Student thesis series INES nr 605

A Comparative Analysis of Decision Tree Models in Identifying Landslide Susceptibility and Type Classification.

Levi Zuiverloon

2023

Department of

Physical Geography and Ecosystem Science

Lund University

Sölvegatan 12

S-223 62 Lund

Levi Zuiverloon (2023).
A Comparative Analysis of Decision Tree Models in Identifying Landslide Susceptibility and Type Classification.
Jämförande analys av två modeller baserade på beslutsträd och deras förmåga att identifiera risk och typ av jordskred.
Bachelor's degree thesis, 15 credits in **Physical Geography and Ecosystem Analysis**
Department of Physical Geography and Ecosystem Science, Lund University

Level: Bachelor of Science (BSc)

Course duration: *March* 2023 until *June* 2023

Disclaimer

A Comparative Analysis of Decision Tree Models in Identifying Landslide Susceptibility and Type Classification.

Levi Zuiverloon

Bachelor thesis, 15 credits, in **Physical Geography and Ecosystem Analysis**

Cecilia Akselsson
Department of Physical Geography and Ecosystem Science

Exam committee:
David Tenenbaum, Department of Physical Geography and Ecosystem Science
Anna Terskaia, Department of Physical Geography and Ecosystem Science

Abstract

Landslides pose a significant risk to human life and infrastructure, especially in Italy, which has a high frequency of landslide occurrences. To mitigate these hazards, Landslide Susceptibility Mapping (LSM) is crucial for identifying risk areas and developing appropriate mitigation strategies. Various methodologies have been adapted to perform LSM, with machine learning models seeing a rise in popularity due their predictive capabilities. The aim of this study is to compare two ensemble tree models, Random Forest (RF) and Extreme Gradient Boosting (XGB), in their predictive performances for landslide susceptibility by type. The typical methodology for assessing landslide type is performing susceptibility assessments individually for every class and aggregating the results. But with the RF and XGB models this process can be simplified by performing one multiclass analysis. The study found that the RF model significantly outperformed the XGB model in multiclass classification, with an overall accuracy of 95.83% compared to the 74.71% of the XGB model. No significant difference was found in the binary classification, with both models having an overall accuracy over 92%. The variables considered most important by both models were found to differ from heuristic models, suggesting a potential bias or incompleteness of the landslide inventory which should be considered in future studies. In conclusion, the RF model demonstrated its proficiency at making maps and high accuracy predictions for each landslide type.

Contents

1. Introduction

Landslides are one of the most destructive natural hazards, causing severe economic damages and can lead to loss of human life (Chaudhary & Piracha, 2021). Italy has the highest total area exposed to landslides and the largest population living in these exposed areas in all of Europe (Jaedicke et al., 2014). Since 1950, Italy has spent more than 52 billion euros on the damages cause by landslides and flooding, and the Ministry of Environment estimates that an additional 40 billion euros are necessary to mitigate all the risks posed by these natural hazards (Trezzini et al., 2013). The production of Landslide Susceptibility Maps (LSMs) is necessary to identify potentially hazardous areas to develop suitable mitigation strategies.

It is important to differentiate between landslide types because each type has varying triggers, characteristics, and potential impacts on human life and infrastructure. For example, fast moving slides, such as rock falls and debris flows, can cause severe and sudden damage to infrastructure and buildings. Sudden landslides are often caused by seismic activity (Wu et al., 2022) or heavy precipitation events (Zêzere, 2002). Slow mass movements, such as creeps and rotational flows, are often the consequence of soil instability caused by prolonged precipitation events (Zêzere et al., 1999). Because of their difference in triggers, the mitigation strategies must be adapted to the type of mass movement.

There are many different methodologies used to perform LSM (Merghadi et al., 2020; Pradhan & Kim, 2016; Reichenbach et al., 2018). In general, the methods of LSM can be divided into four categories, namely; expert-based, deterministic, statistical, and machine learning models (Ado et al., 2022; Merghadi et al., 2020; Pradhan & Kim, 2016; Xing et al., 2021). Machine Learning Models (MLM) have seen a considerable increase in popularity over the recent years (Ado et al., 2022). Merghadi et al. (2020) did a comparative analysis of the binary predictive capabilities of MLM showing significant differences between popular models, with ensemble tree methods (e.g., Random Forest, Extreme Gradient Boosting) significantly outperforming the other regression models. Another benefit of using ensemble tree models is that they are not solely limited to binary predictions and can perform multi-class classifications. The methodology for identifying the typology of landslides is often performed by doing individual assessments for each landslide type and aggregating the results (Clerici et al., 2006; Loche et al., 2022). Research from Taalab et al. (2018) has displayed the predictive capabilities of MLM to determine landslide type in the Piedmont region. The paper used the Random Forest (RF) model to produce high accuracy maps over large heterogenous areas and predict the occurrence of landslide typology, suggesting that the use of machine learning models can simplify the process of assessing multitype landslide susceptibility.

Within the field of LSM, few studies have been conducted to determine landslide susceptibility by typology using MLM and, to my knowledge, no comparative analyses between models have been done. Just like the RF model, the Extreme Gradient Boosting (XGB) model is an ensemble tree model. As mentioned before, both these models performed similarly in terms of binary landslide mapping and studies outside of the field have shown similar performances in multiclass classification between the RF and XGB models (Niu et al., 2019; Salauddin Khan et al., 2023; Wang & Liu, 2020). However, the performances of these models are highly dependent on the complexity of the datasets and can vary significantly based on their specific context.

Therefore, the aim of this study is to do a comparative analysis between the RF and XGB model, with a specific focus on their predictive performance for landslide type identification. This comparative analysis is expected to demonstrate similar performances in both binary and multiclass classification, consistent with previous studies. Due to the added complexity of multiclass analysis, and the region-specific context, their performances may deviate from those expectations in this study. Additionally, the importance of each predicting variable will also be calculated. Studies doing susceptibility analysis using MLMs have not displayed consistent results regarding the predictive capabilities of each variable. Therefore, a substantial variation compared to current literature is expected in the importance of each variable.

1.1 Background

1.1.1 Study Area

The catchment is situated in the north of the Aosta Valley, a region located in the north-western part of Italy. To the north it borders Switzerland, with the Alps forming a natural boundary. The highest point of the catchment is situated on Dent D'herens at 4,171 m a.s.l., and lowest point is in the south at 554 m a.s.l., where it connects with the Dora Baltea River. Due to its topography, there is a high degree variability of climate in the catchment. Most notably the precipitation, with a low in the east of 117 mm/y and a high in the west of 2,215 mm/y.

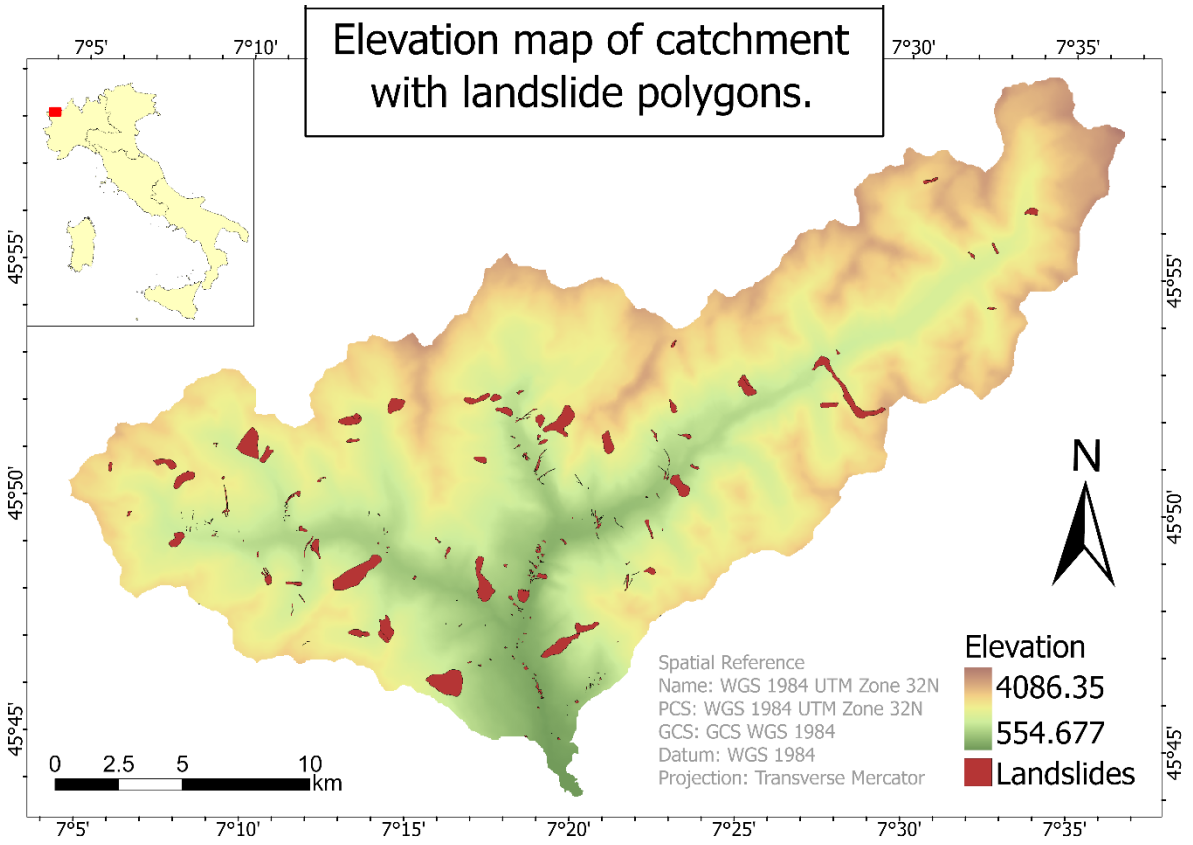

Figure 1: Elevation map of the catchment in the Aosta Valley. Overlayed are the landslide polygons from the IFFI inventory (Trigila et al., 2007).

