1   Assessing the importance of conditioning factor selection in Landslide

2   Susceptibility for Belluno province (Veneto Region, NE Italy)

4   Sansar Raj Meena[1,2] *, Silvia Puliero[1], Kushanav Bhuyan[1,2], Mario Floris[1], Filippo Catani[1]

6   [1] Department of Geosciences, University of Padova, Padova, Italy.

7   [2] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente,

8   Enschede, Netherlands.

10   * Corresponding author Email: sansarraj.meena@unipd.it

12   Abstract

13   In the domain of landslide risk science, landslide susceptibility mapping (LSM) is very

14   important as it helps spatially identify potential landslide-prone regions. This study used a

15   statistical ensemble model (Frequency Ratio and Evidence Belief Function) and two machine

16   learning (ML) models (Random Forest and XG-Boost) for LSM in the Belluno province

17   (Veneto Region, NE Italy). The study investigated the importance of the conditioning factors

18   in predicting landslide occurrences using the mentioned models. In this paper, we evaluated

19   the importance of the conditioning factors in the overall prediction capabilities of the statistical

20   and ML algorithms. By the trial-and-error method, we eliminated the least "important" features

21   by using a common threshold. Conclusively, we found that removing the least "important"

22   features does not impact the overall accuracy of the LSM for all three models. Based on the

23   results of our study, the most commonly available features, for example, the topographic

24   features, contributes to comparable results after removing the least "important" ones. This

25   confirms that the requirement for the important conditioning factor maps can be assessed based

on the physiography of the region. Based on the analysis of the three models, it was observed that most commonly available feature data can be useful for carrying out LSM at regional scale. Identifying LSMs at regional scale has implications for understanding landslide phenomena in the region and post-event recovery measures, planning disaster risk reduction, mitigation, and evaluating potentially affected areas.

1. Introduction

Landslides are one of the most frequently occurring natural disasters that cause significant human casualties and infrastructure destruction. Landslides are triggered by several natural and man-made events such as earthquakes, volcanic eruptions, heavy rains, extreme winds, and unsustainable construction activities such as unplanned settlement development and cutting of roads along the slopes (Glade et al., 2006;van Westen et al., 2008). Extreme meteorological events such as the Vaia storm of 2018 triggered landslides and debris flow, destroyed critical infrastructures in the northern parts of Italy (Boretto et al., 2021). As reported by (Gariano et al., 2021) in the last 50 years between 1969-2018, landslides posed a severe threat to the Italian population. Approximately, 1500 out of the 8100 municipalities in Italy have faced landslides with severe fatalities. Between the years of 1990 and 1999, 263 people were killed by landslides. Studies by (Rossi et al., 2019) estimated that approximately 2500 people were killed between 1945-1990. Moreover, predictive modelling of the Italian population at risk to landslides (Rossi et al., 2019) shows massive tendency of risk to the population with data acquired between 1861-2015, emphasizing the necessity of landslide risk studies.

Therefore, to assess landslide risk and plan for suitable risk mitigation measures, it is crucial to realize the significance of landslide studies, particularly Landslide susceptibility Mapping (LSM). LSM is an essential tool that incorporates the potential landslide locations (Senouci et al., 2021). The probability of a landslide occurring in a particular region owing to the effects

of several causative factors is referred to as landslide susceptibility. LSM is an essential step towards landslide risk management and helps in effective mapping of the spatial distribution of probable landslide manifestations (Dai et al., 2002). In the past, researchers have used a range of models to assess landslide susceptibility using technologies such as Earth Observation (EO) and Geographic Information Systems (GIS). The recognition and analysis of slope movements have been going on since the early 1970s (Brabb et al., 1972) and is still one of the most important components to perform LSM (Ercanoglu and Gokceoglu, 2002;Chacón et al., 2006;Guzzetti et al., 2006;Castellanos Abella and Van Westen, 2008;Floris et al., 2011;Catani et al., 2013;Pham et al., 2015;Reichenbach et al., 2018;Youssef and Pourghasemi, 2021;Liu et al., 2021).

Traditional methods such as the expert-based Analytical Hierarchy Process (AHP), multi-variate statistics, data-driven Frequency Ratio (FR) have been employed for landslide susceptibility for many years, with satisfactory results (Pradhan, 2010;Castellanos Abella and Van Westen, 2008;Komac, 2006). A use case of such approaches is given by Floris et al (2011) which apply traditional LSM methods (FR) for mapping landslide susceptibility in a case study in Veneto Region, Italy. Afterwards, with the development of new approaches, susceptibility modelling has advanced from traditional approaches. Presently, two approaches: (1) statistical and (2) machine learning, are practised for LSM at investigating the landslide predisposing factors and to map the geographical distribution of landslide processes. (Reichenbach et al., 2018) classified landslide susceptibility models into six main groups: (1) classical statistics, (2) index-based, (3) machine learning, (4) multi-criteria analysis, (5) neural networks, and (6) others. Research by (Reichenbach et al., 2018) also depicted that before 1995, only five models were used for LSM, but in recent times, an investigation of 19 other models was carried out, which yielded good results. More than 50 per cent of the methods consisting of the first five models mentioned above accounted for landslide susceptibility studies. Recent work of

76   (Stanley et al., 2021) emphasized the importance of data-driven methods in global LSM,

77   trained to report landslide spatial occurrences between the periods of 2015-2018. The first

78   version of the Landslide Hazard Assessment for Situational Awareness (LHASA) from their

79   work for NASA, reported landslide occurrences with a decision tree model that first defines

80   the intensity of one week of rainfall. LHASA version 2 used the data-driven model of XG-

81   Boost by adding two dynamically varying factors: snow and soil moisture. However, despite

82   advances in LSM, the importance of the conditioning factors in the prediction capability of a

83   model is not discussed enough. The need of increasing our control over the model sensitivity

84   to system parameters changes, including those induced by anthropogenic and climate-change

85   dynamics, is becoming a key factor in the implementation of truly efficient LSM for risk

86   mitigation purposes. The Vaia windstorm of 2018 (Forzieri et al, 2020), as a typical extreme

87   weather event, may easily escape traditional statistical prediction schemes and represent,

88   therefore, a challenging test for exploring the sensitivity of the various LSM models to

89   changing factors and conditions.

90   One goal of this research is to look into the relative changes in LSM accuracy when the least

91   "important" conditioning factors are removed. Feature selection in LSM is an approach in

92   reducing landslide conditioning factors to improve model performance and reduce

93   computational time. The purpose of this approach is to find the optimal set of conditioning

94   factors that will provide the best fit for the model to yield higher accuracy as predictions.

95   (Micheletti et al., 2014) emphasized the importance of feature selection in LSM and discussed

96   the use of Machine Learning (ML) models such as Support Vector Machine (SVM), Random

97   Forest (RF), and AdaBoost for LSM, as well as the significance of associated features within

98   the confluence of the ML models for feature importance. However, their study did not consider

99   geological and meteorological features like lithology, land use, and rainfall intensity for both

100   LSM and feature selection. Studies by (Liu et al., 2021) depicted the improvement in the

101 predictive capability of the so-called Feature Selected Machine Learning (FS-ML) model but

102 also remarked on the fact that the same conditioning factors may contribute differently in

103 different ML models. In this study, we want to investigate the prediction capability of the model

104 after removing conditioning factors as an approach to improve LSM accuracy in contrast to

105 what has been done in literature like (Liu et al., 2021), where they assess conditioning factor

106 importance using approaches like multi-collinearity analysis, variance inflation factor before

107 prediction of the susceptibility. The identification of the most crucial features can help in

108 monitoring the effect of extreme events (such as Vaia) on the changes in the evolution of

109 landslide hazard.

110 We present a study in the province of Belluno (Veneto Region, NE Italy) with the comparison

111 of the conditioning factor importance of statistical and ML models for LSM before the Vaia

112 storm event. The results from the LSM will be then validated using the IFFI landslide inventory

113 data for testing the various models' prediction capability with/without certain factors. We also

114 investigate whether many of the latter conditioning factors are crucial for LSM. As in many

115 regions over the world, the same data or factor maps might not be available.

116

117 2.   Study area and Data

118 2.1 Study area

119 The area of the Belluno Province (Veneto Region, NE Italy) is part of the tectonic unit of the

120 Southern Alps. The territory is 3,672 km² wide, stretching from north to south between the

121 Dolomite Alps and the Venetian Pre-Alps, with elevations ranging from 42 to 3325 m above

122 mean sea level. From a geological point of view, Dolomite Alps comprises the Hercynian

123 crystalline basement consisting of micaschists and phyllites intruded by the Permian

124 ignimbrites (Doglioni, 1990;Schönborn, 1999). These Paleozoic units are mainly outcropping

125 in the NE and central-West sectors. The Middle-Upper Triassic includes carbonate, volcanic

126   and dolomitic formations. In particular, the Upper Triassic Main Dolomite covers 14% of the

127   whole province. Jurassic-Cretaceous limestone and marls are especially located between the

128   Valsugana and Belluno thrusts (Sauro et al., 2013). Moreover, in the Belluno valley and in the

129   southern part of the area, Cenozoic sediments, i.e., flysch and molasse and Quaternary glacial,

130   alluvial and colluvial deposits are largely present. Instead, Venetian Prealps are characterized

131   by Jurassic-Cretaceous sedimentary cover, such as layered limestones and dolomites with

132   cherts (Compagnoni et al., 2005;Corò et al., 2015). Because of its morphological

133   characteristics, the study area is affected by slope instability, which overlay an area of 165 km²

134   corresponding to 6% of the province (Baglioni et al., 2006). Most of the landslides are located

135   in the NW (Upper basin of Cordevole River) and SE (Alpago district) sectors of the province

136   (Figure 1). The dominant landslide types are slides (47%), rapid flows (20%), slow flows

137   (12%), and shallow soil slips (7%) (Iadanza et al., 2021). The climate of the province of Belluno

138   is continental. The mean annual temperature recorded in the period 1961–1990 is 7°C and the

139   mean precipitation is 1284 mm/year (Desiato et al., 2005) with two peaks distributed in spring

140   and autumn. In the last 27 years, temperature and rainfall intensity in the study area have

141   increased due to climatic changes leading to more frequent meteorological conditions

142   (ARPAV, 2021).

143

144   2.2 Landslide inventory data

145   The inventory of landslide phenomena in Italy (IFFI) conducted by the Italian Institute for

146   Environmental Protection and Research (ISPRA) and the Regions and Autonomous

147   Provinces was used in this study (Trigila et al., 2010). The IFFI Project was financed in 1997.

148   Since 2005, the catalogue is available online and consists of point features indicating the

149   scarp of the landslides and polygon features delineating the instabilities. The archive stores

150   the main attributes of the landslides, such as morphometry, type of movement, rate, involved

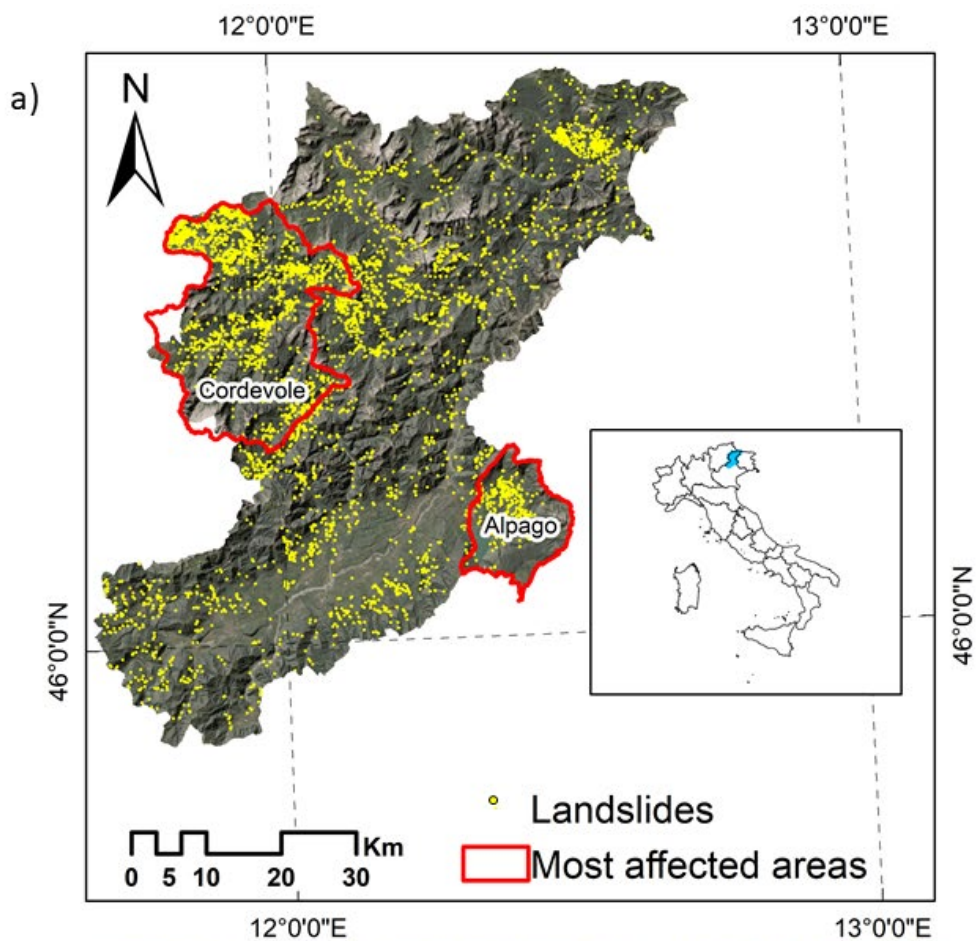
151 material, induced damages and mitigation measures. The inventory currently holds 620,808

152 landslides collected from historical documents, field surveys and aerial photointerpretation,

153 covering an area of 23,700 km^2, which corresponds to the 7.9% of the Italian territory (Trigila

154 and Iadanza, 2018). In the Belluno province, the IFFI inventory consists of 5934 points of

155 landslides occurred before 2006 (Baglioni et al., 2006).

