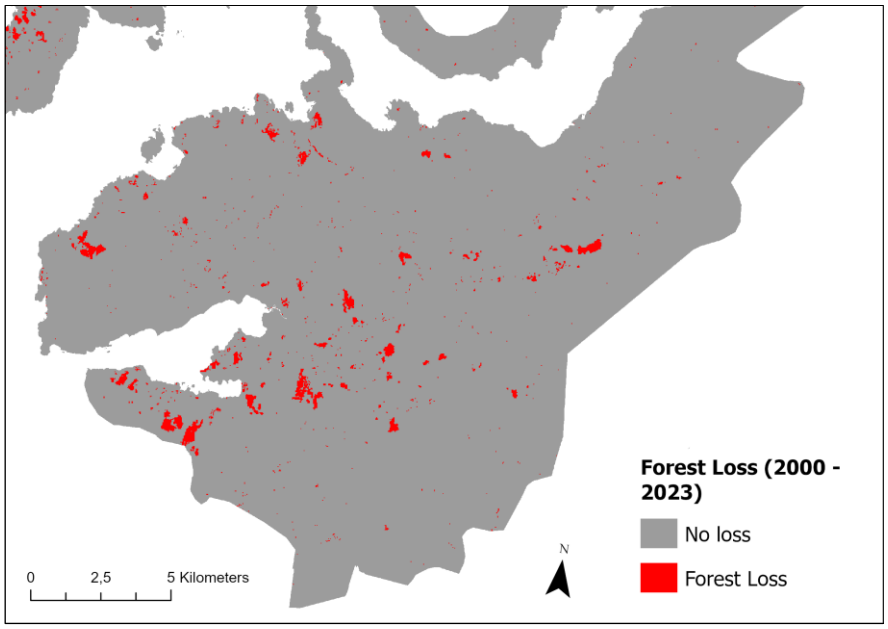


Figure 20: Map snippet showing deforested areas between 2000 and 2023

4.2.2.11 Average annual temperature and precipitation

To map the spatial patterns of precipitation and temperature for the AOI, the newest climatological normals (CN) for the 30-year period 1991 to 2020 were used. A climatological normal refers to the average of weather variables over a period of 30 years (Tveito, 2021). The CN for annual temperature and annual precipitation were downloaded from The Norwegian Meteorological Institutes THREDDS data server as NetCDF format – a file format

[20] Global forest watch website: https://data.globalforestwatch.org/documents/941f17325a494ed78c4817f9bb20f33a/explore

for storing multidimensional scientific data[21]. The multidimensional data was converted to TIFF format, resampled to a resolution of 5 meters and clipped to the AOI, as seen in figure 21.

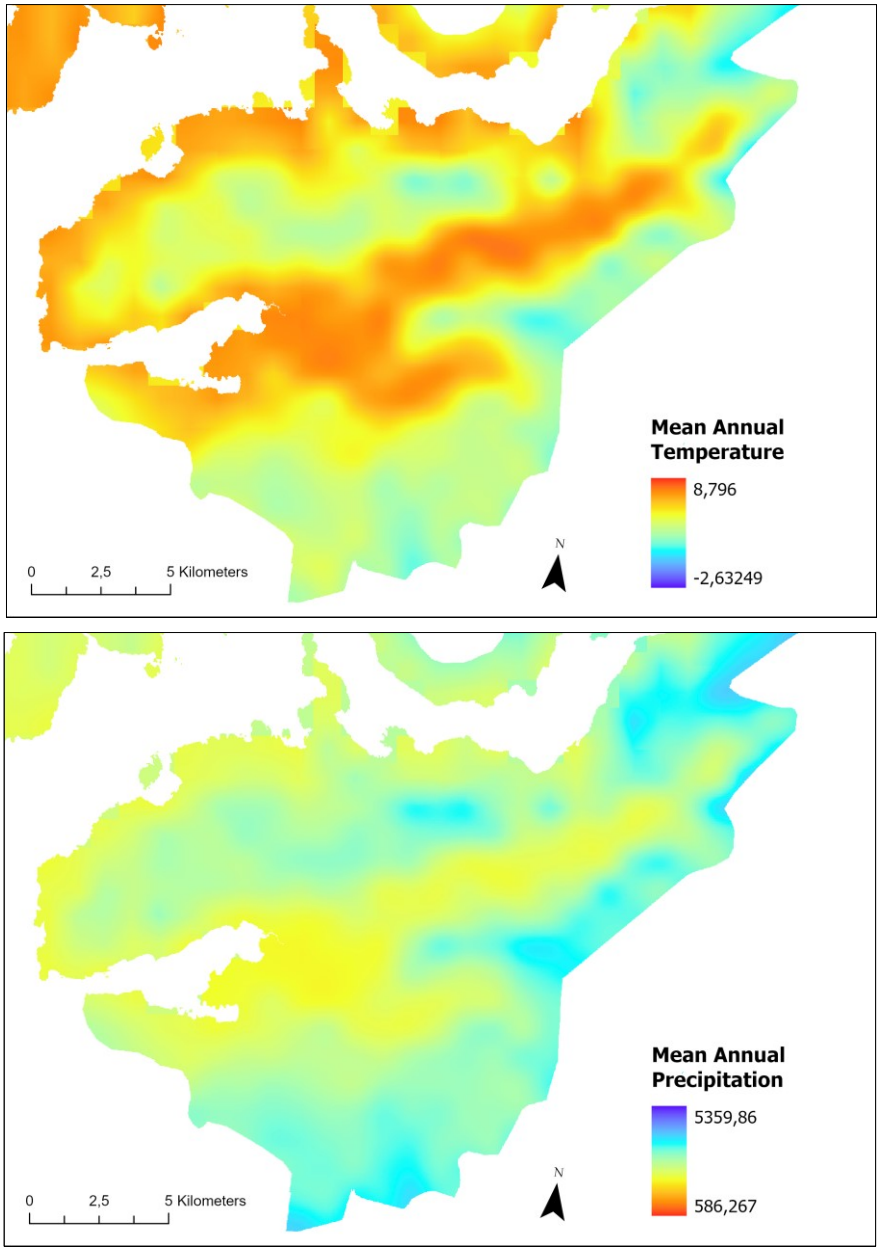


Figure 21: Mean annual temperature and precipitation raster data

[21] CN for temperature: https://thredds.met.no/thredds/catalog/KSS/Gridded_climate_normals_1991-2020/temperature/catalog.html
CN for precipitation: https://thredds.met.no/thredds/catalog/KSS/Gridded_climate_normals_1991-2020/precipitation/catalog.html

4.3. Data Sampling

Following the generation of the influencing factors and preparing both the landslide inventory and non-landslide point data, these elements were combined into tabular data that can be used for the machine learning classification. For classification tasks using the Scikit-learn machine learning library in Python [22], the classifier has to be fitted with two arrays: one array holding the influencing factors, and an array holding the target values, in this case landslide presence/absence (Breiman, 2001). This process essentially bridges the preparation phase using geospatial data with the machine learning phase that necessitates tabular data.

In section 4.2.1, a combined feature class is described that contains points labeled as landslides and non-landslides, with landslides represented by 1 and non-landslides by 0. To include the influencing factors associated with each landslide / non-landslide point in the dataset, the cell values for the influencing factors need to be extracted and mapped to the corresponding landslide / non-landslide point. This process was carried out using the *Extract Multi Values To Points* geoprocessing tool in ArcGIS Pro, which is designed to extract cell values from specified raster data based on the locations of points in a feature class (Esri, n.d-e). As seen in code block 8, the tool uses the shapefile representing landslides / non-landslides as the input point feature class and the collection of all influencing factor raster data to extract cell values. A new output field name is given to each new column in the landslide / non-landslide feature class (see table 6). Following data extraction, the feature class underwent a final transformation into a machine learning-compatible format. By utilizing the GeoPandas library in Python, the point feature class was converted to a CSV file.

Table 6: A subset of the resulting attribute table to the landslide / non-landslide shapefile after the cell value for the corresponding influencing factor is extracted.

Landslide	eastness	profile	northernes	spi	tpi	tri	twi	roughness	slope	elevation	d_river	landforms	d_roads	losmasse	forestloss	ndvi	a_temp	a_perc	landuse	Soil_con
0	0,717467	0,087356	-0,696593	-9,75531	0,064631	0,241446	2,86326	0,173743	3,26132	9,52517	65,3753	5	497,529	130	0	0,000237	7,88751	1932,23	3	4
0	0,088614	0,032632	-0,996066	1,62223	0,029209	0,517594	4,77922	0,43837	11,8578	30,3799	56,8953	6	91,7769	90	0	0,271143	7,63179	2017,62	5	4
0	0,943201	-0,008432	-0,332221	2,84369	-0,07357	0,872203	5,08092	0,965244	18,2242	18,4212	918,297	6	436,111	130	0	0,094776	7,57406	2043,92	3	3
0	0,861211	0,134619	0,508248	3,88042	-0,012337	0,501435	7,21879	0,477104	10,6584	33,507	46,9129	6	395,233	130	0	0,324957	7,82118	1931,86	5	3
0	0,76994	0,000484	-0,638116	0,59063	-0,013577	0,086718	7,67831	0,078838	1,63769	39,1515	464,079	5	600,347	130	0	0,28482	7,57121	2026,98	2	3
1	0,921964	0,012124	0,387276	-11,5053	0,002853	0,025865	4,69109	0,024441	0,520042	44,4742	512,117	5	0	130	0	0,309592	7,49208	2058,17	4	7
0	-0,724829	-0,046523	0,688929	-0,187864	-0,037079	0,534695	3,18841	0,638496	11,6449	33,9022	319,934	6	298,765	130	0	0,28353	7,83003	1922,1	5	3
1	-0,986135	0,012216	-0,165943	-3,36194	0,007128	0,087333	4,08245	0,085531	1,92626	18,7269	348,878	6	0	130	0	0,303982	7,79063	1930,74	4	7
1	-0,84511	0,015661	-0,534592	-2,42172	0,000109	0,108209	4,23452	0,114244	2,48281	18,5004	316,881	6	0	130	0	0,28353	7,80081	1929,1	4	7
1	-0,840918	0,019229	-0,541162	-11,6843	0,013458	0,049144	4,88956	0,036873	0,425398	19,108	330,965	2	0	130	0	0,303982	7,79063	1930,74	4	7
1	0,944995	0,006471	0,327085	-10,7009	0,007921	0,058695	3,83392	0,047176	1,23308	18,6636	335,758	2	0	130	0	0,303982	7,80079	1929,71	4	7

[22] Scikit-learn is an open-source library in python for machine learning, including random forest classification. https://scikit-learn.org/stable/modules/ensemble.html

```
import arcpy

arcpy.env.workspace = "<folder_path_to_raster_data>"
landslide_training = ("<input_combined_landslide/non-landslide_feature_class>")

arcpy.sa.ExtractMultiValuesToPoints(landslide_training, [
    ["<input_eastness_raster>", "eastness"],
    ["<input_profilecurvature_raster>", "profile"],
    ["<input_northerness_raster>", "northerness"],
    ["<input_spi_raster>", "spi"],
    ["<input_tpi_raster>", "tpi"],
    ["<input_tri_raster>", "tri"],
    ["<input_twi_raster>", "twi"],
    ["<input_roughness_raster>", "roughness"],
    ["<input_slope_degrees_raster>", "slope"],
    ["<input_dtm_raster>", "elevation"],
    ["<input_distance_rivers_raster>", "d_river"],
    ["<input_landforms_raster>", "landforms"],
    ["<input_distance_roads_raster>", "d_roads"],
    ["<input_temperature_raster>", "a_temp"],
    ["<input_percipitation_raster>", "a_perc"],
    ["<input_surficial_deposits_raster>", "losmasse"],
    ["<input_forest_loss_raster>", "forestloss"],
    ["<input_land_use_raster>", "landuse"],
    ["<input_soil_con_raster>", "soil_con"],
    ["<input_ndvi_raster>", "ndvi"]
], "NONE")
```

Code 8: Code for extracting the corresponding cell value for the influencing factors using the Extract Multi Values to Points tool.

4.4. Feature Selection

Following the generation of influencing factors, preparing the landslide inventory and combining the two into a machine learning friendly format, one final preprocessing step needs to be completed; refining the dataset by analyzing which influencing factors adversely affect the machine learning process either through multicollinearity or redundant features. This process is called feature selection, where the primary aim is to find the optimal set of relevant features without losing the salient characteristics of the data (Büyükkeçeci & Okur, 2022). Feature selection is important for this study because it can simplify the model by reducing the number of influencing factors, thus decreasing the training time and reducing overfitting (Chen et al., 2020). Overfitting is an undesirable machine learning behavior where the model gives accurate predictions based on the training data, but not on new data, often as a result of large amount of irrelevant data in the training data. According to Yu and Lie, "*An optimal subset should include all strongly relevant features, none of irrelevant features, and a subset of weakly relevant features*" (Yu and Lie, 2004, p. 1208).

Building upon the method proposed by Ageenko et al. (2022), the landslide / non-landslide dataset was checked for linear relationships between the influencing factors. The concept of multicollinearity can be defined as: "*a condition where there is an approximately linear relationship between two or more independent variables*" (Chan et al., 2022, p.2). Multicollinearity becomes a problem mainly because it makes the estimates of the model unreliable and less precise because the variables affect each other. This also makes it difficult to measure the individual effect of a variable. Consequently, the model might not do well with new, unseen data because it is overfitted to the data it was trained on (Chan et al., 2022).

To check the linear relationship between the influencing factors, a correlation matrix was plotted for the data (see figure 22). However, as pointed out by Chan et al. (2022), a problem with such a measurement of multicollinearity is that the correlations don't necessarily mean multicollinearity. In other words, a correlation matrix measures how strongly two variables are related to each other. However, just because two variables have a high correlation, it doesn't automatically mean there is multicollinearity. This is because multicollinearity involves more than just two variables—it concerns the relationship among all variables in the model. High correlations between a couple of variables don't always impact the model in the same way that multicollinearity among several variables do (Chan et al., 2022). To address this problem, this study also uses the most widely used indicator of multicollinearity, the Variation Inflation Factor (VIF). As recommended by Chan et al. (2022), a VIF value of 10 was used. This threshold indicates that the uncertainty of the expected impact of one influencing factor on predicting landslide versus non-landslide outcomes is 10 times greater than it would be if the influencing factor were completely uncorrelated with the other influencing factors. Theoretically, there is no upper limit to how high a VIF can go. Starting at a score of 1, a VIF score lower than 5 usually indicate low multicollinearity, a score between 5 and 10 indicate moderate multicollinearity and values greater than 10 usually mean a high multicollinearity (Chan et al.,2022).

To compute the correlation matrix, a python script was used. The method is based on the work by Ageenko et al. (2022) and uses the python library pandas to load the CSV file for the training data into a pandas DataFrame. Initial preprocessing involves removing irrelevant columns such as "geometry" and several categorical variables that are not suitable for correlation analysis. The correlation matrix is then plotted using the Pearson's correlation coefficient and visualized through a heatmap, allowing easy identification of highly correlated variables (see figure 22).

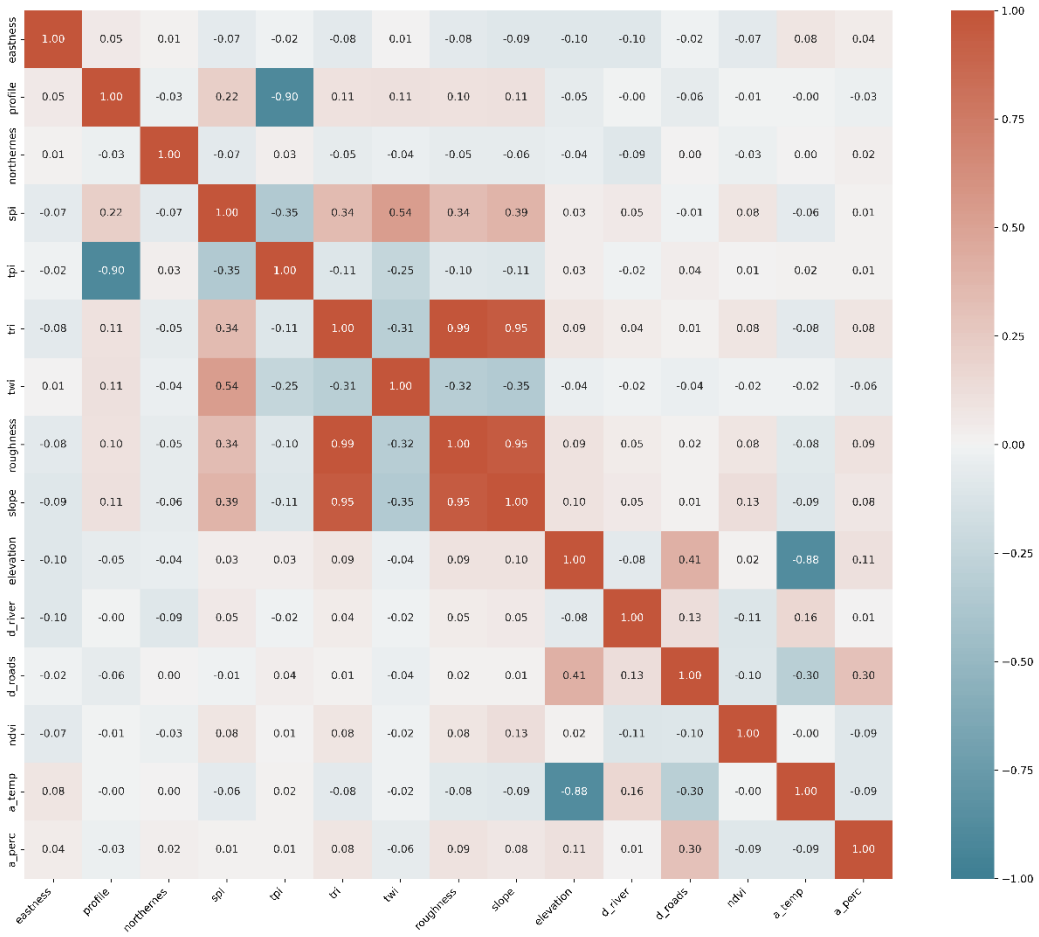


Figure 22: Correlation matrix with the Pearson correlation coefficient. The heatmap's colors range from blue (negative correlation) to red (positive correlation), with annotations providing the correlation values

As recommended by Ageenko et al. (2022), a correlation threshold of 0.75 was used to remove correlated influencing factors. A function was made in python that prints out pairs of influencing factors that that exceed this threshold. Once highly correlated pairs are identified, the script proceeds to select one influencing factor from each pair to remove, aiming to reduce multicollinearity. A total of five correlated pairs were identified: tpi and profile curvature, roughness and tri, slope and tri, slope and roughness and average temperature and elevation. This makes sense as tpi, profile curvature, slope and roughness are all derivative of a DEM that measures elevation differences between cells. As seen in figure 22, temperature and elevation show a negative correlation. This is as expected as temperature normally decrease as elevation increases. The reduction in multicollinearity is confirmed by plotting a new correlation matrix on the new variable where “annual temperature”, “roughness”, “slope”, “tpi” have been removed, as seen in figure 23.