Geologically, the Aosta Valley is complex, with several lithological units that have undergone intense tectonic deformation during the Alpine orogeny (Bistacchi et al., 2010; Ellero & Loprieno, 2018). The lithology consists of various types of sedimentary and metamorphic rocks, such as marls, limestones, shales, and schists. The geology of the area is characterized by high lithological variation and structural discontinuities, both created by the numerous faults and folds (Bistacchi et al., 2010; Ellero & Loprieno, 2018). The presence of weak lithological layers, such as marls and shales, increases the susceptibility to landslides. Additionally, the structural discontinuities, such as faults and joints, function as weaknesses, facilitating the detachment and movement of unstable masses (Bistacchi et al., 2010; Ellero & Loprieno, 2018).

1.1.2 Landslide Inventory

The IFFI (Inventario dei Fenomeni Franosi in Italia) project is Italy's national landslide inventory (Trigila et al., 2007). It is a collaborate effort of geological institutions at both national and regional levels to identify and map landslides in Italy. The database was compiled using remote sensing techniques, field surveys, and historical archives (Trigila et al., 2007). Landslide inventories are a crucial component for landslide susceptibility mapping, as models relate the explanatory variables with the occurrence of landslides and uses these relations to predict future landslides. The inventory distinguishes between eight distinct types of mass movements, six of which are found within the study area. Table 1 gives a description of each landslide relevant to this study according to a translation of Trigila (2005).

Table 1: IFFI Landslide classes and their descriptions (Trigila, 2005).

Landslide Type	Class	Description
Collapse / Rockfall	C1	The mass moves predominantly through the air, by free fall, jumping, bouncing, and rolling, breaking into various elements of variable size, and is generally characterized by extremely rapid movement.
Rotational / Translational sliding	C2	The movement involves displacement along one or more surfaces where shear resistance is exceeded or within a relatively thin zone characterized by intense shear deformation.
Slow creep	C3	Where movements are generally characterized by low speed and involve terrains with high clay content and mostly low water content. These phenomena, even of large dimensions, mainly affect moderately steep slopes composed of clayey rocks or altered rocks with a clayey matrix.
Fast sliding	C4	movements are generally characterized by high speed and primarily affect loose soils with a significant water content. These phenomena, usually of insignificant size, are triggered by intense precipitation and typically involve loose covering soils of slopes with rather steep inclines across their entire grain size range.
Complex	C5	The movement results from the combination of two or more of the previously described movements.
Not defined	C6	This class is not defined within the inventory.

1.1.3 Decision Trees

RF and XGB both use an ensemble of decision trees to make predictions (Breiman et al. 2001). Decision trees are models used for both classification and regression tasks. They are a visual representation of the decision-making process the algorithm undertakes, resembling a tree like structure with branches, exemplified in Figure 2. The root node presents an initial split of the dataset based on an explanatory variable. Each subsequent split at the internal nodes divides the dataset into smaller subsets until the leaf nodes represent a homogeneous subset of classification.

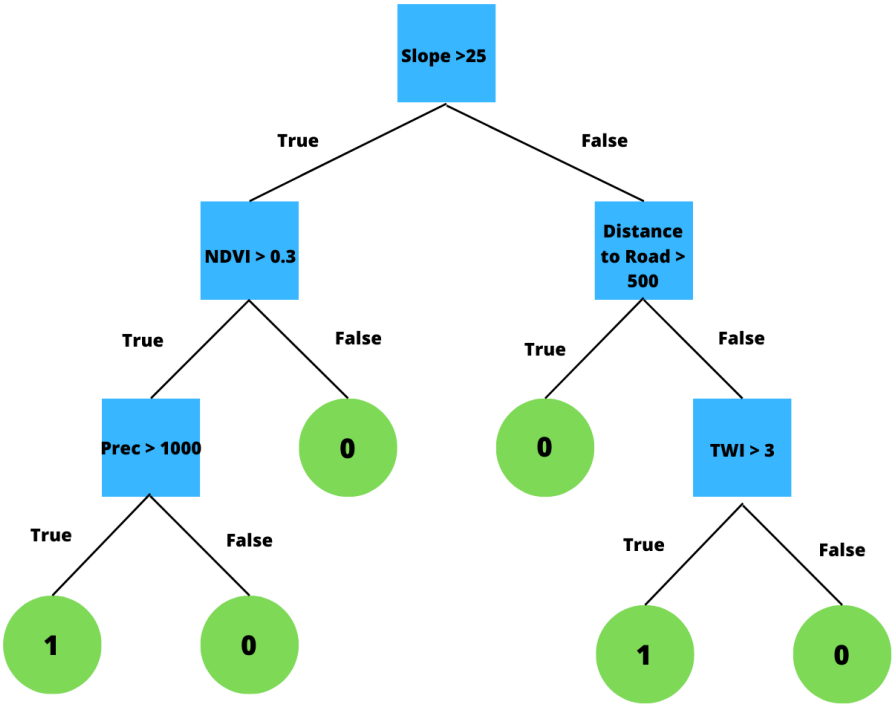

Figure 2: Example of a single decision tree produced by the models. Note that the variable and associated value threshold are randomly selected. The red, orange, and green boxes indicate the root, internal, and leaf node, respectively.

For classification tasks for complex problems such as LSM, a singular tree is a weak predictor, therefore the models aggregate the results of an ensemble of trees instead. When a tree becomes too complex (i.e., too many internal nodes), it can result in having a high variance. High variance means that the model overfits the training data, making it unable to perform accurate predictions on new data sets. On the other hand, when the tree becomes too simple, it will result in having a high bias, making the tree overgeneralize the relationship between the explanatory variables and the target variables.

Most ensemble algorithms eliminate these uncertainties using two techniques: Bootstrapping and aggregation (the combination of these two techniques is commonly referred to as bagging). Bootstrapping is a technique for resampling the dataset. It creates a new dataset of equal size by randomly sampling observations from the original dataset with replacement, causing some observations to become duplicates while roughly a third of observations are omitted. It then builds decision trees based on the explanatory variables, runs new samples through every tree, and combines the results of all trees (aggregation). The aggregation and randomized selection reduce both the bias and variance, giving the models the ability to accurately make predictions on unseen data.

Both models use the majority vote of all trees to determine the classification. When the majority of the decision trees assign a particular class to the data sample (e.g., 80 out of 100 trees determined the presence of a landslide), then the output will be classified as that particular class. For the probability predictions, it uses the percentage of trees that voted for that classification. With the example from before, if 80 out of 100 trees vote for the presence of a landslide, the output probability will be 80% for that instance.

The RF model incorporates an extra element of randomness by basing the creation of decision trees on a random selection of explanatory variables. This prevents the algorithm from always selecting the same subset of variables which helps reduce overfitting and variance. XGB, just like RF, is a decision tree-based algorithm that incorporates bagging to reduce bias and variance. But unlike RF, which creates every tree independently, XGB uses a boosting method. Boosting means that the model trains each new tree sequentially, adjusting the weights of the training samples to account for misclassified occurrences. It uses a loss function to calculate the new weights of each sample. This causes later iterations of the trees to become more accurate, and the ensemble to become stronger. One problem that can arises from boosting, compared to the randomized nature of RF, is overfitting and high variance. To prevent these errors, the algorithm has regulation parameters, such as the learning rate, which controls the step size during the boosting process. A lower learning rate can help prevent overfitting by gradually adjusting the model's weights and reducing the impact of each individual tree in the ensemble, leading to improved generalization performance. Additionally, the XGB model incorporates regression parameters such as 'lambda' and 'alpha'. These parameters control the complexity of the model by adding penalties to the loss function, discouraging the model from fitting irrelevant and noisy patterns present in the training data. With the way trees are constructed in the XGB model compared to the RF model, it is expected that the validation metrics will slightly favour the XGB model.

2. Methodology

2.1 Data Preparation

To achieve the aim of this paper, the RF and XGB models related the landslide inventory to a set of explanatory variables. The dataset containing landslide occurrences and explanatory variables was split into a training dataset and a test dataset. Thus, seventy percent of the available data samples were used to train the model and the remaining thirty percent was used to evaluate its accuracy. The models for binary and multiclass predictions were trained and validated separately. The trained models then classified a dataset of the entire region to produce the LSMs.

The relevant data was visualized and managed in GIS-software (ESRI, ArcGIS Pro 2.7.0). More explanatory variables were derived from the relevant data sources in ArcGIS. The dataset consisted of both continuous and discreet variables and was converted into raster format. The resolution of the grid cells has been shown to influence the accuracy of classifier models (Catani et al., 2013), with a 50 m resolution resulting in the highest accuracy for the RF model. Therefore, all conditioning variables were resampled to a 50 m resolution grid.

The explanatory data which the models in this study were trained on can be categorized into four categories: Topographical (i.e. elevation, slope, curvature profile), geological (i.e. soil

texture, soil parent material, soil depth), hydrological (i.e. precipitation, Topographical Wetness Index, distance to streams), and human influence (i.e. distance to roads and land use). Table 2 displays information about the variables used in this study and their sources. A test was performed to determine each explanatory variable's relative importance, using the "gain" objective. This means that it evaluates the average accuracy that is gained from the inclusion of a variable. Redundant features (gain < 0.5) were removed from the analysis, which helps decrease the complexity of the models, and in turn increases their reliability. The final variables used in the analysis are displayed in Table 2 below.

Table 2: Explanatory variables used for the susceptibility analysis.

