POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA CIVILE,
AMBIENTALE E TERRITORIALE

Master of science in Geoinformatics Engineering

Dipartimento di Ingegneria Civile e Ambientale

Landslide susceptibility mapping using ensemble methods: a case study in Lombardy, Northern Italy

Supervisor: Prof. Maria Antonia Brovelli
Co-supervisor: Dr. Vasil Yordanov

Thesis by: Xu Qiongjie
Student ID: 10757496

Academic year 2022 - 2023

Contents

Abstract

Landslide hazards are pervasive throughout the world, and landslide susceptibility mapping (LSM) provides crucial information that aids various authorities in managing landslide-prone areas. In this study, the performance of various ensemble methods—including stacking, blending, and soft voting—was evaluated by utilizing the best fundamental classifiers for LSM in Lombardy, Northern Italy. The first step is to create a spatial database with 11 landslide influencing elements and historical landslide records. Second, five fundamental classifiers (Bagging, Random Forests, AdaBoost, Gradient Tree Boosting, and Neural Networks) are built at a basin level (i.e., Val Tartano, Upper Valtellina, and Val Chiavenna) and later transferred at a regional level for Lombardy. Third, three ensemble models are created by combining the fundamental classifiers with the greatest ability to generalize and test again on local basin scale. Further, the constructed ensemble models are assessed in the same basins as well as Lombardy. Four, the best model is selected as the final model to produce the LSM for Lombardy. The fundamental classifiers for ensemble models are chosen to be Random Forest, AdaBoost, and Neural Networks as they outperformed other models in terms of generalization. The soft voting model exhibits the best generalization performance when compared to the other ensemble models. The final model to generate the LSM for Lombardy is a Neural Network model (accuracy=0.93 in whole Lombardy) that was trained using data gathered from three basins and performs the best in Lombardy. According to the landslide susceptibility maps created, roughly 37% of the entire Lombardy region fell into the "very high" and "high" categories.

KEYWORDS

Landslide susceptibility mapping; Ensemble methods; Random Forests; AdaBoost; Bagging; Neural Network;

1. Introduction

Landslides are geological events characterized by the downslope movement of various materials, including rock masses, earth materials, or debris (Cruden, 1991). These movements can be induced by several factors, such as excessive precipitation, earthquakes, human activities, or volcanic eruptions (Turner, 2018). Landslides are considered one of the primary natural hazards (Yordanov and Brovelli, 2021), leading to numerous casualties and significant economic losses worldwide (Budimir et al., 2015). For this reason, numerous past and current research endeavors focus on this subject. Identifying landslide-prone areas is essential for policy-makers, scientists, engineers, and the general public to prevent catastrophic landslides. A landslide susceptibility map (LSM) can be used for this purpose, representing the spatial likelihood levels of a specific area being prone (susceptible) to mass movements based on environmental conditions. The foundational principles and assumptions of the landslide susceptibility maps are that future slope instabilities are more probable to occur under the same conditions that led to past and current instabilities and that the occurrence of landslides in space can be inferred through heuristic investigations, environmental information analysis, or physical modeling. Additionally, the modeling of landslide occurrence can be constructed by considering the unstable factors related to slope failure (Guzzetti et al., 1999).

Research methods for landslide susceptibility assignment can be divided into two primary categories (Guzzetti et al., 1999). The first category comprises qualitative methods that are subjective and employ descriptive (qualitative) terminology to express sensitivity levels. The second category comprises quantitative methods that generate numerical estimations, specifically probabilities of landslide occurrence in the studied region.

Reichenbach et al. (2018) further classified approaches and techniques suitable for assigning landslide susceptibility into five categories:

- Geomorphological mapping that relies on the expertise and experience of an investigator to assess and map the actual and potential slope instability conditions. The quality of the mapping also depends on the complexity of the study area.
- Analysis of landslide inventories that can be used for modeling future spatial occurrences of landslides based on the known distribution of past and present landslides. The quality of the predictions depends on the completeness and quality of the inventories.
- Heuristic or index-based approaches that rank and weigh known instability factors based on the investigators' expected or assumed importance in causing landslides. The quality of the predictions primarily depends on the investigators' understanding of the real causes and the instability factors causing landslides in a given area.
- Process-based methods that analyze the stability and instability conditions using simple limit equilibrium models, such as the "infinite slope stability" model, based on simplified, physically-based landslide modeling schemes.
- Statistically-based modeling methods that analyze the functional relationships between known or inferred instability factors and the distribution of past and present landslides. The most widely used statistical techniques for generating maps of landslide susceptibility encompass logistic regression, neural network analysis, data

overlay, index-based approaches, and weight of evidence analyses with a growing trend toward employing machine learning (ML) methods recently.

The objective of the research outlined in this thesis is to employ different ensemble methods to generate landslide susceptibility maps by utilizing the most suitable base ML estimators and to validate the results produced by the final model. Initially, this approach is implemented and evaluated at the basin level in three areas (i.e., Val Tartano, Upper Valtellina, and Val Chiavenna) located in Northern Lombardy. Subsequently, the methodology is extended to cover the entire region of Lombardy.

1.1 Thesis outline

The thesis is organized in the following manner:

Chapter 1 presents an introduction to the work.

Chapter 2 offers an overview of state-of-the-art techniques and previous studies on landslide susceptibility analysis. It outlines the methodology employed and explains the choices made.

Chapter 3 describes the areas of interest and their geomorphic features in detail to provide a comprehensive study framework. Additionally, it is included a description of the collected data and the preparatory steps applied to the dataset.

Chapter 4 introduces the software tools and their applications. It also offers a theoretical explanation of the analytical techniques employed in this study.

Chapter 5 provides a detailed account of the workflow, including the data preprocessing procedure and the scenarios for processing.

Chapter 6 illustrates the outcomes of the study, including the LSM and its evaluation.

Chapter 7 presents the conclusions derived from the study.

2. State of the art

The purpose of this section is to review state-of-the-art techniques concerning the use of machine learning (ML) models in generating LSM.

2.1 Workflow

In the field of landslide susceptibility mapping using ML models, the majority of studies have adopted a workflow comprising four steps as depicted in Figure 2.1. These steps are as follows:

1. Data preparation involves creating a geospatial database to be used for model building. During this process, the relevant influencing factors and landslide inventory for a study area must be selected and preprocessed. Moreover, a mapping unit (such as pixels, slope units, or unique condition units) must be determined.
2. Model building typically involves the selection and application of the modeling methods employed. The selection of a modeling method is influenced by various factors, such as the availability and quality of data, the complexity of the system under investigation, and the level of accuracy and precision desired in the model output. Once a suitable modeling method has been chosen, the model-building process may include several stages, such as implementing the model, training it using training data sampled from the obtained database in the former step, and selecting and adjusting hyperparameters to optimize performance.
3. Model validation evaluates the performance of the building model using different indices and metrics based on the testing dataset sampled from the obtained database. Reichenbach et al. (2018) summarized four types of evaluation criteria:
    a. Model validation verifies the relationships between dependent and independent variables using the validation dataset.
    b. Model fitting performance is assessed by comparing the model predictions with the training dataset. This evaluation can be carried out using metrics such as success rate curve, landslide density or frequency, and Receiver Operating Characteristic (ROC).
    c. Model prediction performance is similar to the evaluation of model fitting performance but based on the comparison of the model's predictions with the testing dataset. The most common metric used is the prediction rate curve, landslide density or frequency, and Receiver Operating Characteristic (ROC).
    d. Model uncertainty assessments are the least common.

4. Mapping can serve as an output and evaluation step following the model-building process. Alternatively, it may occur after model evaluation, with maps generated based on the models exhibiting the most favorable performance. The resultant map is then divided into various susceptibility classes to produce a meaningful susceptibility map.

Figure 2.1: common workflow for landslide susceptibility mapping

2.2 Review of machine learning methods for susceptibility mapping

2.2.1 Used factors

The aim of utilizing an ML model for landslide susceptibility mapping (LSM) is to identify a correlation between landslide occurrence and predisposing factors. Subsequently, these relationships are employed to create LSM for determining the areas prone to the hazard.

The selection of factors used for the generation of landslide susceptibility maps (LSM) varies depending on data availability, study area, local setting, and the landslide types to be modeled. Ado et al. (2022) classified popular landslide influencing factors into five groups, namely topography (e.g., slope, aspect, elevation, plan curvature, profile curvature, sediment transport index), hydrology (e.g., rainfall, solar radiation, stream power index, topographic wetness index, distance to rivers, density of the river), geological (e.g., lithology, distance to faults, density of fault), land use/cover (e.g., land use/land cover, normalized difference vegetation index), and man-made factors (e.g., distance to roads and density of the road). Commonly used factors in LSM include slope, elevation, rainfall, distance to rivers, Land Use/Land Cover (LULC), Normalized Difference Vegetation Index (NDVI), and distance to roads.

However, there is no universal standard for selecting influencing factors. To improve the performance of ML models, many studies have integrated feature selection techniques in their workflow to filter out irrelevant factors. For instance, Micheletti et al. (2014) explored three ML models, namely adaptive Support Vector Machines, Random Forests, and AdaBoost, to perform feature selection for LSM. Wang et al. (2020) applied Random Forest methods and the Pearson correlation coefficient (PCC) to select landslide-affecting factors. Yordanov and Brovelli (2021) refined the factors by considering the weights of the input data derived from a statistical index.

2.2.2 Employed models

Ado et al.(2022) have classified ML models used for landslide susceptibility mapping into four groups: conventional, hybrid, ensemble, and deep learning methods. Conventional models are standalone models that can be served as a benchmark for evaluating new models or combined with other models in hybrid or ensemble methods. The most popular

conventional methods include Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), and Artificial Neural Network (ANN).

In a comparative study by Yordanov and Brovelli (2021), Statistical Index, LR, and RF were evaluated for their effectiveness in generating susceptibility maps. Using 11 predefined terrain variables and one precipitation variable, the study produced 79 susceptibility maps with varying ratios between training and validation datasets. The input data was refined to enhance model performance, and the results indicated that RF and LR are reliable modeling approaches. However, RF outperformed LR in most of cases.

Amici (2021) explored the use of RF for generating LSM in additional areas with 11 influencing factors and redefined the previously proposed No Landslide zone based on geological criteria. According to the study's results, The RF method demonstrated satisfactory accuracy with an AUC value of 0.97.

In another study, Yilmaz (2010) compared conditional probability (CP), LR, ANNs, and SVM for generating LSM using an 11-factor spatial database. The results showed that ANN had the highest accuracy, with an AUC value of 0.846.

Hybrid techniques, which integrate feature selection and optimization methods with ML models, have demonstrated significant performance improvements compared to conventional ML methods due to their ability to address feature selection challenges (Wang et al., 2020; Ado et al., 2022).

Zhao et al. (2021) investigated the performance of five models, including a traditional statistical Certainty Factor (CF) model, Support Vector Machine (SVM), Random Forest (RF), and two hybrid models: CF-SVM and CF-RF. The researchers selected 10 influencing factors and used slope units as the basic mapping units. The SVM model, utilizing the Gaussian radial basis kernel function, achieved a prediction success rate of 71%, while the RF model outperformed the SVM model with a prediction success rate of 78%. The hybrid models, constructed by applying SVM or RF on the sample dataset generated by calculating the CF values of the original training samples, showed further improvement. Specifically, the CF-RF model achieved a prediction success rate of 81%, slightly outperforming the CF-SVM model with a prediction success rate of 77.5%.

Ensemble techniques combine multiple conventional ML models using different averaging or voting methods to produce more accurate predictions than conventional ML methods (Fang et al., 2021; Li et al., 2022), compensating for the limitations or biases of individual algorithms with others (Ado et al., 2022).

Kumar et al. (2023) explored the use of ensemble methods for regional landslide susceptibility modeling (LSM). The researchers evaluated the performance of ten individual ML models, including linear discriminant analysis, mixture discriminant analysis, Bagged Cart, Boosted Logistic Regression, K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), Rotation Forest, and C5.0, using different sets of landslide influencing factors ranked by an ensemble feature selection method. Subsequently, different ensemble ML models were developed and evaluated based

on the suitable combination of individual ML models. The results indicated that the ensemble models of KNN+RTF, KNN+ANN, and ANN+RTF performed the best when developed using the top five landslide influencing factors.

Deep learning methods are representation-learning techniques with multiple layers of representation that can accurately predict LSM with less uncertainty (Wang et al., 2019; Thi Ngo et al., 2021; Habumugisha et al., 2022). However, they may have low model variance and limited generalization capabilities, which can be addressed by using hybrid and ensemble setups (Kavzoglu et al., 2021). Besides, the combination of hybrid and ensemble methods with DL models could also increase prediction accuracy (Azarafza et al., 2021; Li et al., 2022).

2.2.3 Validation metrics

The validation of the model is a crucial step of the modeling process. A common evaluation method used for classification problems in ML is constructing the confusion matrix using a testing dataset that model has not been feed with during the training phase (Table 2.1).

	Actually Positive	Actually Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Table 2.1. Confusion Matrix

where True Positive (TP) and False Positive (FP) are the number of samples correctly classified and misclassified as positive, respectively, while True Negative (TN) and False Negative (FN) are the number of samples correctly classified and misclassified as negative, respectively.

From the confusion matrix, the following statistical indexes can be derived and used for evaluation.

$$Overall\ Accuracy\ (OA) = \frac{TP + TN}{TP + FP + TN + FP}$$

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

$$Recall\ (Sensitivity) = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$FPR = \frac{FP}{FP + TN}$$

$$TNR\ (Specificity)\ =\ \frac{TN}{TN+FP}$$

Besides, Receiver Operating Characteristic (ROC) curve (Bradley, 1997) and Precision-Recall Curve (PRC) are commonly adopted as accuracy statistics for landslide susceptibility modeling in related studies (Park et al., 2013; Lin et al., 2017; Pourghasemi and Rahmati, 2018; Yordanov and Brovelli, 2021; Amici, 2021).

The ROC curve is constructed by plotting the true positive rate (TPR, also known as recall) against the false positive rate (FPR) at various thresholds.

The area under the curve of ROC (AUC-ROC) is widely adopted for model comparison and considered the best evaluation metric (Ado et al., 2022). This value varies between 0 and 1, where 0.5 indicates the classifier is uninformative whilst 1 represents perfect performance.

An optimum ROC-based threshold providing the best trade-off between sensitivity and specificity can be determined by maximizing Youden's J-statistic, a statistic that captures the performance of dichotomous diagnostic tests and can be calculated using the formula below

$$J\ =\ sensitivity\ +\ specificity\ -\ 1$$

PRC is created by plotting the precision against the recall at various threshold settings. Similar to the ROC curve, the area under the curve of PR (AUC-PR) can be computed to indicate the model performance with a random classifier has a value close to 0.5.

Another optimum PR-based threshold considering both precision and recall can be determined by maximizing the F_1 score since an F_1 score value of 1.0 indicates perfect precision and recall.

3. Areas of interest and data

3.1 Areas of interest

Several small basins within the Italian Lombardy region were selected for training, testing, and validation of the model as shown in Figure 3.1. These models were later applied to a bigger area, i.e., Lombardy, to generate landslide susceptibility maps.

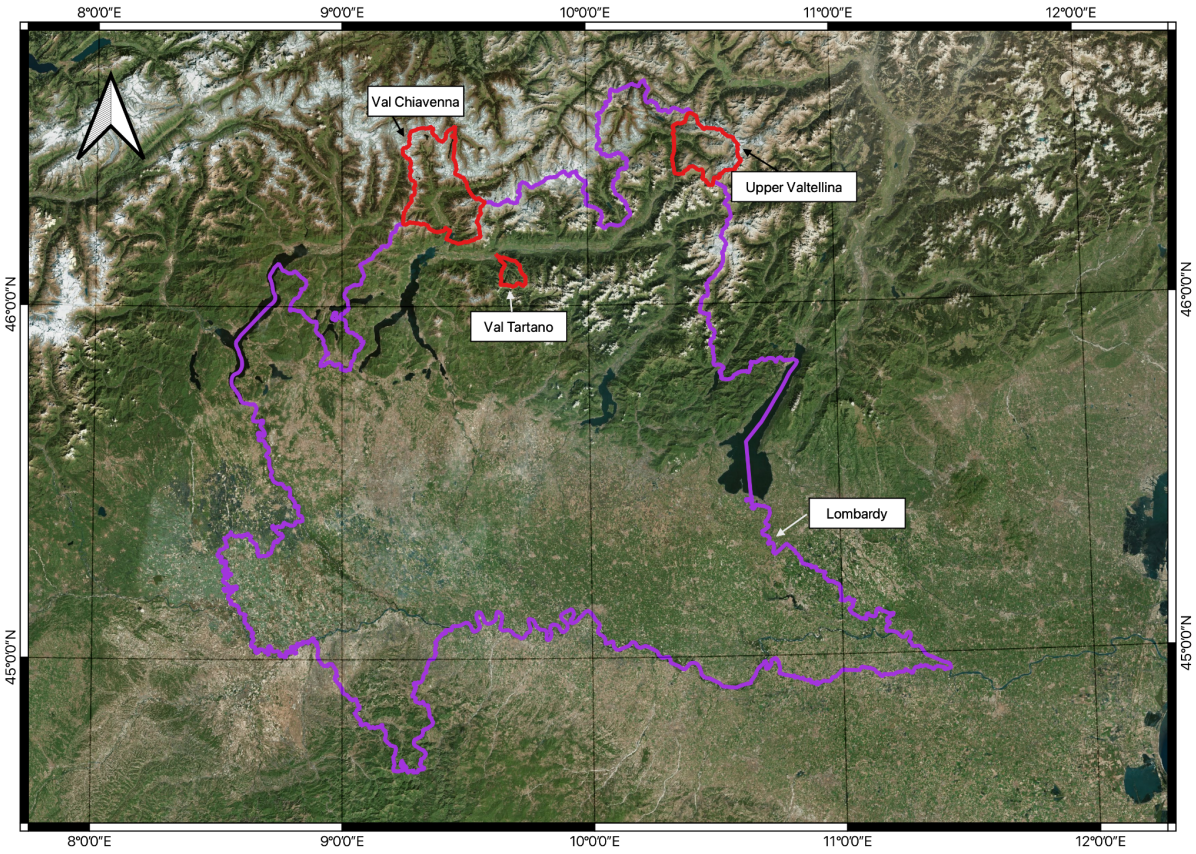


Figure 3.1: The four areas of interest

3.1.1 Lombardy

Lombardy is an administrative region that spans 23,863 km^2 and is situated in the northern central part of Italy between the Alps mountain range and the flow of the Po river. This region boasts a populace of roughly 10 million individuals with a population density of 420.2 inhabits/km^2 (Istat (Italian National Institute of Statistics), 2020). Lombardy is composed of 11 provinces, 1 metropolitan area (Milan), and 1523 municipalities, with Milan serving as its capital city, according to Regione Lombardia (2020).

Lombardy's varied landscape can be broadly categorized into three sections: the northern mountainous region of the Alps, the central foothills, and the southern Lombard portion of the Bataan Plain. This rich topography has a profound impact on Lombardy's climate, which is predominantly humid subtropical, particularly in the plains. In the Alpine foothills, the climate is oceanic, while the climate in the hills and mountains is humid continental.

The Lombardy region has 141,970 landslides recorded in the most recent version of the IFFI database (ISPRA, 2014). Among them, the Oltrep Pavese area and the northernmost portion of the territory have the highest density of the phenomenon (Antonielli et al., 2019). Figure 3 summarizes the frequency of various landslide types in Lombardy. Rapid debris flows, which account for 41.6% of all landslides in Lombardy, the area affected by numerous rockfalls/topples, which account for 29.7%, and rotational/translational slides, which account for 15.2%, are the main types of movement.

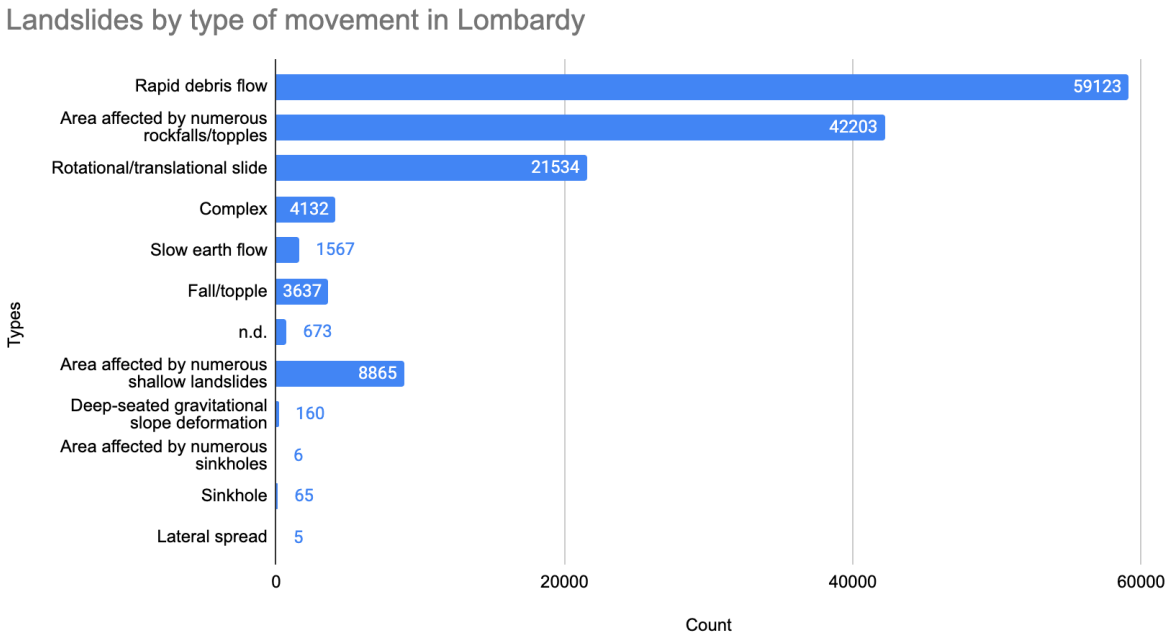


Figure 3.2: Landslides type in Lombardy Region (Italy). Data from ISPRA, in the framework of IFFI project.

3.1.2 Val Tartano

The Tartano Valley, situated in the province of Sondrio in Lombardy, Italy (46.1075° N, 9.6791° E), is an Alpine catchment area characterized by an Alpine continental climate. It covers an area of 51 km^2 and has an elevation range of 250 to 2250 m a.s.l. Due to its distinctive hydrogeological characteristics, the Tartano Valley has piqued the interest of numerous authors (Colombera and Bersezio, 2011; Longoni et al., 2016; Yordanov and Brovelli, 2021; Amici, 2021).

This valley has experienced several flood events, including a catastrophic one in 1987, caused by heavy rains and snowmelt that triggered flooding, mudslides, and mass movements, leading to 20 fatalities and significant damage to river bank protection. Additionally, the landslide inventory of ISPRA (2014) documents over 1000 landslide events in this basin. The most frequent type of landslide (about 51.2%) in Val Tartano is rapid debris flow, as displayed in Figure 3.3.

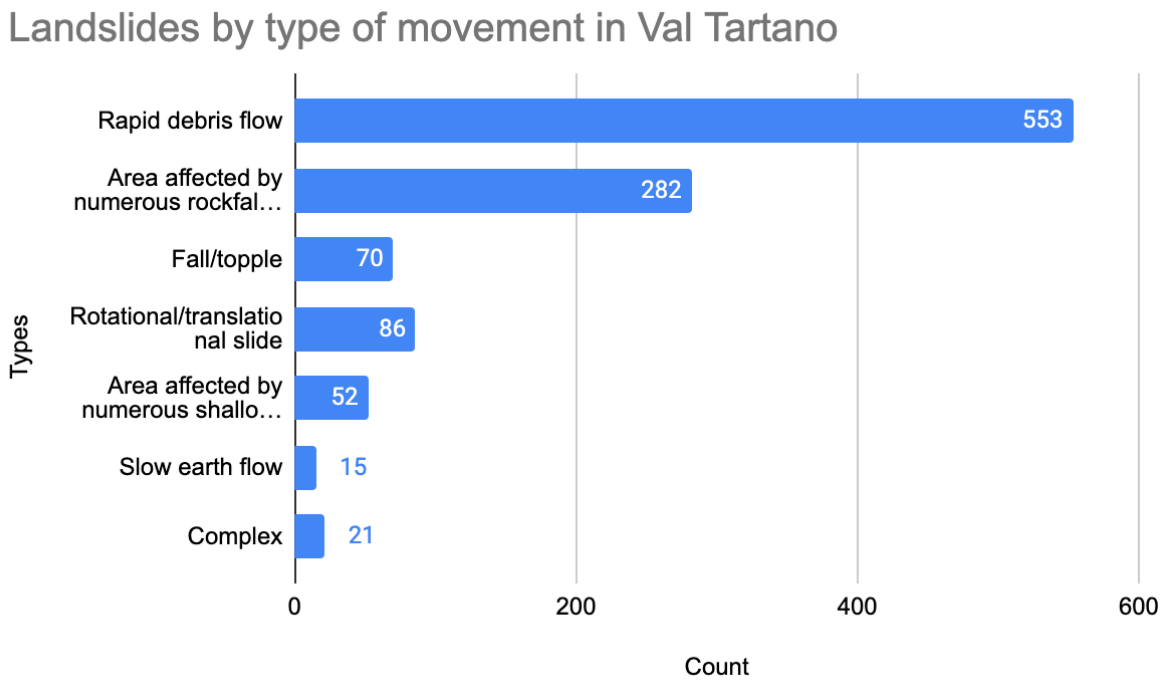


Figure 3.3: Landslides type in Val Tartano. Data from ISPRA, in the framework of IFFI project.

3.1.3 Upper Valtellina

The Upper Valtellina Valley, located close to the border of Trentino-Alto Adige, comprises the second region of interest. Spanning an area of 295 km^2 and an elevation range of 900 to 3800 m a.s.l., this area experiences frequent instabilities due to decompression caused by melting glaciers. According to ISPRA (2014), over 3600 individual landslide events have been recorded in this region, and the main category of landslides is rapid debris flow which is about 41% of the total landslides (Figure 3.4).

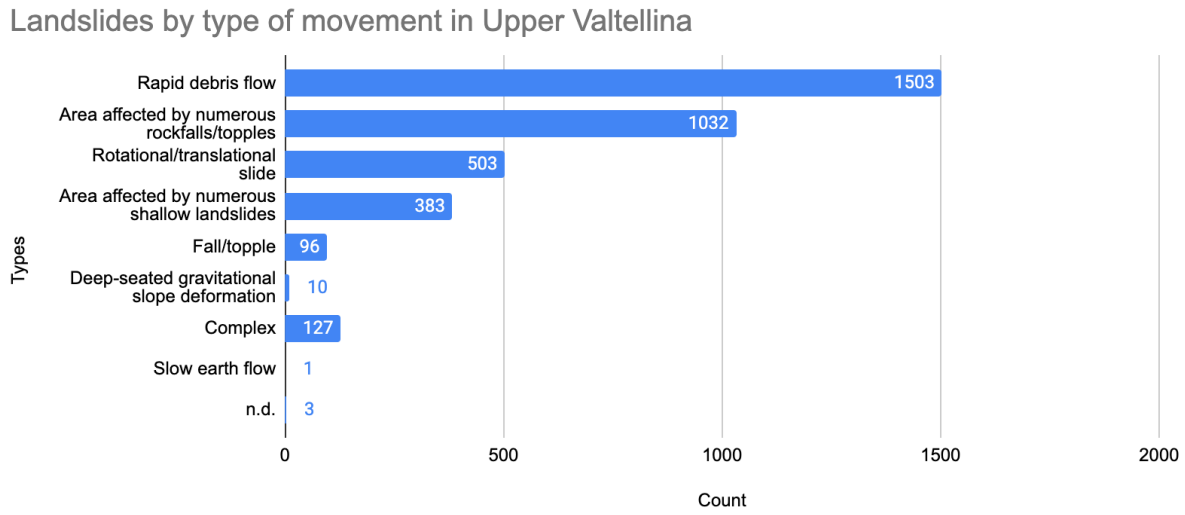


Figure 3.4: Landslides type in Upper Valtellina. Data from ISPRA, in the framework of IFFI project.

3.1.4 Val Chiavenna

The Chiavenna Valley, located in the province of Sondrio, is an alpine valley covering an area of 578 km^2, comprising two orthogonal branches: the San Giacomo valley, oriented north to south, and the Bregaglia valley, oriented east to west. These two branches converge near Chiavenna, forming a broader main valley running from north to south. Within these valleys, cliffs, cirques, and narrow tributary valleys can be found. The Mera River flows southward through the Bregaglia Valley and Valchiavenna before emptying into Lake Como at the confluence of the Valtelline and Adda Rivers, the main tributaries of Lake Como. The complex geography of the valley has attracted the interest of several scholars (Bajni et al., 2021; Tantardini et al., 2022).

Figure 3.5 illustrates the types of landslides that occurred in Val Chiavenna, with data from the landslide inventory of ISPRA (2014), where fast debris flows were the main type of landslides, accounting for 51.6% of the total number of landslides.

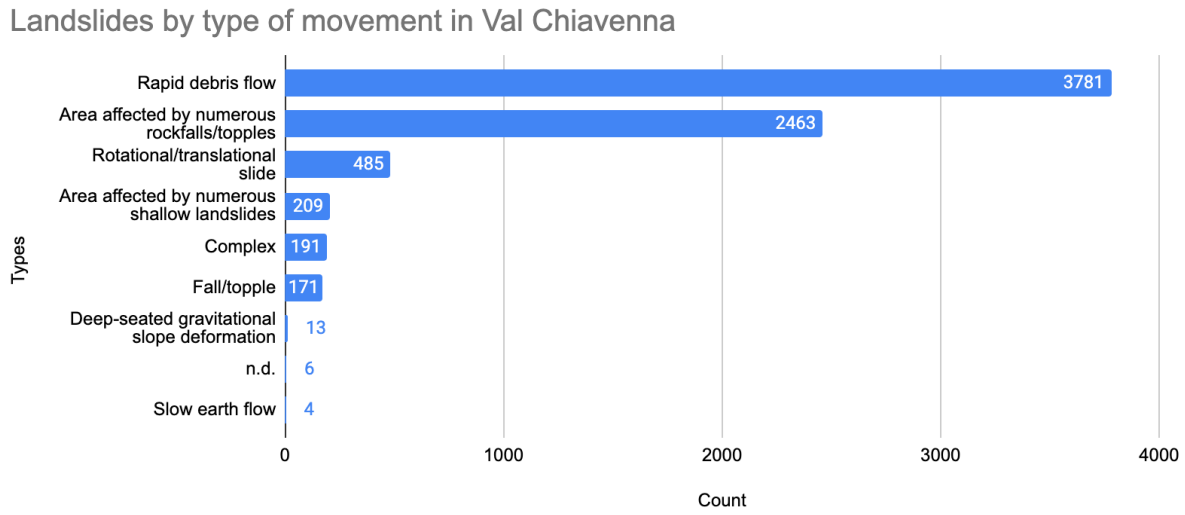


Figure 3.5: Landslides type in Val Chiavenna. Data from ISPRA, in the framework of IFFI project.

3.2 Data

It is essential to select, explore, download, and organize data before analyzing. In this section, the data preparation steps are described and the final selected data are presented.

3.2.1 Data Preparation

The data preparation process usually consists of three parts, data acquisition, data exploration, and data preprocessing. Data acquisition performs data retrieval and obtains corresponding data with specific spatiotemporal attributes according to research needs. The two primary types of data are raster and vector data. Data exploration is to visualize the data through a graphical interface, while the purpose of data preprocessing is to transform the collected data into a form more suitable for research purposes. This study adopts the following geospatial operations in data preprocessing:

- **Clipping** which clips the raw data layers to the areas of interest;
- **Reprojection** which changes the data reference system. This operation is employed when the raw data layers are not consistent in the Coordinate Reference System (CRS). In this study, *WGS 84 / UTM zone 32 N (EPSG:32632)*, the official reference system for Northern Italy, is selected as the coordinate system.
- **Styling** that modifies the display form of layers to make them more readable, clear, and understandable.

This study employed QGIS, a free and open-source Geographic Information System, for data exploration and data preprocessing.

3.2.2 Data Overview

The data required in this thesis is free and open source and can be accessed from the web illustrated in Table 3.1.

Data Source	License
GeoPortale Lombardia (https://www.geoportale.regione.lombardia.it/)	IODL (Italian Open Data License) 2.0. It allows the user to use, modify and share the data with the obligation to cite the original source.
Istituto Superiore per la Protezione e la Ricerca Ambientale IdroGEO (https://idrogeo.isprambiente.it/)	CC BY-SA 4.0 (Attribution-ShareAlike 4.0 International CC BY-SA): It allows the user to share and adapt the data with the obligation of attributing and sharing-alike any work.
Agenzia Regionale per la Protezione Ambientale ARPA Lombardia (https://www.arpalombardia.it/)	CC BY 4.0 (Attribution-ShareAlike 4.0 International CC BY): It allows the user to share and adapt the data with the obligation of attributing any work.
GeoNode (https://www.geolmiv-geonode.polimi.it/)	–

Table 3.1: Data source.

3.2.2.1 Vector data

A vector in a geoinformatics context is a coordinated-based data model that presents discrete geographic features. There are three vector data types: points, lines, and polygons. In this study, the GeoPackage (GPKG) format, a platform-independent lightweight database container, is used. Table 3.2 shows all the vector data used in this work.

Data	Type	Source	Description
Area of Interest	Polygon	Custom made	This layer contains the boundaries of the area of interest, which can be used to clip other datasets.
Land Use and Land Cover	Polygon	GeoNode - Val Tartano, Upper Valtellina; GeoPortale Lombardia - Val Chiavenna, Lombardy	This layer contains territory classification with respect to the main categories.
Lithology	Polygon		This layer contains the lithology classification

Fault Lines	Line		This layer contains fault lines.
River Network	Polygon		This layer shows areas that contain water and are part of stream beds. The stability of slopes can be significantly affected by water drainage systems due to their ability to erode the slope. In the case of LSM, the slopes can be categorized into buffer zones, which aid in assessing the impact of the drainage system on slope stability (Ado et al., 2022).
Road Network	Polygon		This layer shows the road network.
Landslide Inventory	Multiple Type	GeoNode - Val Tartano, Upper Valtellina; IFFI (Inventario dei Fenomeni Franosi in Italia) (https://idrogeo.isprambiente.it/app/page/open-data) - Val Chiavenna, Lombardy	This layer contains the records of the landslides for the corresponding area.

Table 3.2: Vector data

The landslide inventory provided by IFFI is the national and official database on landslides with a scale of 1:10,000 online since 2005. It was carried out by *Institute for Environmental Protection and Research* (*ISPRA*) in cooperation with the regions and autonomous provinces and provides a detailed description of the distribution of landslides in Italy. According to the spatial allocation of existing landslide phenomena, the inventory divided the data into five types:

- **Landslide Identification Points**: include georeferenced points located at the highest point of the landslide crown.
- **Linear landslides**: consist of landslides whose width can not be mapped, for example, rapid debris flows.
- **Polygonal landslides**: contain landslides whose surface can be mapped at the adopted survey scale. The following definition of landslide types is based on the multiple landslide classifications developed by Cruden and Varnes (1996) and Hungr et al.(2001).
    - Rockfall: a very rapid movement of detached materials (soil or rock).
    - Rotational/translational slide: A slide is a downslope movement of mass, mainly on a ruptured surface or a relatively thin area of intense shear strain. The surface of rotational slides is curved and concave, whilst the surface of translational slides is planar or undulating.

    - Debris flow: a very rapid to extremely rapid (>5m/s) flow of saturated non-plastic debris in a steep channel (with an angle between 20 and 45 degrees) that was caused by high-intensity rainfall.
    - Complex: a combination of falls, topples, slides, spreads, and flows.
- **Landslide areas**: include areas with numerous landslide phenomena.
- **Deep-seated Gravitational Slope Deformations (DGSDs)**: slowly developing phenomena of large ground deformations that are very common in alpine regions. Their long-term evolution and continued deformation rates may represent natural hazards capable of endangering various human structures and infrastructures.

3.2.2.2 Raster data

A raster is made up of equally sized pixels (also referred to as grid cells) and is often used to present continuous data. Raster data utilized in this study are summarized in Table 3.3.

Data	Resolution [m]	Source
Digital Terrain Model (DTM)	5x5	GeoNode - Val Tartano, Upper Valtellina; GeoPortale Lombardia - Val Chiavenna, Lombardy;
Slope	5x5	GeoNode - Val Tartano, Upper Valtellina; Computed - Val Chiavenna, Lombardy;
Aspect	5x5	
Eastness	5x5	
Northness	5x5	
Plan curvature	5x5	
Profile curvature	5x5	
Topographic Wetness Index (TWI)	5x5	
Normalized Difference Vegetation Index (NDVI)	5x5	
Precipitation	1500x1500	ARPA Lombardia

Table 3.3: Raster data

Digital Terrain Model (DTM): the 2015 Digital Terrain Model for the Lombardy region is used. Based on this raster, the following raster data were computed.

- **Slope**: a map that represents the change rate of elevation for each DTM pixel that can be used to measure the steepness of the surface. The slope angle is a crucial parameter in evaluating the resilience of a slope, as it is directly linked to the likelihood of landslide events (Yordanov and Brovelli, 2021).

- **Aspect**: a map that shows the orientation of the slope. It is measured clockwise in degrees with a range between 0° and 360°, where 0° is north-facing, 90° is east-facing, 180° is south-facing, and 270° is west-facing.
- **Eastness**: a map that stresses the east-west gradient, where 1 means a slope facing the East direction while -1 means the West direction. It can be calculated using the formula
  $Eastness = sin(Aspect)$

- **Northness**: a map that indicates the north-south gradient, where 1 means a slope facing the North direction while -1 means the South direction. It can be calculated using the formula
  $Northess = cos(Aspect)$

- **Profile curvature**: An indicator that is parallel to the slope and represents the direction of maximum slope. A negative value (Figure 3.6 A) indicates that the cell is convex upward and the flow velocity will decrease. A positive value (Figure 3.6 B) indicates that the cell is concave upwards and the flow velocity will increase. A value of zero indicates that the cell is linear (Figure 3.6 C).

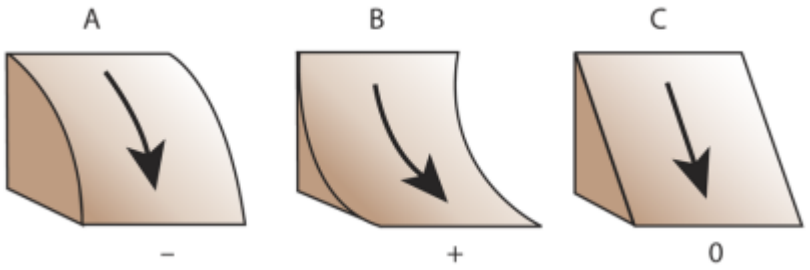


Figure 3.6: Profile curvature

- **Plan curvature**: An indicator that is perpendicular to the direction of maximum slope and relates to the convergence and dispersion of flow across a surface, which influences the characteristics of slope erosion or surface runoff. A positive value (Figure 3.7 A) indicates that the cell is convex laterally. A negative value (Figure 3.7 B) indicates that the cell is concave laterally. A value of zero indicates that the cell is linear (Figure 3.7 C).

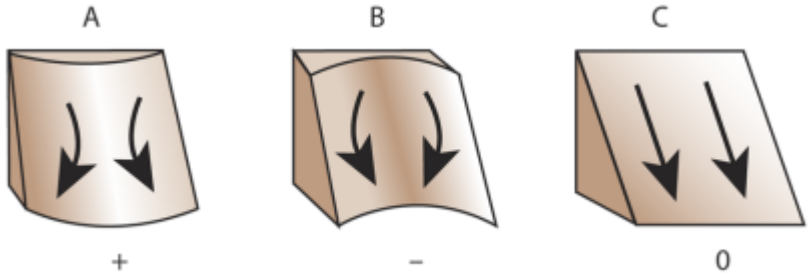


Figure 3.7: Plan curvature

- **Topographic Wetness Index (TWI)**: A hydrological index that is commonly used to quantify topographic control on hydrological processes (Sørensen et al., 2006) and can be a measurement of the degree of water accumulation at a size. It is calculated using the formula

$$TWI = ln(\frac{\alpha}{tan(\beta)})$$

where α is the Specific Catchment Area and $tan(\beta)$ is the slope in radians.

Normalized Difference Vegetation Index (NDVI): a spectral index that is used to quantify vegetation greenness. I can be computed from the satellite images using the formula

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

where NIR and Red are spectral reflectance measured in the Near InfraRed and Red (visible) regions, respectively.

Precipitation: This raster was derived using data that was interpolated from hourly observations made on a 1.5x1.5km grid by the regional meteorological network of ARPA for the Lombardy region. The data was collected in txt format. The information can be averaged to get annual average precipitation data, or the 90th percentile of each grid can be calculated to produce annual 90th percentile precipitation data. The data was afterwards transformed into a raster for further processing.

4. Tools and Techniques

This section focuses on the tools and techniques employed for analysis in this study.

4.1 Software tools

High-quality multispectral images are the primary requirement for the analysis in this study. In addition, software and tools with a high level of robustness and reliability are essential for the consistent processing of images.

Table 4.1 briefly illustrates the software and tools discussed in this section.

Sofware	Purpose	License
QGIS	Data visualization and preprocess	Free and Open Source Software (FOSS). It allows users to freely run, copy, distribute, study, modify and improve the software.
Python	Data preprocessing, model generation, and evaluation	–

Table 4.1: Software tools and programming languages used in this study.

4.1.1 Python

Python is a powerful and friendly open-source programming language (Python, 2023). There are many libraries and frameworks available in Python that can be used for machine learning purposes. For example, NumPy adds support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays (Harris et al., 2020). Matplotlib is a plotting library that allows the creation of visualizations. And the sciki-learn machine learning library (Scikit-learn, 2023) provides various classification, regression, and clustering algorithms and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Python is used throughout the entire research process, including data preparation, model training, model evaluation, and LSM generation.

4.1.2 QGIS

QGIS is a free and open-source cross-platform desktop geographic information system (GIS) application that supports viewing, editing, printing, and analysis of geospatial data (QGIS, 2023). With Python, users can automate tasks on QGIS and extend its functionality.

In this study, QGIS was used together with Python to preprocess the initial data and to view and explore the data at all steps of the analysis.

4.2 Analysis techniques

4.2.1 Base classifiers

4.2.1.1 Bootstrap aggregating (Bagging)

Bootstrap aggregation, also known as bagging, is a machine learning algorithm that improves the accuracy and stability of machine learning algorithms used in statistical classification and regression, as well as reducing variance and avoiding overfitting. Various subsets of the training data are randomly drawn with replacements from the entire training dataset, and each bootstrap sample is used to train a different classifier with the same type. The individual classifiers are then combined by taking a simple majority vote on their decisions to produce the final prediction. (Breiman, 1996)

4.2.1.2 Random Forests

Random Forests (RF) is one of the most widely used machine learning algorithms for classification and regression, which creates multiple decision trees by introducing randomness at the training phase. These decision trees are used to determine the final prediction result by a majority vote.

In practice, the Scikit-learn package implemented in Python was utilized. In contrast to the original publication (Breiman, 2001), the scikit-learn implementation gives the prediction as the averaged prediction of the individual classifiers.

4.2.1.3 AdaBoost

AdaBoost, which stands for adaptive boosting, is a statistical classification meta-algorithm introduced by Yoav Freund and Robert Schapire in 1995. (Freund and Schapire, 1997). The principle of AdaBoost is to fit a sequence of weak learners which are only slightly better than random prediction such that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. Using a weighted majority vote or sum, all the weak learners' predictions are combined to produce the final prediction.

4.2.1.4 Gradient Tree Boosting

Gradient Tree Boosting or Gradient Boosted Decision Trees (GBDT) is a generalization of boosting to arbitrary differentiable loss functions (Friedman, 2001), which can be used for both regression and classification problems. In this algorithm, the model is built in a forward stage-wise fashion and it allows for the optimization of arbitrary differentiable loss functions. In each stage, multiple regression trees are fit on the negative gradient of the loss function.

4.2.1.5 Neural Network

Neural network (NN) algorithms can be used for both classification and regression problems. It starts with the perceptron. A perceptron takes inputs, multiplies them by some weights, and passes them to an activation function to produce an output. By adding the layers of these perceptrons together, a neural network is created, which is known as a multi-layer perceptron (MLP) model. There are three layers of a neural network - the input, hidden, and output layers. The input layer receives data directly, whereas the output layer creates the desired output. The layers in between are known as hidden layers, where the intermediate computation takes place.

4.2.1.6 Logistic Regression

Logistic Regression (LR), also known as logit regression, is a widely used technique in landslide susceptibility modeling (Ayalew and Yamagishi, 2005; Park et al., 2013; Lin et al., 2017). The output prediction of Logistic Regression is defined as follows (King and Zeng, 2001):

$$p = \frac{1}{1+e^{z}}$$

where p is the probability and z denotes the linear combination of variables as follows:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N$$

where $(\beta_0, \beta_1, \beta_2, \ldots, \beta_N)$ are the regression coefficients and $X = X_1, X_2, \ldots, X_N$ represents the explanatory variables.

4.2.2 Ensemble-learning methods

4.2.2.1 Stacking

Stacking is a method for combining estimators to reduce their biases (Wolpert, 1992). The flowchart of the stacking ensemble model is shown in Figure 4.1. It contains the following

steps: (1) using the training dataset to train different base estimators through k-fold cross-validation; (2) the predictions of each estimator are stacked together and used as input to a final estimator; (3) train the final estimator through cross-validation.

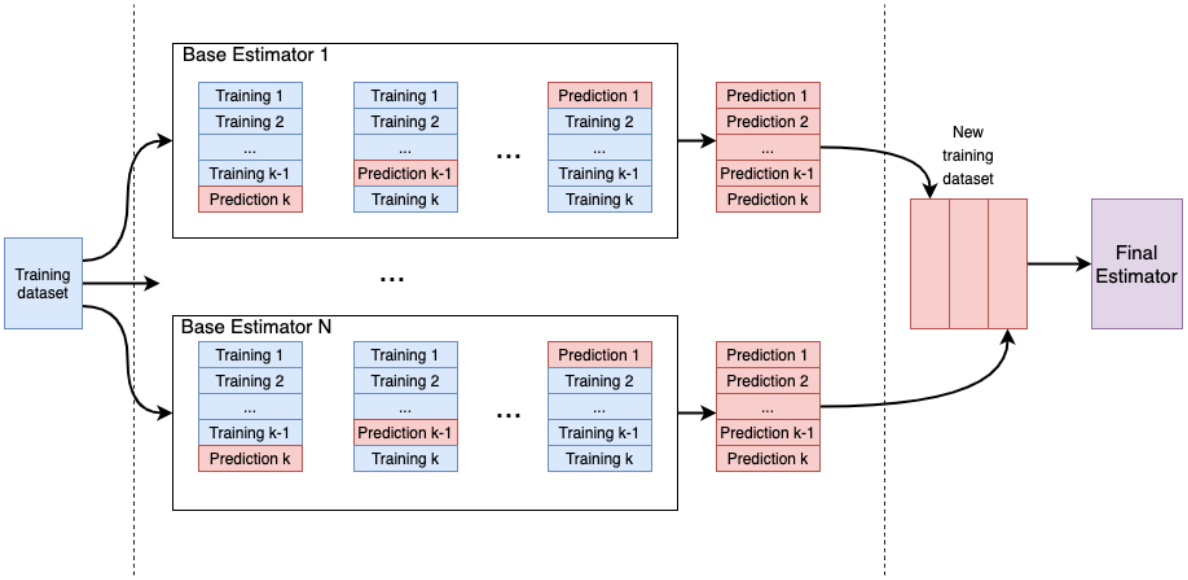


Figure 4.1 Flowchart of stacking ensemble.

4.2.2.2 Blending

A blending ensemble is a variation of the stacking ensemble originally introduced in the Netflix competition (Toscher and Jahrer, 2009). The flowchart of the blending ensemble model is shown in Figure 4.2. It consists of the following steps: (1) split the training dataset into new training and validation datasets by a hold-out method; (2) train all the base estimators using the new training dataset and make predictions of the validation dataset to form training dataset used as input to a final estimator; (3) train the final estimator.

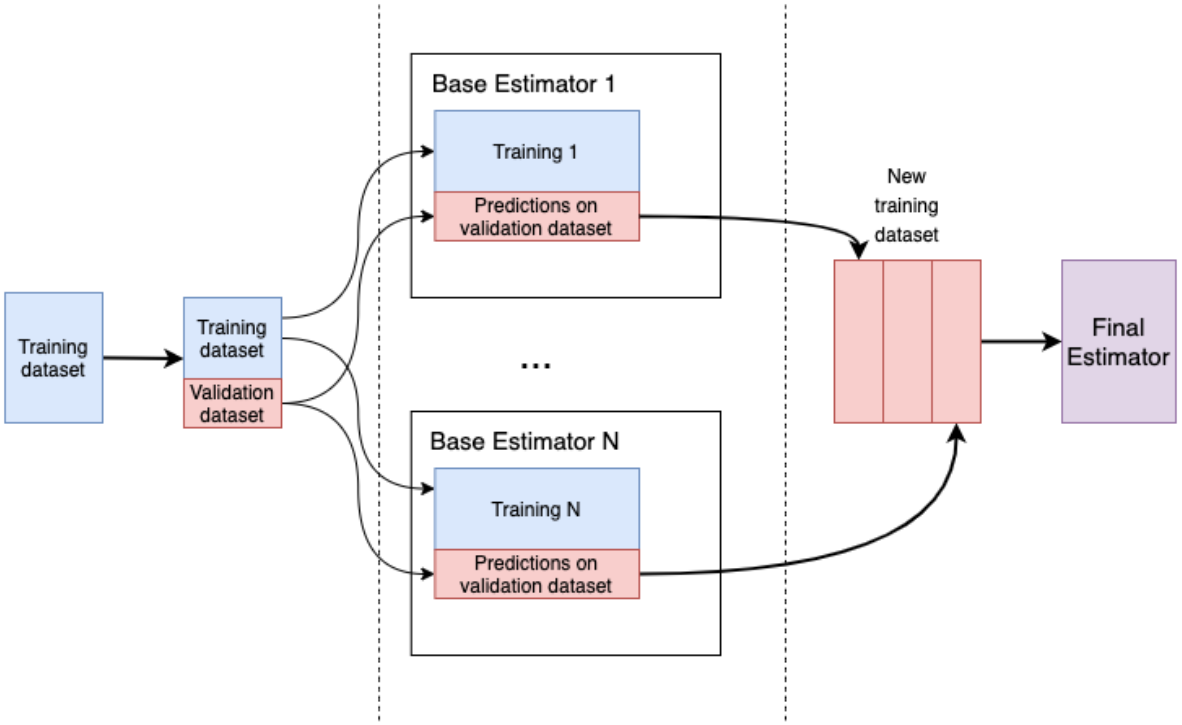


Figure 4.2 Flowchart of blending ensemble.

4.2.2.3 Voting

The principle of Voting is to combine different estimators and use a majority vote (hard voting) or the average predicted probabilities (soft voting) to generate the final prediction.

In the study, soft voting (SV) is adopted which makes the prediction based on the maximum value of the weighted average of predicted probabilities.

4.3 Probability Calibration

Probability calibration is employed when the machine learning model cannot yield good estimates of the class probabilities or even does not support the class probability prediction.

A calibration plot is an indicator of whether to calibrate the prediction result, which can be created by plotting the true frequency of the positive label against probabilities predicted by the classifier, and the diagonal line indicates perfectly calibrated. The close a model's curve is to the diagonal line, the better calibrated it is.

