



# Earthquake-triggered landslide susceptibility in Italy by means of Artificial Neural Network

Gabriele Amato<sup>1</sup> · Matteo Fiorucci<sup>1,4</sup> · Salvatore Martino<sup>1</sup> · Luigi Lombardo<sup>2</sup> · Lorenzo Palombi<sup>3</sup>

Received: 31 January 2022 / Accepted: 8 March 2023 / Published online: 6 April 2023  
© Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

The use of Artificial Neural Network (ANN) approaches has gained a significant role over the last decade in the field of predicting the distribution of effects triggered by natural forcing, this being particularly relevant for the development of adequate risk mitigation strategies. Among the most critical features of these approaches, there are the accurate geolocation of the available data as well as their numerosity and spatial distribution. The use of an ANN has never been tested at a national scale in Italy, especially in estimating earthquake-triggered landslides susceptibility. The CEDIT catalogue, the most up-to-date national inventory of earthquake-induced ground effects, was adopted to evaluate the efficiency of an ANN to explain the distribution of landslides over the Italian territory. An ex-post evaluation of the ANN-based susceptibility model was also performed, using a sub-dataset of historical data with lower geolocation precision. The ANN training highly performed in terms of spatial prediction, by partitioning the Italian landscape into slope units. The obtained results returned a distribution of potentially unstable slope units with maximum concentrations primarily distributed in the central Apennines and secondarily in the southern and northern Apennines. Moreover, the Alpine sector clearly appeared to be divided into two areas, a western one with relatively low susceptibility to earthquake-triggered landslides and the eastern sector with higher susceptibility. Our work clearly demonstrates that if funds for risk mitigation were allocated only on the basis of rainfall-induced landslide distribution, large areas highly susceptible to earthquake-triggered landslides would be completely ignored by mitigation plans.

**Keywords** Artificial Neural Network · Landslide susceptibility · Slope unit partition · CEDIT catalogue · Italy

## Introduction

The concept of landslide susceptibility defines the expectation of where landslides may occur in a given landscape, thus providing information on the spatial component of the landslide hazard definition (Varnes and the IAEG Commission

on Landslides and Other Mass-Movements, 1984; Guzzetti et al. 1999; Lombardo et al. 2020a). The numerical expression of a landslide susceptibility corresponds to the probability of landslide occurrences within a given mapping unit (Lima et al. 2017; Broeckx et al. 2018; Lombardo and Mai 2018). As reported by Bird and Bommer (2004), earthquake-triggered landslides often cause the largest damage related to earthquakes and affect transportation routes inhibiting recovery and safety operations during emergency phases (Martino et al. 2019; Mantovani et al. 2019). In detail, earthquake-triggered landslides caused more than 50% of the total worldwide losses due to landslides in the last decades (Petley 2012).

To preemptively reduce the risk associated with these processes, predictive models have been proposed to estimate the distribution of earthquake-triggered ground effects scenarios (Sassa 1996; Jibson et al. 2000; Prestininzi and Romeo 2000; Romeo 2000; Wasowski and Del Gaudio 2000; Del Gaudio et al. 2003; Jibson 2007; Hsieh and Lee 2011; Lombardo and Tanyaş 2020, among others) representative

✉ Gabriele Amato  
gabriele.amato@uniroma1.it

<sup>1</sup> “Sapienza” University of Rome, Department of Earth Sciences and Research Centre for Geological Risks (CERI), P. le A. Moro 5, 00185 Rome, Italy

<sup>2</sup> Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7500 Enschede, AE, Netherlands

<sup>3</sup> “Nello Carrara” Applied Physics Institute - National Research Council of Italy (IFAC-CNR), Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy

<sup>4</sup> University of Cassino and Southern Lazio, Department of Civil and Mechanical Engineering, Cassino, Italy

of a uniform hazard distribution or seismic shaking scenarios. Out of several available options proposed to preemptively estimate earthquake-triggered effects, the proposed approaches essentially boil down to two types: physically based approaches (Van Westen et al. 2006) and statistically based ones (Guzzetti et al. 2005). The first type of approach implies that slope stability analyses are performed to quantify safety factors (Martino 2016) and/or the expected seismically induced displacements of the landslide masses and include Newmark's method (Del Gaudio and Wasowski 2004). The recently proposed PARSIFAL (Probabilistic Approach to pRo-viding Scenarios of earthquake-Induced slope FAiLures) approach allows obtaining probability maps of expected Newmark's displacements at regional scales (Esposito et al. 2016; Martino et al. 2018, 2019).

As regards the statistically based counterpart, the general framework is quite similar when data-driven (statistical and machine learning) models are used. Typically, a mapping unit is chosen between grid-cells and slope-units and a dichotomous status expressing the absence or presence of landslides (or 0/1) is assigned. In a subsequent step, the binary status is fitted to a set of predictors chosen to represent predisposing factors of slope instability and the outcome of the modeling procedure is a probability (Amato et al. 2019). However, the algorithmic architecture one chooses to implement has notable repercussions on the performance each model provides. For instance, simple bivariate statistical models provide quite straightforward interpretation of the functional relations existing between factors and landslides (e.g. Weight of Evidence; Bonham-Carter 1989; Van Westen 2002; Martino et al. 2019). But, this is achieved at the expense of statistical rigour (the model does not assume any underlying probability distribution nor the interaction among explanatory variables) and performances. For instance, multivariate statistical routines assume that landslides are distributed over space according to the Bernoulli probability distribution (Lombardo et al. 2019). And, they allow to model linear relations (in the case of Generalised Linear Models; Ayalew and Yamagishi 2005; Castro Camilo et al. 2017) or a combination of linear and nonlinear relations (in the case of Generalised Additive Models; Brenning 2008; Goetz et al. 2011) between predisposing factors and landslides occurrences. These models offer excellent performance while keeping a clear interpretability at each step and for each model component (Lombardo et al. 2014; Frattini et al. 2010). Ultimately, machine learning methods provide equally and often even higher performance than the other two approaches mentioned above, this time though at the expense of the interpretability of each step. The reason behind this characteristic is due to the fact that machine learning algorithms are often based on the combination of highly nonlinear functions which are difficult to be individually and multivariately traced as the model evolves converging to the best solution (Liu et al. 2014; Zhou

et al. 2016, 2018). Because of the high performance provided, the number of machine learning applications in landslide science rose in recent years (Marjanović et al. 2011; Huang et al. 2017; Zhu et al. 2017; Amato et al. 2021). And, Artificial Neural Networks (ANNs) (e.g. Ermini et al. 2005; Gomez and Kavzoglu 2005), including their more recent convolutional extensions (e.g. Wang et al. 2019) have demonstrated to be a valid tool for landslide susceptibility assessment. Neural networks (Hassoun et al. 1995) are characterised by the possibility of modelling the relationship between independent and dependent variables in a complex non-linear way and are by nature prone to overparameterization of the model itself. These aspects lead to the advantages that ANN have the ability to model complex relations when these are not known a priori and in the fact that, thanks to overparameterization, they are very little sensitive to problems of collinearity (De Veaux and Ungar 1994) between independent variables (i.e. the internal dependence among variables that may cause that one variable explains the variability of another, or more than, one). Thus, ANNs approaches are particularly suitable for big data.

Most of the works that adopt ANN approaches to analyse earthquake-triggered landslide susceptibility perform analysis at a regional scale (Song et al. 2012; Umar et al. 2014; Zhou and Fang 2015) using input landslide inventories that are limited in time and space to single earthquakes (Tanya et al. 2017; Shrestha and Kang 2019; Tanya and Lombardo 2020). A classical approach to perform earthquake-triggered landslide susceptibility analysis by means of ANN techniques usually sees, as a first step, inventorying landslides triggered by a selected large earthquake. Afterwards, a susceptibility model, driven by the inventoried landslides, is produced for the region hit by the selected earthquake (Song et al. 2012; Umar et al. 2014; Zhou and Fang 2015; Shrestha and Kang 2019). As a result, the susceptibility analysis ends to have a relatively local impact, being hardly exportable to other seismic regions, even within the same country. This is mainly because the model is earthquake-specific. To generalize the model and make it more exportable, landslides triggered by other relevant earthquakes, even from surrounding regions, should be included in the starting inventory.

In this study, we performed a susceptibility analysis of earthquake-triggered landslides by implementing an Artificial Neural Network approach for the whole Italian territory. Slope units have been adopted as mapping units (Alvioli et al. 2020; Titti et al. 2022a; Loche et al. 2022a). The input landslides inventory used to train the network has been accessed via the Italian Catalogue of Earthquake-Induced Ground Failures (CEDIT; Fortunato et al. 2012; Martino et al. 2014; Caprari et al. 2018), which collects ground effects caused by earthquakes occurred over the whole Italian territory from the XII century to present days. As far as the authors know, this represents the first study dealing with earthquake-triggered landslide (EQtLs) susceptibility for the

whole Italian territory. To clearly evaluate and present the achieved results, we quantified the ANN classification performances through commonly adopted metrics, and we generated the first Italian EQtLS susceptibility map. Besides, we investigated the importance of the predictors performing a Permutation Feature Importance (PFI) of single predictors and explored how the classification performance varies by selecting all the possible combinations among predictors groups (i.e. terrain, seismic, geothematic, hydrological, and anthropic predictors). Ultimately, we checked the obtained susceptibility map for every Italian administrative region, by using an additional landslide dataset, which was not included during the ANN training phase. The resulting percentage of unstable territory for every Italian region has been computed to highlight priorities in land management practices at more local scales other than the national one.

## Material and methods

### Italian morphotectonic settings

Italy is the European country most affected by landslides (Herrera et al. 2018), with over 620,000 landslides recorded in the framework of the IFFI dataset, the most complete and detailed landslides inventory existing in Italy (Trigila et al. 2013). The main triggering factors for landslides in Italy are intense rainfalls and earthquakes. And in recent years, anthropogenic factors such as road cuts have been assumed to also play an increasing role. From 1999 until December 2019, almost two billion Euros were dedicated by Italian institutions to interventions for landslide risk mitigation, which great majority was realised within or close to areas classified with a high landslide hazard (link here). This highlights the high relevance that the classification of the territory assumes to establish remediation funding priorities. Italy is also characterised by active geodynamics related to the geological evolution of the two major mountain chains, i.e. the Alps in the north and the Apennines throughout the peninsula, as testified by the distribution of earthquakes and volcanic activity. More specifically, the Alps' chain shows a double-verging growth, involving the exhumation of metamorphic rocks. Conversely, the Apennines chain consists of a single-east-verging belt, mostly characterised by thin-skinned tectonics. As a consequence, earthquakes show prevalent compressional focal mechanisms at the fronts of the two chains and extensional mechanisms along the Apennines backbone (Carminati et al. 2010). The highest magnitude seismic events, with peak ground acceleration (PGA) values higher than 0.225 g and a return time of 475 years, are expected in the central-southern Apennines, Calabria region (on the southwest of the Italian peninsula), in the southeastern part of Sicily island and in the north-eastern sector of

the Alps chain. Medium to low seismic acceleration values (PGA up to 0.225 g) are expected with a return time of 475 years along the entire Alpine Arch, along the entire western Italian coast and the peri-Adriatic regions (eastern Italian coast). Ultimately, the Sardinia island is the only sector with very low seismic hazard (link here).

### CEDIT catalogue

The EQtLS susceptibility model we built in this study is based on data collected in the CEDIT catalogue (Prestininzi and Romeo 2000; Fortunato et al. 2012; Martino et al. 2014, 2019, 2020a, b; Caprari et al. 2018), which is available online at (link here). The CEDIT collects all the ground-failure effects produced by earthquakes that occurred in Italy taking information from historical documents and literature, especially in the case of earthquakes that occurred before 1980, while ground effects induced by more recent events have been surveyed directly on the field by the CERI (Research Centre for the Geological Risks of Sapienza University of Roma) working group (see Martino et al. 2017, for more details on the standard cataloguing procedure). Among the many information contained in the CEDIT, the error estimation assigned to each ground effect location according to the following scheme (Martino et al. 2014) was found particularly useful for this work:

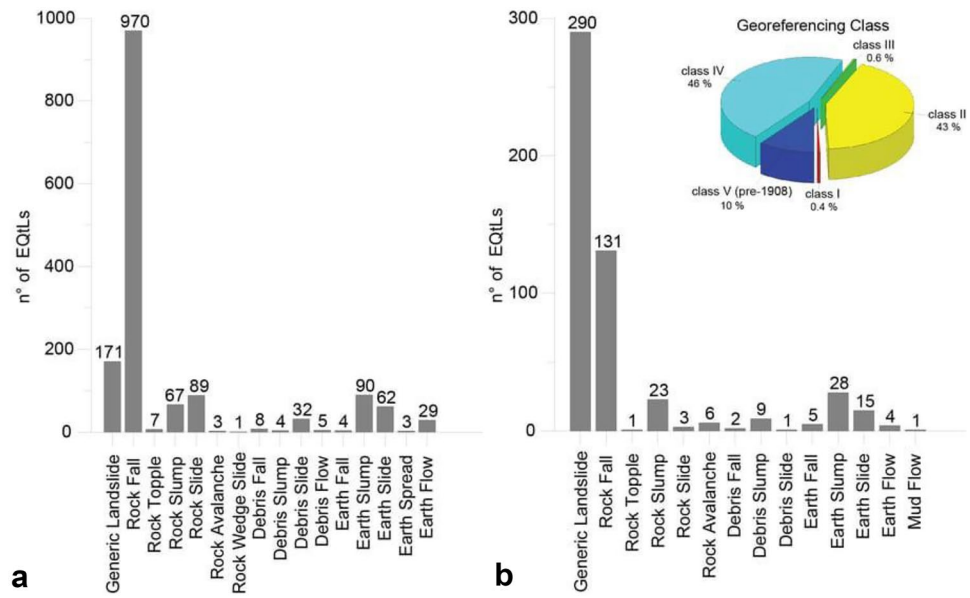
- Class 5: high-quality location from historical documents or GPS measurement with any or negligible associated error;
- Class 4: location coordinates with an average error of 1 km;
- Class 3: location coordinates with an average error of 3 km;
- Class 2: location coordinates with an average error of 10 km;
- Class 1: location coordinates with an average error of 30 km.