Variable and Spatial Resolution	**Description**	**Data source**
Elevation *25 x 25 m*	*Elevation* is commonly used in LSM as it is related to the potential energy of landforms and is indicative of climate (Costanzo et al., 2012; Dai & Lee, 2002).	*EU-DEM v1.1, Copernicus Land Monitoring Service, European Union Copernicus Land Monitoring Service (2016)*
Slope *25 x 25 m*	*Slope* angle is typically considered the main determinants of slope stability. This is because the slope angle is directly related to the gravitational force acting on the surface. It also influences the movement of surface runoff.	*Derived from Copernicus DEM.*
Curvature *25x25 m*	*Curvature* is the derivative of the slope and influences the acting forces on the hillside, affecting the acceleration patterns of mass movements (Ohlmacher, 2007).	*Derived from Copernicus DEM.*
Parent Material *Shapefile* *(1: 1 000 000)*	The dominant *parent material* is considered as different rock types have different physical and chemical properties (van Westen et al., 2008). The classification made is based on the mechanical properties such as composition and shear strength (Carraraet al., 1999).	*European Soil Data Base from European Commission and the European Soil Bureau Network (2004)*
Soil Texture *Shapefile* *(1: 1 000 000)*	The *Soil texture* determines the forces acting between soil particles and contributes to the drainage patterns of the soil (Fan & Or, 2016).	*European Soil Data Base from European Commission and the European Soil Bureau Network (2004)*
Soil Depth *Shapefile* *(1: 1 000 000)*	*Soil depth* is related to the stability of slopes and the shear strength. It also influences water retention and the development of vegetation (Fan & Or, 2016).	*European Soil Data Base from European Commission and the European Soil Bureau Network (2004)*
TWI *25 x 25 m*	*TWI* is the topographical wetness index and is an index for water accumulation patterns in a landscape. In LSM, it is used as a proxy for soil moisture conditions (Sharma, 2010).	*Derived from Copernicus DEM.*
Distance to Stream *Shapefile* *(1: 250 000)*	The *distance to a stream* can be an indicator of the landform characteristics. Streams tend to follow natural drainage paths, which can coincide with areas of concentrated erosion or geologically weak zones.	*Derived from Ecrins, The European Environment The European Environment Agency (2012)*
Precipitation	The mean annual *precipitation* is a trigger	*Derived from WorldClim2, Fick*

1 x 1 km	for landslides. It increases the gravitational forces on slopes as it infiltrates the soil. The erosion of slope material caused by precipitation can lower the shear strength.	*and Hijmans (2017)*
Land Use *100 x 100 m*	*Land use* influences slope stability as it can be used as an index for vegetation cover, surface runoff, erosion, and human land modification.	*Corine Land Cover, Copernicus Land Monitoring Service, European Union Copernicus Land Monitoring Service (2018)*
Distance to Road *Shapefile* *(1: 250 000)*	The *distance to roads* can be considered an explanatory variable as heavy traffic can cause vibrations in the hillside. Also, the construction of roads in mountainous regions often require modifications to the hillside and can disrupt drainage patterns (Collins, 2008).	*Derived from gROADSv1, Center for International Earth Science Information Network - CIESIN - Columbia University and Information Technology Outreach Services - ITOS - University of Georgia (2013)*
NDVI *250 x 250 m*	*NDVI* is the Normalized Difference Vegetation Index and is a measure of live green vegetation. Roots can positively impact soil stability and change infiltration patterns (Peduzzi, 2010).	*MODIS vegetation indices, ORNL DAAC (2018)*

2.2 Sampling Methods

The sampling points for the landslide category were extracted by converting the polygons to raster, where each landslide cell represented a sample point. For the non-landslide category there are different sampling methods, each affecting the results of the susceptibility map (Yilmaz, 2010). The sampling can be done completely at random, or by a heuristic selection of areas not susceptible to landslides as demonstrated in Gómez and Kavzoglu (2005). For this study, a 200 m buffer is created around landslide polygons and the non-susceptible samples are extracted at random outside of the buffer zone. Figure 3 illustrates the sampling technique used in this study.

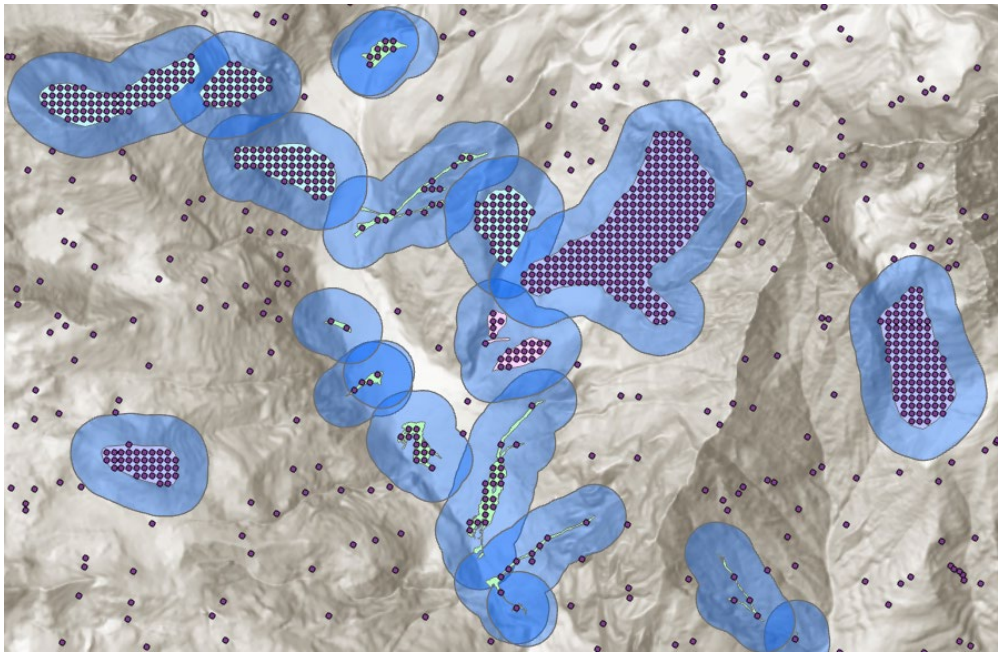

Figure 3: Distribution of the sampling points (purple). Susceptible class is sampled from inside the polygons. Non-susceptible class extracted from areas outside the blue buffer zones.

Catani et al. (2013) found that the accuracy of classifier models increased along with the number of sample points used. Both models need a balanced dataset to perform optimally, meaning that the class distribution is within 5-10% of each other. Because of this, the number of sampling points is limited by the landslide inventory (Petschko et al., 2014). In the study area, the landslide inventory identified 221 landslides of seven varying types, with sizes ranging from 141 m^2 to 1,064,443 m^2 (median 4,498 m^2, total 11,839,643 m^2). As landslides only cover 2.6% of the total catchment area, the maximum number of sample points can only cover 5.2% of the total area or 9,422 cells.

2.3 Modelling

From these sample cells, the values of the overlaying explanatory datasets were extracted into a CSV (comma-separated values) file which was accessed using the pandas library in the python interpreter, PyCharm (JetBrains, 2017). The RF model was trained in the python environment (version 3.9) using the sk-learn library (Pedregosa et al., 2011), and the XGB model was trained using the XGBoost library (Chen & Guestrin, 2016). Both models have hyperparameters (e.g., Number of trees, learning rate, and more that are expanded on in the Appendix) that required tuning for optimal performance. This was done by cross evaluating each parameter within their realistic ranges over a thousand iterations. The RF model has built-in optimizations for unbalanced datasets, the XGB model however does not and requires extra tuning. Weights were assigned to each sample based on the number of occurrences in their landslide classification and were considered in the modelling. After training the binary model and doing predictions on the training dataset, the results were filtered to minimize the non-susceptible class, as only samples with a probability above 0.3 were exported to a new training dataset for the multiclass models. The process of hyperparameter optimization was repeated for the multiclass model on the new training data. After training the models, a dataset containing the predictive variables for the entire region was processed in Python and subsequently exported back to ArcGIS to produce a landslide susceptibility map for the catchment.

2.4 Evaluation Metrics

To assess the performance of the RF and XGB models in predicting landslide susceptibility, several evaluation metrics were calculated. These metrics demonstrate the models' predictive accuracy and their ability to identify landslide prone areas. The calculations are performed on the test dataset to evaluate the models' robustness on new data.

The evaluation metrics employed in this study are accuracy, precision, recall, and the F-1 score. These metrics are widely used in LSM. Accuracy is the proportion of correctly classified samples over the number of total samples and provides an overall measure of the models' correctness. Precision is the proportion of correctly predicted positive samples over the total predicted positive samples and is an indication of the models' ability to correctly identify landslide occurrences. Recall on the other hand, is the proportion of correctly predicted positive samples over the total observed positive samples. In addition to these metrics, the kappa statistic was calculated from the confusion matrixes. It is generally considered more robust than the aforementioned statistics as it considers the possibility of

correct classifications happening by chance. The kappa value ranges from -1 to 1, where a value of 1 indicates perfect agreement, 0 represents agreement equivalent to chance, and -1 denotes complete disagreement.

The strengths and weaknesses of the models were analysed using these metrics, and the interpretability of the produced LSMs. The relative importance of each explanatory variable was also considered. The following section will present the results from this analysis.

3. Results

The LSMs produced by the models are displayed in Figures 4 (XGB) and 5 (RF). The maps indicate the susceptibility of landslides on a continous scale from 1 (highly susceptible) to 0 (not susceptible). The model outputs are quite similar over the catchement and both models classify roughly the same areas as susceptible. XGB has higher probability values than RF.