There are several ways that can be adopted to do the calibration (Niculescu-Mizil and Caruana, 2005):

- Signoid regressor based on Platt Calibration:

$$p(y = 1|f) = \frac{1}{1 + exp(Af+B)}$$

where y is the true label and f is the output of the uncalibrated classifier. A and B are parameters to be determined when fitting the regressor via maximum likelihood.

- Isotonic Regression has fit the isotonic (monotonically increasing) function $\widehat{m}$ such that

$$\widehat{m} = argmin_z \Sigma(y - z(f))^2$$

where y is the true label and f is the output of the uncalibrated classifier. z is the calibrated classifier.

5. Methodology

Figure 5.1 illustrates the study flow in this thesis.

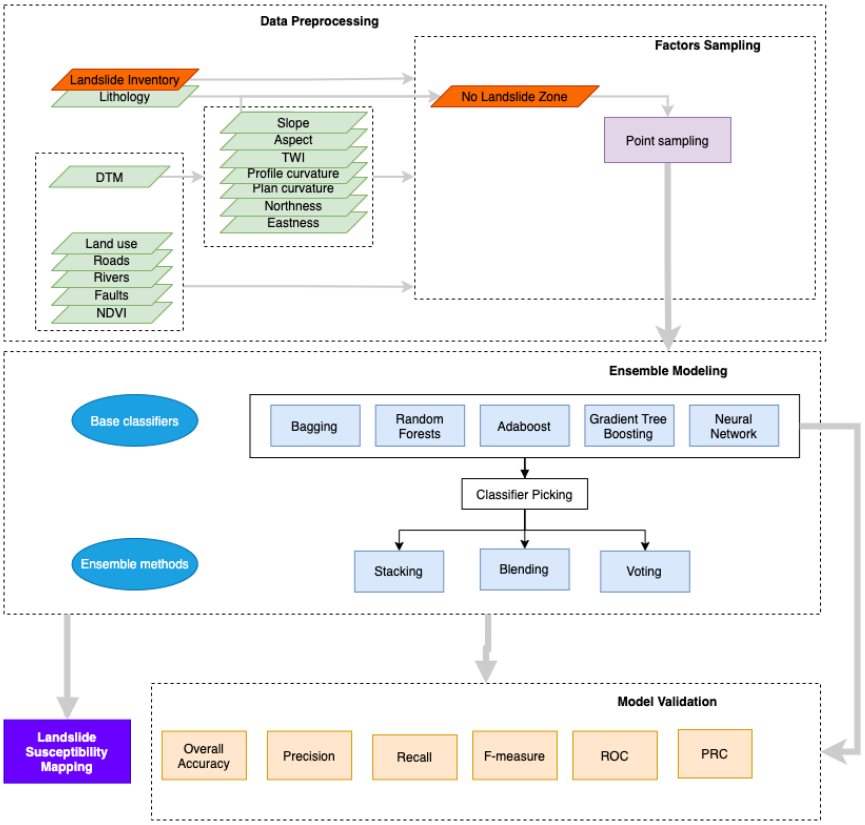


Figure 5.1: Study flow chart.

5.1 Data preprocessing

Creating and preparing the dataset for models is a fundamental step in machine learning analysis. For Landslide Susceptibility analysis, it is crucial to select a suitable set of environmental variables that are considered to be related to mass movement, as well as a sufficiently large list of landslide phenomena inventory.

5.1.1 Factors selection and preparation

In this study, 11 factors were selected according to the previous studies in the similar area (Yordanov and Brovelli, 2021; Amici, 2021):

- Elevation (Raster Data)
- Eastness (Raster Data)
- Northness (Raster Data)
- Plan curvature (Raster Data)
- Profile curvature (Raster Data)
- Topographic Wetness Index (Raster Data)

- NDVI (Raster Data)
- Precipitation (Raster Data)
- Distance from roads (Vector Data)
- Distance from rivers (Vector Data)
- Distance from faults (Vector Data)
- Land use (Vector Data)

Slope and lithology layers were excluded from the factors considered because they were exploited for the definition of No Landslide zone (NoLS zone). Including them produces maps with a high prevalence of these factors relative to others (Amici, 2021).

The precipitation information was included as additional variable only to as a last iteration to the best performing model for Lombardy region. Two-fold modelling was included for the preciptation - on one hand, was included the average interpolated hourly precipitation for the year 2020 and on the other, was added only the 90th percentile which is represeting exceptional intensive events for the same year.

To perform sampling, all of these layers must be in raster format with the same extent, pixel size, and CRS.

Therefore, vector data should be converted to raster data and used as discrete factors. They are

- **Distance from roads**: first buffer the road element layer at 50m, 100m, 250m, 500m, remaining zone with a distance >500m. Then convert the buffered layer to a raster using QGIS.
- **Distance from rivers**: similar to the “distance from roads” factor, the process was applied to the river network layer.
- **Distance from faults**: similar to the “distance from roads” factor, the process was applied to the fault lines layer.
- **Land use**: rasterize the land use layer directly using the Level 2 classification.

It should be noted that in this study, both discrete and continuous variables were employed. Table 5.1 describes the categorization of discrete factors. Both Upper Valtellina and Val Tartano have fewer land-use classes than Val Chiavenna and Lombardy, where “permanent crops” is absent in Upper Valtellina while “non-agricultural green”, “arable land”, “permanent crops”, and “inland wetlands” are missing in Val Tartano.

Factor	Number of classes	Classification parameter
Distance from roads	5	Buffered distance
Distance from rivers	5	Buffered distance
Distance from faults	5	Buffered distance
Land use	11 - Upper Valtellina 8 - Val Tartano	Level 2 classification of land use

	12 - Val Chiavenna 12 - Lombardy	

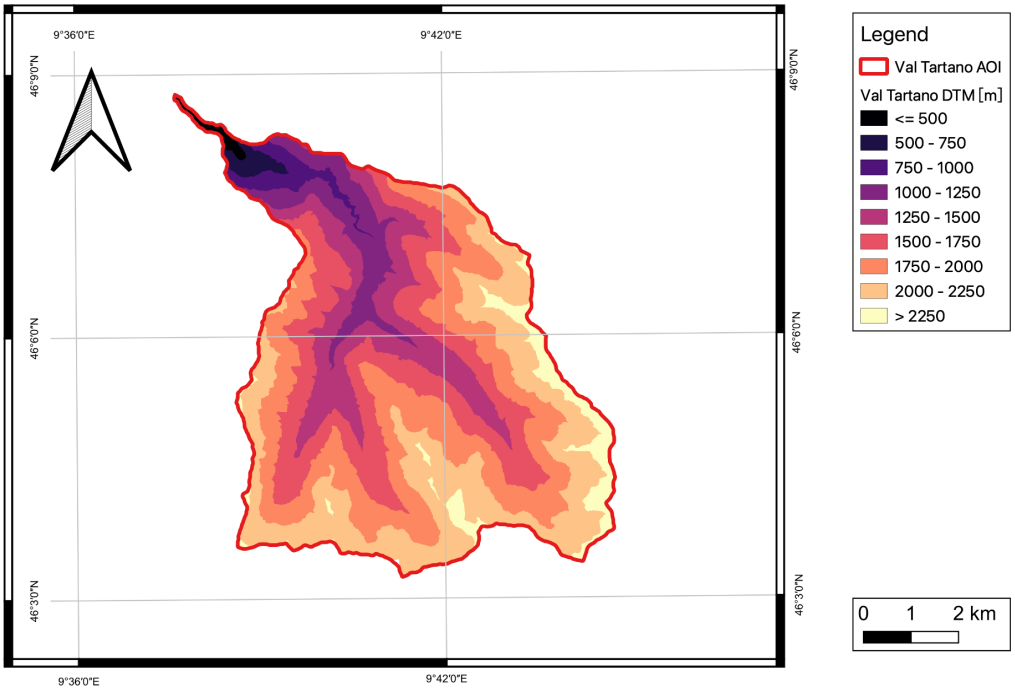
Table 5.1: Categorical environmental variables.

Selecting a suitable mapping unit is a crucial prerequisite for LSMs (Guzzetti et al., 1999; Reichenbach et al., 2018). In this study, the pixel size, equivalent to the DTM resolution (5m × 5m), was employed as the mapping unit since most terrain variables are calculated from DTM, and this scale is sufficient for local factors such as road and river networks. Additionally, this scale is appropriate for polygonal landslides and approximated linear landslides, taking into account the scale of the landslide inventory (1:10,000) (Yordanov and Brovelli, 2021).

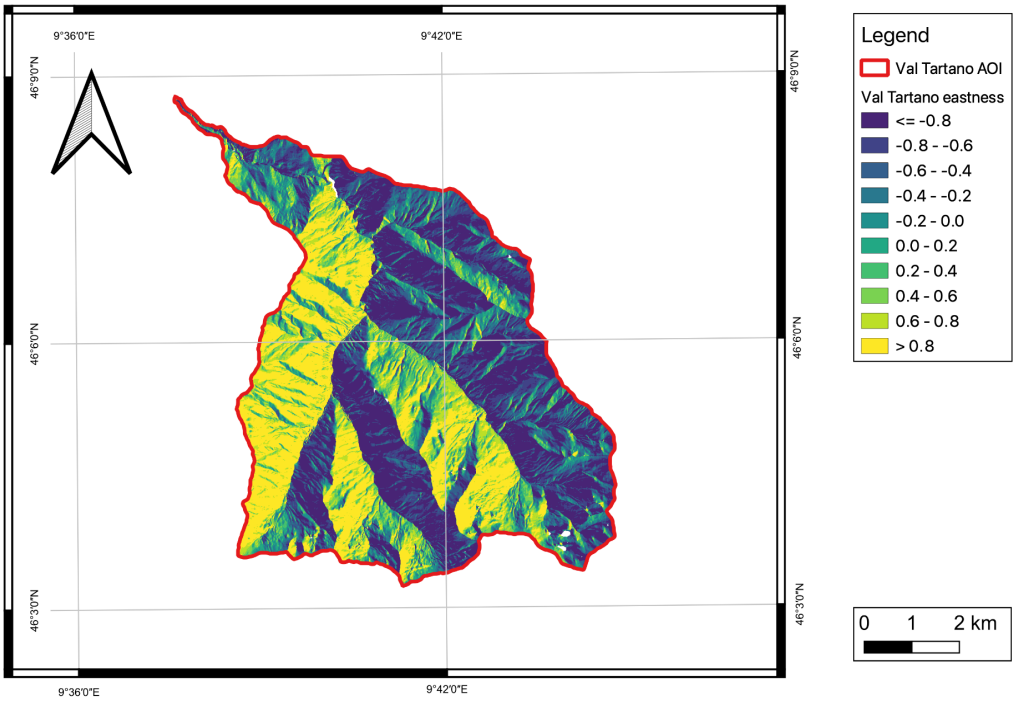
The final preprocessed factors which were used later on in the analysis are shown below.

5.1.1.1 Val Tartano

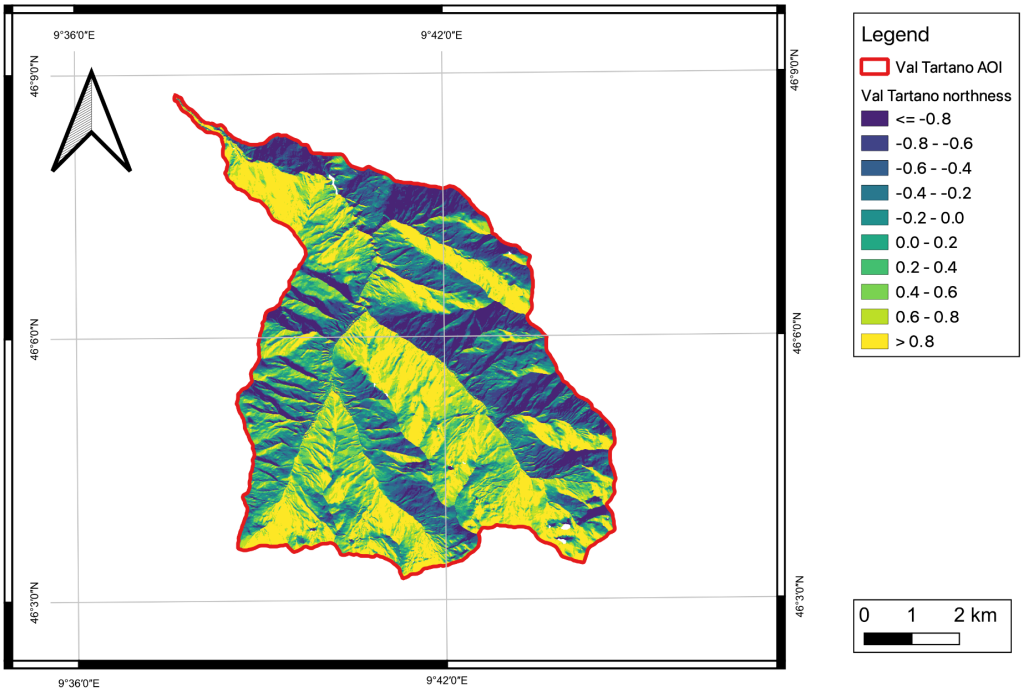
Figure 5.2 describes 11 factors in Val Tartano.



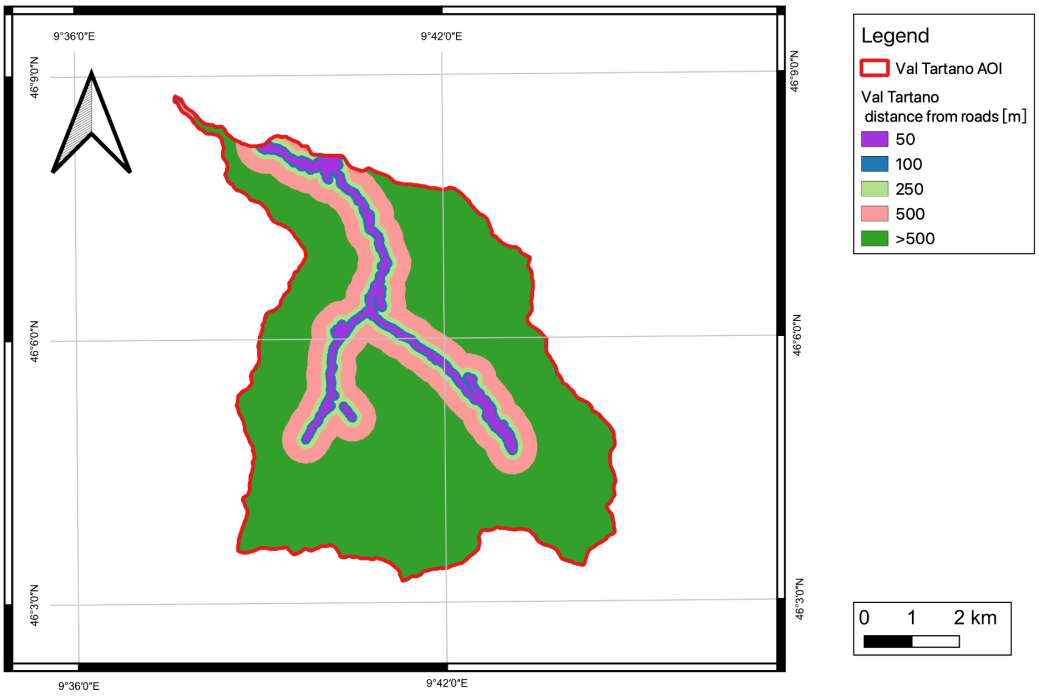
(a) Val Tartano DTM



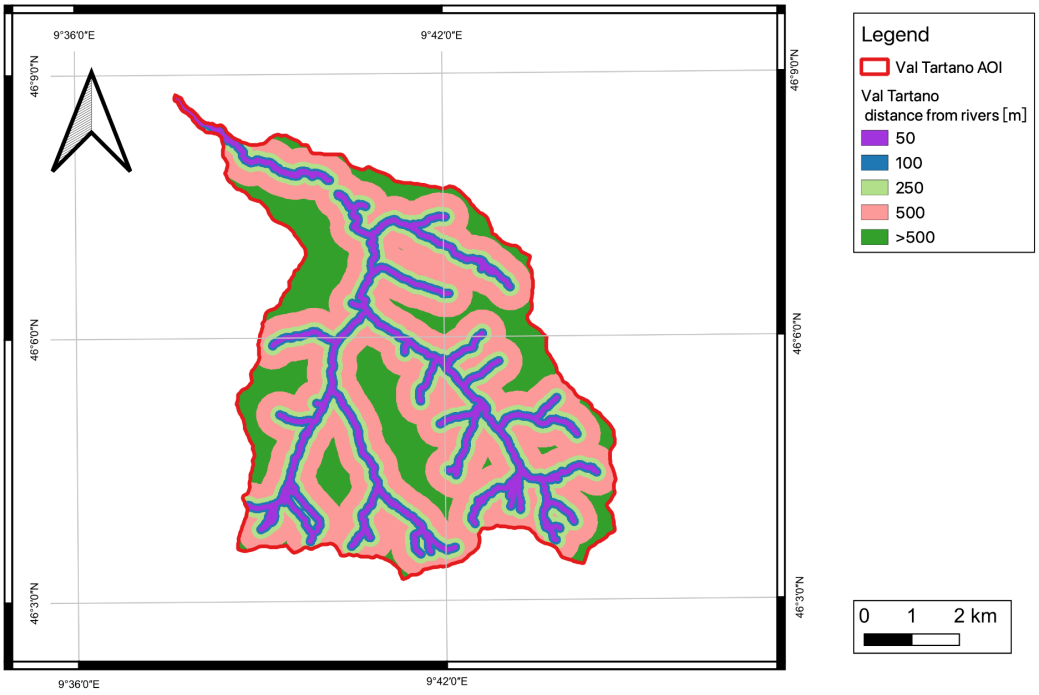
(b) Val Tartano Eastness



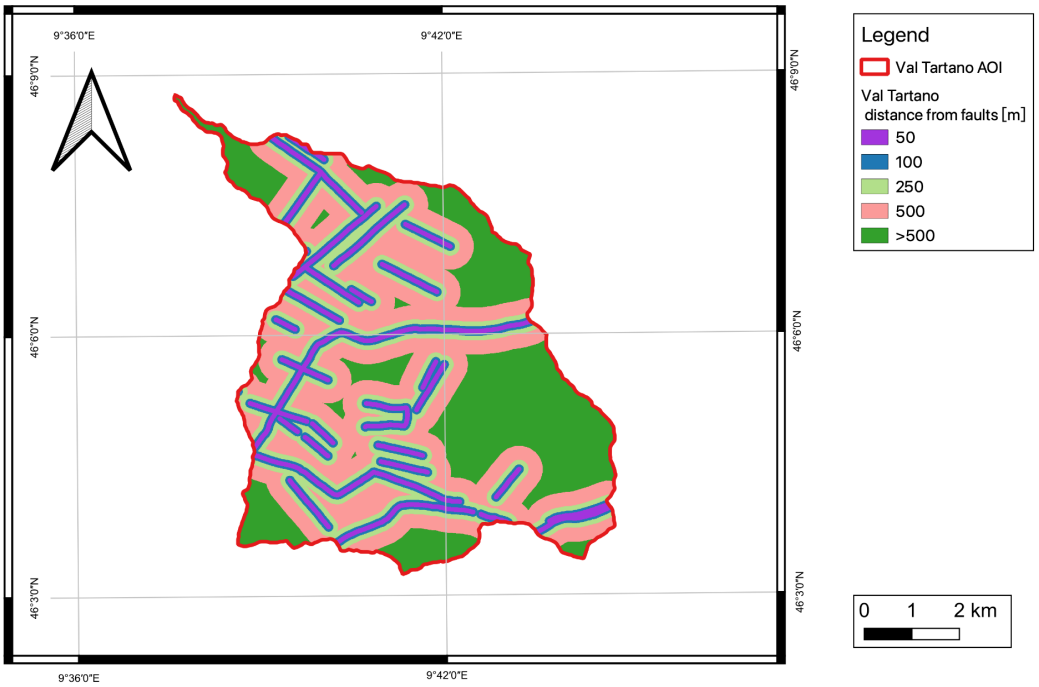
(c) Val Tartano Northness



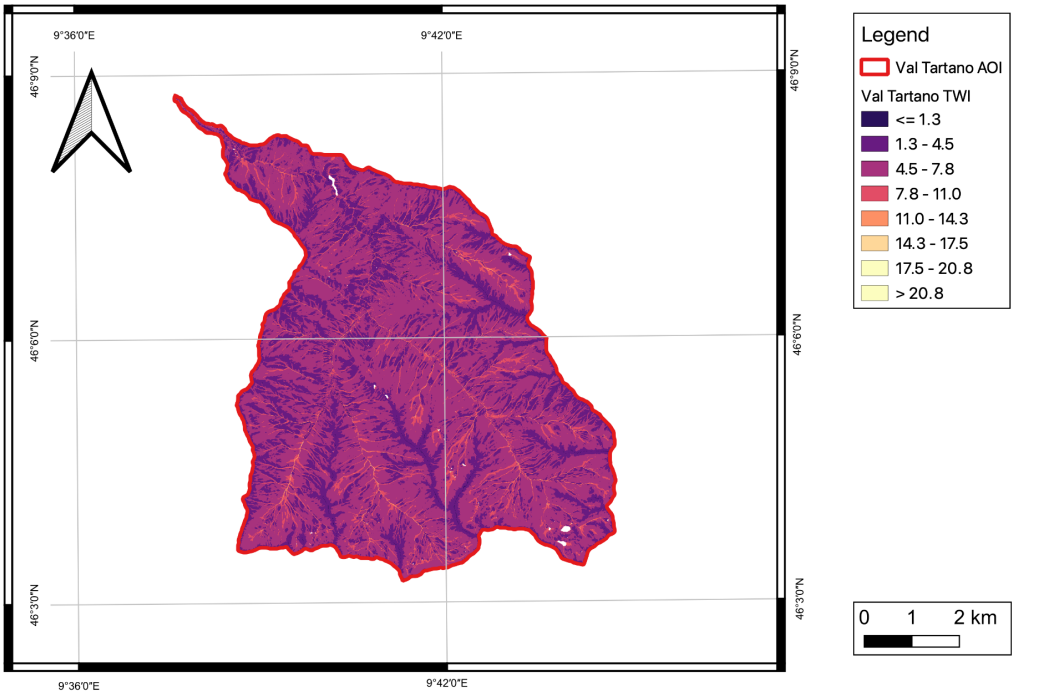
(d) Val Tartano Distance from roads



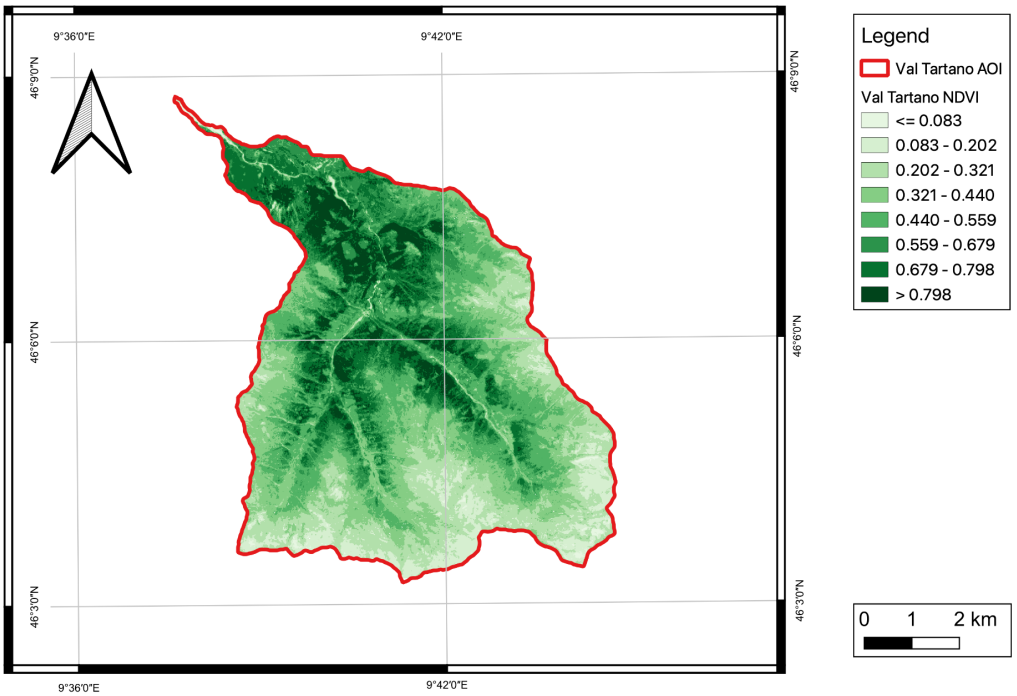
(e) Val Tartano Distance from rivers



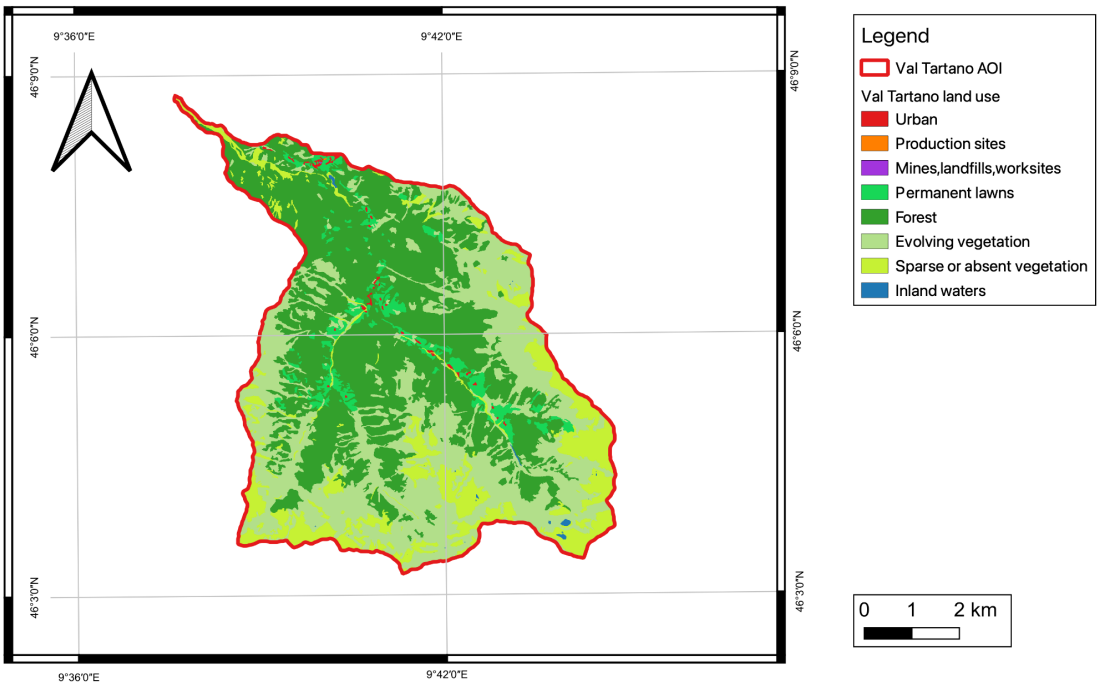
(f) Val Tartano Distance from faults



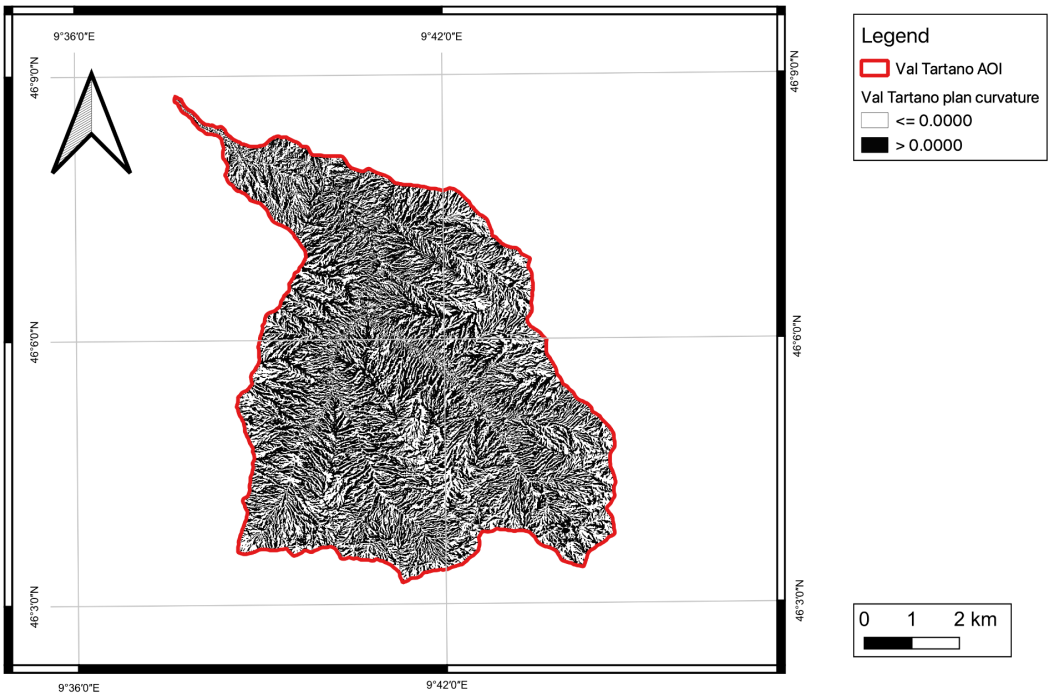
(g) Val Tartano Topographic Wetness Index



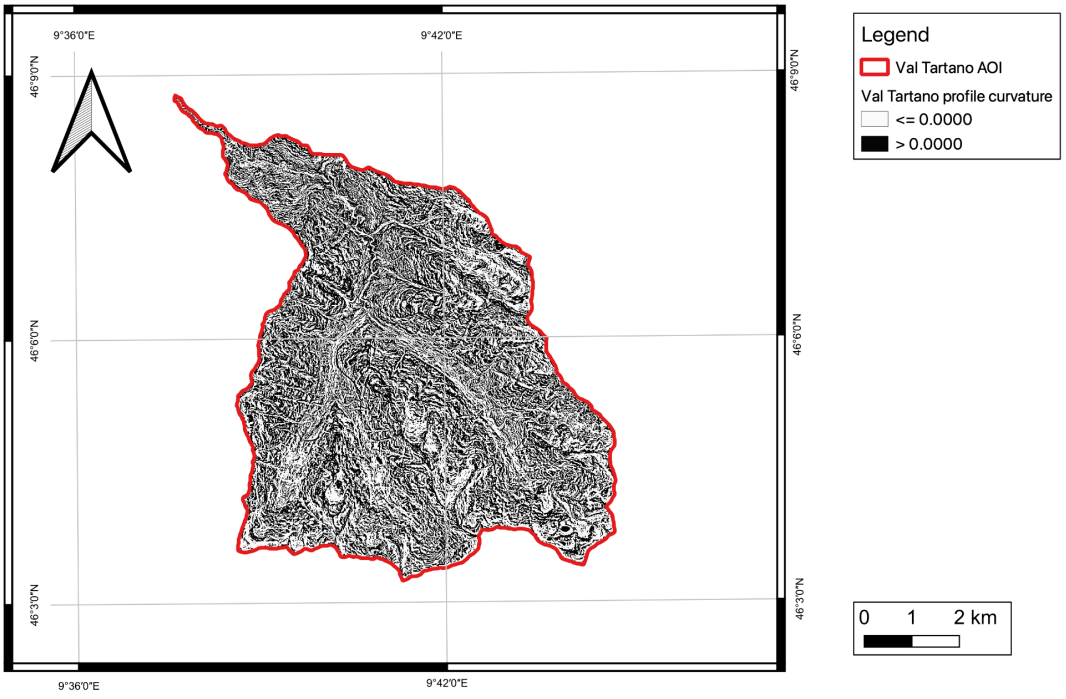
(h) Val Tartano NDVI



(i) Val Tartano Land use



(j) Val Tartano Plan curvature

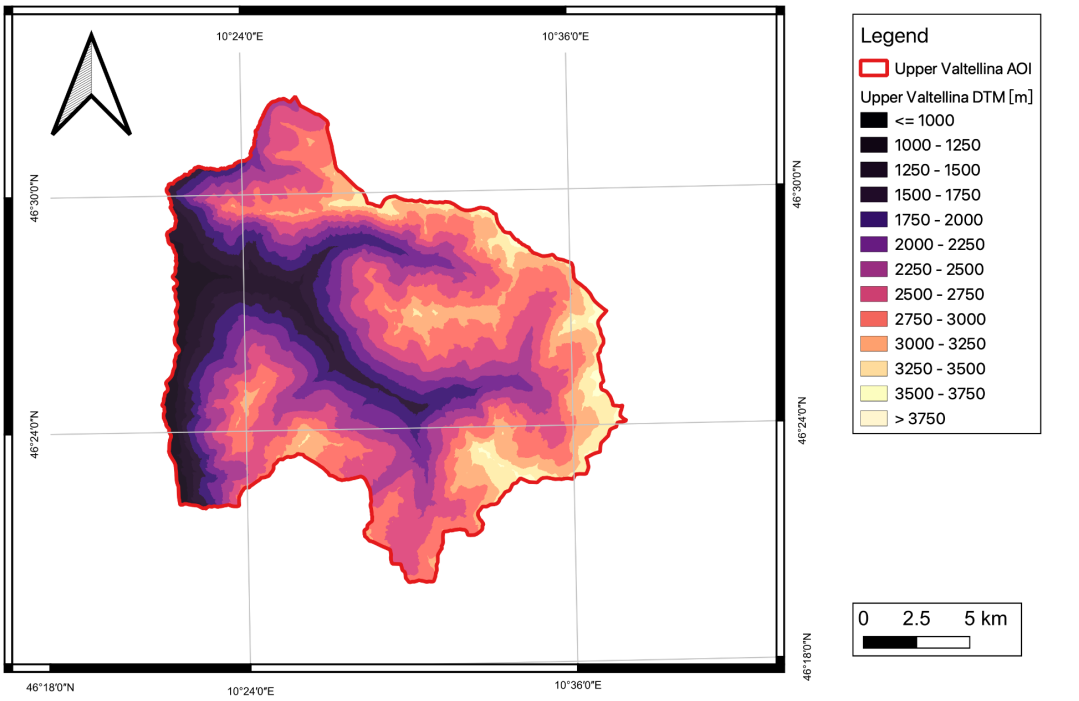


(k) Val Tartano Profile curvature

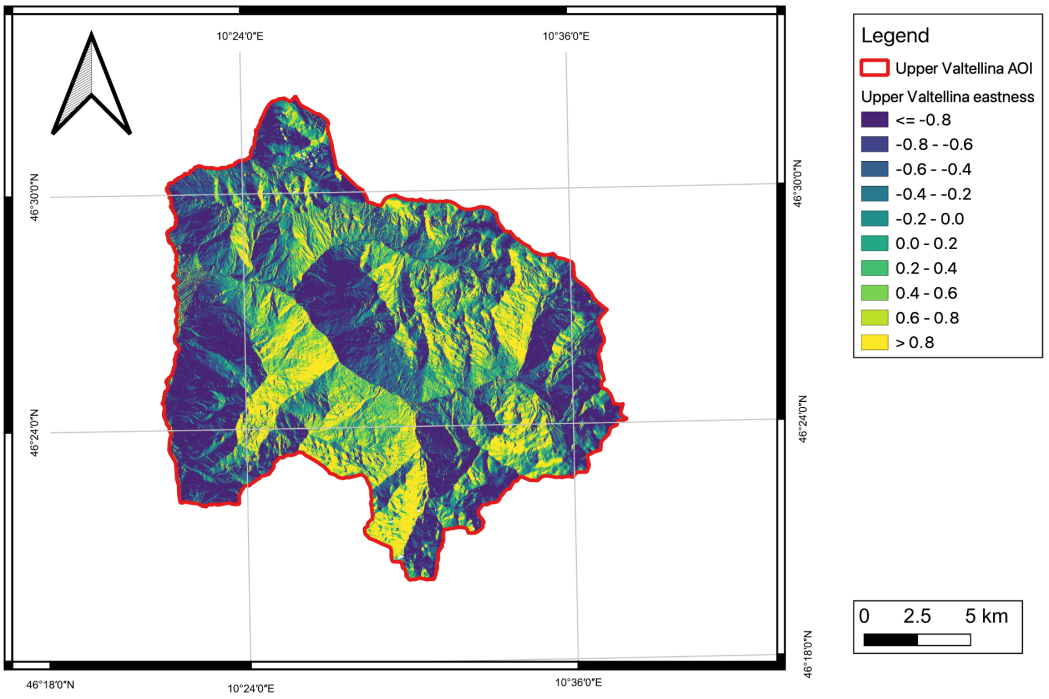
Figure 5.2: Val Tartano terrain variables

5.1.1.2 Upper Valtellina

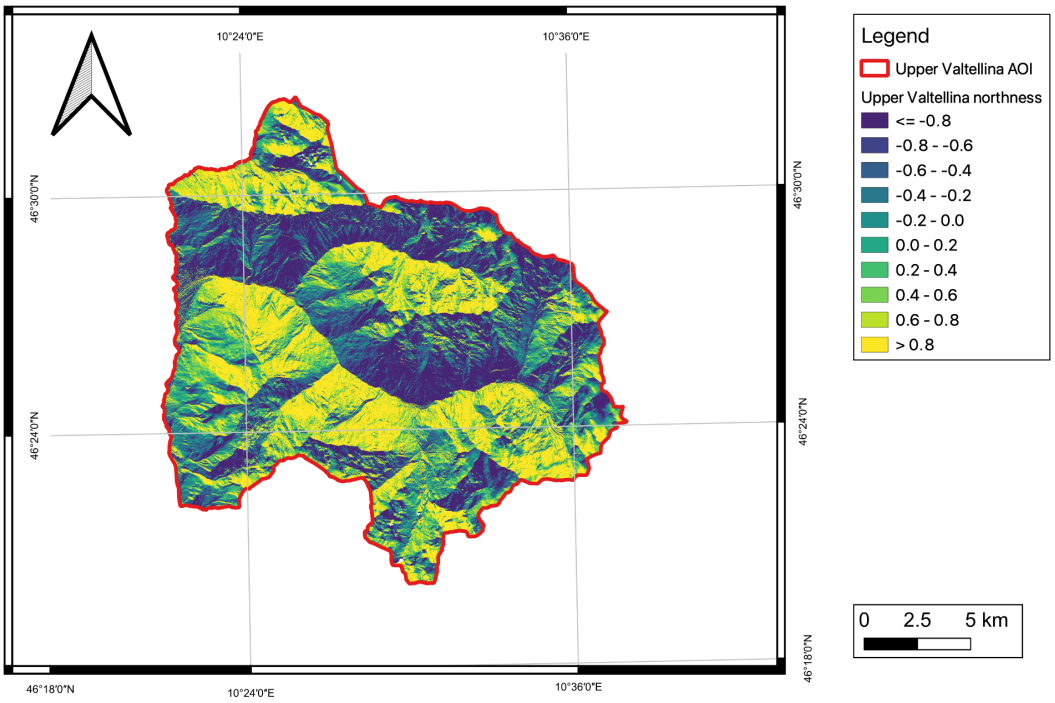
Figure 5.3 illustrates 11 factors in Upper Valtellina.



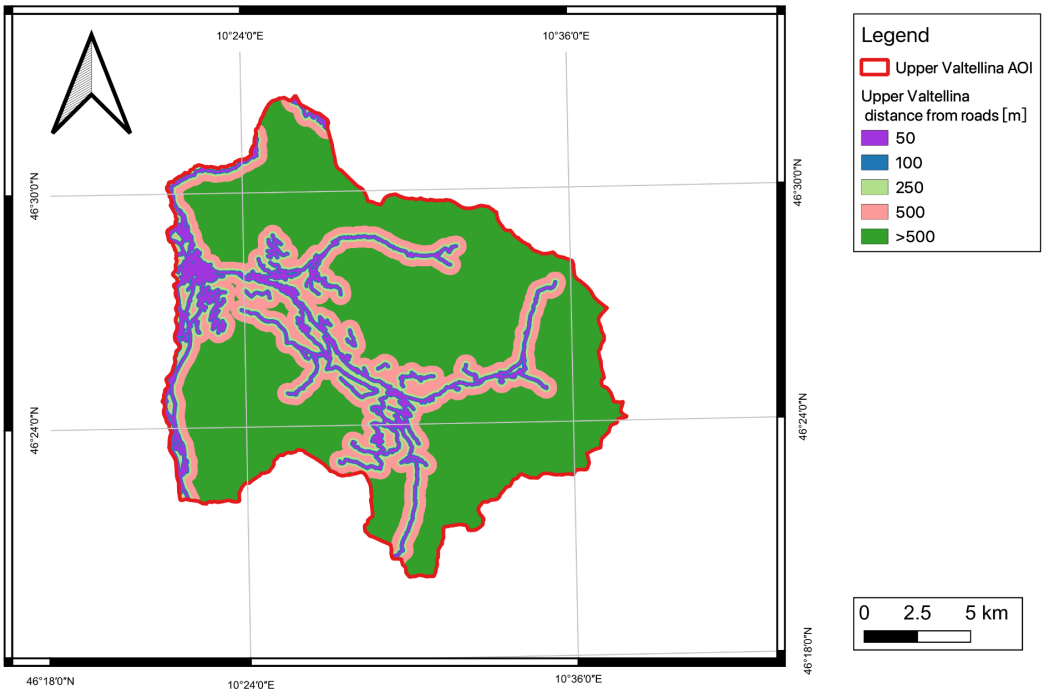
(a) Upper Valtellina DTM



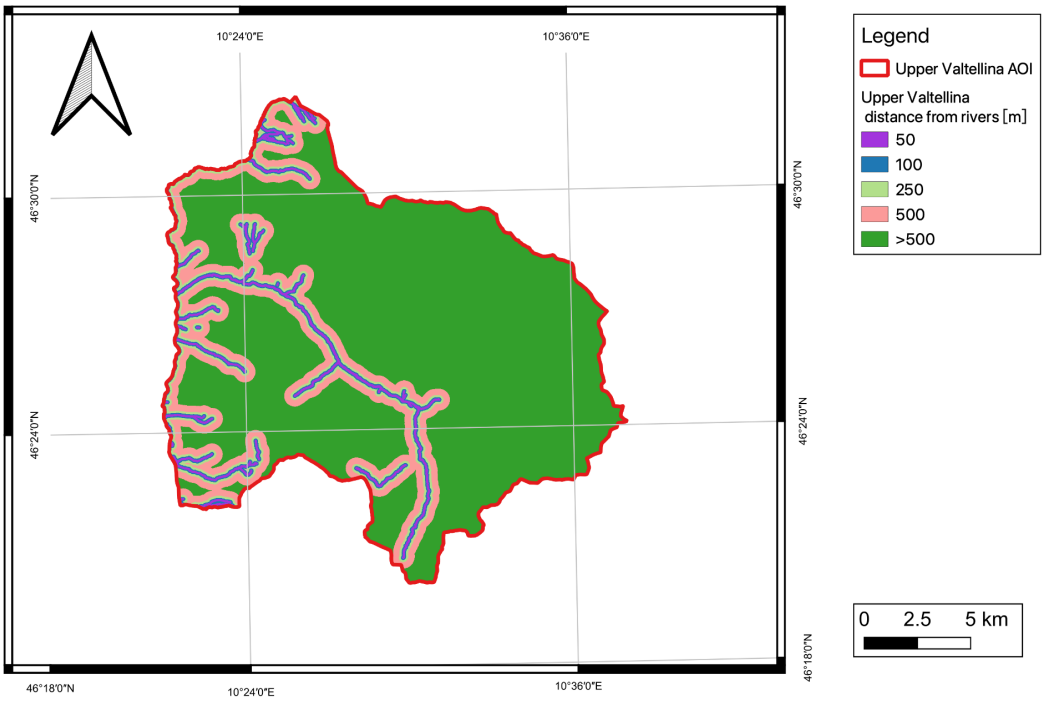
(b) Upper Valtellina Eastness



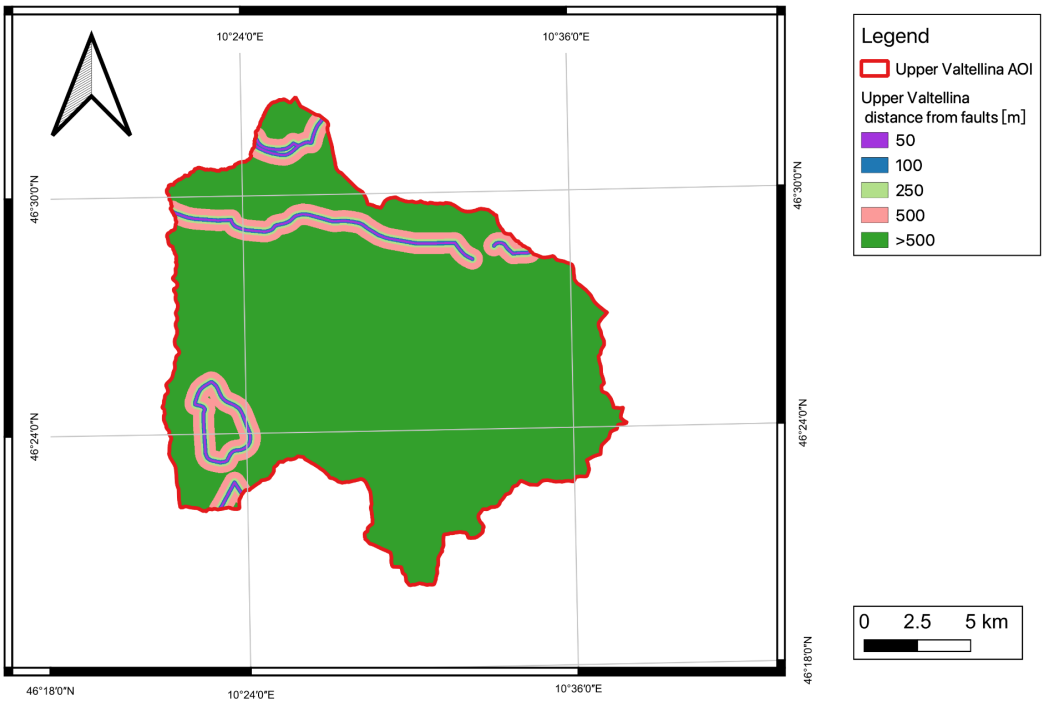
(c) Upper Valtellina Northness



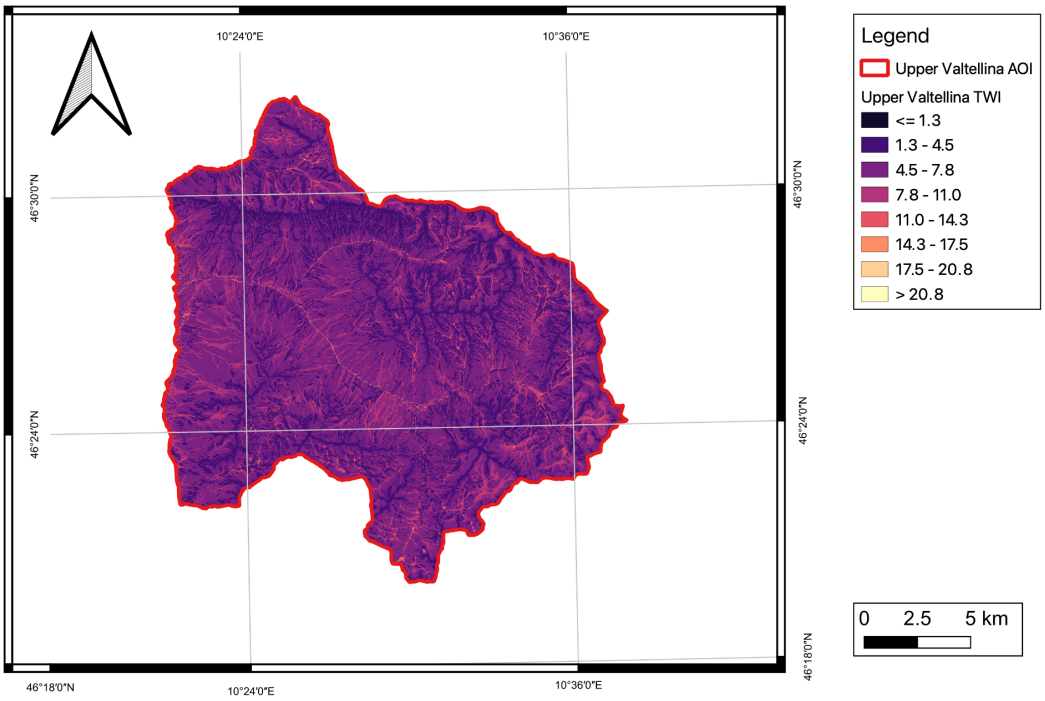
(d) Upper Valtellina Distance from roads



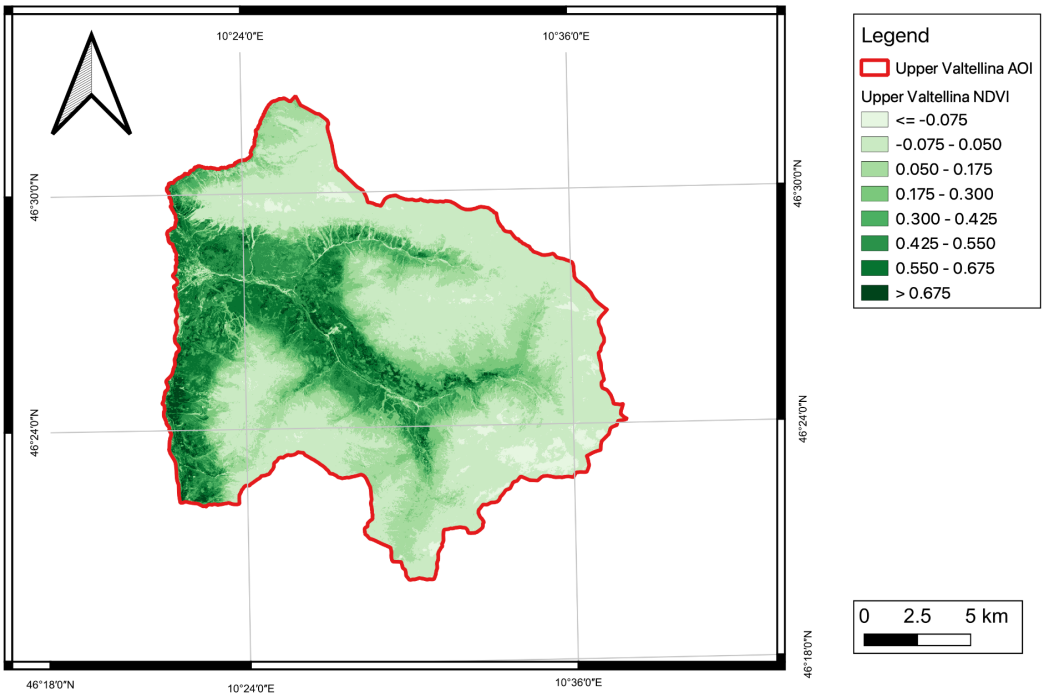
(e) Upper Valtellina Distance from rivers