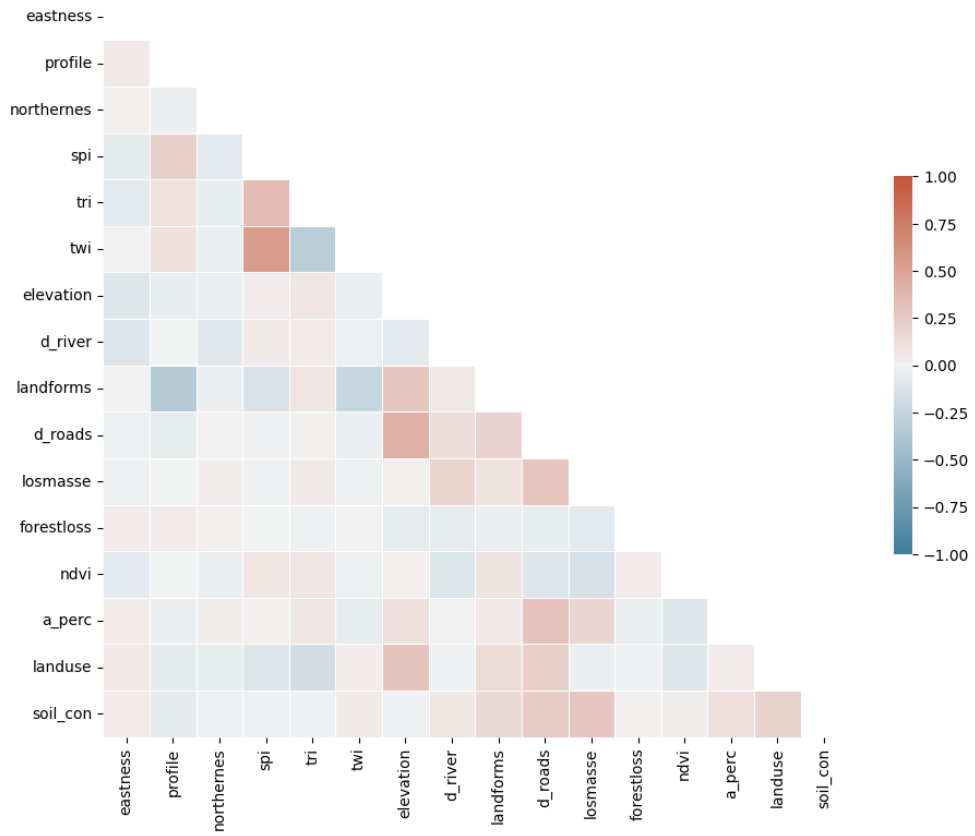


Figure 23: New correlation matrix without highly correlated features.

To investigate the data further for correlated influencing factors, the Variation Inflation Factor (VIF) was also used on the training data. The python module "statsmodels" was used, specifically the "variance_inflation_factor" tool [23]. The prepared python script loads the raw training data into a pandas DataFrame. For each column in the dataset, The VIF score is calculated. This DataFrame is sorted and printed to show which features have the highest VIF values, indicating where multicollinearity is most problematic. The code then enters a loop where it continuously recalculates the VIF value as variables are removed. The variable with the highest VIF value is removed if it exceeds the threshold of 10. This process repeats until no variables exceed this threshold. Based on this method, the script removed the variable "roughness" (VIF value of 107). "Average temperature", "slope" and "average precipitation" was also removed, having VIF values of 27, 18 and 13, respectively. Based on these two tests for multicollinearity, roughness, average temperature, slope and tpi was ultimately removed

[23]Further information can be found from the statsmodels 0.14.1 manual: https://www.statsmodels.org/stable/generated/statsmodels.stats.outliers_influence.variance_inflation_factor.html

from the training data. Data on precipitation was kept because it showed very low correlation with any other influencing factor in the correlation matrix.

In addition to the issue of multicollinearity (redundant features), the training data might also contain irrelevant features, defined as features that provide no useful information to the machine learning model (Büyükkeçecİ & Okur, 2022). Chan et al. (2022) and Ageenko et al. (2022) both utilized the Random Forest (RF) algorithm to enhance feature selection in classification tasks. Specifically, Chan et al. focused on selecting important features, while Ageenko et al. (2022) employed RF to eliminate irrelevant features using the algorithm's feature importance method.

To identify potential irrelevant features, the dataset was divided into training and testing subsets (70% as training and 30% for testing). Subsequently, a Random Forest (RF) classifier was trained using the training set. Post-training, the feature importance scores were extracted. These scores are a measure of each feature's contribution to the accuracy of the model's predictions and is provided by the attribute *feature_importance_* for the Scikit-Learn Python library [24]. As seen in figure 24, the derived scores were visualized in a bar chart, allowing for an identification of both the most and least influential features. Consequentially, the influencing factor “forestloss”, representing the “Hansen global forest change” was removed from the training data because it had very little significance for the model. The final tabular data is shown in table 7.

[24] Manual for feature importance with random forest: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

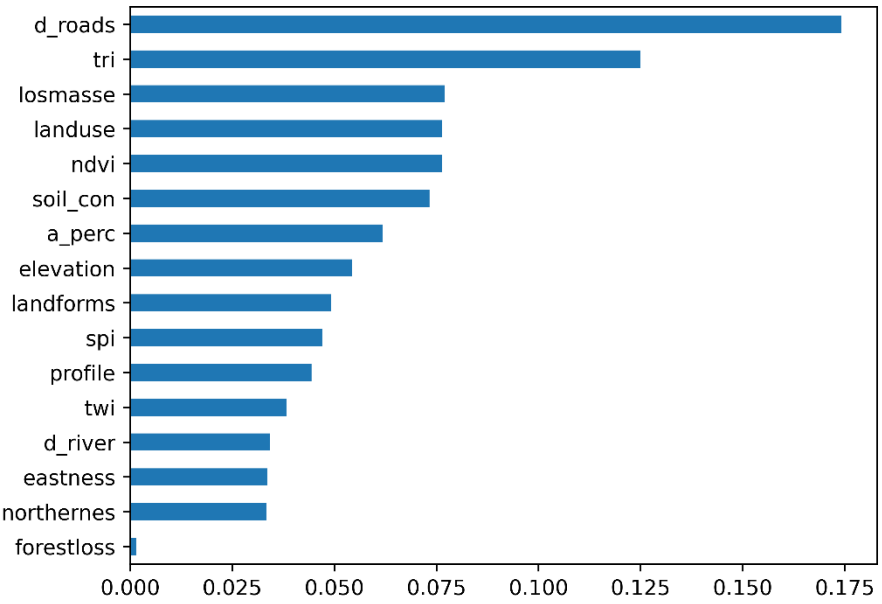


Figure 24: Plot of the feature importance for the training data. The variable "forestloss" is removed as it has very little significance for the model.

Table 7: Subset of the final attribute table after feature selection. The variables: tpi, roughness, annual temperature and forest loss have been removed from the original training dataset.

Landslide	eastness	profile	northernes	spi	tri	twi	elevation	d_river	landforms	d_roads	losmasse	ndvi	a_perc	landuse	Soil_con
0	0,717467	0,087356	-0,696593	-9,75531	0,241446	2,86326	9,52517	65,3753	5	497,529	130	0,000237	1932,23	3	4
0	0,088614	0,032632	-0,996066	1,62223	0,517594	4,77922	30,3799	56,8953	6	91,7769	90	0,271143	2017,62	5	4
0	0,943201	-0,008432	-0,332221	2,84369	0,872203	5,08092	18,4212	918,297	6	436,111	130	0,094776	2043,92	3	3
0	0,861211	0,134619	0,508248	3,88042	0,501435	7,21879	33,507	46,9129	6	395,233	130	0,324957	1931,86	5	3
0	0,76994	0,000484	-0,638116	0,59063	0,086718	7,67831	39,1515	464,079	5	600,347	130	0,28482	2026,98	2	3
1	0,921964	0,012124	0,387276	-11,5053	0,025865	4,69109	44,4742	512,117	5	0	130	0,309592	2058,17	4	7
0	-0,724829	-0,046523	0,688929	-0,187864	0,534695	3,18841	33,9022	319,934	6	298,765	130	0,28353	1922,1	5	3
1	-0,986135	0,012216	-0,165943	-3,36194	0,087333	4,08245	18,7269	348,878	6	0	130	0,303982	1930,74	4	7
1	-0,84511	0,015661	-0,534592	-2,42172	0,108209	4,23452	18,5004	316,881	6	0	130	0,28353	1929,1	4	7
1	-0,840918	0,019229	-0,541162	-11,6843	0,049144	4,88956	19,108	330,965	2	0	130	0,303982	1930,74	4	7
1	0,944995	0,006471	0,327085	-10,7009	0,058695	3,83392	18,6636	335,758	2	0	130	0,303982	1929,71	4	7

4.5. Machine Learning

After completing the data preparation steps, the training data, which includes both the influencing factors and the target variable, is now ready to be used for training a machine learning model. The Random Forest algorithm is selected based on its robust performance characteristics and suitability for handling complex datasets with multiple input features, as detailed in section 2.3.4. However, to substantiate the choice and ensure the selection of the most effective machine learning technique for our dataset, Auto-sklearn will also be used to train on the dataset. Auto-sklearn automatically implements optimal algorithm selection and hyperparameter tuning (Feurer et al., 2022). By using the built-in leaderboard function which

facilitates a comparison of all evaluated models, the best performing model type can be selected [25](See chapter 2.4.3 for more detail).

To run the Random Forest workflow, the python-based machine learning library scikit-learn was used. Scikit-learn is a free and open-source and features various classification and regression algorithms, including Random Forest (Pedregosa et al., 2011). Firstly, the csv file including predictor variables and the target variable is imported using pandas. Then, all influencing factor raster data are stored as TIFF-files in a list; these raster files are then stacked into a single raster object using the PySpatialML library, facilitating the manipulation and analysis of multi-band raster data [26]. Because the columns in the tabular training data is based on the influencing factors in the raster stack, we can train the model on the tabular data and use the trained model to predict on the raster stack (Ageenko et al., 2022).

Moreover, the tabular data underwent preprocessing where categorical and numerical features are handled differently. Categorical features (landforms, surficial deposits, land use and soil condition) are transformed so that each categorical class are represented as a new binary feature coded presence (1) or absence (2) of the category. This is beneficial because it removes any ordinal relationship and treats each class separately [27]. Numerical features are standardized using a *StandardScaler* to normalize their distributions, improving the performance and stability of the machine learning model [28].

Next, the landslide prediction model is constructed using a Random Forest Classifier. The tabular training data is divided into two arrays: one for the predictor variables and one for the target variable. The model utilizes these arrays to learn the relationships between the values in the predictor variables and the binary outcomes in the target variable. Furthermore, the data was partitioned into a training and a testing subset, containing 70 and 30 % of the data, respectively. The splitting was stratified after classes to ensure a balanced representation of landslide and non-landslide samples. The result is 977 samples used for training and 419 samples for testing.

[25] The leaderboard gives an overview of all models trained during the search process along with various statistics about their training: https://automl.github.io/auto-sklearn/master/api.html
[26] Pyspatialml is a Python module for applying scikit-learn machine learning models to 'stacks' of raster datasets: https://pypi.org/project/pyspatialml/
[27] The sklearn.preprocessing.OneHotEncoder is used: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html
[28] StandardScaler standardizes features by removing the mean and scaling to unit variance: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

Based on the work by Ageenko et al. (2022), a grid-search cross validation is implemented to optimize model performance. According to Liu et al. (2023), choosing the right hyperparameters is always helpful to improve performance. The Grid Seach method is the easiest and most direct form of hyperparameter optimization and is widely used in LSM (Liu et al., 2023). The grid search provided by the *GridSearchCV* function in the Scikit-learn library generates candidates from a grid of parameter values. The grid search explores all possible combinations of hyper-parameters to find the best model configuration based on accuracy [29]. After training, the model's performance is evaluated on the test subset, comprising 30% of the original data. The test set is withheld during the training phase and is therefore an unbiased benchmark to evaluate the performance of the model on unseen data.

According to Reichenbach et al. (2018), evaluating the models prediction performance (the ability of the model to predict new landslides) is recommended in contrast to evaluating the model fit (the model's ability to describe the known distribution of landslides, assessed by comparing its outcomes with the landslide data used for training) (Reichenbach et al., 2018). As pointed out by Liu et al. (2023), the most used evaluation methods for classification problems include accuracy, precision, recall, F1-score, and ROC_AUC (see section 2.4.2 for more detail). As for LSM, the accuracy score and ROC_AUC are one of the recommended metrics to evaluate the performance of a trained model (Liu et al., 2023). The accuracy score has been shown to be a useful metric for binary classification tasks with balanced negative/positive samples (Liu et al., 2023).

The performance evaluation began with the generation of a classification report using the *classification_report* function from scikit-learn *sklearn.metrics*-module. This report computes the precision, recall, and F1-score for each class ("landslide" and "non-landslide"). Additionally, the overall accuracy of the model was calculated using the *accuracy_score* function. To further visually assess the model's performance, a confusion matrix was plotted using *ConfusionMatrixDisplay.from_predictions*, illustrating the counts of true positive, false positive, true negative, and false negative predictions. Furthermore, as recommended by Liu et al. (2023), The Receiver Operating Characteristic (ROC) curve was generated to evaluate the model's diagnostic ability. By plotting the true positive rate against the false positive rate at various threshold settings, this curve provides a graphical representation of the trade-off

[29] https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html#sklearn.model_selection.GridSearchCV

between sensitivity and specificity. The ROC was plotted with *RocCurveDisplay* function from *sklearn.metrics* module.

The auto-sklearn workflow was run on a Linux operating system, through a Windows Subsystem for Linux (WSL) [30]. The workflow for auto-sklearn is similar to that of Random Forest, as auto-sklearn is built around the scikit-learn library. However, as auto-sklearn automatically searches for the right learning algorithm and optimizes its hyperparameters, no preprocessing is necessary. The training data was imported using pandas, split into training and testing sets, and fitted to the autosklearn classifier, as shown in code block 9. A classification report, including overall accuracy, precision, recall and F1-score was computed, as well as a confusion matrix and ROC_AUC curve [31]. Finaly, the trained auto-sklearn classifier is tested on the external validation data.

```
cls = autosklearn.classification.AutoSklearnClassifier(
    time_left_for_this_task=240*60,
    per_run_time_limit=None,
    n_jobs=8
)
cls.fit(X_train, y_train, X_test, y_test)
```

Code 9: The auto-sklearn classifier fitted to the training data. The classifier was run for two hours on all 8 of the computer's CPU cores.

4.6. External Validation

During the development of the nationwide landslide susceptibility mapping of Norway, Fischer et al. (2014) used different threshold values to define susceptible start zones for landslides in different regions. Due to the highly varying topography, geomorphology, geology, surficial sediment cover, and climate in Norway, a model trained on data from one region in the country is not expected to have the same predictive accuracy when fitted to unseen data in a different region (Fischer et al., 2014). Subsequently, when selecting model hyperparameters, the risk of overfitting (overlearning the relationship between the predictor variables and the landslide/non-landslide class in the training data) can become a problem; Over-learning the patterns in the training data results in the degradation of the model's predictive ability, when exposed to new data (Ageenko et al., 2022). To avoid this, and to explore how well the model can generalize on data from a different geographical region, the

[30] The full system requirements for Auto-Sklearn can be found here: https://automl.github.io/auto-sklearn/master/installation.html

[31] Auto-sklearn uses the same functions to compute the accuracy metrics as the scikit-learn library.

model with the tuned hyperparameters is validated on external data. Crucially, the external data has not been used in the training of the model or tuning of the hyperparameters.

As shown in figure 25, a dataset outside of the AOI is used for this purpose, as described in section 4.1, with a total of 466 landslide points, with an equal distribution of landslide registries and randomly generated non-landslide points, as described in section 4.2.1. The process for generating external validation data follows a similar workflow to that of the training data. First, the AOI is defined. Then, point data for landslides is generated using the same SQL query, and point data for non-landslides is created using the same Python script, as outlined in section 4.2.1. New influencing factors are generated for the external validation area, using the same Python scripts employed for the training data (see figure 26). However, the input data are adjusted to reflect the new AOI. This adaptation involves using a different DEM and updating data sets for rivers, roads, surficial deposits, land use, precipitation, and Sentinel 2 satellite imagery specific to the new geographical area. Because the external validation is used to validate the trained RF-model, we do not want to introduce new predictive variables. Hence, the influencing factors removed during the feature selection process is similarly ignored for the external validation. The appropriate raster data is then mapped onto the landslide / non-landslide point data using the code described in section 4.3.

For the machine learning task, the external dataset is prepared by splitting it into predictive variables and the target variable, mirroring the data preparation done for the training dataset. This ensures that the RF model is tested under consistent conditions. Using this external validation set, the model predicts landslide locations, and the results are analyzed using the same performance metrics. This approach provides a direct comparison between the performance of the model on the training dataset and its effectiveness in a new region in Norway. This will be discussed in chapter 5.