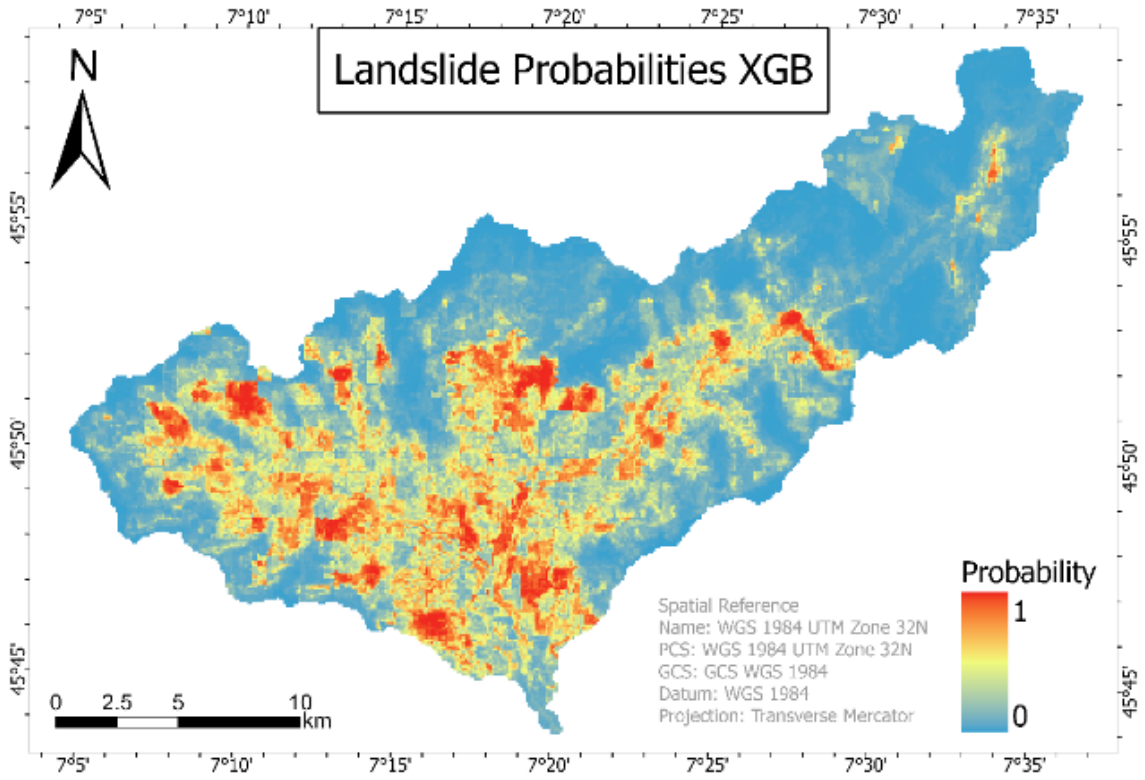

Figure 4: Overall susceptibility map produced by XGB model.

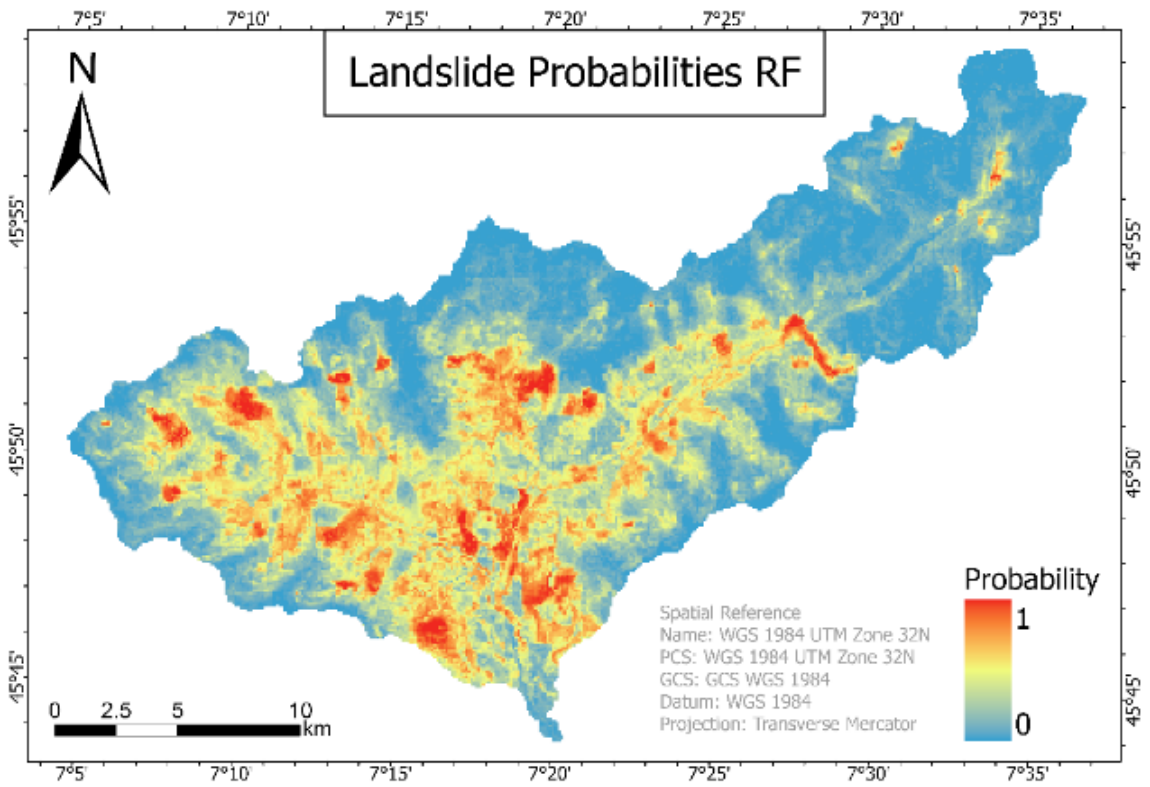

Figure 5: Overall susceptibility map produced by RF model.

Tables 3 (XGB) and 4 (RF) display the confusion matrices for the models. Both models performed similarly with an overall accuracy of 91.00% and 91.25% respectively. The F1-score, which represents the harmonic mean of recall and precision, exhibits a close similarity between the two models. For the XGB model, the F1-score is recorded at 91.11%, while the RF model achieves a slightly higher F1-score of 91.42%. The models also had similar kappa statistics, 0.82 and 0.83 for XGB and RF respectively. The models exhibited comparable kappa statistics, with XGB achieving a value of 0.82 and RF obtaining a slightly higher value of 0.83.

Table 3: Confusion matrix of XGB model for the binary predictions. Average accuracy in the bottom right.

XGB binary		**Predicted**			**Recall**	**Average Recall**
	Type	**Landslide**	**No Landslide**	**Total**		
Observed	**Landslide**	2096	295	2391	87.66%	91.05%
	No Landslide	129	2191	2320	94.44%	
	Total	2225	2486	4711		
Precision		94.20%	88.13%		91.00%	
Average Precision		91.17%				

Table 4: Confusion matrix of RF model for the binary predictions. Average accuracy in the bottom right.

RF binary		**Predicted**			**Recall**	**Average Recall**
	Type	**Landslide**	**No Landslide**	**Total**		
Observed	**Landslide**	2082	309	2391	87.08%	91.32%
	No Landslide	103	2217	2320	95.56%	
	Total	2185	2526	4711		
Precision		95.29%	87.77%		91.25%	
Average Precision		91.53%				

The models also calculate the importance of each explanatory variable, displayed in Figures 6 (XGB) and 7 (RF). It ranks the variables based on the mean decrease in accuracy when a variable is omitted. XGB considers elevation to be the most important factor, followed by the soil parent material and depth, precipitation, and the distance to roads. TWI and curvature are considered the least important predictors according to the XGB. The RF model ranks distance to roads as the best predictor, followed by precipitation, elevation, and NDVI. Whereas the soil properties rank lowest.

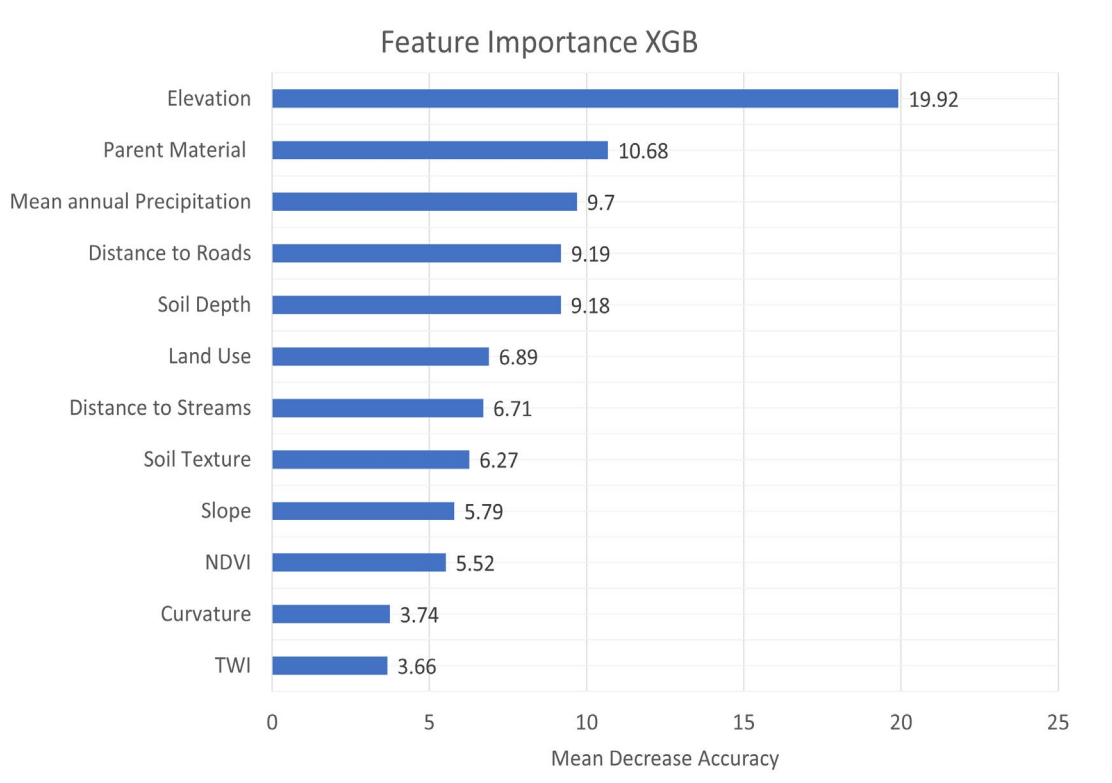

Figure 6: Importance of every feature according to the XGB model.