The EQtLS were extracted from the CEDIT catalogue and were split into two different subsets (Fig. 1):

1. An "Input dataset" containing 1545 landslides, all belonging to the georeferencing class 5. These were induced by the earthquakes that occurred in Italy from 1908 to 2018.
2. A "Check dataset" containing 465 landslides with georeferencing classes ranging from 1 to 4, induced by all the earthquakes contained in the CEDIT catalogue, and 54 landslides belonging to the georeferencing class 5, induced by earthquakes that occurred in Italy before 1908.

With the aim to provide the susceptibility analysis with a high-quality geolocalised dataset, we used the Input

**Fig. 1** Bar chart showing the distribution of the type of EQtLs for Input dataset (a) and for Check dataset (b), together with the georeferencing class distribution for the Check dataset only (b). “Generic Landslides” means that the landslide type is not known

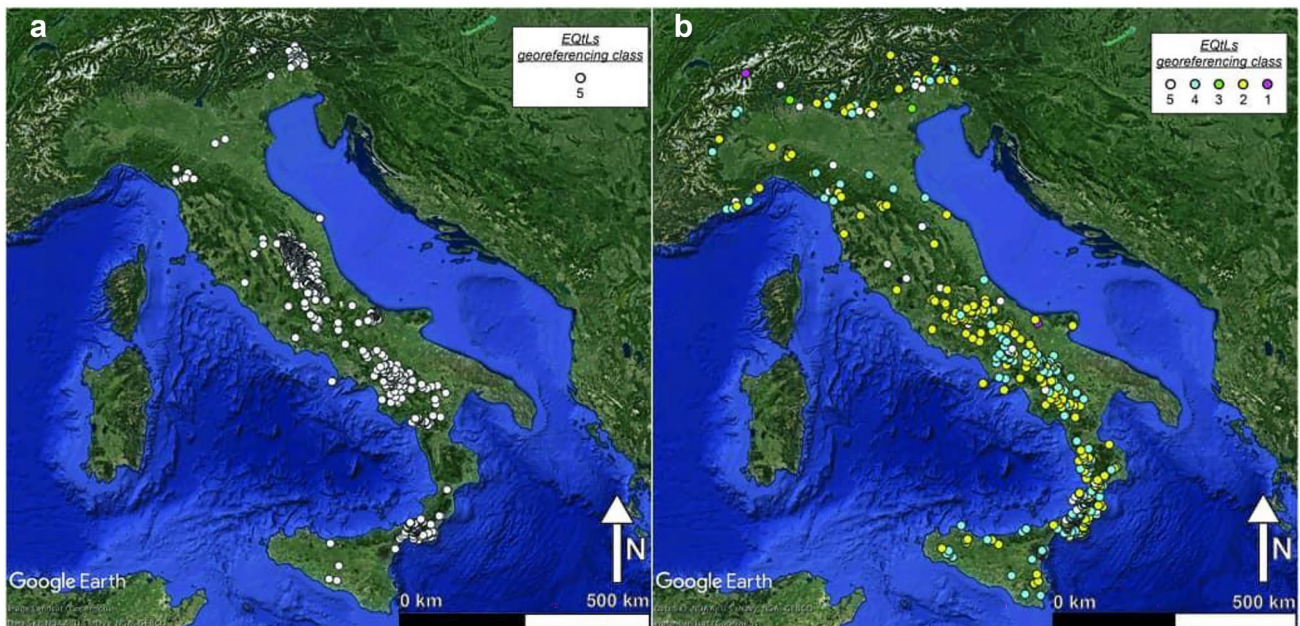


dataset for the training and cross-validation-test cycles of the neural network, whereas we used the Check dataset to perform an a-posteriori and independent verification of the EQtLs susceptibility map of Italy. The spatial distribution of the two here considered datasets (i.e. earthquake-induced landslides for Input and Check) are shown in Fig. 2.

**Model building strategy**

**Mapping unit**

A mapping unit in landslide science is considered to be a geographical object upon which the landscape is partitioned (Carrara 1988). The vast majority of the landslide



**Fig. 2** a Spatial distribution of EQtLs belonging to the Input dataset and b of EQtLs belonging to the Check dataset, coloured on the basis of their georeferencing class

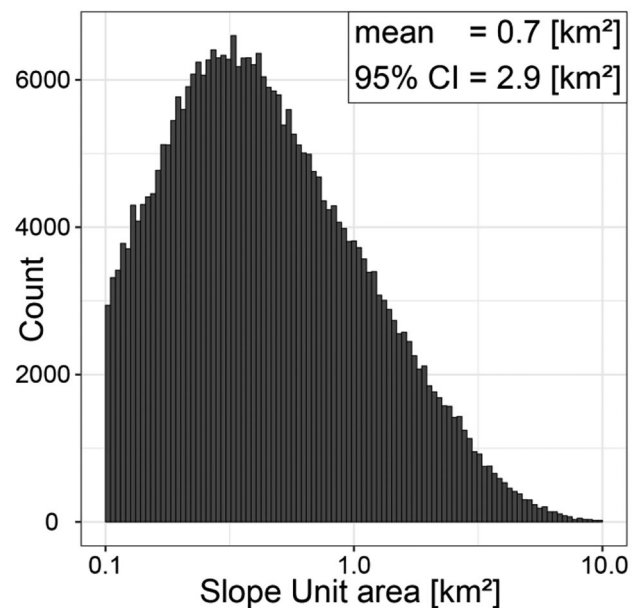
susceptibility literature is based on regular mapping units shaped as a squared (e.g. Jibson et al. 2000; Steger et al. 2020) or hexagonal (Avolio et al. 2013; Lupiano et al. 2018) lattice. However, when it comes to statistically based applications, the way these units are used is generally flawed for a few reasons. These have been extensively described in Reichenbach et al. (2018) and we direct the reader to this article for more details. Most of these issues do not affect a valid alternative represented by a slope unit (SU) partition which is mapping units bounded by ridges catchment/sub-catchment divides and streamlines (Carrara et al. 1991, 1995). SUs are able to capture the variability of the landscape associated with the failure process, by maximising homogeneity of slope steepness and aspect within a single unit and heterogeneity of the same between adjacent SUs (Alvioli et al. 2016; Titti et al. 2021). On the other hand, a SU choice requires an additional step which corresponds to the aggregations or upscaling of properties that are represented over space with a much higher resolution. Oftentimes, mean and standard deviation values are extracted for numerical properties at the scale of the single SU (e.g. Guzzetti et al. 2006). But, these could also be expressed via different summary distribution metrics, e.g. such as quantiles (Amato et al. 2019). Similarly, it is not standardised the way categorical properties such as lithology or land use are aggregated at the SU scale and different approaches can be found in the literature (e.g. Castro Camilo et al. 2017; Schlögel et al. 2018).

In this study, we select a SU partition of the Italian territory. In addition to the above-mentioned reasons, for such a large study area, choosing a small regular lattice would have inevitably produced several tens of millions of grid-cells. The alternative of seeking a reasonable size of the dataset would have instead produced grid-cells which would have been individually very coarse, neglecting any geomorphological representation of the landscape under study. The SUs we used were made available by Alvioli et al. (2020) at the following address (link here). In Fig. 3, we summarised the distribution of all the SUs' planimetric areas, ranging from approximately 0.1 to 10 km<sup>2</sup>.

To support the analyses in this study, we assigned unstable conditions' labels to SU intersected by landslides contained in the Input dataset while we assigned a stable label to all the others.

### Predictor variables

To support the modeling protocol at a scale comparable to the whole Italian territory, we selected a broad set of predictors aimed at expressing properties known or assumed to influence landslide occurrences. In Table 1 in the Appendix, we provide a general overview of these predictors by grouping them into geological, seismic, anthropic, terrain, and hydrological



**Fig. 3** Distribution of SU planimetric areas. The  $x$ -axis is plotted in a logarithmic scale to improve the figure readability. The 95% confidence interval is calculated as the difference between the 97.5 and the 2.5 percentiles of the SU area

macro-classes. Before describing these properties, we stress to the reader that in order to aggregate each predictor at the SU scale, we used the mean and standard deviation criterion for continuous properties as well as the dominant class for categorical properties. And we recall that, if collinearity among predictors, which may negatively affect the model by inflating the variance estimates (McElroy and Jach 2019), exists, ANNs are able to handle it by spreading the estimated weights over the collinear variables to take into account the different noise levels, taking actually advantage in terms of predictive performance. Therefore, we have chosen to keep the whole predictor set which consists of 167 layers (more details will be provided in “Artificial Neural Network”).

### Geothematic predictors

We considered three geo-thematic properties, detailed below:

1. Landforms are specific geomorphic features on the earth's surface which encompass both large-scale terrains such as plains or mountain ranges and small-scale characteristics such as single hills or valleys (Jacek 1997). The work of Guisan et al. (1999) first and Jenness (2006) later has pioneered the automatic extraction of such features from DEMs. More recently, Jasiewicz and Stepinski (2013) have implemented an efficient automatic classification tool for landforms named geomophon, which returns 10 terrain

morphologies in 10 classes: (1) flat, (2) summit, (3) ridge, (4) shoulder, (5) spur, (6) slope, (7) hollow, (8) footslope, (9) valley, (10) depression. In this study, we used geomorphon to initially calculate the ten landforms, and in a subsequent step, we have aggregated this information at the SU scale by assigning to a given mapping unit the class with the largest planimetric extent.

2. Similarly, we have assigned to each SU the predominant lithological type. This geological information was retrieved from the Geological Map of Italy at 1:500,000 scale. This map was based on 1:100,000 and 1:50,000 national geological cartography or geological maps (Tacchia et al. 2005). Overall, after the aggregation step, 21 lithology classes have been assigned to SUs across the whole Italian territory. Table 2 in the Appendix offers a description of each class.
3. The predominant soil type was assigned to each SU on the basis of the European Soil Map compiled by the European Commission—Joint Research Centre (Finke and Montanarella 2001). In this map, soil type classes are classified according to the World Reference Base (WRB) system, which consists of a two-levels terminology. The first level defines the Reference Soil Groups whereas the second level is nested within the first and consists of a set of principal and supplementary qualifiers (for more details, see link here). In this study, SUs have been classified on the basis of 91 soil type classes, which have been used for modeling purposes and mainly belong to the Reference Soil Groups reported in Table 3 in the Appendix.

Concerning the categorical predictors, slope units have been labelled with “1” in correspondence with the predominant classes of geomorphon, soil type and lithology, and “0” for all the other classes.

#### Seismic predictors: distance to seismogenic features

Seismic information has been considered in the form of Euclidean distance to the nearest active fault and the Euclidean distance to the nearest seismogenic source. Specifically, for each SU, the mean distance value and its standard deviation have been computed. Data required to produce these predictors have been accessed from the Italian Database of Individual Seismogenic Sources (DISS) catalogue (link here), which continuously updates the study on the Italian seismogenic sources. An Individual Seismogenic Source is obtained by parameterising the geometry and kinematics of large active faults considered capable of generating earthquakes with a magnitude ( $M_w$ ) greater than 5.5 (Basili et al. 2008; DISS-Working-Group 2018). This corresponds to an active fault that has accumulated some displacement in the recent past and can be

considered very likely to produce a new offset in the near future (link here).