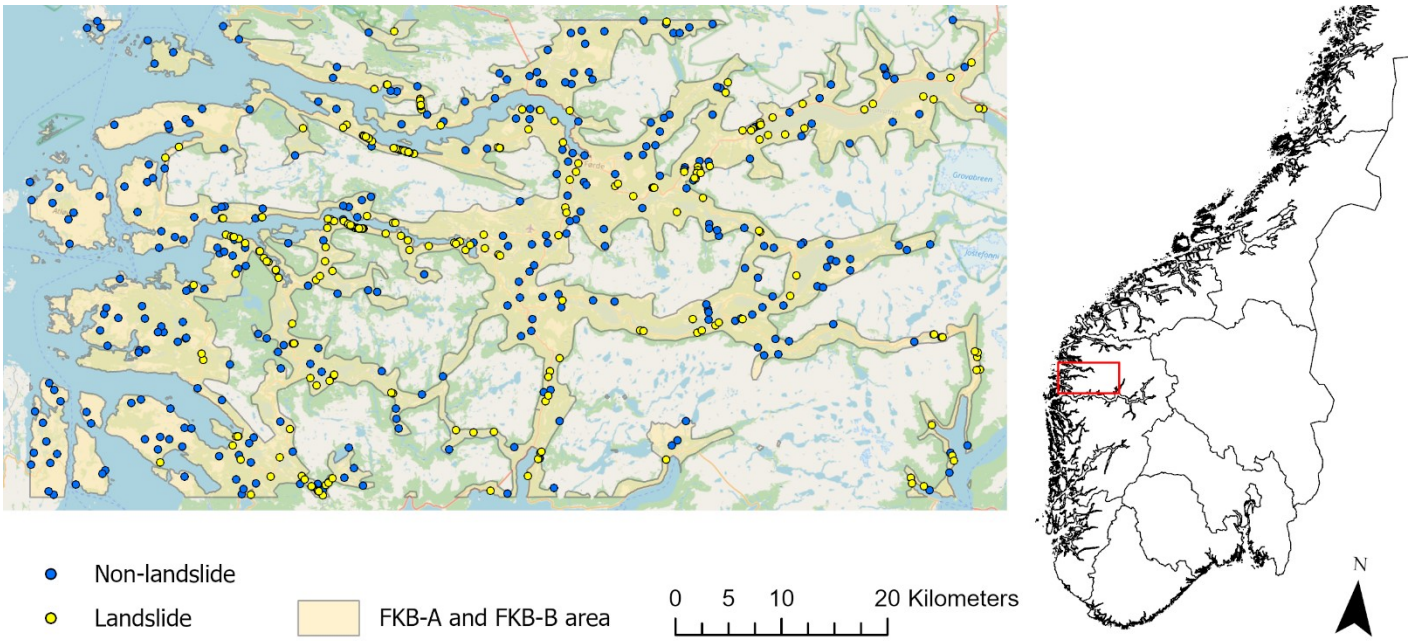
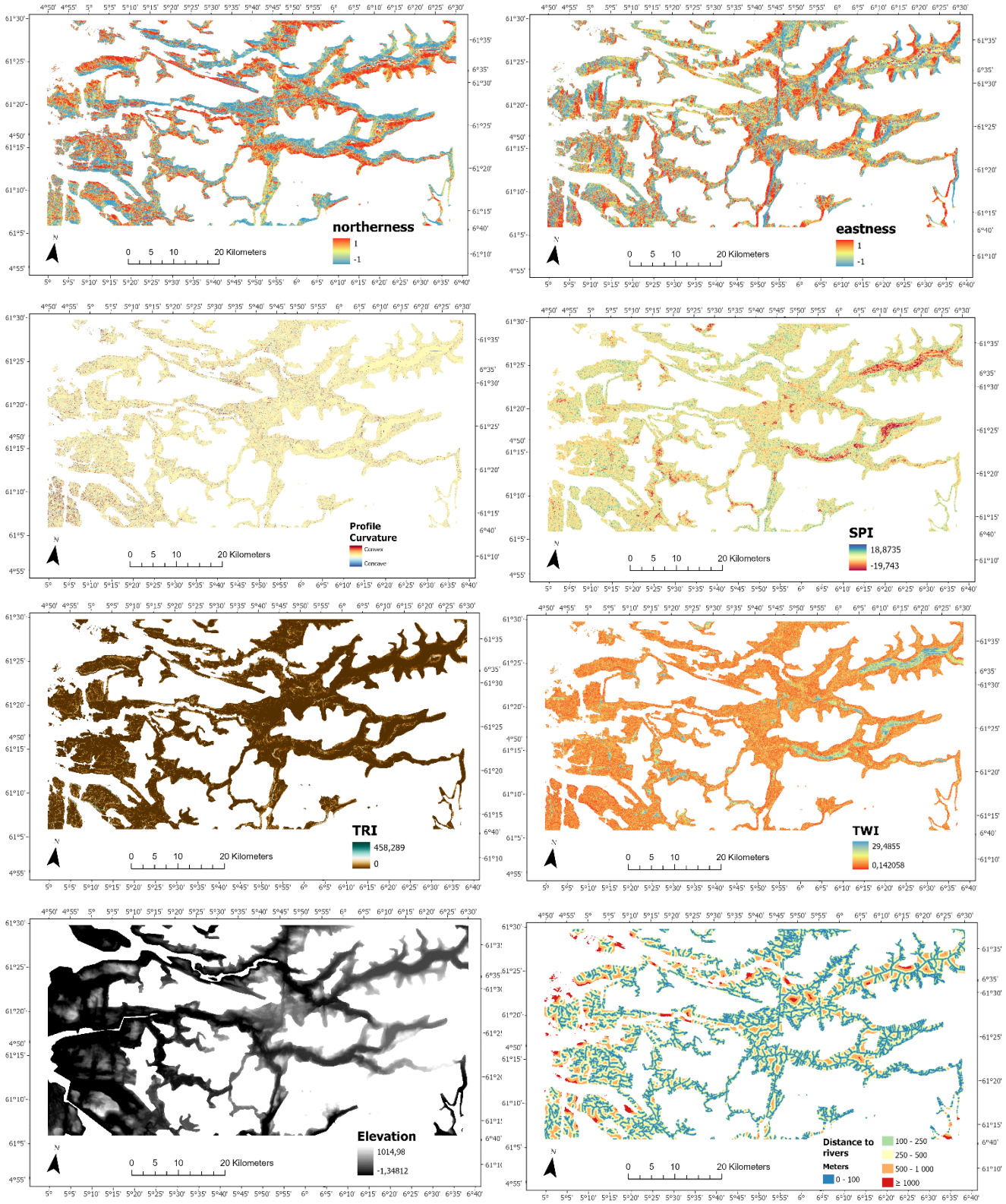


Figure 25: Area for external validation, including the 250 landslide points and the 250 randomly generated non-landslide points.


northerness
1
-1
0 5 10 20 Kilometers
eastness
1
-1
0 5 10 20 Kilometers
Profile Curvature
Convex
Concave
0 5 10 20 Kilometers
SPI
18,8735
-19,743
0 5 10 20 Kilometers
TRI
458,289
0
0 5 10 20 Kilometers
TWI
29,4855
0,142058
0 5 10 20 Kilometers
Elevation
1014,98
-1,34812
0 5 10 20 Kilometers
Distance to rivers
Meters
0 - 100
100 - 250
250 - 500
500 - 1 000
≥ 1000
0 5 10 20 Kilometers

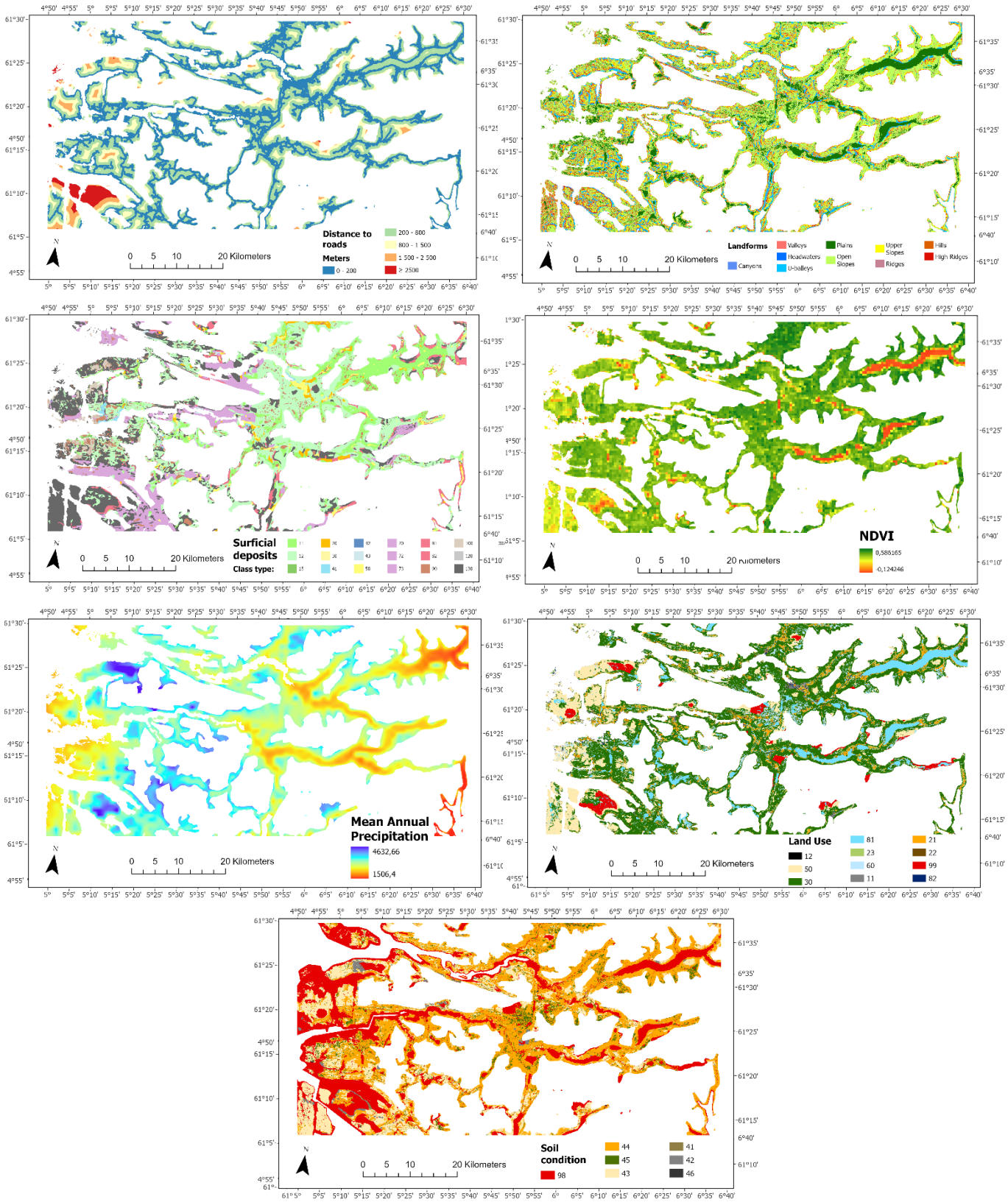


Figure 26: Conditioning factors for the external validation area.

4.7. Landslide susceptibility map

To gain an understanding of which areas are vulnerable to landslides, the trained machine learning model alone is not sufficient. We need a visual representation of landslide susceptibility, a landslide susceptibility map. A landslide susceptibility map serves the purpose of showing users and planners which areas are most susceptible to landslides, aiding policymakers and planners in developing proactive measures to mitigate the risks associated with landslides (Reichenbach et al., 2018).

To produce the LSM, the trained RF model is used to predicts the probability of a landslide for each cell in the stack of influencing factor raster data. This is possible because the columns in the training data align with the order of the raster data in the stack. The Python library *pyspatialml* is specifically designed for applying a fitted machine learning model to make predictions across pixels in a raster stack. Unlike Python's *numpy* module, which requires data to be held in memory, *pyspatialml* can operate directly on raster datasets stored on disk [32]. This capability is particularly useful for handling large datasets. As a result, the probability score was calculated for each cell in the raster stack using the *predict_proba* method as shown in code 12 [33].

```
predictors = ["eastness.tif",
              "profile.tif",
              "northernes.tif",
              "spi.tif",
              "tri.tif",
              "twi.tif",
              "elevation.tif",
              "d_river.tif",
              "landforms.tif",
              "d_roads.tif",
              "losmasse.tif",
              "ndvi.tif",
              "a_perc.tif",
              "landuse.tif",
              "soil_con.tif",
             ]
stack = Raster(predictors)
stack.names

result_rf = stack.predict_proba(model_rf)
result_rf.write("<output_LSM_raster>")
```

Code 10: Code to predict landslide probability scores on the raster stack, based on the trained Random Forest model, called "model_rf" in the script.

[32] More detailed description of pyspatialml: https://pypi.org/project/pyspatialml/
[33] The predict_proba method outputs the class probability as a multi-band raster (a band for each class probability).

According to Reichenbach et al. (2018), landslide susceptibility maps often face challenges in practical application due to several key reasons. Non-specialists often find the resulting maps difficult to interpret, as they use descriptive terms without quantitative measures; Susceptibility levels are commonly displayed in descriptive terms (very high, high, medium, low, and very low) but areas defined as having “very low” susceptibility can still receive landslides. Similarly, the outputs of statistical classification models, such as probability values around 0.5 (in the range from 0 to 1), are commonly misunderstood; this value does not necessarily mean that the area has a moderate risk of landslides; rather, it indicates that the model is unsure about the outcome of the prediction (landslide or non-landslide area).

Taking this into consideration, the output LSM raster data is classified to get a prediction pattern for better visualization. The LSM is classified based on the equal-interval method, where the range of all the pixel values is divided into several equal intervals with identical range (Chung & Fabbri, 2003). The classification is set to five intervals, breaking based on the value range (0 to 1) and visualized with red colors for susceptible areas, green colors for non-susceptible areas and yellow for areas where the model is uncertain in its prediction.

Chapter 5

Results and Discussion

The previously described methodology was applied to the presented area to map the likelihood of landslides in soil occurring on the basis of the proposed influencing factors and the obtained landslide data. To address the research questions, a random forest model was used to test how well machine learning performs for LSM in Vestland county. The random forest algorithm was applied on the data to evaluate the feature importance of the influencing factors. Additionally, the performance of the Random Forest algorithm was compared with that of the top-performing auto-sklearn algorithm to determine whether automated machine learning can surpass traditional approaches in this context. Finally, both machine learning strategies was tested using an external validation dataset to assess their predictive accuracy in identifying landslide-prone areas across a different geographical region.

5.1. Machine learning for Landslide Susceptibility Mapping

The nature of machine learning, fundamentally detached from specific scientific disciplines, is to optimize the prediction accuracy based on the patterns discerned from data. This process is less about adhering to the established physical or geological laws and more about recognizing and leveraging statistical correlations that the data presents (Langley, 1988). For supervised learning, the trial-and-error method is mainly used (Aurélien, 2019). In other words, an experimental approach. There are an endless amount of possible variable interactions and correlations that can explain the formation of landslides; the goal of machine learning is to best attempt to capture this structure. The job of the machine learning algorithm is not to discover why landslides occur. Rather, its role is to optimize a function that predicts the likelihood of a landslide occurring, based on the data provided.

There is no standard guideline for selecting the right influencing factors, and the importance of influencing factors differs from one study area to another (Ado et al., 2022). Similarly, advanced models generated with ensemble techniques have been shown to improve the mapping accuracy of LSM. However, there is no existing series of standards to determine

what type of machine learning-based methods should be used on a specific task (Liu et al., 2023; Reichenbach et al., 2018).

The research in this thesis addresses several questions: Firstly, can machine learning be effectively used to predict landslides in the distinct geological context of Vestland county? Secondly, to explore whether a model trained on this localized dataset can extend its predictive capabilities to another geographical region, thus examining the generalizability of the machine learning model. Thirdly, the thesis compared the performance of automated machine learning against the traditional random forest algorithm to explore whether state of the art methods can improve the predictions. Lastly, the random forest model was used to explore which local terrain conditions contribute to landslide formation.

5.2. Landslide Susceptibility Mapping using Random Forest

As discussed in chapter 2.3.4, the random forest algorithm, together with other ensemble techniques, have been proven to be one of the best performing algorithms for LSM (Ado et al., 2022; Ageenko et al. 2022; Liu et al., 2023). However, no such guarantee exists when predicting landslides in a new geographical area with an unproven landslide inventory and a new set of influencing factors. To evaluate the performance of the model, multiple performance metrics was used (see chapter 2.4.2). *Overall accuracy* reflects how often the model correctly predicts the outcome. A *confusion matrix* represents the accuracy of model performance by displaying true positives, false positives, true negatives, and false negatives in a table format. Precision is the accuracy score of the positive prediction. Recall is the ratio of positive instances that are correctly classified. The F1 score is the harmonic mean of precision and recall. The *receiver operating characteristic curve* (ROC) plots the true positive rate against the false positive rate to visualize the ability of the model to differentiate between classes. According to Ado et al. (2022), a ROC curve can represent the summary of overall performance and is the most popular method to evaluate performance for LSM.

The accuracy metrics is seen in table 8. The accuracy metrics are plotted using the classification_report function for the scikit-learn python library [34]. The confusion matrix is

[34] Creates a text report showing the main classification metrics. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html

plotted using the *ConfusionMatrixDisplay* function and the ROC curve is plotted using the *RocCurveDisplay* function [35].

Table 8: Overall accuracy, precision, recall and F1-score for the Random Forest model on the test set.

Model	Overall accuracy	Precision	Recall	F1-score
Random Forest	**0,88**	**0,87**	**0,90**	**0,88**

The random forest classifier distinguished well between the landslide-class and the non-landslide class. Out of 419 cases in the test set (30% of the training data), 180 were true negatives, 189 were true positives, 29 were false positives and 21 were false negatives (see figure 27). This means that 189 real landslide records were correctly classified and 180 false landslide records were correctly classified as such. However, the classifier was not perfect. 21 real landslide records were incorrectly classified as landslides and 29 non-landslide records were incorrectly classified as landslides. From these values, the overall accuracy was computed to approximately 88,07%. This metric indicates a high level of accuracy in the model's ability to distinguish between landslide and non-landslide instances.