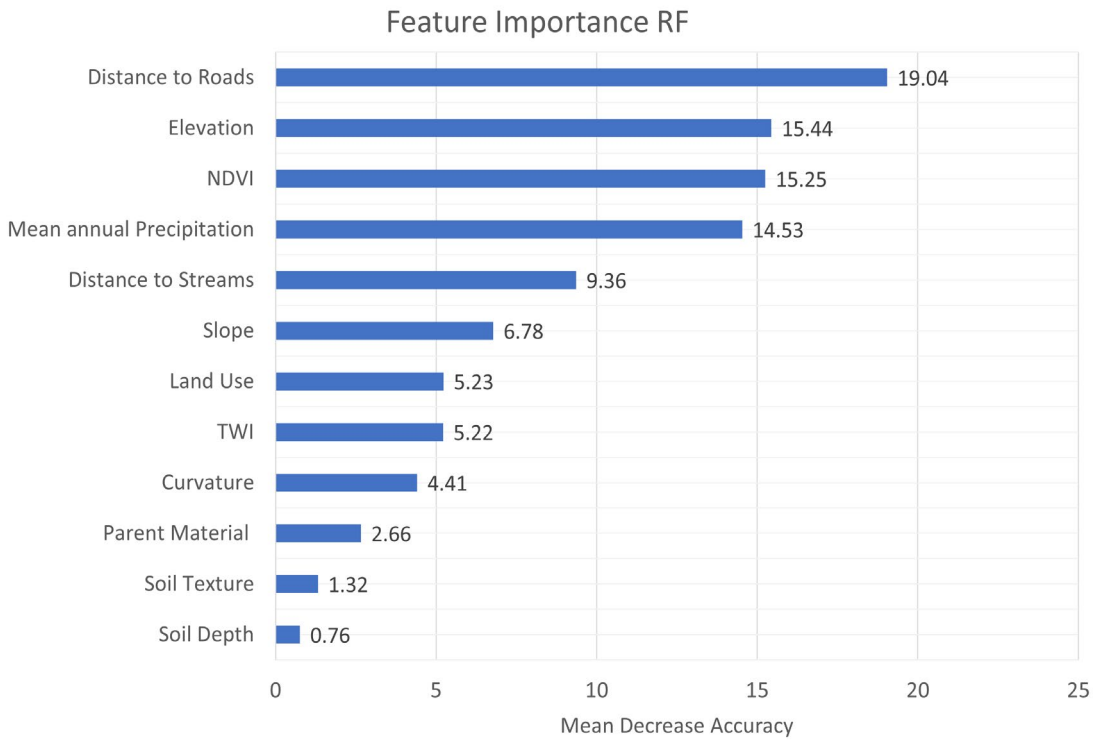

Figure 7: Importance of every feature according to RF model.

The maps in Figures 8 (XGB) and 9 (RF) provide a visual presentation of the spatial distribution for the likelihood of each type of landslide occurrence. The number of cells with medium to high probability appear to be linked to the number of occurrences of each type, most notably the sparse categories C3 (Slow creep) and C4 (Fast sliding) have little presence on the maps.

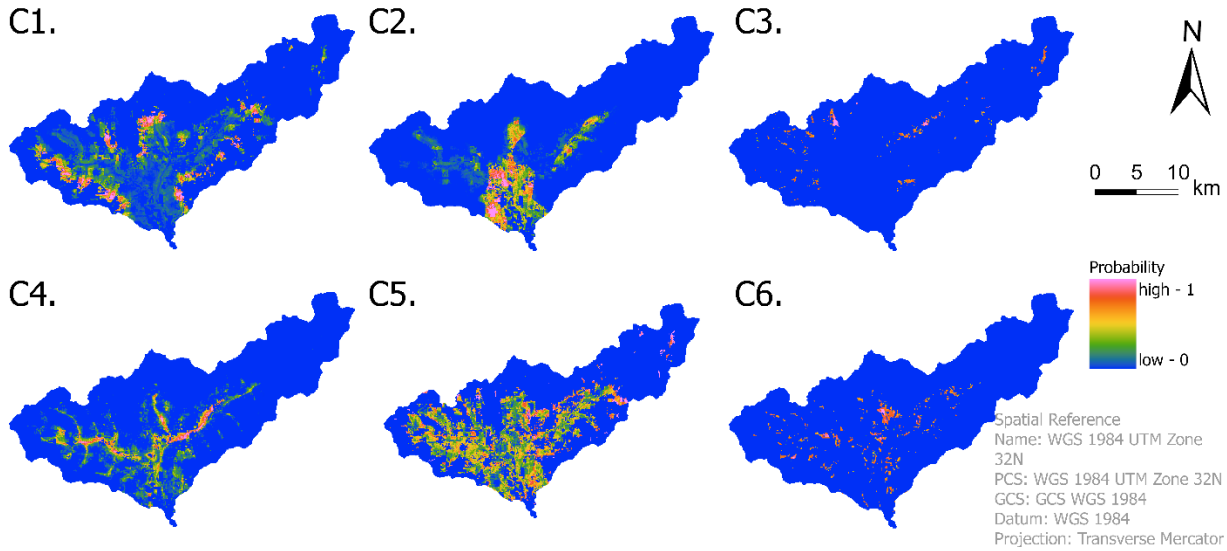

Figure 8: LSMs produced by the XGB model for each landslide type. C1) Collapse / Rockfall C2) Rotational / Translational sliding C3) Slow creep C4) Fast sliding C5) Complex C6) Not defined.

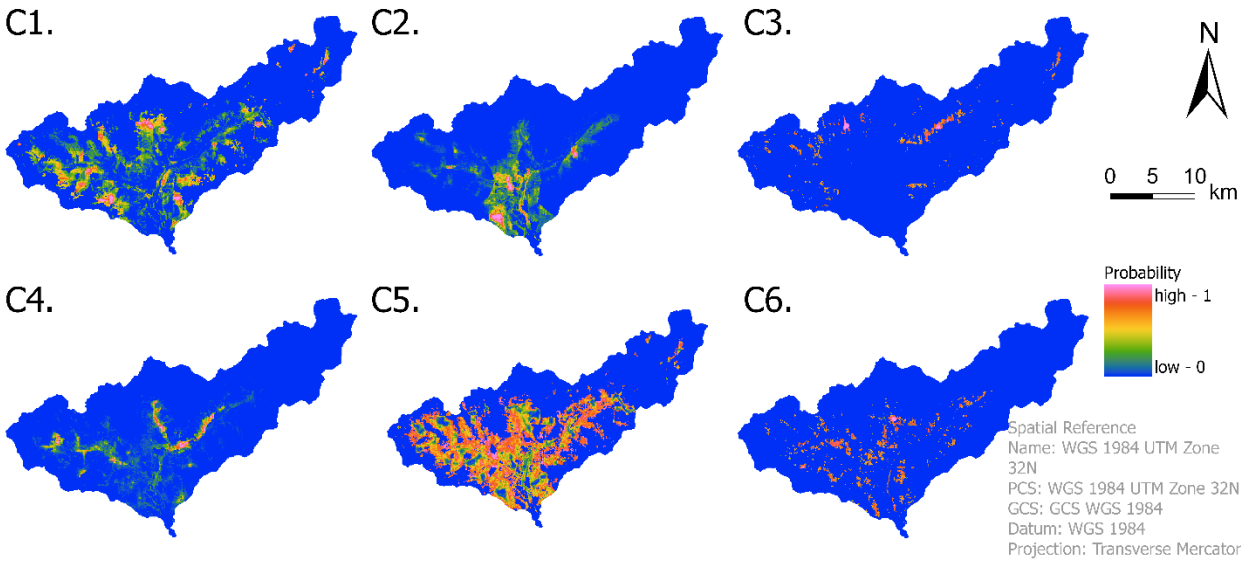

Figure 9: LSMs produced by the RF model for each landslide type. C1) Collapse / Rockfall C2) Rotational / Translational sliding C3) Slow creep C4) Fast sliding C5) Complex C6) Not defined.

In the IFFI landslide inventory, six different landslide types were identified in the area as shown in table 1. XGB's predictions done on the test dataset, return an overall accuracy, recall, and precision of 74.71%, 68.20%, and 79.10%, respectively. According to table 5, C1 (Collapse / Rockfall) and C4 (Fast sliding) have the lowest recall value of 46.98% and 34.38% respectively. And C4 (Fast sliding) having the lowest precision at 62.86%. There is an issue of class imbalance at play here, where the category with the highest precision of 100% merely has ten samples. The kappa statistic calculated from table 5 was 0.61,

Table 5: Confusion matrix of the test dataset by the XGB model. NLS indicated samples where no landslide was detected, the landslides are categorized as follows: C1) Collapse / Rockfall C2) Rotational / Translational sliding C3) Slow creep C4) Fast sliding C5) Complex C6) Not defined.

XGB		NLS	C1	C2	C3	C4	C5	C6	Total	Recall	Average Recall
Observed	NLS	1193	36	55	3	12	125	0	1424	83.78%	68.20%
	C1	126	140	13	0	1	18	0	298	46.98%	
	C2	33	2	240	0	0	9	0	284	84.51%	
	C3	5	0	0	43	0	0	0	48	89.58%	
	C4	32	4	1	0	22	5	0	64	34.38%	
	C5	210	6	15	0	0	464	0	695	66.76%	
	C6	1	0	1	0	0	2	10	14	71.43%	
	Total	1600	188	325	46	35	623	10	2827		
Precision		74.56%	74.47%	73.85%	93.48%	62.86%	74.48%	100.00%		Overall Accuracy:	
Average precision		79.10%								74.71%	

Table 6 shows the confusion matrix of the RF model and has an overall accuracy, recall, and precision of 95.83%, 95.16% and 87.38% respectively. With the lowest recall values found in the C4 (Fast sliding) category with a value of 60.94%. The lowest precision is also found in the C4 (Fast sliding) category with 81.25%, where the model wrongly predicted no landslide occurrence instead of the appropriate landslide type. The kappa value came out to be 0.94.

Table 6: Confusion matrix of the test dataset by the RF model. NLS indicated samples where no landslide was detected, the landslides are categorized as follows: C1) Collapse / Rockfall C2) Rotational / Translational sliding C3) Slow creep C4) Fast sliding C5) Complex C6) Not defined.