#### Terrain predictors

Concerning Terrain predictors, we used the 20m DEM released by the Italian Institute for Environmental Research in 2013 (link here). And, for each slope unit, we calculated the mean value and the standard deviation of the following derivatives:

- Elevation (e.g. Ayalew and Yamagishi 2005) can be considered a proxy for climate-related characteristics (e.g. ground temperature or even the precipitation itself when high ridges play the role of meteorological barriers). And, its standard deviation per slope unit mimics the signal of surface roughness.
- Eastness and Northness, these are computed as the sine and cosine of the aspect expressed in radians, respectively (Lombardo et al. 2018). These are two linear components of the nonlinear slope exposition signal, a common proxy for strata attitude and localised dry/wet soil conditions.
- Slope gradient (Zevenbergen and Thorne 1987) expresses the potential gravitational forces acting over a given slope.
- General, Longitudinal, and Tangential Curvatures (Evans 1980; Wood 1996); Planar and Profile Curvatures (Heerdegen and Beran 1982). Plan and profile curvatures carry the signal of the potential soil availability, and potential small-scale hydraulic and gravitational forces (Ohlmacher 2007). Conversely, cross-sectional curvature measures the curvature perpendicular to the downslope direction. As a result, it detects small-scale features such as channels. Longitudinal curvature plays a similar role but parallel to the downslope direction (Patel and Sotiropoulos 1997).
- Topographic Positioning Index (TPI, De Reu et al. 2013) measures the difference between the elevation of a focal cell and the average elevation within a predetermined radius.
- Topographic Roughness Index (TRI, Riley et al. 1999) expresses rough terrain conditions.
- Topographic Wetness Index (TWI, Beven and Kirkby 1979) expresses the terrain’s tendency to retain water at a given location, as a function of local slope steepness and upslope contributing areas. Therefore, it conveys information related to potential high pore pressure conditions distributed over the landscape or the presence of open floodplains.
- The area of each slope unit ( $A_{SU}$ ) controls the availability of potential material to fail.

### Anthropic predictors: distance to roads

An ideal situation to inform any predictive model of the potentially destabilising effect of road cuts would be to collect the exact location and height of the cut. However, such information is available only for the location component and no height characteristics can be accessed for the whole Italian road network. For this reason, we opt to compute the Euclidean distance from roads at buffers equal to 5, 10, 50, and 100 m. Subsequently, a series of statistical metrics of the distances to roads have been calculated for each SU, namely the mean, maximum, and minimum distance of the unit from the closest road and the portion of the territory extending within certain distance ranges. Therefore, the following statistics have been calculated for every slope unit using the Zonal Statistics Plugin, in QGIS 3.10.4 (Graser 2016).

- Count: the count of the number of pixels at a < 100-m distance;
- Sum: the sum of the pixel distance values;
- Mean: the mean distance;
- Min: the minimum distance;
- Max: the maximum distance;
- Range: the range (max–min) of distance;
- Majority: the most represented distance within a slope unit;
- Count < 5: the count of the number of pixels at a < 5 m distance.

### Hydrological predictors: distance to watercourses

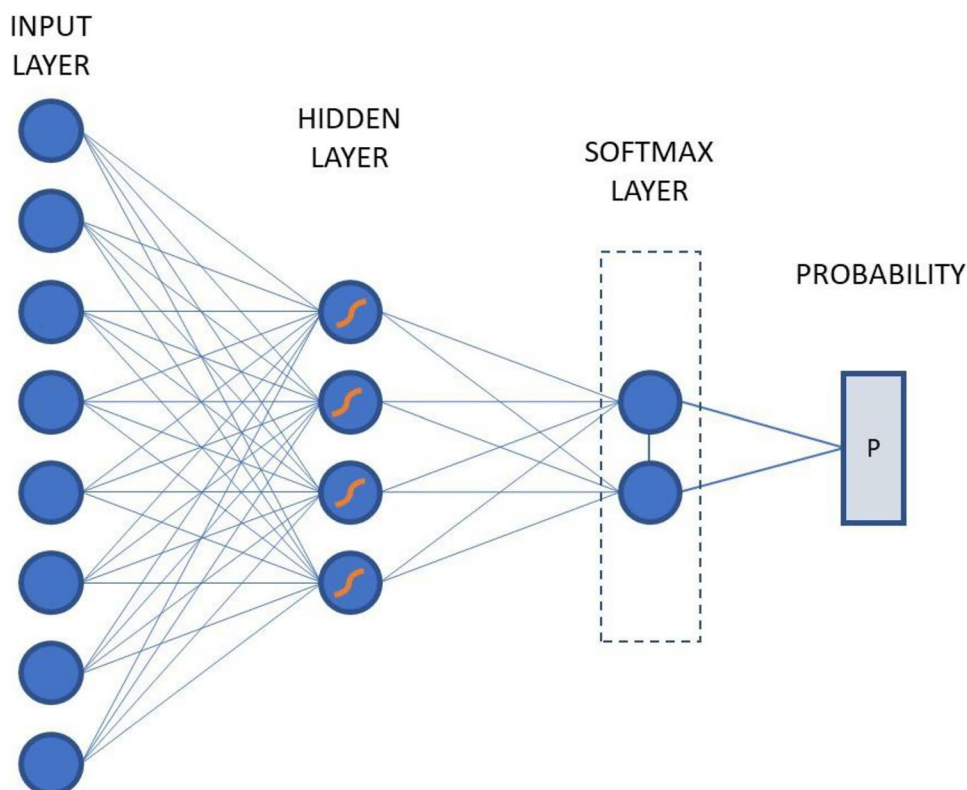
The Euclidean distance from watercourses has been computed similarly to the road network case. This time though, we extracted ten equally spaced (100 m wide) buffer zones from 0 up to 1000 m from each streamline. The same summary statistics calculated for the distance from the road network have been computed also for the hydrological network with respect to each slope unit.

### Artificial Neural Network

The used ANN architecture was previously proposed in Amato et al. (2021) and has been here optimised to perform a binary classification between stable and unstable slope units. Stable slope units are those SUs with no EQtLs while unstable SUs contain at least one landslide of the Input dataset. The ANN training is performed on balanced class datasets (Titti et al. 2022b). The used network is a “shallow” ANN whose architecture is a two-layer fully connected feed-forward network. For the hidden layer, a sigmoid activation function has been considered. The output layer is a “softmax layer”, in which the outputs are normalised into probabilities proportional to the exponentials of the input

values (Fig. 4). The network is trained by scaled conjugate gradient backpropagation. To limit any overfitting effect an “early stopping by validation” training criterion has been adopted. The classification process associates a probability value, from 0 to 1, to each slope unit to be susceptible to EQtLs. In order to be correctly trained to distinguish between stable and unstable slope units, the ANN needed to learn from samples of both classes. We set a fixed number of samples per class (equal to the number of all the slope units with landslides). Therefore, the Input dataset counted 644 positives (i.e. slope units with landslides) and an equal number of negatives (i.e. slope units without landslides), the latter chosen randomly from the larger number available. The Input dataset was then split as follows: 70% of samples were used to train the network, 15% was used for validation and 15% as the test dataset. The training dataset is used to optimise the weights and the bias assigned to each node of the ANN. After each step of the iterative training, the ANN classification is applied also on the validation dataset and the classification performances on the two datasets are monitored. As the classification performance continues to improve on the training dataset but worsens on the validation dataset, the training process is early stopped and overfitting of the model is avoided. Finally, the test dataset is a completely independent dataset used to test the reproducibility of performances obtained on the first two sets. In order to build a statistically significant distribution of the classification results and performance metrics, we replicated the training procedure 100 times. To ensure the maximum statistical independence, for each of the 100 replicates, the training, validation, and test datasets are recreated from scratch as described before. Furthermore, the initial values of ANN weights and biases are randomly changed. Fixed the ANN architecture, some of the operating network hyperparameters, and in particular, the number of nodes in the hidden layer, have been tuned to achieve the best and more reliable performances. In the “tuning” tests, the ANN performance was calculated as True Positive Rate (TPR, or Recall). TPR is the ratio between the number of true positives (i.e. those samples correctly predicted by the model as belonging to the given class) and the sum of true positives and false negatives (i.e. those samples that the model predicted as not belonging to a given class while they belong). A number varying from 1 to 6 nodes in the hidden layer has been tested. It resulted in a TPR increase as the number of nodes increased. The number of nodes was finally set to 4 as being the smallest number of nodes, which still produced a significant increase in performance. At the end of each of the 100 training replicates, the ANN was run on all the SUs, covering the whole national territory. The mean of the probability values output from the 100 classification replicates, as well as their standard deviation, was calculated and was used to plot the Earthquake-induced Landslide Susceptibility Map of Italy.

**Fig. 4** Schematic diagram of the ANN architecture



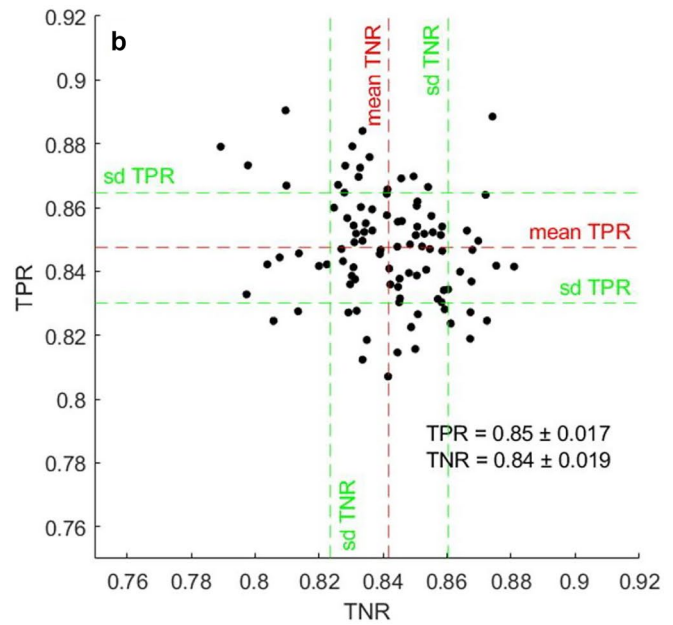
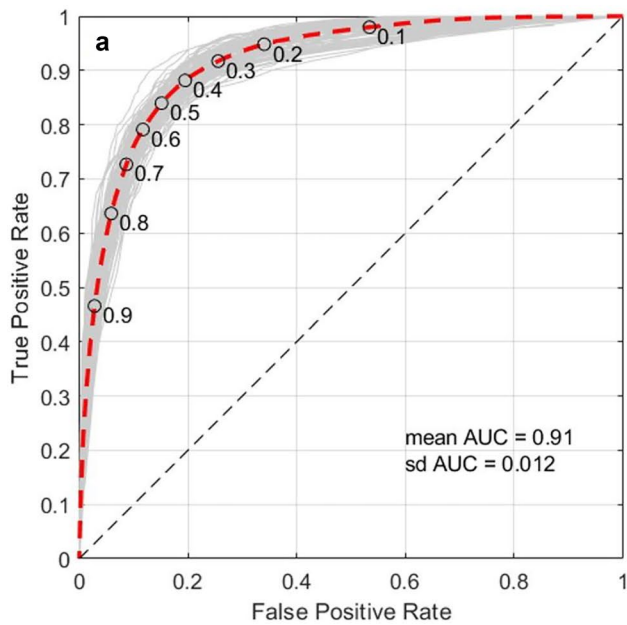
### Performance assessment: validation routines

In the case of binary classification, a probability threshold is set to associate a particular sample to one of the two possible classes. The most appropriate way to investigate the discriminatory capabilities of a binary classifier for each possible value of the discrimination threshold between 0 and 1 is commonly the receiver operating characteristic (ROC; Rah-mati et al. 2019) plot. ROC plots report the TPR on the  $y$ -axis and the false positive rate (FPR or fall-out) on the  $x$ -axis. FPR is defined as the ratio between false positives and all the negatives, namely false positives + true negatives. False positives are samples classified as belonging to the class of interest while they were not, whereas true negatives are those samples correctly predicted by the model as not belonging to the class of interest. The Area Under the Curve (AUC) is strictly linked to the shape of the ROC curve, and it is a good proxy of the overall capability of a model to distinguish between two classes, regardless of what classification threshold is chosen. AUC assumes values between 0 and 1, gradually increasing with the classification capabilities of the model. If AUC is 1, the model is perfectly able to distinguish between positive class and negative class (Hosmer and Lemeshow 2000).