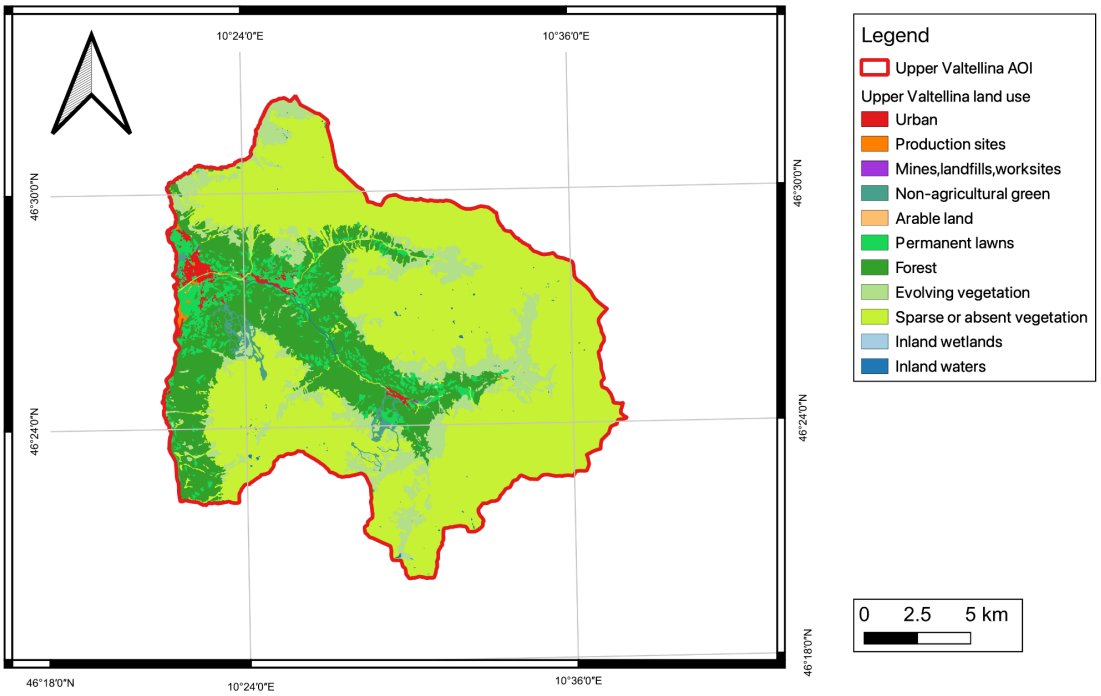
(f) Upper Valtellina Distance from faults



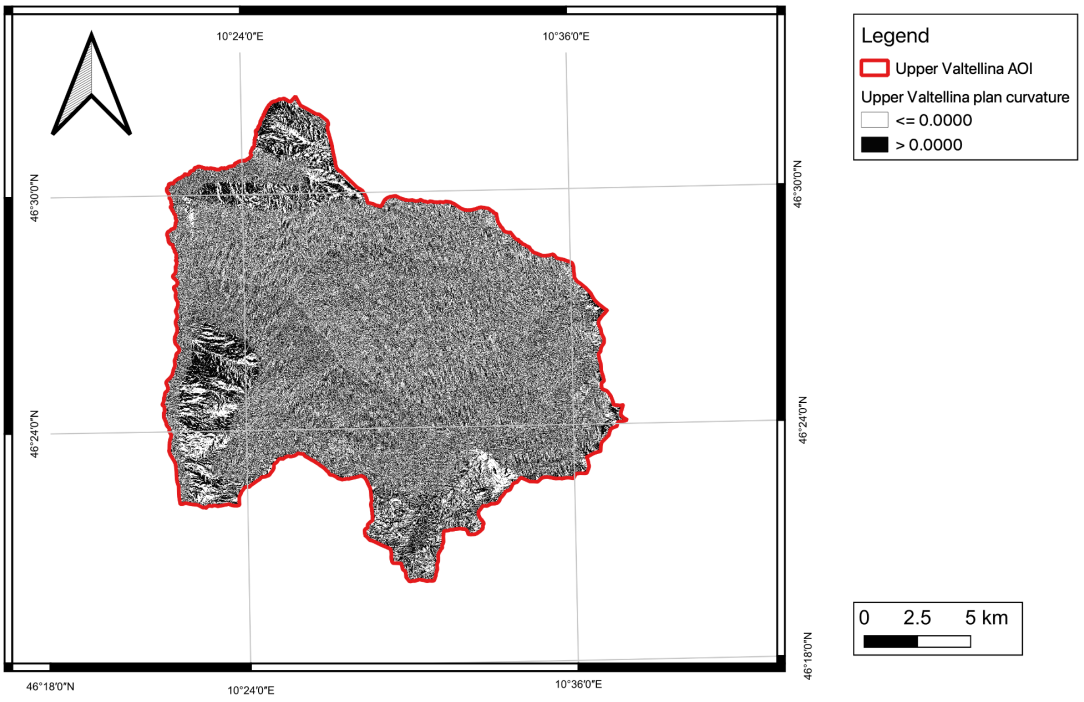
(g) Upper Valtellina Topographic Wetness Index



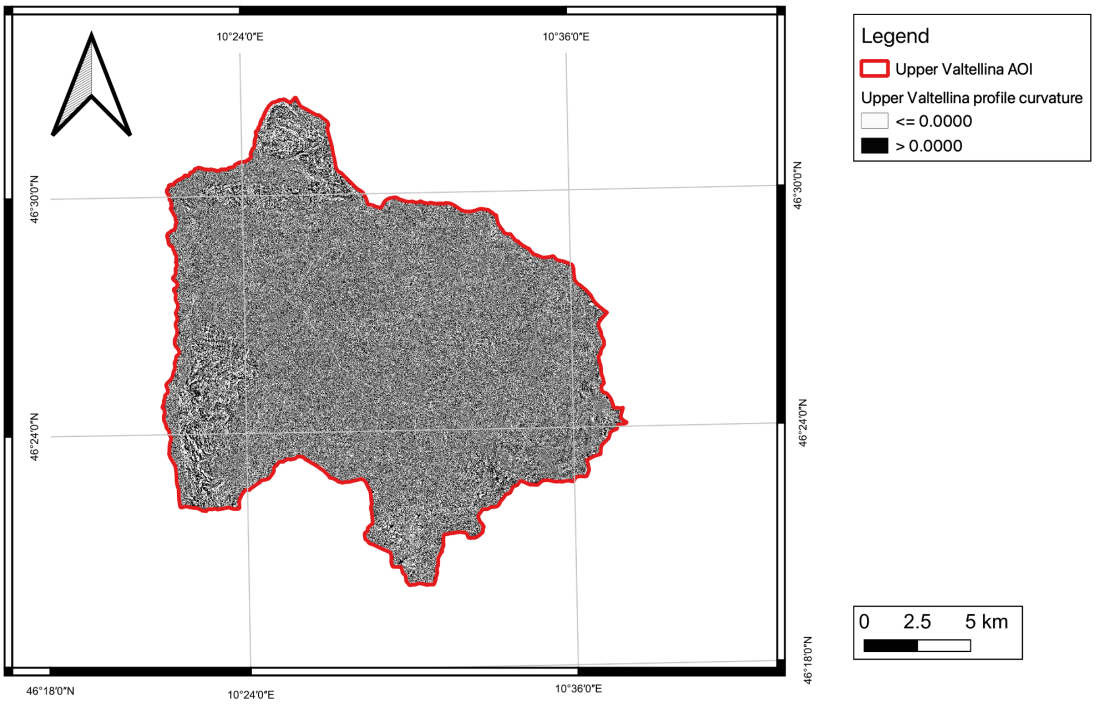
(h) Upper Valtellina NDVI



(i) Upper Valtellina Land use



(j) Upper Valtellina Plan curvature

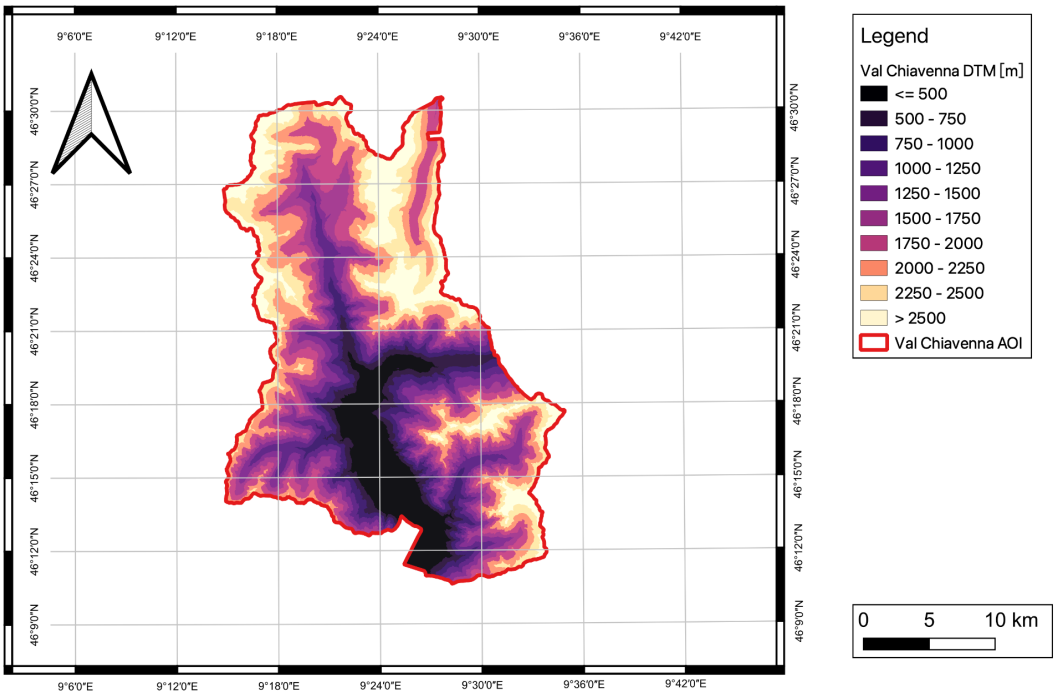


(k) Upper Valtellina Profile curvature

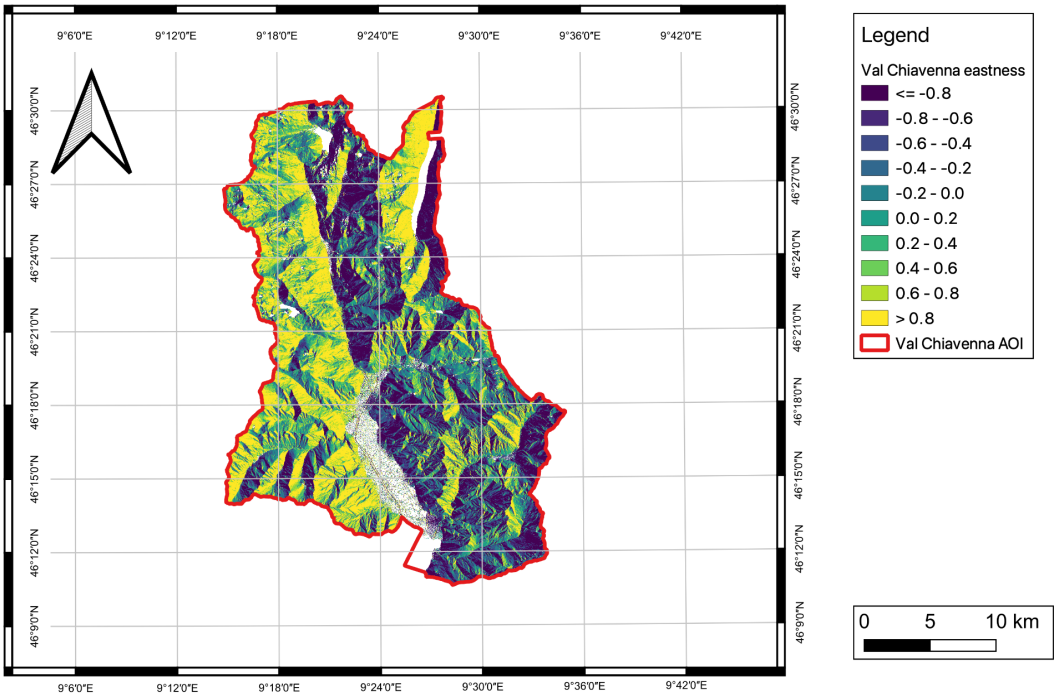
Figure 5.3: Upper Valtellina terrain variables

5.1.1.3 Val Chiavenna

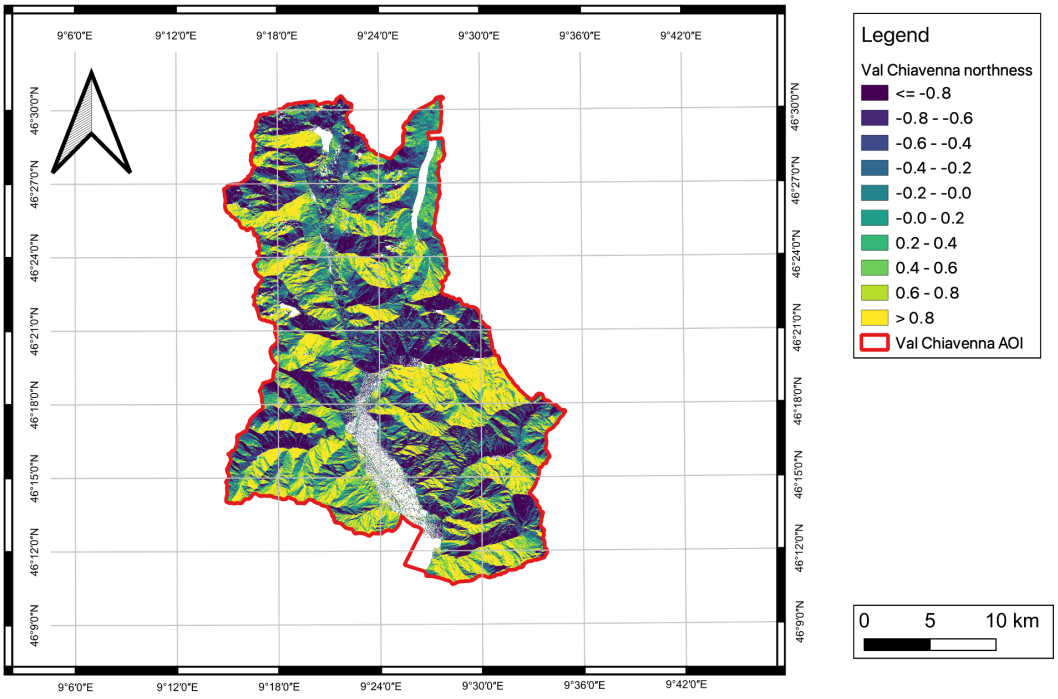
Figure 5.4 shows 11 factors in Val Chiavenna.



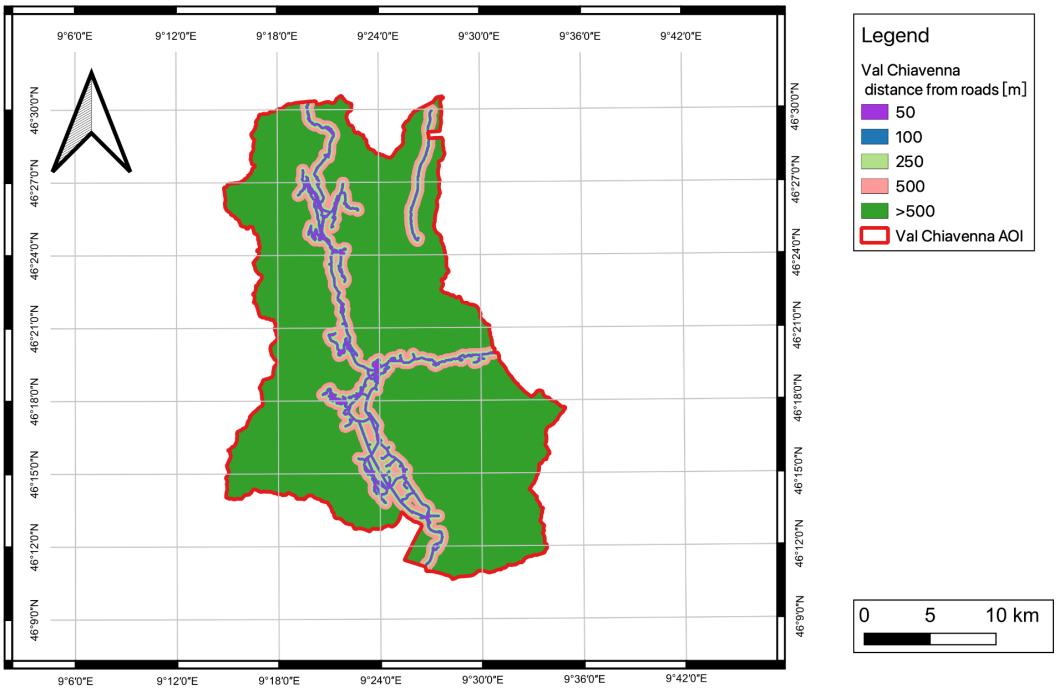
(a) Val Chiavenna DTM



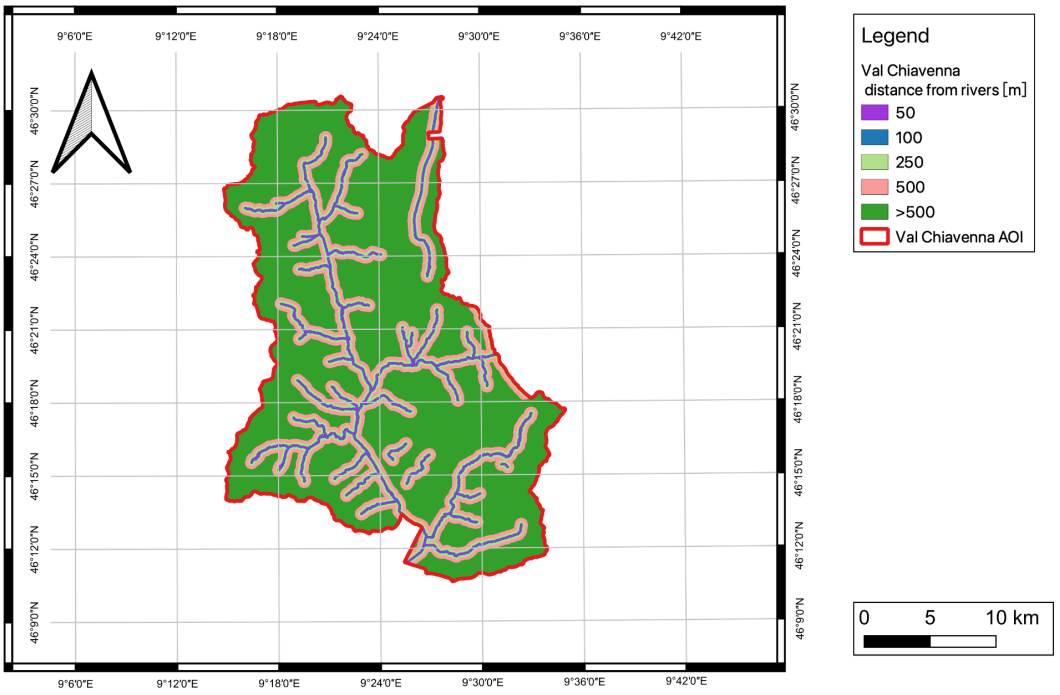
(b) Val Chiavenna Eastness



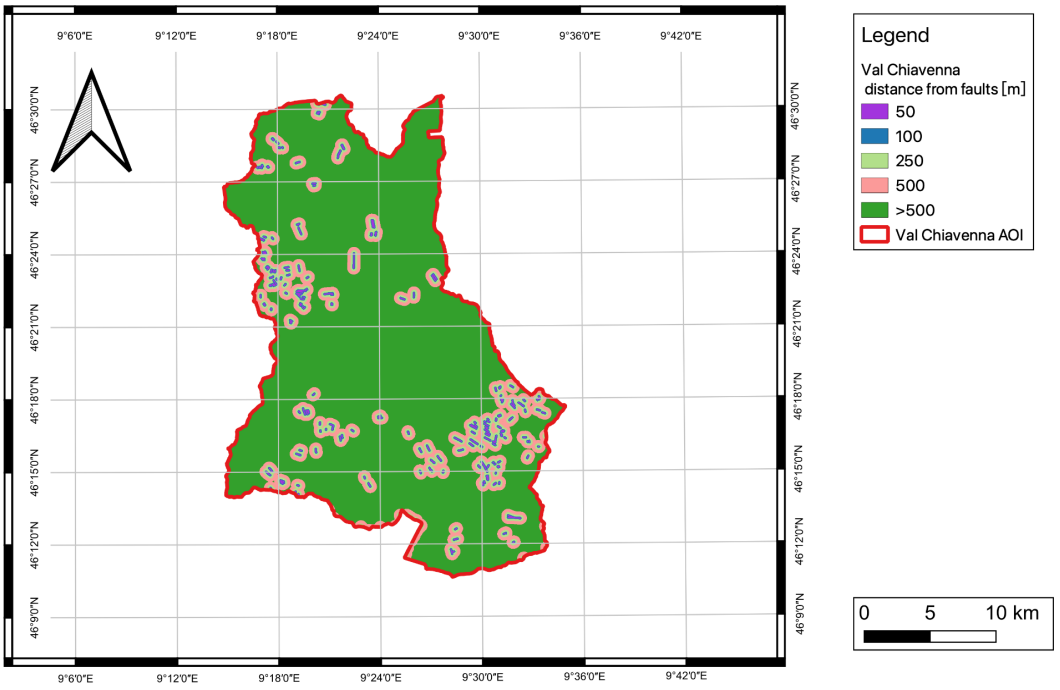
(c) Val Chiavenna Northness



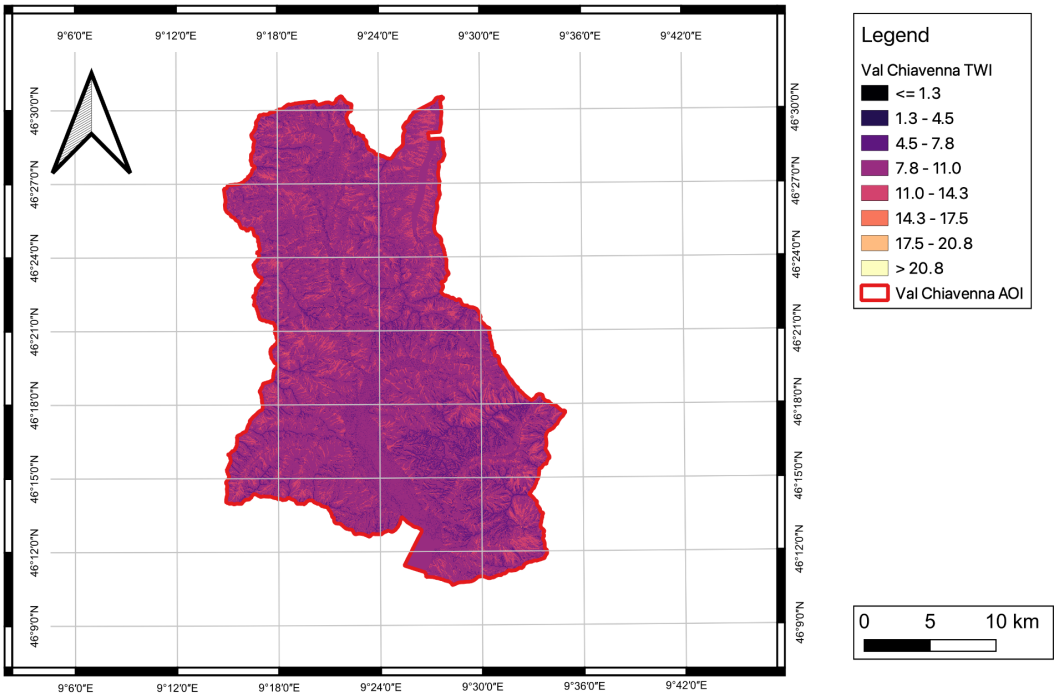
(d) Val Chiavenna Distance from roads



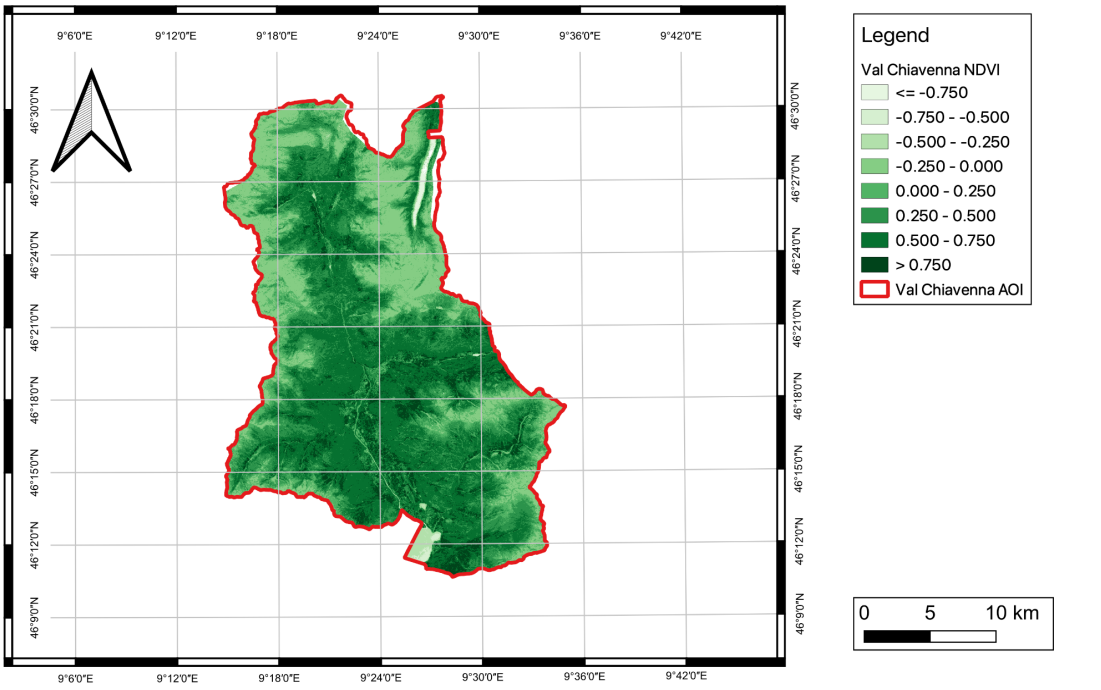
(e) Val Chiavenna Distance from rivers



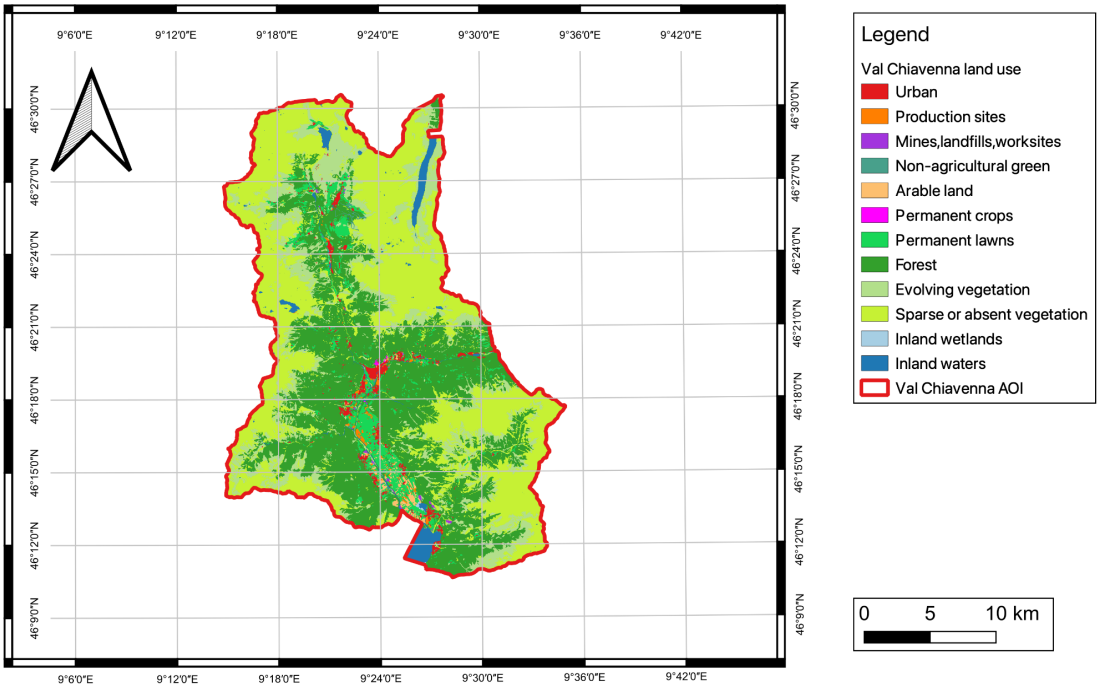
(f) Val Chiavenna Distance from faults



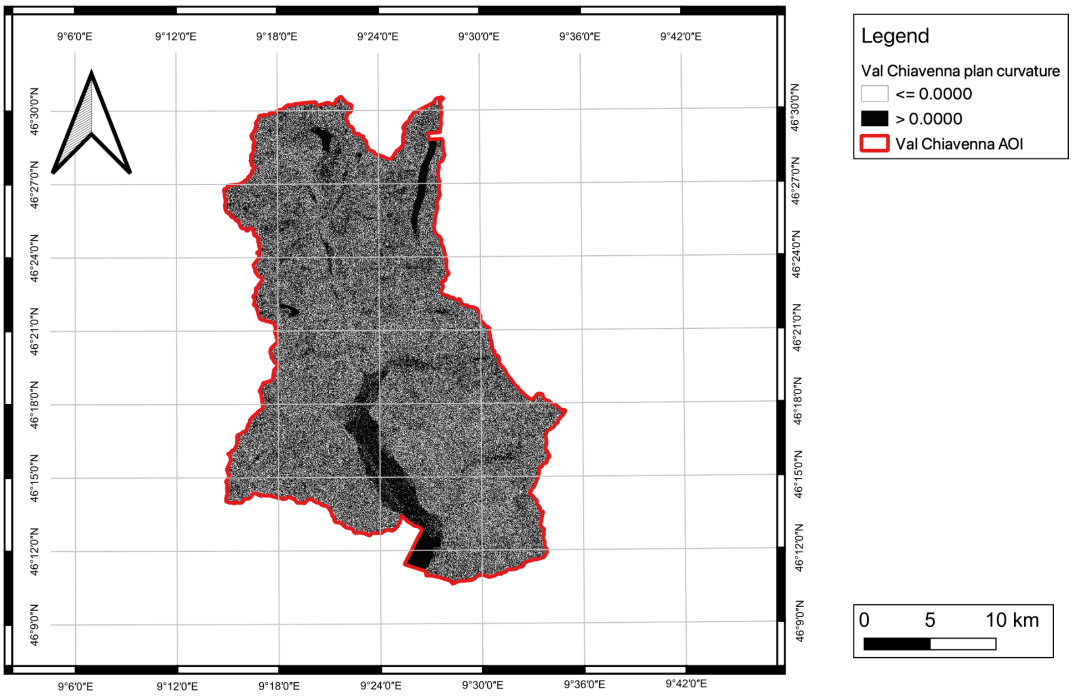
(g) Val Chiavenna Topographic Wetness Index



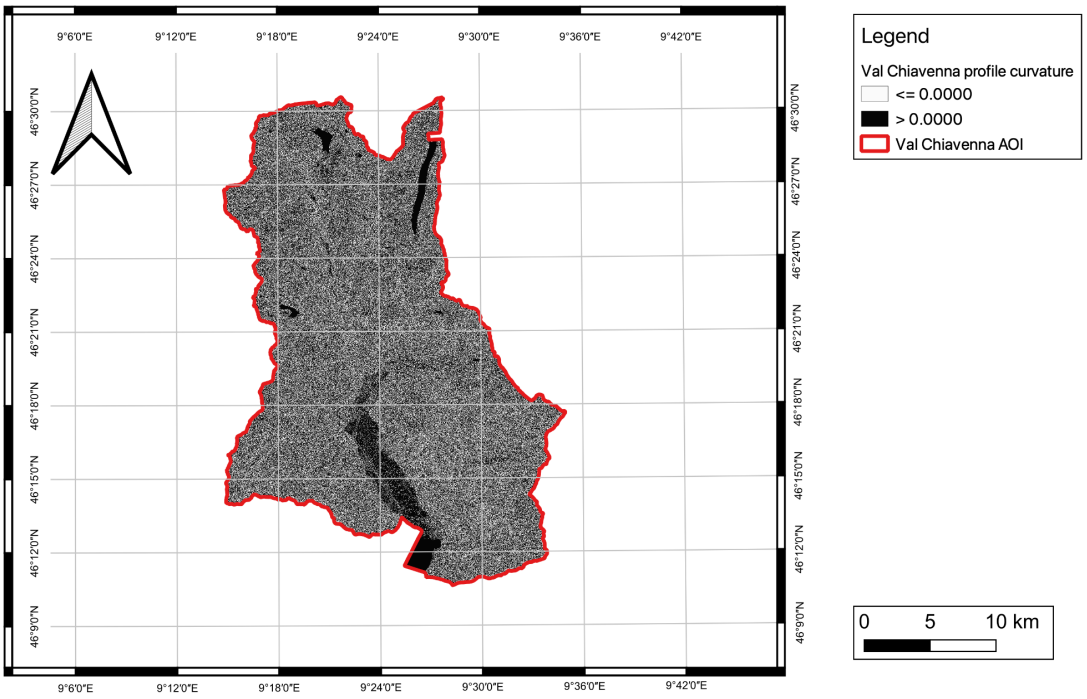
(h) Val Chiavenna NDVI



(i) Val Chiavenna Land use



(j) Val Chiavenna Plan curvature

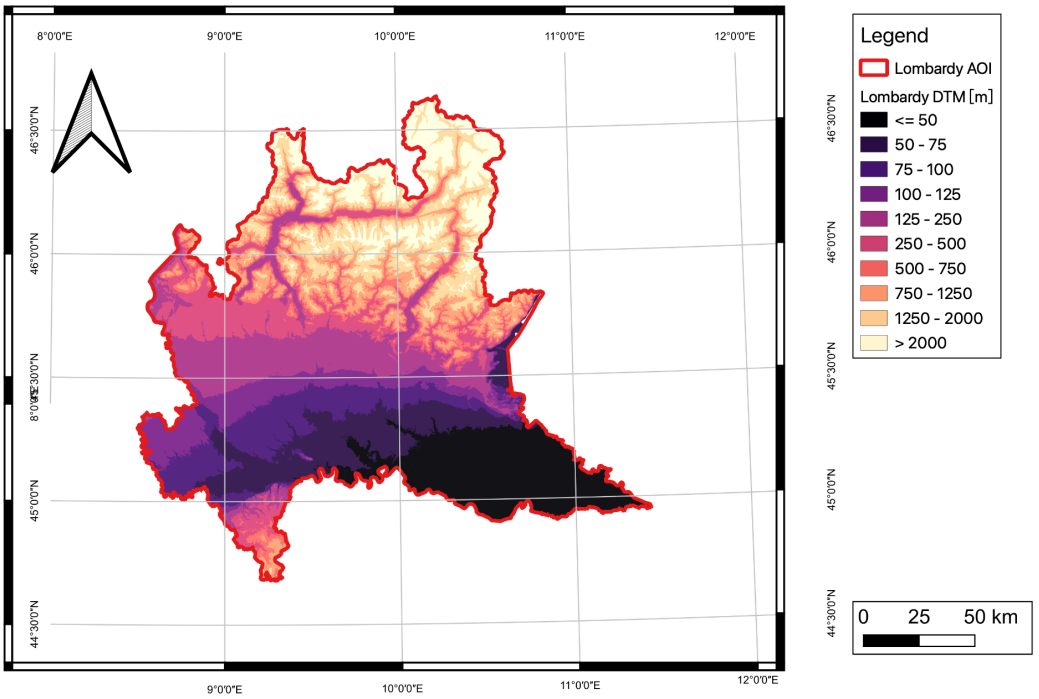


(k) Val Chiavenna Profile curvature

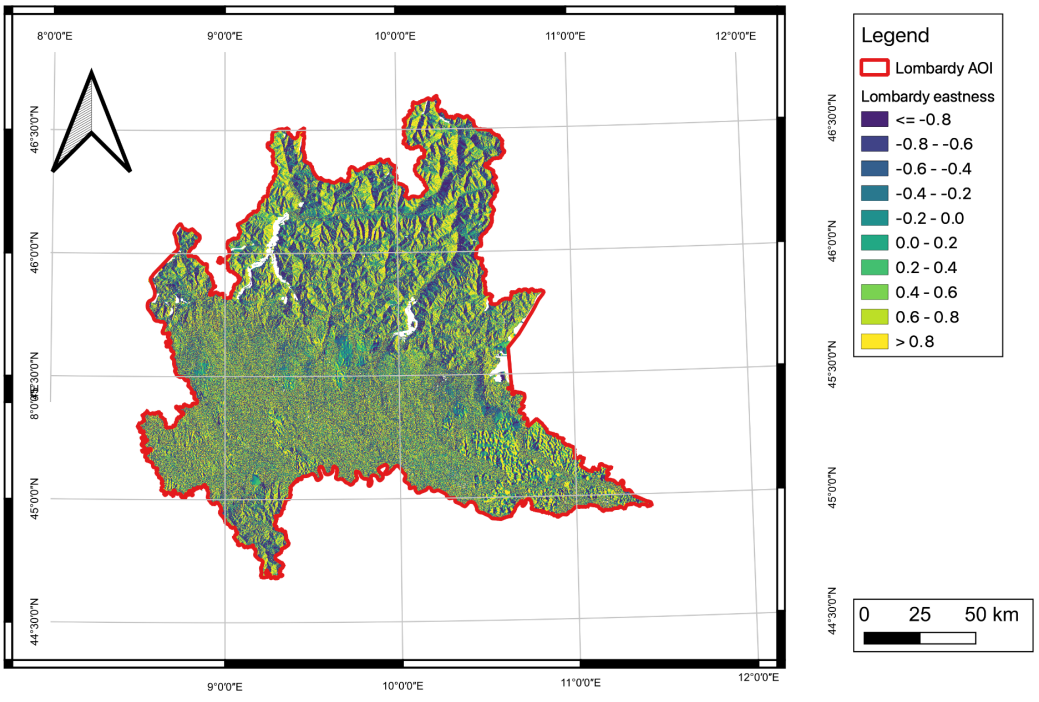
Figure 5.4: Val Chiavenna terrain variables

5.1.1.4 Lombardy

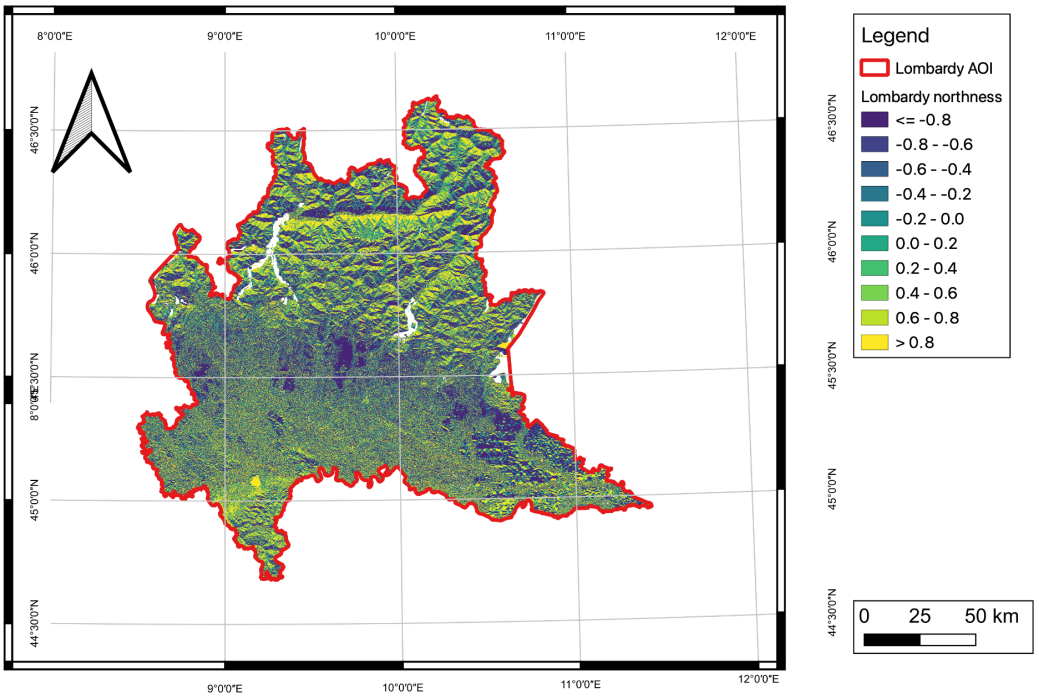
Figure 5.5 displays 11 factors in Lombardy.