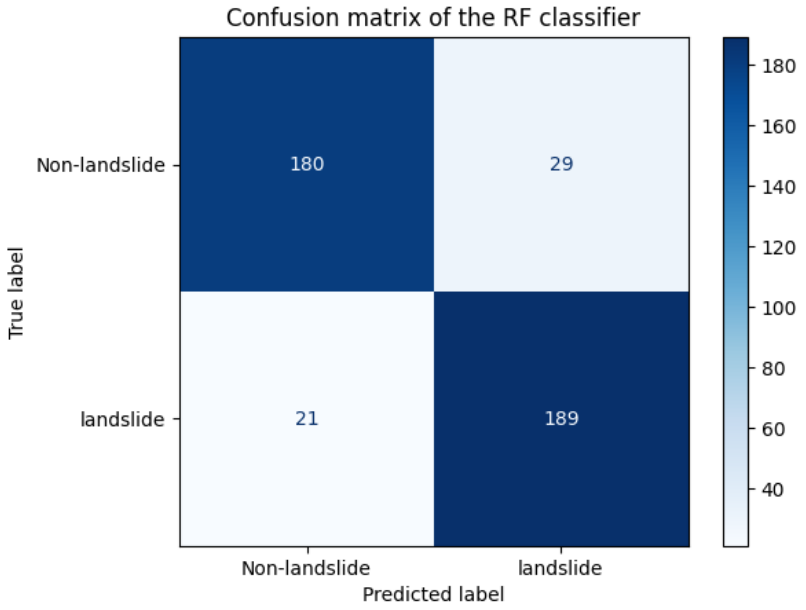


Figure 27: Confusion matrix for the random forest classifier on the test set.

[35] Confusion matrix from scikit-learn: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.ConfusionMatrixDisplay.html#sklearn.metrics.ConfusionMatrixDisplay.from_predictions

The overall accuracy is a valuable metric given that the distribution of landslide and non-landslide classes are balanced in the dataset (Liu et al., 2023). The accuracy compares relatively well with similar studies, namely the Danish case study form 2022 that had an overall accuracy score of 91% for the RF model (Ageenko et al., 2022). Case studies have shown varying predictive accuracies for RF models: Sun et al. (2020) reported slightly better accuracy at 92%, while Kavzoglu & Teke (2022) documented a lower accuracy of 81%.

Furthermore, as shown in Table 6, the model achieved a precision score of 0,87. This high precision indicates that when the model predicts a landslide event, it is likely to be correct 87% of the time. This is particularly important in LSM where false positives can lead to an exaggerated amount of highly susceptible areas. Similarly, a recall of 90% is also significant because it shows that the model is highly effective at identifying most of the actual landslide events. In practical terms, this means the model misses very few actual landslides.

The ROC curve displays the recall (true positive rate) on the Y-axis and the false positive rate (FPR) on the X-axis across various threshold settings. Essentially, the curve shows how well the model distinguishes between landslide and non-landslide events. The threshold settings represent cut-off values used to determine whether a prediction is positive or negative. A binary classification model returns a number from 0 to 1 for each record. A landslide record can have a value of 0.2, 0.7, 0.99, or any other number. At which probability a record is converted to a label (landslide / non-landslide) represents the threshold used for the ROC-curve. The function varies the threshold from 0 to 1 and calculates the true positive rate and false positive rate for each threshold. These rates are then plotted on the ROC curve, providing a visual representation of the trade-offs between detecting true positives and avoiding false positives (Aurélien, 2019).

As presented in figure 28, the sharp rise along the y-axis at lower threshold values of FPR suggest that the model quickly reaches a high recall with a minimal increase in false positives. At a threshold around 0.6, further increase in recall (true positive rate) requires an increase in FPR. However, it is quite steady up to a threshold of 0.8 where further increase in recall significantly increases the FPR. The AUC is the area under the ROC curve and is used to evaluate the performance of classification problems (Aurélien, 2019). AUC measures the capability of ML model in distinguishing the classes. An AUC value greater than 0.7 is generally considered to indicate good accuracy, and it is not uncommon for machine learning models to achieve values above 0.9 (Ado et al., 2022).

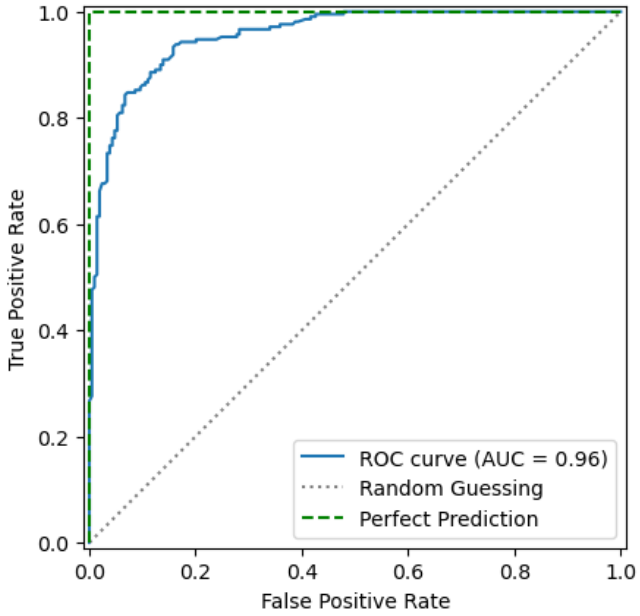


Figure 28: Plot of the ROC curve for the RF model and the corresponding AUC value of 0.96. The diagonal line represents a score of 0.5 (random guess) and the green line represents a score of 1 (perfect prediction)

As seen in figure 28, the ROC curve is way above the "random guessing line" with a score of 0.96, indicating that the model distinguishes very well between the landslide and the non-landslide class. This result compares well with the Danish study that showed an AUC score of 0,98 for the RF model (Ageenko et al., 2022).

The ROC plot illustrates the trade-off between correctly predicting landslides and incorrectly predicting landslides at different thresholds. When the recall increases, the more false positives the classifier produces (Aurélien, 2019). As shown in figure 28, at higher threshold values, the model becomes more conservative in assigning the true label, but the recall decreases. Lowering the threshold the model becomes less strict in assigning the true label, thus the recall increases. However, the model now makes more false positive predictions. This trade-off represents a dilemma; should the model have as high of a recall as possible in order to predict as many landslides as possible while consequently accepting that a substantial amount of these predictions are false positives, or should the model be conservative in its prediction? This is a question deemed outside the scope of this thesis; however, it is an important decision that affects the resulting susceptibility map. The balanced threshold of 0,5 was chosen for this study. As shown in figure 29, threshold values of 0.3 and 0.5 results in more correct landslide predictions but more non-landslide record incorrectly labelled as landslide. In contrast, high threshold values of 0.7 and 0.9 almost removes all false positive

but now the amount of correct landslide prediction goes down, as well as false negatives going up dramatically.

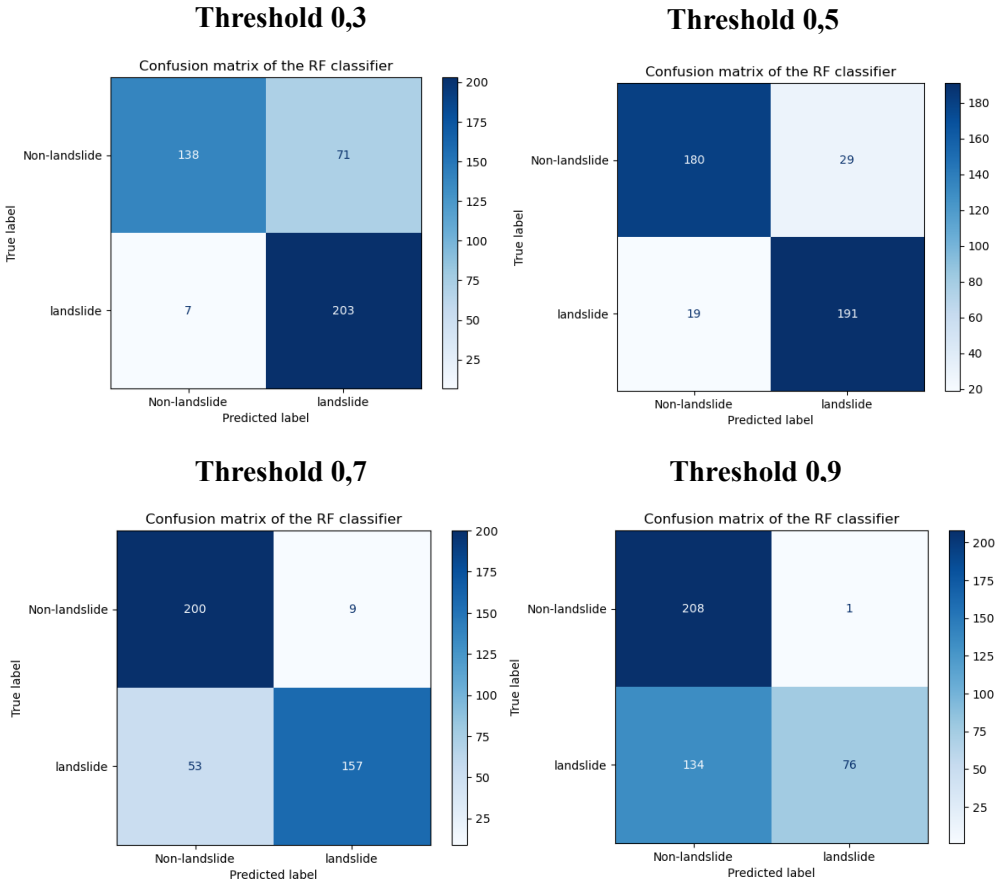


Figure 29: Confusion matrix for the RF model with classification thresholds of 0.3, 0.5, 0.7 and 0.9.

5.2.1. External Validation

The inclusion of external validation is proposed as a way to explore the generalizability of the machine learning model. This method was used to evaluate if the random forest algorithm, trained on data in one geographical region, is capable of predicting landslides in another region. The overall accuracy of the model on the external validation dataset is 76%, which, while indicating a decrease, still represents a reasonable level of prediction accuracy (see table 7). The ROC curve demonstrates that the false positive rate increases more at lower classification thresholds. The AUC score also drops from 0.96 to 0.87 (see Figure 31). Despite this decrease, the AUC score is still considered good, surpassing the threshold of 0.7 indicating good predictive accuracy (Ado et al., 2022).

Table 9: Overall accuracy, precision, recall and F1-score for the Random Forest model on the external validation data set.

Model	Overall accuracy	Precision	Recall	F1-score
Random Forest	**0,76**	**0,80**	**0,69**	**0,74**

As seen in figure 30, out of 466 cases (100% of the external validation data), 195 were true negatives, 160 were true positives, 39 were false positives and 72 were false negatives. The classifier is now worse at correctly predicting landslides with 72 actual landslide records incorrectly labelled as “non-landslide”. This can be attributed to overfitting and to the different spatial distribution of the surficial deposit classes. Class 30 (lake deposits), 72-73 (type of weathering materials) and 100 (thin peat cover) are present in the external validation data but not in the training data. Furthermore, classes 70 (weathering material) and 81 (landslide material) are characterized by an overweighting of landslide points in the training AOI. However, in the external validation area, these classes predominantly contain non-landslide points, or the classes themselves are rare. Only 50 out of a total of 232 landslide points are found inside these classes. For the external validation area, the classes 11 and 12 (thick and thin moraine material, respectively) contain most of the landslide points. However, for the training data, only 191 out of the 699 total landslide points are found inside these classes. A similar problem was reported by Ageenko et al. (2022), but the random forest method showed better generalizability on the external validation. This could be because of higher quality landslide inventory or because regions within Vestland county in Norway possibly exhibit a more varied and distinct geology, climate, and topography compared to different regions in Jutland, Denmark.

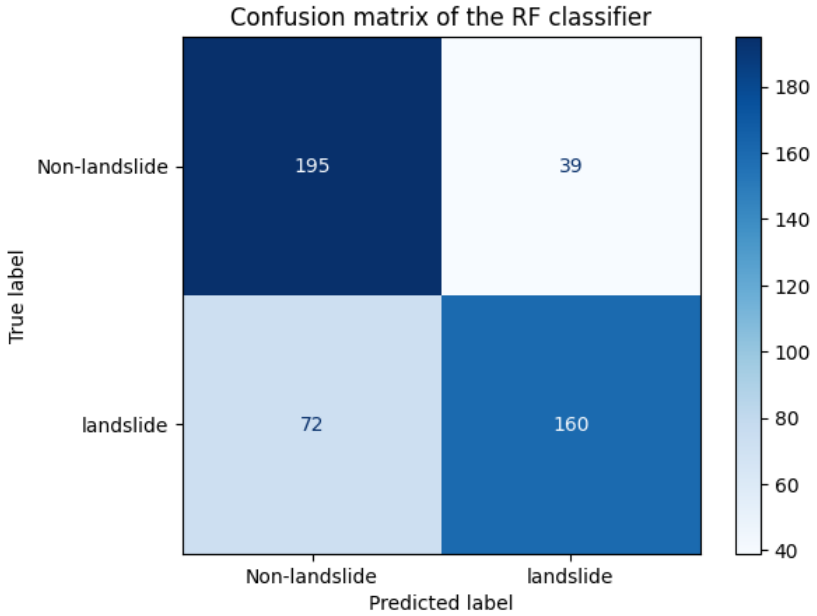


Figure 30: Confusion matrix for the random forest classifier on the external validation dataset.

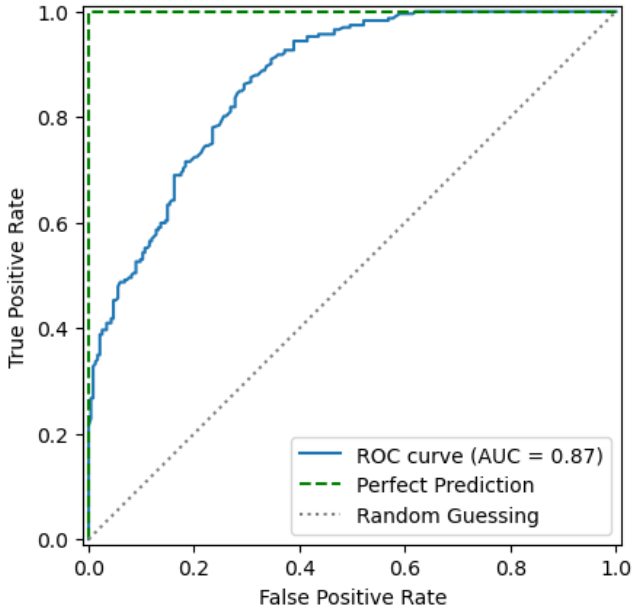


Figure 31: Plot of the ROC curve for the external validation dataset with the corresponding AUC value of 0,87. The diagonal line represents a score of 0,5 (random guess) and the green line represents a score of 1 (perfect prediction).

5.2.2. Susceptibility Map

The accuracy of any machine learning model hinges on the quality of the input data. 'Garbage in, garbage out' is a classic saying in any data modelling, including machine learning. The phrase expresses how problematic input data will produce problematic outputs (Geiger et al., 2021). Even though the RF model used in this study showed good predictive accuracy for the

training area, it does not inherently validate the correctness of the predictions. Instead, this accuracy reflects the model's proficiency in identifying patterns and trends within the specific dataset it has been trained on. Hence, if the training data, such as a landslide inventory, contains inaccuracies or biases - due to uneven mapping of landslide occurrences or other data collection inconsistencies - these imperfections will be mirrored in the output of the landslide susceptibility map. The predictive power of the machine learning model is fundamentally tied to the quality of its training data; it can only predict accurately based on what it has been taught.

As seen in figure 32 (b), the susceptibility map shows high and very high probabilities of landslide along large roads and forests roads and on steep valley sides. The probability values reflect the model's certainty in labelling a cell as either "landslide" or "non-landslide". Thus, values close to 0.5 does not necessarily reflect moderate landslide risk; rather, it indicates that the model is unsure about the outcome of the prediction. Nevertheless, the high and very high probability values (between 0.6 and 1) indicate areas where the model is confident in assigning the "landslide" label. Areas defined as very low and low probability values (between 0 and 0,4) indicate high confidence in assigning the "non-landslide" label.

Furthermore, many of the highly susceptible areas are found within the surficial deposits class 81 and 12 (landslide material and moraine material, respectively). Additionally, as the RF model is trained on a landslide inventory where 61% of the landslide registries are either mapped by the Norwegian Public Roads Administration (SVV) or the Norwegian Railroad Administration (JBV), there is a bias in the distribution of landslide registries across the AOI. The majority of the landslides mapped by SVV and JBV are mapped on, or very close to the road and railroad network (Ganerød et al., 2023). This is because SVV often map landslides if there has been an incident that has affected their road operations. This results in a lack of registrations of the actual trigger area for the event, but rather a point along a road that is located down in the valley (Fischer et al., 2014). Figure 32(a) illustrates an example of this bias, where areas highly susceptible to landslides are shown running parallel to roads. While road networks can directly destabilize hillslopes (Sidle & Ochiai, 2006), it is likely that the overrepresentation of landslide registries along these roads contributes to this trend, rather than proximity to roads being the primary causative factor for landslides.