In this study, a probability threshold of 0.5 has been chosen to classify each slope unit as stable or unstable. This choice is confirmed by examining the point of the average

ROC corresponding to a threshold value of 0.5 (as also reported in Fig. 5a). This point is the closest one to a TPR equal to 1 and an FPR equal to zero. Once the threshold value has been chosen, it is possible to further investigate the obtained discrimination capabilities by the means, for instance, of a Confusion Plot (Rossi and Reichenbach 2016; Lombardo et al. 2020b). Conversely to ROC (and AUC), the Confusion Plot is a threshold-dependent method to evaluate classification performance. It shows, fixed the classification threshold, the value of TPR as a function of the corresponding TNR for each of the 100 ANN replicates, allowing to evaluate the statistical robustness of the classification to different training, validation, and test datasets. In model performance evaluation, TNR stands for True Negative Rate and is the ratio between the number of true negatives and the sum of true negatives and false positives. In this study, TNR refers to the success rate in classifying slope units as belonging to the “stable” class, and TPR refers to the “unstable” one. Besides Confusion Plot and ROC (plus AUC), which are considered good indicators of the general performance of a model and commonly adopted in the scientific literature (Lombardo and Mai 2018), the performance obtained by the network in this study has been evaluated also representing the importance assumed by each predictor during the classification by performing a Feature Importance analysis. This procedure highlights those predictors that gave a major contribution to the success of the susceptibility analysis. To





**Fig. 5** **a** ROC of each of the 100 ANN replicates with, in red, the resulting average. Circles represent different classification thresholds. Also, AUC mean and standard deviation values are reported. **b** TPR/

TNR confusion plot of the 100 ANN replicates for a 0.5 classification threshold. The mean and standard deviation of TPR and TNR are reported

make this, the Permutation Feature Importance (PFI) was adopted. The method is based on the assumption that a random variation of the value of an important predictor has a negative impact on the performance of the model greater than that of the random variation of a less important predictor (Putin et al. 2016). The permutation allows the random variation of the predictor while preserving the natural distribution of the values of the predictor itself (Gao et al. 2020). In the current study, the PFI was applied to each of the predictors. The model reduction, i.e. the PFI score of a predictor, was calculated as the ratio between the TPR of the non-permuted model and the TPR of the permuted model. FPI scores were evaluated for each of the 100 ANN replicates thus allowing the evaluation of a statistical distribution of the predictors' importance. Also, we grouped the 167 predictors into 5 groups (Road, Hydro, Geo, Terrain, Seismic; see Table 1 for more details) and we investigated how the network performance varies by running the classification 20 times with each of all the possible different combinations of the five groups. Finally, the susceptibility map was verified by means of a comparison with the Check dataset, and for

each Italian administrative region, an additional check TPR was calculated, as well as the percentage of territory classified as unstable.

### Results

Table 4 shows the average values and standard deviation of the TPR, TNR, and AUC general performance indicators obtained through the 100 ANN replicates. The results are reported for the three types of datasets we considered, namely training, validation, and test. Furthermore, we also report the values obtained for the dataset composed of the sum of the three subsets (All). The results in the table show high performances for the three indicators considered. The average values for the three datasets are also comparable, demonstrating that the approach followed is able to limit any evident overfitting effect and the consequent loss of generality in the slope unit classification phase. Very limited values of the standard deviations also demonstrate the robustness of the method, which is able to obtain comparable

**Table 1** Performance of the ANN after 100 replicates. For each indicator, mean and standard deviation are provided

	Mean TPR	SD TPR	Mean TNR	SD TNR	Mean AUC	SD AUC
Train	0.86	0.02	0.85	0.02	0.92	0.01
Validation	0.83	0.03	0.84	0.03	0.91	0.02
Test	0.81	0.04	0.79	0.03	0.89	0.02
All	0.85	0.02	0.84	0.02	0.91	0.01

performances regardless of the specific datasets used in each of the 100 replicates.

Considering the comparability of the performances obtained on the three training, validation, and test datasets, for the following results, it was considered appropriate to report those obtained on the overall dataset composed by the three.

Figure 5 a shows (in grey) the ROC obtained for each of the 100 replicates of the ANN training. The average ROC is shown in red. In this study,  $AUC = 0.91$  has been reached on average, with a standard deviation smaller than 0.02 (see Fig. 5a and Table 4). Besides the best classification threshold that resulted in being about 0.5, in Fig. 5a, the TP and FP rates related to other eight different thresholds (from 0.1 to 0.9) are indicated by the means of black circles in order to allow a deeper and direct interpretation of the EQtLs susceptibility probability value that the model associates to each SU. As an example, by choosing a threshold value of 0.8 a very low FPR (about 0.06) is obtained, meaning that only a very limited fraction of the stable SU would be wrongly classified as unstable. As a result, those SUs that have been classified with a probability higher than 0.8 to be susceptible to EQtLs, are statistically very significantly likely to have actually experienced landslides/be true positives. Figure 5 b shows, for the 100 ANN replicates, the values of the TPR parameter according to the TNR parameter. Mean and standard deviation ranges are also reported for both TPR and TNR. On average the classification has a very similar success rate for both classes (about 0.84) with a small standard deviation (0.02), demonstrating that the classification is carried out with the same accuracy for both classes. The low value of the standard deviation and the absence of correlation between the values of TPR and TNR also make it possible to assert that the results obtained are robust with regard to the statistical representativeness of the samples considered and the absence of bias introduced.

## Susceptibility mapping

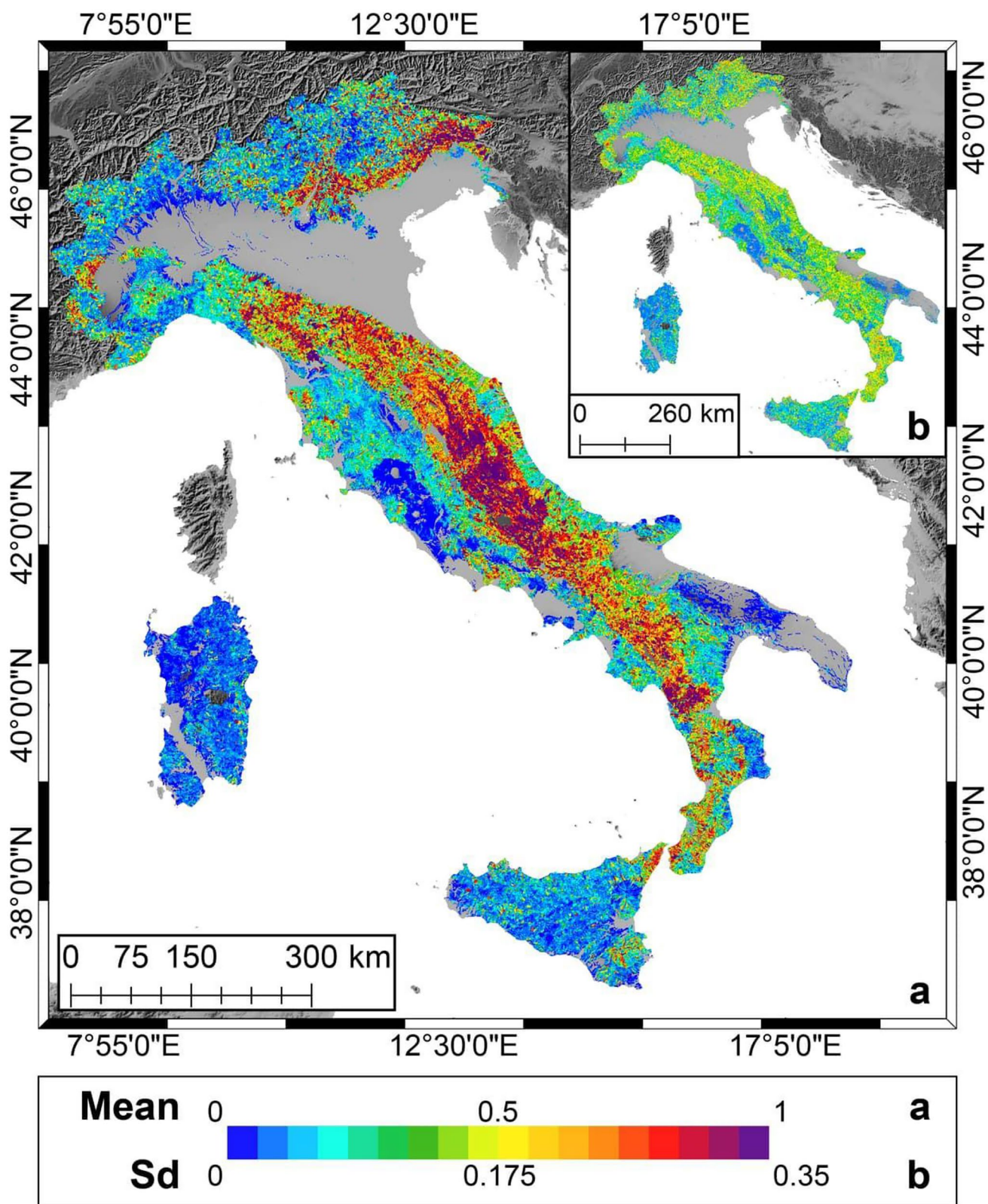
After every training replicate, the ANN was applied to all the slope units of Italy and 100 susceptibility values for each SU have been generated. The mean susceptibility of each SU, and its standard deviation, after 100 replicates has been considered to produce the EQtLs susceptibility map of Italy (Fig. 6a).

In the EQtLs susceptibility map of Italy, flat lowland areas have been taken out from the classification and resulted in grey colour. Orange to red areas represent moderately to highly susceptible slope units (probability  $> 0.5$ ), while green to blue areas have been classified as stable. Susceptible areas are frequent in the northeastern part of Italy and along an NW-SE-oriented longitudinal belt that corresponds to the Apennine mountain chain. In particular, red

areas are located in correspondence with the epicentral area of historically strong earthquakes and a moderate density of unstable slope units is present in the Calabria region, the most southern region of the Italian peninsula. Conversely, most of the western side of the peninsula and of the alpine region, in the north, are lowly susceptible to being affected by EQtLs. Also, the south-east and the two main Italian islands, Sicily and Sardinia, are widely blue-coloured. The standard deviation of the resulting classification (Fig. 6b), associated with the mean susceptibility of every SU, is very low ( $< 0.1$ ) in correspondence with the high susceptibility SUs in central Italy and in the north-east, as well as for most of the highly stable areas. In general, the standard deviation of the susceptibility is low (0.1–0.18) for the overall Italian territory. Higher values are present in limited spotted locations and more concentrated in the Calabria region.

A point of novelty of this study is represented by the comparison of the landslide Susceptibility map of Italy with an EQtLs dataset that was not used to train the network.

The overlapping between checking landslides and unstable SUs has been evaluated to verify the correctness of the susceptibility map. In order to make the checking process reliable, a radius sized as the associated error has been taken into account around the less precisely georeferenced landslides. When more than half of the area of the resulting circle overlapped with unstable slope units, that landslide was considered a true positive (TP). Conversely, when the overlap was limited to less than half of the circle area, landslides were considered false negatives (FN). When some parts of the uncertainty circles included areas with no classification (e.g. lowlands or sea), only the portion overlapping with classified slope units was considered. Consequently, the checking TPR (C-TPR) has been calculated for every Italian region. On the basis of the susceptibility map, also the regional percentage of unstable territory has been computed. As a result, in most of the Italian regions, the number of TP was higher than FN, although not all the regions counted the same number of landslides from the checking dataset. In cases of regions with only a few ( $< 10$ ) checking landslides, C-TPR generally reached very small values. Conversely, Friuli, Veneto, Emilia-Romagna, Tuscany, Abruzzo, Molise, Campania, and Basilicata show very good performances (C-TPR  $\geq 70\%$ ) and a high number of checking landslides ( $> 14$ ), which makes the resulting statistics quite reliable. In these regions, the percentage of unstable territory varies from around 20–40% to more than 60%. Contextually, Lombardy, Latium, Sicily, and Calabria show low to very low C-TPR despite the good number of checking samples. In Calabria, 36% of the regional extent has been classified as unstable, while in the other three regions, the unstable territory is  $< 20\%$ . Nevertheless, considering the low C-TPR values achieved, these percentages might have been probably underestimated.



**Fig. 6** EQtLs susceptibility map of Italy shown as **a** the mean estimated probability of EQtLs occurrence per SU, through the 100 ANN replicates, and **b** as the standard deviation per SU associated with the mean shown in the larger panel to the left

## Predictors' importance

PFI provided an interesting analysis of the importance that the single predictor had in order to achieve the final classification.

In Fig. 8, it can be seen that the ANN mainly relies on five predictors, all listed among Geothematic and Seismic ones. In particular, soil type (code 122), distance from seismogenic sources (123), lithology (31), distance from active faults (125), and geomorphon (10) have the highest PFI score, respectively. The first terrain predictor in order of importance is represented by the mean tangential curvature of a slope unit (code 144). Its importance, however, varied significantly among the 100 replicates. Following, all the other predictors, such as other terrain predictors and the road-related ones, account for very little contribution to the classification and the associated PFI standard deviation is small. On the basis of the EQtLs susceptibility map, the a posteriori distribution of the classes of GEO predictors among the unstable slope units has been analysed at the national level in Fig. 9.