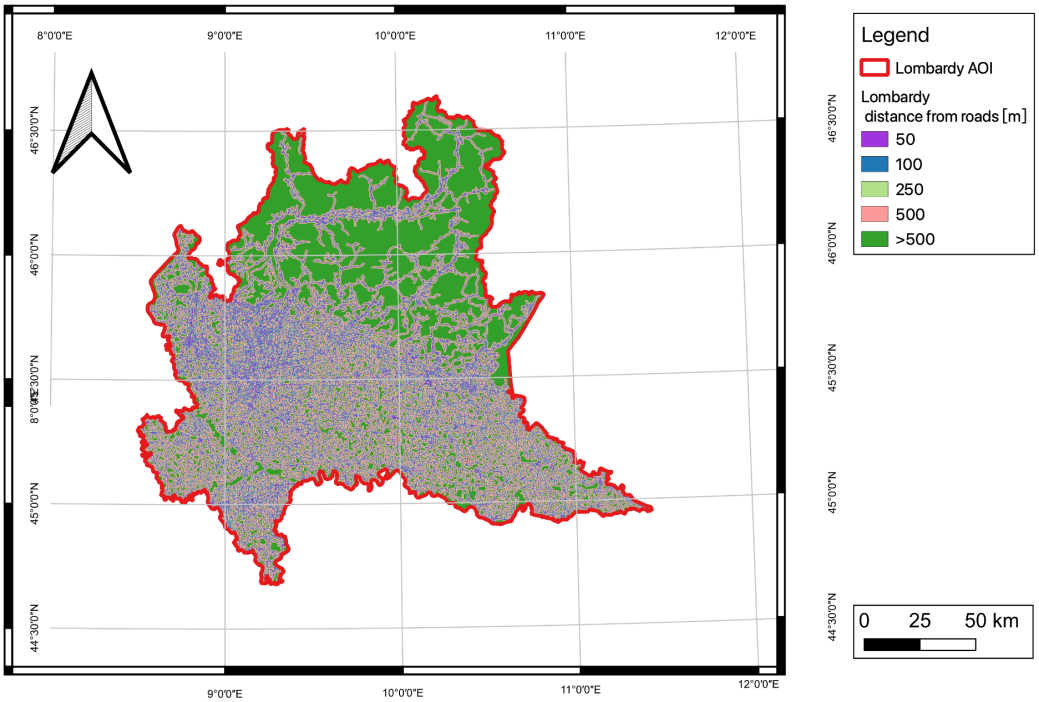
(a) Lombardy DTM



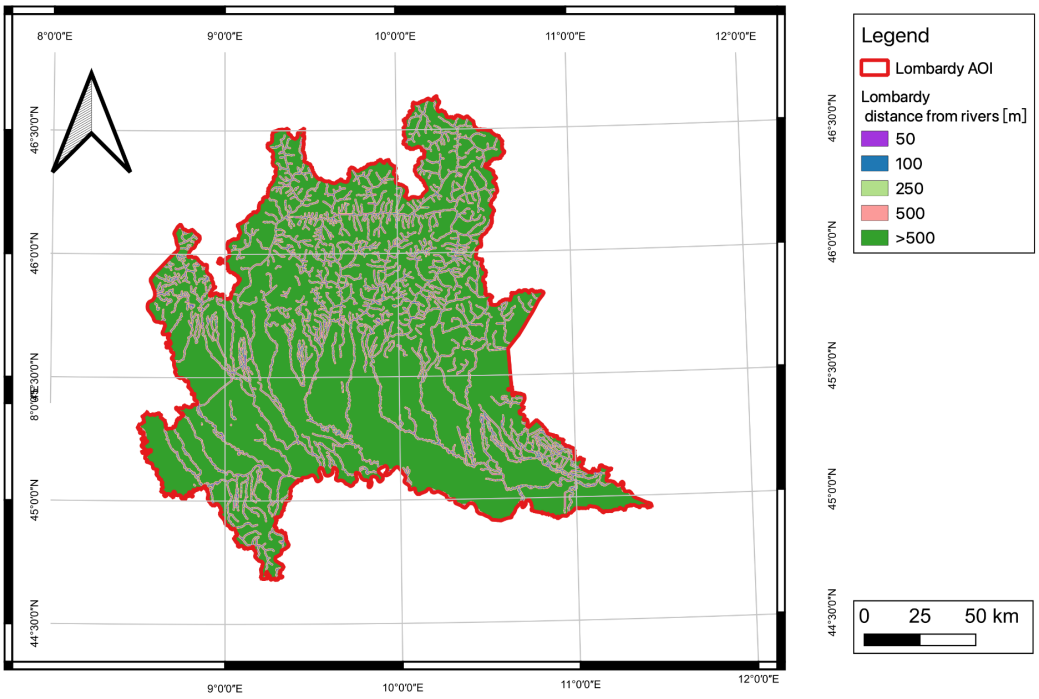
(b) Lombardy Eastness



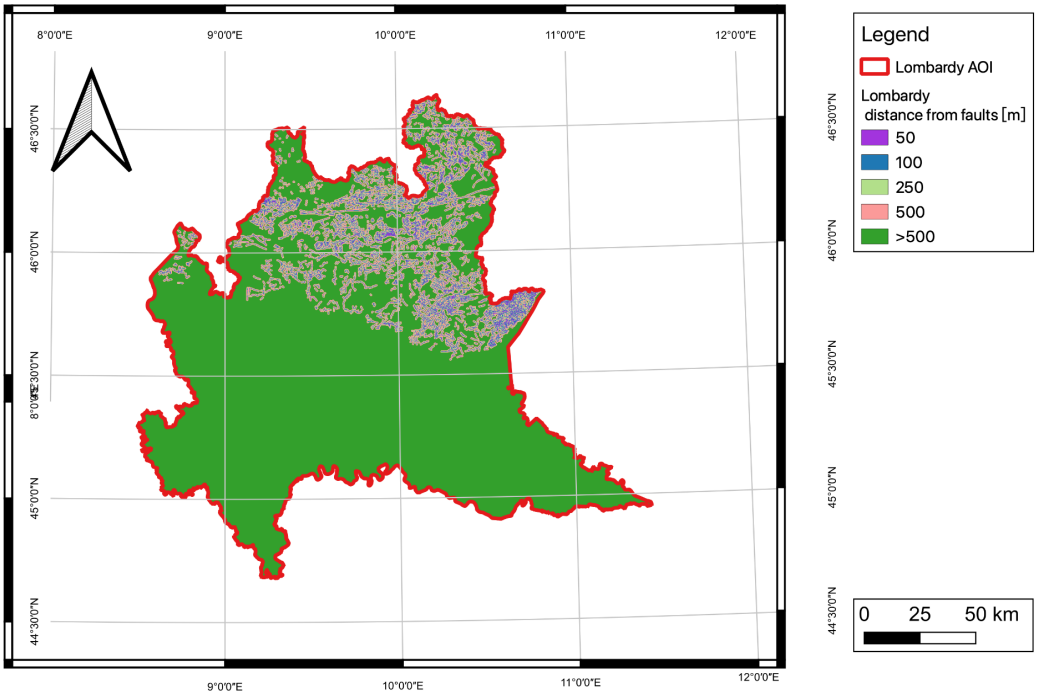
(c) Lombardy Northness



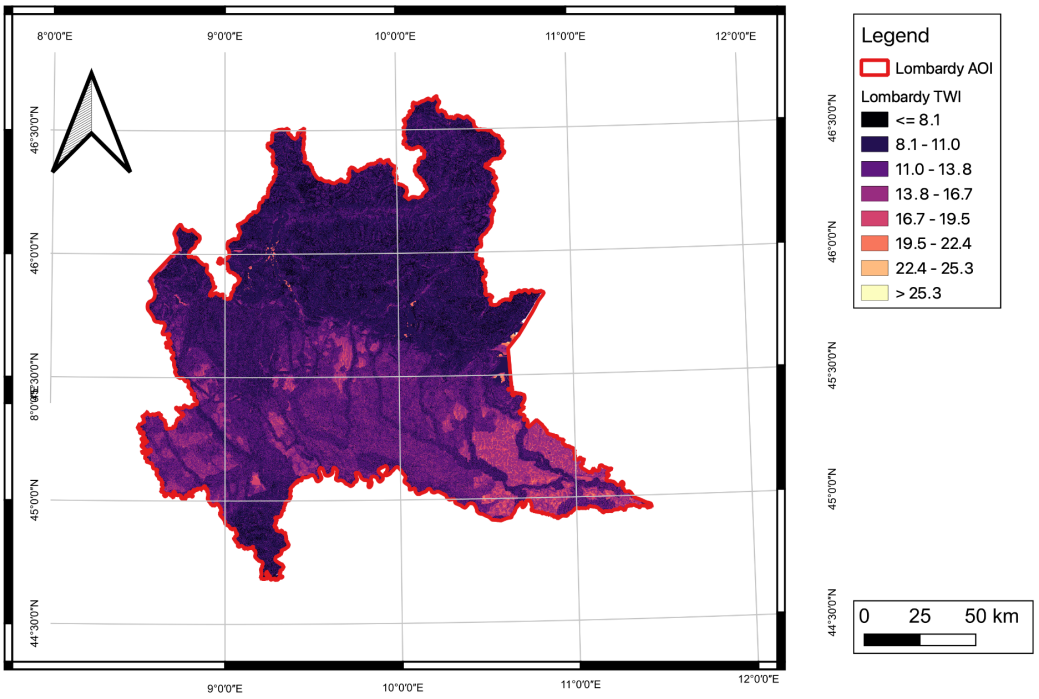
(d) Lombardy Distance from roads



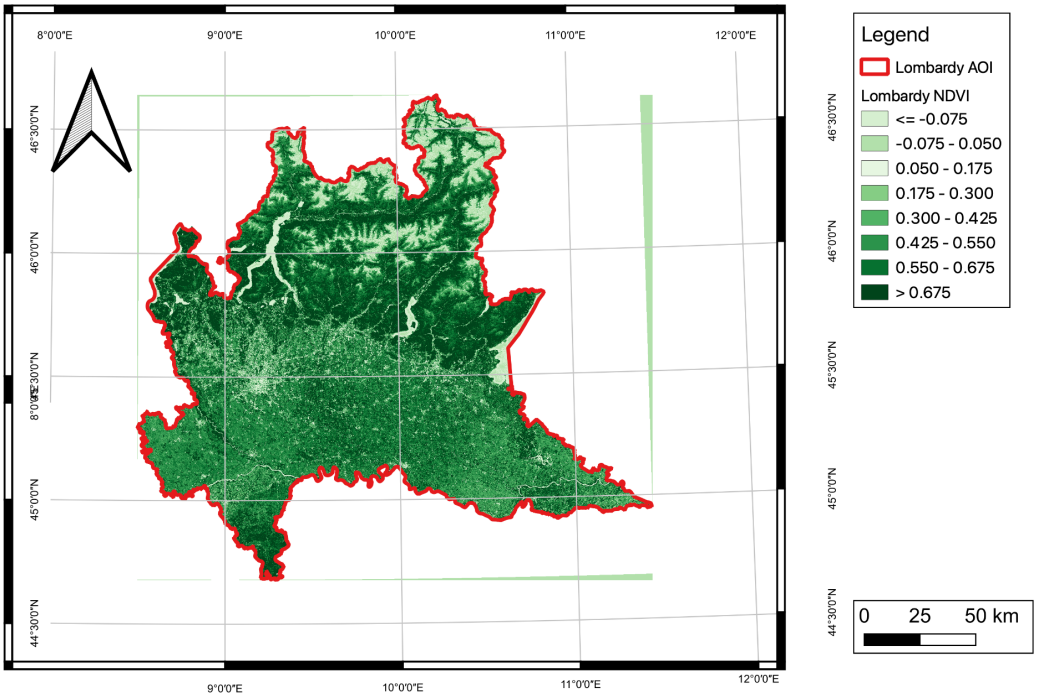
(e) Lombardy Distance from rivers



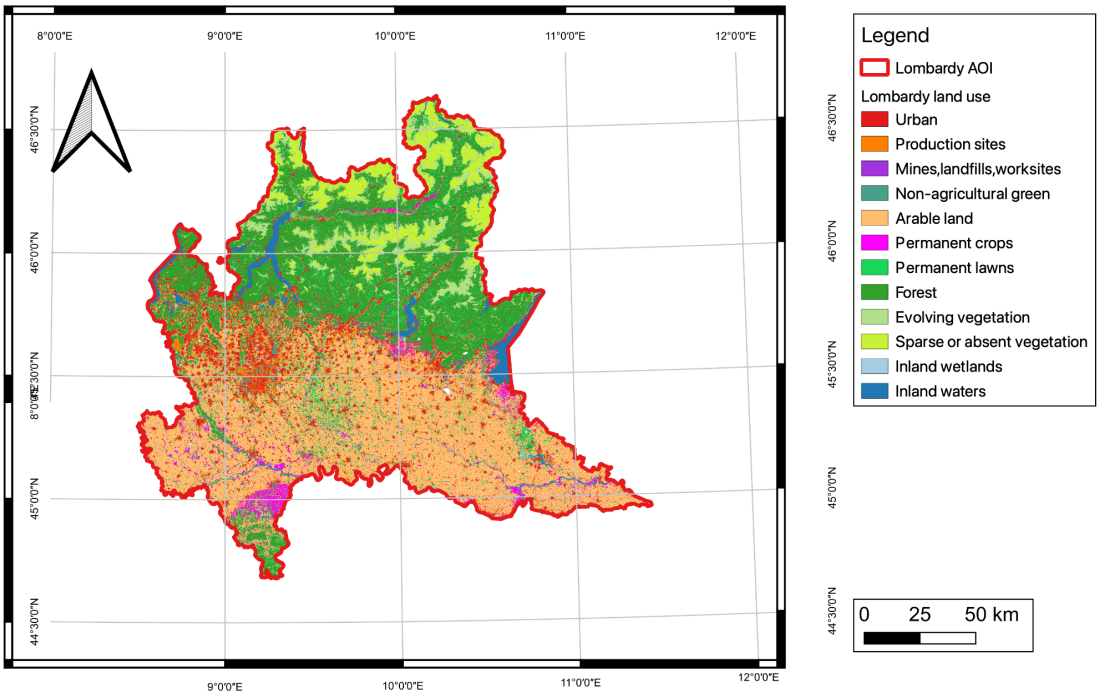
(f) Lombardy Distance from faults



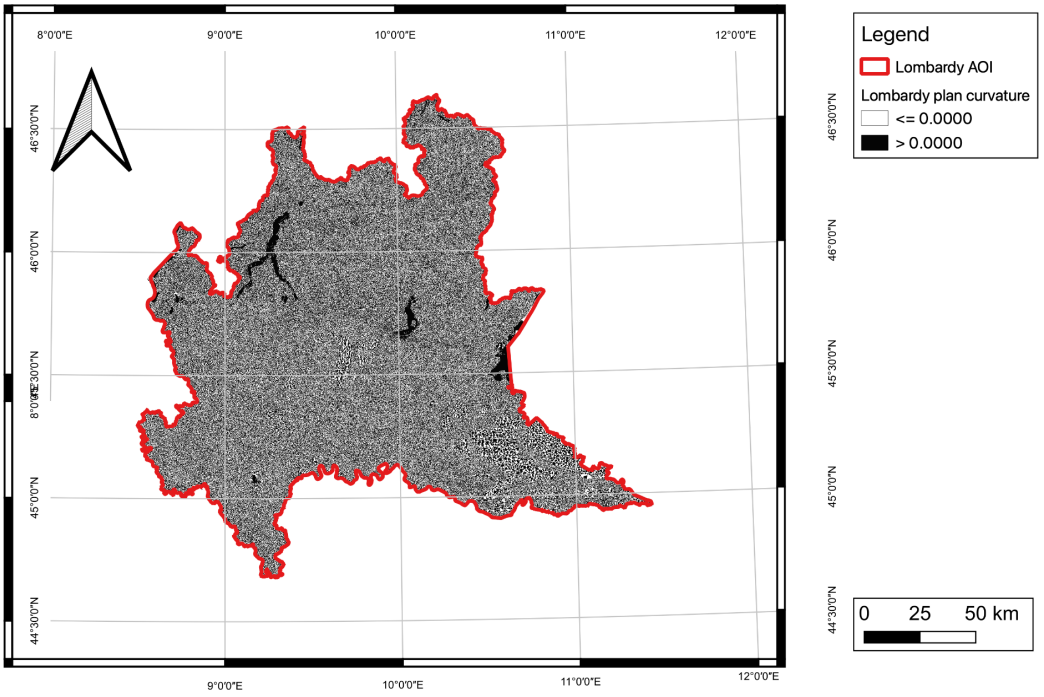
(g) Lombardy Topographic Wetness Index



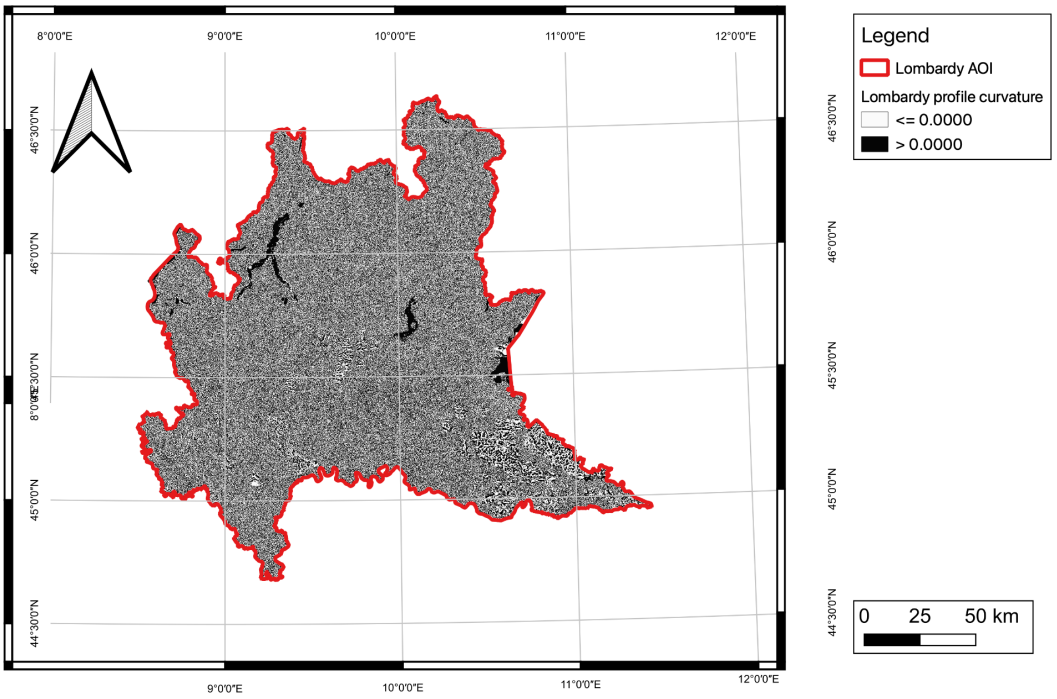
(h) Lombardy NDVI



(i) Lombardy Land use



(j) Lombardy Plan curvature



(k) Lombardy Profile curvature

Figure 5.5: Lombardy terrain variables

5.1.2 Landslide inventory

According to the preceding section, the landslide inventory established by ISPRA was utilized, with solely those layers containing polygonal and linear landslides being incorporated. During this process, an issue that necessitated resolution was that linear landslides have no area, thereby posing difficulties in the subsequent sampling stage. To address this concern, it was recognized that all linear landslides fall under the "debris flow" classification, typically transpiring in narrow valleys and channels. As per the solution proposed by Yordanov and Brovelli (2021), the width of linear landslides was estimated to be approximately 10 meters.

5.1.3 Factor sampling

Upon completion of the data preprocessing stage, it is necessary to generate a set of random data sampling points to sample the terrain variables. This collection comprises two types of data: one is from locations where landslides have taken place in the past (designated as the "Landslide zone"), assigned a value of 1; the other is from regions where the likelihood of landslide incidence is almost negligible (termed as the "No Landslide zone"), assigned a value of 0. A comprehensive account of this procedure, which encompasses the

definition of the Landslide zone and No Landslide zone, along with the point sampling process, is presented in the subsequent section.

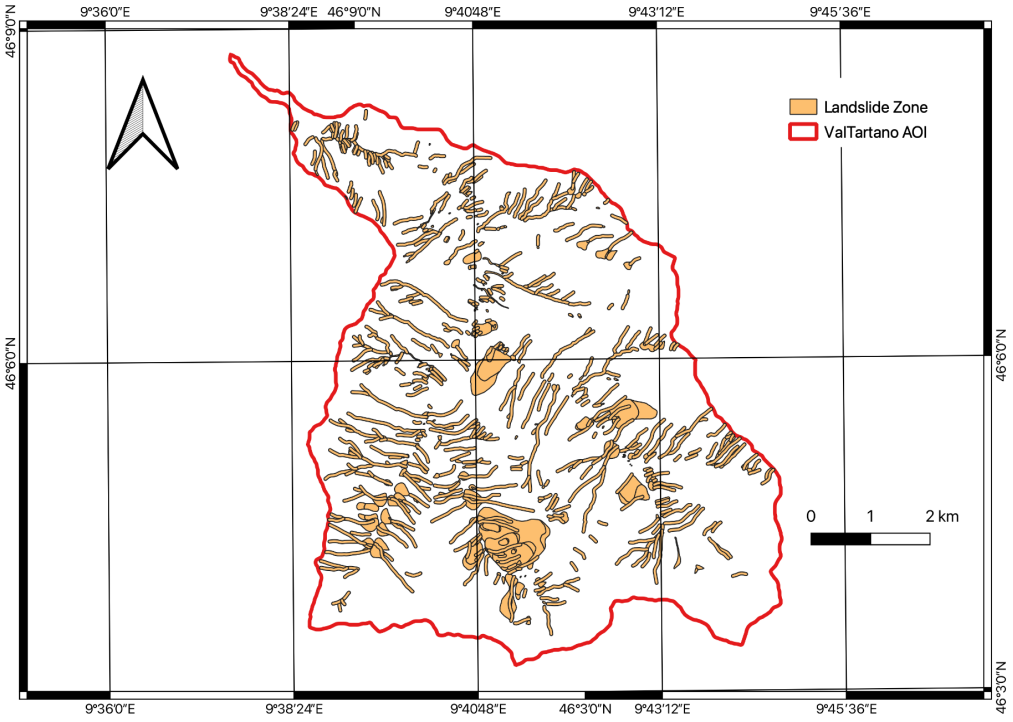
5.1.3.1 Landslide zone and No Landslide zone

This study consider only two types of shallow landslides, namely debris flows and rotational/translational slides, even though there are four types in total. Linear landslides are categorized as debris flow, and therefore, the LS zone was created by merging polygonal debris flows, polygonal rotational/translational slides, and buffered linear landslides. Table 5.2 shows some statistics on the type and number of landslides in all studied areas.

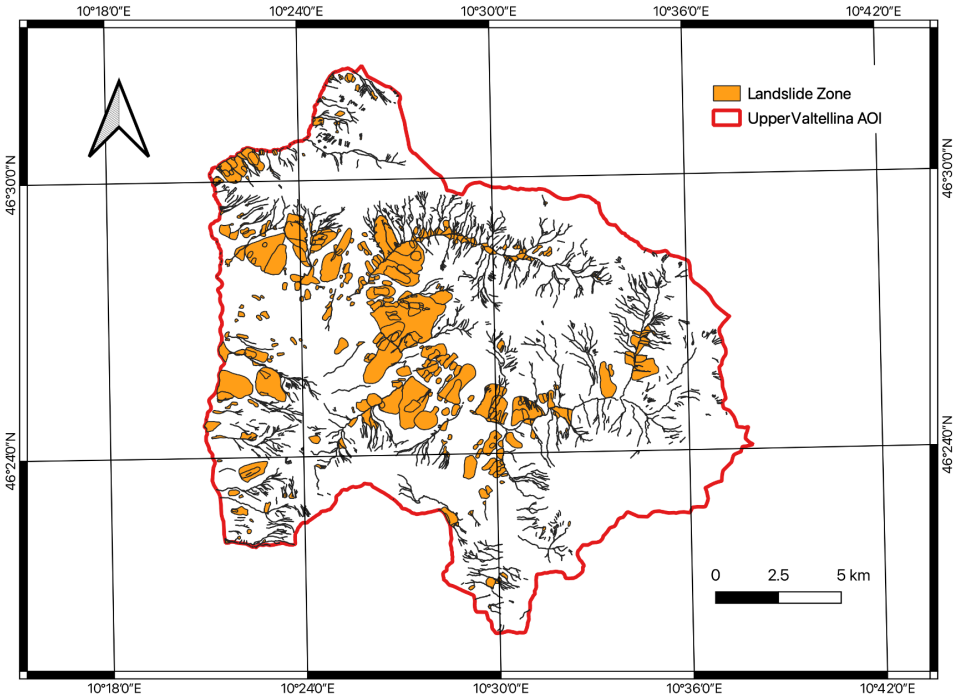
Zone	Landslide Type	Number of features
Val Tartano	Polygonal landslides	213
	Linear landslides	441
Upper Valtellina	Polygonal landslides	637
	Linear landslides	1363
Val Chiavenna	Polygonal landslides	1,050
	Linear landslides	3,641
Lombardy	Polygonal landslides	32,849
	Linear landslides	54,650

Table 5.2: Statistics on the type and number of landslides in the LS zone.

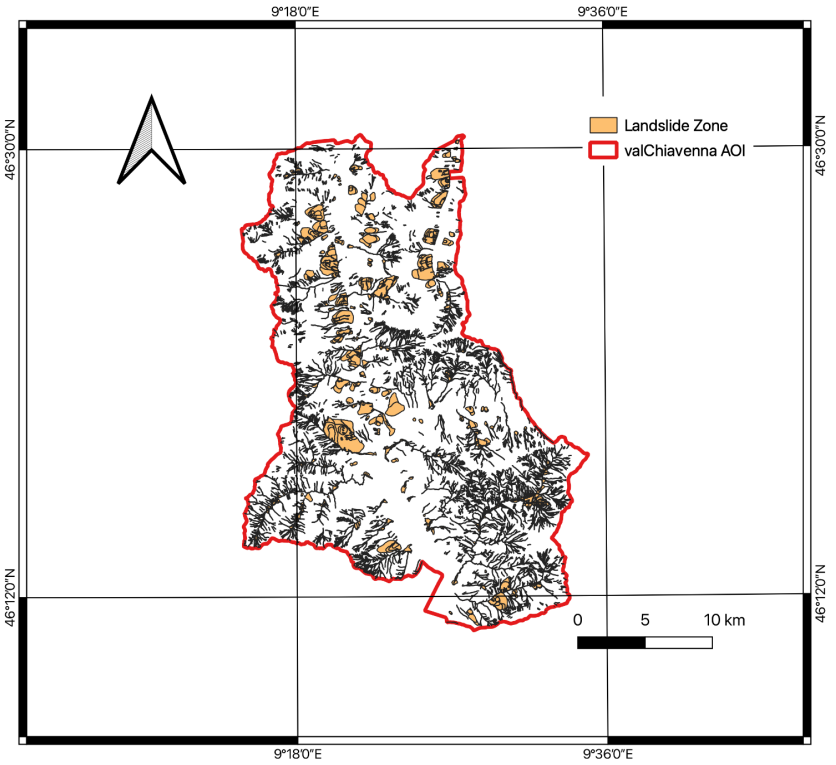
Figure 5.6 illustrates the LS zones for the studied areas.



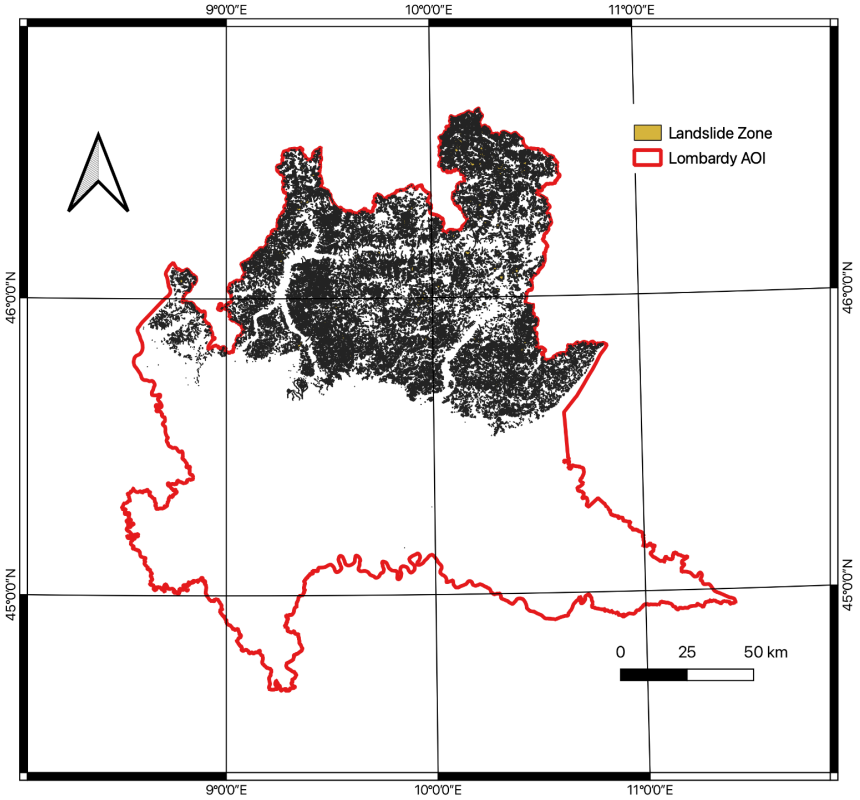
(a) Val Tartano LS zone



(b) Upper Valtellina LS zone



(c) Val Chiavenna LS zone



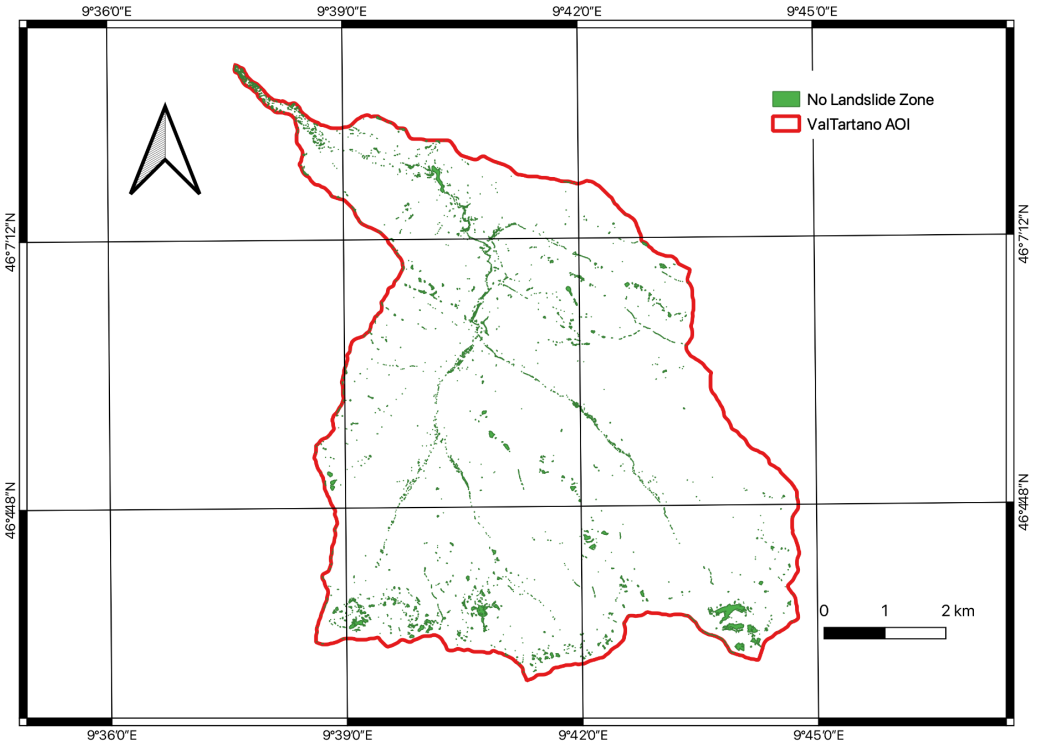
(d) Lombardy LS zone

Figure 5.6: Landslide zone of studied areas

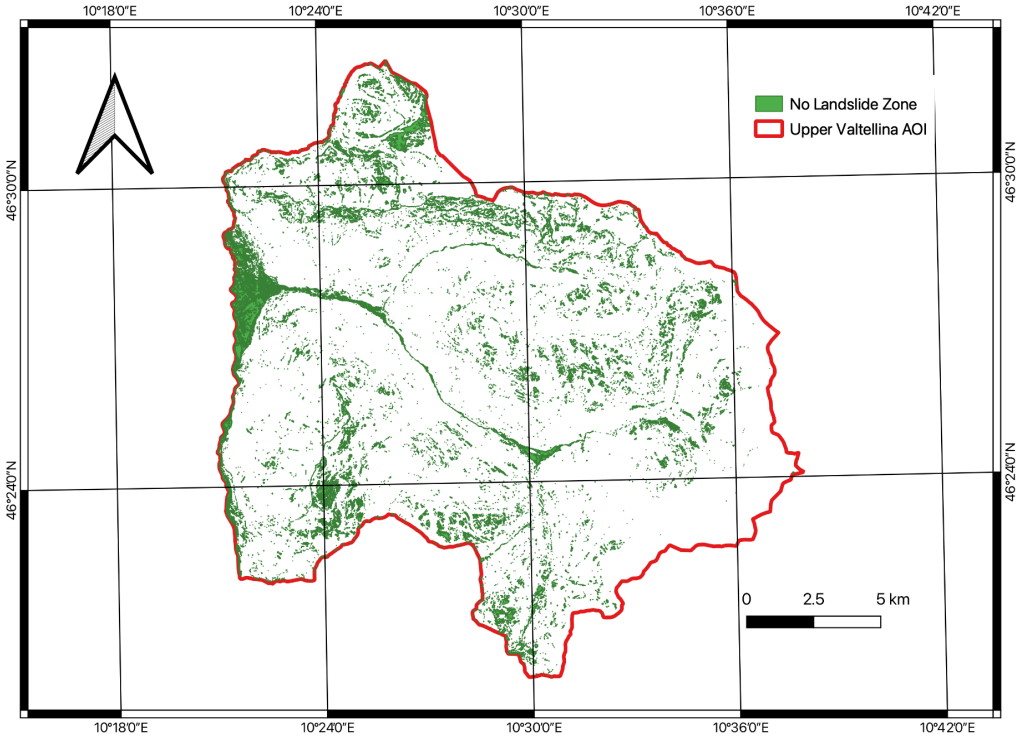
Additionally, the No Landslide zone (or NoLS zone) needs to be defined in order to ensure data integrity for areas where the probability of landslide occurance is very low. In this case, previous studies were taken into consideration,such as, Yordanov and Brovelli (2021) stated the hypothesis that a NoLS zone for shallow landslides should be an area with a slope angle smaller than 20° or larger than 70° which accounts for a very smooth or very steep slope. Amici (2021) made further refinement that the areas with very low (i.e. less than 5°) slopes (e.g. cities) must be included in the NoLS zone definition. Besides, the Intact Uniaxial Compressive Strength (IUCS) of materials, a value obtained by tests that describes the maximum stress that a material can take before failing should also be taken into account based on the terrain lithology. Therefore, the No Landslide zone is defined as follows (Amici, 2021):

$$(slope < 5^{o}) \vee [(5^{o} < slope < 20^{o} \vee slope > 70^{o}) \wedge (IUCS > 100\,MPa)]$$

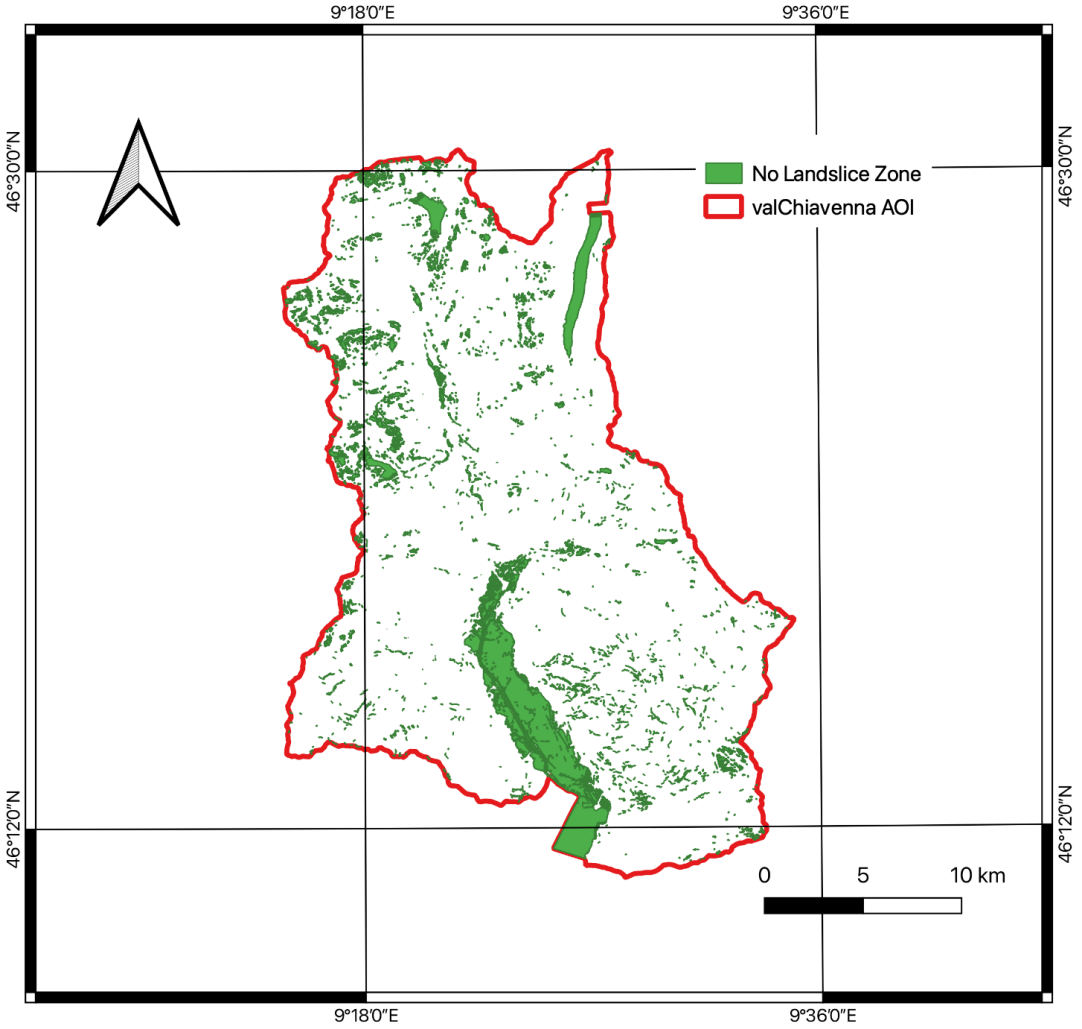
According to Amici (2021), the Upper Valtellina region has an overlapping ratio of 1.7% between the NoLS zone and the Landslide zone, while Val Tartano has a ratio of 0.5%. Similarly, Val Chiavenna has a ratio of 1.8% and Lombardy has a ratio of 2.3%. These overlap areas are negligible and support the NoLS hypothesis. These areas were eliminated from the NoLS zone before conducting the analysis. Figure 5.7 displays the final maps of the NoLS zones.



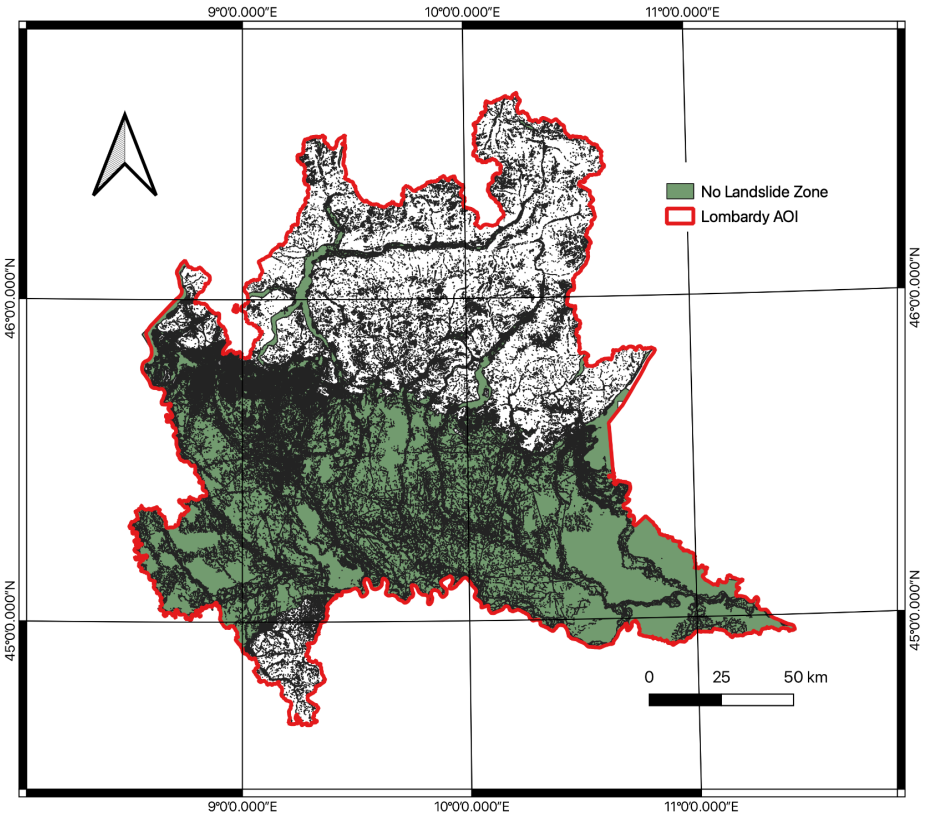
(a) Val Tartano NoLS zone



(b) Upper Valtellina NoLS zone


9°18'0"E
9°36'0"E
46°30'0"N
46°12'0"N
No Landslice Zone
valChiavenna AOI
0
5
10 km

(c) Val Chiavenna NoLS zone



(d) Lombardy NoLS zone

Figure 5.7: NoLS zone of studied areas

5.1.3.2 Point sampling

Having prepared all the necessary layers, the next step is to conduct point sampling. This procedure involved randomly selecting points at a 1:1 ratio within both the landslide and no-landslide zones. These points were then enriched with terrain data. Table 5.3 provides the number of sample points allocated for each area, taking into account the study area's size. Notably, no training dataset was needed for Lombardy since no model was trained for the entire region.

Zone	Number of sample points for training	Number of sample points for Testing	Source
Val Tartano	15,560	3,904	Geonode
Upper Valtellina	79,148	19,793	Geonode
Val Chiavenna	80,000	25,000	Computed
Lombardy	—	400,000	Computed

Table 5.3: Information of sampled points for each studied area, including both points for NoLS (class 0) and LS (class 1) .

5.2 Processing

To select the most suitable base classifiers, five models (Bagging, Random Forests, AdaBoost, Gradient Tree Boosting, and Neural Network) are trained and evaluated on different regions using the Python package scikit-learn. There are several scenarios based on the different training data sets and test sets used for this process (Table 5.5).

- **Val Tartano (VT)**: train and evaluate five models using the dataset sampled from *Val Tartano*;
- **Upper Valtellina (UV)**: train and evaluate five models using the dataset sampled from *Upper Valtellina*;
- **Val Chiavenna**
    - Case 1 (VCC1): train five models using the dataset sampled from *Val Tartano* and *Upper Valtellina*, and evaluate each model using the dataset sampled from *Val Chiavenna*;
    - Case 2 (VCC2): train five models using the dataset sampled from *Val Tartano, Upper Valtellina* and *Val Chiavenna*, and evaluate each model using the dataset sampled from *Val Chiavenna*;
    - Case 3 (VCC3): train and evaluate five models using the dataset sampled from *Val Chiavenna*;
- **Lombardy**
    - Case 1 (LC1): evaluate each model trained in above five cases (Val Tartano, Upper Valtellina, Val Chiavenna Case 1, Val Chiavenna Case 2, Val Chiavenna Case 3) using the dataset that sampled from the northern part of Lombardy but excluded from *Val Tartano, Upper Valtellina,* and *Val Chiavenna*. Northern Italy was chosen as the validation region because Lombardy as a whole has a high percentage of flat, NoLZ (about 49.48%), which biases the validation for class 1 (i.e., LS). Therefore, this portion is eliminated, leaving only the northern alpine region, which actually contains an abundance of LS.
    - Case 2 (LC2): evaluate each model trained in the above five cases (Val Tartano, Upper Valtellina, Val Chiavenna Case 1, Val Chiavenna Case 2, Val Chiavenna Case 3) using the dataset that sampled from Lombardy but excluded from *Val Tartano, Upper Valtellina,* and *Val Chiavenna*.

Case	Trained using data sampled from	Tested on data sampled from
Val Tartano (VT)	Val Tartano	Val Tartano
Upper Valtellina (UV)	Upper Valtellina	Upper Valtellina
Val Chiavenna Case 1 (VCC1)	Val Tartano, Upper Valtellina	Val Chiavenna
Val Chiavenna Case 1 (VCC2)	Val Tartano, Upper Valtellin, Val Chiavenna	Val Chiavenna

Val Chiavenna Case 1 (VCC3)	Val Chiavenna	Val Chiavenna
Northern Lombardy + Val Tartano (LC1+VT)	Val Tartano	Northern Lombardy
Northern Lombardy + Upper Valtellina (LC1+UV)	Upper Valtellina	Northern Lombardy
Northern Lombardy + Val Tartano, Upper Valtellina (LC1+VT+UV)	Val Tartano, Upper Valtellina	Northern Lombardy
Northern Lombardy + Val Chiavenna Case 2 (LC1+VCC2)	Val Tartano, Upper Valtellina, Val Chiavenna	Northern Lombardy
Northern Lombardy + Val Chiavenna Case 3 (LC1+VCC3)	Val Chiavenna	Northern Lombardy
Lombardy + Val Tartano (LC2+VT)	Val Tartano	Lombardy
Lombardy + Upper Valtellina (LC2+UV)	Upper Valtellina	Lombardy
Lombardy + Val Tartano, Upper Valtellina (LC1+VT+UV)	Val Tartano, Upper Valtellina	Lombardy
Lombardy + Val Chiavenna Case 2 (LC2+VCC2)	Val Tartano, Upper Valtellina, Val Chiavenna	Lombardy
Lombardy + Val Chiavenna Case 3 (LC2+VCC3)	Val Chiavenna	Lombardy

Table 5.5: source of training dataset and testing dataset

Then, three selected models based on accuracy are used in different ensemble methods (stacking, blending, and voting). This process is also based on the scenarios used in model selection. At the same time, LSMs for different regions will also be generated.

The model with the best performance is chosen after a comparison of all the produced models, including base models and ensemble models. Then, this model is trained once more using precipitation data (90th percentile and average) and evaluated.

6. Results

The initial focus of this section is on the outcomes obtained from both the base classifiers and the ensemble models, along with the corresponding decision-making process. Subsequently, the suitability of the models is assessed, and the landslide susceptibility maps generated using the most appropriate ones are presented and analyzed.

6.1 Performance evaluation

6.1.1 Performance evaluation of base classifiers

Effective selection of optimal base classifiers plays a critical role in achieving successful Landslide Susceptibility Mapping (LSM) using ensemble models. To assess the performance of each base classifier, we evaluated its efficacy when applied to different study areas.

Figure 6.1 shows the calibration curves for all base models in Val Tartano, with the average predicted probability for each bin on the x-axis and the fraction of positive classes in each bin on the y-axis.

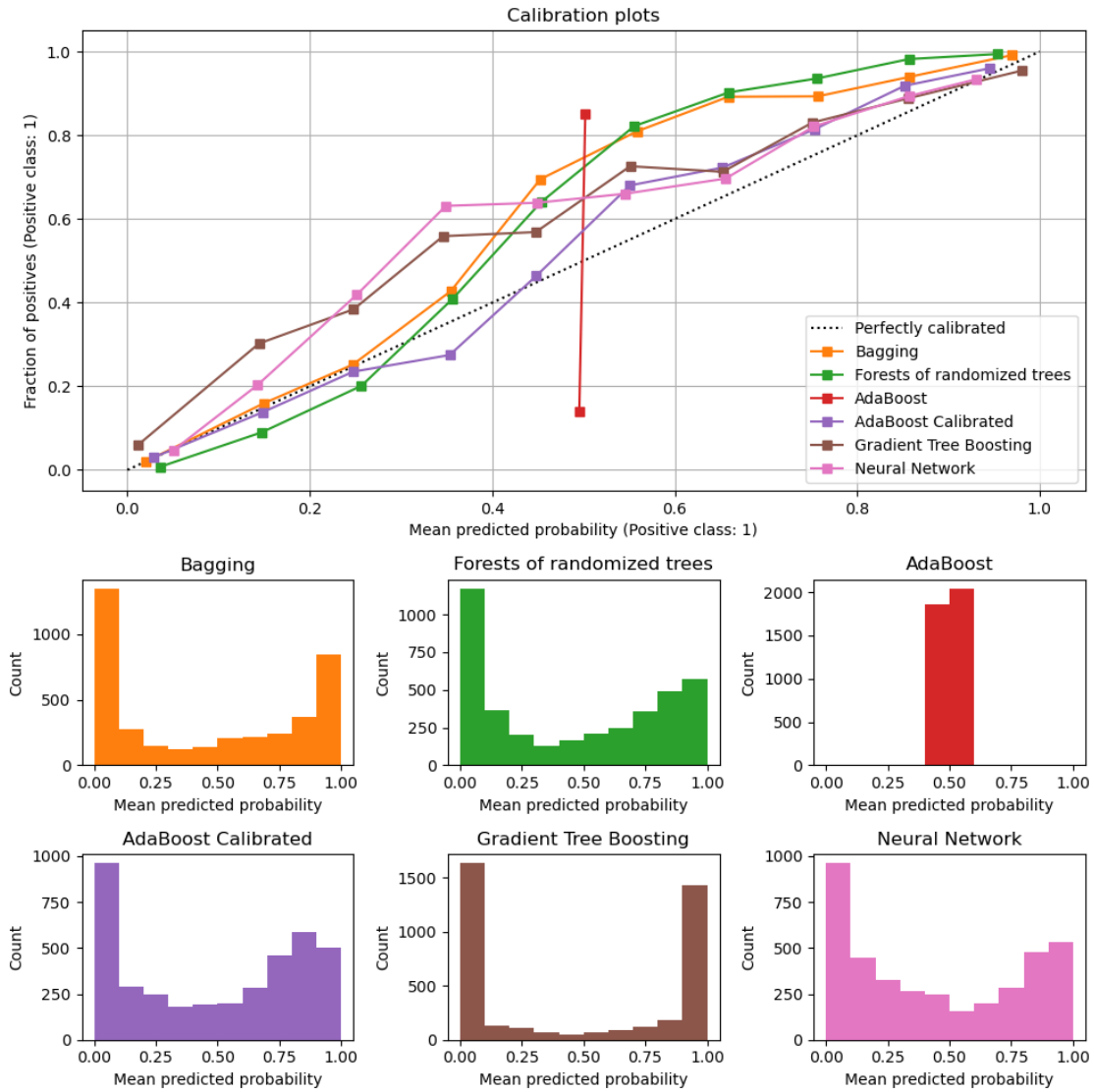


Figure 6.1: Calibration plots of base classifiers for Val Tartano

Please refer to Appendix A for the calibration plots of all classifiers.

In all cases, Bagging, RF, Gradient Tree Boosting, and NN classifiers exhibit well-calibrated predictions. This is evidenced by the proximity of all curves generated by these models to the diagonal line, which represents perfect calibration.

In contrast to other classifiers, AdaBoost exhibits a distinct pattern where the probability histograms display peaks at around 0.4 and 0.6 probabilities, while probabilities at other levels are relatively infrequent. A possible explanation for this phenomenon is that AdaBoost can be viewed as an additive logistic regression model, in which the predictions made by boosting attempt to fit a logit of the true probabilities instead of the actual probabilities themselves (Friedman et al., 2000). As a result, it is essential to calibrate this classifier in all cases. The calibration result of AdaBoost with sigmoid regression is also illustrated in Figure 6.1 and shows well-calibrated behavior.

Different accuracy statistics (e.g., OA, precision, recall, and F-measure) of ML base classifiers for different cases were calculated (see Table 6.1 and Figure 6.2). Compared to the uncalibrated version, the calibrated version of AdaBoost does not significantly alter prediction accuracy measures. Therefore, the calibrated AdaBoost will be used in place of the uncalibrated AdaBoost in subsequent analyses.

Cases	Bagging	Forests of randomized trees	AdaBoost	AdaBoost Calibrated	Gradient Tree Boosting	Neural Network
VT	90.37%	91.44%	85.48%	85.60%	89.75%	78.25%
UV	90.70%	90.89%	88.32%	88.57%	89.54%	87.12%
VCC2	94.36%	94.34%	89.61%	89.64%	93.30%	89.01%
VCC3	94.76%	95.07%	92.22%	92.25%	94.76%	86.33%
VCC1	69.28%	75.16%	78.82%	78.81%	68.41%	79.66%
LC1+VT	80.08%	76.58%	78.68%	79.05%	64.77%	71.99%
LC1+UV	66.92%	76.34%	75.89%	77.18%	60.55%	75.91%
LC1+VT+UV	71.66%	77.33%	84.93%	85.04%	70.07%	75.69%
LC1+VCC2	77.11%	81.31%	80.86%	80.87%	73.34%	85.83%
LC1+VCC3	68.10%	67.97%	60.99%	60.62%	61.68%	90.82%
LC2+VT	83.03%	78.47%	80.34%	80.68%	65.89%	73.24%
LC2+UV	66.00%	78.57%	80.99%	81.70%	66.83%	77.38%
LC2+VT+UV	74.17%	79.53%	87.07%	87.20%	71.00%	77.09%
LC2+VCC2	82.84%	89.67%	87.48%	88.05%	80.91%	93.10%
LC2+VCC3	89.91%	81.77%	55.90%	55.56%	72.13%	97.01%

(a) Overall Accuracy

Cases	Bagging		Forests of randomized trees		AdaBoost		AdaBoost Calibrated		Gradient Tree Boosting		Neural Network	
	OA	Thresh old	OA	Thresh old	OA	Thresh old	OA	Thresh old	OA	Thresh old	OA	Thresh old
VT	91.78 %	0.4100	92.90 %	0.3700	85.66 %	0.4998	85.86 %	0.4749	90.24 %	0.3343	82.25 %	0.2637
UV	90.93 %	0.4167	91.36 %	0.3900	88.53 %	0.5001	88.57 %	0.5031	89.66 %	0.4525	87.27 %	0.4637
VCC2	94.40 %	0.5500	94.36 %	0.5200	89.81 %	0.5002	89.78 %	0.5282	93.42 %	0.5596	89.05 %	0.4861
VCC3	94.76 %	0.5500	95.15 %	0.5400	92.29 %	0.5001	92.35 %	0.5151	94.82 %	0.5257	86.36 %	0.4574
VCC1	70.50 %	0.5300	75.17 %	0.5000	79.60 %	0.4990	79.58 %	0.2818	69.23 %	0.5899	86.00 %	0.0018

LC1+VT	83.06 %	0.4300	80.44 %	0.4250	91.14 %	0.4899	90.91 %	0.0030	80.03 %	0.0005	89.95 %	0.0073
LC1+UV	71.34 %	0.3725	85.33 %	0.3600	75.90 %	0.5000	77.28 %	0.5615	62.46 %	0.8353	84.95 %	0.0022
LC1+VT+UV	77.66 %	0.1625	83.39 %	0.3620	91.34 %	0.4839	89.80 %	0.2000	82.69 %	0.0066	92.53 %	0.0307
LC1+VCC2	82.49 %	0.7195	85.35 %	0.6533	82.64 %	0.5010	82.60 %	0.7043	76.05 %	0.7882	87.11 %	0.6571
LC1+VCC3	76.22 %	0.8800	82.43 %	0.8000	70.32 %	0.5026	69.56 %	0.9087	64.13 %	0.9196	92.35 %	0.7087
LC2+VT	88.91 %	0.2000	83.55 %	0.4260	95.03 %	0.4780	94.70 %	0.0015	82.87 %	0.0005	95.84 %	0.0012
LC2+UV	73.37 %	0.2913	88.12 %	0.3667	81.26 %	0.4995	82.17 %	0.3131	68.93 %	0.7926	93.47 %	0.0017
LC2+VT+UV	79.06 %	0.3025	87.39 %	0.3560	95.16 %	0.4827	93.40 %	0.0000	86.80 %	0.0036	96.76 %	0.0020
LC2+VCC2	91.07 %	0.6550	91.48 %	0.6117	88.70 %	0.5004	89.07 %	0.5730	82.53 %	0.7057	93.13 %	0.4843
LC2+VCC3	90.30 %	0.6100	92.38 %	0.6800	68.71 %	0.5029	67.54 %	0.9199	75.75 %	0.8502	97.02 %	0.5148

(b) Accuracy based on optimum ROC-based thresholds

Cases	Bagging		Forests of randomized trees		AdaBoost		AdaBoost Calibrated		Gradient Tree Boosting		Neural Network	
	OA	Threshold	OA	Threshold	OA	Threshold	OA	Threshold	OA	Threshold	OA	Threshold
VT	91.78 %	0.3500	92.90 %	0.3700	85.48 %	0.4996	85.84 %	0.4589	90.22 %	0.2942	82.25 %	0.2576
UV	90.86 %	0.3800	91.32 %	0.3800	88.31 %	0.4998	88.33 %	0.3942	89.48 %	0.3774	87.01 %	0.3553
VCC2	94.39 %	0.5300	94.36 %	0.5200	89.81 %	0.5002	89.78 %	0.5282	93.42 %	0.5545	88.96 %	0.4310
VCC3	94.76 %	0.5100	95.15 %	0.5300	92.24 %	0.5000	92.35 %	0.5151	94.82 %	0.5257	86.34 %	0.2818
VCC1	70.50 %	0.5300	74.68 %	0.4620	78.67 %	0.4846	79.20 %	0.2000	66.66 %	0.1860	85.99 %	0.0016
LC1+VT	82.11 %	0.2000	80.09 %	0.4033	91.13 %	0.4783	90.88 %	0.0022	79.52 %	0.0002	89.90 %	0.0060
LC1+UV	70.86 %	0.2813	85.33 %	0.3467	75.44 %	0.4995	76.84 %	0.3357	51.32 %	0.0000	84.77 %	0.0019
LC1+VT+UV	77.55	0.1525	83.26	0.3325	91.33	0.4837	89.78	0.2000	82.57	0.0031	92.52	0.0221

	%		%		%		%		%		%	
LC1+VCC2	82.21 %	0.6575	85.21 %	0.6333	82.13 %	0.5004	81.99 %	0.5744	75.28 %	0.6544	87.03 %	0.6192
LC1+VCC3	75.63 %	0.8000	82.18 %	0.7600	68.59 %	0.5018	67.54 %	0.8152	61.25 %	0.4377	92.33 %	0.6911
LC2+VT	88.90 %	0.1900	83.24 %	0.4060	95.03 %	0.4780	94.67 %	0.0012	82.43 %	0.0002	95.84 %	0.0011
LC2+UV	73.37 %	0.2913	88.04 %	0.3567	80.89 %	0.4982	81.83 %	0.2002	66.61 %	0.4721	93.45 %	0.0016
LC2+VT+UV	78.95 %	0.1625	87.23 %	0.3350	95.16 %	0.4825	93.39 %	0.0000	86.79 %	0.0033	96.76 %	0.0020
LC2+VCC2	91.05 %	0.6450	91.48 %	0.6033	88.66 %	0.5003	89.05 %	0.5600	82.25 %	0.6214	93.10 %	0.4457
LC2+VCC3	90.28 %	0.5900	92.38 %	0.6700	65.41 %	0.5019	65.17 %	0.8430	75.05 %	0.7317	97.02 %	0.5148

(c) Accuracy based on optimum PR-based thresholds

Cases	Bagging	Forests of randomized trees	AdaBoost	AdaBoost Calibrated	Gradient Tree Boosting	Neural Network
VT	0.9361	0.9483	0.8507	0.8545	0.9197	0.8482
UV	0.9466	0.9482	0.8944	0.9010	0.9164	0.9033
VCC2	0.9378	0.9361	0.8707	0.8718	0.9273	0.8653
VCC3	0.9441	0.9460	0.9109	0.9111	0.9495	0.7972
VCC1	0.6870	0.7501	0.7856	0.7863	0.6821	0.8419
LC1+VT	0.8918	0.9197	0.9386	0.9394	0.9145	0.9451
LC1+UV	0.7618	0.8958	0.7087	0.7297	0.5777	0.9386
LC1+VT+UV	0.8761	0.9147	0.9324	0.9319	0.8974	0.9412
LC1+VCC2	0.7021	0.7505	0.7504	0.7505	0.6681	0.8206
LC1+VCC3	0.6142	0.6131	0.5643	0.5617	0.5731	0.8611
LC2+VT	0.9657	0.9754	0.9865	0.9867	0.9753	0.9887
LC2+UV	0.7400	0.9588	0.7717	0.7895	0.6368	0.9873
LC2+VT+UV	0.9574	0.9777	0.9855	0.9853	0.9299	0.9879
LC2+VCC2	0.7676	0.8668	0.8401	0.8488	0.7498	0.9424
LC2+VCC3	0.8468	0.7416	0.5326	0.5305	0.6592	0.9664

(d) Precision

Cases	Bagging	Forests of randomized trees	AdaBoost	AdaBoost Calibrated	Gradient Tree Boosting	Neural Network
VT	0.8715	0.8810	0.8690	0.8665	0.8765	0.7010
UV	0.8640	0.8662	0.8708	0.8683	0.8716	0.8333
VCC2	0.9502	0.9518	0.9303	0.9294	0.9398	0.9241
VCC3	0.9516	0.9559	0.9361	0.9363	0.9456	0.9746
VCC1	0.7082	0.7544	0.7927	0.7912	0.6896	0.7302
LC1+VT	0.6847	0.5825	0.6139	0.6210	0.3260	0.4669
LC1+UV	0.4923	0.5962	0.8792	0.8633	0.7842	0.5545
LC1+VT+UV	0.5045	0.6029	0.7533	0.7560	0.4534	0.5480
LC1+VCC2	0.9419	0.9379	0.9248	0.9248	0.9279	0.9172
LC1+VCC3	0.9736	0.9743	0.9641	0.9671	0.9162	0.9736
LC2+VT	0.6849	0.5842	0.6153	0.6219	0.3260	0.4702
LC2+UV	0.4933	0.5970	0.8800	0.8646	0.7837	0.5548
LC2+VT+UV	0.5059	0.6045	0.7524	0.7553	0.4543	0.5485
LC2+VCC2	0.9419	0.9375	0.9258	0.9259	0.9278	0.9182
LC2+VCC3	0.9746	0.9750	0.9651	0.9681	0.9164	0.9740

(e) Recall

Cases	Bagging	Forests of randomized trees	AdaBoost	AdaBoost Calibrated	Gradient Tree Boosting	Neural Network
VT	0.9026	0.9134	0.8598	0.8605	0.8976	0.7676
UV	0.9034	0.9054	0.8825	0.8843	0.8935	0.8669
VCC2	0.9439	0.9439	0.8995	0.8997	0.9335	0.8937
VCC3	0.9478	0.9509	0.9233	0.9235	0.9475	0.8770
VCC1	0.6974	0.7523	0.7892	0.7887	0.6858	0.7821
LC1+VT	0.7746	0.7133	0.7423	0.7477	0.4806	0.6250
LC1+UV	0.5981	0.7159	0.7848	0.7909	0.6653	0.6972
LC1+VT+UV	0.6403	0.7268	0.8333	0.8348	0.6024	0.6927
LC1+VCC2	0.8045	0.8338	0.8286	0.8286	0.7768	0.8662
LC1+VCC3	0.7532	0.7526	0.7119	0.7106	0.7051	0.9139
LC2+VT	0.8014	0.7307	0.7579	0.7629	0.4887	0.6373
LC2+UV	0.5920	0.7358	0.8223	0.8253	0.7026	0.7104
LC2+VT+UV	0.6620	0.7470	0.8533	0.8551	0.6104	0.7054
LC2+VCC2	0.8459	0.9007	0.8809	0.8857	0.8294	0.9301

LC2+VCC3	0.9062	0.8424	0.6864	0.6854	0.7668	0.9702

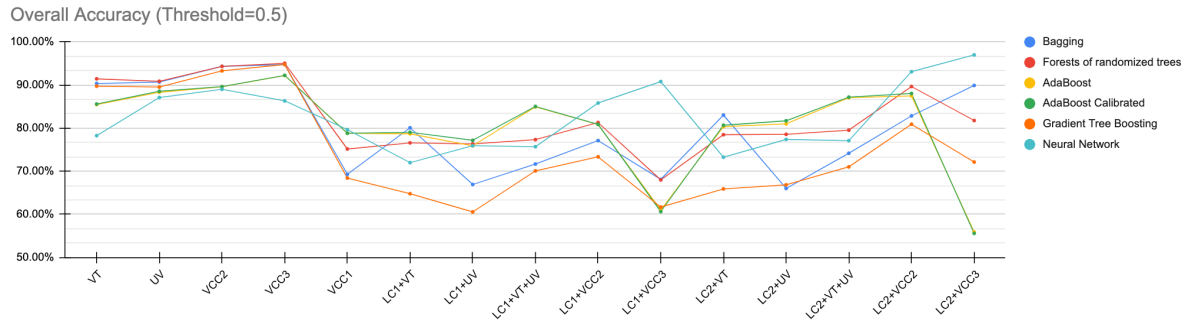
(f) F_1 score

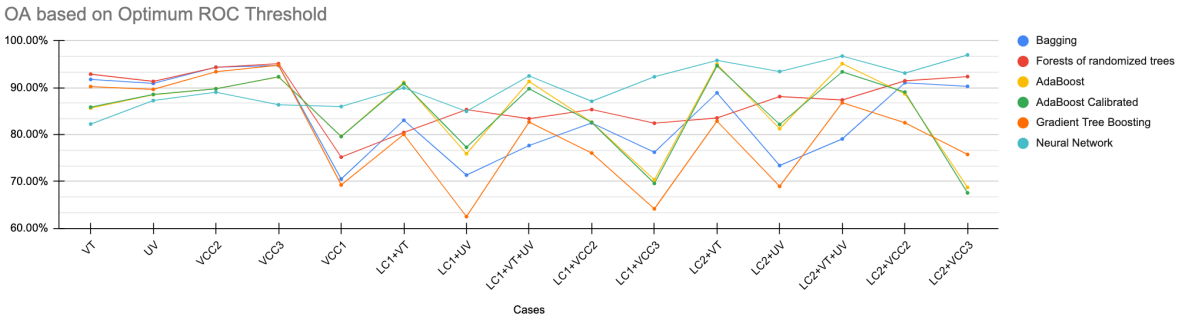
Table 6.1: Accuracy statistics of base classifiers for different cases. Table cells are colored based on each region. The darker the color, the higher the value

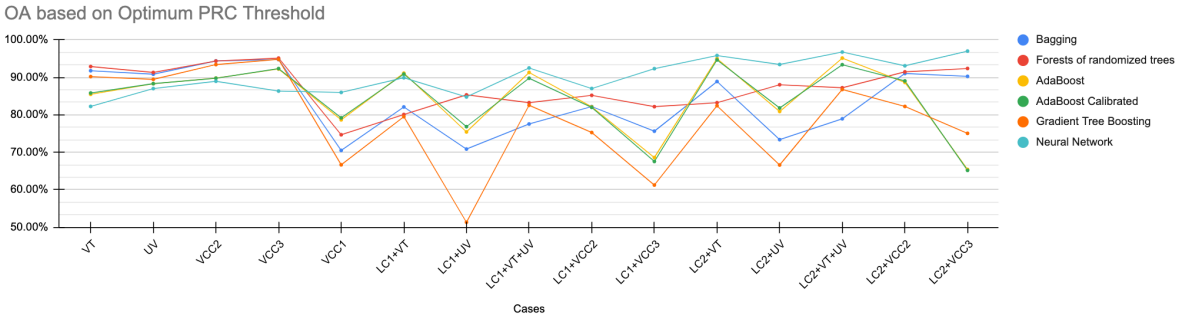
Please refer to Appendix C for the ROC and PRC of all classifiers.