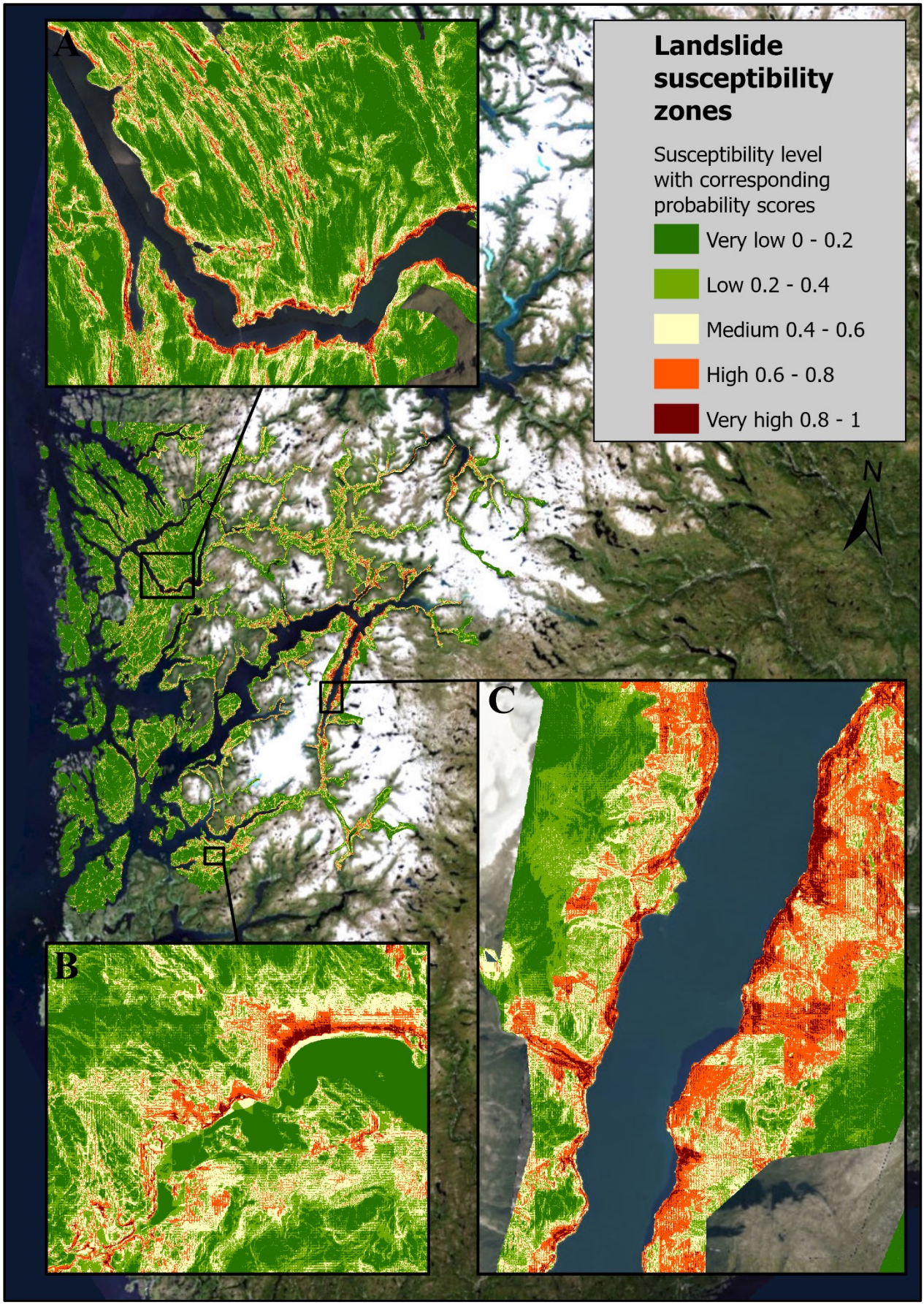


Figure 32: LSM of the training AOI using the random forest classifier. Cell values represent the probability score by the classifier; Between 0 and 1.

5.3. Landslide Susceptibility Mapping using Auto-Sklearn

To investigate whether automated machine learning can surpass the established random forest algorithm, based on predictive performance on the training data and on the external validation data, an Auto-sklearn classifier was run on the same training and external validation data as the RF model, described in chapter 4.5. However, due to the automated nature of Auto-sklearn, the training data only need to be split into a training subset (70% of the data) and testing subset (30% of the data). There is no need to preprocess the data; Auto-Sklearn automatically searches for the right algorithm and optimizes its hyperparameters based on the data it is trained on (Feurer et al., 2022). The process was executed using the *AutoSklearn2Classifier* function, running for 2 hours[36]. Results were visualized on a leaderboard showcasing the top-performing algorithms. For the training dataset, the best-performing model was a multilayer perceptron (MLP), a type of artificial neural network. Notably, MLPs have been previously applied in landslide susceptibility modelling (LSM), as demonstrated by Azarafza et al. (2021), who compared various deep learning methods for LSM.

The MLP approach used the same performance metrics as the RF approach. The MLP model, fine-tuned by Auto-Sklearn for optimal performance, excelled at correctly predicting landslide records with a total of 198 true positives, surpassing the RF model's 189. However, the MLP model identified fewer non-landslide records correctly, with only 171 compared to the RF model's 180. Additionally, the MLP model had fewer false positives but slightly more false negatives, as shown in figure 33. The overall accuracy score of 88% was identical to the RF mode (see table 10). Furthermore, the MLP showed an AUC score of 0.95, indicating that the model distinguishes very well between landslide and non-landslide events. The AUC score is very similar to the RF model.

Table 10: Overall accuracy, precision, recall and F1-score for the Multilayer perceptron artificial neural network model on the test set.

Model	Overall accuracy	Precision	Recall	F1-score
MLP	**0.88**	**0.89**	**0.88**	**0.89**

[36] Manual for Auto-Sklearn 2.0: https://automl.github.io/auto-sklearn/master/manual.html

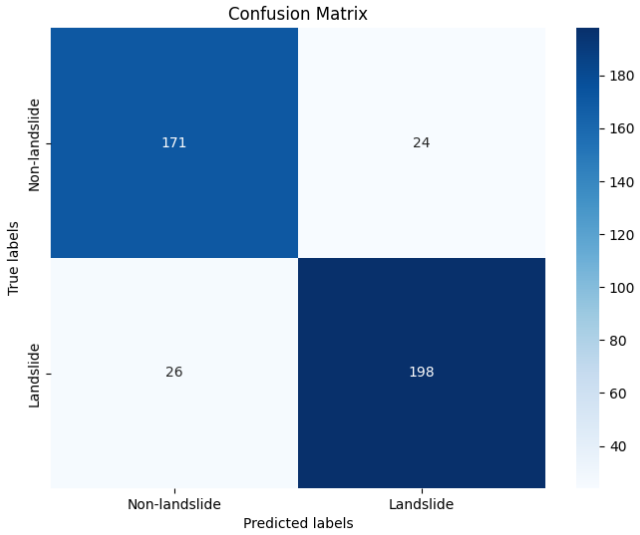


Figure 34: Confusion matrix for the multilayer perceptron model on the test set.

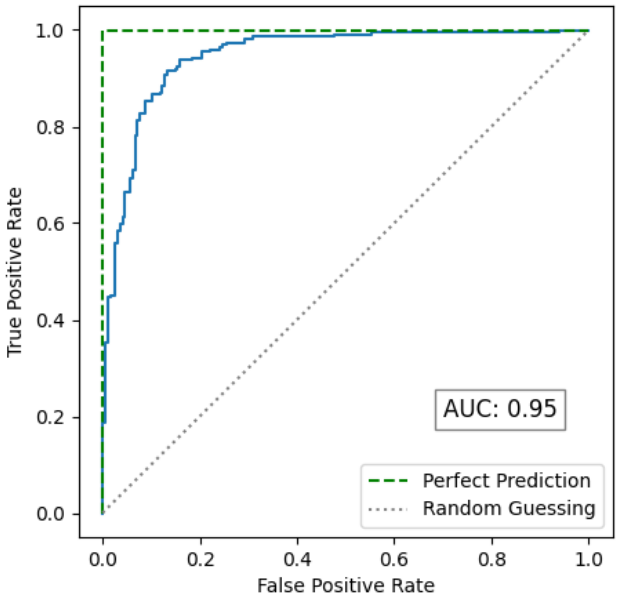


Figure 34: Plot of the ROC curve for the multilayer perceptron model and the corresponding AUC value of 0.95. The diagonal line represents a score of 0.5 (random guess) and the green line represents a score of 1 (perfect prediction).

5.3.1 External validation:

The multilayer perceptron model was also tested on external validation data. This approach allowed for a comparison of the two models not only on the training data but also in

predicting landslides in a different geographical region, using unseen data. The overall accuracy of the MLP model on the external validation dataset was 81% which indicates better performance compared to the RF model. As seen in table 9, F1 score was significantly better for the MLP model; 0.81 compared to 0.74.

Table 11: Overall accuracy, precision, recall and F1-score for the Multilayer perceptron artificial neural network model on the external validation data.

Model	Overall accuracy	Precision	Recall	F1-score
MLP	**0.81**	**0.80**	**0.81**	**0.81**

As shown in Figure 35, out of 466 cases, there were 187 true negatives, 189 true positives, 47 false positives, and 43 false negatives. Although the MLP model's predictions were less accurate compared to its performance on the training data, it demonstrated significantly better generalization on unseen data - especially in correctly identifying landslides - compared to the RF model. The MLP model yielded more true positives and fewer false negatives. Its AUC score was also higher at 0.9, compared to 0.87 for the RF model. This underscores the effectiveness of automated machine learning in selecting and optimizing the ideal algorithm for specific datasets (Feurer et al., 2022). Interestingly, Azarafza et al. (2021) reported very low AUC score of 0.51 using an MLP approach for a case study in Iran, highlighting that there is no standard method for determining the most suitable machine learning-based approach for a specific task.

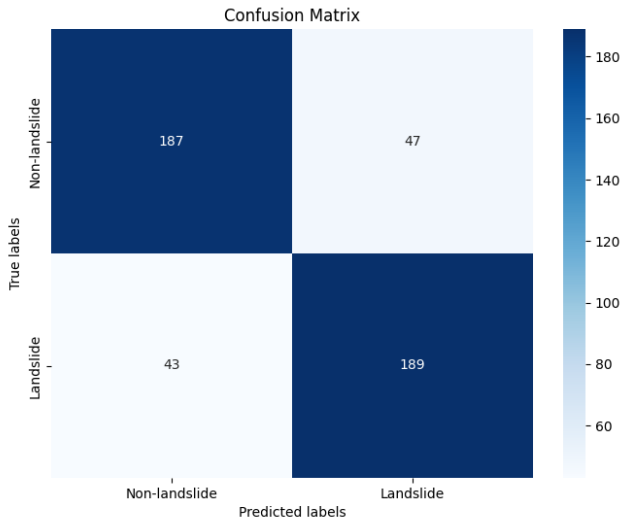


Figure 35: Confusion matrix for the multilayer perceptron model on the external validation data.

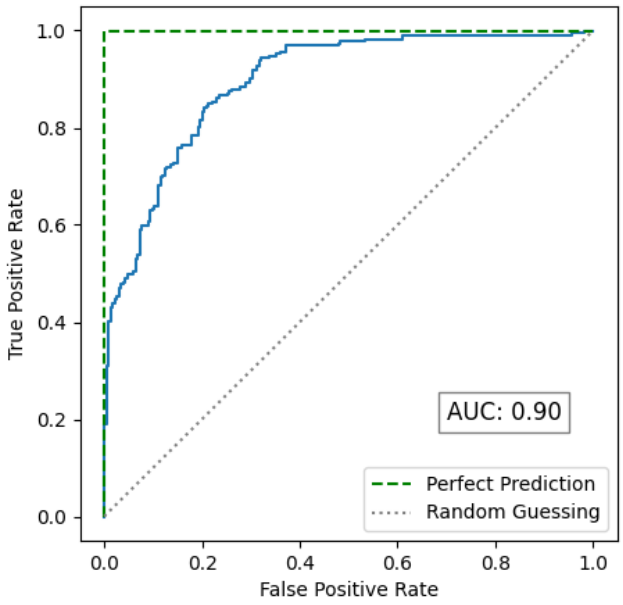


Figure 36: Plot of the ROC curve for the multilayer perceptron model and the corresponding AUC value of 0.9. The diagonal line represents a score of 0.5 (random guess) and the green line represents a score of 1 (perfect prediction).

5.3. Comparing machine learning and NGU's method for landslide susceptibility mapping.

To investigate how well machine learning can predict landslide-susceptible areas in Vestland county, this study compares the susceptibility map created by the RF model to the established nationwide susceptibility map for Norway, developed by NGU in 2014[37]. The NGU approach identifies potential release areas and delineates the downslope runout area for these points (Fischer et al., 2014). In contrast, the machine learning approach used in this study predicts landslide-susceptible areas by learning from the patterns of 'ground truth' landslide points, retrieved from the NSDB database. This prediction reflects the spatial bias in the landslide inventory; some points are release points, others are mapped within the entire landslide release zone, and many registries are located along the road network (Ganerød et al., 2023). Ganerød et al. (2023) concluded that the limitations in locational and qualitative accuracy rendered the landslide inventory unsuitable for statistical methods, leading them to develop a

[37] The nation-wide LSM for landslides in soil can be found from the NVE atlas: https://temakart.nve.no/tema/jordflomskredaktsomhet

continuous landslide detection system using satellite data. However, statistical based landslide susceptibility mapping is a widely used method to estimate where landslides are likely to occur (Reichenbach et al., 2018). Comparing results is still interesting and can help shed light on the strengths and weaknesses of both methods.

As shown in Table 12, the susceptible landslide zones from the NGU susceptibility map cover 816 km² of the total AOI of 6046 km². In comparison, the zones from the RF model, which include areas with a probability score of 0.6 or higher, cover 500 km². Within the AOI, the NGU susceptibility zones encompass 43% of the total landslide registries for the region, while the zones from the RF model contain 94% of the registries. This suggests that the RF model's susceptibility map is more conservative in assigning susceptible areas, yet captures a greater portion of areas that have actually experienced a registered landslide. However, the acknowledged limitations of the landslide inventory reduce the significance of these findings for real-world applications, but shows the potential for the ML approach.

Table 12: Coverage of susceptibility zones and percentage of landslide records intersecting with NGU LSM and RF LSM, illustrating mapping accuracy of both methods.

	Area (km^2)	Percentage of landslide records within LSM zones
NGU - LSM	816	43%
RF - LSM	500	94%

The similarities and differences between the two methods are visible in Figure 37. Figure 37(c) demonstrates how NGU's method classifies most steep valley sides as susceptible areas, while the RF model is more conservative, assigning a high probability of landslides based on the intricate relationships between the landslide registries and the influencing factors in the region. Figure 37(b) shows an example where the RF model has defined an area as highly susceptible that is not included in NGU's susceptibility map. Due to the unpredictable nature of landslide initiation, it is difficult to determine which of the two models is more accurate. However, it is interesting to observe the differences between the two methodologies and the possibilities of a machine learning approach with a landslide inventory devoid of locational and qualitative limitations.

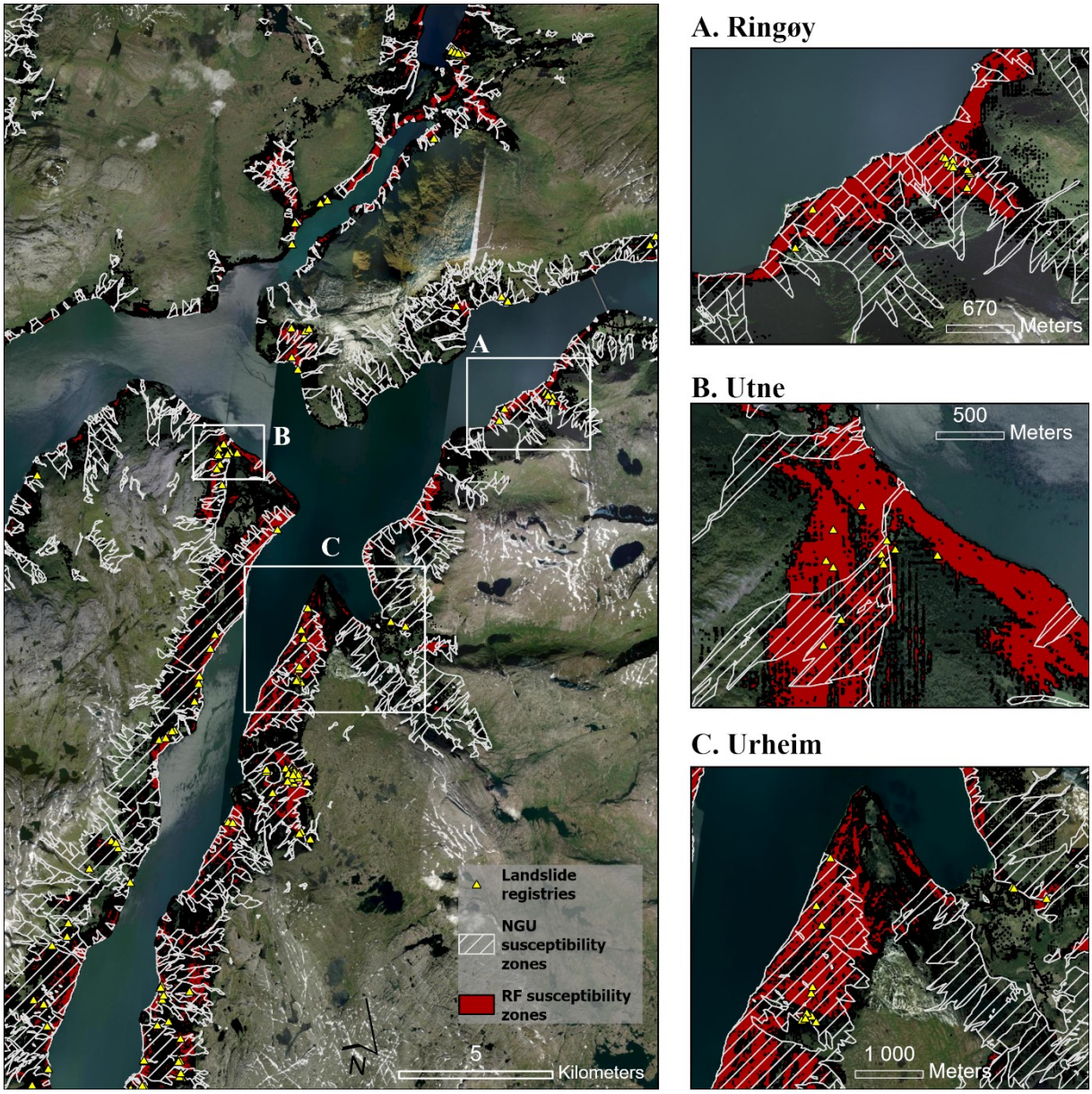


Figure 37: Comparison between the NGU LSM and the RF LSM. The zones from the RF model include all areas with a probability score of 0.6 or greater. Landslide registries are added to better illustrate the relationship between the predictions and the 'ground truth'.