To make the chart clearer, only soil types with unstable slope units higher than 10% have been reported. Concerning soil types, slope units mainly covered by Dystric Cambisol resulted highly susceptible to EQtLs, and 75% of them have been classified as unstable, although they are not numerous (< 5000 in the whole national territory). In the WRB system, "Dystric" indicates a soil with base saturation of less than 50% at a given depth, and Dystric Cambisol is located in small parts of central and south Apennine, in seismically very active areas, which have been historically hit by strong earthquakes. Further, more than 60% of the slope units composed of Rendzic Leptosol have been classified as unstable. Rendzic Leptosol is described as very shallow soils immediately overlying highly calcareous material and is quite frequent in Italy, particularly in central and south Apennine as well as in Friuli and Veneto regions. Almost 50% of slope units characterised by the main presence of Chromi-calcaric Luvisol have been classified as unstable. Chromi-calcaric Luvisol is defined by the WRB system as a reddish calcareous with a marked textural differentiation whose surface horizon was depleted of clay, which accumulated more in-depth. And, according to the pedological map of Europe, it is very rare in Italy. Finally, almost 40% of slope units were dominated by Lithic Leptosol, a very shallow soil with continuous hard rock within 10 cm from the soil surface (Table 3), resulted susceptible to EQtLs. In Italy, its occurrence is limited to central Apennine, between Latium and Abruzzo, and in Sicily island. Concerning lithology, 75% of slope units are mainly constituted by chaotic sedimentary complexes, and 50% of those composed by marls have been classified as unstable. The first lithology is composed of turbiditic sandstones and clays, locally with evaporites and limestones. It is mainly spread in central Italy, along the eastern side of the Apennine chain. The second type of rock is spread in

central Italy and the north-west. Successively, 25–40% of arenaceous and limestone slope units resulted susceptible to EQtLs. Arenaceous formations crop out all over the Italian territory, mainly in mountain areas, while Limestones are spread in seismic regions of central Italy, such as Umbria and Abruzzo, in the southern part of the Alps, and along the coasts of south Italy. Finally, metamorphic rocks, mainly granitoid gneiss, whose almost 20% of slope units are considered unstable, crop out only in the northern part of the Alps and in small parts of Calabria and Sicily regions. Concerning the slope morphology, valley, and concave slope units interestingly resulted to be relatively more unstable than slope units located in other parts of the slope. In detail, the 25–35% of hollow, valley, and depression slope units have been classified as unstable against the 15–20% of summit, ridge, spur, and slope classes. Finally, slope units which are linked with flat areas, such as flat, shoulder and footslope, are generally stable.

In this paragraph, the analysis of how the classification performance changes varying the combination of groups of predictors used by the ANN is also provided. Predictors have been grouped as Terrain, Seismic, Geo (i.e. Geothematic), Hydrological and Roads (Anthropic) as described in "Material and methods." All possible combinations made up of a variable number of groups have been taken into account (one group at a time up to all five groups together). For each of the possible combinations among these groups, the ANN has been run 20 times, and the related AUCs have been calculated. Figure 10 shows the box plot of the AUC values distribution among the 20 replicates and for all the possible combinations of predictor groups.

Combinations are ordered by the medians of the AUC distributions. The background colour varies according to the quartiles of the distribution of the median AUCs calculated over the 20 replicas per combination. The median quartiles are at AUC values of 0.84, 0.88 and 0.89. Lower performances (AUC < 0.84) are generally achieved with only one or two groups, or with 3-groups combinations that contain Hydrology and Roads but not Geo. Good performances (0.84 < AUC < 0.89) are achieved with all the 2-groups combinations that include Geo. In this regard, Geo+Seismic performs the best. Also, combinations with three or four groups achieve good performances. Finally, those combinations of predictors groups whose AUC is entirely included in the dark red band and greater than the third quartile (AUC > 0.89) can be considered the best-performing ones. Among these, two 3-groups' combinations are listed.

1. Geo + Seismic + Road
2. Geo + Seismic + Terrain

whereas the three 4-groups' combinations are:

1. Geo + Seismic + Terrain + Hydrology
2. Geo + Seismic + Hydrology + Road

### 3. Geo + Seismic + Terrain + Road

and only one combination exists with all five groups. From the analysis of the best-performing combinations, it is clear as Geo and Seismic predictors must be both considered in order to achieve a median AUC higher than 0.89, and that at least another group is also needed. The importance of Geo (i.e. lithology, soil type and geomorphon of slope units) and Seismic (i.e. distances from active faults and seismogenic sources) predictors was previously indicated also by the PFI analysis. Nevertheless, what and how many predictor groups are needed besides Geo and Seismic was not straightforward. Related to this, on the basis of the interquartile range and the median of AUC values, Geo+Seismic+Terrain+Road and Geo+Seismic+Terrain seem to perform slightly better than all the other combinations.

Figure 11 represents a heatmap of the mean AUC value obtained by adding one of the five groups of predictors to each of all their possible combinations.

Each row contains one of the possible combinations and is sorted from top to bottom by the increasing number of groups. In each column, one of the five groups is present. The mean AUC obtained after 20 ANN replicates considering the combination in the row and the addition of the group in the column is reported in each cell of the heatmap. “Null” row and column respectively indicate that none of the possible combinations has been considered and that no groups have been added. Figure 10 confirms what has been previously seen in Fig. 10 that the higher the number of groups within a combination, the higher the performance. Nevertheless, not all the groups have the same effect. When the classification has been carried out taking only one group at a time (first row on the top), Terrain and Geo performed the best, with mean AUC = 0.84, and significantly better than Seismic (mean AUC = 0.79) although some of the Seismic features resulted among the most important in the full model PFI analysis. Nevertheless, Terrain + Seismic reaches AUC > 0.9 only when Geo is added while, conversely, Geo + Seismic reaches AUC > 0.9 also with Roads appearing that, when combined with Geo, Seismic provides a bigger contribution than Terrain. This led to infer that Terrain and Geo groups might bring partially overlapping information and that those brought by Seismic better combine with Geo than with Terrain features. In general, when the Geo group is added to whatever combination (second column from the left in Fig. 11), the mean AUC reaches 0.9 in seven cases and it never goes below 0.8. This means that the Geo predictors have a high importance for the ANN and their presence ensures very good performances, whatever another group is added to the combination. Similarly, Seismic predictors allow reaching mean AUC  $\geq$  0.9 when added to six different combinations. Further, when they are present, performance decreases below 0.8 only in one case and the combination

Geo + Seismic achieves AUC = 0.89. Conversely to Geo and Seismic, only four combinations that include Hydrological predictors allow achieving a mean AUC of at least 0.9 and, in all these cases, Geo is present. Also, two combinations that include Hydrological predictors do not reach AUC = 0.8. Finally, five combinations containing Roads and five combinations containing Terrain reach a mean AUC = 0.9. This means that the probability to reach very good performance by a combination that contains Road-related predictors is the same as a combination that includes terrain predictors. However, when Roads is added to Geo and Seismic, AUC arrives at 0.91 and vice-versa, when Terrain is added to Geo and Seismic, AUC averagely arrives at 0.92. Adding Roads to Geo + Seismic + Terrain brings a contribution lower than 0.01 while, adding Hydrology, performance decreases to 0.90. Besides, 0.92 is the highest mean AUC reached by the ANN and is due to the main contribution of Geo and, secondly, Seismic information. Terrain predictors would have a much higher importance, when grouped, than that resulting from the single predictors’ analysis. But its information might be partially provided also by Geo predictors and, when combined with other groups, it accounts for slope units variability less than Geo group, ending to provide only + 0.03 to the combination Geo + Seismic. From this analysis, the key role of Geo and Seismic predictors is confirmed and emphasised. Also, a significant contribution of Terrain has been proven. At the same time, the non-significance of distance to rivers as a predictor for EQtLs susceptibility resulted and a not ignorable contribution to improving the classification performance given by the presence of roads. Finally, concerning what can be selected as the most performing combination among all the possible and tested ones, it should be noted that the differences between the mean AUC values for the three best median AUC combinations, which are Geo + Seismic + Terrain, Geo + Seismic + Terrain + Road, and Geo+Seismic+Terrain+Road+Hydro, are not statistically significant ( $p$ -value = 0.86 with one-way ANOVA test).

## Discussions

The sections below are meant to provide the reader with an overview of the strengths and potential weaknesses of the modeling protocol we implemented, these discussed both from the data as well as the modeling strategy perspectives.

## Supporting arguments

### ANN performance overview

The ANN performance was very good. In detail, after 100 replicates mean AUC was 0.91 and the associated standard

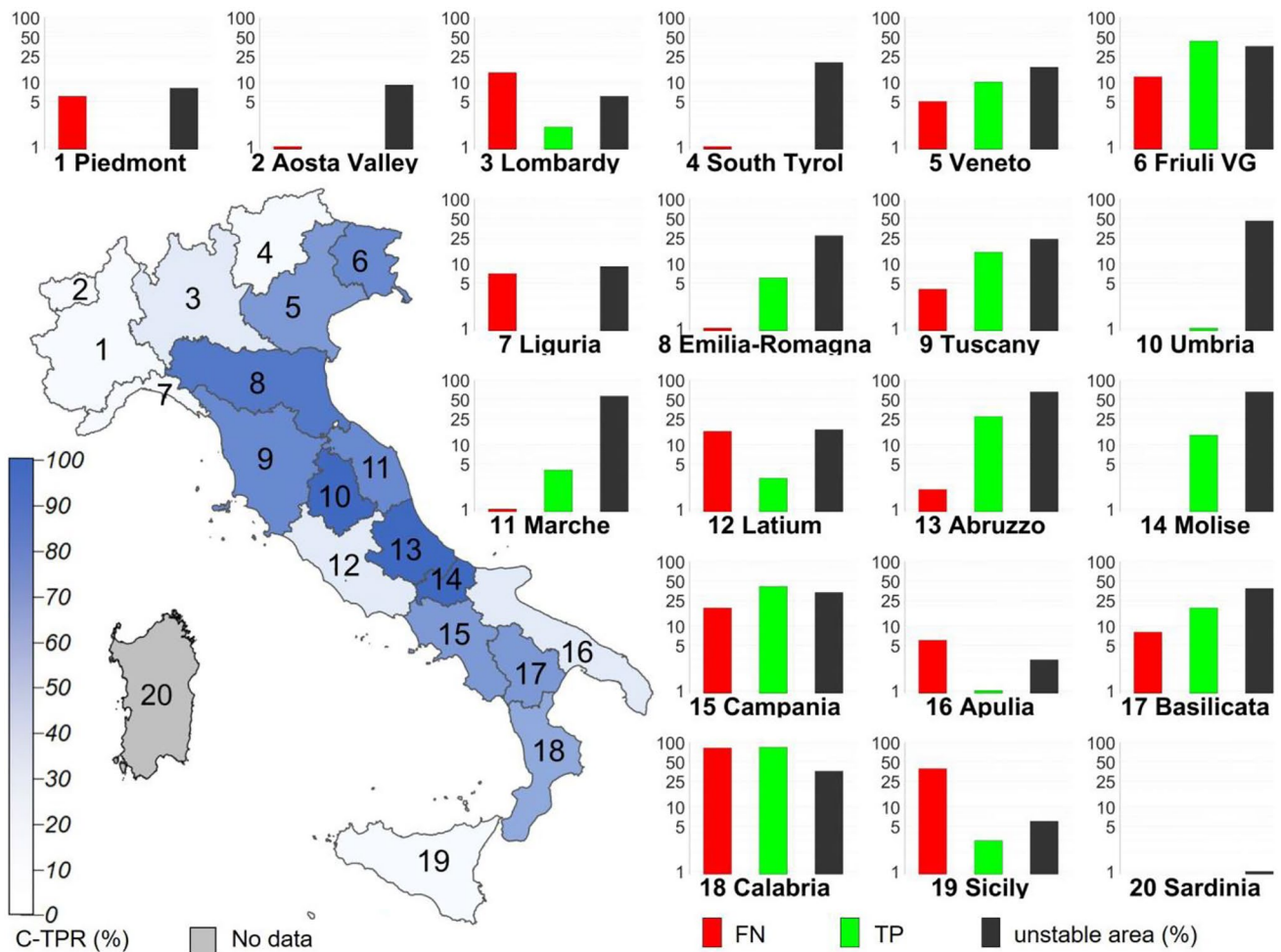


Fig. 7 C-TPR value (in map), number of False Negatives and True Positives (red and green bars) per region with respect to the checking dataset. The dark bars represent the regional estimated percentage of unstable territory according to the susceptibility map

deviation was 0.01. Considering that both positive and negative samples (i.e. slope units with and without landslides) within training, test, and validation datasets changed at every replicate, the very low standard deviation is an excellent

result, which demonstrates the solid stability of the network. Also, the ability to distinguish between the two classes was high: averagely, TPR, namely the ability to correctly classify unstable slope units, was 0.85 while TNR, proficiency in