159

160   2.3 Landslide conditioning factors

161   Based on the regional environmental characteristics of the study area and the scientific

162   literature, fourteen landslide conditioning factors were selected, including: (i) topographical

163   factors such as elevation, slope angle, slope aspect, topographical wetness index (TWI),

164   topographical position index (TPI), topographical roughness index (TRI), profile curvature,

165   and plan curvature; (ii) hydrological factors (i.e., distance to drainage, precipitation);

166   geological factors (lithology); (iii) anthropogenic factors (distance to roads); and (iv)

167   environmental factors like Normalized Difference Vegetation Index (NDVI) and landcover

168   (see figure 2). A freely accessible digital elevation model (DEM) with a spatial resolution of

169   25 metres and was downloaded from the Veneto Region cartographic portal

170   (https://idt2.regione.veneto.it), was used to derive the topographic layers. Refer to table 1 for a

171   detailed description of the conditioning factors. Land cover, lithology maps, road network and

172   drainage maps were downloaded from the same portal. Rainfall data was downloaded from the

173   Regional Agency for the Environmental Prevention and Protection of Veneto (ARPAV:

174   https://www.arpa.veneto.it/) web site. We resampled the conditioning factor maps to 25 meter

175   pixels in order to do the analysis.

176   Table 1: Description of the conditioning factors for landslide occurrences.

Sl No.	Conditioning Factor	Data Range	Description/Justification
1	Elevation	42 m to 3325 m	The geomorphological and geological processes are affected by elevation (Raja et al., 2017). It has an impact on topographic characteristics, which contribute to spatial differences in many landform processes, as well as the distribution of vegetation.

2	Slope	Flat areas to very high slopes till 86.48°	Slope is a derivative of the DEM which can cause failure of slope (Pham et al., 2018). Landforms having a higher angle of slope are usually more susceptible to collapse, which is closely correlated to landslides.
3	Aspect	North (0 degrees) to North (360 degrees)	Aspect has a correlation with other geo-environmental factors is a crucial factor for LSM that describes the slope direction (Dahal et al., 2008). The slope direction to a degree dictates the frequency of landslides.
4	Topographic wetness index	-2.12 to 20.06	The influence of topography on the location and amount of saturated runoff source areas is an essential conditioning factor (Pourghasemi et al., 2012). TWI measures the amount of accumulated water and distribution of soil moisture at a location. Higher TWI values can relate to higher chances of landslide occurrence.
5	Topographic Position Index	-1143.68 to 243.84	The topographic position index (TPI) shows the difference between the elevation of a point and its surrounding. Lower values represent the plausibility of features lower than the surrounding, thus possibly relating to higher odds of landslide occurrence.

6	Topographic Roughness Index	0 to 1077.30	Topographic Roughness Index (TRI) calculates the difference in elevation between adjacent pixels in a DEM which depicts the terrain fluctuation (Riley et al., 1999). As the slope of a landscape moves, the TRI decreases, relating to slope movement.
7	Profile Curvature	Concave Flat Convex	The driving and resisting forces within a landslide in the slope direction are affected by profile curvature.
8	Plan Curvature	Concave Flat Convex	The direction of landslide movement is controlled by the plan curvature, which regulates the convergence or divergence of landslide material (Dury, 1972;Meten et al., 2015).
9	Drainage	0 to 400	Drainage transports water, which induces material saturation, culminating in landslides in valleys. (Shahabi and Hashim, 2015).
10	Rainfall	84 to 1198.05 (mm/month)	Precipitation characteristics shift by climatic conditions and geographical characteristics, resulting in significant temporal and geographical variations in rainfall quantity and intensity. This can lead to the triggering of landslides across large areas but also for specific smaller areas.

11	Lithology	Volcanites, Pre-Permian, metamorphic, sequence Morainic, Gravels, etc.	The geological strength indices, failure susceptibility, and permeability of lithological units differ (Yalcin and Bulut 2006), where changes in the stress-strain behaviour of the rock strata can be caused by lithological unit variation. Slope failure typically occurs on a slope with low shear strength.
12	Distance to Roads	0 to 200	A crucial manmade element impacting the occurrence of landslides is roads because of road clear-cutting and construction activities (Dunning et al., 2009).
13	Landcover	Rock, Forest, Urban cover etc.	Because land cover may influence the hydrological functioning of slopes, rainfall partitioning, infiltration properties, and runoff, as well as the soil shear strength, different land cover types may affect slope stability.
14	NDVI	-0.66 to 0.66	NDVI is important in realizing the amount of vegetation cover which can be interpreted to understand the strength of the slope and the landslide occurrences. The NDVI reflects the inhibitory effect of landslide occurrence (Huang et al., 2020).