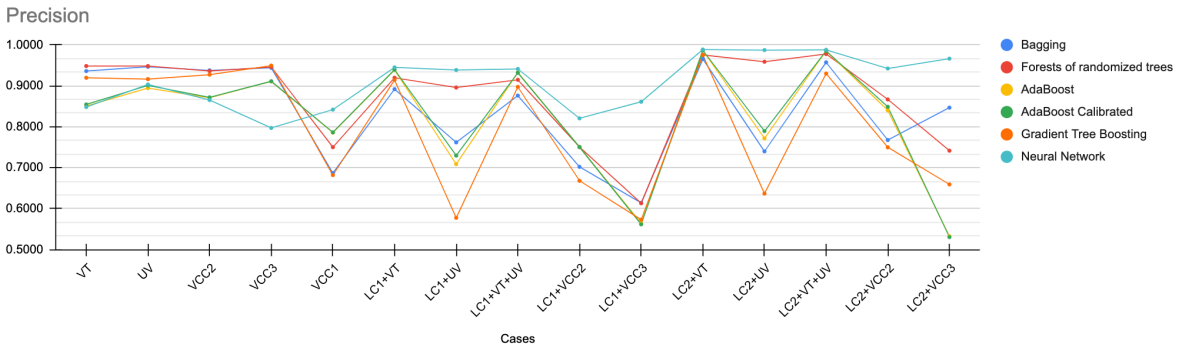
Bagging, RF, and Gradient Tree Boosting achieved the highest accuracy (i.e. 0.9~0.93 for VT, 0.9~0.91 for UV, 0.93~0.94 for VCC2, and 0.95 for VCC3) when sampled data from the relevant area (i.e., VT, UV, VCC2, and VCC3) were included for model training, and the accuracy of the NN model is the worst (i.e., 0.82 based on an optimum threshold for VT, 0.87 for UV, 0.89 for VCC2, and 0.86 for VCC3). The most accurate model was switched to NN (i.e., 0.86 for VCC1, 0.85~0.93 for LC1, and 0.93~0.97 for LC2 based on the optimum threshold), followed by RF (i.e., around 0.75 for VCC1, 0.8～0.85 for LC1 based on optimum threshold, and 0.83~0.92 for LC2 based on an optimum threshold) when the sampling data of the pertinent areas (i.e., VCC1, LC1, and LC2) were not used to train the model. The third model with the best performance in terms of overall accuracy is calibrated AdaBoost with around 0.79 for VCC1. In the case of Lombardy, the accuracy varies when using calibrated AdaBoost models, with LC1 having the highest accuracy of 0.91 and LC2 having the highest accuracy of 0.95.

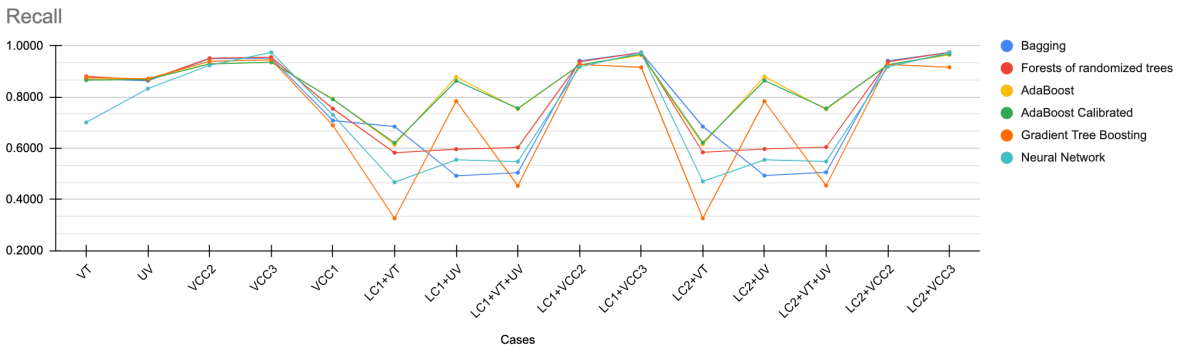
In Val Tartano and Upper Valtellina, RF generated the highest results (0.93 when using the best threshold for VT and 0.91 for UV regardless of the threshold selected), but it was only marginally superior to Bagging (i.e., just a 1% improvement for VT and the maximum increase of 0.4% for UV). The most accurate models for Val Chiavenna were Bagging, RF, and Gradient Tree Boosting models trained in VCC3. The accuracy of the three models is extremely close to 0.95 regardless of the threshold that is chosen. When adopting the optimum threshold, the NN models trained in VCC3 and VT+UV showed relatively similar accuracy and the best performance for Lombardy. The NN model that was trained using VT+UV, however, is predicated on a high accuracy on a threshold that is near 0. As a result, while taking into account the rationality of the threshold, the NN model trained in VCC3 performs better.




OA based on Optimum ROC Threshold
100.00%
90.00%
80.00%
70.00%
60.00%
VT
UV
VCC2
VCC3
VCC1
LC1+VT
LC1+UV
LC1+VT+UV
LC1+VCC2
LC1+VCC3
LC2+VT
LC2+UV
LC2+VT+UV
LC2+VCC2
LC2+VCC3
Cases
Bagging
Forests of randomized trees
AdaBoost
AdaBoost Calibrated
Gradient Tree Boosting
Neural Network


OA based on Optimum PRC Threshold
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
VT
UV
VCC2
VCC3
VCC1
LC1+VT
LC1+UV
LC1+VT+UV
LC1+VCC2
LC1+VCC3
LC2+VT
LC2+UV
LC2+VT+UV
LC2+VCC2
LC2+VCC3
Cases
Bagging
Forests of randomized trees
AdaBoost
AdaBoost Calibrated
Gradient Tree Boosting
Neural Network


Precision
1.0000
0.9000
0.8000
0.7000
0.6000
0.5000
VT
UV
VCC2
VCC3
VCC1
LC1+VT
LC1+UV
LC1+VT+UV
LC1+VCC2
LC1+VCC3
LC2+VT
LC2+UV
LC2+VT+UV
LC2+VCC2
LC2+VCC3
Cases
Bagging
Forests of randomized trees
AdaBoost
AdaBoost Calibrated
Gradient Tree Boosting
Neural Network


Recall
1.0000
0.8000
0.6000
0.4000
0.2000
VT
UV
VCC2
VCC3
VCC1
LC1+VT
LC1+UV
LC1+VT+UV
LC1+VCC2
LC1+VCC3
LC2+VT
LC2+UV
LC2+VT+UV
LC2+VCC2
LC2+VCC3
Cases
Bagging
Forests of randomized trees
AdaBoost
AdaBoost Calibrated
Gradient Tree Boosting
Neural Network

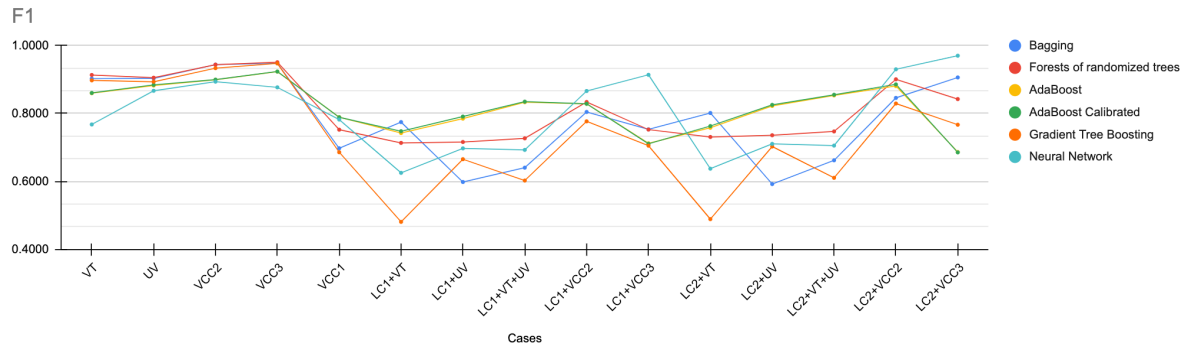


Figure 6.2: Accuracy statistics of base classifiers for different cases

Compare model evaluation results for specific basins (i.e., VT, UV, VCC3) using the training dataset with using the test dataset to investigate why Neural Network perform worst when training models using sampled data from relevant regions but perform best in the opposite scenario. The evaluation results (Table 6.2) shows that, regardless of the type of dataset used to conduct the evaluation, the accuracy performance is not significantly different. This demonstrates that there are no overfitting issues with the trained model.

Cases	Evaluation based on training dataset			Evaluation based on testing dataset		
	Neural Network			Neural Network		
	Base OA	Optimal ROC OA (Threshold)	Optimal PRC OA (Threshold)	Base OA	Optimal ROC OA (Threshold)	Optimal PRC OA (Threshold)
VT	86.08%	86.15%	86.01%	78.25%	82.25%	82.25%
UV	87.47%	87.50%	87.38%	87.12%	87.27%	87.01%
VCC3	86.39%	86.45%	86.44%	79.66%	86.00%	85.99%

Table 6.2: Accuracy evaluation for Neural Network on single basin cases

The performance of Valchiavenna case 3 (VCC3) in overall exceeds that of Valchiavenna case 2 (VCC2), despite the larger sample size of training points in VCC2. To explore this trend further, we analyzed the distribution of input factors in the training dataset for both VCC2 (UpperValtellina + ValTartano + ValChiavenna) and VCC3 (ValChiavenna), as well as in the testing dataset for both cases. It is noteworthy that the testing dataset remains constant for both cases. We also compared these distributions to the training dataset in VCC2, excluding the ValChiavenna part (UpperValtellina + ValTartano), which is depicted in Figure 6.3. For distribution of other factors, refer to Appendix B.

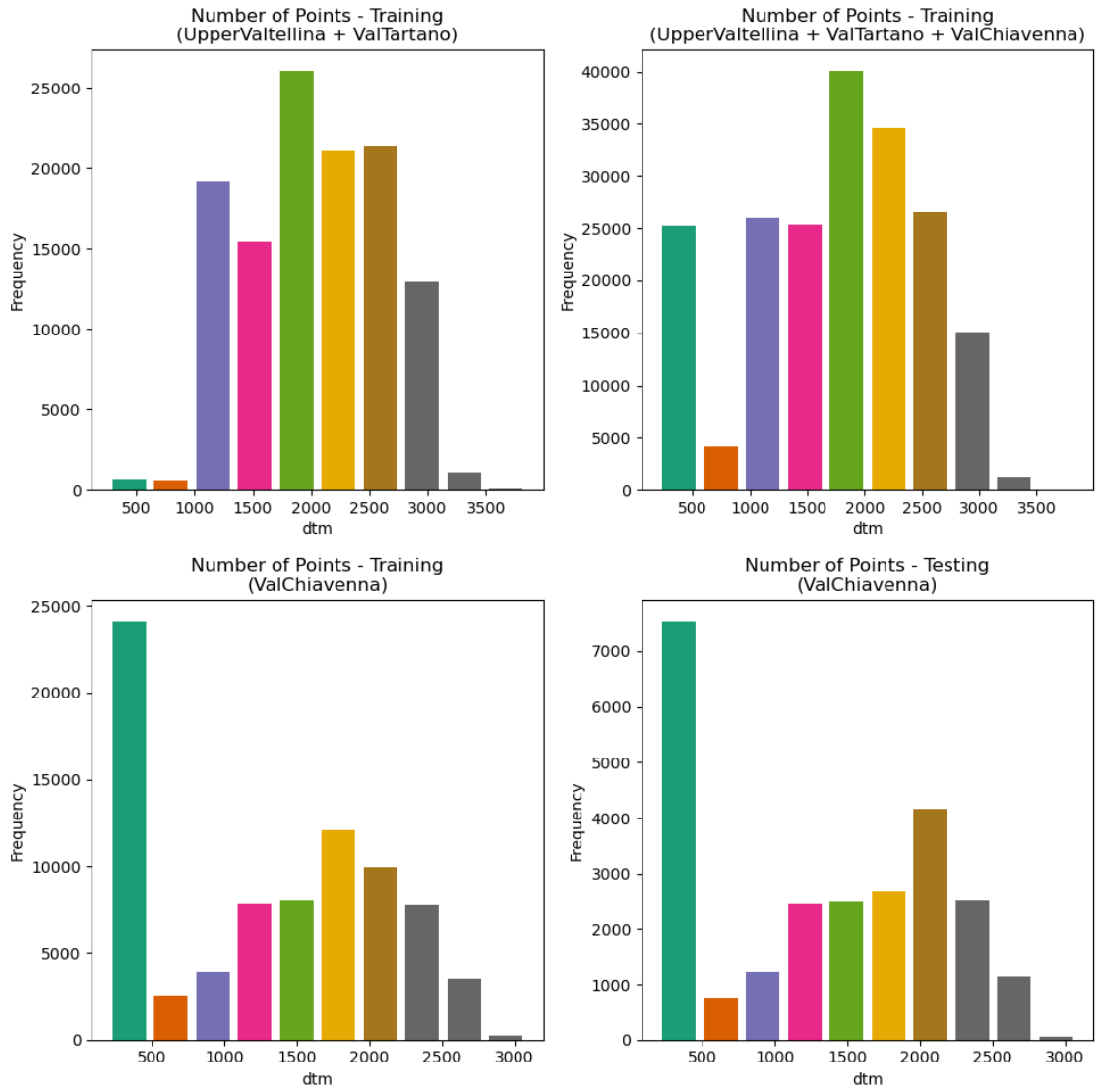


Figure 6.3: Distribution of DTM in the training dataset and testing dataset for ValChiavenna Case 1, 2, 3

The heterogeneity in the distribution of input factors observed between the training and testing datasets of VCC2, as opposed to the homogeneity observed in VCC3, indicates that the inclusion of additional samples from UpperValtellina and ValTartano introduces some noise into the training model. This finding helps explain why the data volume increase does not necessarily lead to improved accuracy.

6.1.2 Performance Evaluation of ensemble models

We intended to achieve the best performance in the unseen areas. Based on this criterion, the top three performing models (RF, AdaBoost with calibration, and NN) in the cases where sampled data from the corresponding area were not used for model training were considered in developing the ensemble ML models. Table 6.3 and Figure 6.4 display the accuracy statistics of different ensemble ML models.

Cases	Ensemble Stacking	Ensemble Blending	Ensemble Soft Voting
VT	91.14%	91.32%	88.58%
UV	91.34%	91.32%	89.39%
VCC2	94.57%	94.21%	91.76%
VCC3	95.00%	94.91%	92.86%
VCC1	67.40%	71.31%	77.18%
LC1+VT	71.88%	71.77%	79.93%
LC1+UV	78.33%	79.78%	80.02%
LC1+VT+UV	75.59%	77.97%	81.94%
LC1+VCC2	75.86%	71.69%	85.67%
LC1+VCC3	65.53%	71.83%	77.50%
LC2+VT	73.00%	73.46%	81.44%
LC2+UV	79.43%	82.55%	82.10%
LC2+VT+UV	78.46%	81.53%	83.64%
LC2+VCC2	81.69%	73.91%	92.23%
LC2+VCC3	81.48%	88.20%	92.00%

(a) Overall Accuracy

Cases	Ensemble Stacking		Ensemble Blending		Ensemble Soft Voting	
	OA	Threshold	OA	Threshold	OA	Threshold
VT	92.73%	0.1890	92.52%	0.2374	89.37%	0.4280
UV	91.44%	0.4418	91.45%	0.4624	90.41%	0.3869
VCC2	94.65%	0.5482	94.29%	0.5396	91.88%	0.4822
VCC3	95.01%	0.4996	94.92%	0.4876	93.91%	0.6190
VCC1	70.21%	0.6496	72.47%	0.5889	77.39%	0.4860
LC1+VT	72.59%	0.5117	73.96%	0.3790	88.94%	0.1935
LC1+UV	80.56%	0.3285	81.94%	0.3527	84.20%	0.4307
LC1+VT+UV	76.84%	0.3484	80.99%	0.3536	87.98%	0.3487
LC1+VCC2	83.48%	0.8515	80.70%	0.8597	86.46%	0.6030
LC1+VCC3	81.33%	0.9483	83.30%	0.9044	86.48%	0.7534
LC2+VT	73.99%	0.5117	75.72%	0.3790	92.64%	0.1741
LC2+UV	79.94%	0.3959	83.97%	0.4008	87.25%	0.4205
LC2+VT+UV	79.29%	0.4394	85.53%	0.3064	92.71%	0.2061
LC2+VCC2	90.34%	0.7842	87.56%	0.8324	92.25%	0.4898
LC2+VCC3	91.97%	0.8249	92.78%	0.7787	94.88%	0.6049

(b) Accuracy based on optimum ROC-based thresholds

Cases	Ensemble Stacking		Ensemble Blending		Ensemble Soft Voting	
	OA	Threshold	OA	Threshold	OA	Threshold
VT	92.73%	0.1850	92.52%	0.2170	89.34%	0.3917
UV	91.38%	0.3000	91.44%	0.3396	90.42%	0.3680
VCC2	94.64%	0.5137	94.29%	0.5396	91.86%	0.4398
VCC3	95.01%	0.4996	94.92%	0.4876	93.91%	0.6190
VCC1	67.31%	0.4629	70.63%	0.4044	77.39%	0.4860
LC1+VT	69.73%	0.3419	73.64%	0.3296	88.94%	0.1932
LC1+UV	79.65%	0.2273	81.69%	0.2854	84.01%	0.4175
LC1+VT+UV	76.40%	0.2863	80.18%	0.2429	87.98%	0.3487
LC1+VCC2	83.22%	0.8012	80.39%	0.8162	86.30%	0.5537
LC1+VCC3	80.94%	0.9204	83.10%	0.8775	86.36%	0.7293
LC2+VT	72.46%	0.4245	75.00%	0.3296	92.64%	0.1739
LC2+UV	79.44%	0.3478	83.64%	0.3525	86.05%	0.2647
LC2+VT+UV	78.16%	0.3270	85.35%	0.2630	92.68%	0.1926
LC2+VCC2	90.31%	0.7708	87.44%	0.8156	92.23%	0.4691
LC2+VCC3	91.96%	0.8111	92.76%	0.7627	94.87%	0.5949

(c) Accuracy based on optimum PR-based thresholds

Cases	Ensemble Stacking	Ensemble Blending	Ensemble Soft Voting
VT	0.9456	0.9472	0.9164
UV	0.9371	0.9394	0.9386
VCC2	0.9354	0.9355	0.9055
VCC3	0.9477	0.9461	0.8973
VCC1	0.6464	0.6805	0.7761
LC1+VT	0.7670	0.8520	0.9457
LC1+UV	0.8539	0.8522	0.9157
LC1+VT+UV	0.8441	0.8613	0.9421
LC1+VCC2	0.6876	0.6495	0.8176
LC1+VCC3	0.5949	0.6447	0.6938
LC2+VT	0.7873	0.8997	0.9882
LC2+UV	0.8776	0.9113	0.9701
LC2+VT+UV	0.9133	0.9462	0.9875
LC2+VCC2	0.7509	0.6699	0.9249
LC2+VCC3	0.7388	0.8230	0.8717

(d) Precision

Cases	Ensemble Stacking	Ensemble Blending	Ensemble Soft Voting
VT	0.8775	0.8795	0.8550
UV	0.8875	0.8845	0.8443
VCC2	0.9576	0.9498	0.9324
VCC3	0.9525	0.9525	0.9679
VCC1	0.7682	0.8032	0.7640
LC1+VT	0.6284	0.5269	0.6351
LC1+UV	0.6836	0.7205	0.6614
LC1+VT+UV	0.6278	0.6666	0.6806
LC1+VCC2	0.9479	0.9420	0.9181
LC1+VCC3	0.9733	0.9726	0.9843
LC2+VT	0.6302	0.5282	0.6363
LC2+UV	0.6840	0.7212	0.6624
LC2+VT+UV	0.6290	0.6686	0.6815
LC2+VCC2	0.9482	0.9427	0.9193
LC2+VCC3	0.9741	0.9734	0.9850

(e) Recall

Cases	Ensemble Stacking	Ensemble Blending	Ensemble Soft Voting
VT	0.9103	0.9121	0.8846
UV	0.9116	0.9111	0.8890
VCC2	0.9464	0.9426	0.9188
VCC3	0.9501	0.9493	0.9313
VCC1	0.7021	0.7368	0.7700
LC1+VT	0.6908	0.6511	0.7599
LC1+UV	0.7593	0.7808	0.7680
LC1+VT+UV	0.7201	0.7516	0.7903
LC1+VCC2	0.7970	0.7689	0.8650
LC1+VCC3	0.7385	0.7754	0.8139
LC2+VT	0.7000	0.6656	0.7742
LC2+UV	0.7688	0.8052	0.7872
LC2+VT+UV	0.7449	0.7835	0.8064
LC2+VCC2	0.8381	0.7832	0.9221

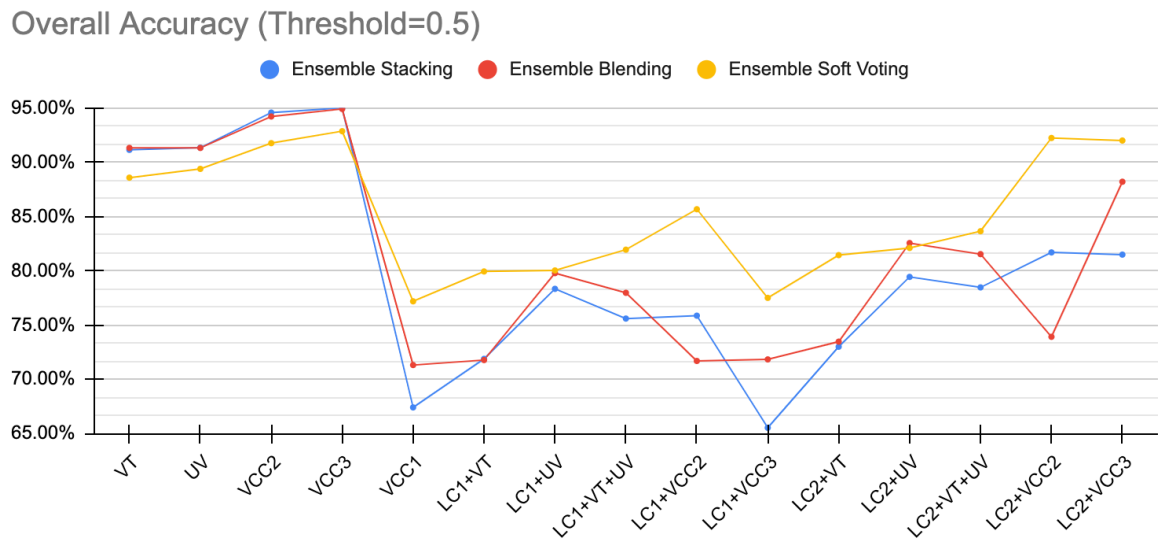
LC2+VCC3	0.8403	0.8919	0.9249

(f) F_1 score

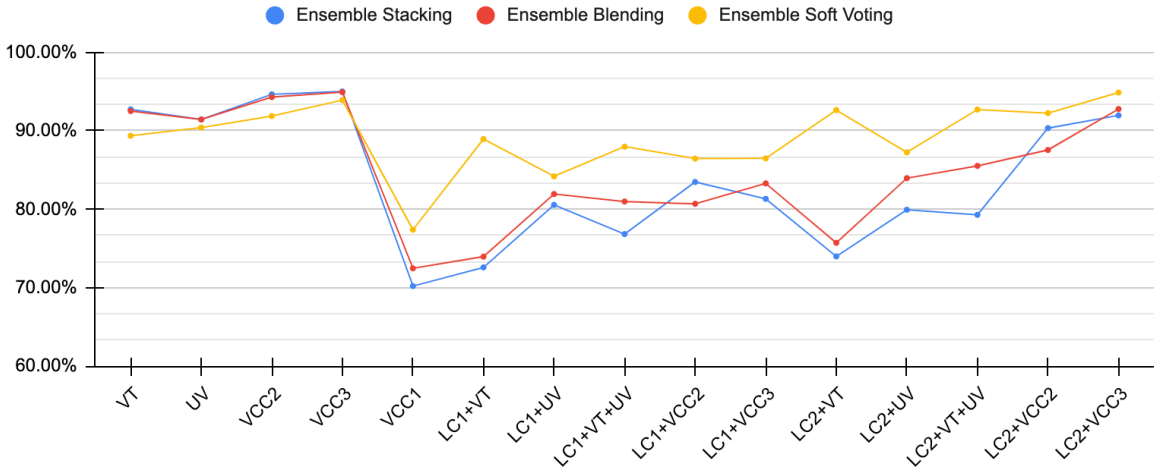
Table 6.3: Accuracy statistics of ensemble models for different cases. Table cells are colored based on each region. The darker the color, the higher the value

Disregardign the chosen threshold, stacking and blending methods produced similar performance and displayed the best accuracy (i.e., 0.91~0.93 for VT, 0.92 for UV, and 0.94~0.95 for both VCC2 and VCC3) compared to SF (i.e., 0.88~0.89 for VT, 0.89~0.9 for UV, and 0.91~0.94 for both VCC2 and VCC3) when the data sampled in a region was used to train the model (i.e., VT, UV, VCC2, and VCC3). When the data sampled in an area was not used to train the model (i.e., VCC1, LC1, and LC2), the SF model outperformed other models. Blending performed somewhat well in terms of accuracy (i.e., 0.71~0.72 for VCC1, and 0.72~0.93 for both LC1 and LC2) in this case and was marginally superior to stacking (i.e., 0.67~0.7 for VCC1, and 0.66~0.92 for both LC1 and LC2).

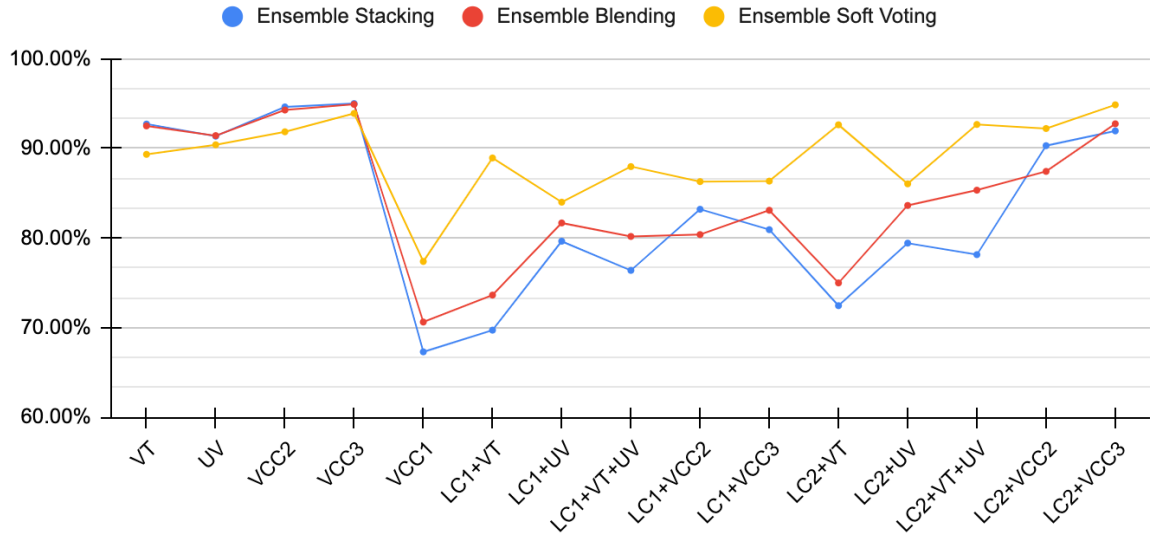
Both the staking and blending models offered comparable accuracy for Tartano and Upper Vatellina and outperformed the SF model. With a maximum accuracy of 0.95 when utilizing the optimum threshold, stacking model training on the dataset only sampled in ValChiavenna (VCC3) demonstrated the best performance for ValChiavenna. With a maximum accuracy improvement of 0.12, the SV model trained in VCC3 outperformed the stacking model trained in VCC3 in Lombardy instances and provided the best performance there.



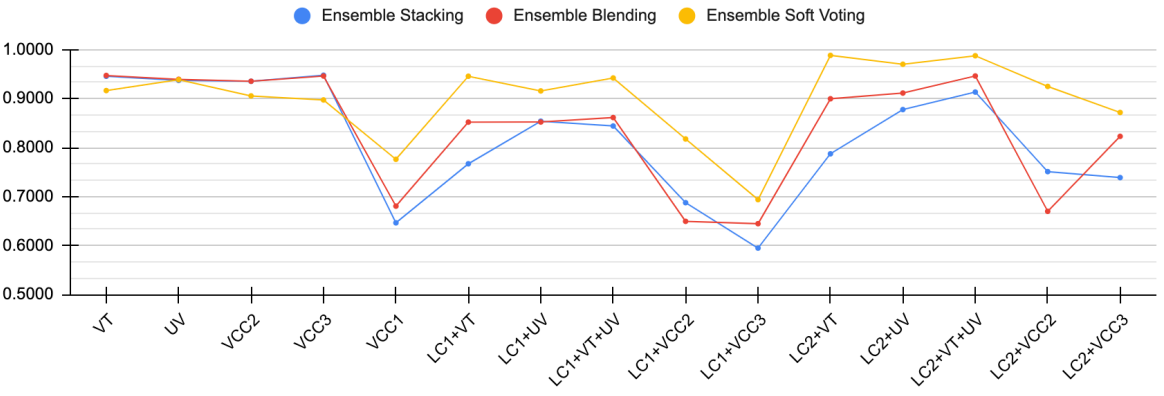
OA based on Optimum ROC Threshold



OA based on Optimum PRC Threshold



Precision



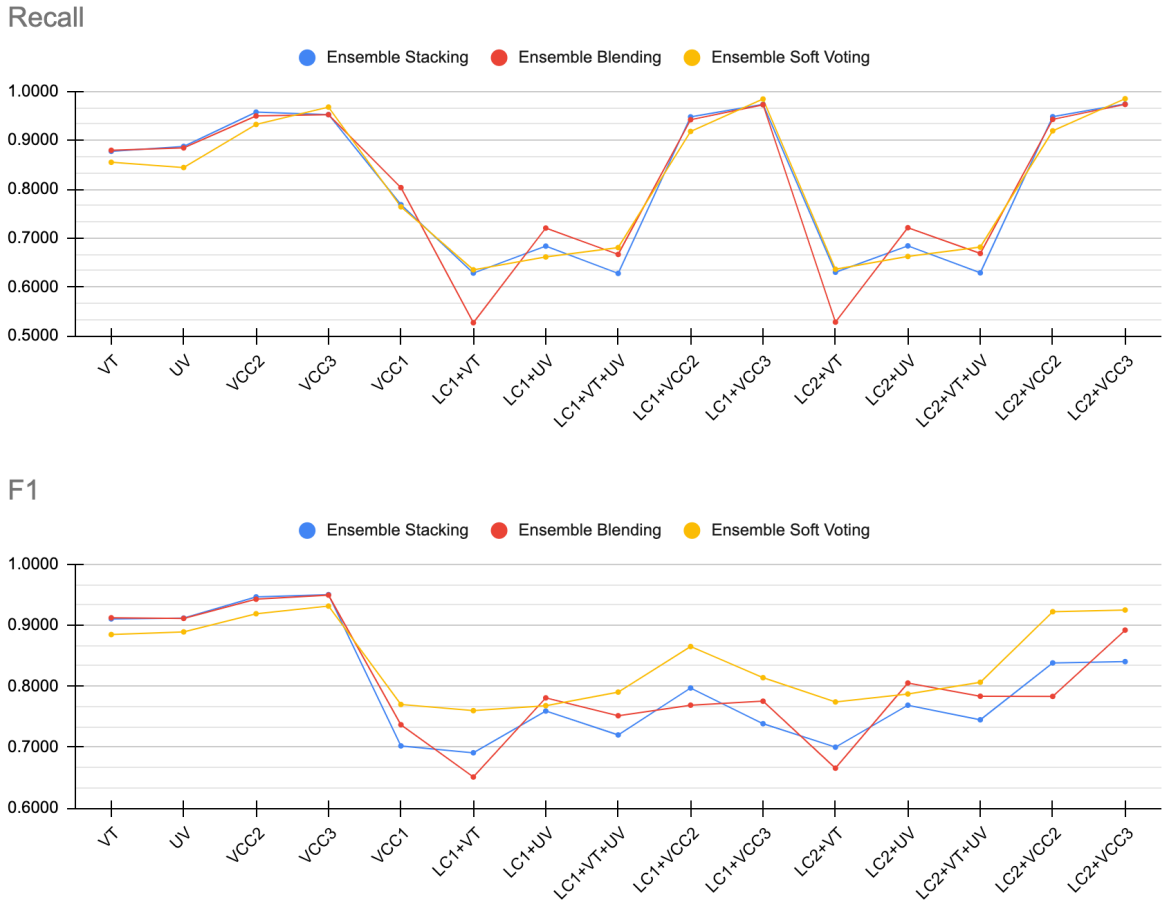


Figure 6.4: Accuracy statistics of ensemble models for different cases

The Neural Network outperformed all other models when all evaluation outcomes on Lombardy cases were compared, with the highest accuracy of 0.97. Since the goal of this research was to produce LSMs, it was essential to take into account whether the corresponding model gave a good estimate of the likelihood that a landslide would occur when deciding on the final model for Lombardy. The calibration plots for the various neural network models evaluated in Lombardy are depicted in Figure 6.5. This figure shows that only the neural network model learned in VCC2 displays well calibration behavior. The Neural Network model trained in VCC2 was therefore chosen as the final model to generate LSM for Lombardy.

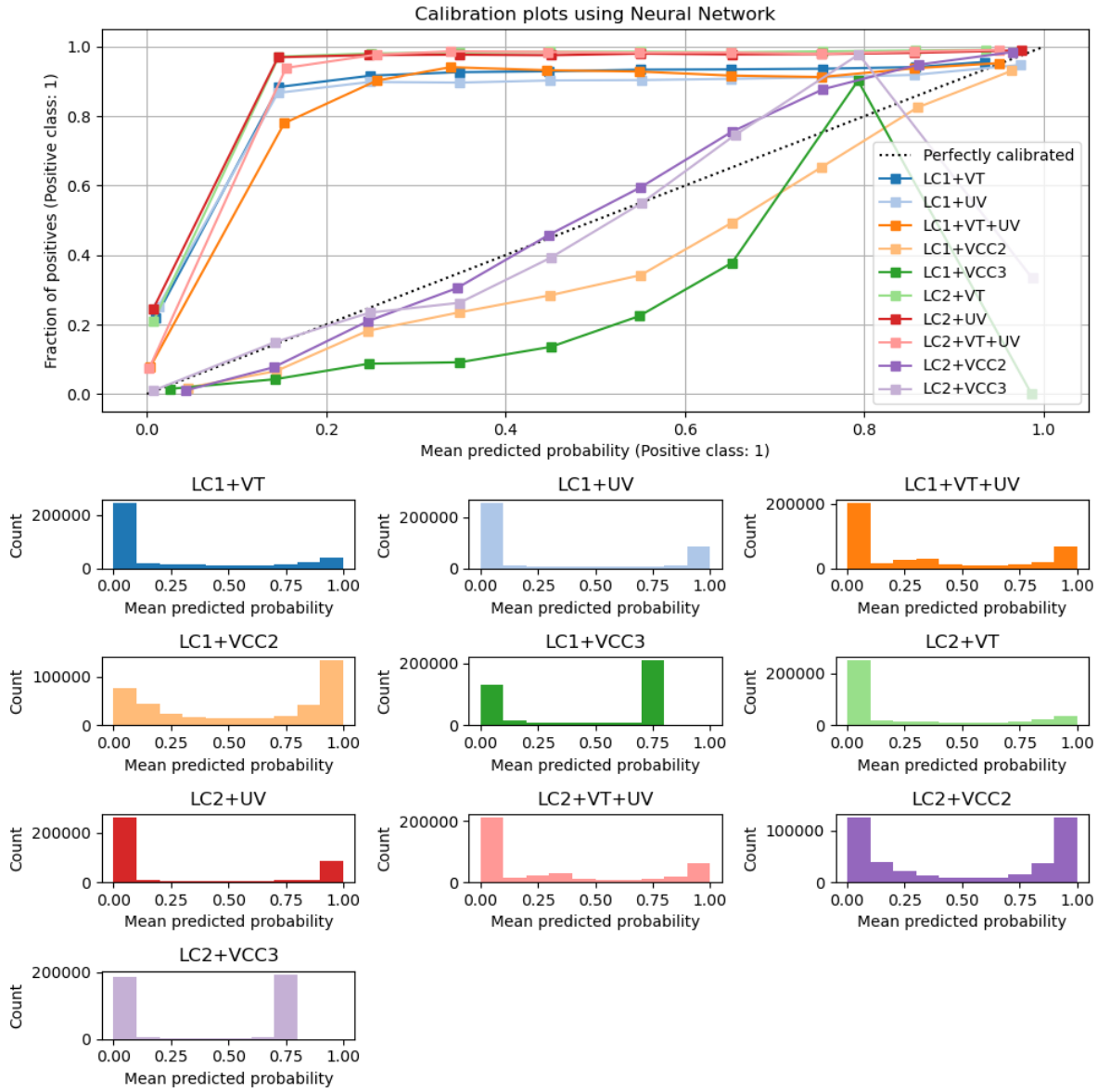


Figure 6.5: Calibration plots for Neural Network models applied on Lombardy case.

6.2 Result of introducing precipitation as an environmental factor

Precipitation was added to the Neural Network as a new factor in VCC2. In this situation, three additional NN models were trained individually utilizing average precipitation, the 90th percentile of precipitation, and both average and 90th percentile precipitation. The performance of the constructed models was then assessed by applying them to the Lombardy cases (i.e., LC1 and LC2). Table 6.4 showed the accuracy estimation based on several threshold types with or without input from precipitation.

Cases	Neural Network	Neural Network with	Neural Network with	Neural Network with

	without precipitation			average precipitation			90th percentile precipitation			both average and 90th percentile precipitation		
	Base OA	Optimal ROC OA	Optimal PRC OA	Base OA	Optimal ROC OA	Optimal PRC OA	Base OA	Optimal ROC OA	Optimal PRC OA	Base OA	Optimal ROC OA	Optimal PRC OA
VCC2	89.01%	89.05 %	88.96 %	87.72 %	88.46 %	88.46 %	87.92 %	88.54 %	88.37 %	87.96 %	88.74 %	88.72 %
LC1+VCC2	85.83%	87.11 %	87.03 %	87.96 %	90.16 %	90.16 %	87.57 %	89.54 %	89.53 %	88.03 %	88.86 %	88.85 %
LC2+VCC2	93.10%	93.13 %	93.10 %	91.08 %	95.93 %	95.93 %	91.04 %	95.29 %	95.29 %	91.80 %	95.77 %	95.77 %

Table 6.4: Accuracy statistic of Neural Network with or without precipitation as input based on threshold 0.5 and optimum thresholds based on ROC and PRC.

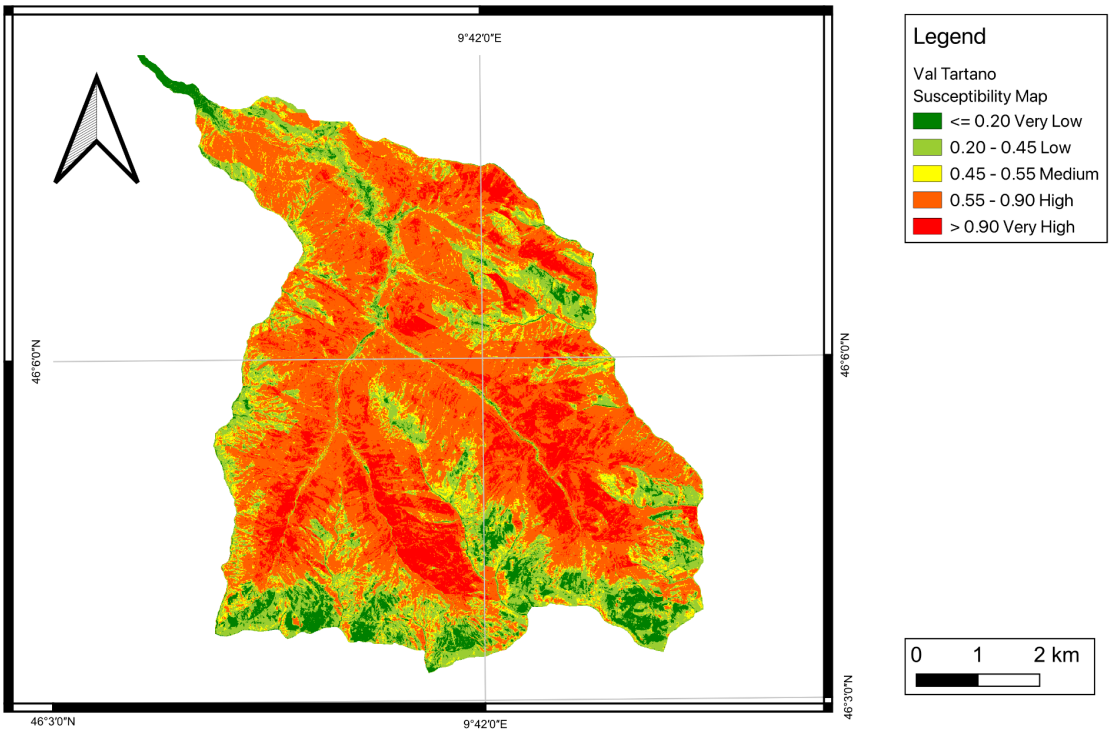
When the built model was applied to the area from which the sampled points were also used for training (i.e., VCC2), there was no comparable accuracy difference with or without precipitation as one factor; however, the model trained with precipitation shows an improvement in accuracy of about 2–3% in the Lombardy case (i.e., LC1 and LC2), and the maximum accuracy improvement was found in the model trained only with average precipitation.

6.3 Landslide susceptibility maps

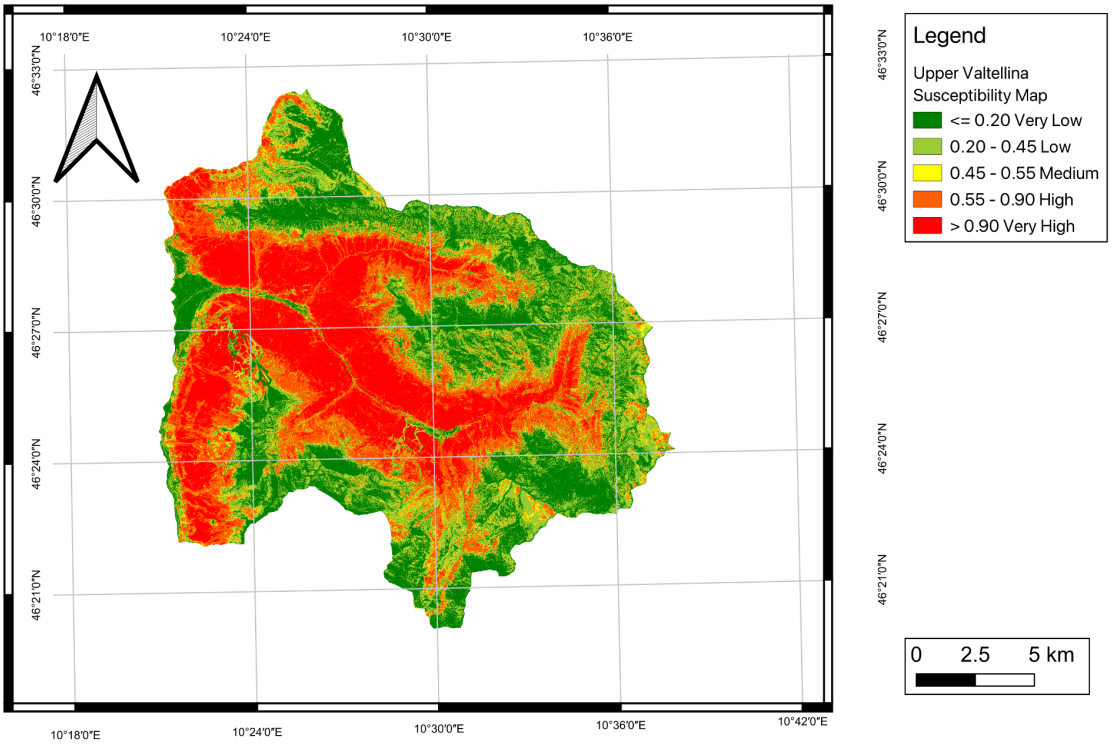
Two different schemas were used to categorize the probability values of ML models into different classes in QGIS:

- Four classes: low (<=0.25), medium (0.25-0.5), high (0.5-0.75), very high (>0.75);
- Five classes: very low (<=0.2), low (0.2-0.45), medium (0.45-0.55), high (0.55-0.9), very high (>0.9).

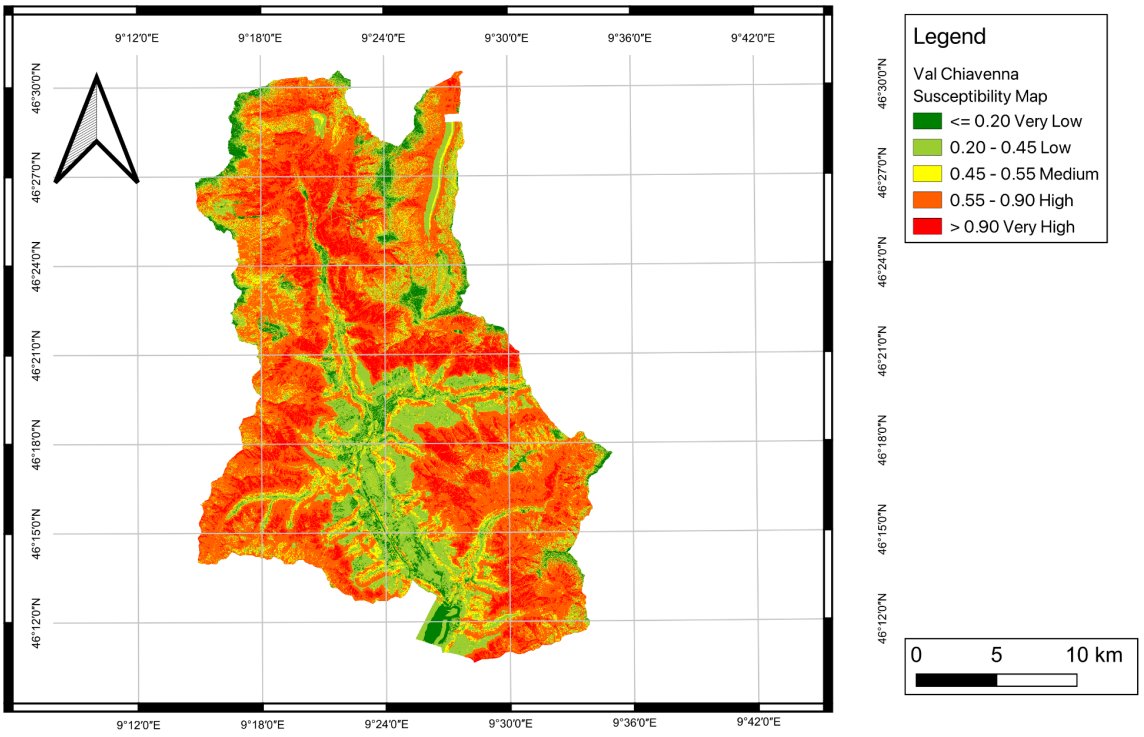
Figure 6.6 displays the landslide susceptibility maps for each basin based on the top-performing ML models (i.e., Random Forests) for each basin scenario. Figure 6.6(c) also displays the outcomes of first-time application of a model that was trained on one region to another (i.e., VCC1). Although the model was not built using any data from this area, the accuracy for this situation is close to 0.75, demonstrating good generalizability.