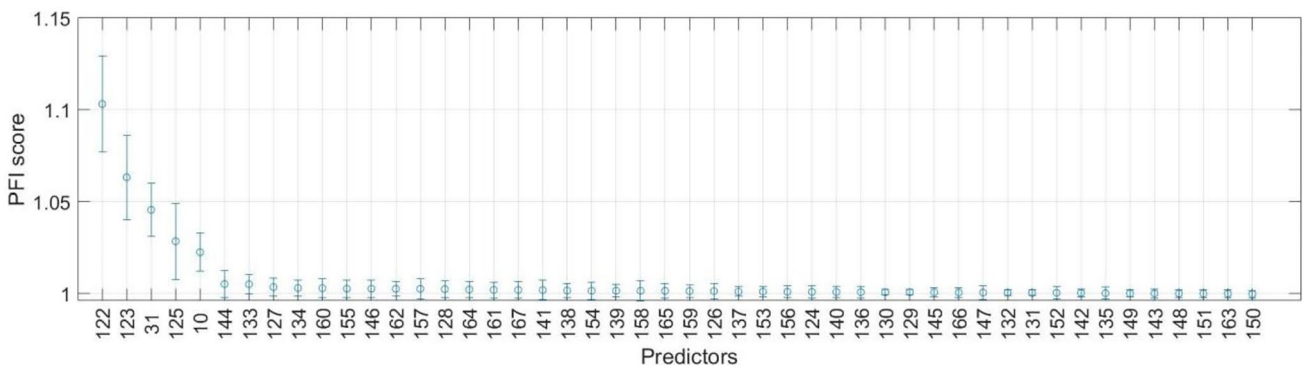
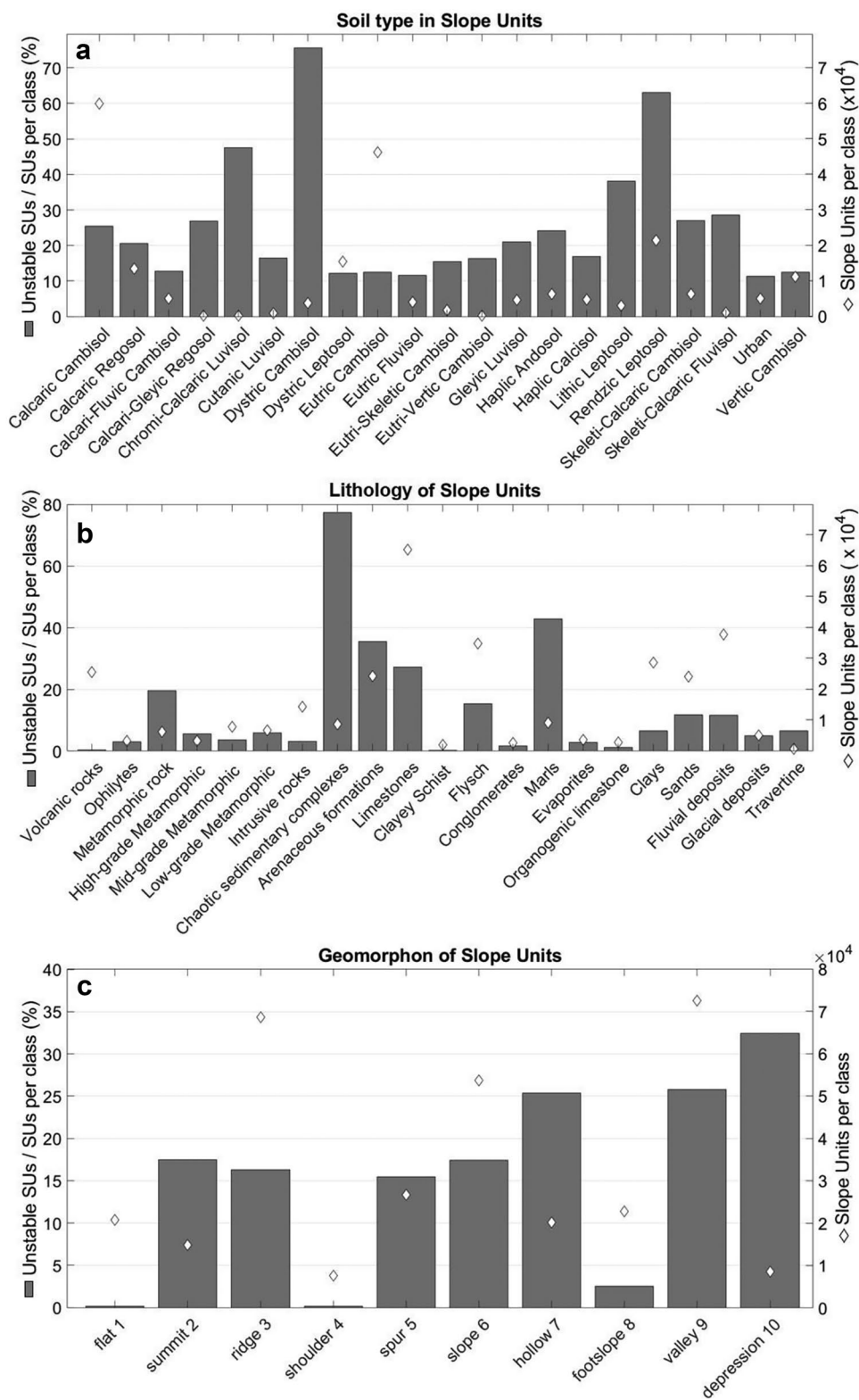


Fig. 8 Resulting scheme of the Permutation Feature Importance analysis. Predictors codes are provided in Table 1

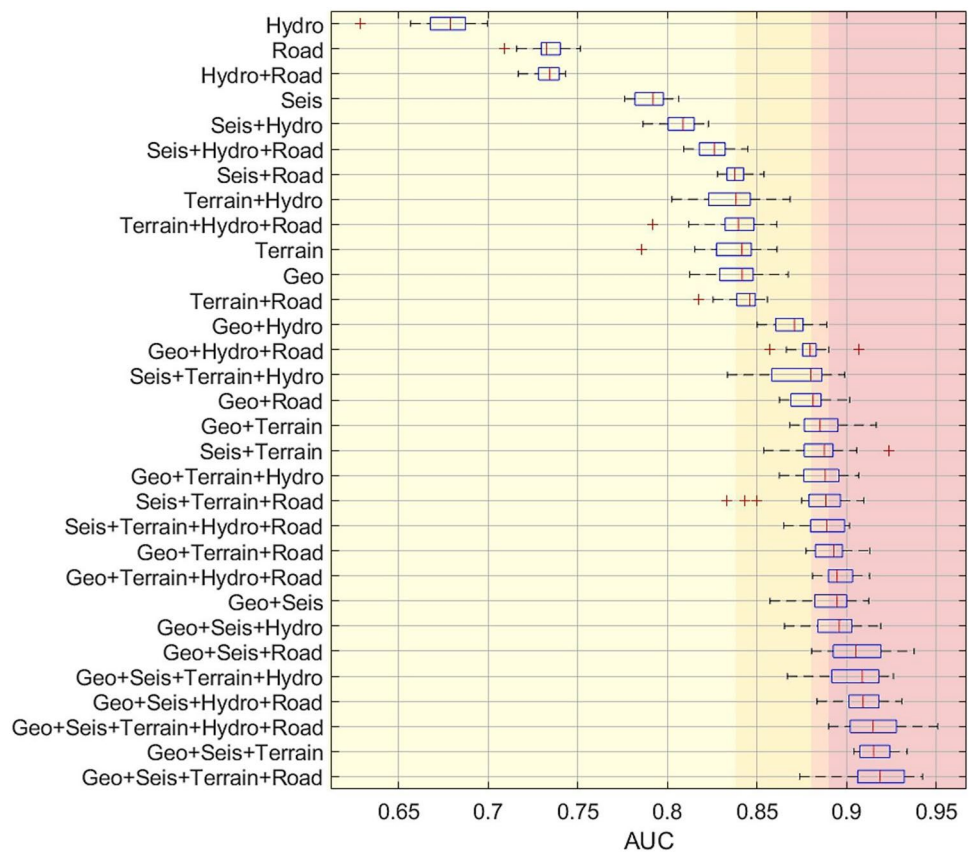
**Fig. 9** Distribution of slope units among the three geothematic variables classes: bars refer to the percentage of unstable slope units out of the total number of slope units, per class; diamonds indicate the total number of slope units, per class. **a** Refer to soil type classes. To make the chart clearer, only soil types with unstable slope units higher than 10% have been reported. **b** Refer to lithology classes and **c** to geomorphon classes



classifying stable slope units, was 0.84. Both metrics show a standard deviation lower than 0.02 after 100 replicates confirming the robustness of the classification. In particular, the classification error plot shows low standard deviations

especially for those SUs classified as extremely stable (mean susceptibility < 0.25) or unstable (mean susceptibility > 0.75), giving rise to high reliability of the final susceptibility model. These outputs fulfil the aim of the work to perform a

**Fig. 10** Box plot of the AUC distribution among 20 replicates and for all the combinations of predictors groups. Combinations are ordered by the medians of the AUC distributions. The background colour varies according to the quartiles of the distribution of the median AUCs calculated over the 20 replicates per combination. The median quartiles are at AUC values of 0.84, 0.88 and 0.89



robust susceptibility analysis of earthquake-triggered landslides at the national scale which, being trained on landslides distributed over more than one century and over the whole Italian territory, could serve as a basis to prioritise funds for remedial interventions at national to regional levels.

### EQtLs susceptibility patterns

The EQtLs susceptibility map of Italy obtained by the means of the neural network approach was compared with a landslide distribution map of Italy derived from the IFFI inventory, which reports landslides whose genesis is linked to rainfall, earthquake, snowmelt, and anthropic effects.

The comparison reveals an interesting output which regards the main distribution of earthquake-induced landslide all along a more internal portion of the Apennine Chain backbone (Fig. 12). As a result, the eastern coastal zone is less predisposed to landslide triggering due to earthquakes.

On the north, along the Alps Chain, the highest susceptibility zone corresponds to the eastern area, namely parts of Veneto and Friuli regions, where seismogenic sources are more concentrated.

The average dimensions of the chosen mapping unit, i.e. 0.7-km<sup>2</sup> slope units, provide a detailed level of spatial resolution to the susceptibility map, which represents an accurate

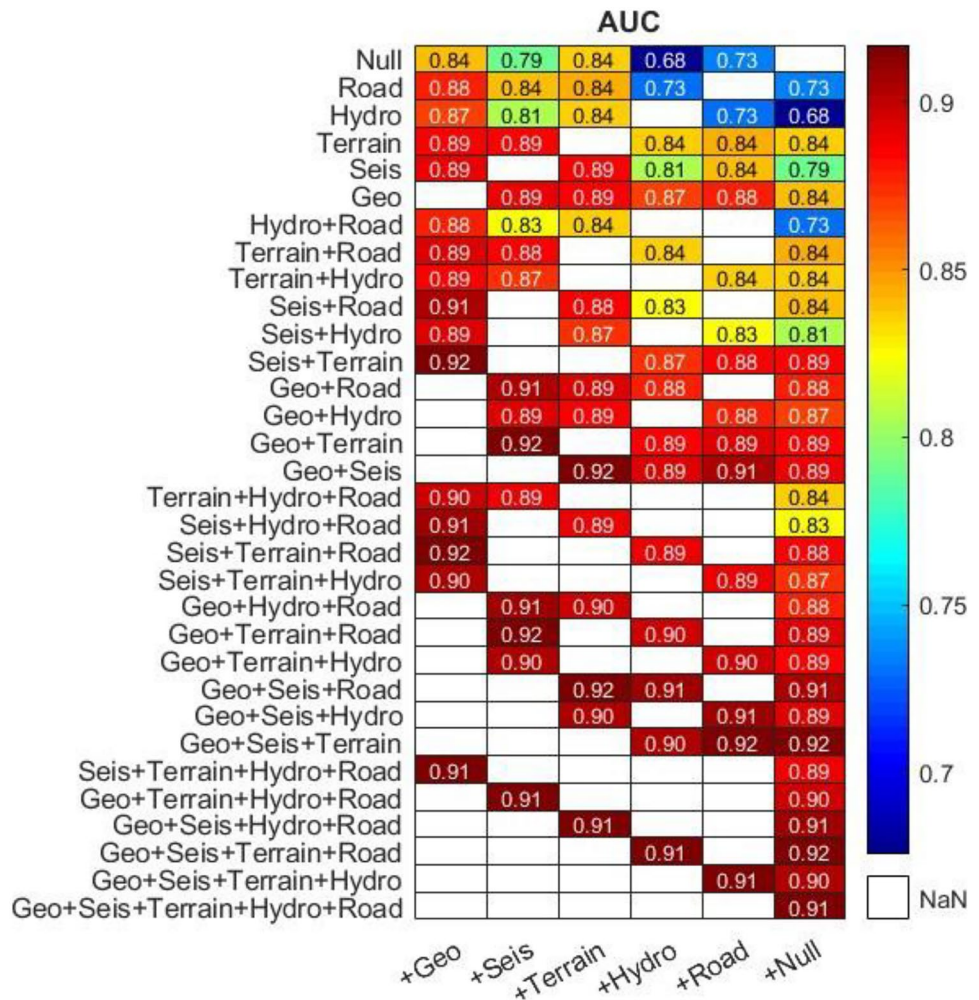
model when observed at the regional scale and clearly identifies what are the more susceptible areas with respect to the more stable ones. In those regions where the a-posteriori model check reaches high performance ( $C\text{-TPR} > 70\%$  in Fig. 7), such as Veneto, Tuscany, Friuli, Abruzzo, Emilia-Romagna, Molise, Campania, and Basilicata regions, the produced EQtLs susceptibility map can be taken as a reliable instrument to drive the decision makers toward appropriate funding management, i.e. in order to provide priority lists of local interventions. The comparison between the overall landslides density map and the here presented EQtLs susceptibility map clearly indicates that areas highly susceptible to earthquake-triggered landslides might be not taken into account in the frame of landslide mitigation National funds since not necessarily exposed to a high generic landslide hazard, e.g. rainfall-induced. Contextually, in these areas, the likely dedicated funds for earthquake risk mitigation might tend to be used primarily for building reinforcement, keeping on ignoring the significant slope stability matter demonstrated with this study.