RF		NLS	C1	C2	C3	C4	C5	C6	Total	Recall	Average Recall
Observed	NLS	1388	6	13	0	6	11	0	1424	97.47%	87.38%
	C1	30	265	0	0	1	2	0	298	88.93%	
	C2	9	0	273	0	2	0	0	284	96.13%	
	C3	1	0	0	47	0	0	0	48	97.92%	
	C4	23	2	0	0	39	0	0	64	60.94%	
	C5	7	0	1	0	0	687	0	695	98.85%	
	C6	1	0	1	0	0	2	10	14	71.43%	
	Total	1459	273	288	47	48	702	10	2827		
Precision		95.13%	97.07%	94.79%	100.00%	81.25%	97.86%	100.00%		Overall accuracy:	
Average precision		95.16%								95.83%	

The variable of importance for the multiclass classification is displayed in Figures 10 (XGB) and 11 (RF). Note that the function for determining feature importance for the XGB model is still "gain" but in with multiple classes it no longer displays the mean decrease in accuracy but an F-score instead. The ranking and ratios between variables however remained consistent. The XGB model saw a significant increase in the importance of the parent material and TWI. While the most key factor in the binary classification, elevation, is shown to be a bad predictor of landslide type. For the RF model the rankings are very similar to the binary classification, with the order and scores of the four most predicting variables changing slightly.

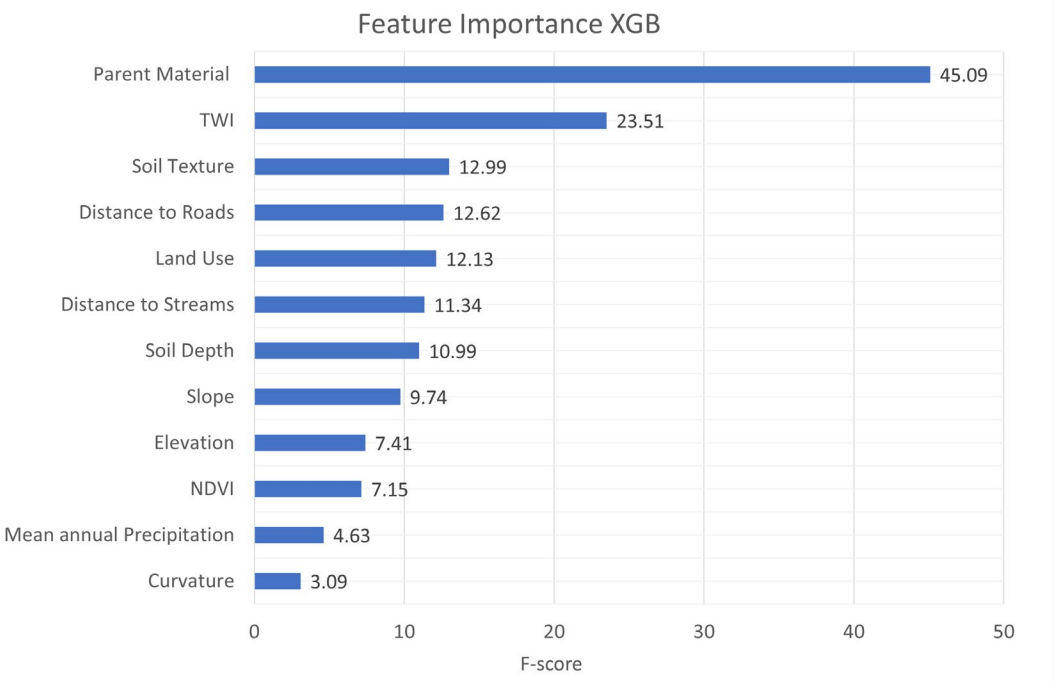

Figure 10: Feature importance according to the XGB model. Variables are displayed on the y-axis and the F-score on the x-axis.

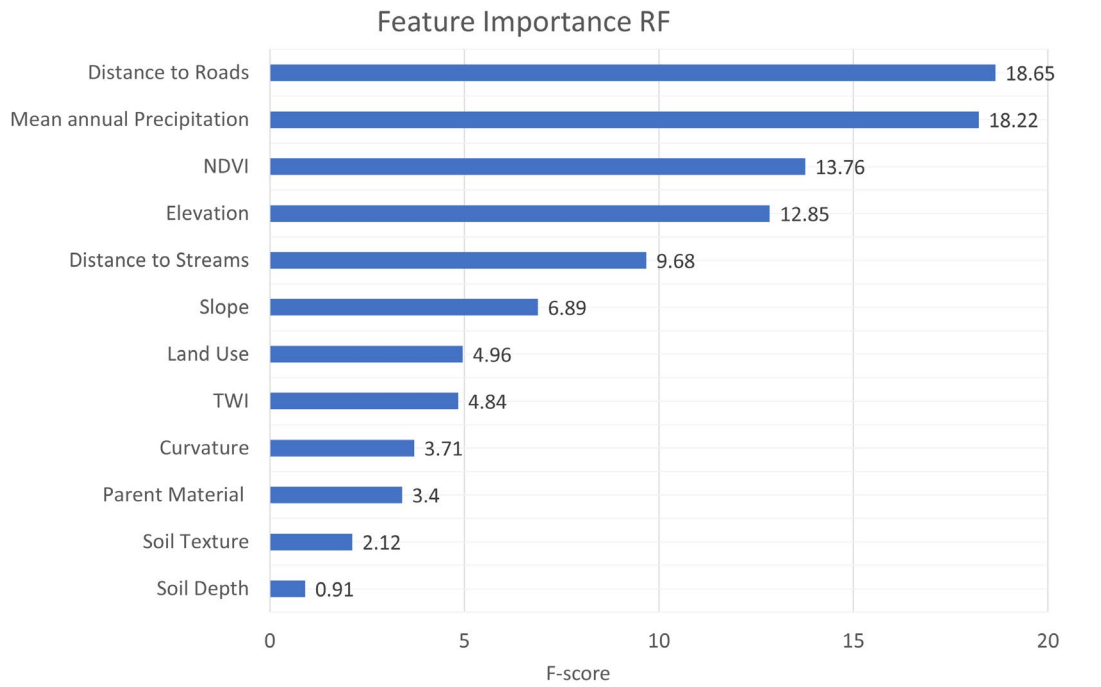

Figure 11: Feature importance according to the RF model.

4. Discussion

4.1 Interpretation and Analysis

The aim of this study was to compare the multiclass predictive capabilities of the RF and XGB models for the prediction of multitype landslide susceptibility. The results indicate that while there was a minimal difference in the evaluation metrics for the binary prediction of landslide presence, a notable disparity emerged when performing predictions on landslide typology. The RF model outperformed the XGB model, achieving an overall accuracy of 95.83% compared to the XGB model's accuracy of 74.71%.

The study highlighted minimal differences between the RF and XGB models' binary capabilities, aligning with previous research in the field. Multiple recent studies, such as Pradhan and Kim (2020), Hussain et al. (2022), and Ado et al. (2022), have reported similar results, with both models' overall accuracies' operation within few percentage points of each other. Along with this study, these studies also demonstrate that the XGB model identifies a larger proportion of susceptible areas as having a very high susceptibility, as compared to the RF model. This suggests that there is less variation between the ensemble of decision trees created by the XGB model, which could be due to the iterative nature of the construction of trees. Because every tree built is based on previous iterations, the likelihood of agreement between consecutive trees on classification is higher than the RF model, which builds every tree independently.

There is a notable difference in the relative importance of each feature for each model as demonstrated in Figures 6, 7, 10, and 11. Figure 6 shows that the omission of elevation in the

dataset would result in a 19.92% decrease in accuracy, making it the most important predictor of landslide occurrence for the XGB model. While the ranking of importance is the same as the findings in Catani et al. (2013), the reasoning is contradictory. The XGB model found an inverse correlation between landslide occurrence and elevation, mapping the lower regions of the catchment as more susceptible. Following elevation in importance are the parent material, precipitation, distance to road, and soil depth. Both statistical and heuristic studies have also demonstrated the importance of the three former variables in LSM, the significance of soil depth however is still debated (Carrara et al., 1995; Chung et al., 1995; Corominas, 1996; Martelloni et al., 2012). The importance of soil depth found by the XGB model is also contradicted by the results of the RF model. The second least important variable according to XGB is curvature. Carrara et al. (1995) debates the practicality of using higher derivatives of elevation, such as curvature, stating that their effectiveness is highly dependent on methodology and regional context. The least important factor according to XGB is TWI and this finding is contradictory to the results from the RF model and Taalab et al. (2018). Figure 10 shows that variables used for predicting presence of landslides do not translate to the prediction of landslide type for the XGB model. These high variances indicate that the variable selection for multiclass analysis using the XGB model could be optimized in the future.

Figure 7 displays the importance of variables according to the RF model. The distance to roads has been used successfully used in LSM (Devkota et al., 2013) and is the best predicting feature for the RF model in this study. It should be noted however that there is a potential bias in the inventory of landslides in proximity to roads. In Europe, the parties responsible for the mapping of landslides are often road management companies or collaborate closely with them (Van Den Eeckhaut & Francisco, 2012). This can cause a shift in the prioritization of mapping landslides close to existing or future road infrastructure. While there is evidence supporting the impact of the anthropogenic influences caused by roads on landslides (Collins, 2008; Ramakrishnan et al., 2013), heuristic methods suggest other factors, mainly topographical and hydrological, to be better predictors (Bourenane et al., 2015; Sezer et al., 2017). This is also demonstrated by Figure 7, displaying elevation, NDVI, and precipitation to be the next most predicting variables. Once again, the relationship between landslide occurrence and elevation is inverse and in contradiction to current literature. The ranking of vegetation and precipitation however corresponds to the findings of Jaafari et al. (2014). Figure 11 demonstrates that the variables RF used for the binary classification are also good predictors of landslide type, with minimal differences showing between the importance of explanatory variables in binary and multiclass classification.

The results for both models regarding variable importance could potentially indicate a bias in the landslide inventory. The northeast side of the catchment is characterized by factors (i.e., slope, elevation, vegetation, lithology) that heuristically would be more favourable to landslide occurrence than the landslide dense centre of the catchment. The inventory upon which the MLMs are trained however has not recorded many occurrences in the northeast, causing a discrepancy between the results of MLMs and typical heuristic based approaches of landslide susceptibility. The potential lack of recordings could be due to several reasons: Accessibility for field research, prioritization of infrastructure, or the area not being susceptible to landslides due to variables not included in this study. The incompleteness of the IFFI inventory is also questioned by Loche et al. (2022), who found discrepancies in

landslide recordings between the inventory and region specific studies. In future research, the validity of the landslide inventory should always be carefully considered and rectified if necessary, before analysis is performed.