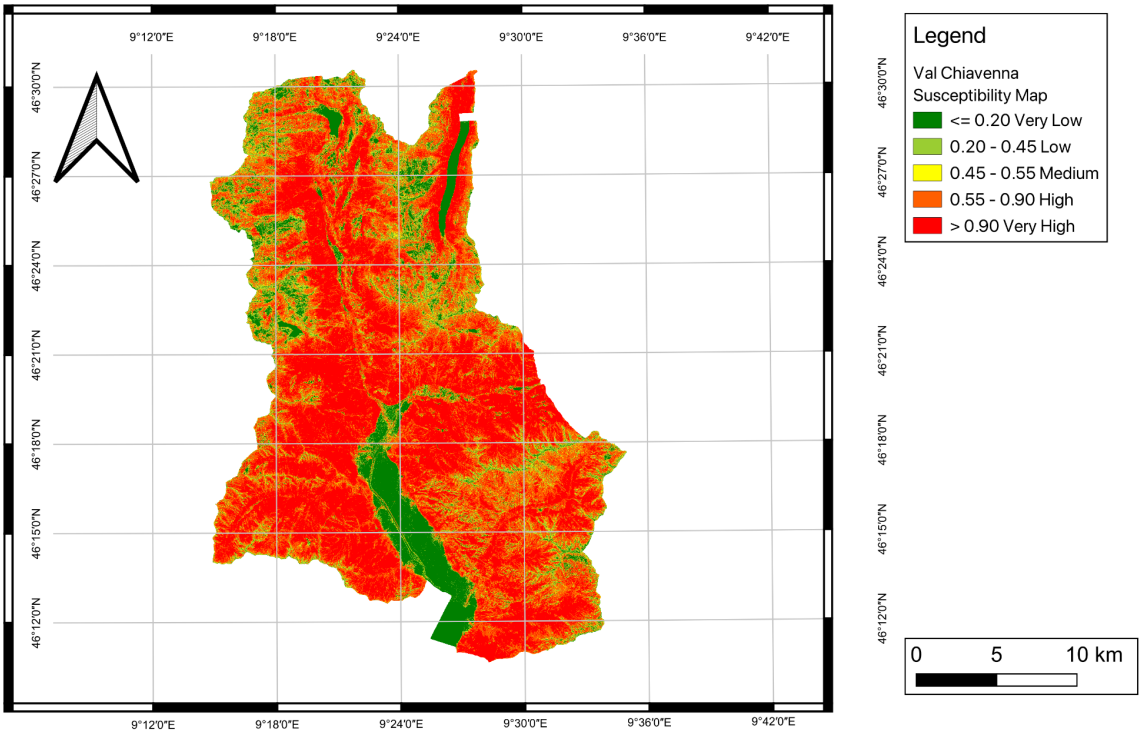
a. LSM for Val Tartano derived from Random Forests model trained in VT without precipitation data



b. LSM for Upper Valtellina derived from Random Forests model trained in UV without precipitation data



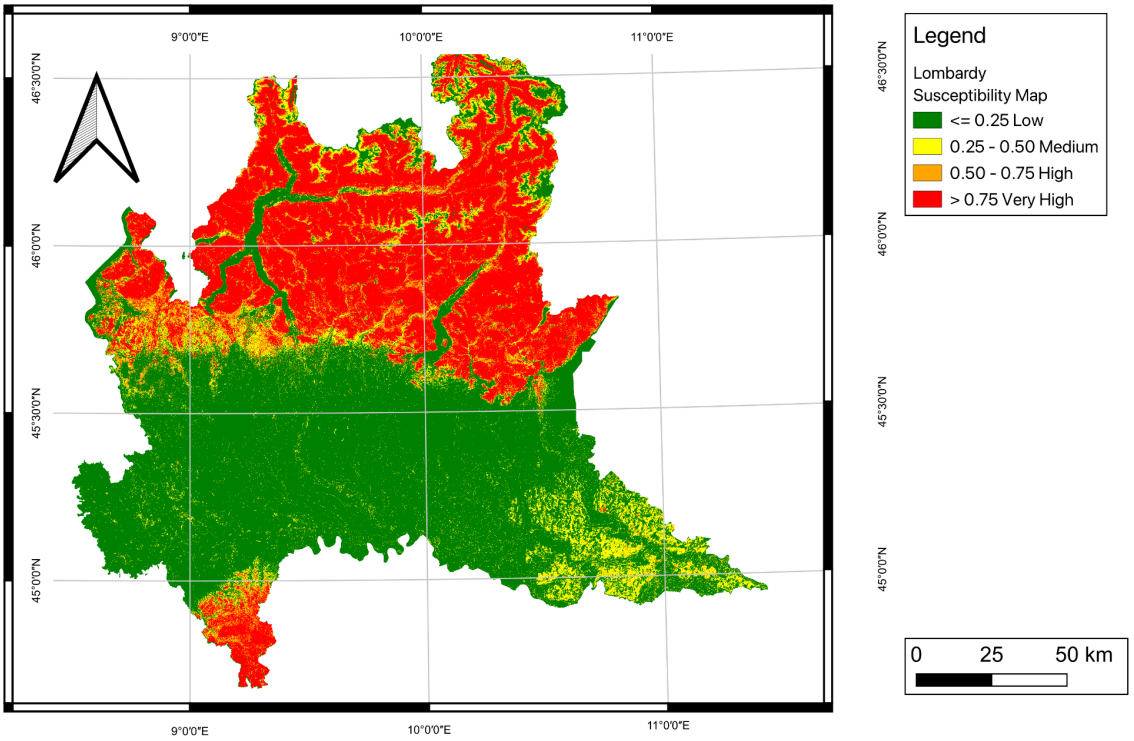
c. LSM for Val Chiavenna derived from Random Forests model trained in VCC1 without precipitation data



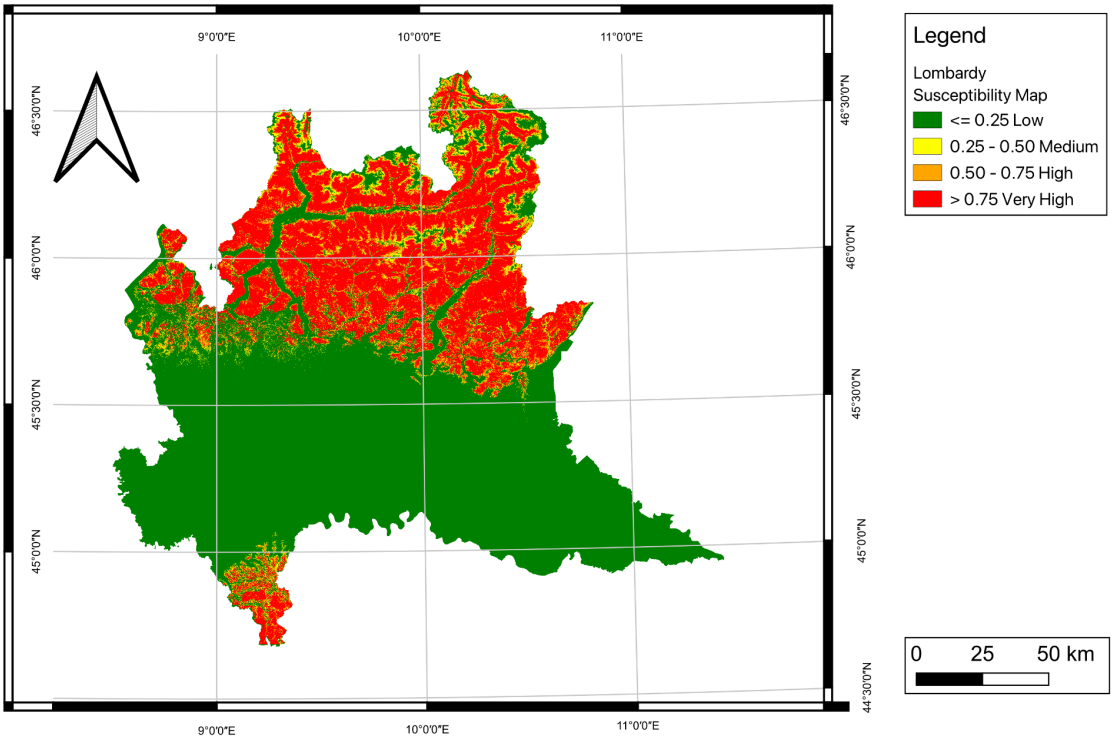
d. LSM for Val Chiavenna derived from Random Forests model trained in VCC3 without precipitation data

Figure 6.6: LSMs for three basins

The best-performing ML models (i.e., Neural Network) developed in VCC2 with average precipitation data and without precipitation data were used to map the landslide susceptibility of Lombardy (Figure 6.7 - 6.8).

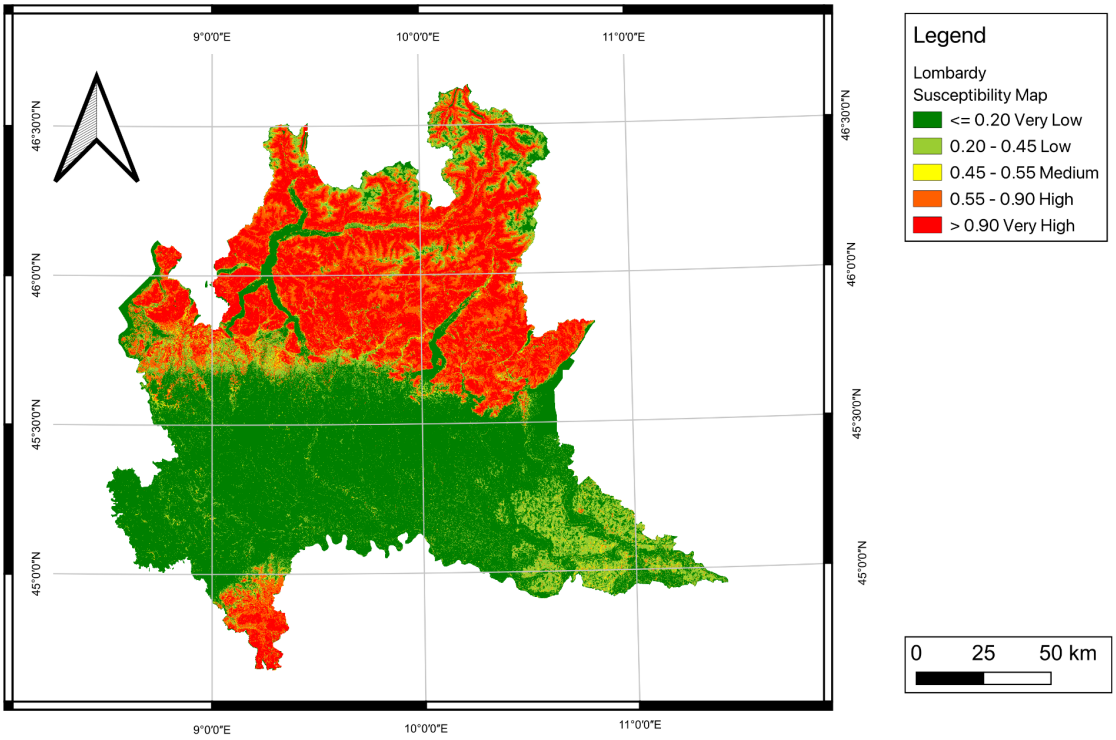


a. LSM derived from Neural Network model trained in VCC2 without precipitation data

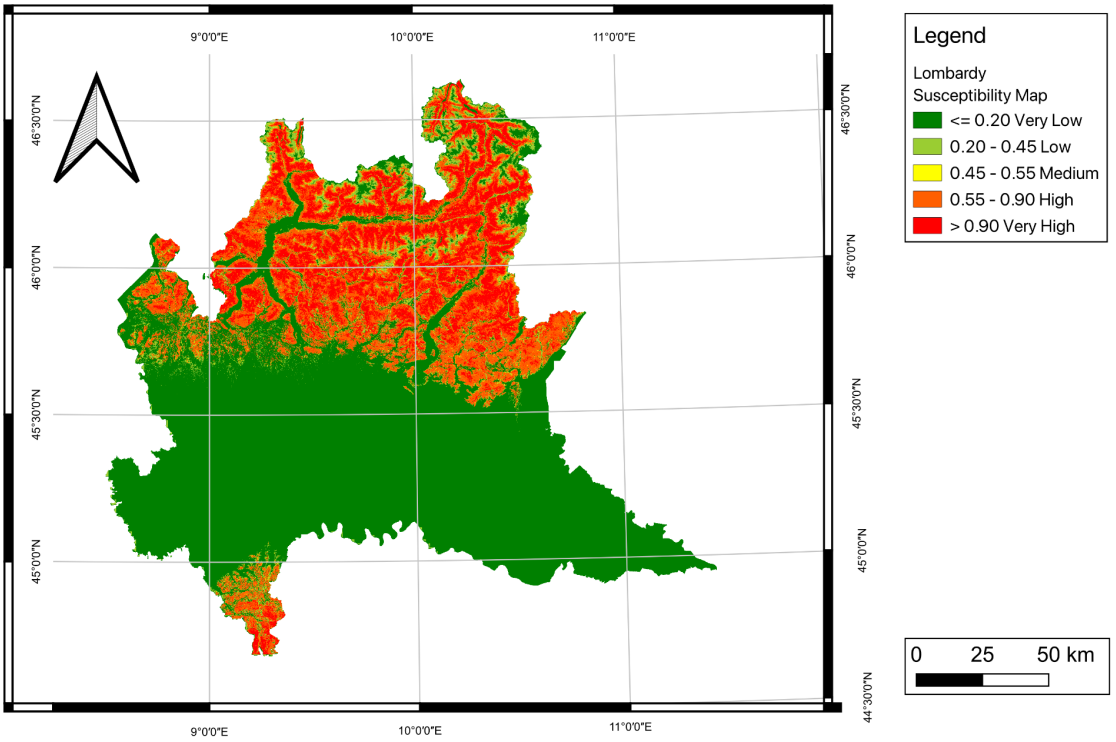


b. LSM derived from Neural Network model trained in VCC2 with average precipitation data

Figure 6.7: Landslide susceptibility maps of Lombardy using 4-classes schema



a. LSM derived from Neural Network model trained in VCC2 without precipitation data

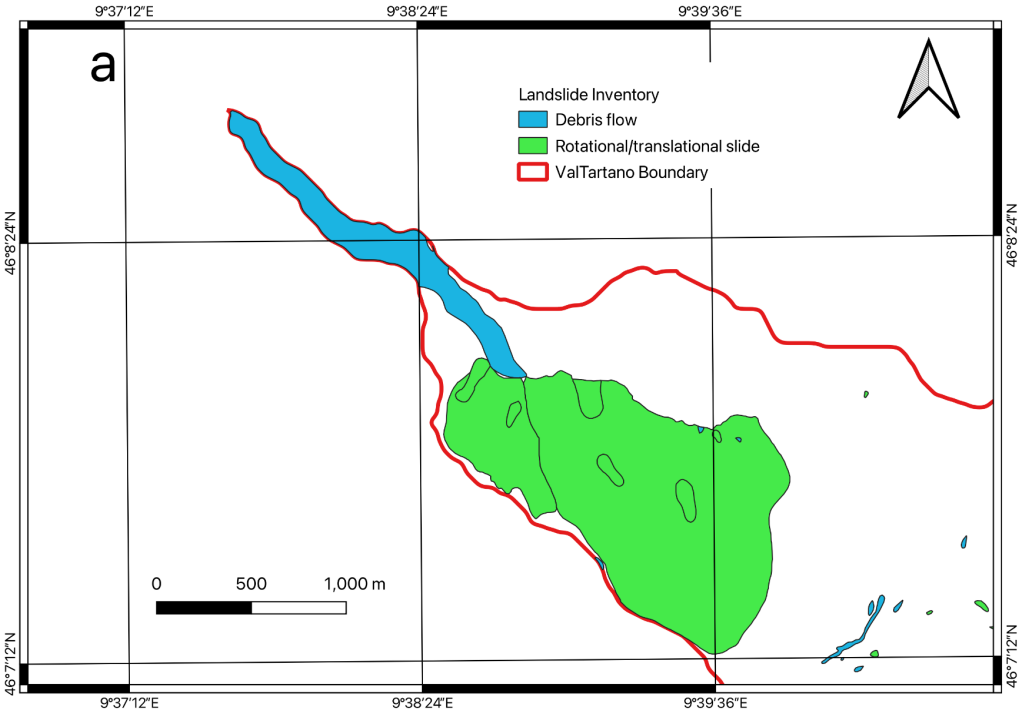


b. LSM derived from Neural Network model trained in VCC2 with average precipitation data

Figure 6.8: Landslide susceptibility maps of Lombardy using 5-classes schema

The resulting LSMs were consistent with the fact that all considered landslides documented in the landslide inventory occurred north of Lombardy. A high and very high susceptibility is also present in the south part of Lombardy (the southern part of the Po River), despite the fact that there are no records of landslides there. The central foothills of Lombardy are in regions of extremely low or low susceptibility when the precipitation data was not taken into account in the model. Using the same model, some land parcels in southeast Lombardy are classed as medium-scale under the 4-classes schema as opposed to the 5-classes schema, which classifies the region as low susceptibility. Both the central foothills and the southeast region of Lombardy were classed as very low susceptibility using the 5-classes schema or low susceptibility under the 4-classes schema and showed no variation when the average precipitation was included in the model.

In order to evaluate the quality of the LSM, the Pruna landslide, the largest landslide in Val Tartano, is used. Regardless of the model and schema employed, the region of this landslide is accurately classified as high and very high risk (Figure 6.9). More area was categorized as having a high susceptibility when precipitation data was incorporated, particularly when a 5-class schema was used.



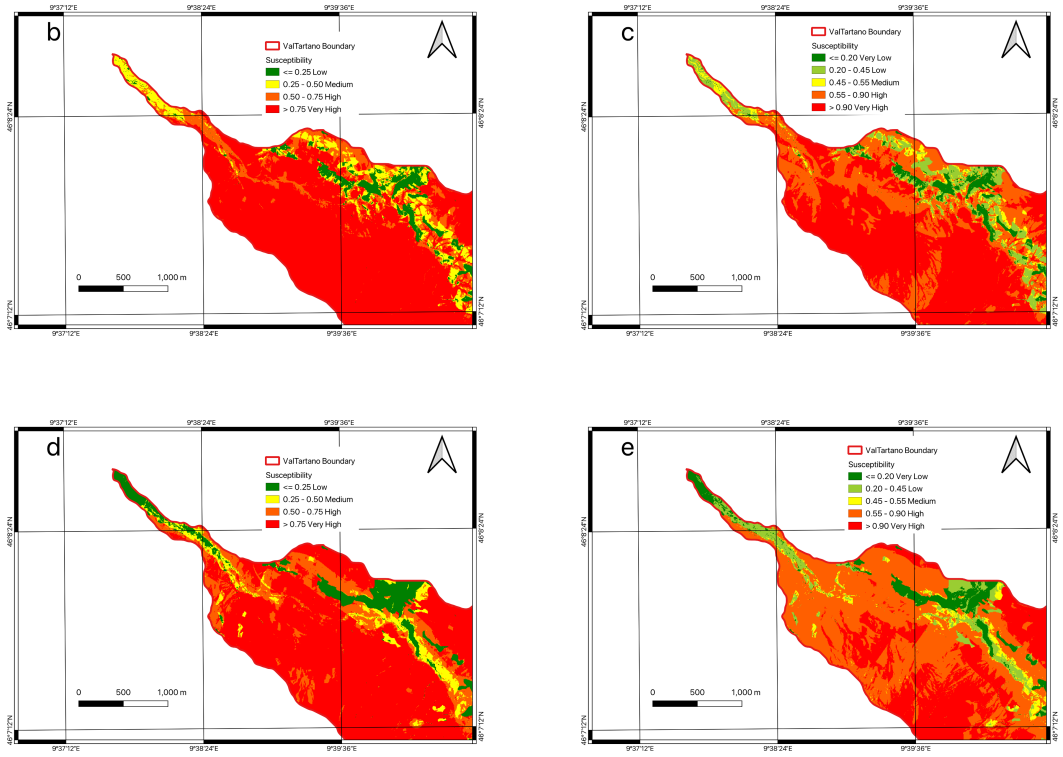


Figure 6.9: Pruna landslides: a is representing the Pruna landslides and the debris flow; b the LSM derived from the VCC2+NN models without precipitation using a 4-classes schema; c the LSM derived from same model as b but using a 5-classes schema; d the LSM derived from the VCC2+NN models with average precipitation using a 4-classes schema; e the LSM derived from same model as d but using a 5-classes schema.

The spatial statistics of landslide susceptibility and the corresponding precentage of documented landslide in each susceptibilitiy class derived from these models are presented in Table 6.5. The spatial statistics of susceptibility maps indicated that when precipitation data was excluded for model construction, 21.21% (i.e., 5070.66 km^2) and 15.46% (i.e., 3696.34 km^2) of the total area fell within very high and high landslide susceptibility categories, respectively when adopting the 5-classes schema. Under this schema, very low, low, and medium susceptible zones make up 48.93%, 11.93%, and 2.48%, respectively. When the 4-class schema was used, Lombardy had a susceptibility to landslides of 52.65%, which was 13.19% lower than the 4-class schema when taking into account the extremely low and low categories jointly. Only a little section of the Lombary region was categorized as medium, with 9.47% of the overall area utilizing a 4-classes schema that was larger than a 5-classes schema and about 37.88% in regions with high and very high susceptibility that was also somewhat larger than 5-classes schema.

When average precipitation data was taken into account, a larger region was categorized as low (i.e., 63.78% in 4-classes schema) and very low (i.e., 62.18% in 5-classes schema) susceptible area. The high and very high susceptible areas, on the other hand, were reduced (i.e., 5.4% of high susceptible area and 26.41% of very high susceptible area in the 4-classes schema, and 12.93% of very high susceptible area in the 5-classes schema).

Susceptibility Class	LSM derived using VCC2+NN without precipitation			LSM derived using VCC2+NN with average precipitation		
	Area [km^2]	Area Percentage [%]	Landslide Percentage [%]	Area [km^2]	Area Percentage [%]	Landslide Percentage [%]
Low	12589.93	52.65%	6.03%	15250.58	63.78%	11.43%
Medium	2263.52	9.47%	8.04%	1055.55	4.41%	10.25%
High	1593.00	6.66%	11.71%	1290.97	5.40%	13.21%
Very High	7464.30	31.22%	74.22%	6313.64	26.41%	65.10%

a. using 4-classes schema

Susceptibility Class	LSM derived using VCC2+NN without precipitation			LSM derived using VCC2+NN with average precipitation		
	Area [km^2]	Area Percentage [%]	Landslide Percentage [%]	Area [km^2]	Area Percentage [%]	Landslide Percentage [%]
Very Low	11698.81	48.93%	3.79%	14867.34	62.18%	8.31%
Low	2852.74	11.93%	8.74%	1253.03	5.24%	11.42%
Medium	592.22	2.48%	3.16%	378.88	1.58%	3.92%
High	3696.34	15.46%	35.04%	4320.16	18.07%	45.09%
Very High	5070.66	21.21%	49.27%	3091.34	12.93%	31.26%

b. using 5-classes schema

Table 6.5: Distribution of susceptibility classes and landslides for Lombardy

Almost 94% of the historical landslide points in the 4-classes schema and 87% in the 5-classes schema fall into the range of medium to extremely high landslide susceptibility derived from the model built without using precipitation data. The percentage dropped to about 89% in the 4-class schema and 80% in the 5-class schema when average precipitation data were taken into account.

7. Conclusions

Identification of landslide-prone areas can be valuable for land use planners or disaster management agencies in allocating resources to forecast and mitigate landslide impacts. The presented work covers a wide range of susceptibility modeling scenarios using five base classification methods and three ensemble techniques on three basins and extends to the whole Lombardy region. 11 factors (e.g., elevation, eastness, northness, plan curvature, profile curvature, TWI, NDVI, Distance from roads, distance from rivers, distance from faults, and land use) were firstly used to train and determine the final model and then the precipitation data was incorporated into the final model.

Using several validation indices, such as overall accuracy, precision, recall, ROC, and PRC, the performance of each individual ML model was assessed in a variety of scenarios, including VT, UV, VCC1, VCC2, VCC3, LC1, and LC2. The models with the best performance in the cases where sampled data from the corresponding area were not used for model training (RF, Calibrated AdaBoost, and NN) were chosen as the base classifiers for ensemble models. When compared to their base classifiers, the ensemble models' statistical performance did not significantly improve. Although the Neural Network model had lower performance when training using points from the relevant regions, it outperformed all other generated models in terms of generalization ability.

When precipitation data was not taken into account, the susceptibility map created using Neural Network constucted in VCC2 for the Lombardy region indicated that almost 37% of the whole study area fell between the "very high" and "high" landslide susceptibility categories. When the average precipitation data were added, this percentage dropped to 31%.

Although only the overall accuracies at various thresholds were taken into account when choosing base classifiers for ensemble models, future research can look into other methodologies or indices to determine the best combination of ML models to produce more robust and accurate ensemble ML models.

In terms of model validation, we evaluated the performance of ML models using a split of training and testing data. Future research can, however, consider an intensive iteration-based cross-validation strategy to evaluate the robustness of ML models. This might reveal further details about the models' robustness and potential for generalization.

Acknowledgement

I would like to express my gratitude to all those who helped me during the writing of this thesis.

First of all, I would like to thank my supervisors Prof. Maria Antonia Brovelli and Dr. Vasil Yordanov for their constant support and guidance. They were very patient with me as I worked through each phase of drafting this paper. They showed me the proper way to conduct research with a serious mindset. Throughout the dissertation preparation process, they consistently offered me inspiring guidance and assisted me in identifying and resolving issues so that I could make my dissertation better. Without them, I would not have been able to finish this thesis.

I also want to express my gratitude to my professors who taught me during my academic career. I have gained a lot from each of you, especially from the rigor and attention to detail with which you teach. After graduation, I'll miss you all a lot.

Finally, I want to thank my loving family. Without their careful attention and support, I would not have been able to complete my studies without concern. In particular, I want to thank my dear sisters, without whom I would not be able to study and live in Italy. They took on the difficult responsibility of caring for my parents, and they showed me a lot of kindness and helpful advice when I was lost because of the pandemic and because I was in a foreign country.

Bibliography

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, 3029. https://doi.org/10.3390/rs14133029

Amici, L., 2021. Landslide susceptibility mapping and displacement monitoring: a case study in Northern Italy [WWW Document]. URL http://hdl.handle.net/10589/183109 (accessed 3.10.23).

Antonielli, B., Mazzanti, P., Rocca, A., Bozzano, F., Dei Cas, L., 2019. A-DInSAR Performance for Updating Landslide Inventory in Mountain Areas: An Example from Lombardy Region (Italy). Geosciences 9, 364. https://doi.org/10.3390/geosciences9090364

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65, 15–31. https://doi.org/10.1016/j.geomorph.2004.06.010

Azarafza, Mohammad, Azarafza, Mehdi, Akgün, H., Atkinson, P.M., Derakhshani, R., 2021. Deep learning-based landslide susceptibility mapping. Sci Rep 11, 24112. https://doi.org/10.1038/s41598-021-03585-1

Bajni, G., Camera, C.A.S., Brenning, A., Apuani, T., 2021. Rock mass geomechanical properties to improve rockfall susceptibility assessment: a case study in Valchiavenna (SO). IOP Conf. Ser.: Earth Environ. Sci. 833, 012180. https://doi.org/10.1088/1755-1315/833/1/012180

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 30, 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman, L., 1996. Bagging predictors. Mach Learn 24, 123–140. https://doi.org/10.1007/BF00058655

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Budimir, M.E.A., Atkinson, P.M., Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. Landslides 12, 419–436. https://doi.org/10.1007/s10346-014-0550-5

Colombera, L., Bersezio, R., 2011. Impact of the magnitude and frequency of debris-flow events on the evolution of an alpine alluvial fan during the last two centuries: responses to natural and anthropogenic controls. Earth Surface Processes and Landforms 36, 1632–1646. https://doi.org/10.1002/esp.2178

Cruden, D.M., 1991. A simple definition of a landslide. Bulletin of the International Association of Engineering Geology - Bulletin de l'Association Internationale de Géologie de l'Ingénieur 43, 27–29. https://doi.org/10.1007/BF02590167

Cruden, D., Varnes, D., 1996. Landslide types and processes. Transportation research board, US National Research Council, Special Report 247, 36–75.

EVANS, S., Bovis, M. J., Hutchinson, J. N., 2001. Landslides of the Flow Type. Environmental and Engineering Geoscience, 7(3), 221-238.

Fang, Z., Wang, Y., Peng, L., Hong, H., 2021. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. International Journal of Geographical Information Science 35, 321–347. https://doi.org/10.1080/13658816.2020.1808897

Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences 55, 119–139. https://doi.org/10.1006/jcss.1997.1504

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). The Annals of Statistics 28, 337–407. https://doi.org/10.1214/aos/1016218223

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 29, 1189–1232.

Habumugisha, J.M., Chen, N., Rahman, M., Islam, M.M., Ahmad, H., Elbeltagi, A., Sharma, G., Liza, S.N., Dewan, A., 2022. Landslide Susceptibility Mapping with Deep Learning Algorithms. Sustainability 14, 1734. https://doi.org/10.3390/su14031734

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

ISPRA, 2014. IdroGEO - Open Data [WWW Document]. URL https://idrogeo.isprambiente.it/app/page/open-data

Istat (Italian National Institute of Statistics), 2020. Territory and population. - Regional [WWW Document]. URL https://www.asr-lombardia.it/asrlomb/en/opendata/Territory_and_population___Regional.

Kavzoglu, T., Teke, A., Yilmaz, E.O., 2021. Shared Blocks-Based Ensemble Deep Learning for Shallow Landslide Susceptibility Mapping. Remote Sensing 13, 4776. https://doi.org/10.3390/rs13234776

King, G., Zeng, L., 2001. Logistic Regression in Rare Events Data. Political Analysis 9, 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Kuhlman, D., 2012. A Python Book: Beginning Python, Advanced Python, and Python Exercises [WWW Document]. URL https://web.archive.org/web/20120623165941/http://cutter.rexx.com/~dkuhlman/python_book_01.html#important-features-of-python (accessed 3.9.23).

Kumar, C., Walton, G., Santi, P., Luza, C., 2023. An Ensemble Approach of Feature Selection and Machine Learning Models for Regional Landslide Susceptibility Mapping in the Arid Mountainous Terrain of Southern Peru. Remote Sensing 15, 1376. https://doi.org/10.3390/rs15051376

Li, W., Fang, Z., Wang, Y., 2022. Stacking ensemble of deep learning methods for landslide susceptibility mapping in the Three Gorges Reservoir area, China. Stoch Environ Res Risk Assess 36, 2207–2228. https://doi.org/10.1007/s00477-021-02032-x

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, 1411–1424. https://doi.org/10.5194/nhess-17-1411-2017

Longoni, L., Papini, M., Brambilla, D., Barazzetti, L., Roncoroni, F., Scaioni, M., Ivanov, V.I., 2016. Monitoring Riverbank Erosion in Mountain Catchments Using Terrestrial Laser Scanning. Remote Sensing 8, 241. https://doi.org/10.3390/rs8030241

Niculescu-Mizil, A., Caruana, R., 2005. Predicting good probabilities with supervised learning, in: Proceedings of the 22nd International Conference on Machine Learning, ICML '05. Association for Computing Machinery, New York, NY, USA, pp. 625–632. https://doi.org/10.1145/1102351.1102430

QGIS, 2023. QGIS - A Free and Open Source Geographic Information System. https://www.qgis.org/en/site/

Park, S., Choi, C., Kim, B., Kim, J., 2013. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environ Earth Sci 68, 1443–1464. https://doi.org/10.1007/s12665-012-1842-5

Pourghasemi, H.R., Rahmati, O., 2018. Prediction of the landslide susceptibility: Which algorithm, which precision? CATENA 162, 177–192. https://doi.org/10.1016/j.catena.2017.11.022

Python, 2023. Python Homepage. https://www.python.org/about/

Regione Lombardia, 2020. Lombardy. In the north of Italy, in the heart of Europe. https://www.en.regione.lombardia.it/wps/portal/site/en-regione-lombardia/discover-lombardy/territory-and-population

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/10.1016/j.earscirev.2018.03.001

Scikit-learn, 2023. Scikit-learn Homepage. https://scikit-learn.org/stable/

Sørensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrology and Earth System Sciences 10, 101–112. https://doi.org/10.5194/hess-10-101-2006

Tantardini, D., Stevenazzi, S., Apuani, T., 2022. The Last Glaciation in Valchiavenna (Italian Alps): maximum ice elevation data and recessional glacial deposits and

landforms. Italian Journal of Geosciences 141, 259–277. https://doi.org/10.3301/IJG.2022.13

Thi Ngo, P.T., Panahi, M., Khosravi, K., Ghorbanzadeh, O., Kariminejad, N., Cerda, A., Lee, S., 2021. Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. Geoscience Frontiers 12, 505–519. https://doi.org/10.1016/j.gsf.2020.06.013

Töscher, A., Jahrer, M., 2009. The BigChaos Solution to the Netflix Grand Prize. Netflix prize documentation, 1-52.

Turner, A.K., 2018. Social and environmental impacts of landslides. Innov. Infrastruct. Solut. 3, 70. https://doi.org/10.1007/s41062-018-0175-y

Wang, Y., Fang, Z., Hong, H., 2019. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of The Total Environment 666, 975–993. https://doi.org/10.1016/j.scitotenv.2019.02.263

Wang, Y., Feng, L., Li, S., Ren, F., Du, Q., 2020. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. CATENA 188, 104425. https://doi.org/10.1016/j.catena.2019.104425

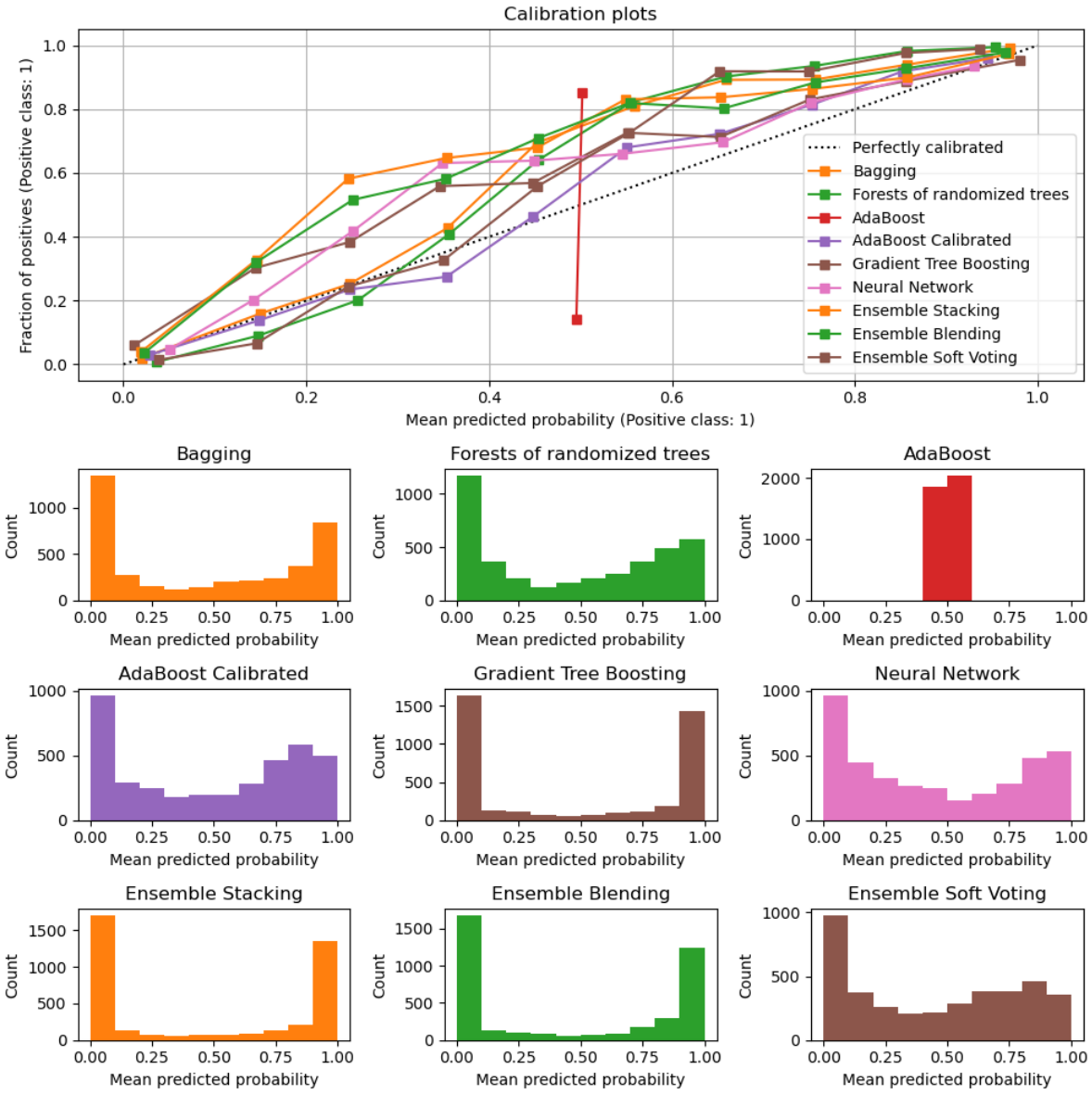
Wolpert, D.H., 1992. Stacked generalization. Neural Networks 5, 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Yordanov, V., Brovelli, M.A., 2021. Application of various strategies and methodologies for landslide susceptibility maps on a basin scale: the case study of Val Tartano, Italy. Appl Geomat 13, 287–309. https://doi.org/10.1007/s12518-020-00344-1

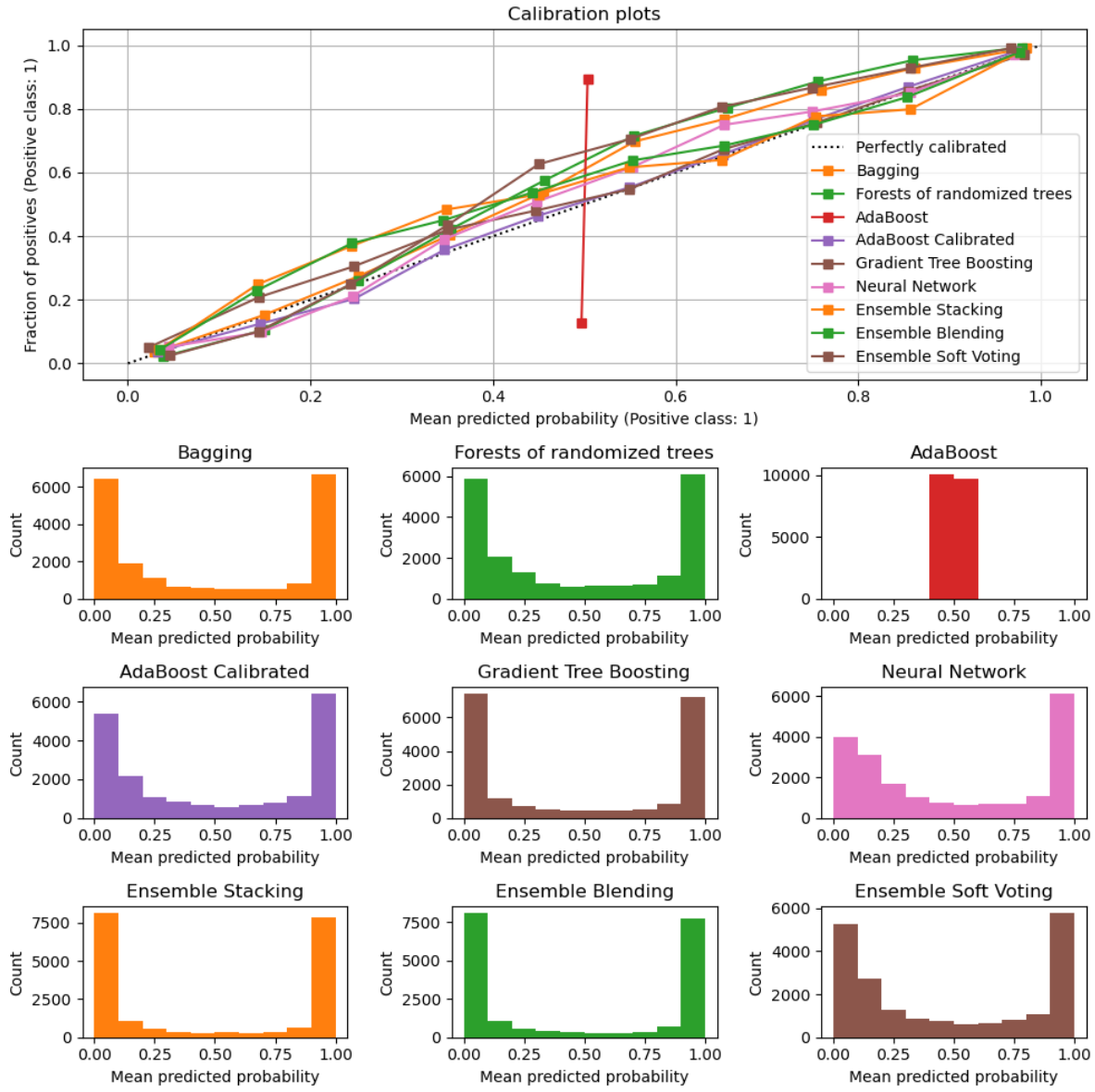
Zhao, Z., Liu, Z. yuan, Xu, C., 2021. Slope Unit-Based Landslide Susceptibility Mapping Using Certainty Factor, Support Vector Machine, Random Forest, CF-SVM and CF-RF Models. Frontiers in Earth Science 9.

Appendix

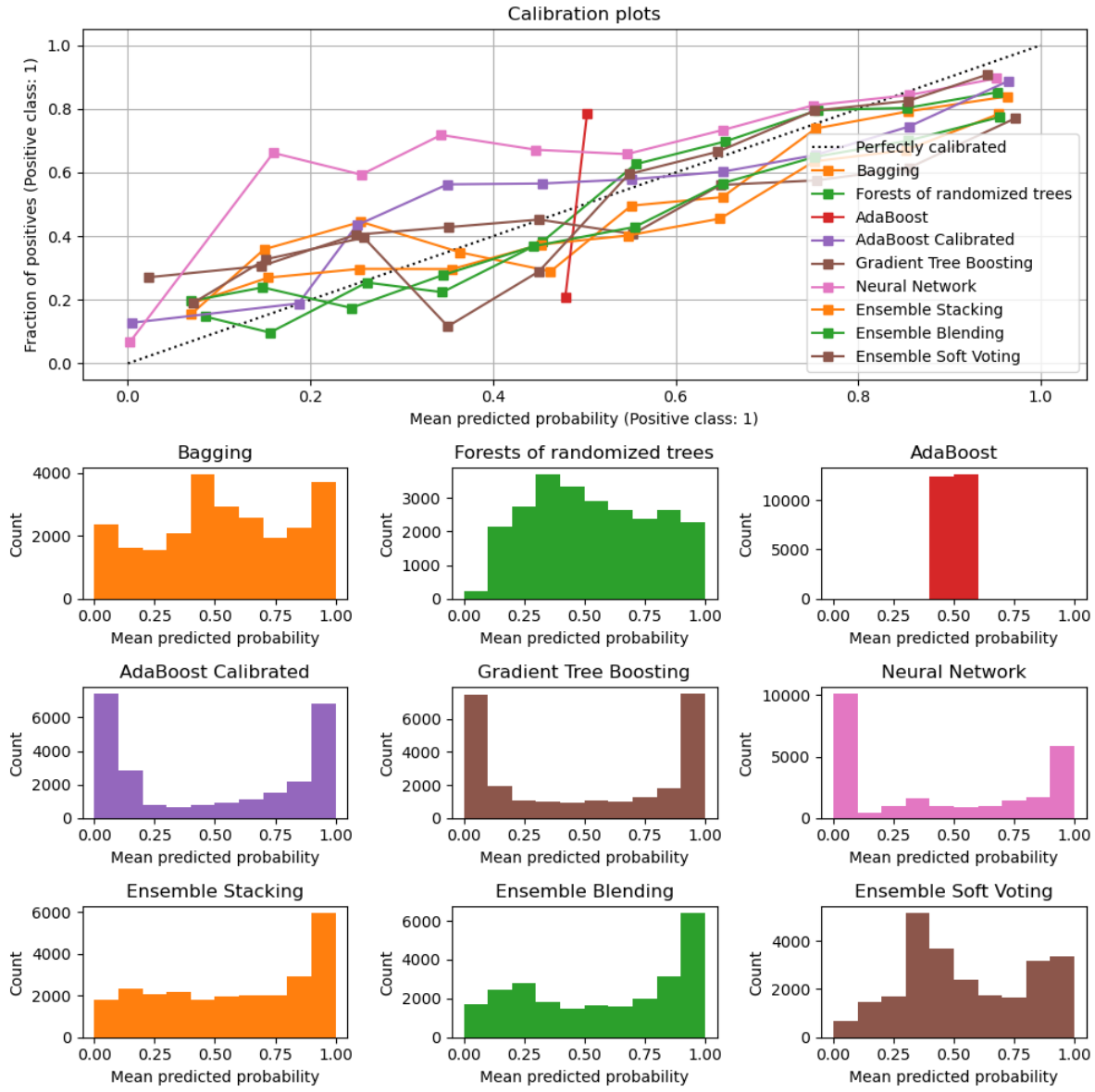
A. Calibration plots of all classifiers for different cases



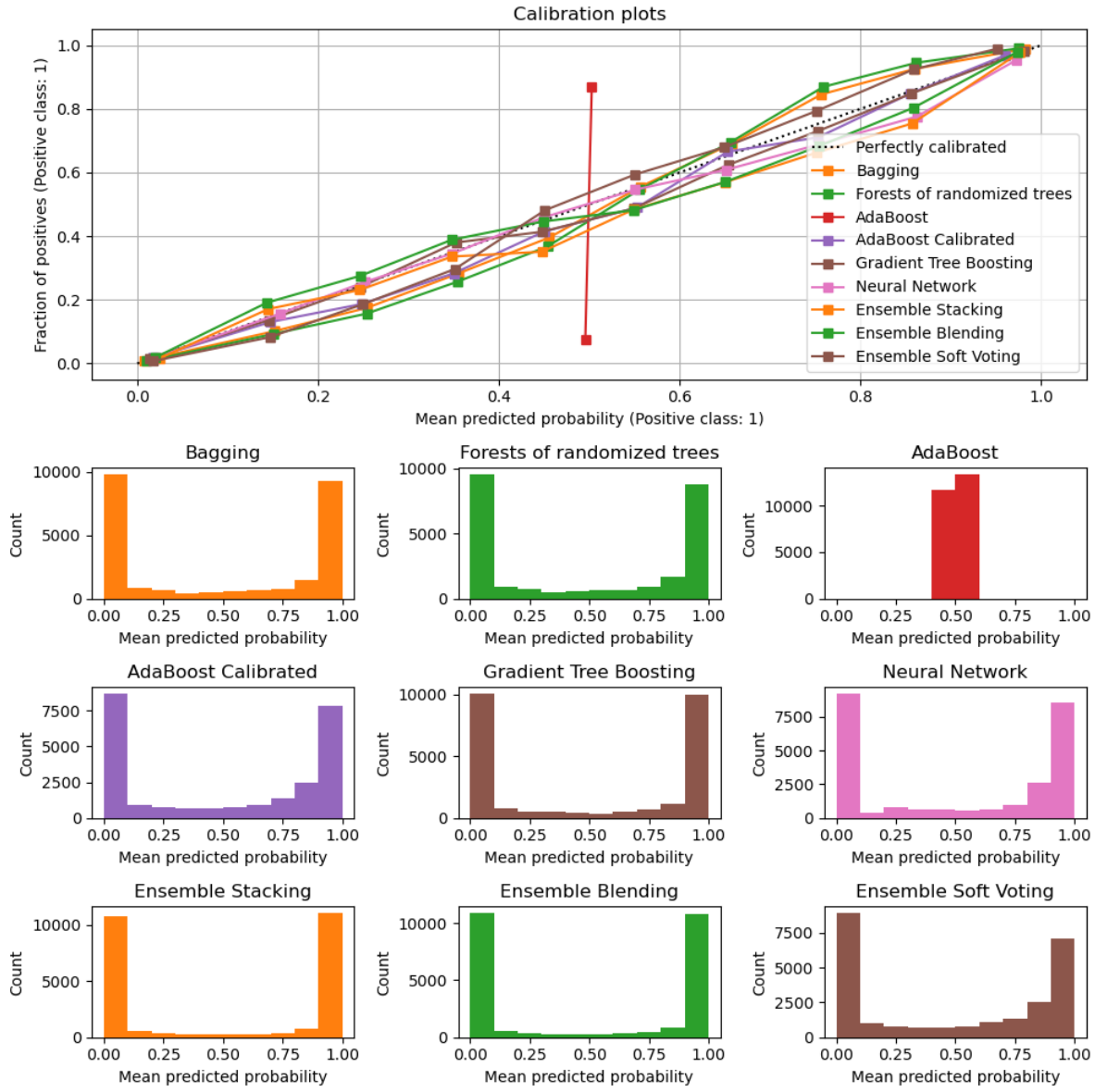
a. Val Tartano (Case: VT)