177

178

179

182

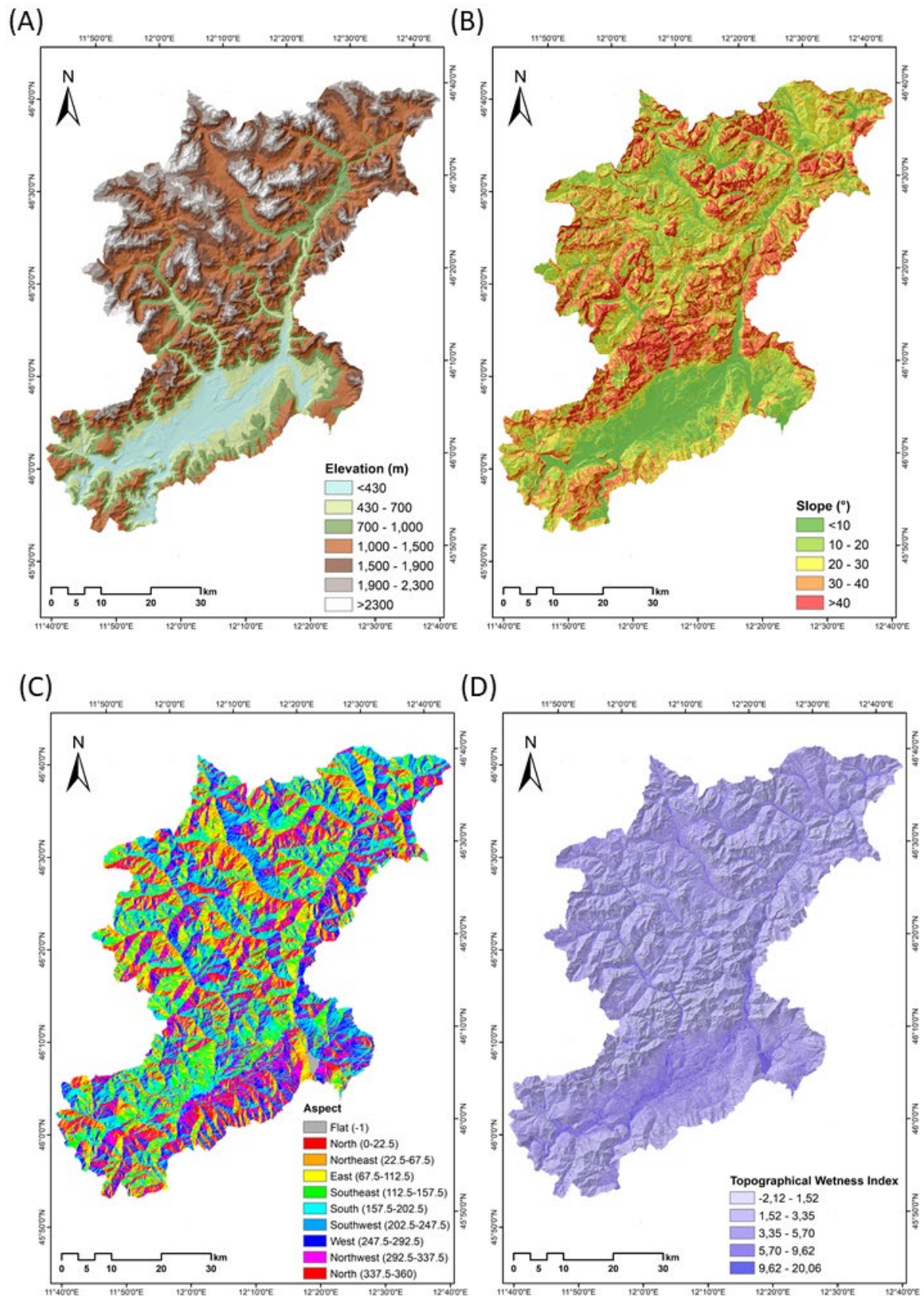
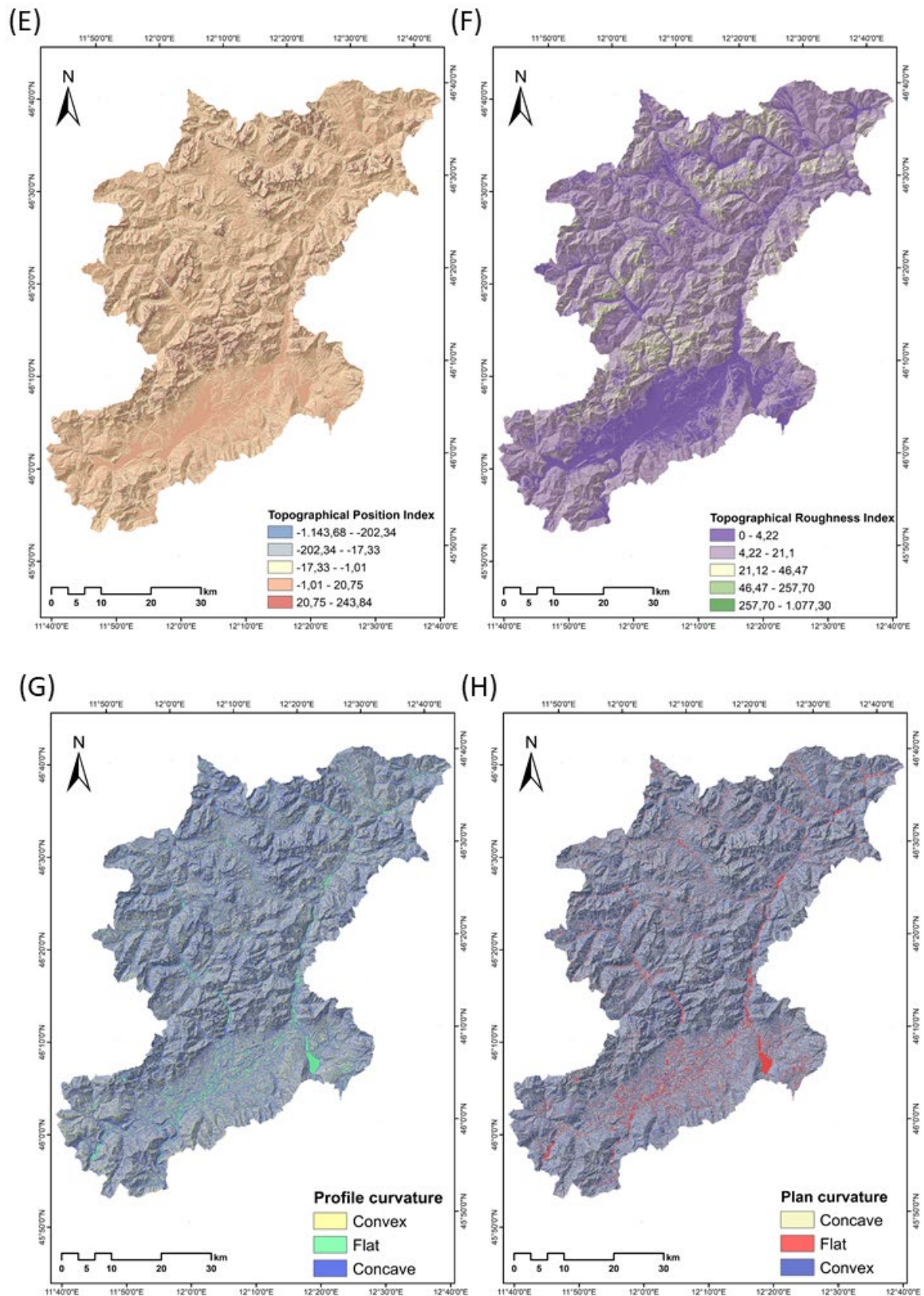
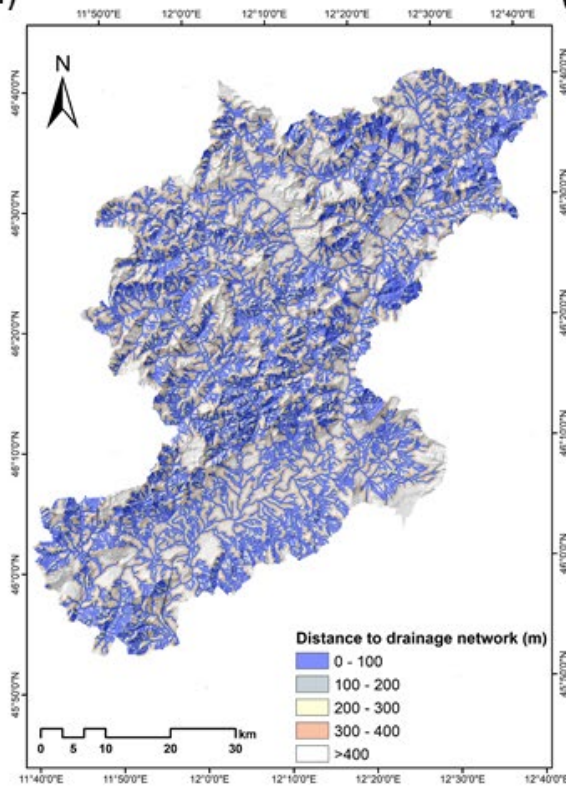
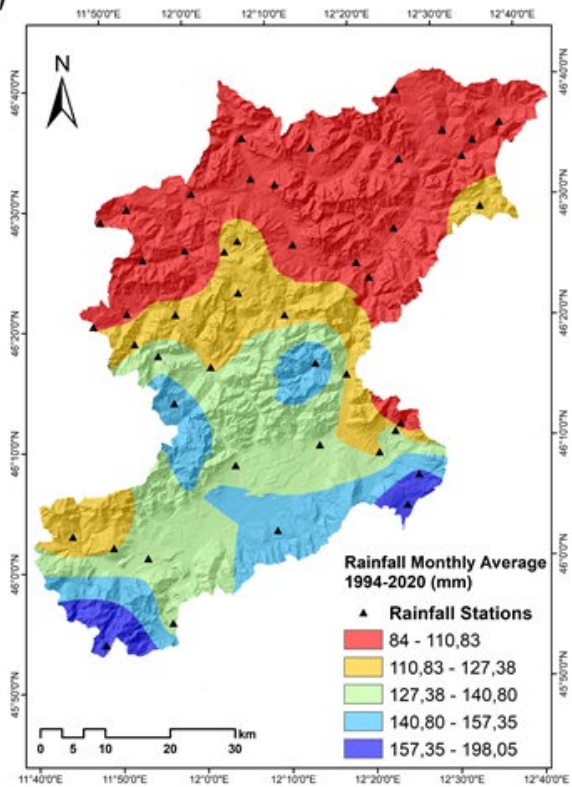
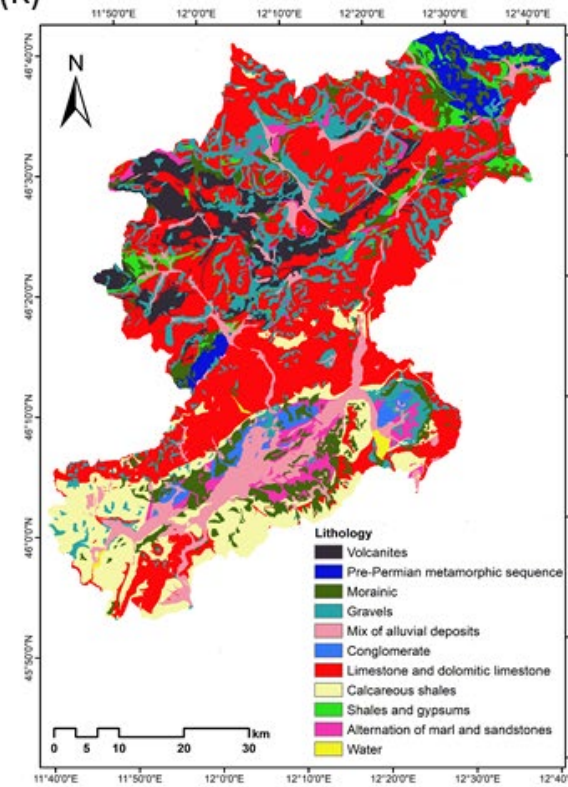
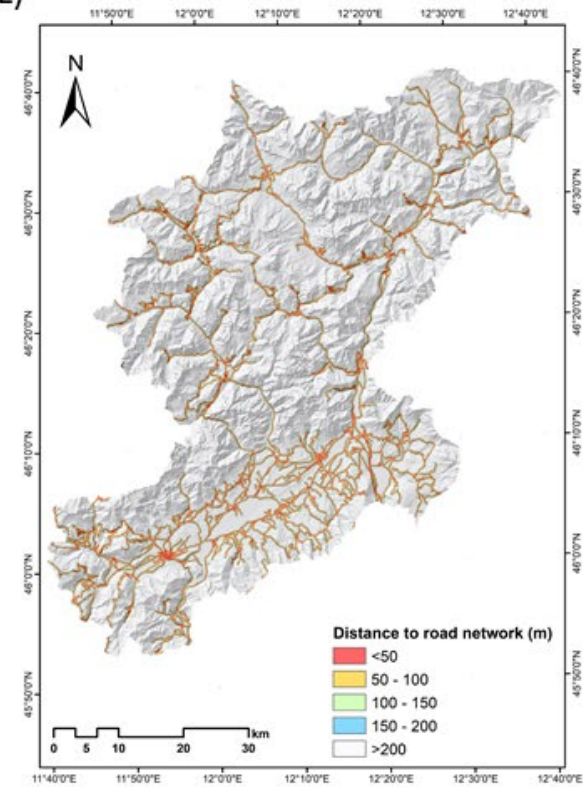
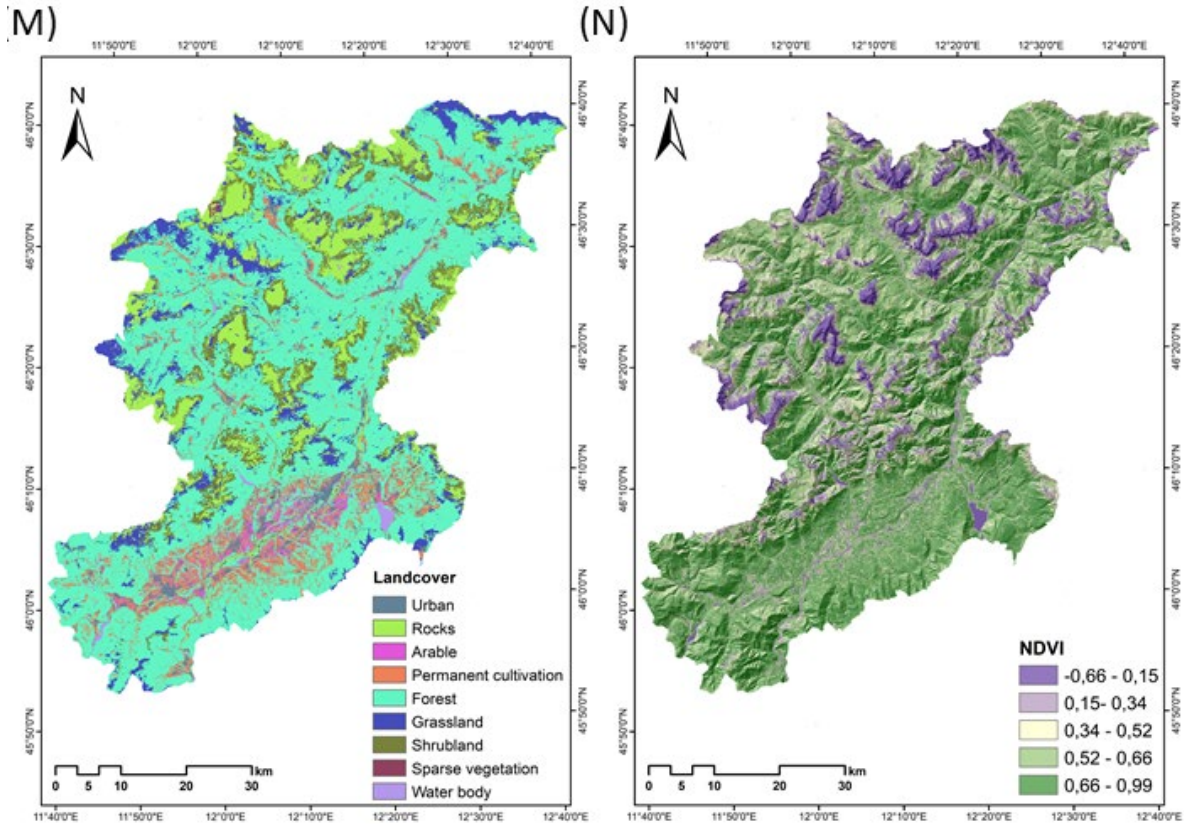



Figure 2: Maps of the conditioning factors used in this study: (A) Elevation, (B) Slope, (C)

Aspect, (D) Topographical wetness index, (E) Topographical position index, (F) Topographical

roughness index, (G) Profile curvature, (H) Plane curvature, (I) Distance to drainage networks,

(J) Rainfall monthly average (1994-2020) mm, (K) Lithology, (L) Distance to road network

(M) Landcover, (N) NDVI

188

3. Methodology

We propose an approach that helps assess importance of the conditioning factors, which can

help improve the susceptibility results by removing the less "important" factors throughout the

statistical and ML models. As stated previously, the study attempts the application of sensitivity

analysis to understand relative importance of the conditioning factors as a preliminary step

towards improving the landslide susceptibility prediction capability. In this study, the LSM

195   was obtained by the combination between IFFI landslide inventory and the conditioning factors

196   through statistical methods such as FR-EBF and ML models, i.e. Random Forest and XG-Boost

197   (Figure 3). The successive sub-sections address the definitions of the statistical and ML models

198   for LSM.

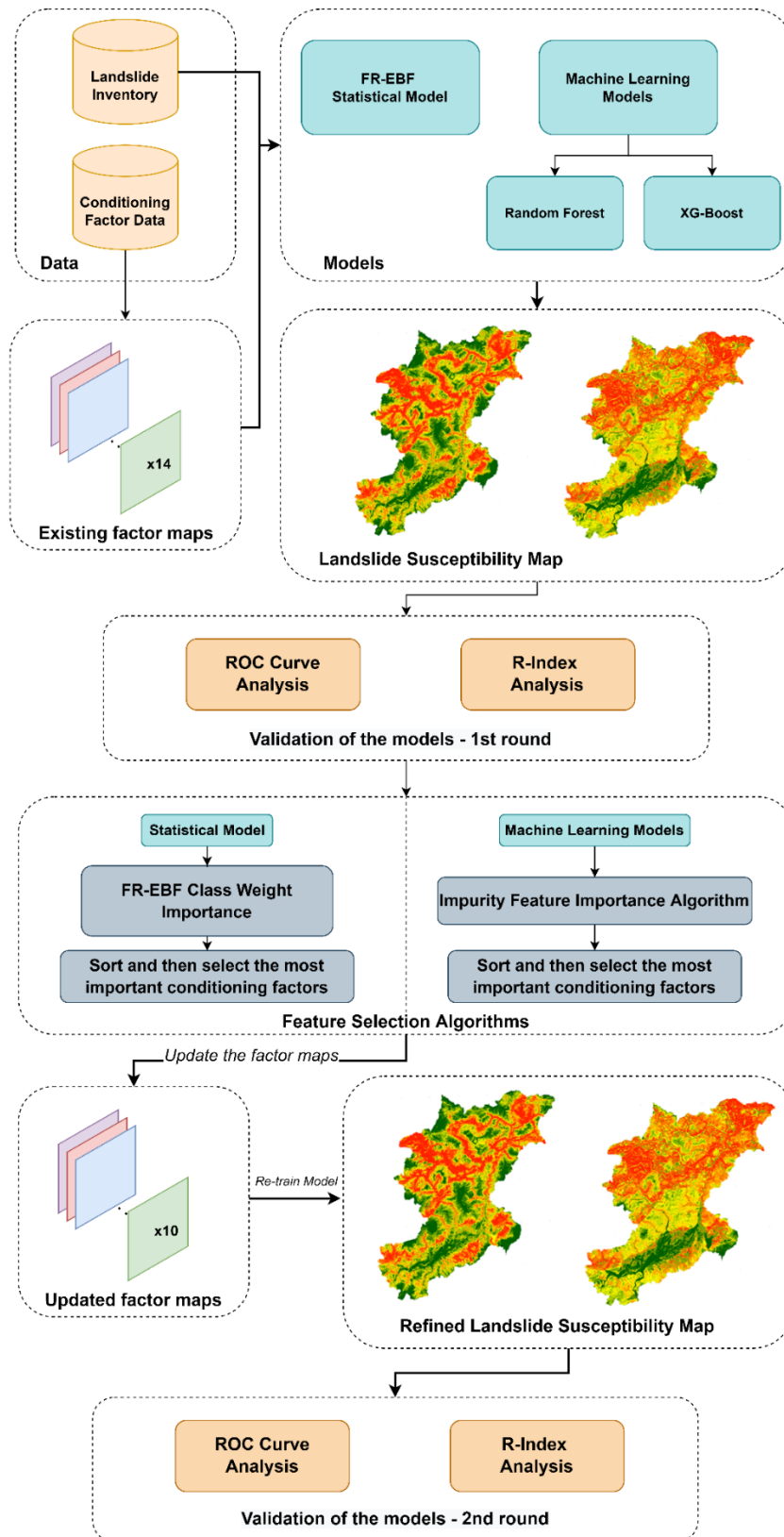



Figure 3: Overview of the conceptual workflow of methodology for landslide susceptibility assessment.