5.4. Exploring the importance of the influencing factors for prediction accuracy

Exploring the statistical relationships between the landslide inventory and topographical, geological, hydrological, vegetation and meteorological factors, not only makes it possible to predict landslide susceptible areas, but also to evaluate the importance of the features for the prediction (Ageenko et al., 2022). Chen et al. (2020) concluded that random forest methods are extremely useful and efficient in selecting the important features for a classification task.

As shown in figure 38, the variable “d_roads”, representing the distance to nearest road, is most important to the model, with a score of 0,14. This score represents the normalized importance score, scaled so that they all sum up to 1 (Chen et al., 2020). Following closely is the variable “tri” (Topographic Ruggedness Index) with a score of approximately 0,11. Subsequently, variables such as “losmasse” (surficial deposits), “a_perc” (average annual precipitation) and “ndvi” (Normalized Difference Vegetation Index) exhibit similar importance scores, at around 0.06. Other variables demonstrate importance scores ranging between 0.06 and 0.03.

These results further demonstrate some of the spatial bias in the landslide inventory. Firstly, distance to roads is shown to have the most affect in predicting landslides. This is likely due to the overrepresentation of landslide registries along roads, not necessarily because proximity to roads is the primary causative factor for landslides. Furthermore, the high importance of land use class 1 can also be attributed to the spatial bias in the landslide inventory as land use class 1 represents transport infrastructure. Additionally, the surficial deposits class 130 (bare rock) is identified as the fourth most important feature (see Figure 38). According to Fischer et al. (2014), landslides occur in soils with varying surficial deposit thicknesses. The emphasis on class 130 by the Random Forest (RF) model can be attributed to the coarse mapping of these deposits by NGU, where class 130 encompasses not only bare rock but also vegetated areas and other surficial deposits.

Furthermore, terrain ruggedness is highlighted as a significant factor in landslide formation. The Topographic Ruggedness Index (TRI) describes terrain heterogeneity, reflecting elevation differences between cells (Riley et al., 1999). The Topographic Position Index (TPI) strongly correlates with slope and is thus expected to be one of the models most critical features. The "ndvi" feature, ranking third in importance, represents vegetation cover by quantifying vegetation density. Vegetation cover enhances slope stability by removing soil moisture through evapotranspiration and providing root cohesion to the soil mantle (Sidle & Ochiai, 2006, p.89). Sidle & Ochiai (2006) also emphasize negative effects of vegetation on slope stability, like surcharge of tree cover increasing the normal and downhill force components. The average Normalized Difference Vegetation Index (NDVI) for the susceptibility zones is 0.4, which is relatively high given the range of -0.19 to 0.6 for the AOI. This indicates that

landslides in soil typically occur in vegetated areas within the AOI. Vegetation cover could serve as a proxy for other environmental variables, or it might simply reflect correlations identified by the RF model that do not necessarily imply causation.

In addition, average annual precipitation and elevation is also one of the most important features for the RF model. According to Sidle and Ochiai (2006), the special patterns of rainfall are closely associated with landslide initiation with higher mountain elevations typically experiencing larger volumes of precipitation. However, total rainfall is only one rainfall attribute; short-term intensity, antecedent storm precipitation and storm duration also influence landslides (Sidle & Ochiai, 2006). For the AOI, the annual mean precipitation recorded between 1991 and 2020 ranges from 586 to 5,359 mm/year. Within zones of high landslide susceptibility (defined by a probability score of 0.6 or greater), the mean annual precipitation is 2,355 mm. Similarly, the mean annual precipitation recorded in landslide registries within the AOI is 2,267 mm. These findings suggesting some correlation between landslide initiation and high precipitation in the region.

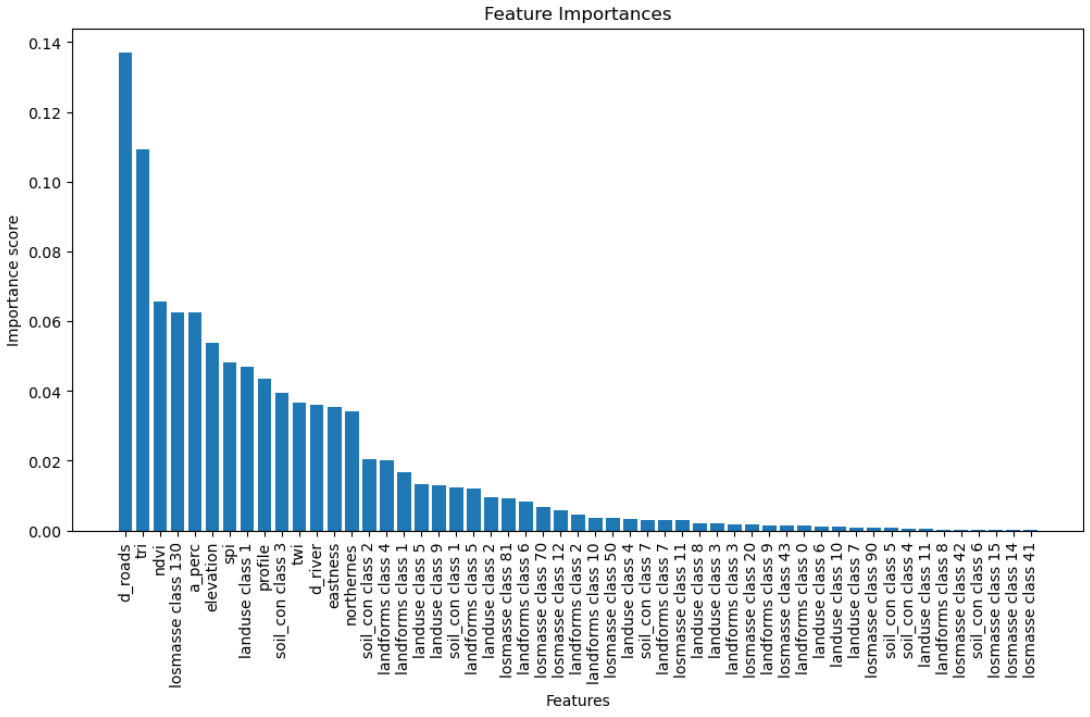


Figure 38: The feature importance of all the features used by the RF model.

5.4.1 Assessing the Performance of the random forest model with Targeted Topographic Factors

The NGU approach identifies potential release areas for landslides in soil using an index-based approach where thresholds for the three topographic factors slope, planform curvature and water contributing area are used in combination (Fischer et al., 2014). To test how well these three influencing factors can predict landslides using a machine learning approach, the random forest model was tested on the landslide inventory, but only using raster data representing slope in degrees, planar curvature and water contributing area as influencing factors[38]. This test, unlike the expert-based NGU method, did not require selecting specific threshold values and was designed to evaluate the effectiveness of the focused topographic inputs. However, when compared to the comprehensive method employed primarily in this study, which utilizes a broader set of influencing factors, the predictive power of this focused model decreased significantly, particularly for the external validation dataset. Overall accuracy for the training data dropped from 88% to 72%, and for external validation, it decreased from 81% to 59% (see table 13).

Table 13: Overall accuracy, precision, recall and F1-score for the RF model only trained on the influencing factors used by the NGU approach: slope, planform curvature and water contributing area.

	Overall accuracy	Precision	Recall	F1-score
RF - Training	**0,72**	**0,69**	**0,79**	**0,73**
RF – External Validation	**0,59**	**0,60**	**0,51**	**0,55**

Furthermore, AUC scores also saw a notable reduction; the score for the training dataset fell from 0.96 to 0.8, and for the external validation area, it plummeted from 0.9 to 0.62, as detailed in Figure 39. This outcome is somewhat surprising, considering that these three influencing factors were specifically utilized by NGU to identify starting points for landslides in soil. One might expect more consistent results between the training area and the external

[38] To represent water contributing area, a flow accumulation raster was generated with a similar method described in section 4.2.2.2

validation dataset. However, the NGU method employs these factors in combination, using different thresholds values for different regions in Norway; the findings suggest that when used separately, these factors do not predict landslides effectively. This underscores the complexity of interactions among influencing factors in landslide prediction.

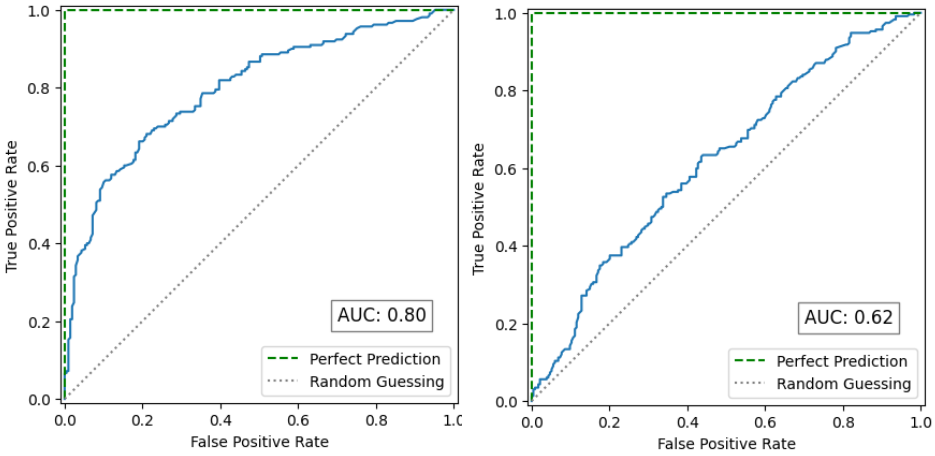


Figure 39: Plot of the ROC curve for the RF model, trained only on the influencing factors slope, planar curvature and water contributing area.

A visual representation of the comparative results is shown in figure 40. The RF model trained on only slope, planform curvature and water contributing area covers a much larger and arbitrary area, likely because of underfitting. Underfitting occurs because the model was trained on a limited set of factors, leading it to not capture the complex patterns in the data, hence predicting larger areas as susceptible due to its inability to differentiate effectively (Cunningham & Delany., 2021).

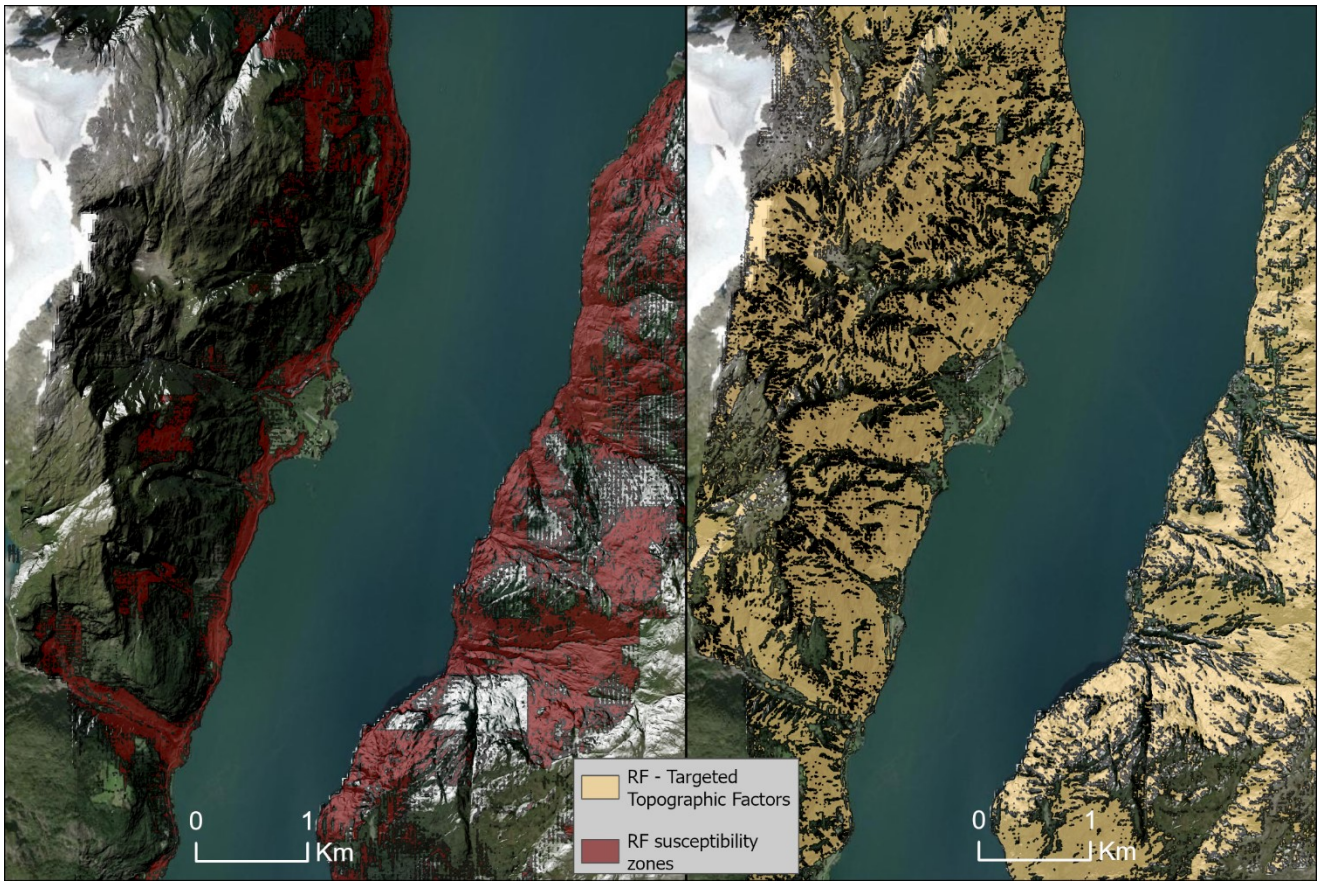


Figure 40: A snippet of the AOI, showing a comparison between LSM derived from the RF model trained on the targeted topographic factors used by NGU, and the comprehensive method employed primarily in this thesis. Both LSM include all areas with a probability score greater or equal to 0.6.

5.4.2 Exploring spatial bias in the landslide inventory.

To further demonstrate the spatial bias in the landslide inventory, the Random Forest model was fitted to both the training and external validation datasets with specific influencing factors removed. These factors include "d_roads," which measures proximity to the road network, and "landuse," which include a land use class representing transport infrastructure (see section 5.4). The performance metrics for this test is presented in table 14.

Table 14: Overall accuracy, precision, recall and F1-score for the RF model where the influencing factors "d_roads" and "landuse" are removed from training.

	Overall accuracy	Precision	Recall	F1-score
RF - Training	**0,82**	**0,80**	**0,86**	**0,83**
RF – External Validation	**0,73**	**0,80**	**0,69**	**0,74**

Compared to the primary method, removing the features "d_roads" and "landuse" negatively affect the performance of prediction. For the training data, the overall accuracy dropped from 88% to 82%. Recall dropped from 0.9 to 0.86 and the AUC score dropped from 0.96 to 0.92. These changes are not massive however. An accuracy decreases of 6%, while notable, does not suggest a catastrophic degradation in model performance. It points to some dependency on these features, but the model still retains a good level of accuracy without them. Similarly, for the external validation, the overall accuracy drops from 76% to 73%. Recall is identical at 0.69, but the AUC score drops from 0.87 to 0.81 (see figure 41). The ROC curve, presented in figure 41, shows that even at low threshold values, further increase in recall (true positive rate) requires an increase in FPR. Nevertheless, the AUC score of 0.81 still indicate a good performing model that distinguishes well between the two classes. An AUC value greater than 0.7 is generally considered to indicate good accuracy (Ado et al., 2022). These findings, while indicating some spatial bias, thus underscores the robustness of the model when relying on other geographical and environmental factors.

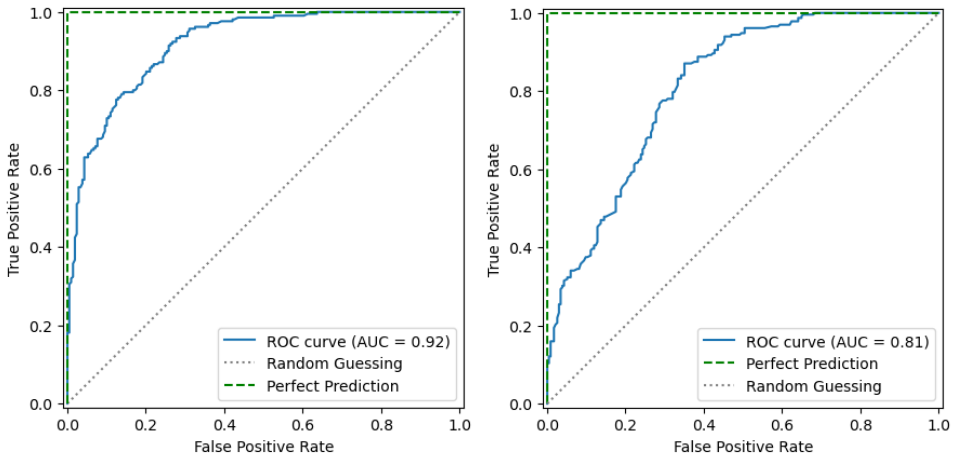


Figure 41: Plot of the ROC curve for the RF model where the influencing factors "d_roads" and "landuse" are removed from training.