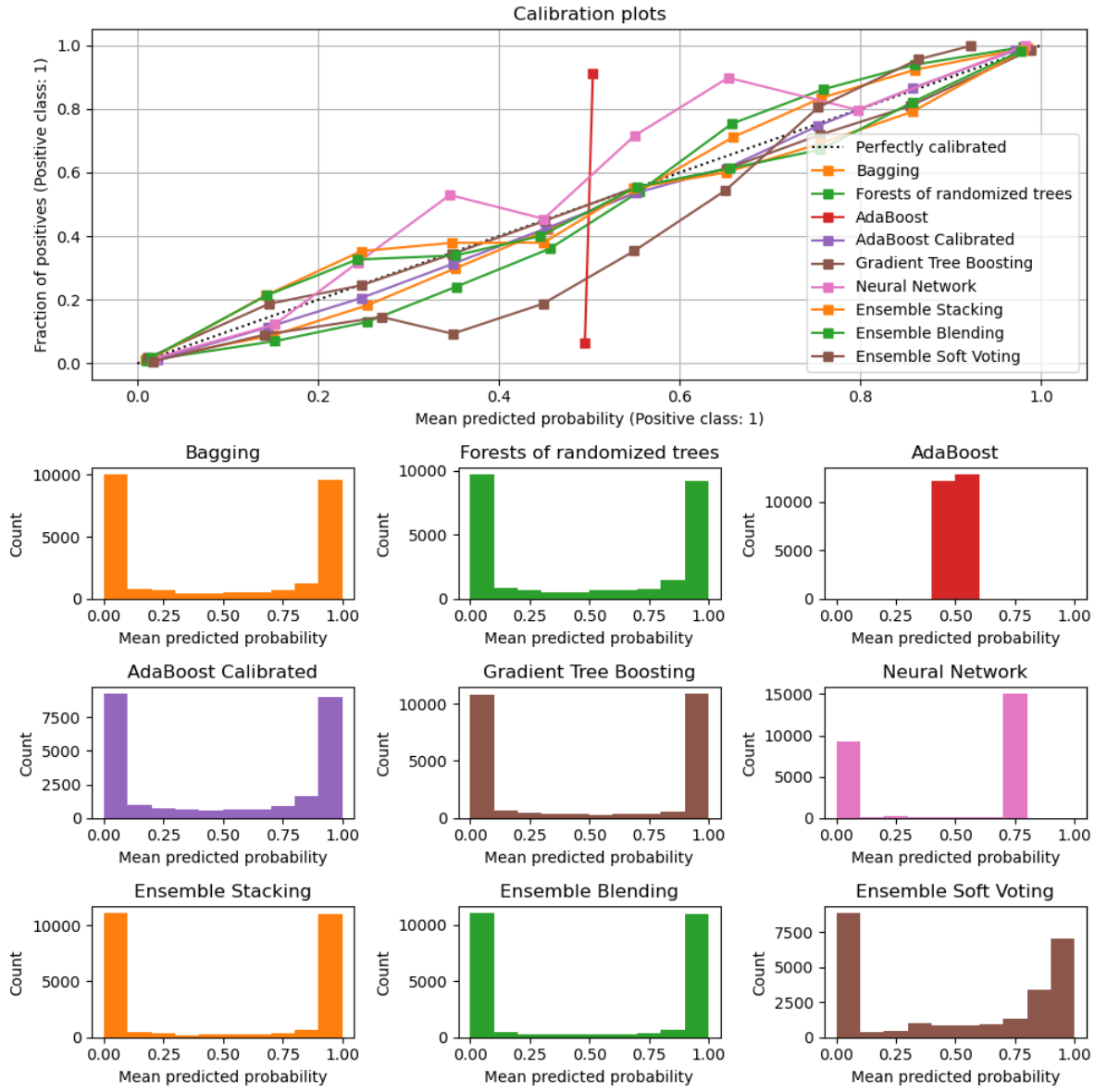
b. Upper Valtellina (Case: UV)



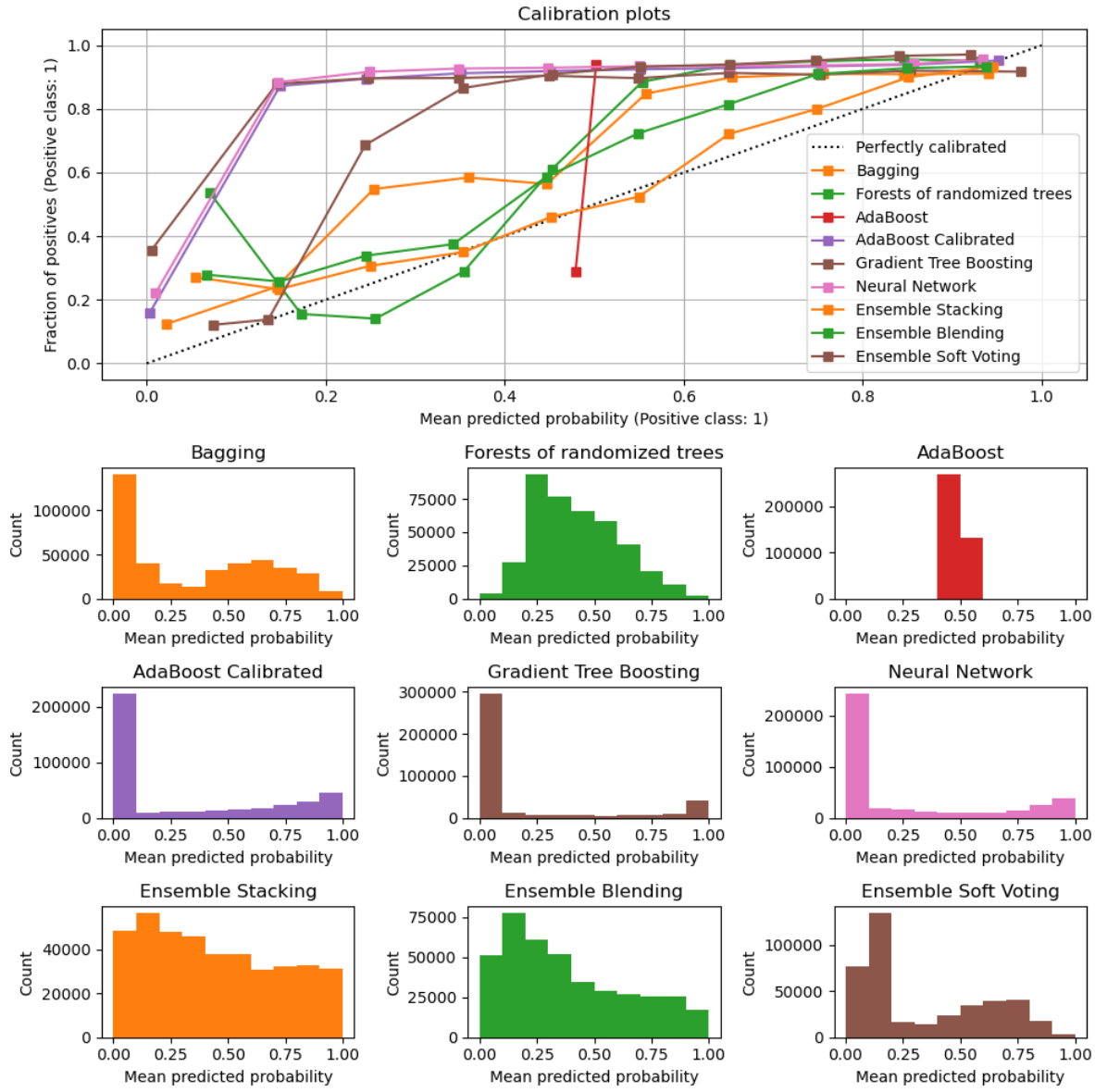
c. ValChiavenna Case 1 (Case: VCC1)



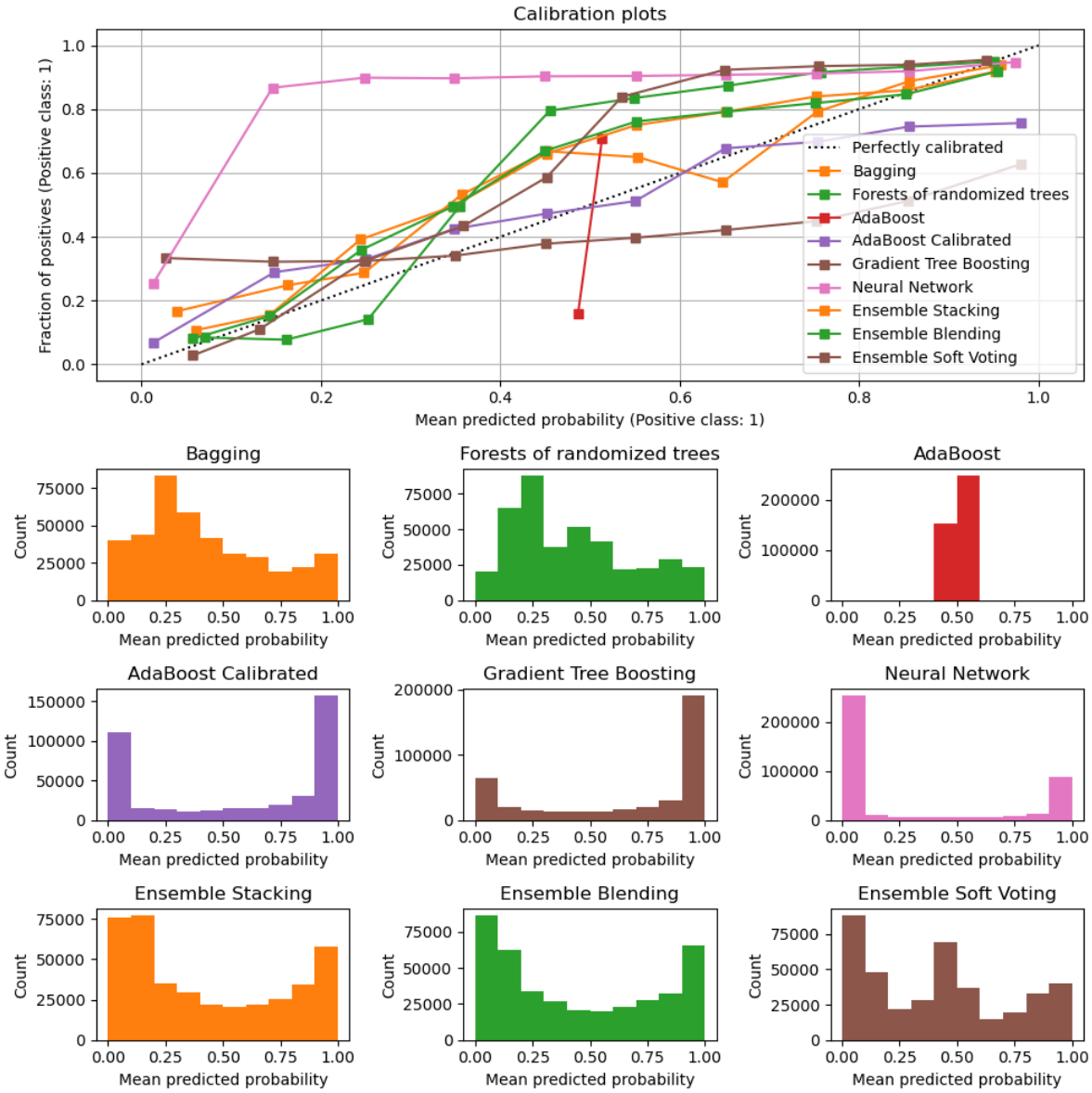
d. ValChiavenna Case 2 (Case: VCC2)



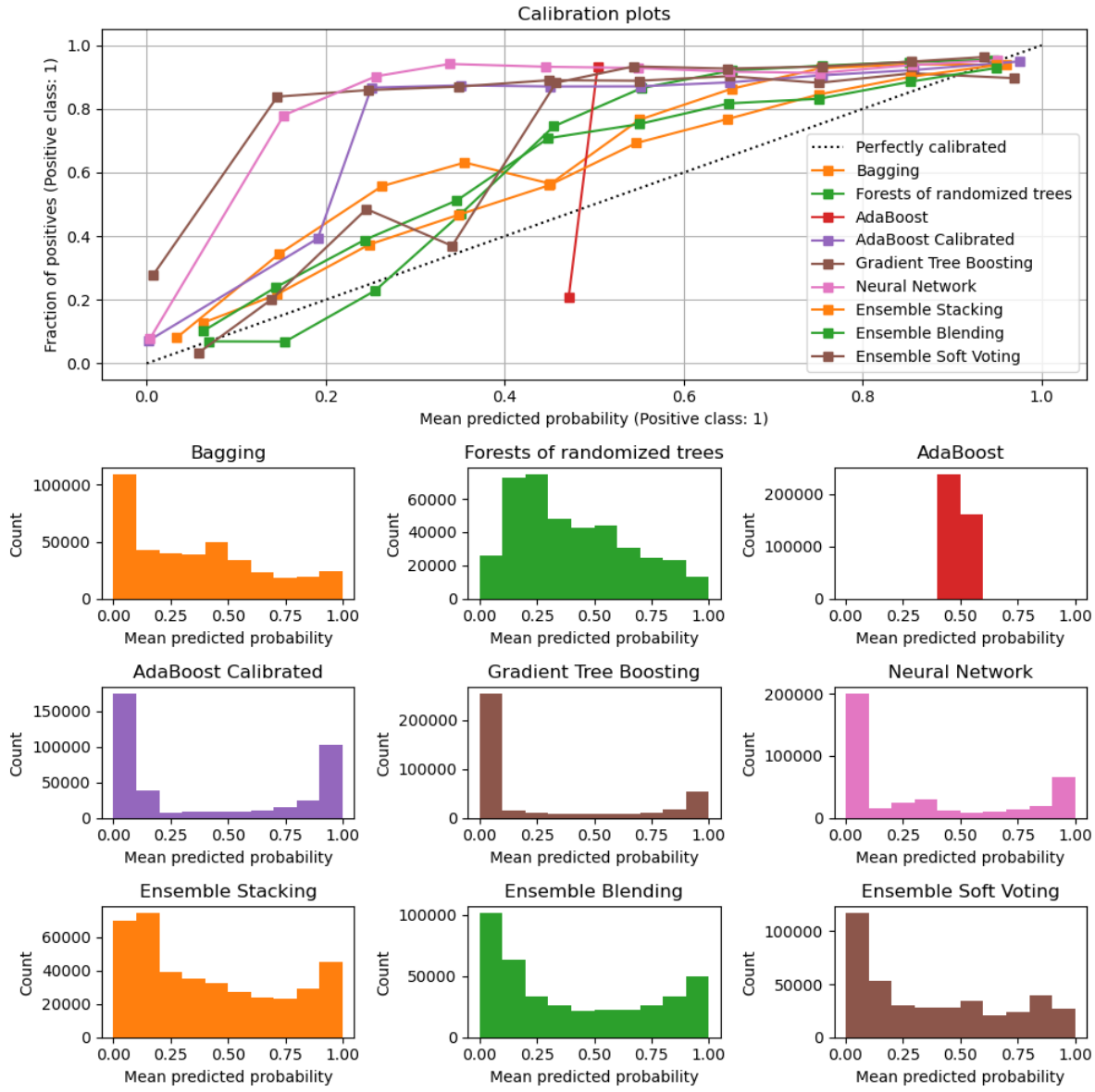
e. ValChiavenna Case 3 (Case: VCC3)



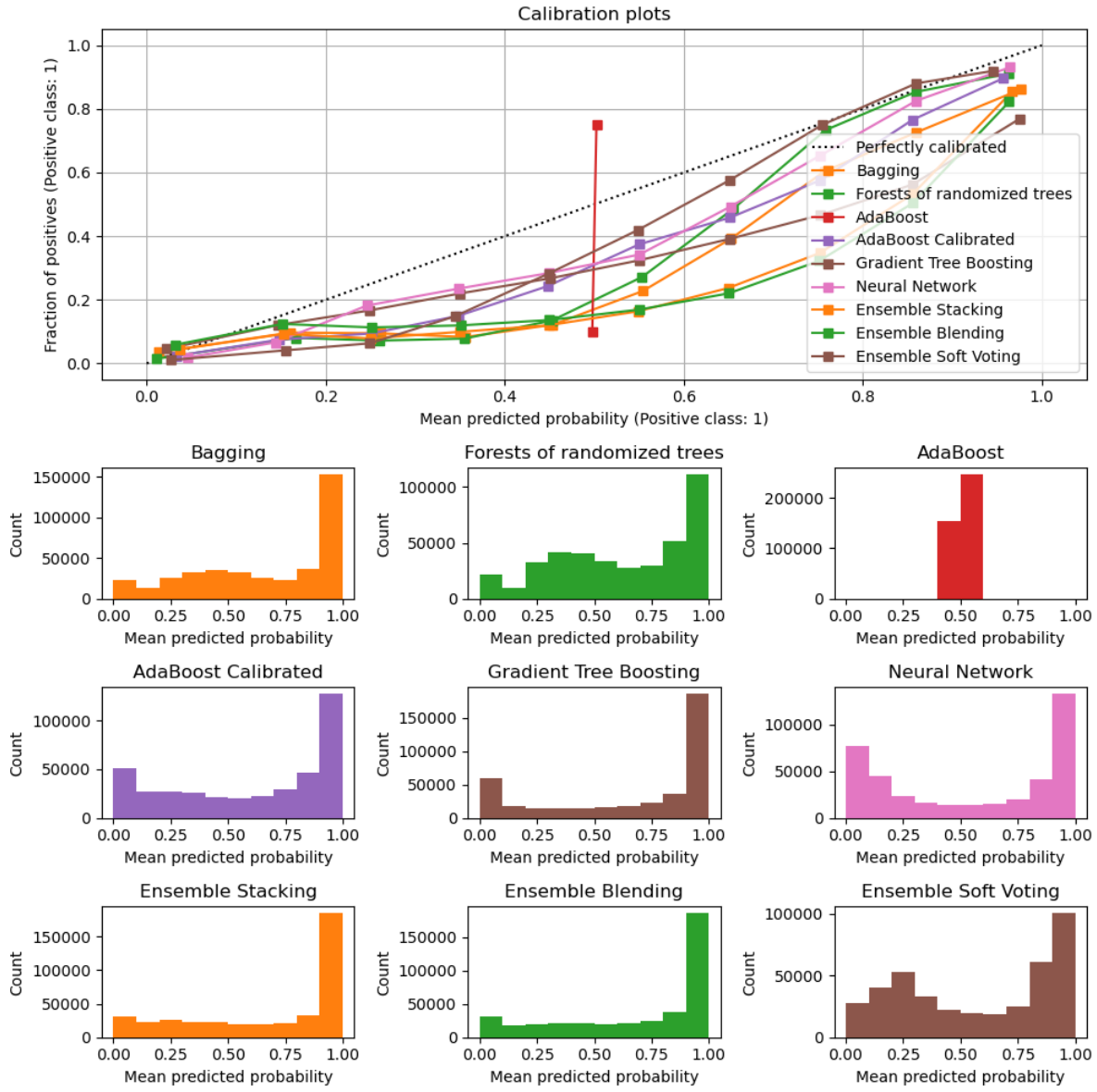
f. Northern Lombardy + Val Tartano (Case: LC1 + VT)



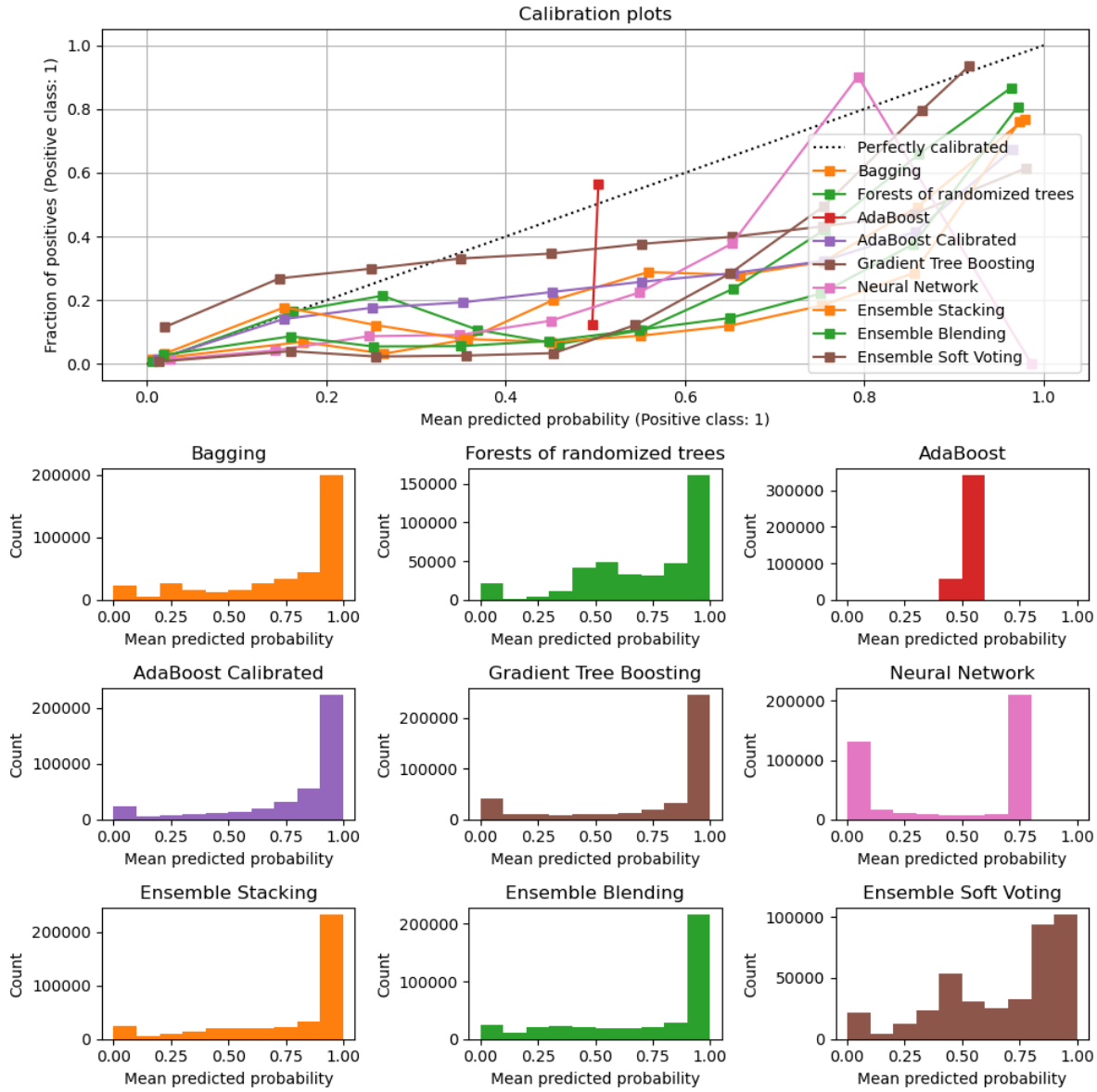
g. Northern Lombardy + Upper Valtellina (Case: LC1 + UV)



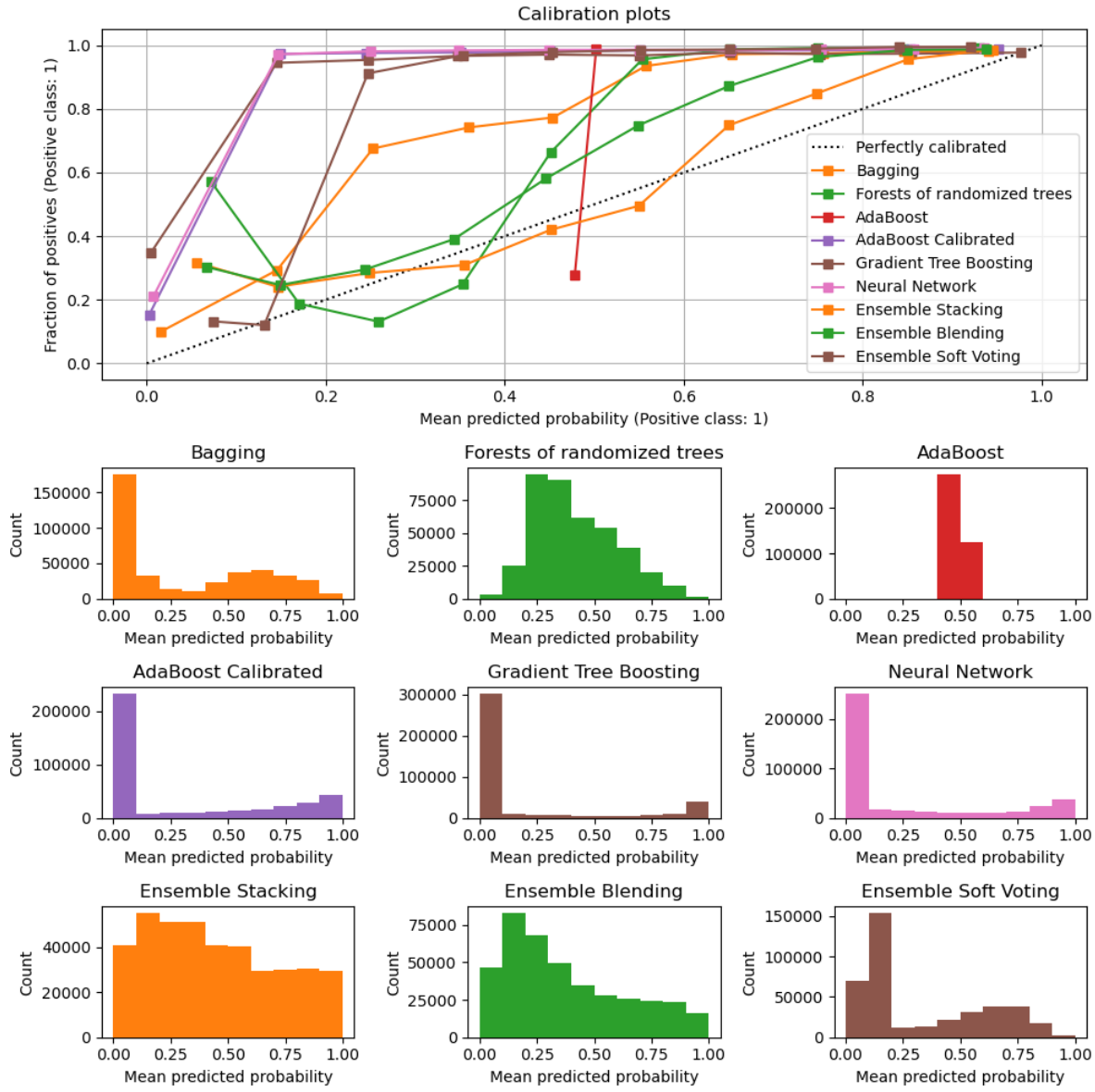
h. Northern Lombardy + Val Tartano, Upper Valtellina (Case: LC1 + VT+ UV)



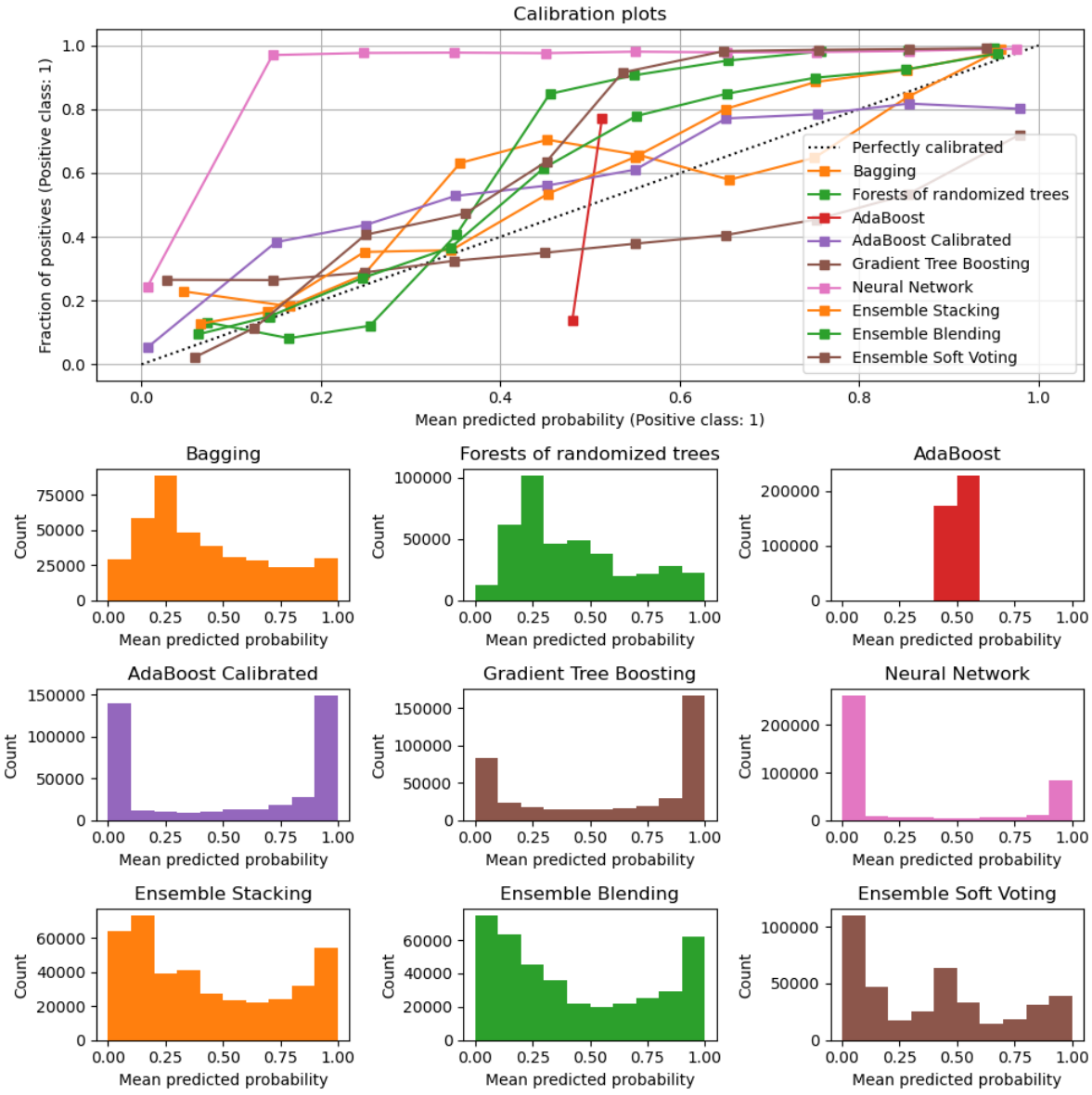
i. Northern Lombardy + ValChiavenna Case 2 (Case: LC1 + VCC2)



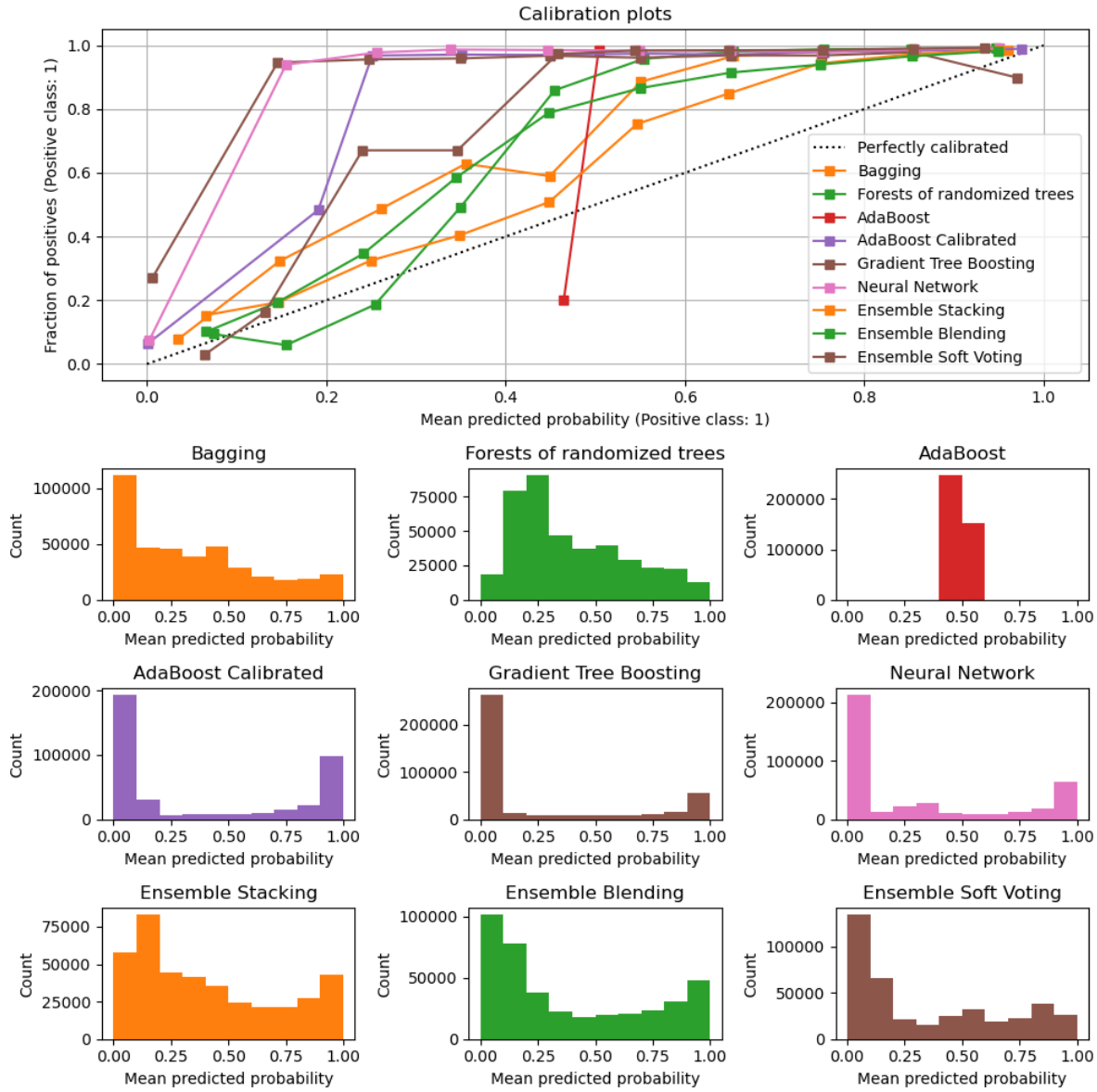
j. Northern Lombardy + ValChiavenna Case 3 (Case: LC1 + VCC3)



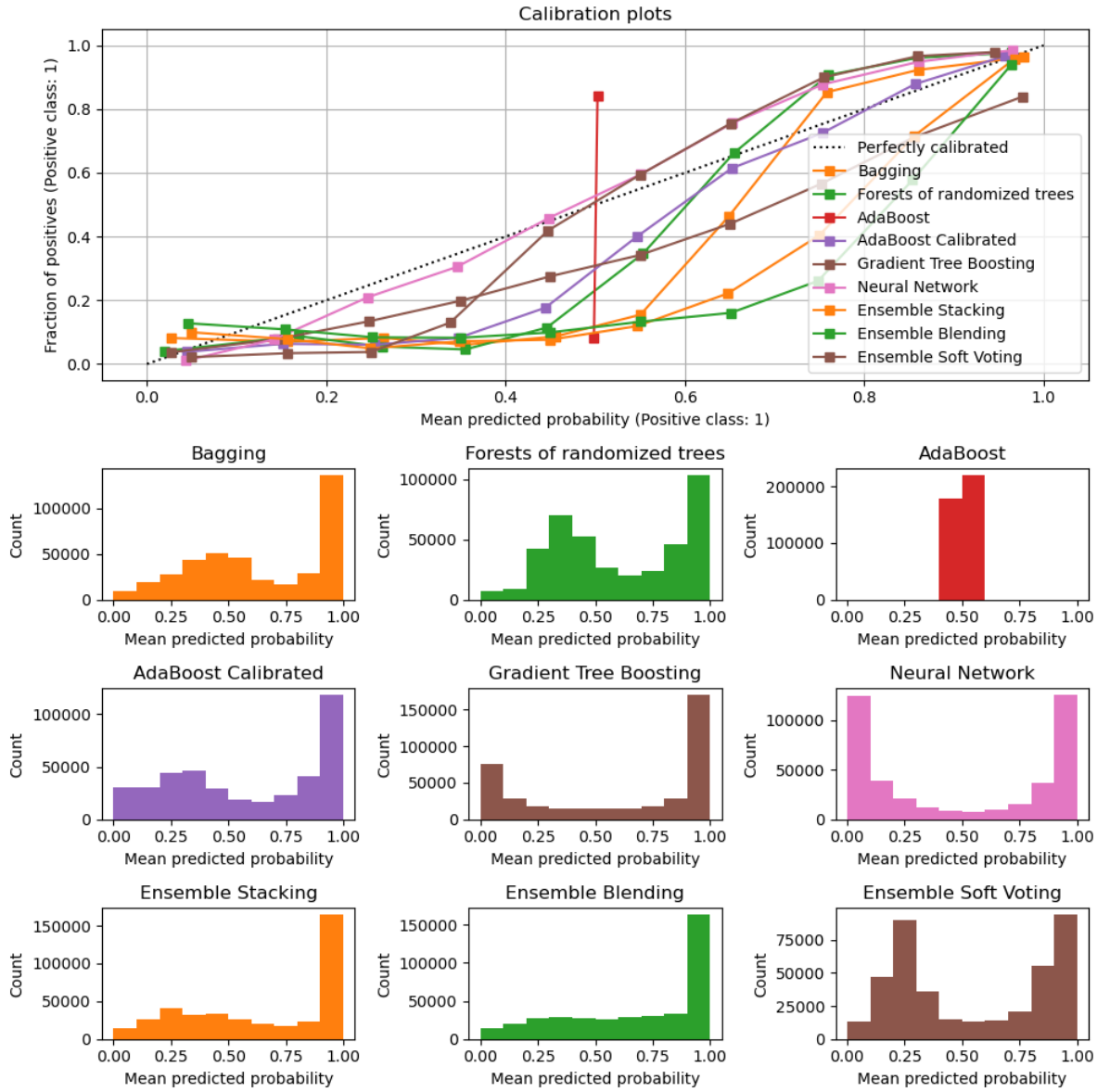
k. Lombardy + Val Tartano (Case: LC2 + VT)



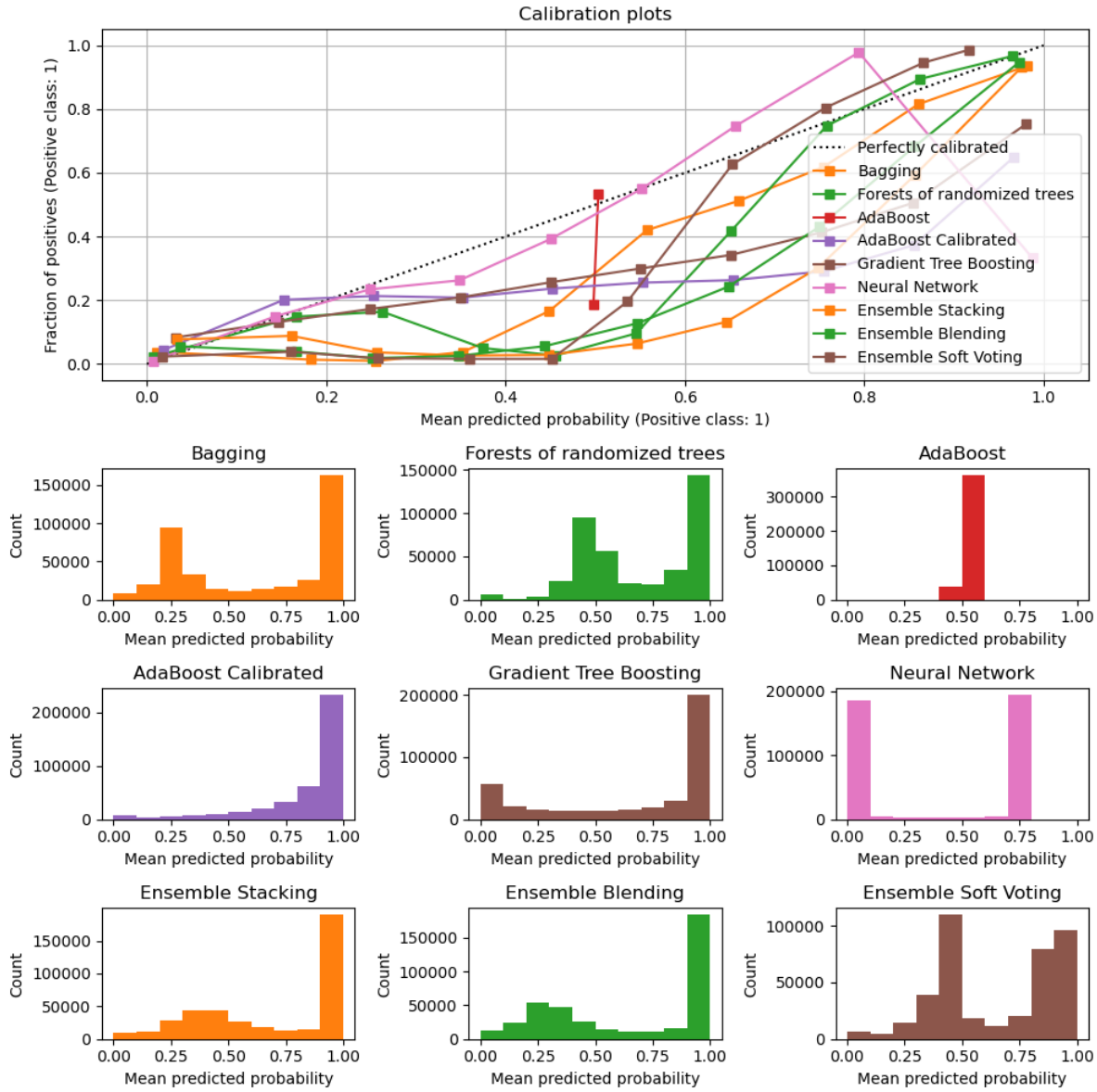
I. Lombardy + Upper Valtellina (Case: LC2 + UV)



m. Lombardy + Val Tartano + Upper Valtellina (Case: LC2 + VT+ UV)

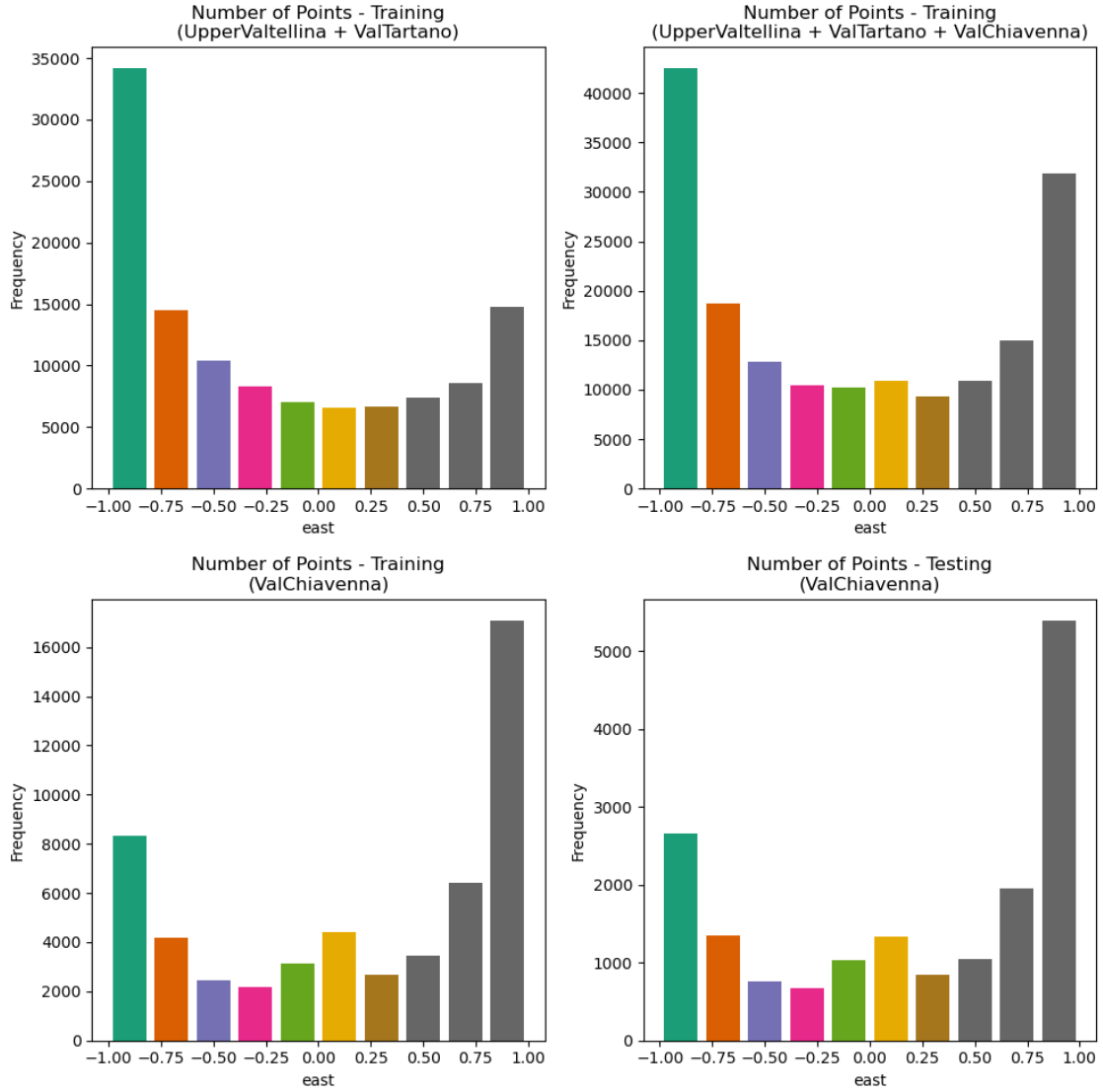


n. Lombardy + ValChiavenna Case 2 (Case: LC2 + VCC2)

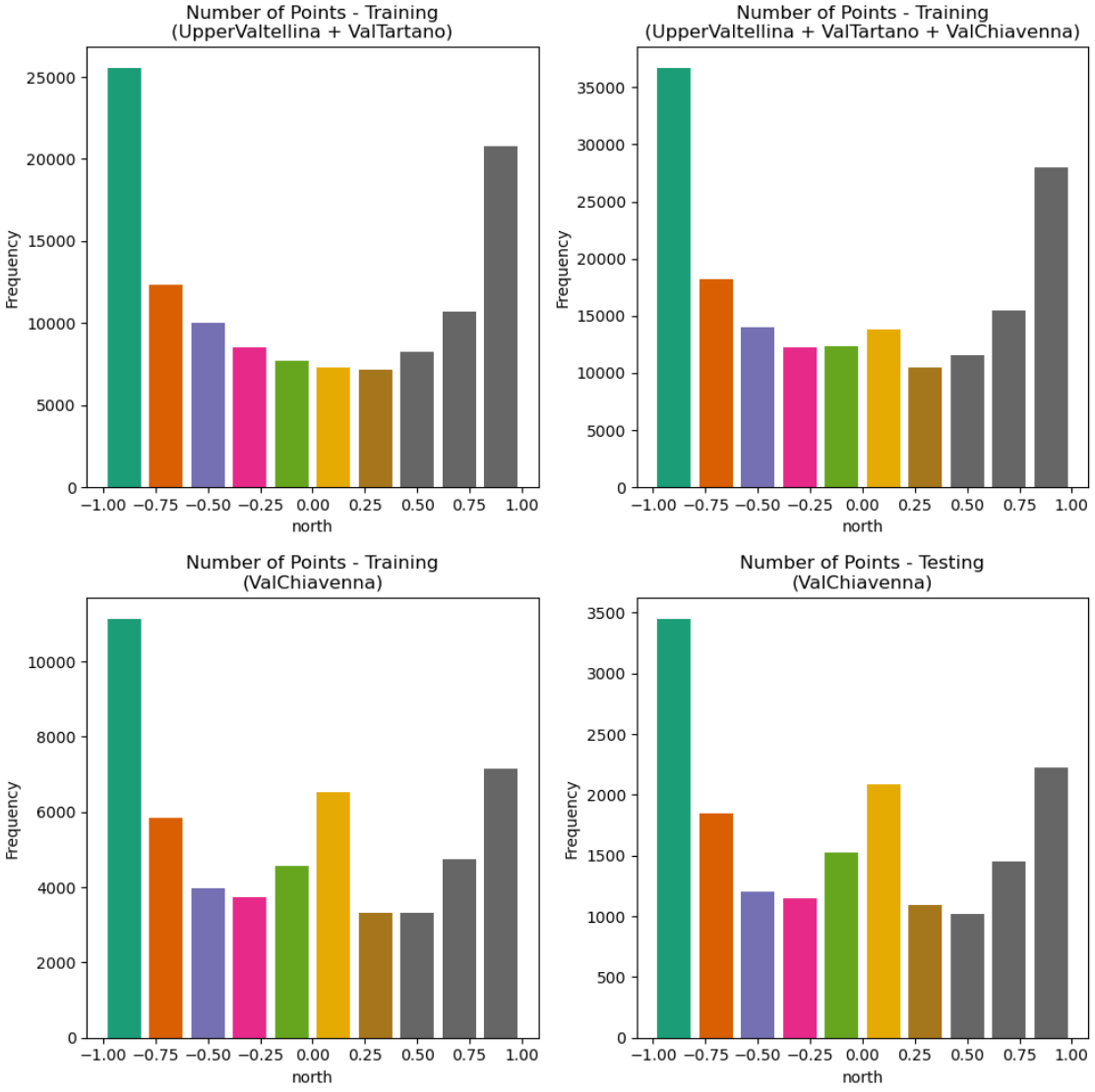


o. Lombardy + ValChiavenna Case 3 (Case: LC2 + VCC3)
Figure A.1 Calibration plots of all classifiers

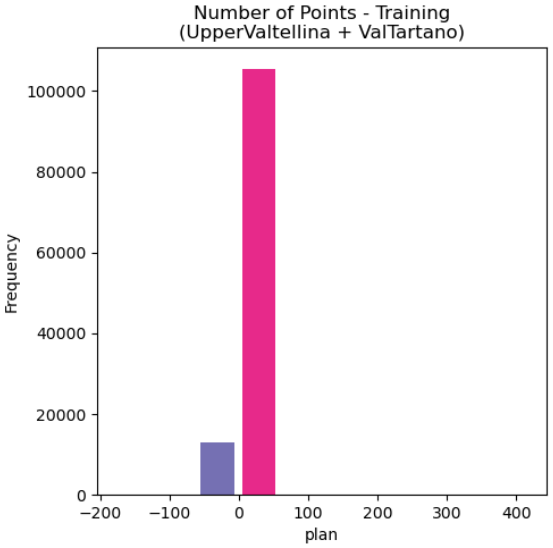
B. Distribution of different factors in the training dataset and testing dataset for ValChiavenna Case 1, 2, 3