203 3.1 Statistical approach

204 3.1.1 Ensemble Frequency Ratio - Evidence Belief Function

205 In landslide susceptibility studies, the frequency ratio (FR) model is often applied. This is an

206 evaluation method which calculates the likelihood of landslide occurrence and non-occurrence

207 for each conditioning factor. (Lee, 2013;Mondal and Maiti, 2013;Shahabi et al., 2014). For

208 each landslide conditioning factor, the FR is a probabilistic model based on observed

209 correlations between landslide distribution and related parameters (Lea Tien Tay 2014). The

210 model depicts the relationship between spatial locations and the factors that determine the

211 occurrence of landslides in a specific area. Spatial phenomenon and factor classes correlation

212 can be found through FR and is very helpful for geospatial analysis (Mahalingam et al. 2016;

213 Meena et al. 2019b). Figure 3 gives an overview of the methodology employed in this study.

214 FR weights can be computed using the ratios of landslide inventory points of all classes within

215 each factor. The landslide inventory points are then overlaid with the conditioning factors to

216 obtain the area ratio for each factor class to the total area. The FR weights are then obtained by

217 dividing the landslide occurrence ratio in a class by the area in that class (Demir et al. 2012).

218

219 Using the Eq. 1, the Landslide Susceptibility Index (LSI) was computed by summing the values

220 of each factor ratio (Lee, 2013):

221

222 $$LSI = \sum FR \quad (Eq.1)$$

223

224 LSI= (DEM*wi)+(slope*wi)+(aspect*wi)+(Topographic Wetness Index*wi)+(Topographic

225 Roughness Index*wi)+(Topographic Position Index*wi)+(Distance to road*wi)+(Distance to

226 drainage*wi)+(Land Cover*wi)+(Lithology*wi)+(NDVI*wi)+(Rainfall*wi)+(Profile

227 Curvature*wi)+(Plain Curvature*wi)

228

229 Where LSI is the landslide susceptibility index, FR is the frequency ratio of every factor type

230 or class, and *wi* is the weight of each conditioning factor. The higher the LSI value, the higher

231 the susceptibility to landslides.

232 We integrated the LSI results with Evidence Belief Function (EBF) derived predictor values.

233 The EBF uses the conditioning factors defined by FR as the input data. Eq. (2) was applied to

234 the rating of every spatial factor.

235

236 $$PR = \frac{SAmax - SAmin}{SAmax - SAmin} min \ (Eq.2)$$

237 where SA is the indicator of Spatial Association between spatial factors and landslides,

238 whereas PR is the Prediction Rate. The lowest absolute difference of all factors is divided by

239 the computed absolute difference between the maximum and the least SA values (Table 2).

240 Pairwise comparison of the PR values of the slope failure predictors yielded the pairwise rating

241 matrix of the predictor rating. We used PR values for assigning weights of the factors for

242 susceptibility analysis.

243

244 3.2 Machine learning models

245 3.2.1 Random Forest model

246 Random Forest (RF) is based on the concept of the "wisdom of crowds" where multiple

247 decision trees, introduced by (Breiman, 2001), has been utilized in a number of remote sensing

248 research for a variety of applications (Melville et al., 2018). RF creates many deep decision

249 trees using the training data and it can overcome the overfitting problem mostly resulting from

250 complex datasets better than other decision trees. Each RF decision tree gives a prediction,

251 which is then weighted according to the value created from votes from each tree leading to

252 generation of the susceptibility map (see figure 4). Since the RF has shown an impressive

253 performance for classification purposes, it is regarded as one of the most efficient non-

254 parametric ensembles models (Chen et al., 2017). Based on the advantages listed above, the

255 RF model is used to assess landslide susceptibility. Landslide inventories along with the

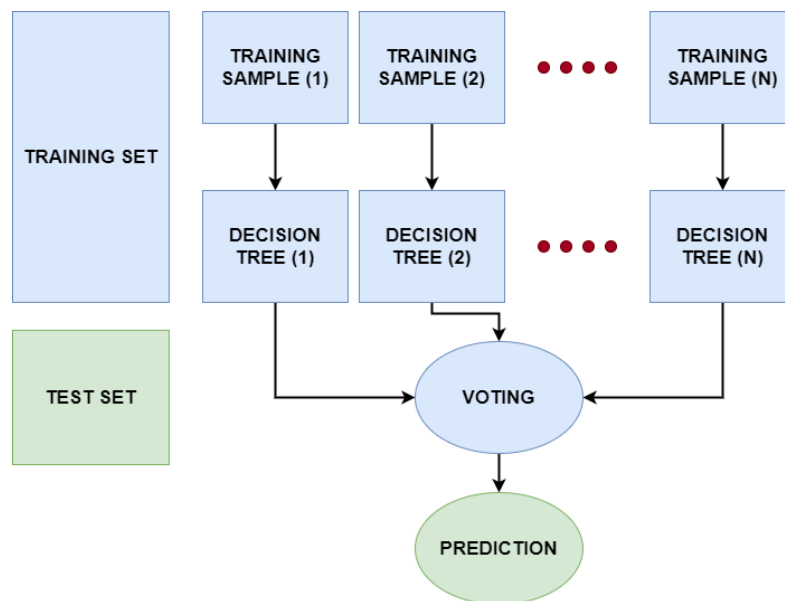
256 conditioning factors are divided into training and testing data as seen in figure 4. Using the

257 bagging technique, the training data is divided into training subsets, generally about one-third

258 of the total training samples. A decision tree is created for each subset based on the training

259 subset defined in the first stage and accordingly, votes as implemented that outputs the

260 landslide susceptibility.

262

263   Figure 4: Conceptual diagram of the Random Forest model.

264

265 3.2.2 XG-Boost model

266 Extreme gradient boosting or commonly known as the XG-Boost ML model is an optimized

267 gradient boosting algorithm that is designed for optimum speed and performance and boosting

268 ensembles are used to generate a prediction model (Sahin, 2020). The core idea of a boosting

269 algorithm is to combine the weaker learners to improve accuracy (Can et al., 2021), meaning

270 that different models with lower susceptibility accuracies are "boosted" by combining them to

271 achieve an ensembled higher susceptibility accuracy. The model is known for its fast-training

272 speed for classification tasks. In the study, we use training parameters to adjust the XG-Boost

273 algorithm like learning rate, subsample ratio, maximum depth of the tree and others. It uses

274 boosting techniques to reduce overfitting problems to improve accuracy results (figure 5). The

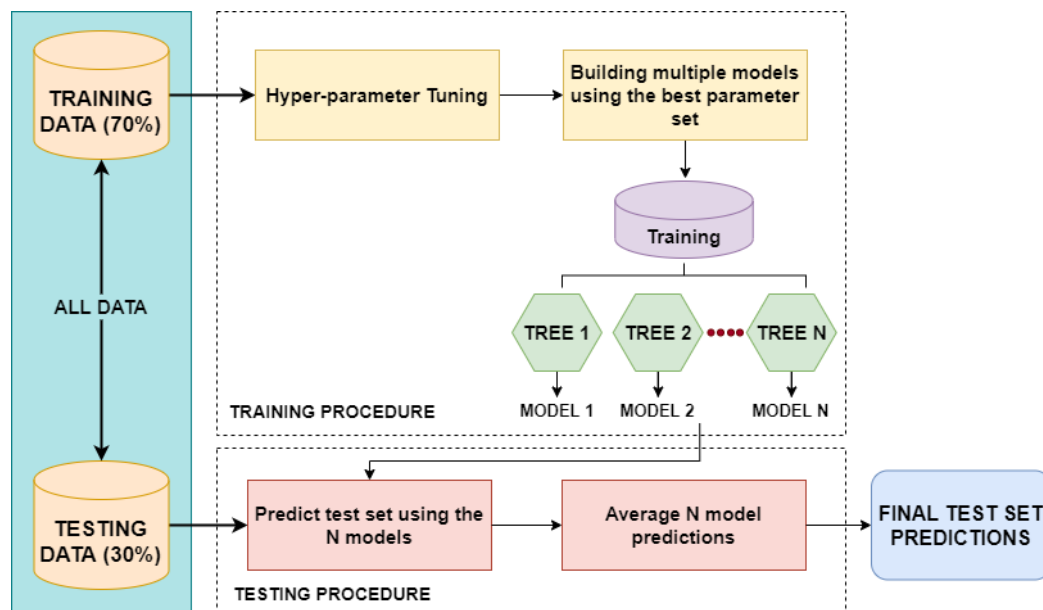
275 training data is divided into subsets which are then trained using a tree ensemble model. This

276 means that every weight derived from each model training of landslide instances in the area are

277 added and then predicted on the test set with the average landslide susceptibility scores of the

278 ensemble models.

279

281

282 Figure 5: Training and testing procedure of the XG-Boost model.

283 3.3 Feature selection algorithms

284 The goal of feature selection is to remove the least important conditioning factors in order to

285 increase the generalisability in landslide prediction. This selection help eliminates the

286 irrelevant (less important) conditioning factors to obtain optimal prediction accuracy

287 (Micheletti et al., 2014). For the statistical model, we used class weights obtained from

288 frequency ratio and used them as input for generating predictor rate from FR-EBF model which

289 gives the final weights of the conditioning factors. So, we used the predictor rate weights to

290 select the suitable features.

291 In terms of the feature importance for selecting the right set factors for both RF and XG-Boost,

292 we use the in-built impurity feature importance algorithm which is performed on the training

293 set (refer to feature selection in figure 3). Based on the results of the feature selection

294 algorithms for the conditioning factors for each model, the most important factors will be

295 selected to investigate the improvement of model performance. With this, we can understand

296 which of the conditioning factors played the most important roles in giving the highest accuracy

297 for each ML model.

298 4.   Results

299 4.1 Statistical model

300 The class weights were derived from data driven FR model and the final weights of the factors

301 were derived by using predictor rate from evidence belief function given in Table 2. The class

302 and factor weights were calculated using equations 1 and 2. The final weights of landslide

303 conditioning factors were calculated using an ensemble of FR-EBF, and then utilised to create

304 the final LSM. Because there is no common approach for identifying landslide susceptibility

305 classes in the final LSM, we normalised the findings to 0 to 100 for uniformity and

306 comparability. Using a natural breaks classification, which separates the values into groups

307 with random number of values, the resultant LSM was classified into five classes: very low,

308 low, moderate, high, and very high, as shown in figure 7 (Chung and Fabbri, 2003). This

309 method of classification gives a better distribution of values in each class than common

310 approaches such as natural breaks, which can result in certain classes having limited or

311 excessive data.

312 In terms of the feature importance that we observe in figure 6 and Table 2 (normalized weights),

313 based on the trial-and-error approach, factors (or features) under the threshold of 0.3 were

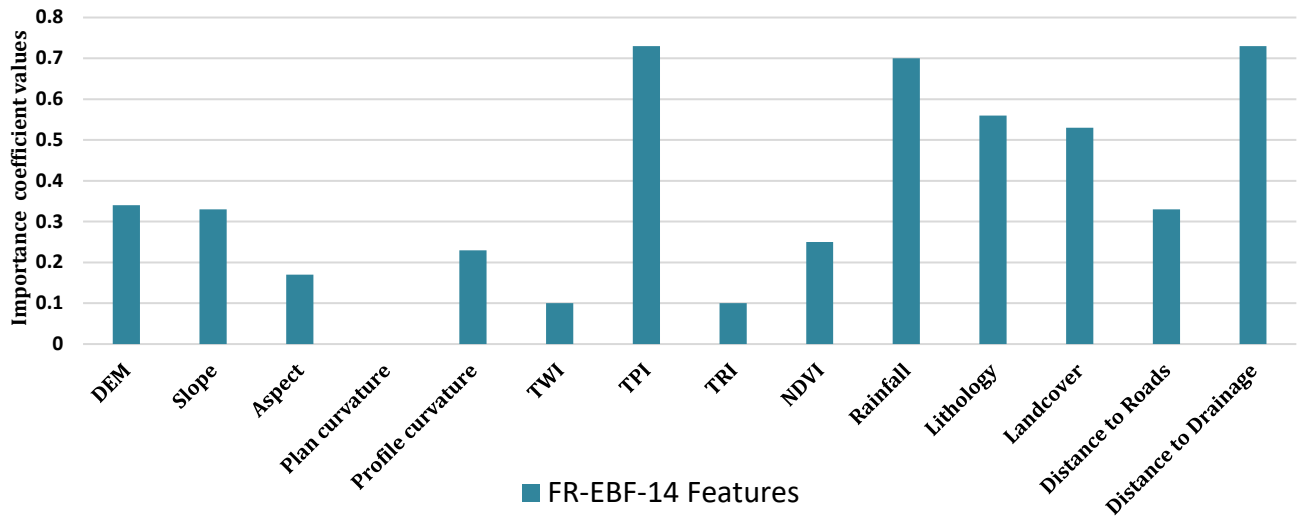
314 discarded as they did not make much of a difference in terms of predicting landslide

315 occurrences in the study area. Therefore, five conditioning factors having coefficient values

316 lower than 0.30 were dropped and overall, the area under the curve (AUC) accuracy still

317 remained similar to the original accuracy with the 14 factors.