### Predictors' role

From predictors importance analyses, Geothematic and Seismic resulted in being the most important predictors, as largely supported in the literature by other susceptibility

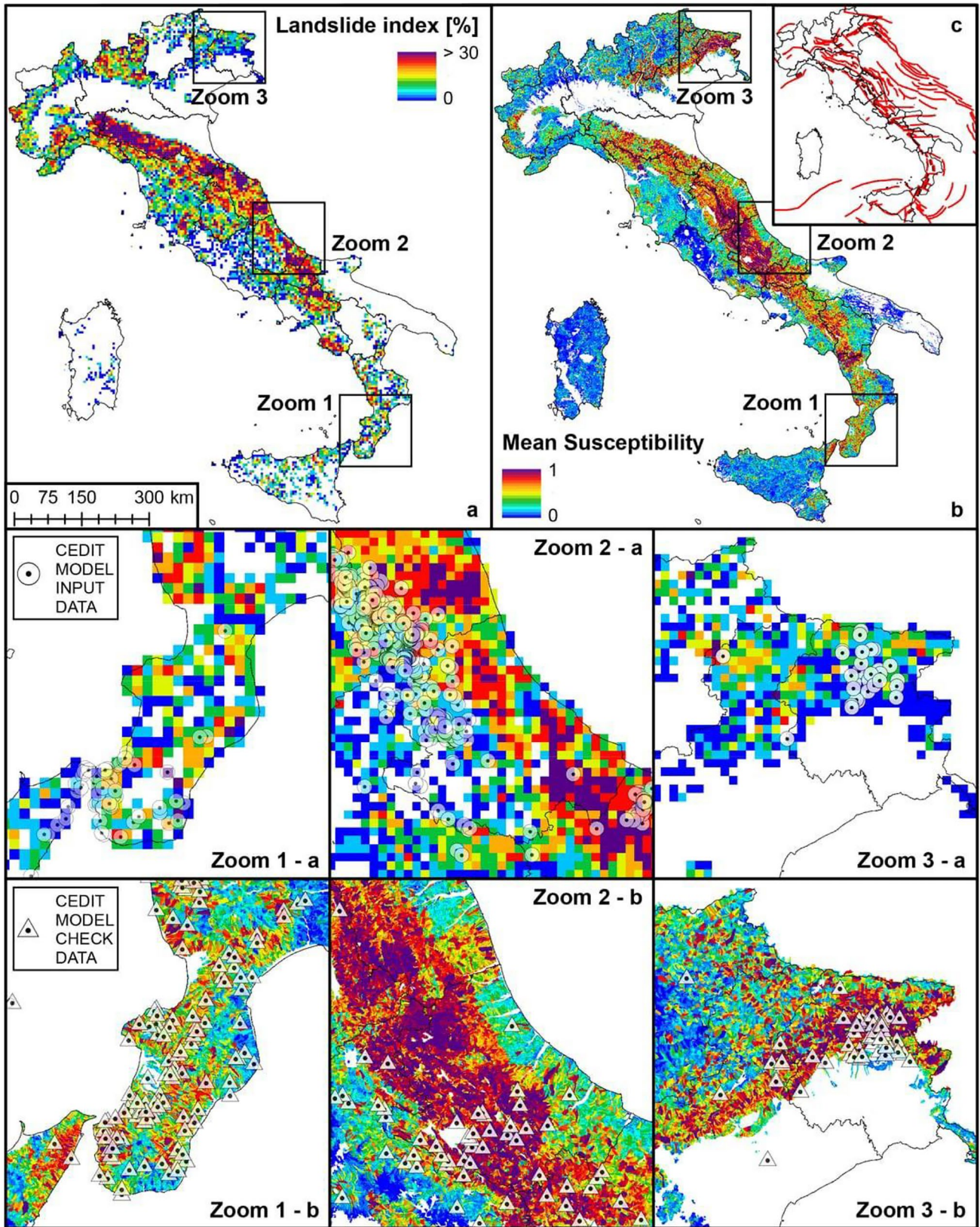


**Fig. 11** Heatmap of the mean AUC values after 20 replicates for all the combinations of predictors groups. Combinations are obtained by adding the group in the column to the combination in the row. Combinations in rows are sorted from top to bottom by the increasing number of groups. In each column, one of the five groups is present. “Null” row and “null” column indicate that none of the possible combinations has been considered and that no groups have been added, respectively



analyses of EQtLs (Fan et al. 2019; Lombardo et al. 2019; Tanyaş et al. 2019; Loche et al. 2022b). From the PFI analysis, it resulted that soil type, distance from seismogenic sources, lithology, distance from active faults, and geomorphon are the most important predictors for the network and, as consequence, for the good result of the classification between stable and unstable slope units. As one can expect from an application on earthquake-triggered landslides, distance from seismogenic sources and distance from active faults played a key role in the classification, demonstrating to well represent the slope units variability due to seismic predictors. In this regard, the seismogenic source represents the portion of a fault that is more likely to enucleate a  $M_w > 5.5$  earthquake (Basili et al. 2008; DISS-Working-Group 2018). Nevertheless, landslides can occur even along the tip portions of a fault, after  $M_w < 5.5$  seismic events or in correspondence with secondary segments; therefore, distances from both source and fault line have been considered. Taking into account seismogenic sources and active faults, the susceptibility analysis presented in this study resulted in an

inclusive model that is not bound to some specific seismic events and can be applied to the whole National territory accounting for more local variability than that provided by the national PGA. Soil type resulted to be the most important predictor from the PFI analysis. Statistically speaking, this may be partially due to the fact that slope units have been characterised on the basis of 91 different soil types, giving rise to a high, detailed, pedological variability. This represents an impressive quantity of data for the ANN to take useful information from in order to perform the classification. Related to this, lithology and geomorphon, which only count 21 and 10 classes respectively, might have provided lesser, albeit meaningful, information, that resulted as third and fifth more important predictors, respectively, among all the considered ones. In percentage, the most unstable soil categories resulted in poorly developed pedotypes, generally thin, and derived from the alteration of rocky or highly calcareous bedrocks. This result is in line with what resulted from the analysis of the instability percentage per lithology class, which shows as calcareous and arenaceous formations, besides



**Fig. 12** **a** Map of landslide density derived by the IFFI inventory. Each pixel is 5×5 km. **b** Earthquake-triggered susceptibility map of Italy produced in this study. **c** Map of active faults in Italy. Zooms of the maps in **a** and in **b** are shown in the upper and the lower bands, respectively. Circles represent the landslides of the Input dataset used in the ANN training process. Triangles represent the landslides of the checking dataset used for an ex-post evaluation of the susceptibility map

clayey and marly lithologies, are largely present within the unstable slope units. These results reflect the high abundance of disrupted failures that affect rock masses during an earthquake, like rock fall, which also represents the most numerous landslide type in the Input dataset (Fig. 2). It is particularly relevant that slope units with prevalent Chromi-calcaric Luvisol, although they are very rare in Italy, are in percentage rather unstable. They are located in the Veneto region, within a restricted area quite close to a few landslides that occurred as a consequence of an earthquake that struck the region in 1936 and had an epicentre 60 km away. In this case, a clustering effect cannot be ruled out: it may have been the earthquake, with consequent EQtLs, to occur in correspondence with areas with Chromi-calcaric Luvisol rather than the presence of this soil favouring the trigger of EQtLs.

Concerning lithology, 75% of slope units are mainly constituted by chaotic sedimentary complexes, and 50% of those composed by marls have been classified as unstable. The first lithology is composed of turbiditic sandstones and clays, locally with evaporites and limestones. It is mainly spread in central Italy, along the eastern side of the Apennine chain. The second type of rock is spread in central Italy and the north-west. Successively, 25–40% of arenaceous and limestone slope units resulted susceptible to EQtLs. Arenaceous formations crop out all over the Italian territory, mainly in mountain areas, while Limestones are spread in seismic regions of central Italy, such as Umbria and Abruzzo, in the southern part of the Alps, and along the coasts of south Italy. Moreover, an exploratory analysis of the data highlights chaotic sedimentary complexes as the lithotype is mostly prone to slope failures. This may be due to the unstructured bond between constituents, which under seismic stresses tend to generate mass movements more than in other more cohesive/massive lithologies.

Concerning geomorphon, its importance may be linked to several aspects. Intuitively, slope morphology is strictly correlated with the probability of landslide occurrence. Related to this, geomorphon classes could properly represent the most important morphological features of a slope, accounting for the contribution given by the terrain predictors to the ANN, which did not result as important as one could expect. In particular, our analysis highlights that hollow, valley, and depression slope units have been classified as more frequently

unstable than summit and ridge areas. This aspect can be partially linked with the presence of minor scarps along the slope (due to strata bedding, faults and fractures, terraces, and other natural reasons) and with the presence of roads in the lower part of the valley sides, as it occurs in many mountainous regions of Italy. Martino et al. (2019) and Tanyaş et al. (2022) found out that the presence of road cuts at the bottom of deeply incised V-shaped valleys played a conditioning role in the spatial distribution of EQtLs triggered in 2016 in Central Italy higher than road cuts located elsewhere. In this regard, from the analysis of the predictor group combination, the contribution of the distance from roads to the classification performance resulted to be not negligible. And cases of EQtLs mostly occurred along slopes modified by road cuts are documented in the literature (Keefer et al. 2006; Delgado et al. 2015). Therefore, despite the role of roadcuts in favouring EQtLs occurrence is probably more appreciable in applications in small study areas, this study also contributes to the ongoing debate drawing attention to the potential importance of this predictor.

### Opposing arguments: Validation routine through the Check data

The Italian EQtLs susceptibility map, albeit resulting from a robust iterative training-validation-test procedure, shows heterogeneous results between different regions from the comparison with the Check EQtLs dataset, which was not used to train the ANN. This can be partially due to a low number of check landslides in regions like Piedmont, Aosta Valley, Liguria, and Apulia (Fig. 7). Nevertheless, Calabria and Sicily show low C-TPR in spite of a high number of checking landslides. A reason for this could be that, as shown in zoom 1 in Fig. 12, the Input landslides from which the ANN was trained are concentrated in the area of the Strait among these two regions, while the checking EQtLs are more spread over the regional territory. This may have led to a too-low-generalised training of the network. Further, the exact location of seismogenic sources in Calabria is an argument of debate in the scientific community. Considering the importance assumed by the Seismic predictors in our analysis, their potential location's uncertainty can significantly affect the accuracy of our model.

Ultimately, an ideal model should holistically include all the factors which are known to control the slope failure mechanisms. Therefore, one should theoretically build an analogous ANN by featuring soil moisture characteristics or at least a proxy of this property, usually expressed in terms of precipitation. However, our training dataset spans multiple earthquakes, even historical ones. For this reason, we did not include any information related to soil moisture and its preparatory effect for slope failures under seismic shaking. It is our intention to fulfil this requirement for a subset of the data we considered here. Specifically, we are already

working on the earthquake sequence that affected central Italy in 2016. This case would offer a perfect stage to include soil moisture in the analyses.

### Concluding remarks

This study represents the first example of susceptibility analysis to Earthquake-triggered Landslides built at the national scale in Italy by using an ANN approach. At this aim, we exploited the CEDIT catalogue, and we implemented an ANN at the Slope Unit scale featuring predictors that take into account predisposing factors of morphological, lithological, and seismotectonic characteristics of the Italian territory. To train the ANN, a sub-dataset made of 1545 EQtLs, most accurate in terms of geolocalisation, and related to strong earthquakes from 1908 to date, was extracted from CEDIT.

The ANN was optimised and trained for the classification of Slope Units in terms of susceptibility to earthquake-triggered landslides, based on 167 predictors. The performances of the ANN have been evaluated by carrying out 100 training on independent datasets to assess its robustness. These produced a remarkable performance distinguishing between the stable and unstable SU classes. We also tested our model via an external dataset composed of 465 EQtLs, which were not featured in the model-building phase due to a lower geolocation accuracy. This ex-post verification confirmed the overall suitability of our ANN analytical protocol.