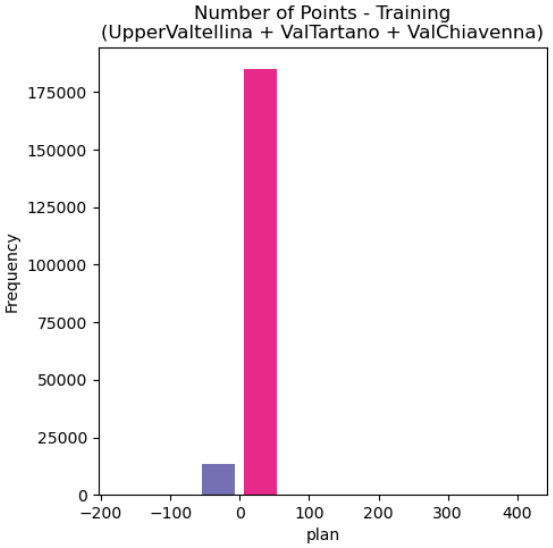


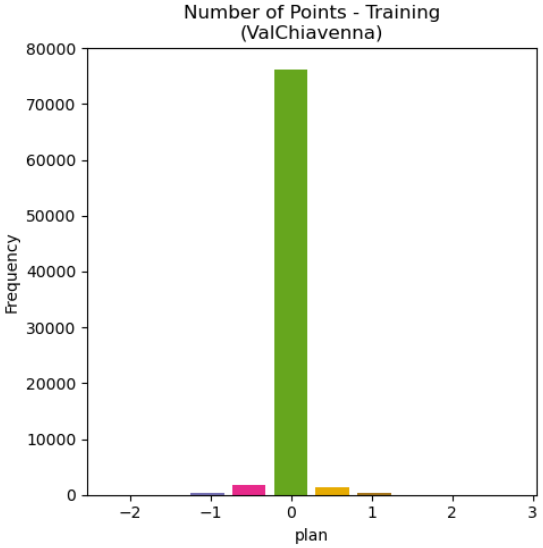
a. east

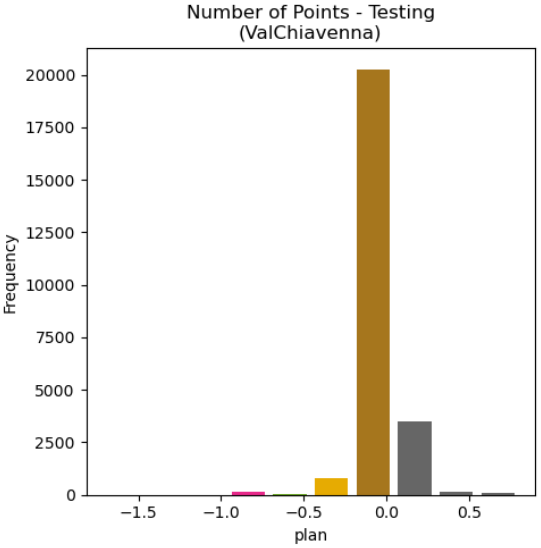


a. north

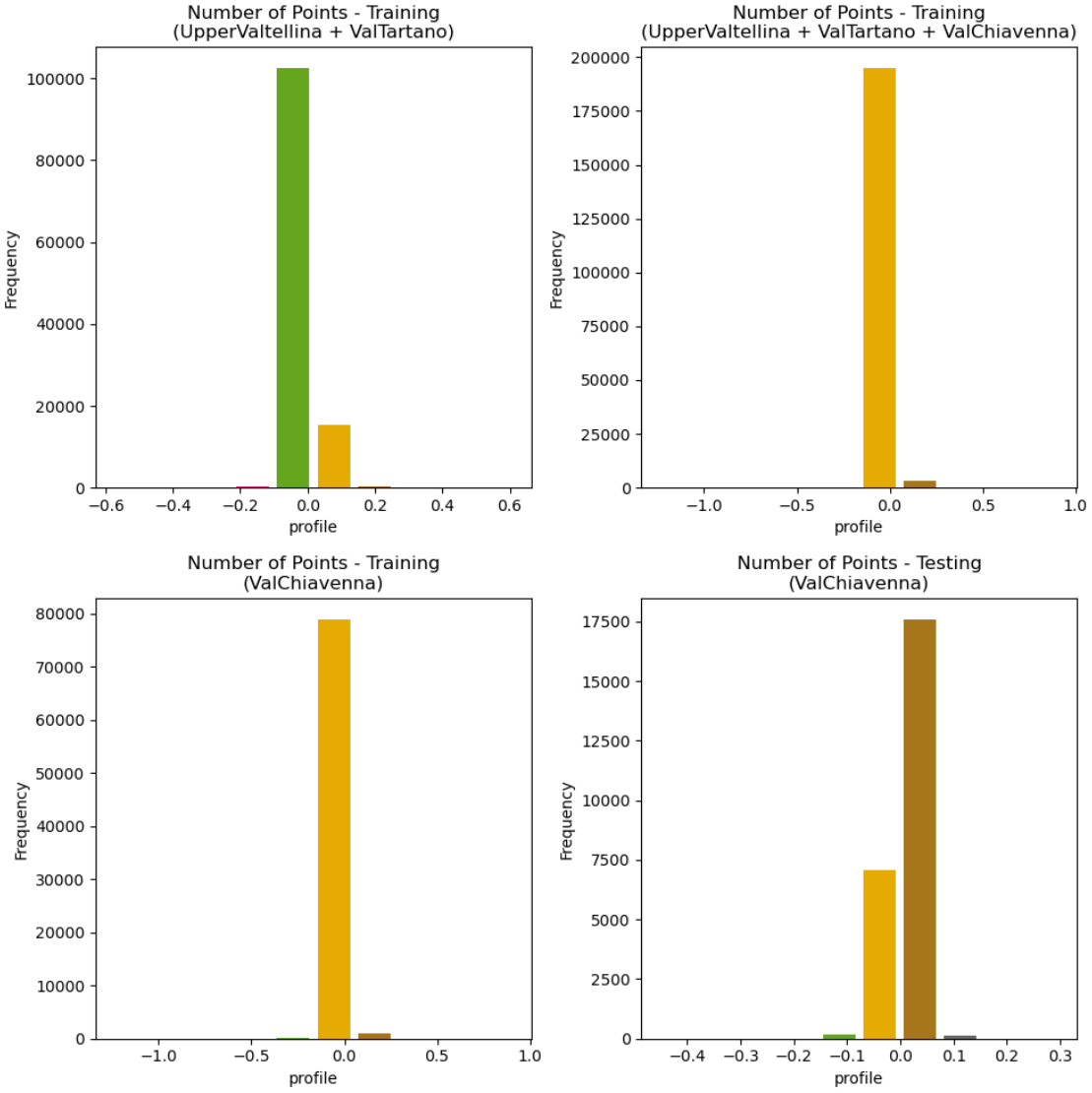

Number of Points - Training
(UpperValtellina + ValTartano)
Frequency
plan


Number of Points - Training
(UpperValtellina + ValTartano + ValChiavenna)
Frequency
plan

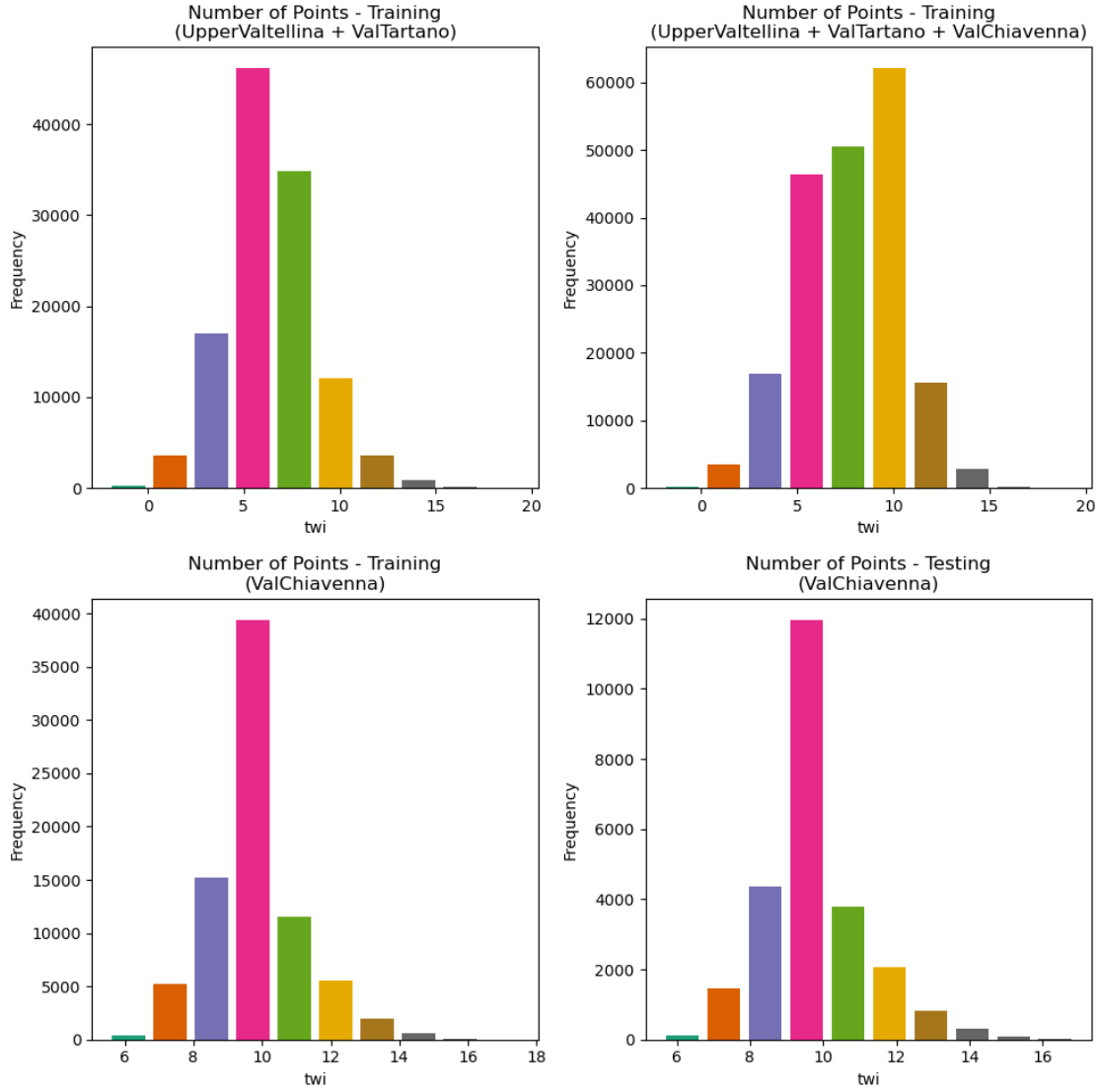

Number of Points - Training
(ValChiavenna)
Frequency
plan


Number of Points - Testing
(ValChiavenna)
Frequency
plan

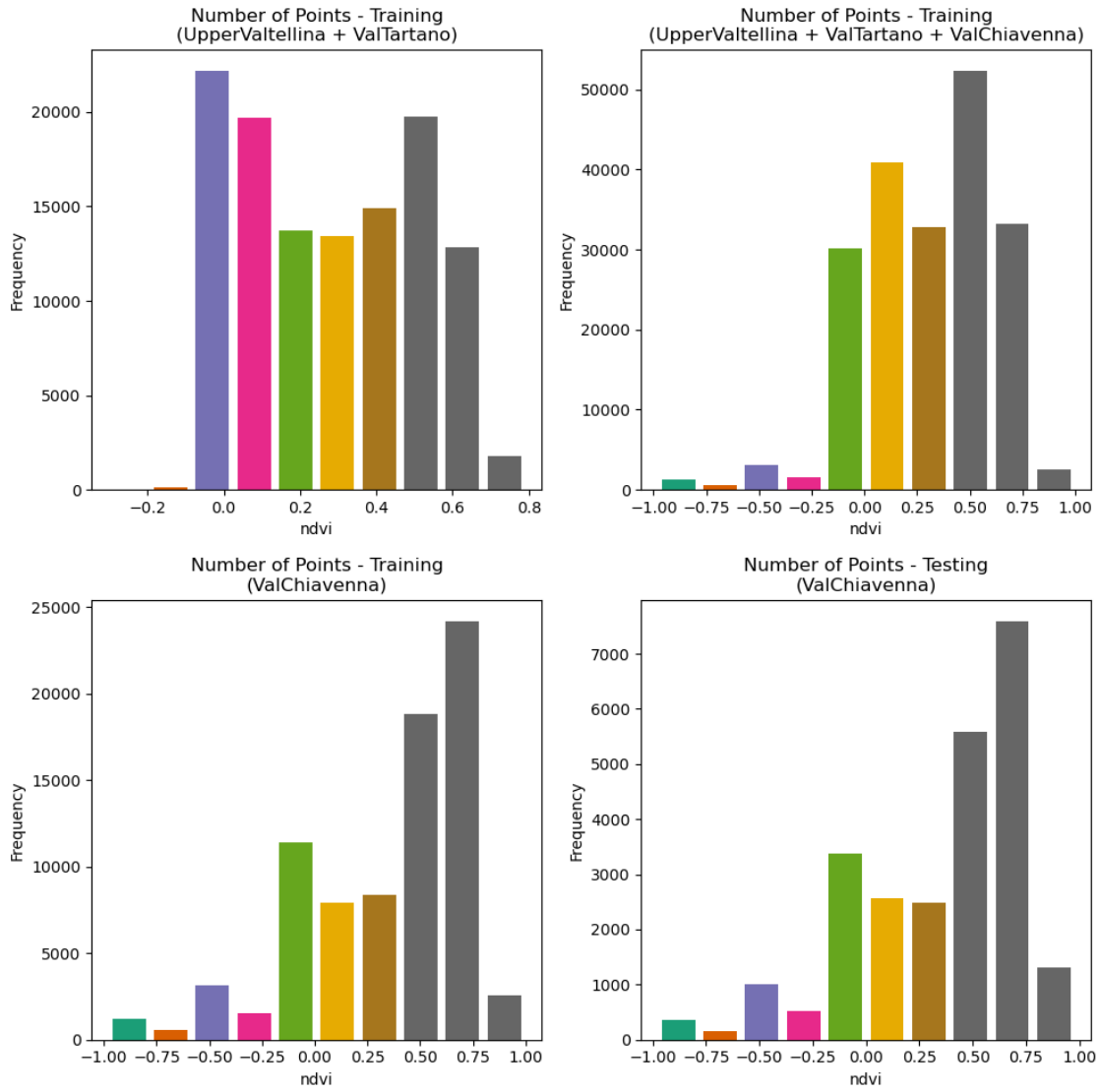
c. north



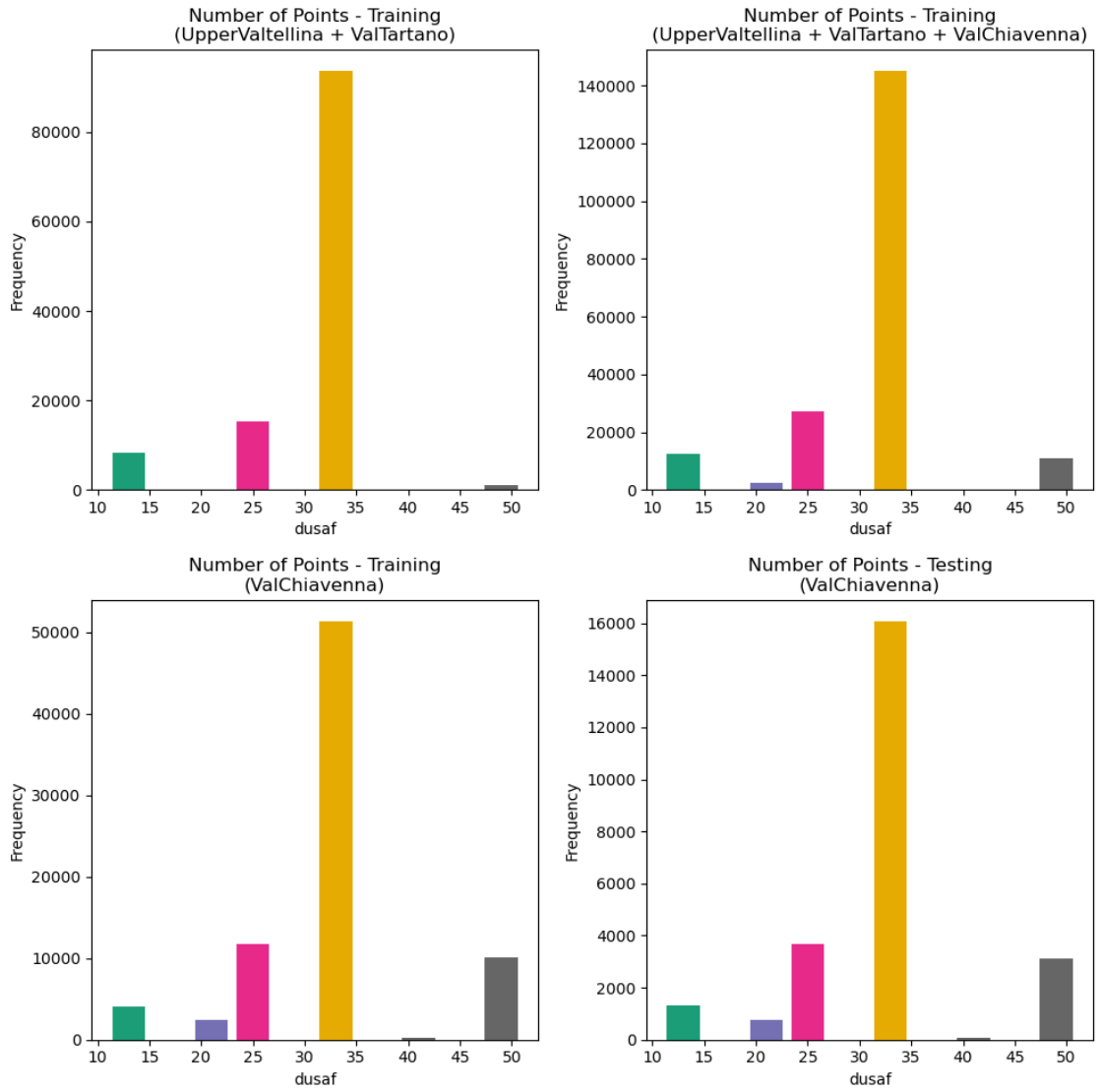
d. profile


Number of Points - Training
(UpperValtellina + ValTartano)
Frequency
twi
Number of Points - Training
(UpperValtellina + ValTartano + ValChiavenna)
Frequency
twi
Number of Points - Training
(ValChiavenna)
Frequency
twi
Number of Points - Testing
(ValChiavenna)
Frequency
twi

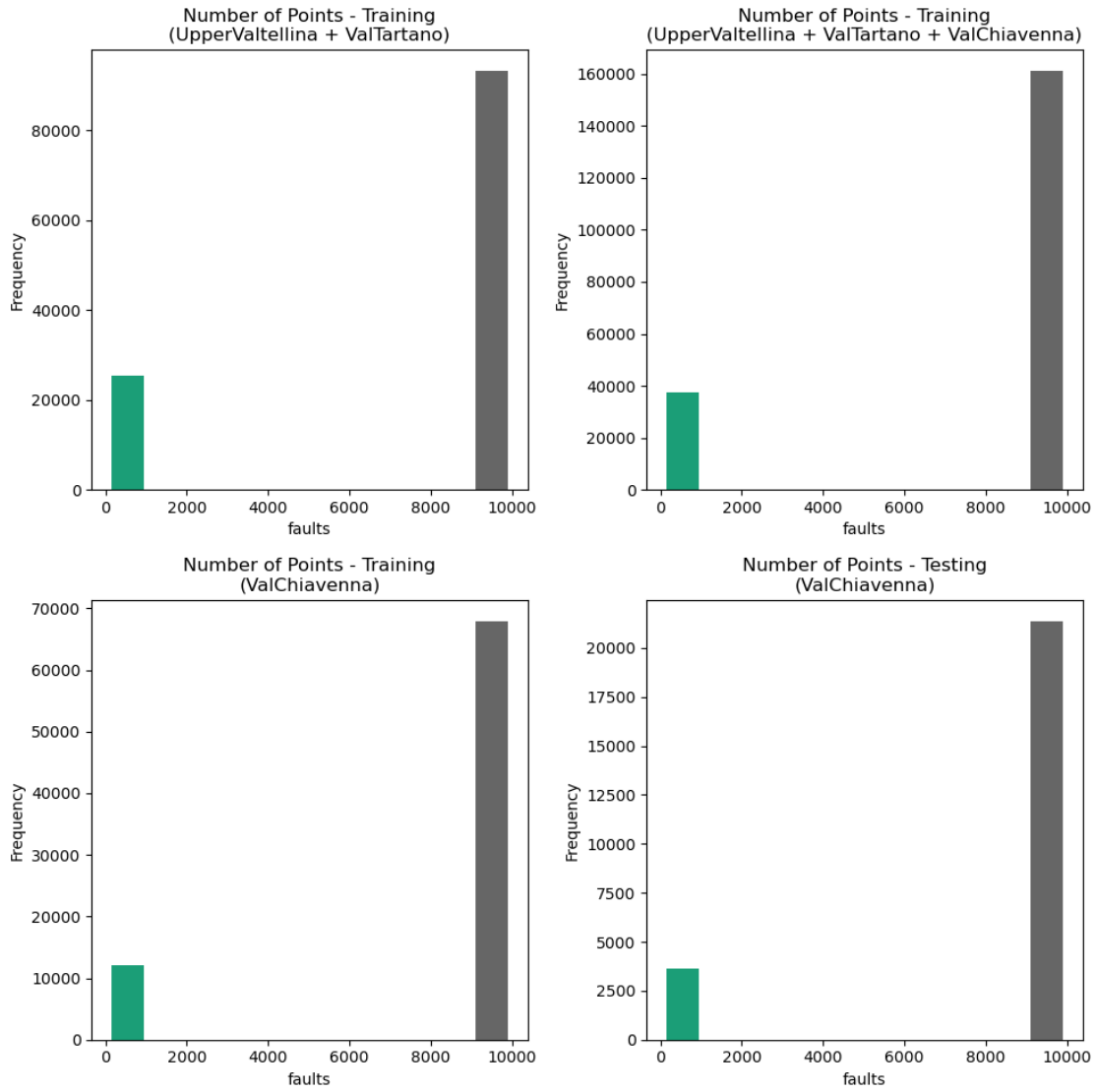
e. twi


Number of Points - Training
(UpperValtellina + ValTartano)
Frequency
ndvi
Number of Points - Training
(UpperValtellina + ValTartano + ValChiavenna)
Frequency
ndvi
Number of Points - Training
(ValChiavenna)
Frequency
ndvi
Number of Points - Testing
(ValChiavenna)
Frequency
ndvi

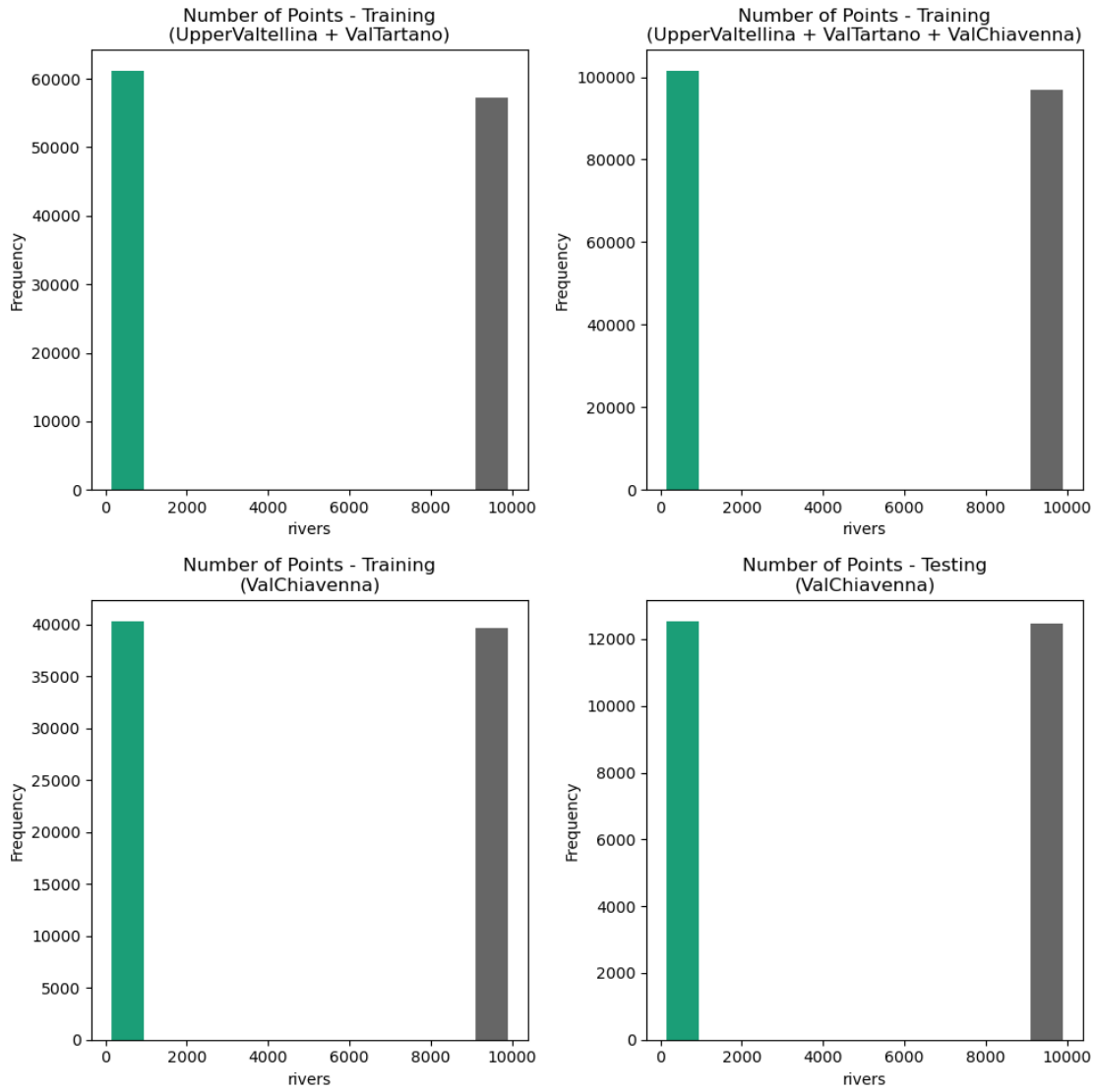
f. ndvi


Number of Points - Training
(UpperValtellina + ValTartano)
Frequency
dusaf
Number of Points - Training
(UpperValtellina + ValTartano + ValChiavenna)
Frequency
dusaf
Number of Points - Training
(ValChiavenna)
Frequency
dusaf
Number of Points - Testing
(ValChiavenna)
Frequency
dusaf

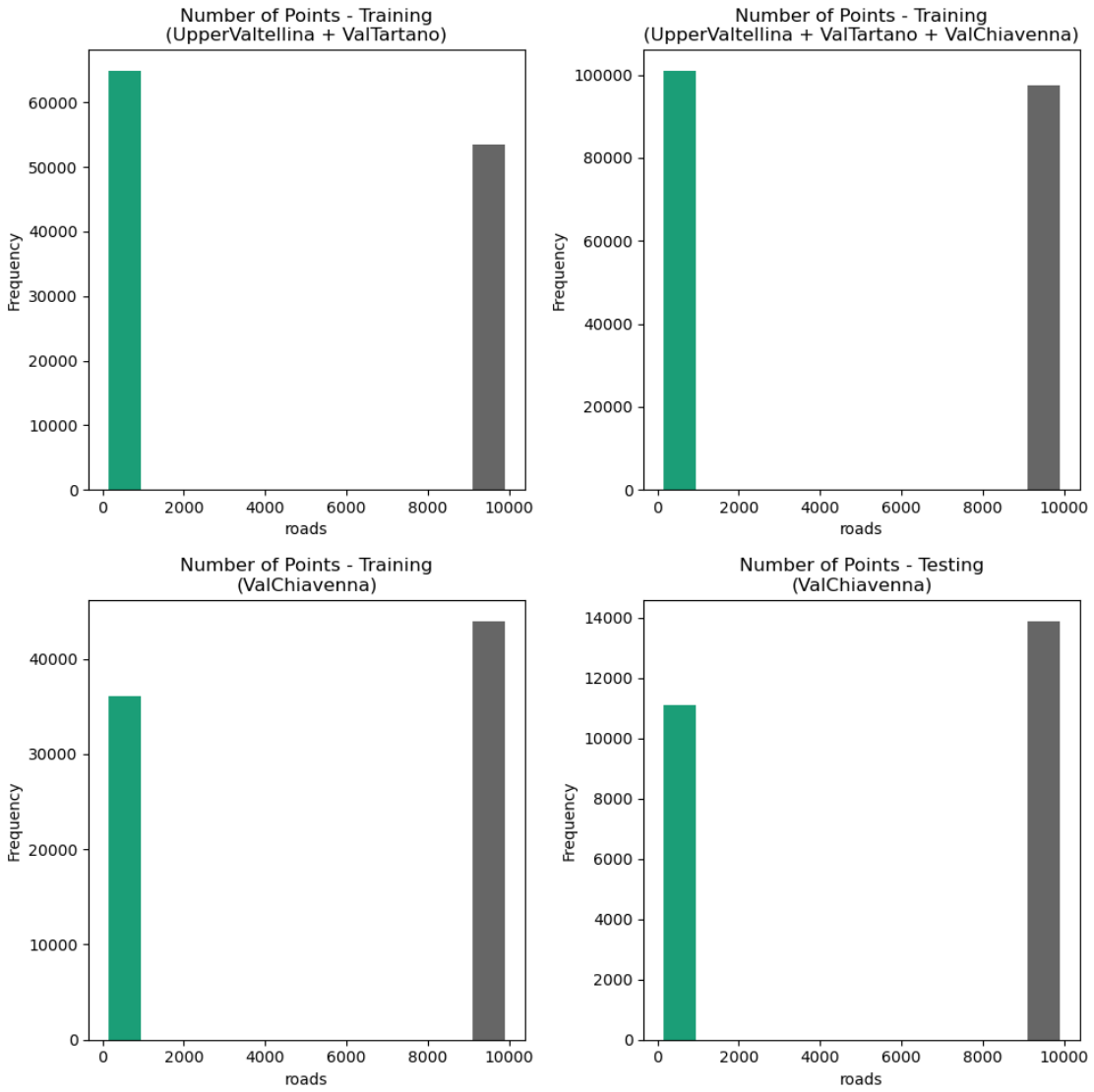
g. dusaf


Number of Points - Training
(UpperValtellina + ValTartano)
Frequency
faults
Number of Points - Training
(UpperValtellina + ValTartano + ValChiavenna)
Frequency
faults
Number of Points - Training
(ValChiavenna)
Frequency
faults
Number of Points - Testing
(ValChiavenna)
Frequency
faults

h. faults


Number of Points - Training
(UpperValtellina + ValTartano)
Frequency
rivers
Number of Points - Training
(UpperValtellina + ValTartano + ValChiavenna)
Frequency
rivers
Number of Points - Training
(ValChiavenna)
Frequency
rivers
Number of Points - Testing
(ValChiavenna)
Frequency
rivers

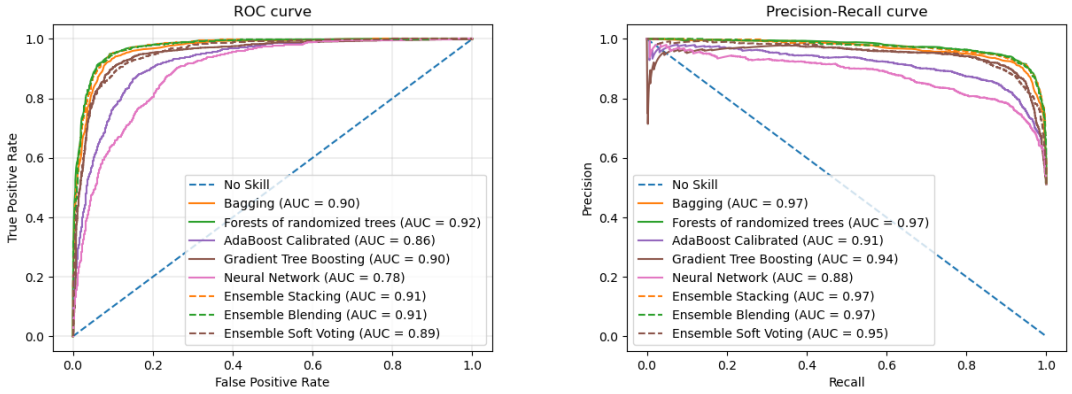
g. rivers



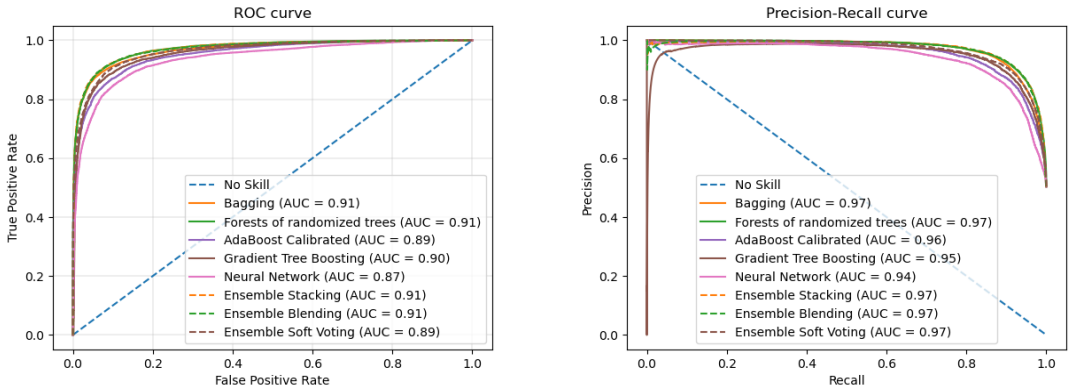
h. roads

Figure B.1: distribution of different factors in the training dataset and testing dataset for ValChiavenna Case 1, 2, 3

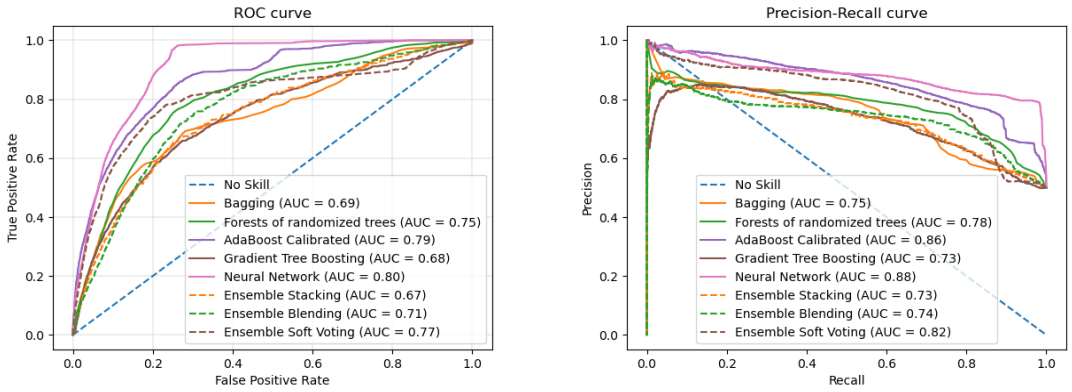
C. Accuracy statistics of all classifiers



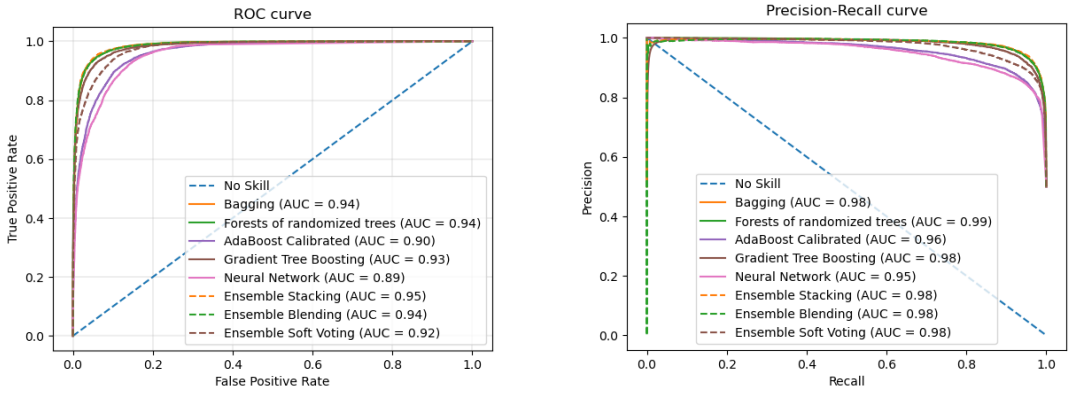
a. Val Tartano (Case: VT)



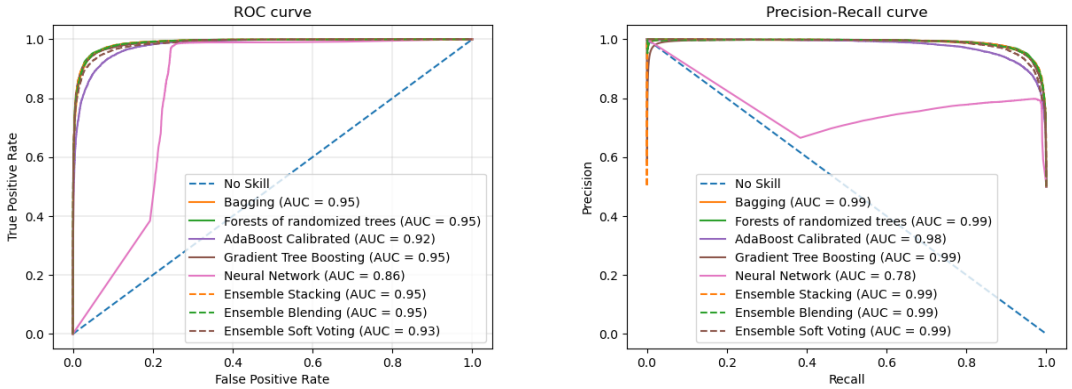
b. Upper Valtellina (Case: UV)



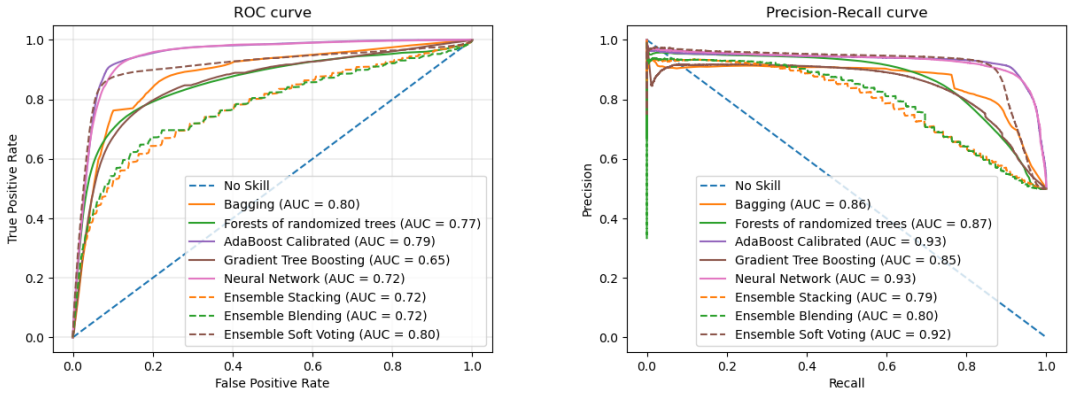
c. ValChiavenna Case 1 (Case: VCC1)



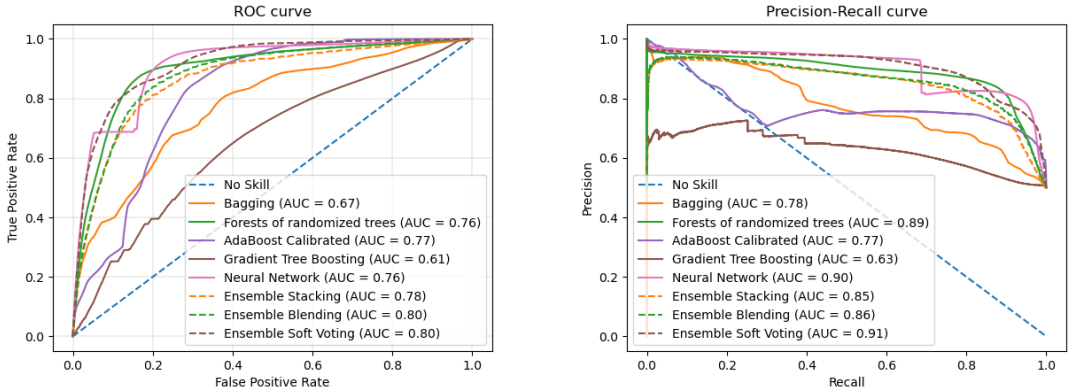
d. ValChiavenna Case 2 (Case: VCC2)



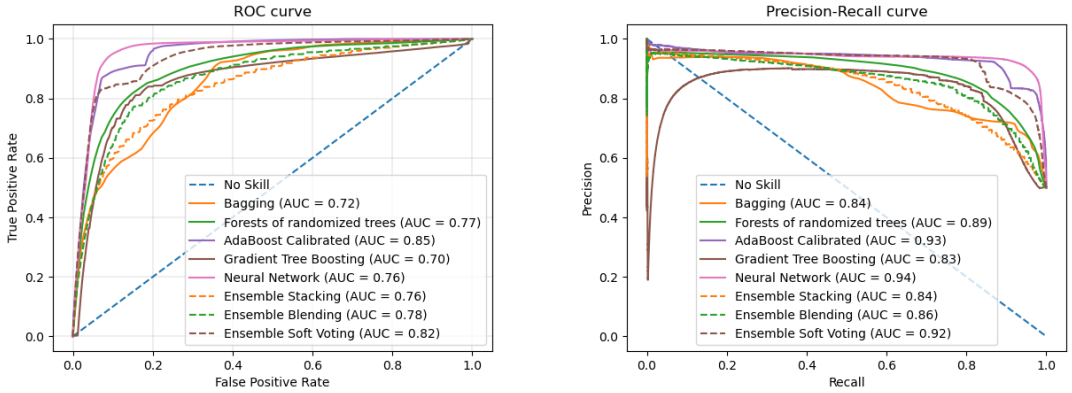
e. ValChiavenna Case 3 (Case: VCC3)



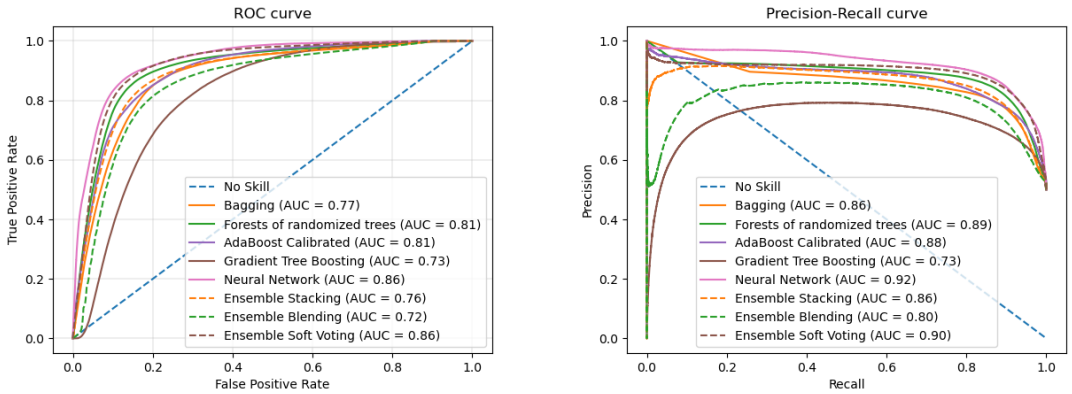
f. Northern Lombardy + Val Tartano (Case: LC1 + VT)



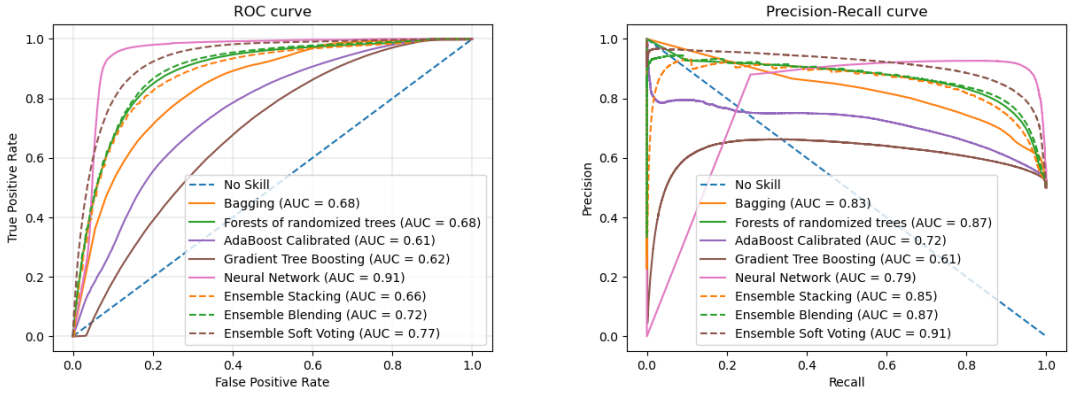
g. Northern Lombardy + Upper Valtellina (Case: LC1 + UV)



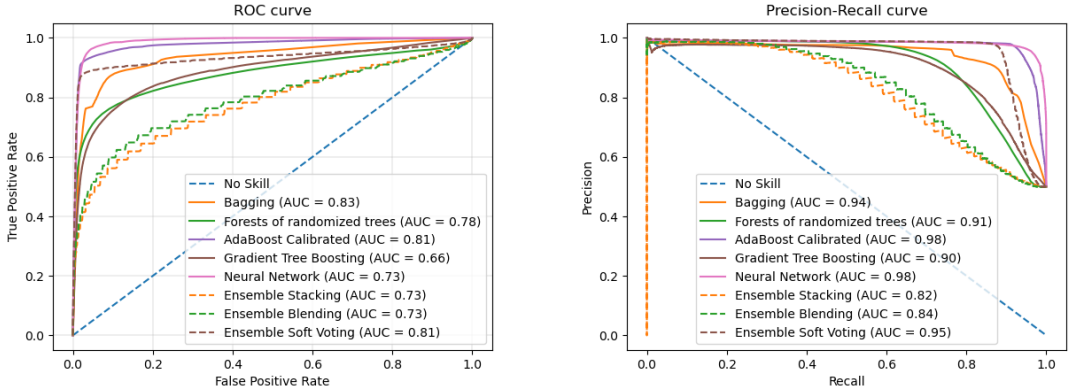
h. Northern Lombardy + Val Tartano, Upper Valtellina (Case: LC1 + VT+ UV)



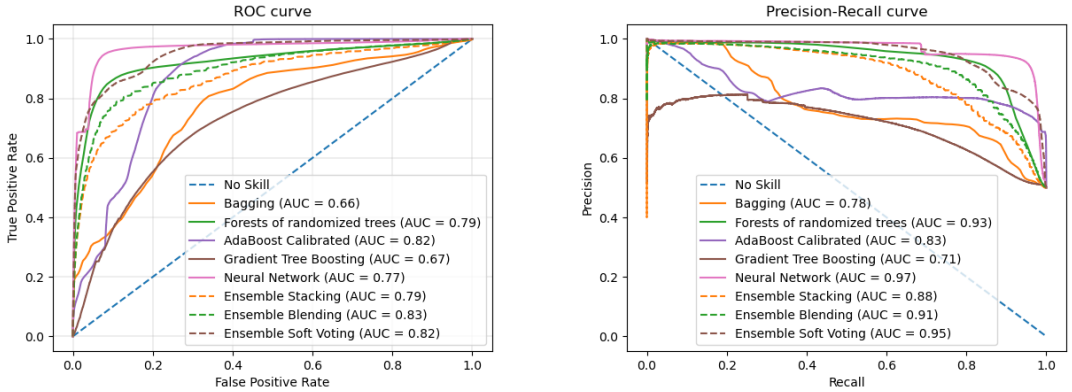
i. Northern Lombardy + ValChiavenna Case 2 (Case: LC1 + VCC2)



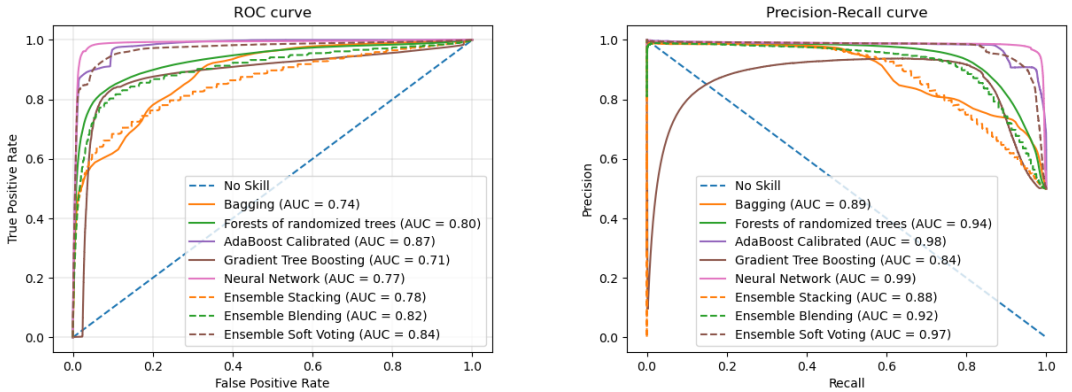
j. Northern Lombardy + ValChiavenna Case 3 (Case: LC1 + VCC3)



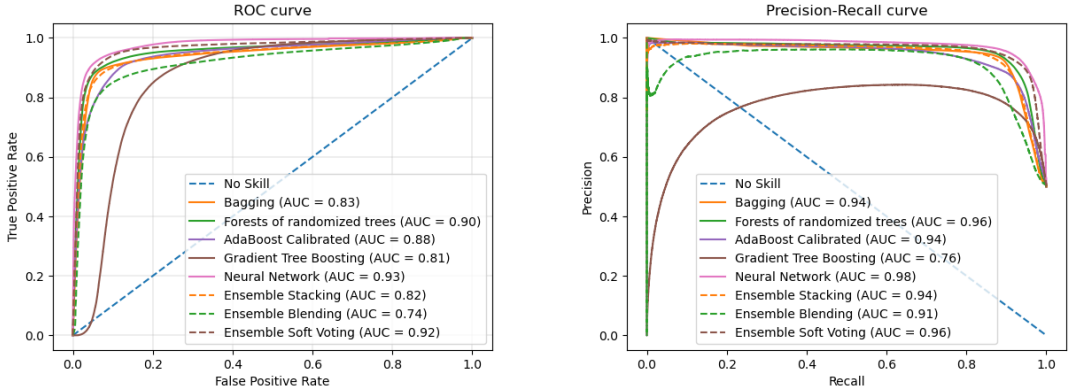
k. Lombardy + Val Tartano (Case: LC2 + VT)



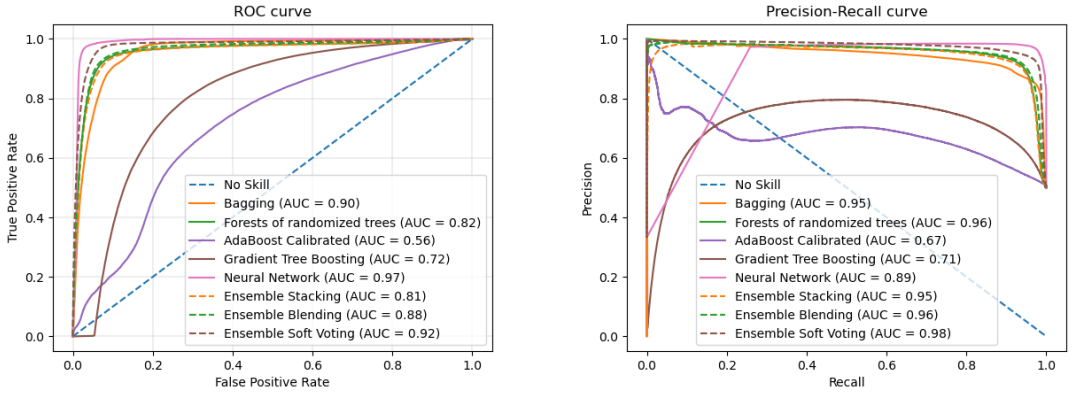
l. Lombardy + Upper Valtellina (Case: LC2 + UV)



m. Lombardy + Val Tartano + Upper Valtellina (Case: LC2 + VT+ UV)



n. Lombardy + ValChiavenna Case 2 (Case: LC2 + VCC2)



o. Lombardy + ValChiavenna Case 3 (Case: LC2 + VCC3)

Figure C.1: ROC curve (left) and PRC curve (right) of all models for different cases

D. Codes

All the codes used in this thesis can be found at https://github.com/ictar/master_thesis.