318

319

320 Figure 6: Feature importance of the statistical model

321

322 Table 2: Frequency ratio values for spatial factors class weighting and EBF coefficients for

323 predictor rates (PR) based on degrees of spatial associations.

Factors and classes	Bel	Min	Max	[Max-Min]	Predictor Rate	FR Weights	Normalized weights
Elevation		0.07	0.24	0.17	0.73		
<430	0.07					0.50	0.06
430 - 700	0.15					1.13	0.20
700 - 1000	0.13					0.96	0.19
1000 - 1500	0.12					0.86	0.15
1500 - 1900	0.11					0.81	0.12
1900 - 2300	0.24					1.72	0.17
>2300	0.18					1.31	0.12

Profile Curvature		0.00	0.53	0.53	2.30		
Concave	0.53					1.05	0.40
Flat	0.00					0.00	0.30
Convex	0.47					0.95	0.30
Plan Curvature		0.00	0.52	0.52	2.26		
Concave	0.52					1.03	0.35
Flat	0.00					0.00	0.33
Convex	0.48					0.97	0.32
Slope		0.14	0.25	0.11	0.48		
<10	0.14					0.70	0.14
10 - 20	0.23					1.11	0.22
20 - 30	0.25					1.25	0.27
30 - 40	0.20					0.99	0.20
>40	0.17					0.86	0.17
Distance from drainage		0.02	0.36	0.34	1.49		
0 - 100	0.36					1.15	0.28
100 - 200	0.30					0.97	0.19
200 - 300	0.23					0.74	0.12
300 - 400	0.10					0.31	0.07
>400	0.02					0.06	0.34
Distance from roads		0.08	0.24	0.15	0.67		
0 - 50	0.36					1.15	0.27
50 - 100	0.30					0.97	0.19
100 - 150	0.23					0.74	0.17
150 - 200	0.10					0.31	0.16
>200	0.02					0.06	0.13

Landcover		0.01	0.24	0.23	2.98		
Urban	0.17					1.48	0.17
Rocks	0.10					0.90	0.09
Arable	0.01					0.07	0.01
Permanent cultivation	0.10					0.92	0.13
Forest	0.11					0.95	0.11
Grassland	0.24					2.11	0.14
Shrubland	0.04					0.37	0.04
Sparse vegetation	0.12					1.08	0.21
Water body	0.12					1.05	0.09
TWI		0.17	0.25	0.08	1.00		
-2.12 - 1.52	0.19					1.01	0.20
1.52 - 3.35	0.20					1.04	0.20
3.35 - 5.70	0.18					0.92	0.18
5.70 - 9.62	0.17					0.90	0.18
9.62 - 20.06	0.25					1.30	0.24
TPI		0.00	0.31	0.31	1.35		
-1143.68 - -202.34	0.00					0.00	0.00
-202.34 - -17.33	0.18					0.74	0.21
-17.33 - -1.01	0.26					1.06	0.27
-1.01 - 20.75	0.24					0.98	0.26
20.75 - 243.84	0.31					1.24	0.27
TRI		0.00	0.34	0.34	1.47		
0 - 4.22	0.22					0.73	0.23
4.22 - 21.1	0.34					1.11	0.35
21.12 - 46.47	0.25					0.82	0.22

46.47 - 257.70	0.20					0.65	0.20
257.70 - 1077.30	0.00					0.00	0.00
Rainfall intensity		0.00	0.81	0.81	3.54		
84 - 110.83	0.81					11.29	0.32
110.83 - 127.38	0.08					1.15	0.27
127.38 - 140.80	0.05					0.70	0.15
140.80 - 157.35	0.06					0.81	0.19
157.35 - 198.05	0.00					0.00	0.06
NDVI		0.14	0.25	0.11	0.48		
-0.66 - 0.15	0.14					0.70	0.13
0.15 - 0.34	0.22					1.13	0.21
0.34 - 0.52	0.25					1.26	0.25
0.52 - 0.66	0.21					1.07	0.21
0.66 - 0.99	0.18					0.89	0.20
Aspect		0.05	0.15	0.09	0.41		
Flat (-1)	0.11					1.02	0.10
North (0-22.5)	0.08					0.75	0.07
Northeast (22.5-67.5)	0.09					0.84	0.09
East (67.5-112.5)	0.11					1.08	0.11
Southeast (112.5-157.5)	0.14					1.31	0.14

South (157.5-202.5)	0.15					1.40	0.14
Southwest (202.5-247.5)	0.14					1.33	0.14
West (247.5-292.5)	0.08					0.76	0.09
Northwest (292.5-337.5)	0.05					0.50	0.07
North (337.5-360)	0.06					0.58	0.06
Lithology		0.04	0.26	0.22	2.84		
Volcanites	0.26					3.45	0.16
Pre-Permian metamorphic sequence	0.11					1.50	0.11
Morainic	0.06					0.85	0.15
Gravels	0.04					0.52	0.04
Mix of alluvial deposits	0.05					0.70	0.03
Conglomerates	0.21					2.84	0.21
Limestone and dolomitic limestone	0.13					1.76	0.16
Calcareous shales	0.08					1.04	0.08
Shales and gypsums	0.06					0.76	0.07

Alternation of marls and sandstones	0.07		0.91	0.06
Water body	0.22		2.97	0.00

324

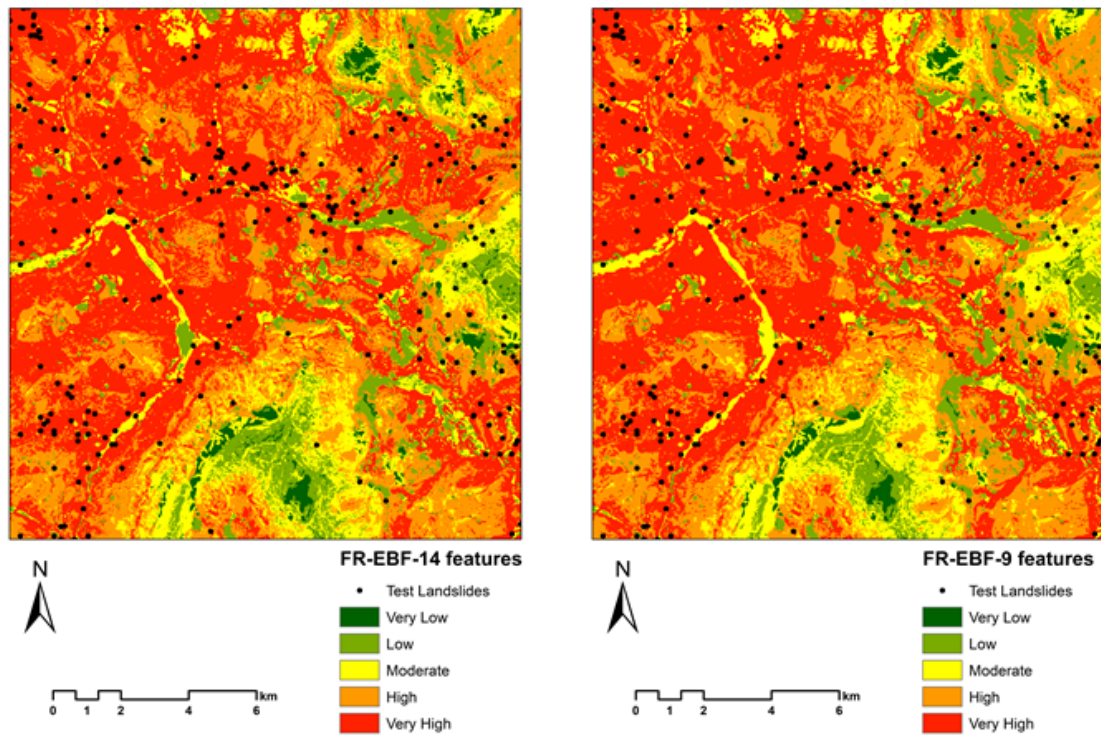
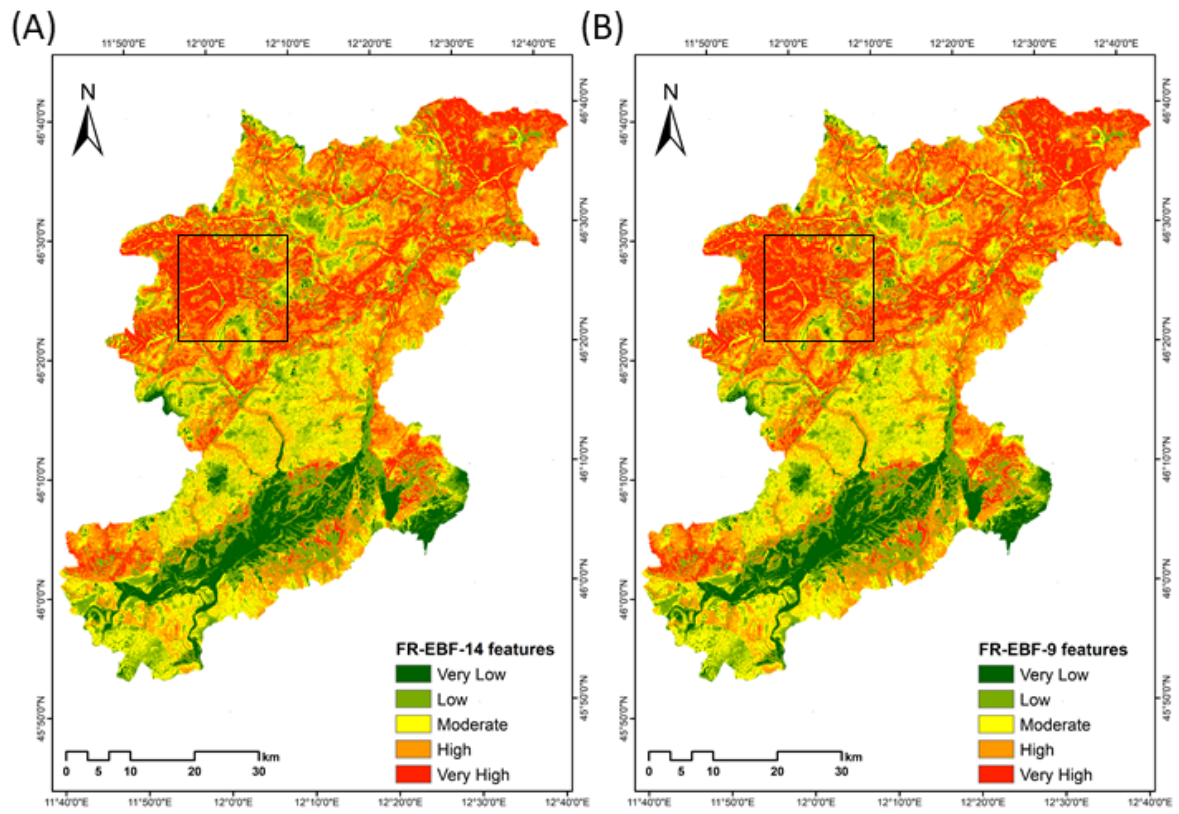



Figure 7: Landslide susceptibility maps derived using the ensemble of FR-EBF approaches for (A) 14 landslide features and (B) 9 landslide features (Black square represents the enlarged area).

4.2 Machine learning models

The LSM was generated based on the conditioning factor data, where the model learnt the information from the feature maps, which helped identify areas of susceptibility. The final results of the ML models in generating the LSM are given in Table 3. We observe that the AUC scores of RF are not much apart from the XG-Boost model, indicating similar predictive skills of both the modelsVisually the results show more susceptible areas near the landslide features (figures 8 and 9).

The model performance in terms of the accuracy of AUC is relatively similar to the results after eliminating the lower degree of feature importance for both RF and XG-Boost. As discussed previously in section 3.3, the feature importance for the ML models is carried out using the impurity feature importance algorithm that enables to assess the relative relevance of the conditioning factors in the optimal prediction of the landslides in terms of accuracy. As seen in figure 10, the factors of Landcover, Profile Curvature, Plan Curvature, TWI and TPI have the lowest values for the RF model. We examined various values as a cut-off for choosing the "important" conditioning factors and after much trial-and-error, a value of 0.03 was chosen as the threshold. Any factors above this value were considered as "important" factors for landslide susceptibility, hence, in figure 8, we see that the five factors mentioned above are removed and giving us 0.906 AUC as accuracy, which is better in AUC accuracy without removing the five factors (0.902 AUC as seen in Table 3).