Chapter 6

Conclusion and future work

In this thesis, the feasibility of using machine learning to predict areas susceptible to landslides in soil for Vestland county in Norway was investigated. An area covering 6 478 km^2 was used for training/prediction. Another area, covering 1 798 km^2 was used for external validation. A dataset of historic landslide registries, based on the NSDB database, together with event inventories, were used as the landslide inventory. Multiple conditioning factors was used as the explanatory variables for the classification task. Feature selection was conducted to understand the relationship between the influencing factors. A correlation matrix and Variance Inflation Factor was used to ultimately remove the variables “roughness”, “average annual temperature”, “slope” and “topographic position index”, because of multicollinearity. In addition, Random Forest was used to evaluate the feature importance; checking for irrelevant features, ultimately removing the variable “forest loss” as it had very little significance for the model.

Two different machine learning approaches was explored; The established RF algorithm and an AutoML algorithm Auto-Sklearn. For evaluation of the predictive accuracy of the models, accuracy, precision, recall, F1-score and ROC_AUC was used. An independent dataset was used to assess the model’s generalizability on unseen data. The trained random forest classifier was fitted to the entire raster stack of influencing factors to produce a landslide susceptibility map for the AOI.

The use of the two ML algorithms showed promising results for landslide susceptibility mapping in the region of Vestland county. The two classifiers (RF and Auto-Sklearn) trained in this study and then tested on the test data both showed a high overall accuracy of 88%. The RF model showed an AUC score of 0.96. Auto-Sklearn, whose best performing algorithm was a multilayer perceptron feedforward artificial neural network, had an AUC score of 0.95, indicating that both models distinguish very well between the landslide and non-landslide classes. For the test data, the multilayer perceptron model predicted more actual landslides and had fewer false positives, compared to the RF model.

To evaluate the generalizability of the ML method, the models were tested on the external validation area. For the RF model, the overall accuracy dropped to 76% and the AUC dropped to 0.87. For the multilayer perceptron model, the overall accuracy dropped to 81% and the AUC dropped to 0.9. This underscores the effectiveness of AutoML in selecting and optimizing the ideal algorithm for specific datasets. The AutoML method conclusively outperforms the established random forest model in this case. Furthermore, even though the predictive accuracy drops for both models when the classifiers are exposed to data in a different geographical region, the retention of good predictive accuracy indicates a robust model that can predict landslides across different regions.

The RF model allows for a variable importance analysis, where it was found that the most significant variable related to landslides was distance to roads, followed by Topographic Ruggedness Index, surficial deposit class 130 (bare rock/thin turf cover), average annual precipitation and NDVI. These results demonstrate undeniable spatial bias in the landslide inventory with an overrepresentation of landslide registries along roads, rather than proximity to roads being a primary causative factor for landslides in soil. This spatial bias was further explored by training the RF model without road-associated variables, which resulted in a slight performance degradation. However, the models still performed well, especially on the test data. Furthermore, the RF model trained solely with slope, planform curvature, and water contributing area underperformed significantly in predicting landslides, underscoring that incorporating a broader selection of diverse and relevant influencing factors is crucial for predictive performance. The variable importance should be assessed in the context of the given case study and is not necessarily generalizable to other regions as anthropogenic activities, topography, and geological characteristics are not identical.

The susceptibility map developed in this thesis shows strong potential for accurately predicting landslide-prone areas, particularly within the regions where the machine learning models were trained. This could make it a valuable tool for municipalities needing detailed, region-specific landslide susceptibility assessments, offering a more detailed and area-specific result than the current nationwide susceptibility map available for Norway. However, the usefulness of the LSM is limited by the spatial biases in the landslide data. Better landslide inventory is thus of paramount importance for better statistical landslide prediction.

6.1 Future Work

To enhance the generalizability and precision of machine learning models for landslide susceptibility mapping across Norway, future studies should consider employing a more diverse set of external validation areas. This approach would not only help in understanding the machine learning models' adaptability nationwide but also illuminate the regional variations in geoenvironmental factors influencing landslide formation.

Future studies should consider applying the methodology outlined in this thesis using slope units as the mapping unit, instead of the grid cells utilized in this study. Given that landslides predominantly occur on hillslopes, the use of slope units could theoretically enhance the precision and relevance of the LSM.

Finally, recent advancements in machine learning for continuous landslide detection present an opportunity to develop a more realistic and comprehensive landslide inventory. This updated database could serve as a refined input for more accurate LSM, ensuring that the models are working with the best available data. Furthermore, if improved data can be collected specifically for landslide initiation areas, machine learning models could be deployed to pinpoint these high-risk areas. Coupling this with modeling techniques for the runout zones could make for an integrated approach that could significantly advance the capabilities of LSM for Norway.

References

Ado, M., Amitab, K., Maji, A. K., Jasińska, E., Gono, R., Leonowicz, Z., & Jasiński, M. (2022). Landslide susceptibility mapping using machine learning: A literature survey. *Remote Sensing*, *14*(13), 3029. https://doi.org/https://doi.org/10.3390/rs14133029

Ageenko, A., Hansen, L. C., Lyng, K. L., Bodum, L., & Arsanjani, J. J. (2022). Landslide susceptibility mapping using machine learning: A Danish case study. *ISPRS International Journal of Geo-Information*, *11*(6), 324. https://doi.org/https://doi.org/10.3390/ijgi11060324

Alcántara-Ayala, I., Cui, P., & Pasuto, A. (2022). Disaster risk reduction in mountain areas: a research overview. *Journal of mountain science*, *19*(6), 1487-1494.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, *9*(11), 3975-3991. https://doi.org/https://doi.org/10.5194/gmd-9-3975-2016, 2016.

Arnone, E., Francipane, A., Scarbaci, A., Puglisi, C., & Noto, L. V. (2016). Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. *Environmental Modelling & Software*, *84*, 467-481. https://doi.org/10.1016/j.envsoft.2016.07.016

Aurélien, G. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc.

Azarafza, M., Azarafza, M., Akgün, H., Atkinson, P. M., & Derakhshani, R. (2021). Deep learning-based landslide susceptibility mapping. *Scientific reports*, *11*(1), 24112. https://doi.org/https://doi.org/10.1038/s41598-021-03585-1

Ba, Q., Chen, Y., Deng, S., Yang, J., & Li, H. (2018). A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment. *Earth Science Informatics*, *11*, 373-388. https://doi.org/https://doi.org/10.1007/s12145-018-0335-9

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological sciences journal*, *24*(1), 43-69. https://doi.org/https://doi.org/10.1080/02626667909491834

Blahut, J., Horton, P., Sterlacchini, S., & Jaboyedoff, M. (2010). Debris flow hazard modelling on medium scale: Valtellina di Tirano, Italy. *Natural Hazards and Earth System Sciences*, *10*(11), 2379-2390. https://doi.org/https://doi.org/10.5194/nhess-10-2379-2010, 2010.

Brabb, E. E. (1985). Innovative approaches to landslide hazard and risk mapping. International landslide symposium proceedings, Toronto, Canada,

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32. https://doi.org/https://doi.org/10.1023/A:1010933404324

Bruland, O., Frengstad, B., Eiksund, G., Felix, M., Fredin, O., Hermanns, R., Holmøy, K., Osmundsen, T., Per, Rød, K., Jan, & Sægrov, S. (2023). *Innspel til Flaum - og Skredmedldinga*. https://www.regjeringen.no/no/aktuelt/skriftlige-innspill-til-stortingsmeldingen-for-flom-og-skred/id3015357/

Büyükkeçecİ, M., & Okur, M. C. (2022). A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning. *Gazi University Journal of Science*, *36*(4), 1506-1520. https://doi.org/https://doi.org/10.35378/gujs.993763

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., & Reichenbach, P. (1991). GIS techniques and statistical models in evaluating landslide hazard. *Earth surface processes and landforms*, *16*(5), 427-445. https://doi.org/https://doi.org/10.1002/esp.3290160505

Carrara, A., Catalano, E., Reali, C., & Osso, I. (1978). Digital terrain analysis for land evaluation. http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCALGEODEBRGM8020490289

Carrara, A., Pugliese Carratelli, E., & Merenda, L. (1977). Computer-based data bank and statistical analysis of slope instability phenomena. *Zeitschrift für Geomorphologie*, 187-222. https://doi.org/ 10.1127/zfg/21/1977/187

Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*, *10*(8), Article 1283. https://doi.org/10.3390/math10081283

Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1), 52. https://doi.org/https://doi.org/10.1186/s40537-020-00327-4

Chen, T., Zhong, Z., Niu, R., Liu, T., & Chen, S. (2020). Mapping landslide susceptibility based on deep belief network. *Geomatics and Information Science of Wuhan University*, *45*(11), 1809-1817. https://doi.org/10.13203/j.whugis20190144

Chung, C.-J. F., & Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, *30*, 451-472. https://doi.org/https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b

Cruden, D. M. (1993). Cruden, DM, Varnes, DJ, 1996, Landslide Types and Processes, Transportation Research Board, US National Academy of Sciences, Special Report, 247: 36-75. *Landslides Eng. Pract*, *24*, 20-47.

Cruden, D. M., & Varnes, D. J. (1996). Landslides: investigation and mitigation. Chapter 3-Landslide types and processes. *Transportation research board special report*(247).

Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning. In M. Cord & P. Cunningham (Eds.), *Machine learning techniques for multimedia: case studies on*

organization and retrieval (pp. 21-49). Springer, Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-540-75171-7_2

Cunningham, P., & Delany, S. J. (2021). Underestimation Bias and Underfitting in Machine Learning. In (pp. 20-31). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-73959-1_2

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., & Finke, P. (2013). Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, *186*, 39-49. https://doi.org/https://doi.org/10.1016/j.geomorph.2012.12.015

Devoli, G., Jarsve, K., Mongstad, H., Sandboe, K., & Jensen, O. (2020). Kontroll av registrerte skredhendelser og tildeling av kvalitetsnivå for utglidning, jordskred, flomskred og sørpeskred. *NVE rapport*, *31*, 2020.

Dinov, I. D. (2023). Supervised Classification. In (pp. 275-340). Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-031-17483-4_5

Esri. (n.d-a). *Surface Parameters (Spatial Analyst)*. Esri. https://pro.arcgis.com/en/pro-app/3.1/tool-reference/spatial-analyst/surface-parameters.htm

Esri. (n.d-b). *Flow Accumulation (Spatial Analyst)*. Esri. https://pro.arcgis.com/en/pro-app/3.1/tool-reference/spatial-analyst/flow-accumulation.htm

Esri. (n.d-c). *Focal Statistics (Spatial Analyst)*. Esri. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/focal-statistics.htm

Esri. (n.d-d). *Distance Accumulation (Spatial Analyst)*. Esri. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/distance-accumulation.htm

Esri. (n.d-e). *Extract Multi Values to Points (Spatial Analyst)*. Esri. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/extract-multi-values-to-points.htm

Esri. (n.d-f). *Curvature function*. Esri. https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/curvature-function.htm

Fabbri, A. G., Chung, C.-J. F., Cendrero, A., & Remondo, J. (2003). Is prediction of future landslides possible with a GIS? *Natural Hazards*, *30*, 487-503. https://doi.org/https://doi.org/10.1023/B:NHAZ.0000007282.62071.75

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering geology*, *102*(3-4), 85-98. https://doi.org/https://doi.org/10.1016/j.enggeo.2008.03.022

Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M., & Hutter, F. (2022). Auto-sklearn 2.0: Hands-free automl via meta-learning. arXiv 2020. *arXiv preprint arXiv:2007.04074*.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. (2015). Efficient and robust automated machine learning. *Advances in neural information processing systems*, *28*.

Fischer, L., Rubensdotter, L., & Stalsberg, K. (2014). *Aktsomhetskart jord - og flomskred: Metodeutvikling og landsdekkende modellering* (2014.019).

Furseth, A. (2006). *Skredulykker i Norge*. Tun Forlag.

Ganerød, A. J., Lindsay, E., Fredin, O., Myrvoll, T.-A., Nordal, S., & Rød, J. K. (2023). Globally vs. Locally Trained Machine Learning Models for Landslide Detection: A Case Study of a Glacial Landscape. *Remote Sensing*, *15*(4), 895. https://doi.org/https://doi.org/10.3390/rs15040895

Gao, B., & Wang, X. (2019). Risk zoning of landslide based on SINMAP model in Yan'an City. *Bull. Soil Water Conserv*, *39*, 211-216.

García-Rodríguez, M. J., Malpica, J., Benito, B., & Díaz, M. (2008). Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. *Geomorphology*, *95*(3-4), 172-191. https://doi.org/https://doi.org/10.1016/j.geomorph.2007.06.001

Geiger, R. S., Cope, D., Ip, J., Lotosh, M., Shah, A., Weng, J., & Tang, R. (2021). "Garbage in, garbage out" revisited: What do machine learning application papers report about human-labeled training data? *Quantitative Science Studies*, *2*(3), Article 00144. https://doi.org/10.1162/qss_a_00144

Geovekst. (2016). *SOSI Del 3 Produktspesifikasjon for Felles KartdataBase (FKB)*. https://www.kartverket.no/geodataarbeid/geovekst/fkb-produktspesifikasjoner

Goodchild, M. (2010). Twenty years of progress: GIScience in 2010. *Journal of spatial information science*(1), 3-20.

Google. (n.d). *Google Earth Engine*. Google. https://www.google.com/earth/education/tools/google-earth-engine/

Graziella, D., Jarsve;, K. T., Mongstad;, H. H., Sandboe;, K. S., & Jensen, O. A. (2020). *Kontroll av registrerte skredhendelser og tildeling av kvalitetsnivå for utglidning, jordskred, flomskred og sørpeskred*.

Guo, Z., Tian, B., Li, G., Huang, D., Zeng, T., He, J., & Song, D. (2023). Landslide susceptibility mapping in the Loess Plateau of northwest China using three data-driven techniques-a case study from middle Yellow River catchment. *Frontiers in Earth Science*, *10*, 1033085. https://doi.org/https://doi.org/10.3389/feart.2022.1033085

Guzzetti, F. (2006). *Landslide hazard and risk assessment* Universitäts-und Landesbibliothek Bonn]. https://nbn-resolving.org/urn:nbn:de:hbz:5N-08175

Hamel, S., Garel, M., Festa-Bianchet, M., Gaillard, J. M., & Côté, S. D. (2009). Spring Normalized Difference Vegetation Index (NDVI) predicts annual variation in timing of peak faecal crude protein in mountain ungulates. *Journal of Applied Ecology*, *46*(3), 582-589. https://doi.org/https://doi.org/10.1111/j.1365-2664.2009.01643.x

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., & Loveland, T. R. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850-853. https://doi.org/DOI: 10.1126/science.1244693

Himan, S., Saeed, K., Baharin, A., & Mazlan, H. (2014). Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. *Catena*, *115*, 55-70. https://doi.org/10.1016/j.catena.2013.11.014

Holden, J. (2017). *An Introduction to Physical Geography and the Environment*. Pearson Education Limited. https://books.google.no/books?id=32f7MAAACAAJ

Hong, H. Y., Liu, J. Z., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Zhu, A. X., Chen, W., & Bin Ahmad, B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, *163*, 399-413. https://doi.org/10.1016/j.catena.2018.01.005

Huang, F., Cao, Z., Guo, J., Jiang, S.-H., Li, S., & Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena*, *191*, 104580. https://doi.org/https://doi.org/10.1016/j.catena.2020.104580

Huang, Y., & Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena*, *165*, 520-529. https://doi.org/https://doi.org/10.1016/j.catena.2018.03.003

Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, *11*(2), 167-194. https://doi.org/10.1007/s10346-013-0436-y

I. Hanssen-Bauer, E. J. F., I. Haddeland, H. Hisdal, S. Mayer, A. Nesje, J.E.Ø. Nilsen, S. Sandven,, & A.B. Sandø, A. S. o. B. Å. (2015). *Klima i Norge 2100 -*

kunnskapsgrunnlag for klimatilpasning oppdatert i 2015. N. Klimaservicesenter. https://www.miljodirektoratet.no/globalassets/publikasjoner/M741/M741.pdf

Kamp, U., Growley, B. J., Khattak, G. A., & Owen, L. A. (2008). GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. *Geomorphology*, *101*(4), 631-642. https://doi.org/https://doi.org/10.1016/j.geomorph.2008.03.003

Kavzoglu, T., Sahin, E. K., & Colkesen, I. (2014). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, *11*(3), 425-439. https://doi.org/https://doi.org/10.1007/s10346-013-0391-7

Kavzoglu, T., & Teke, A. (2022). Predictive Performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arabian Journal for Science and Engineering*, *47*(6), 7367-7385. https://doi.org/https://doi.org/10.1007/s13369-022-06560-8

Kreuzer, T. M., & Damm, B. (2020). Automated digital data acquisition for landslide inventories. *Landslides*, *17*(9), 2205-2215. https://doi.org/https://doi.org/10.1007/s10346-020-01431-5

Langley, P. (1988). Machine learning as an experimental science. *Machine learning*, *3*(1), 5-8.

Lin, L., Lin, Q., & Wang, Y. (2017). Landslide susceptibility mapping on a global scale using the method of logistic regression. *Natural Hazards and Earth System Sciences*, *17*(8), 1411-1424. https://doi.org/https://doi.org/10.5194/nhess-17-1411-2017

Liu, S., Wang, L., Zhang, W., He, Y., & Pijush, S. (2023). A comprehensive review of machine learning-based methods in landslide susceptibility mapping. *Geological Journal*. https://doi.org/https://doi.org/10.1002/gj.4666

Mandal, J. K., & Bhattacharya, D. (2019). Supervised Classification Algorithms in Machine Learning: A Survey and Review. In (Vol. 937, pp. 99-111). Singapore: Springer Singapore Pte. Limited. https://doi.org/10.1007/978-981-13-7403-6_11

Mandal, J. K., & Bhattacharya, D. (2020). *Emerging technology in modelling and graphics*. Springer.