Figures 8 (XGB) 9 (RF) depict the LSMs for each class and show highly similar results in spatial distribution between the two models, with varying degrees of probability. The spatial distribution of complex landslides is remarkably similar to the overall susceptibility maps produced by the binary model. This is due to the definition and high frequency of complex landslides. A landslide is classified as complex when it shares characteristics or causes with at least two different landslide types. Therefore, it overlaps with all susceptible areas specific to other landslide typologies. The highest susceptibility of collapse / rockfall landslides is on the upper slopes surrounding the valley. These areas facilitate the occurrence of the collapse / rockfall type due to their topography and position above the treeline, allowing the masses to move freely through the air, which is the basis of their classification. The lower slopes, near the bottom of the valley, have the most susceptibility to rotational / translational sliding. Similarly, fast sliding landslides are also predicted to occur mostly on those lower slopes but are more concentrated. There are few occurrences in the inventory of the not defined and slow creep landslides, making their distributions limited to the landslide samples they were trained on. With an increase in samples, the “not defined” category would probably share the same spatial distribution as the complex category, as their causes and mechanisms are not specified nor region specific.

The discrepancy found in the accuracies of the multiclass predictions is something that has not been shown in previous research. Not much research has been done comparing the performances of the models’ multiclass capabilities in the field of LSM, but outside of the field it was found that the models performed similarly to each other in terms of accuracy (Niu et al., 2019; Salauddin Khan et al., 2023; Wang & Liu, 2020). Their research focused on variety of different topics, from credit card fraud to microbiology, suggesting that the relative performance of these models can vary based on the specific context and characteristics of the dataset. The difference in results presented in those studies and this one, highlights that the efficacy of the model is dependent on the data and methods used in the research.

The use of the majority vote in multiclass analysis by ensemble tree methods introduces uncertainties that should be considered. Although both models leverage the aggregation of multiple trees to enhance their accuracies, the majority vote concept used to predict the class assumes equal importance and reliability of each tree. In cases where there is an uneven distribution of occurrences within the classes, such as in this study, the majority vote scheme can lead to the dominance of the most prevalent class, lowering the overall accuracy of the model. The RF model has built-in optimizations to account for unbalanced class distributions, and weight assignment based on class frequency in the XGB model reduces the problems associated with an uneven dataset. Even with these optimizations the results will be inferior to a balanced dataset. When looking at Table 5 and 6, the main error made by the model is misclassifying the landslide type as the no landslide occurrence, highlighting the skewing towards the majority class. Also, the proportion of occurrences necessary to reach the majority vote decreases with an increased number of classes. This leads to instances of classes being predicted with less than the majority and only a plurality of the vote (e.g., a class has 30% of the vote and the other 70% is evenly distributed between three other classes, the model will predict the class with 30% even though it is 70% certain it’s a different class.).

The results indicate that the RF model is better at accounting for these uncertainties presented by the majority vote scheme and the imbalance of classes than the XGB model.

4.2 Limitations

Future studies could improve on the validation method of this study to depict the models' robustness more accurately on new datasets by incorporating an independently sampled dataset for testing. With the method presented in this study, 30% of the sampling points are randomly sampled for testing purposes, meaning that both training and testing datapoints could potentially be extracted from the same landslide polygon with highly similar explanatory variables. If the test data is independently sampled in a way that entire polygons are omitted from the training, the validation metrics would better reflect the models' capabilities to predict new landslide occurrences. This approach would offer a more reliable evaluation of the models' performance on unseen data.

In the context of LSM, the completeness of the inventory plays a vital role in ensuring the accuracy of the modelling process. As highlighted by Loche et al. (2022), the IFFI inventory is derived from 20 different sub-inventories and has been subject to discussion regarding its completeness. As mentioned above, there is reason to believe a spatial bias is present in the dataset, but further research is needed to thoroughly investigate and understand the extent of this spatial bias and its potential impact on the results and conclusions of the study. While these areas should be prioritized as mitigation measures there are the most effective to reduce damages to infrastructure and human life, it results in the inventory potentially displaying discrepancies compared to reality. The model output is only as good as the input and an incomplete inventory will lead to an inaccurate depiction of the reality of landslides and their causal factors; therefore, the validity of the inventory must be established in future research.

Another limitation to consider is the computational power necessary to perform LSM. The scope of this study was severely limited by the available computational resources and time needed to run the algorithms. The study would have benefited from a larger number of samples to prevent overfitting on sparse categories. With the current sample sizes, it is probable that the models tend to overfit their classification. As regions where landslides occur are small in frequency and extent, the models become specialized and have trouble processing previously unseen data. An increase in area and thus sample size would likely result in a lower overall accuracy. This however would not necessarily entail that the model would be worse, as it would give a more accurate depiction of the region's susceptibility and the underlying relationships between causal factors.

5. Conclusion

In conclusion, the study aimed to compare the multiclass predictive capabilities of the Random Forest (RF) and Extreme Gradient Boosting (XGB) models as applied to landslide susceptibility modelling. The results showed that both models perform similarly on binary classification problems, but discrepancies emerged with respect to multiclass predictions, where RF outperformed the XGB model with an overall accuracy of 95.83% compared to XGB's 74.71%. This finding refutes my hypothesis, which expected the XGB model to

perform better due to tree boosting. The superior performance of RF model can be attributed to it being able to better manage imbalanced datasets. While the importance of certain variables differed from what is found in other similar studies in the literature, they demonstrated their efficacy in prediction not only the presence of landslides, but also their typology.

While the LSMs produced for each landslide type proved useful for some categories, the study would have benefited from a bigger scope with more samples, as this would have reduced overfitting on sparse landslide categories. While a larger scope would result in lower accuracy metrics, it would help visualize the susceptibility to each landslide type better and make the results more useful for decision makers. The study could have also been improved by implementing independent test datasets to validate the models' robustness on unseen data. The importance of complete, high resolution landslide inventories can also not be understated when it comes to accurately modelling real world scenarios and should be considered in future research.

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://www.mdpi.com/2072-4292/14/13/3029

Bistacchi, A., Massironi, M., & Menegon, L. (2010). Three-dimensional characterization of a crustal-scale fault zone: The Pusteria and Sprechenstein fault system (Eastern Alps). *Journal of Structural Geology*, *32*, 2022-2041. https://doi.org/10.1016/j.jsg.2010.06.003

Bourenane, H., Bouhadad, Y., Guettouche, M. S., & Braham, M. (2015). GIS-based landslide susceptibility zonation using bivariate statistical and expert approaches in the city of Constantine (Northeast Algeria). *Bulletin of Engineering Geology and the Environment*, *74*(2), 337-355. https://doi.org/10.1007/s10064-014-0616-6

Carrara, A., Cardinali, M., Guzzetti, F., & Reichenbach, P. (1995). Gis Technology in Mapping Landslide Hazard.

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.*, *13*(11), 2815-2831. https://doi.org/10.5194/nhess-13-2815-2013

Center for International Earth Science Information Network - CIESIN - Columbia University, & Information Technology Outreach Services - ITOS - University of Georgia. (2013). *Global Roads Open Access Data Set, Version 1 (gROADSv1)* NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4VD6WCT

Chaudhary, M. T., & Piracha, A. (2021). Natural Disasters—Origins, Impacts, Management. *Encyclopedia*, *1*(4), 1101-1131. https://www.mdpi.com/2673-8392/1/4/84

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785

Chung, C.-J., Fabbri, A. G., & Westen, C. J. v. (1995). Multivariate Regression Analysis for Landslide Hazard Zonation.

Clerici, A., Perego, S., Tellini, C., & Vescovi, P. (2006). A GIS-based automated procedure for landslide susceptibility mapping by the Conditional Analysis method: the Baganza valley case study (Italian Northern Apennines). *Environmental Geology*, *50*(7), 941-961. https://doi.org/10.1007/s00254-006-0264-7

Collins, T. K. (2008). Debris flows caused by failure of fill slopes: early detection, warning, and loss prevention. *Landslides*, *5*(1), 107-120. https://doi.org/10.1007/s10346-007-0107-y

Corominas, J. (1996). The angle of reach as a mobility index for small and large landslides: Reply. *Canadian Geotechnical Journal*, *33*, 260-271. https://doi.org/10.1139/t96-005

Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J. D., & Chacón, J. (2012). Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: application to the river Beiro basin (Spain). *Nat. Hazards Earth Syst. Sci.*, *12*(2), 327-340. https://doi.org/10.5194/nhess-12-327-2012

Dai, F. C., & Lee, C. F. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology*, *42*, 213-228. https://doi.org/10.1016/S0169-555X(01)00087-3

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., Dhital, M. R., & Althuwaynee, O. F. (2013). Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Natural Hazards*, *65*(1), 135-165. https://doi.org/10.1007/s11069-012-0347-6

Ellero, A., & Loprieno, A. (2018). Nappe stack of Piemonte–Ligurian units south of Aosta Valley: New evidence from Urtier Valley (Western Alps). *Geological Journal*, *53*(5), 1665-1684. https://doi.org/https://doi.org/10.1002/gj.2984

European Commission and the European Soil Bureau Network. (2004). *The European Soil Database distribution version 2.0* https://esdac.jrc.ec.europa.eu/

European Union Copernicus Land Monitoring Service. (2016). European Digital Elevation Model (EU-DEM), version 1.1.