Similarly, the same was repeated for the XG-Boost ML model and referring to Table 3, and despite removing the lower valued conditioning factors of Profile Curvature, TPI, and Plan

351 Curvature, the AUC accuracy score was similar (Table 3). We observe that Slope and Distance

352 to Roads had a much bigger impact on the RF mode than the XG-Boost model. On the other

353 hand, Lithology played a bigger role in estimating landslide occurrences in the XG-Boost

354 model. These observations indicate interesting results which will be discussed further in the

355 discussion section.

356

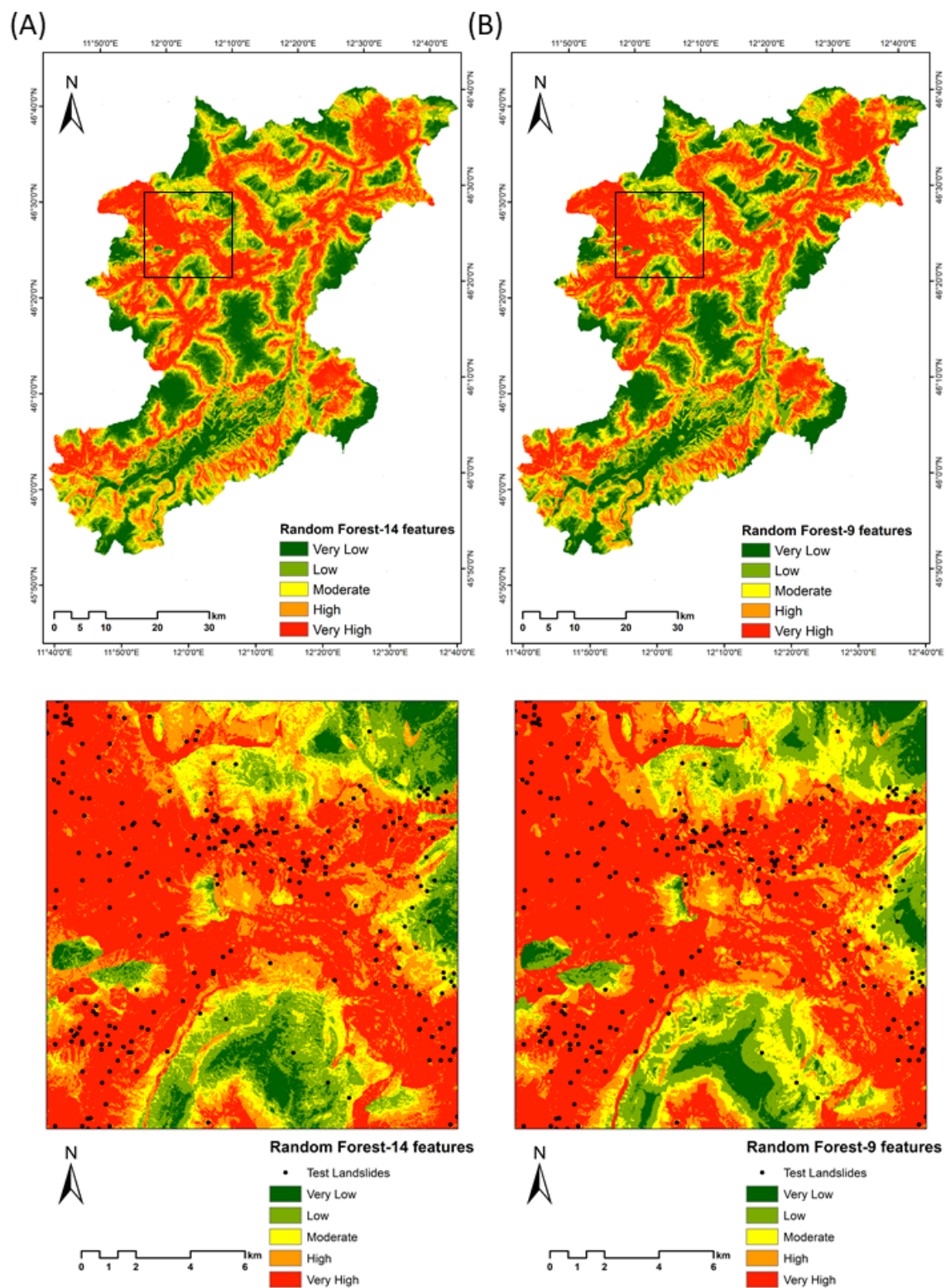
357 Table 3: Overall table with AUC results for landslide susceptibility of Belluno.

358

No.	Model	AUC
1	FR-EBF 14 features	0.836
2	FR-EBF 9 features	0.834
3	RF 14 features	0.902
4	RF 9 features	0.906
5	XG-Boost 14 features	0.910
6	XG-Boost 10 features	0.907

359

360

362

363   Figure 8: LSMs derived using the Random Forest approach for (A) 14 landslide features and

364   (B) 9 landslide features (Black square represents the enlarged area).

365

366

368

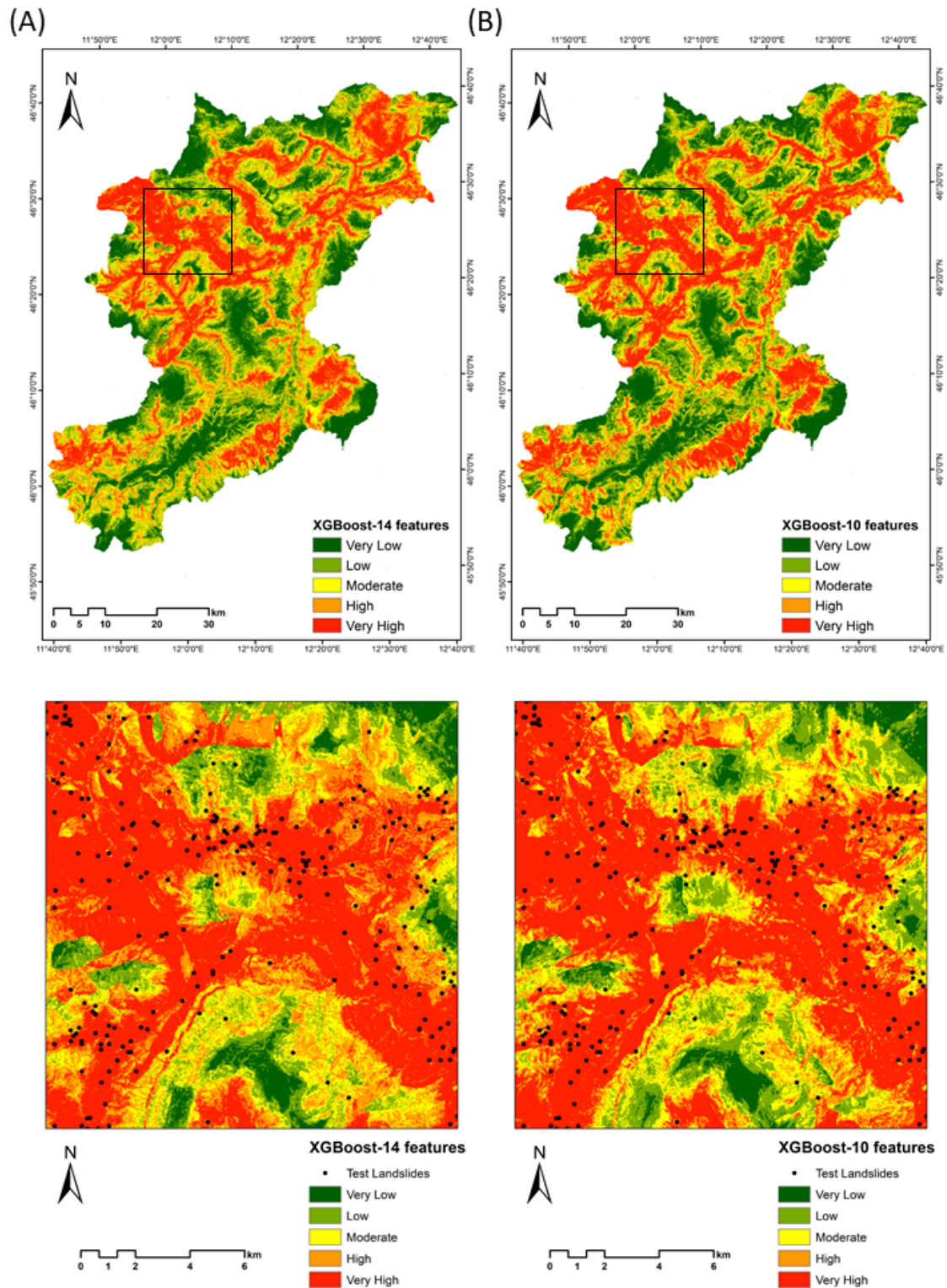



Figure 9: LSMs derived using the XG-Boost approach for (A) 14 landslide features and (B) 9

landslide features (Black square represents the enlarged area).

371

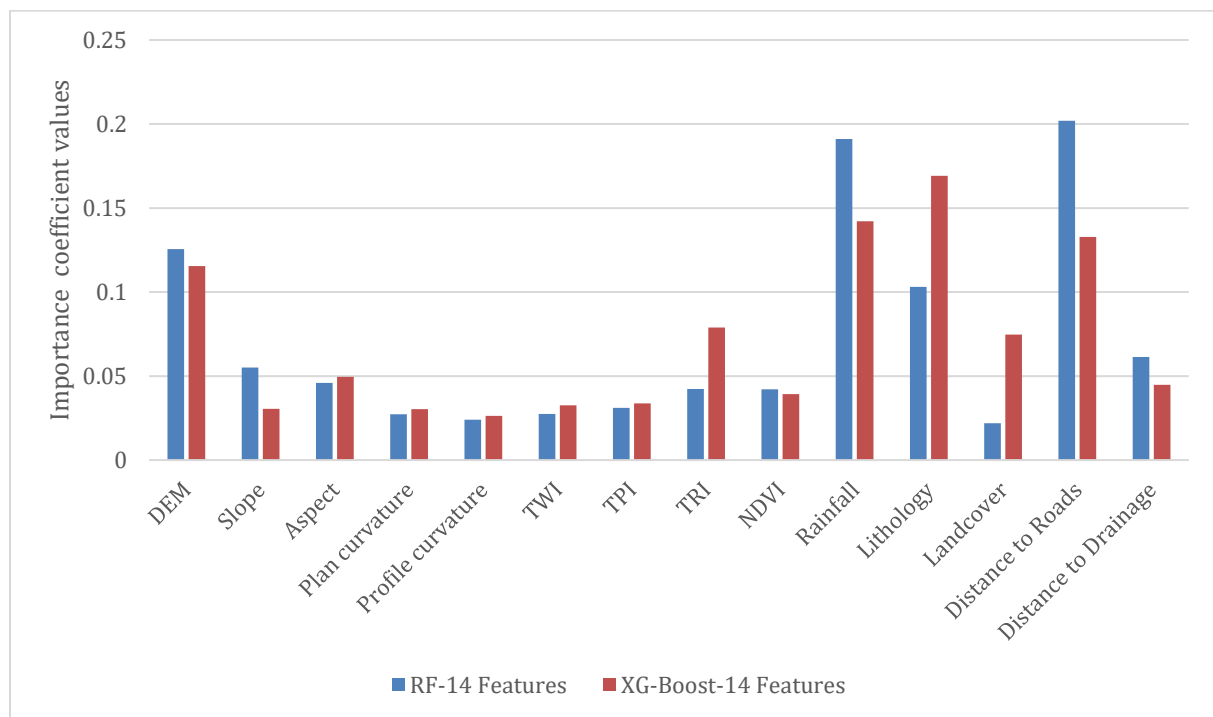



Figure 10: Feature importance of the RF and XG-Boost models.

5. Accuracy Assessmen

Accuracy assessment is crucial in producing quality LSMs for natural hazards where the information presented in the map is beneficial for planners (Goetz et al., 2015) A number of accuracy assessment approaches may be used to assess the quality of the LSMs. We compare the landslide inventory data to the resultant maps derived using the ensemble of FR-EBF, machine learning RF and XG-Boost models. The efficiency of any model for LSM is calculated by comparing the inventory data to the produced maps. This reflects if the models in use can accurately forecast which areas are susceptible to landslides (Pourghasemi et al., 2018). The findings from the total landslide input events were tested using 30% of the landslide occurrences. Testing for this study was done using the Receiver Operating Characteristics (ROC) and the Relative Landslide Density (R-Index) approaches.

387

5.1 Receiver Operating Characteristics (ROC)

The test dataset was used to corroborate the six resultant LSMs from statistical and machine learning using the receiver operating characteristics (ROC) approach. The ROC approach shows how to evaluate the true positive rate (TPR) and false positive rate (FPR) in the LSMs (Ghorbanzadeh et al., 2018; Linden, 2006). TPRs are pixels that are correctly labeled as high susceptibility in the landslide validation data, whereas FPRs are pixels that are incorrectly labeled. ROC curves are created using TPRs versus FPRs. The accuracy of the generated LSMs is determined by the AUC. The AUC shows whether there were more correctly labeled pixels than incorrectly labeled pixels. Greater AUC values suggest a more accurate susceptibility map, and vice versa. The susceptibility map is meaningful if the AUC values are close to unity or one. A map with a value of 0.5 is considered insignificant since it was created by chance. (Baird, 2013).