Mathew, J., Jha, V., & Rawat, G. (2009). Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides*, *6*, 17-26. https://doi.org/https://doi.org/10.1007/s10346-008-0138-z

Meld.St.15. (2012). *Hvordan leve med farene - om flom og skred, white paper from the Ministry of Petroleum and Energy*. https://www.regjeringen.no/no/dokumenter/meld-st-15-20112012/id676526/?ch=1

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Moore, I., O'loughlin, E., & Burch, G. (1988). A contour-based topographic model for hydrological and ecological applications. *Earth surface processes and landforms*, *13*(4), 305-320. https://doi.org/https://doi.org/10.1002/esp.3290130404

Nefeslioglu, H. A., Gokceoglu, C., & Sonmez, H. (2008). An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering geology*, *97*(3-4), 171-191. https://doi.org/https://doi.org/10.1016/j.enggeo.2008.01.004

Neuhäuser, B., & Terhorst, B. (2007). Landslide susceptibility assessment using “weights-of-evidence” applied to a study area at the Jurassic escarpment (SW-Germany). *Geomorphology*, *86*(1-2), 12-24. https://doi.org/https://doi.org/10.1016/j.geomorph.2006.08.002

NGU. (n.d). *Kvartærgeologisk kartlegging*. NGU. Retrieved 16.04 from https://www.ngu.no/geologisk-kartlegging/kvartaergeologisk-kartlegging

NIBIO. (2020). *Grunnforhold*. NIBIO. Retrieved 14.05.2024 from https://www.nibio.no/tema/jord/arealressurser/arealressurskart-ar5/grunnforhold

Nordrum, R., Jensen, O. A., Lindgaard, A., & Nilsen, J. E. (2020). Skog som vern mot naturfarer: kunnskapssammenstilling og tilpasning til Natur i Norge (NiN).

NVE. (2014). *Flaum- og skredfare i arealplanar* [guideline]. http://www.nve.no/arealplan

NVE. (2018). Presentasjonsregler: Skredhendelser. In. Geonorge.no: The Norwegian Water Resources and Energy Directorate.

NVE. (2021). *Aktsomhetskart for jord - og flomskred*. The Norwegian Water Resources and Energy Directorate. Retrieved 11.12.2023 from https://www.nve.no/naturfare/utredning-av-naturfare/om-kart-og-kartlegging-av-naturfare/om-kartlegging-av-skredfare-i-bratt-terreng/aktsomhetskart-for-jord-og-flomskred/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

Pham, Q. B., Achour, Y., Ali, S. A., Parvin, F., Vojtek, M., Vojteková, J., Al-Ansari, N., Achu, A. L., Costache, R., Khedher, K. M., & Anh, D. T. (2021). A comparison among fuzzy multi-criteria decision making, bivariate, multivariate and machine learning models in landslide susceptibility mapping. *Geomatics Natural Hazards & Risk*, *12*(1), 1741-1777. https://doi.org/10.1080/19475705.2021.1944330

Promper, C., Puissant, A., Malet, J.-P., & Glade, T. (2014). Analysis of land cover changes in the past and the future as contribution to landslide risk scenarios. *Applied Geography*, *53*, 11-19. https://doi.org/https://doi.org/10.1016/j.apgeog.2014.05.020

Pullarello, S., Jose. (2021). *Liten utgliding - starting points mapped on the detailed quaternary map at NGU* [Shapefile].

Qasimi, A. B., Isazade, V., Kaplan, G., & Nadry, Z. (2022). Spatiotemporal and multi-sensor analysis of surface temperature, NDVI, and precipitation using google earth engine cloud computing platform. *Russian Journal of Earth Sciences*, *22*(6), Article Es6002. https://doi.org/10.2205/2022es000812

Ramiz, M., Siddiqui, M. A., Salman, M. S., Siddiqui, L., Tahir, M., Naqvi, H. R., & Shakeel, A. (2023). Landslide susceptibility mapping along Rishikesh-Badrinath national highway (Uttarakhand) by applying multi-criteria decision-making (MCDM) approach. *Environmental Earth Sciences*, *82*(24), Article 591. https://doi.org/10.1007/s12665-023-11268-5

Regmi, A. D., Devkota, K. C., Yoshida, K., Pradhan, B., Pourghasemi, H. R., Kumamoto, T., & Akgun, A. (2014). Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal of Geosciences*, *7*, 725-742. https://doi.org/https://doi.org/10.1007/s12517-012-0807-z

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-science reviews*, *180*, 60-91. https://doi.org/https://doi.org/10.1016/j.earscirev.2018.03.001

Renza, D., Cárdenas, E. A., Jaramillo, C. M., Weber, S. S., & Martinez, E. (2021). Landslide Susceptibility Model by Means of Remote Sensing Images and AutoML. In J. C. Figueroa-García, Y. Díaz-Gutierrez, E. E. Gaona-García, & A. D. Orjuela-Cañón, *Applied Computer Sciences in Engineering* Cham.

Rickenmann, D., & Zimmermann, M. (1993). The 1987 debris flows in Switzerland: documentation and analysis. *Geomorphology*, *8*(2-3), 175-189.

Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). Index that quantifies topographic heterogeneity. *intermountain Journal of sciences*, *5*(1-4), 23-27.

Ruther, D. C., Hefre, H., & Rubensdotter, B. L. E. F. (2022). Extreme precipitation-induced landslide event on 30th July 2019 in Jølster, western Norway.

Rød, J. K., Berthling, I., Lein, H., Lujala, P., Vatne, G., & Bye, L. M. (2012). Integrated vulnerability mapping for wards in Mid-Norway. *Local Environment*, *17*(6-7), 695-716. https://doi.org/https://doi.org/10.1080/13549839.2012.685879

Sahin, E. K., Colkesen, I., & Kavzoglu, T. (2020). A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto International*, *35*(4), 341-363. https://doi.org/https://doi.org/10.1080/10106049.2018.1516248

Saleem, N., Huq, M. E., Twumasi, N. Y. D., Javed, A., & Sajjad, A. (2019). Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: a review. *ISPRS International Journal of Geo-Information*, *8*(12), 545. https://doi.org/https://doi.org/10.3390/ijgi8120545

Sidle, R. C., & Ochiai, H. (2006). *Landslides : processes, prediction, and land use* (Vol. 18). American Geophysical Union.

Skempton, A., & Hutchinson, J. (1969). Stability of natural slopes and embankment foundations. Soil Mech & Fdn Eng Conf Proc/Mexico/,

Smith, H. G., Spiekermann, R., Betts, H., & Neverman, A. J. (2021). Comparing methods of landslide data acquisition and susceptibility modelling: Examples from New Zealand. *Geomorphology*, *381*, 107660. https://doi.org/https://doi.org/10.1016/j.geomorph.2021.107660

Stewart, T. J. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *Omega*, *20*(5-6), 569-586. https://doi.org/https://doi.org/10.1016/0305-0483(92)90003-P

Stott, P. (2016). How climate change affects extreme weather events. *Science*, *352*(6293), 1517-1518. https://doi.org/doi:10.1126/science.aaf7271

Sujatha, E. R., & Rajamanickam, V. (2011). Landslide susceptibility mapping of Tevankarai Ar sub-watershed, Kodaikkanal taluk, India, using weighted similar choice fuzzy model. *Natural Hazards*, *59*(1), 401-425. https://doi.org/https://doi.org/10.1007/s11069-011-9763-2

Sun, D., Wen, H., Wang, D., & Xu, J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, *362*, 107201. https://doi.org/https://doi.org/10.1016/j.geomorph.2020.107201

Tehrani, F. S., Calvello, M., Liu, Z., Zhang, L., & Lacasse, S. (2022). Machine learning and landslide studies: recent advances and applications. *Natural Hazards*, *114*(2), 1197-1245. https://doi.org/https://doi.org/10.1007/s11069-022-05423-7

Tveito, O. E. (2021). *Norwegian standard climate normals 1991-2020- the methodological approach* [METreport](05/2021). (Norwegian Meteorological Institute, Issue.

Umar, Z., Pradhan, B., Ahmad, A., Jebur, M. N., & Tehrany, M. S. (2014). Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *Catena*, *118*, 124-135. https://doi.org/https://doi.org/10.1016/j.catena.2014.02.005

Varnes, D. J. (1978). Slope movement types and processes. *Special report*, *176*, 11-33.

Weiss, A. D. (2001). Topographic Positions and Landforms Analysis (poster). In E. I. User (Ed.), *ESRI Conference*. San Diego, CA.

Xie, M. (2014). *Landslide Hazard Assessment Using GIS* (Vol. 1). Science Press and Alpha Science International Ltd.

Yang, C., Liu, L.-L., Huang, F., Huang, L., & Wang, X.-M. (2023). Machine learning-based landslide susceptibility assessment with optimized ratio of landslide to non-landslide samples. *Gondwana Research*, *123*, 198-216. https://doi.org/https://doi.org/10.1016/j.gr.2022.05.012

Yao, J., Yao, X., Zhao, Z., & Liu, X. (2023). Performance comparison of landslide susceptibility mapping under multiple machine-learning based models considering InSAR deformation: A case study of the upper Jinsha River. *Geomatics, Natural Hazards and Risk*, *14*(1), 2212833. https://doi.org/https://doi.org/10.1080/19475705.2023.2212833

Yong, C., Jinlong, D., Fei, G., Bin, T., Tao, Z., Hao, F., Li, W., & Qinghua, Z. (2022). Review of landslide susceptibility assessment based on knowledge mapping. *Stochastic Environmental Research and Risk Assessment*, *36*(9), 2399-2417. https://doi.org/https://doi.org/10.1007/s00477-021-02165-z

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *The Journal of Machine Learning Research*, *5*, 1205-1224.

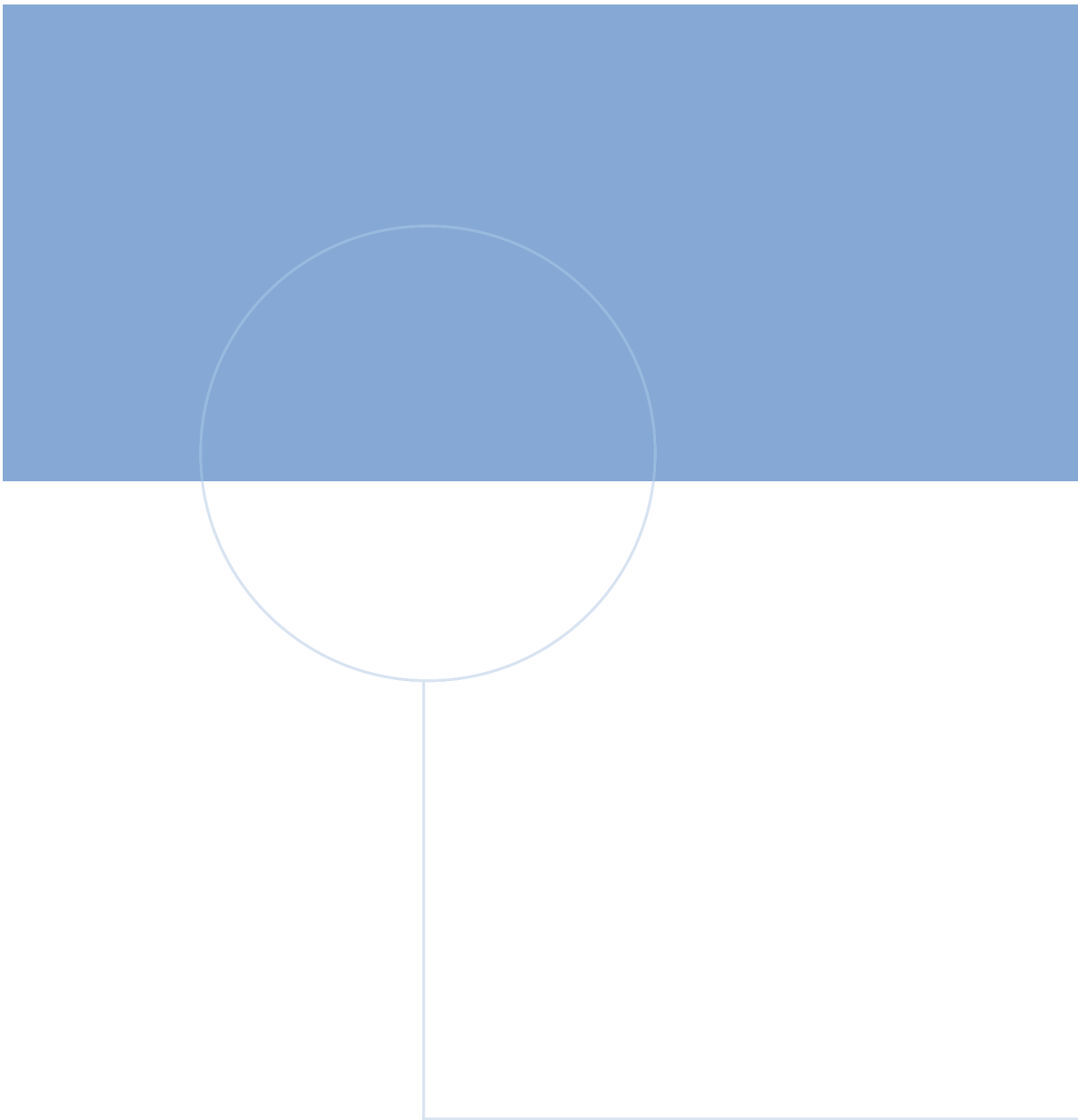
Zhao, X., Zhao, Z., Huang, F., Huang, J., Yang, Z., Chen, Q., Zhou, D., Fang, L., Ye, X., & Chao, J. (2023). Application of environmental variables in statistically-based landslide susceptibility mapping: A review [Review]. *Frontiers in Earth Science*, *11*. https://doi.org/10.3389/feart.2023.1147427

Øydvin., E. K., Graziella Devoli, Terje H. Bargel, Toril Wiig, Andrea Taurisano, Hallvard Berg, Olianne Eikenæs, Einar Lyche, Tharan Fergus, Marte R.Kvakland, Jaran

Wasrud, Tonje E. Helle, Kjartan Orvedal, Ivar O. Peereboom, Øivind B. Andersen, Reginald Hermanns, Jan Høst, Louise Hansen, Halvor Bunkholt, . . . Stalsberg, K. (2011). *Plan for skredfarekartlegging - Status og prioriteringer innen oversiktkartlegging*

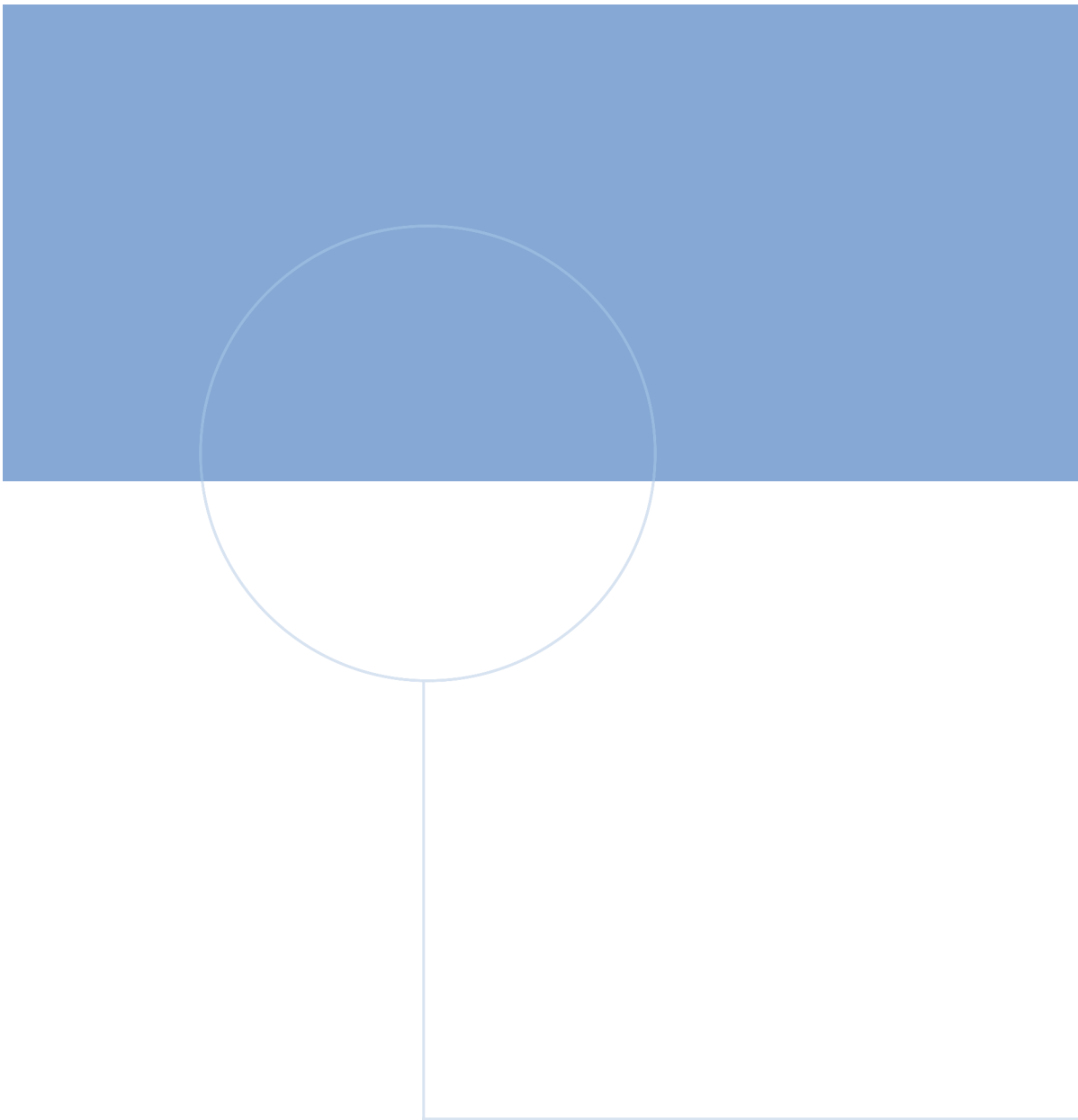
og detaljert skredfarekartlegging i NVEs regi (14/2011). N. v.-o. energidirektorat.

Åberg, S., Amanda. (2021). *Predicting a catastrophe: A geographical and statistical investigation of a precipitation-triggered multi-debris flow event in Jølster, western Norway*.



NTNU

Norwegian University of
Science and Technology



NTNU

Norwegian University of
Science and Technology