European Union Copernicus Land Monitoring Service. (2018). *CORINE Land Cover*

Fan, L., & Or, D. (2016). Effects of soil spatial variability at the hillslope and catchment scales on characteristics of rainfall-induced landslides. *Water Resources Research*, *52*. https://doi.org/10.1002/2015WR017758

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302-4315. https://doi.org/https://doi.org/10.1002/joc.5086

Gómez, H., & Kavzoglu, T. (2005). Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Engineering Geology*, *78*(1), 11-27. https://doi.org/https://doi.org/10.1016/j.enggeo.2004.10.004

Hussain, M. A., Chen, Z., Kalsoom, I., Asghar, A., & Shoaib, M. (2022). Landslide Susceptibility Mapping Using Machine Learning Algorithm: A Case Study Along Karakoram Highway (KKH), Pakistan. *Journal of the Indian Society of Remote Sensing*, *50*(5), 849-866. https://doi.org/10.1007/s12524-021-01451-1

Jaafari, A., Najafi, A., Pourghasemi, H. R., Rezaeian, J., & Sattarian, A. (2014). GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *International Journal of Environmental Science and Technology*, *11*(4), 909-926. https://doi.org/10.1007/s13762-013-0464-0

Jaedicke, C., Van Den Eeckhaut, M., Nadim, F., Hervás, J., Kalsnes, B., Vangelsten, B. V., Smith, J. T., Tofani, V., Ciurean, R., Winter, M. G., Sverdrup-Thygeson, K., Syre, E., & Smebye, H. (2014). Identification of landslide hazard and risk 'hotspots' in Europe. *Bulletin of Engineering Geology and the Environment*, *73*(2), 325-339. https://doi.org/10.1007/s10064-013-0541-0

Loche, M., Alvioli, M., Marchesini, I., Bakka, H., & Lombardo, L. (2022). Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. *Earth-Science Reviews*, *232*, 104125. https://doi.org/https://doi.org/10.1016/j.earscirev.2022.104125

Martelloni, G., Segoni, S., Fanti, R., & Catani, F. (2012). Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides*, *9*(4), 485-495. https://doi.org/10.1007/s10346-011-0308-2

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, *207*, 103225. https://doi.org/https://doi.org/10.1016/j.earscirev.2020.103225

Niu, X., Wang, L., & Yang, X. (2019). A comparison study of credit card fraud detection: Supervised versus unsupervised. *arXiv preprint arXiv:1904.10604*.

Ohlmacher, G. (2007). Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology*, *91*, 117-134. https://doi.org/10.1016/j.enggeo.2007.01.005

ORNL DAAC. (2018). *MODIS and VIIRS Land Products Global Subsetting and Visualization Tool.* https://doi.org/https://doi.org/10.3334/ORNLDAAC/1379

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Peduzzi, P. (2010). Landslides and vegetation cover in the 2005 North Pakistan earthquake: A GIS and statistical quantitative approach. *Natural Hazards and Earth System Sciences*, *10*. https://doi.org/10.5194/nhess-10-623-2010

Petschko, H., Brenning, A., Bell, R., Goetz, J., & Glade, T. (2014). Assessing the quality of landslide susceptibility maps – case study Lower Austria. *Nat. Hazards Earth Syst. Sci.*, *14*(1), 95-118. https://doi.org/10.5194/nhess-14-95-2014

Pradhan, A. M. S., & Kim, Y.-T. (2020). Rainfall-Induced Shallow Landslide Susceptibility Mapping at Two Adjacent Catchments Using Advanced Machine Learning Algorithms. *ISPRS International Journal of Geo-Information*, *9*(10), 569. https://www.mdpi.com/2220-9964/9/10/569

Pradhan, A. M. S., & Kim, Y. T. (2016). Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. *CATENA*, *140*, 125-139. https://doi.org/https://doi.org/10.1016/j.catena.2016.01.022

Ramakrishnan, D., Singh, T. N., Verma, A. K., Gulati, A., & Tiwari, K. C. (2013). Soft computing and GIS for landslide susceptibility assessment in Tawaghat area, Kumaon Himalaya, India. *Natural Hazards*, *65*(1), 315-330. https://doi.org/10.1007/s11069-012-0365-4

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

Salauddin Khan, M., Nath, T. D., Murad Hossain, M., Mukherjee, A., Bin Hasnath, H., Manhaz Meem, T., & Khan, U. (2023). Comparison of multiclass classification techniques using dry bean dataset. *International Journal of Cognitive Computing in Engineering*, *4*, 6-20. https://doi.org/https://doi.org/10.1016/j.ijcce.2023.01.002

Sezer, E. A., Nefeslioglu, H. A., & Osna, T. (2017). An expert-based landslide susceptibility mapping (LSM) module developed for Netcad Architect Software. *Computers & Geosciences*, *98*, 26-37. https://doi.org/https://doi.org/10.1016/j.cageo.2016.10.001

Sharma, A. (2010). Integrating terrain and vegetation indices for identifying potential soil erosion risk area. *Geo-spatial Information Science*, *13*(3), 201-209. https://doi.org/10.1007/s11806-010-0342-6

Taalab, K., Cheng, T., & Zhang, Y. (2018). Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*, *2*(2), 159-178. https://doi.org/10.1080/20964471.2018.1472392

The European Environment Agency. (2012). *European catchments and Rivers network system (Ecrins)* http://www.eea.europa.eu

Trezzini, F., Giannella, G., & Guida, T. (2013). Landslide and Flood: Economic and Social Impacts in Italy. In C. Margottini, P. Canuti, & K. Sassa (Eds.), *Landslide Science and Practice: Volume 7: Social and Economic Impact and Policies* (pp. 171-176). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31313-4_22

Trigila, A. (2005). *Metodologia di lavoro e struttura della banca dati*.

Trigila, A., Iadanza, C., Guerrieri, L., & Hervás, J. (2007). The IFFI project (Italian landslide inventory): Methodology and results. *Guidelines for mapping areas at risk of landslides in Europe*, *23*, 15.

Van Den Eeckhaut, M., & Francisco, H. D. D. (2012). Landslide inventories in Europe and policy recommendations for their interoperability and harmonisation-A JRC contribution to the EU-FP7 SafeLand project.

van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology*, *102*(3), 112-131. https://doi.org/https://doi.org/10.1016/j.enggeo.2008.03.010

Wang, X.-W., & Liu, Y.-Y. (2020). Comparative study of classifiers for human microbiome data. *Medicine in Microecology*, *4*, 100013. https://doi.org/https://doi.org/10.1016/j.medmic.2020.100013

Wu, B. S., Chuang, R. Y., Chen, Y.-C., & Lin, Y.-S. (2022). Characteristics of landslides triggered by the 2013 ML6.5 Nantou, Taiwan, earthquake. *Earth, Planets and Space*, *74*(1), 7. https://doi.org/10.1186/s40623-021-01560-8

Xing, Y., Yue, J., Guo, Z., Chen, Y., Hu, J., & Travé, A. (2021). Large-Scale Landslide Susceptibility Mapping Using an Integrated Machine Learning Model: A Case Study in the Lvliang

Mountains of China [Original Research]. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.722491

Yilmaz, I. (2010). The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks. *Environmental Earth Sciences*, *60*(3), 505-519. https://doi.org/10.1007/s12665-009-0191-5

Zêzere, J. L. (2002). Landslide susceptibility assessment considering landslide typology. A case study in the area north of Lisbon (Portugal). *Nat. Hazards Earth Syst. Sci.*, *2*(1/2), 73-82. https://doi.org/10.5194/nhess-2-73-2002

Zêzere, J. L. s., de Brum Ferreira, A., & Rodrigues, M. L. s. (1999). The role of conditioning and triggering factors in the occurrence of landslides: a case study in the area north of Lisbon (Portugal). *Geomorphology*, *30*(1), 133-146. https://doi.org/https://doi.org/10.1016/S0169-555X(99)00050-1

Appendix

A1: Hyperparameters used in the analysis along with a brief description of their function. Two separate sets of hyperparameters were used for the XGB model, and a single set for the RF model. Parameters not included in the table were set to default. Descriptions derived from Pedregosa et al. (2011) and Chen and Guestrin (2016).

XGB binary hyperparameters		
tree_method	"gpu_hist"	Specifies the algorithm used to build decision trees in XGB
enable_categorical	True	Enables categorical features in the dataset.
learning_rate	0.184	Controls the step size at each iteration during gradient boosting
n_estimators	200.000	Number of trees used
'colsample_bytree'	0.741	Determines the fraction of features (columns) to be randomly sampled for each tree in the gradient boosting process
'gamma'	1.805	The gamma hyperparameter controls the minimum reduction in the loss function required to make a split in a decision tree during the training process of gradient boosting models.
'max_depth'	17.000	Specifies the maximum depth or the maximum number of levels allowed in each decision tree.
'min_child_weight'	1.000	Determines the minimum sum of instance weights required to create a new child node during tree growth
'reg_alpha'	0.970	The L1 regularization term added to the loss function during training, also known as Lasso regression
'reg_lambda'	0.602	The L2 regularization term added to the loss function during training, also known as Ridge regression.
'subsample'	1.000	Specifies the fraction of training instances (rows) to be randomly sampled for each tree
objective	"binary: Logistic"	The objective hyperparameter defines the loss function to be optimized

XGB multiclass hyperparameters		
tree_method	"gpu_hist"	Specifies the algorithm used to build decision trees in XGB
enable_categorical	True	Enables categorical features in the dataset.
'learning_rate'	0.175	Controls the step size at each iteration during gradient boosting
'n_estimators'	350.000	Number of trees used
'colsample_bytree'	0.795	Determines the fraction of features (columns) to be randomly sampled for each tree in the gradient boosting process
'gamma'	1.198	The gamma hyperparameter controls the minimum reduction in the loss function required to make a split in a decision tree during the training process of gradient boosting models.
'max_depth'	1.000	Specifies the maximum depth or the maximum number of levels allowed in each decision tree.
'min_child_weight'	3.000	Determines the minimum sum of instance weights required to create a new child node during tree growth
'reg_alpha'	0.254	The L1 regularization term added to the loss function during training.
'reg_lambda'	0.344	The L2 regularization term added to the loss function during training.
'subsample'	0.969	Specifies the fraction of training instances (rows) to be randomly sampled for each tree
objective'	multi:softmax'	The objective hyperparameter defines the loss function to be optimized

RF hyperparemeters		
'max_depth'	15	Specifies the maximum depth or the maximum number of levels allowed in each decision tree.
'min_samples_leaf'	1	Sets the minimum number of samples required to be at a leaf node
'min_samples_split'	2	Determines the minimum number of samples required to split an internal node
n_estimators'	260	Number of trees used
min_weight_fraction_leaf'	0	Sets the minimum weighted fraction of samples required to be at a leaf node, similar to 'min_samples_leaf'
max_features'	"sqrt"	Determines the maximum number of features considered for splitting at each decision tree node
max_leaf_nodes'	None	Specifies the maximum number of leaf nodes allowed in a decision tree