Figure 11 shows the accuracy values obtained using the ROC technique for the statistical approaches of FR-EBF and machine learning approaches of RF and XG-Boost. XG-Boost shows the highest accurate results with an AUC value of 0.91 and RF with 0.906, and FR-EBF with 0.836 (refer to Table 3). These results are quite good as it is closer to unity or one. The ensemble of FR-EBF shows lower AUC values than the machine learning-based XG-Boost and Random Forest. Machine learning results may differ because the models used landslide and non-landslide features as training data, whereas FR-EBF results are derived solely from landslide data. The results may differ depending on the geographical location and the selection of landslide conditioning factors.

409

410

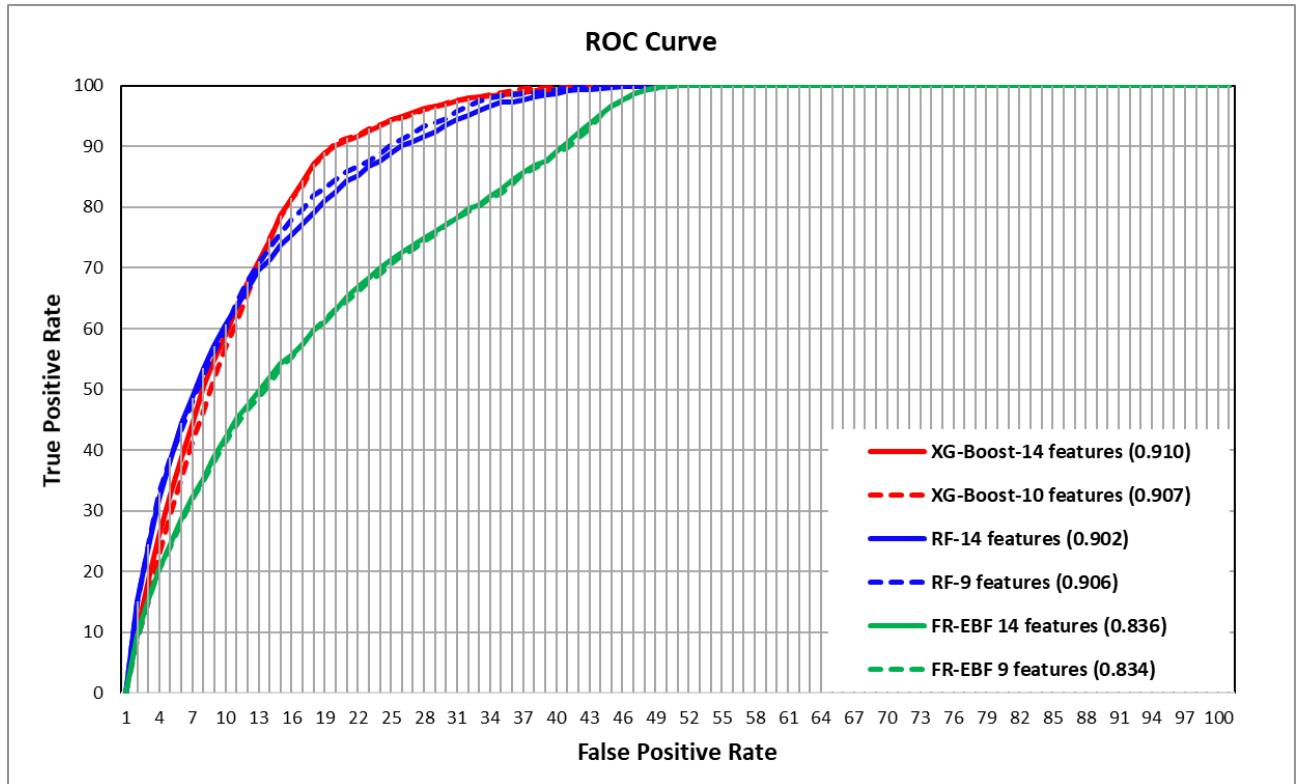



Figure 11. Testing for the performance of the statistical and machine learning models for

LSM in Belluno province, Italy.

5.2 Relative Landslide Density (R-Index)

The relative landslide density index was also used to assess the accuracy of the LSMs (R-index). Equation (4) is used to get the R-index:

$$R = (ni/Ni)/\Sigma(ni/Ni)) \times 100 \text{ (Eq.4)}$$

where Ni is the percentage of landslides in each susceptibility class and ni is the percentage of area susceptible to landslides in each susceptibility class Table 4 shows the quantile classification approach to classify the six landslide susceptibility maps into five susceptible classes. In comparison to the RF and FR-EBF models, the XG-Boost model with 14 and 10 features has a higher R-index for very high susceptibility classes. The R-index findings show

424   that FR EBF has a better R-index value for high susceptibility class than XG-Boost, which has

425   the lowest R-index for high susceptibility class. FR-EBF has a higher r-index value for the high

426   susceptibility class than the other three approaches. In addition, the R-index of FR-EBF is

427   higher for the very low susceptible class. Table 4 shows the R-index values for susceptibility

428   class in FR-EBF, RF, and XG-Boost, as well as plots of the same in figure 12.

429

430   Table 4: R-indices for the FR-EBF, RF, and XG-Boost models' landslide susceptibility

431   mappings (LSMs).

Validation methods	Susceptibility class	Number of pixels	Area (m²)	Area (%) (ni)	Number of landslides	Landslide (%) (Ni)	R- index
FR-EBF-14 Features	Very Low	21875	334248750	9.28	48	2.71	6
	Low	90000	570760000	15.85	171	9.66	13
	Moderate	165000	896709375	24.90	308	17.40	15
	High	263750	1026578125	28.50	460	25.99	20
	Very High	444375	773585000	21.48	783	44.24	45
FR-EBF-9 Features	Very Low	19375	323332500	8.98	38	2.15	5
	Low	91875	541371875	15.03	179	10.11	15
	Moderate	153125	894758125	24.84	289	16.33	15
	High	276875	1041846875	28.93	480	27.12	21
	Very High	443750	800571875	22.23	784	44.29	44
RF-14 Features	Very Low	6875	682346250	18.94	11	0.62	1
	Low	34375	658375000	18.28	55	3.11	4
	Moderate	75625	619031875	17.19	122	6.89	9

Model	Class						
	High	159375	749470625	20.81	264	14.92	17
	Very high	712500	892657500	24.78	1318	74.46	69
RF-9 Features	Very Low	7500	735246875	20.41	12	0.68	1
	Low	30000	632679375	17.57	48	2.71	4
	Moderate	75000	581844375	16.15	120	6.78	10
	High	147500	692276250	19.22	245	13.84	17
	Very High	729375	959834375	26.65	1345	75.99	68
XG-Boost-14 Features	Very Low	11250	1076978750	29.90	18	1.02	1
	Low	6875	330045625	9.16	11	0.62	3
	Moderate	11875	278243750	7.72	19	1.07	5
	High	11250	352568125	9.79	18	1.02	4
	Very High	947500	1564045000	43.42	1704	96.27	87
	Very Low	12500	1094226250	30.38	20	1.13	1
	Low	7500	297782500	8.27	12	0.68	3
XG-Boost-10 Features	Moderate	8125	242914375	6.74	13	0.73	4
	High	15625	314181875	8.72	25	1.41	7
	Very High	945000	1652776250	45.89	1700	96.05	84

432

6. Discussion

433 Landslides are very dynamic in nature, meaning that their behaviour, movement, and spatial

435 distribution changes over space and time. Therefore, it is important to analyse the significance

436 of the conditioning factors that lead to landslide occurrences. The relevance of the conditioning

437 features for LSM is essential to realize which of the features had the biggest impact on the

438 prediction of landslide occurrences. As not all conditioning factor maps be available globally,

439 or sometimes even locally, due to reasons such as non-compliance in sharing data, data

440 unavailability, erroneous data structure, and others, it can be worthwhile to understand which

441 of the available conditioning factors play an important role in LSM. For example, topographical

442 features derived from digital elevation models such as Elevation, Slope, aspect, Plan curvature,

443 Profile curvature, TWI, TPI, TRI are available almost globally because of missions such as the

444 Shuttle Radar Topography Mission (SRTM). Other features, such as distance to roads and

445 drainage networks, that might have direct or indirect influence on the occurrence of landslides,

446 can also be easily accessed through numerous open-source platforms. However, conditioning

447 factor maps of rainfall data derived from rain gauge stations are not easily accessible and

448 available. In this study, we used fourteen features for landslide susceptibility assessment and

449 carried out the feature importance of the conditioning factors for traditional statistical ensemble

450 model of FR-EBF and machine learning models of RF and XG-Boost. The feature selection

451 approach from statistical model is dependent upon the landslide data and its relation to each

452 feature and their classes. On the other hand, feature selection for machine learning models

453 depends upon the landslide and non-landslide samples that are used to train the models. We

454 used the in-built impurity feature importance algorithm to assess the importance of the features

455 during the model training phases. Based on literature review for this sort of study, there is no

456 standard threshold values available for discarding or selection of features for LSM. In this

457 study, we used a trial-and-error approach to determine a threshold of 0.30 for the selection of

458 conditioning factors used for landslide susceptibility for all the three models.

459 Feature importance algorithms used in this study are different, however there is similarity in

460 the importance of the features in both statistical and machine learning algorithms (See figure 6

461 and 10). As we look at the figures 7, 8, and 9 in the enlarged region, we observe that there are

462 not many differences despite removing the least important features. The reason for such

463 observation can be linked to the lower impact of least important factors on overall LSM results.

464 Furthermore, there are several factors that determine the importance of features for carrying

465 out LSM such as (1) completeness and quality of the landslide inventory dataset used for

466 analysis, (2) mapping scale of the features maps like landcover, lithology, or other geological

467 features. If the spatial locations of landslides in an inventory does not represent the ground

468 truth phenomenon, then there can be negative impact of landslide input data for feature

469 selection. Sampling methodology of landslide selection is important, there are various ways to

470 use landslides in carrying out susceptibility assessment, many studies have used 70-30 ratio

471 and others have used random sampling or K-fold sampling methods (Merghadi et al.,

472 2018;Chen et al., 2018). One of the most important observations from this study was the

473 reclusion of the "least important factors" in the context of LSM. The fact that despite removal

474 of certain conditioning factors, we still get very good results or comparable results after

475 removing them, this observation explains employing the important conditioning factors are

476 enough for LSM.

477 The use of landslide samples along with non-landslide samples can affect the landslide feature

478 importance as can be seen in results in this study. In the case of the statistical model, one of the

479 reasons for the lower AUC performance can be accredited to the absence of the non-landslide

480 samples. As the model was trained without non-landslide samples and simply trained with only

481 landslide samples, the model's ability to discriminate between the non-landslide and landslide

482 pixels is affected therefore, predicting landslide occurrences over non-landslide locations.

483 Because of this reason, the statistical model exhibited homogeneous distribution of predicted

484 landslide pixels (see figure 7). We used landslides and non-landslide samples for training the

485 ML models which shows varying results from that of the statistical ensemble model (See figure

486 8 and 9). There is more homogeneous distribution of landslide susceptibility classes in

487 statistical model results, but it is evident from the machine learning results that the non-

488 landslide samples have a greater impact on final landslide susceptibility results.

489

490   7.   Conclusions

491   In the current state-of-the-art approaches for LSM, the contemporary literature lays emphasis

492   on different models for improving accuracy of landslide susceptibility against the test data.

493   However, this study investigated how the conditioning factors affect the overall prediction of

494   landslides in the context of northeast Italy, Belluno province. An important aspect of this study

495   was to identify if at all, removing the "least important" conditioning factors in the modelling

496   process affects the performance in predicting new unknown landslides.

497   As understood, ML models require conditioning factors as input for LSM, however, investing

498   on the importance of the features (conditioning factors) could possibly provide a better

499   understanding of landslide occurrences with respect to the available conditioning factor maps

500   for LSM. This study indicates that various models behave differently with different features,

501   whereby the same features that are important in one instance of a particular model, can be the

502   least important in other models. Therefore, this study gives new insights towards the use of

503   already available conditioning factor maps, without exhausting resources for generating other

504   conditioning factor maps that might not be available.

505   In this study we also concluded that the landslides and non-landslides samples impacts the

506   feature importance, especially in the ML models, and in contrast, the statistical model used

507   only landslide samples. Therefore, it was found to be crucial in asserting a balance between the

508   two data samples to avoid overfitting or underfitting. This study illustrates that feature selection

509   is very important step of carrying out LSMs. We found that there are differences in the final

510   LSMs derived from the statistical and ML models, which are attributed to the above-mentioned

511   sample selection techniques.

512   This research introduces the importance of post-training feature importance algorithms for

513   LSM. This approach can also be used to assess the susceptibility of other natural disasters. The

514 results can eventually comment whether certain conditioning factors can be discarded while

515 modelling landslide occurrences. In many parts of the globe, the availability of data is scarce

516 and therefore, with the ability to model landslides without relying on the conventional factors,

517 we can still predict landslides spatially over a given region. Although there are certain

518 drawbacks like (1) the same factor maps will not be available everywhere, (2) factors that are

519 least important in one region might not repeat the same behaviour in other regions of the world,

520 and (3) model capability changes with respect to different regions, the resulting susceptibility

521 maps can still give quality information for local emergency relief measures, planning of disaster

522 risk reduction, mitigation, and to evaluate potentially affected areas.

523

524 Funding: This research was funded by the Veneto Region, VAIA-LAND project, Research

525 Unit UNIPD-GEO, Principal Investigator Mario Floris.

526

References:

528 Baglioni, A., Tosoni, D., De Marco, P., and Arziliero, L.: Analisi del dissesto da frana in
529 Veneto, 2006.
530 Baird, C.: Comparison of Risk Assessment Instruments in Juvenile Justice, 2013.
531 Boretto, G., Crema, S., Marchi, L., Monegato, G., Arziliero, L., and Cavalli, M.: Assessing the
532 effect of the Vaia storm on sediment source areas and connectivity storm in the Liera catchment
533 (Dolomites), Copernicus Meetings, 2021.
534 Brabb, E. E., Pampeyan, E. H., and Bonilla, M. G.: Landslide susceptibility in San Mateo
535 County, California, Reston, VA, Report 360, 1972.
536 Breiman, L.: Random Forests, Machine Learning, 45, 5-32, 10.1023/A:1010933404324, 2001.
537 Can, R., Kocaman, S., and Gokceoglu, C.: A Comprehensive Assessment of XGBoost
538 Algorithm for Landslide Susceptibility Mapping in the Upper Basin of Ataturk Dam, Turkey,
539 Applied Sciences, 11, 4993, 2021.
540 Castellanos Abella, E. A., and Van Westen, C. J.: Qualitative landslide susceptibility
541 assessment by multicriteria analysis: A case study from San Antonio del Sur, Guantánamo,
542 Cuba, Geomorphology, 94, 453-466, 10.1016/j.geomorph.2006.10.038, 2008.
543 Catani, F., Lagomarsino, D., Segoni, S., and Tofani, V.: Landslide susceptibility estimation by
544 random forests technique: sensitivity and scaling issues, Natural Hazards and Earth System
545 Sciences, 13, 2815-2831, 10.5194/nhess-13-2815-2013, 2013.
546 Chacón, J., Irigaray, C., Fernández, T., and El Hamdouni, R.: Engineering geology maps:
547 landslides and geographical information systems, Bulletin of Engineering Geology and the
548 Environment, 65, 341-411, 10.1007/s10064-006-0064-z, 2006.

Chen, T., Trinder, J. C., and Niu, R.: Object-oriented landslide mapping using ZY-3 satellite imagery, random forest and mathematical morphology, for the Three-Gorges Reservoir, China, Remote sensing, 9, 333, 2017.

Chen, W., Peng, J. B., Hong, H. Y., Shahabi, H., Pradhan, B., Liu, J. Z., Zhu, A. X., Pei, X. J., and Duan, Z.: Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China, Science of the Total Environment, 626, 1121-1135, 10.1016/j.scitotenv.2018.01.124, 2018.

Chung, C.-J. F., and Fabbri, A. G.: Validation of Spatial Prediction Models for Landslide Hazard Mapping, Natural Hazards, 30, 451-472, 10.1023/B:NHAZ.0000007172.62651.2b, 2003.

Compagnoni, B., Galluzzo, F., Bonomo, R., and Tacchia, D.: Carta geologica d'Italia, Dipartimento difesa del suolo, 2005.

Corò, D., Galgaro, A., Fontana, A., and Carton, A.: A regional rockfall database: the Eastern Alps test site, Environmental Earth Sciences, 74, 1731-1742, 10.1007/s12665-015-4181-5, 2015.

Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Masuda, T., and Nishino, K.: GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping, Environmental Geology, 54, 311-324, 10.1007/s00254-007-0818-3, 2008.

Dai, F. C., Lee, C. F., and Ngai, Y. Y.: Landslide risk assessment and management: an overview, Engineering Geology, 64, 65-87, https://doi.org/10.1016/S0013-7952(01)00093-X, 2002.

Desiato, F., Lena, F., Baffo, F., Suatoni, B., Toreti, A., di Ecologia Agraria, U. C., and Romagna, A. E.: Indicatori del clima in Italia, APAT, Roma, 2005.

Doglioni, C.: Thrust tectonics examples from the Venetian Alps, 1990.

Dunning, S., Massey, C., and Rosser, N.: Structural and geomorphological features of landslides in the Bhutan Himalaya derived from terrestrial laser scanning, Geomorphology, 103, 17-29, 2009.

Dury, G.: Hillslope form and Process. M.A. Carson and M.J. Kirkby, 1972. Cambridge University Press, London, vii + 475 pp., £ 6.60, 1972,

Ercanoglu, M., and Gokceoglu, C.: Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach, Environmental Geology, 41, 720-730, 10.1007/s00254-001-0454-2, 2002.

Floris, M., Iafelice, M., Squarzoni, C., Zorzi, L., De Agostini, A., and Genevois, R.: Using online databases for landslide susceptibility assessment: an example from the Veneto Region (northeastern Italy), Nat. Hazards Earth Syst. Sci., 11, 1915-1925, 10.5194/nhess-11-1915-2011, 2011.

Forzieri, G., Pecchi, M., Girardello, M., Mauri, A., Klaus, M., Nikolov, C., Rüetschi, M., Gardiner, B., Tomaštík, J., Small, D., Nistor, C., Jonikavicius, D., Spinoni, J., Feyen, L., Giannetti, F., Comino, R., Wolynski, A., Pirotti, F., Maistrelli, F., Savulescu, I., Wurpillot-Lucas, S., Karlsson, S., Zieba-Kulawik, K., Strejczek-Jazwinska, P., Mokroš, M., Franz, S., Krejci, L., Haidu, I., Nilsson, M., Wezyk, P., Catani, F., Chen, Y. Y., Luyssaert, S., Chirici, G., Cescatti, A., and Beck, P. S. A.: A spatially explicit database of wind disturbances in European forests over the period 2000–2018, Earth Syst. Sci. Data, 12, 257-276, 10.5194/essd-12-257-2020, 2020.

Gariano, S. L., Verini Supplizi, G., Ardizzone, F., Salvati, P., Bianchi, C., Morbidelli, R., and Saltalippi, C.: Long-term analysis of rainfall-induced landslides in Umbria, central Italy, Natural Hazards, 106, 2207-2225, 10.1007/s11069-021-04539-6, 2021.

Ghorbanzadeh, O., Rostamzadeh, H., Blaschke, T., Gholaminia, K., and Aryal, J.: A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-

fold cross-validation approach for land subsidence susceptibility mapping, Natural Hazards, 94, 497-517, 10.1007/s11069-018-3449-y, 2018.

Glade, T., Anderson, M. G., and Crozier, M. J.: Landslide hazard and risk, John Wiley & Sons, 2006.

Goetz, J. N., Brenning, A., Petschko, H., and Leopold, P.: Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling, Computers & Geosciences, 81, 1-11, 10.1016/j.cageo.2015.04.007, 2015.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M.: Estimating the quality of landslide susceptibility models, Geomorphology, 81, 166-184, 10.1016/j.geomorph.2006.04.007, 2006.

Huang, F., Chen, J., Du, Z., Yao, C., Huang, J., Jiang, Q., Chang, Z., and Li, S.: Landslide Susceptibility Prediction Considering Regional Soil Erosion Based on Machine-Learning Models, ISPRS International Journal of Geo-Information, 9, 377, 2020.

Iadanza, C., Trigila, A., Starace, P., Dragoni, A., Biondo, T., and Roccisano, M.: IdroGEO: A Collaborative Web Mapping Application Based on REST API Services and Open Data on Landslides and Floods in Italy, ISPRS International Journal of Geo-Information, 10, 89, 2021.

Komac, M.: A landslide susceptibility model using the Analytical Hierarchy Process method and multivariate statistics in perialpine Slovenia, Geomorphology, 74, 17-28, 10.1016/j.geomorph.2005.07.005, 2006.

Lee, S.: Landslide detection and susceptibility mapping in the Sagimakri area, Korea using KOMPSAT-1 and weight of evidence technique, Environmental Earth Sciences, 70, 3197-3215, 10.1007/s12665-013-2385-0, 2013.

Linden, A.: Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis, Journal of evaluation in clinical practice, 12, 132-139, 2006.

Liu, L.-L., Yang, C., and Wang, X.-M.: Landslide susceptibility assessment using feature selection-based machine learning models, Geomechanics and Engineering, 25, 1-16, 2021.

Melville, B., Lucieer, A., and Aryal, J.: Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia, International journal of applied earth observation and geoinformation, 66, 46-55, 2018.

Merghadi, A., Abderrahmane, B., and Tien Bui, D.: Landslide Susceptibility Assessment at Mila Basin (Algeria): A Comparative Assessment of Prediction Capability of Advanced Machine Learning Methods, ISPRS International Journal of Geo-Information, 7, 10.3390/ijgi7070268, 2018.

Meten, M., PrakashBhandary, N., and Yatabe, R.: Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia, Geoenvironmental Disasters, 2, 9, 10.1186/s40677-015-0016-7, 2015.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., and Kanevski, M.: Machine Learning Feature Selection Methods for Landslide Susceptibility Mapping, Mathematical Geosciences, 46, 33-57, 10.1007/s11004-013-9511-0, 2014.

Mondal, S., and Maiti, R.: Integrating the analytical hierarchy process (AHP) and the frequency ratio (FR) model in landslide susceptibility mapping of Shiv-khola watershed, Darjeeling Himalaya, International Journal of Disaster Risk Science, 4, 200-212, 2013.

Pham, B. T., Tien Bui, D., Pourghasemi, H. R., Indra, P., and Dholakia, M. B.: Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods, Theoretical and Applied Climatology, 128, 255-273, 10.1007/s00704-015-1702-9, 2015.

Pham, B. T., Tien Bui, D., and Prakash, I.: Bagging based Support Vector Machines for spatial prediction of landslides, Environmental Earth Sciences, 77, 146, 10.1007/s12665-018-7268-y, 2018.

Pourghasemi, H. R., Pradhan, B., and Gokceoglu, C.: Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran, Natural hazards, 63, 965-996, 2012.

Pradhan, B.: Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches, Journal of the Indian Society of Remote Sensing, 38, 301-320, 10.1007/s12524-010-0020-z, 2010.

Raja, N. B., Çiçek, I., Türkoğlu, N., Aydin, O., and Kawasaki, A.: Landslide susceptibility mapping of the Sera River Basin using logistic regression model, Natural Hazards, 85, 1323-1346, 10.1007/s11069-016-2591-7, 2017.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Reviews, 180, 60-91, 10.1016/j.earscirev.2018.03.001, 2018.

Riley, S. J., DeGloria, S. D., and Elliot, R.: Index that quantifies topographic heterogeneity, intermountain Journal of sciences, 5, 23-27, 1999.

Rossi, M., Guzzetti, F., Salvati, P., Donnini, M., Napolitano, E., and Bianchi, C.: A predictive model of societal landslide risk in Italy, Earth-Science Reviews, 196, 102849, https://doi.org/10.1016/j.earscirev.2019.04.021, 2019.

Sahin, E. K.: Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest, SN Applied Sciences, 2, 1308, 10.1007/s42452-020-3060-1, 2020.

Sauro, F., Zampieri, D., and Filipponi, M.: Development of a deep karst system within a transpressional structure of the Dolomites in north-east Italy, Geomorphology, 184, 51-63, https://doi.org/10.1016/j.geomorph.2012.11.014, 2013.

Schönborn, G.: Balancing cross sections with kinematic constraints: The Dolomites (northern Italy), Tectonics, 18, 527-545, 1999.

Senouci, R., Taibi, N.-E., Teodoro, A. C., Duarte, L., Mansour, H., and Yahia Meddah, R.: GIS-Based Expert Knowledge for Landslide Susceptibility Mapping (LSM): Case of Mostaganem Coast District, West of Algeria, Sustainability, 13, 630, 2021.

Shahabi, H., Khezri, S., Ahmad, B. B., and Hashim, M.: Landslide susceptibility mapping at central Zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models, Catena, 115, 55-70, 2014.

Shahabi, H., and Hashim, M.: Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment, Scientific reports, 5, 9899, 2015.

Stanley, T. A., Kirschbaum, D. B., Benz, G., Emberson, R. A., Amatya, P. M., Medwedeff, W., and Clark, M. K.: Data-Driven Landslide Nowcasting at the Global Scale, Frontiers in Earth Science, 9, 10.3389/feart.2021.640043, 2021.

Trigila, A., Iadanza, C., and Spizzichino, D.: Quality assessment of the Italian Landslide Inventory using GIS processing, Landslides, 7, 455-470, 10.1007/s10346-010-0213-0, 2010.

Trigila, A., and Iadanza, C.: Landslides and floods in Italy: hazard and risk indicators - Summary Report 2018, 2018.

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

Youssef, A. M., and Pourghasemi, H. R.: Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia, Geoscience Frontiers, 12, 639-655, 10.1016/j.gsf.2020.05.010, 2021.