The analysis shows that a large portion of the Italian national territory is highly prone to earthquake-triggered landslides. This is especially the case throughout the

Apennine arc, where high susceptibility values are associated with more than 50% of Abruzzo, Marche, Molise, and Umbria and ~ 25% of Tuscany, Emilia-Romagna, Campania, and Basilicata regions. Furthermore, high susceptibility values are associated with approximately 25% of the territory of the Friuli region, in the eastern sector of the Alpine arc. Conversely, north-western regions, Sicily, Sardinia, and most of Latium and Apulia appear to be quite stable with minor percentages of the territory characterised by susceptible slopes.

The 0.7-km<sup>2</sup> SU mean size is extremely detailed, even for a regional or provincial scale assessment. For instance, our susceptibility map would be a perfect addition to Civil Protection in order to design emergency plans at the provincial level. Or it could be used at the level of each region to plan long-term risk mitigation actions and allocate funds. Our work demonstrates that if funds were allocated only based on rainfall-induced landslide distribution, large areas highly susceptible to EQtLs would be completely ignored by mitigation plans.

Nevertheless, the resolution of these mapping units is still far from the requirements for planning purposes, or for seismic microzonation studies, at a municipal scale.

As regards future improvements we envision for this study, the main extension consists of scaling down the model to a greater spatial resolution focusing on a specific data-rich sub-region, of particular interest. Here, we envision to run also a suite of physically based models, besides the ANN-base procedure and compare the two outputs in order to investigate potential differences, both in terms of strength and weaknesses, between each modeling routine.

### Appendix. Predictors' information

**Table 2** List of the predictors assigned to each slope unit. Codes reported in "Predictors code and description" have been used to represent the results of the permutation feature importance. Predictors

have been grouped as indicated in "Group" to perform the combination-group analysis. Terrain characteristics were calculated from a 20-m Digital Elevation Model (DEM)

Group	Code	Description
<i>Geothematic predictors</i>		
Geomorphon (10 classes)	10	Class of "geomorphon" which covers most of the area of the Slope Unit (SU)
Lithology (21 classes)	31	Lithology covering most of the SU area; it is taken from the geological map of Italy at a scale of 1:50.000 or 1:100.000
Soil type (91 classes)	31	Type of soil that covers most of the area of the SU; it is taken from the ecopedological map of Italy at a scale of 1:250.000
<i>Seismic predictors</i>		
Distance to seismic features	123	The average distance of a SU from the nearest seismogenic source
	124	The standard deviation of the distance to the nearest seismogenic source
	125	The average distance of a SU from the nearest active fault line (capable or not)
	126	The standard deviation of the distance of a SU from the nearest active fault line (capable or not)

**Table 2** (continued)

Group	Code	Description
<i>Anthropic predictors</i>		
Distance to roads	127	Count of the pixels of a SU covered by any buffer of distance from roads. The buffer ranges are 10, 50, and 100 m
	128	Sum of the buffer values of the pixels of a SU covered by any buffer of distance from roads
	129	The buffer of distance from a road that takes up most of the area of the SU
	130	The maximum value of the buffer of distance from a road within a SU
	131	The average value of the buffer of distance from a road within a SU
	132	The minimum value of the buffer of distance from a road within a SU
	133	Range (max–min) of distance values from roads included in a SU
	134	Count of the pixels of a SU that fall within 5 m of distance from roads
<i>Hydrological predictors</i>		
Distance to watercourses	135	Count of the pixels of a SU covered by any buffer of distance from rivers. The buffer ranges are 10, 50, and 100 m
	136	Sum of the buffer values of the pixels of a SU covered by any buffer of distance from rivers
	137	The buffer of distance from a river that takes up most of the area of the SU
	138	The maximum value of the buffer of distance from a river within a SU
	139	The average value of the buffer of distance from a river within a SU
	140	The minimum value of the buffer of distance from a river within a SU
	141	Range (max–min) of distance values from rivers included in a SU
	142	Count of the pixels of a SU that fall within 5 m of distance from rivers
<i>Terrain predictors</i>		
Curvature	143	The average Tangential Curvature of a SU
	144	The standard deviation of the Tangential Curvature of a SU
	145	The average Profile Curvature of a SU
	146	The standard deviation of the Profile Curvature of a SU
	147	The average Plan Curvature of a SU
	148	The standard deviation of the Plan Curvature of a SU
	149	The average Longitudinal Curvature of a SU
	150	The standard deviation of the Longitudinal Curvature of a SU
	151	The average General Curvature of a SU
	152	The standard deviation of the General Curvature of a SU
Elevation	153	The average Elevation of a SU
	154	The standard deviation of the Elevation of a SU
Exposure	155	The average Exposure of a SU from north to south
	156	The standard deviation of the Exposure of a SU, from north to south
	157	The average Exposure of a SU from east to west
Slope	158	The standard deviation of the Exposure of a SU, from east to west
	159	The average Slope of a SU
	160	The standard deviation of the Slope of a SU
Topographic Wetness Index	161	The average Topographic Wetness Index of a SU
	162	The standard deviation of the Topographic Wetness Index of a SU
Topographic Position Index	163	The average Topographic Position Index of a SU
	164	The standard deviation of the Topographic Position Index of a SU
Topographic Ruggedness Index	165	The average Topographic Ruggedness Index of a SU
	166	The standard deviation of the Topographic Ruggedness Index of a SU
Area	167	SU areal extent

**Table 3** Description of the classes of the categorical predictor “Lithology”

Lithology	Description
Volcanic rocks Ophilytes	Lavas, pyroclastic rocks, and ignimbrites Gabbri and anorthosites; Basalts, spilites, hyaloclastites; Serpentine, serpentines schists, and chloritoscists; Metabasites, eclogites, amphibolites, green stones s.l
Metamorphic rock	Granitoid gneiss
High-grade metamorphic	Acid granulites and biotitic-sillimantic granatiferous gneisses (at times with cordierite), with marbles, amphibolites
Mid-grade metamorphic	Mica schists and paragneisses, amphibolites, phyllites, quartzites, and marbles
Low-grade metamorphic	Phyllites with albitic paragneisses, porphyroids, marbles, and green schists
Intrusive rocks	Granites, granodiorites, tonalites and rare diorites
Chaotic sedimentary complexes	Sandstones (including turbidite) and clays, in places with evaporites and subordinately limestone
Arenaceous formations	Sandstones and conglomerates, sometimes turbidites
Limestones	Limestones, sometimes arenaceous, and marl metamorphosed into carbonate schists (marbles, phyllites, etc.)
Clayey schist	Clayey schist, at times carbonaceous
Flysch	Clayey and clayey-calcareous units often with turbidite character, sometimes including the lower Miocene
Conglomerates	Clastic deposits locally with marl including, at times, the Carboniferous
Marls	Pelagic facies marls, sometimes with flint
Evaporites	Chalky-sulphurous formation
Organogenic limestone	Debris and organogenic limestones, “bench” type
Clays	Clays and marls, locally with holistostromes
Sands	Conglomerates and sandstones, sometimes including the Upper Permian
Fluvial deposits	Debris accumulations, alluvial and fluviolacustri deposits, current beaches
Glacial deposits	Glacial deposits
Travertine	Travertines (sometimes Holocene)

**Table 4** Description of the reference soil groups compared in the classes of the categorical predictor “Soil type”

Soil type	WRB code	Description
Cambisols	CM	Soils with little or no profile differentiation—Moderately developed. Cambisols are developed in medium and fine-textured materials derived from a wide range of rocks, mostly in alluvial, colluvial, and aeolian deposits
Regosols	RG	Soils with little or no profile differentiation—No significant profile development. Regosols are developed in unconsolidated materials. Regosols are extensive in eroding lands, in particular in arid and semi-arid areas and in mountain regions
Luisols	LV	Soils with clay-enriched subsoil—High-activity clays, high base status. The main characteristic is an argic horizon, a subsurface zone with higher clay content than the material above it. This typically arises as clay is washed downward by water and accumulates at greater depth
Andosols	AN	Soils distinguished by Fe/Al chemistry—Allophanes or Al-humus complexes. Andosols are generally quite young soils found in volcanic areas formed in volcanic tephra. Andosols are usually defined as soils containing high proportions of glass and amorphous colloidal materials, including allophane, imogolite and ferrihydrite
Calcisols	CL	Accumulation of moderately soluble salts or non-saline substances—Accumulation of secondary carbonates. Calcisols are developed in mostly alluvial, colluvial, and aeolian deposits of base-rich weathering material. They are found on level to hilly land in arid and semi-arid regions. The natural vegetation is sparse and dominated by xerophytic shrubs and trees and/or ephemeral grasses
Leptosols	LP	Soils with limitations to root growth—Thin or with many coarse fragments. Leptosols are very shallow soils over hard rock or a deeper soil that is extremely gravelly and/or stony. Leptosols can be found on hard rocks or where erosion has kept pace with soil formation or removed the top of the soil. The very shallow, less than 10 cm deep, Lithic Leptosols in mountain regions are the most extensive Leptosols on Earth
Fluvisols	FL	Soils with little or no profile differentiation—Stratified fluvial, marine, and lacustrine sediments. Fluvisols are found on alluvial plains, river fans, valleys, and tidal marshes on all continents and in all climate zones. Under natural conditions, periodical flooding is fairly common. The soils have a clear evidence of stratification. Soil horizons are weakly developed, but a distinct topsoil horizon may be present

**Acknowledgements** Authors acknowledge the project "Distribution analysis of earthquake-induced instability effects based on a national scale inventory for the probabilistic definition of multi-hazard scenarios" (University of Rome Sapienza - RM120172A2582F52 - year 2020) supervised by Prof. Salvatore Martino.

**Data Availability** Landslide data can be found in the CEDIT catalogue, available at the link <https://gdb.ceri.uniroma1.it/index.php/view/map/?repository=cedit&project=Cedit> (as mentioned in the text)

## References

- Alvioli M, Guzzetti F, Marchesini I (2020) Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology* 107124
- Alvioli M, Marchesini I, Reichenbach P, Rossi M, Ardizzone F, Fiorucci F, Guzzetti F (2016) Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geosci Model Dev* 9(11), 3975–3991
- Amato G, Eisank C, Castro-Camilo D, Lombardo L (2019) Accounting for covariate distributions in slope-unit-based landslide susceptibility models. a case study in the alpine environment. *Eng Geol* 260, In print
- Amato G, Palombi L, Raimondi V (2021) Data-driven classification of landslide types at a national scale by using artificial neural networks. *Int J Appl Earth Obs Geoinf* 104:102549
- Avolio MV, Di Gregorio S, Lupiano V, Mazzanti P (2013) SCID-DICA-SS 3: a new version of cellular automata model for simulating fast moving landslides. *J Supercomput* 65(2):682–696
- 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(1):15–31
- Basili R, Valensise G, Vannoli P, Burrato P, Fracassi U, Mariano S, Tiberti MM, Boschi E (2008) The Database of Individual Seismogenic Sources (DISS), version 3: summarizing 20 years of research on Italy's earthquake geology. *Tectonophysics* 453(1):20–43. *Earthquake Geology: Methods and Applications*
- Beven K, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrol Sci J* 24(1):43–69
- Bird JF, Bommer JJ (2004) Earthquake losses due to ground failure. *Eng Geol* 75(2):147–179
- Bonham-Carter GF (1989) Weights of evidence modeling: a new approach to mapping mineral potential. *Stat Appl Earth Sci* 171–183
- Brenning A (2008) Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie* 19(23–32):410
- Broeckx J, Vanmaercke M, Duchateau R, Poesen J (2018) A data-based landslide susceptibility map of Africa. *Earth Sci Rev* 185:102–121
- Caprari P, Della Seta M, Martino S, Fantini A, Fiorucci M, Priore T (2018) Upgrade of the CEDIT database of earthquake-induced ground effects in Italy. *Italian Journal of Engineering Geology and Environment* 2:23–39
- Carminati E, Lustrino M, Cuffaro M, Doglioni C (2010) Tectonics, magmatism and geodynamics of Italy: what we know and what we imagine. *J Virtual Explor* 36(8):10–3809
- Carrara A (1988) Drainage and divide networks derived from high-fidelity digital terrain models. In: Quantitative analysis of mineral and energy resources, 581–597. Springer
- Carrara A, Cardinali M, Detti R, Guzzetti F, Pasqui V, Reichenbach P (1991) GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf Proc Land* 16(5):427–445
- Carrara A, Cardinali M, Guzzetti F, Reichenbach P (1995) GIS technology in mapping landslide hazard. In *Geographical information systems in assessing natural hazards*, pp 135–175. Springer
- Castro Camilo D, Lombardo L, Mai P, Dou J, Huser R (2017) Handling high predictor dimensionality in slope-unit-based landslide susceptibility models through LASSO-penalized Generalized Linear Model. *Environ Model Softw* 97:145–156
- De Reu J, Bourgeois J, Bats M, Zwertvaegher A, Gelorini V, De Smedt P, Chu W, Antrop M, De Maeyer P, Finke P et al (2013) Application of the topographic position index to heterogeneous landscapes. *Geomorphology* 186:39–49
- De Veaux RD, Ungar LH (1994) Multicollinearity: a tale of two nonparametric regressions. In: *Selecting models from data*, pp 393–402. Springer
- Del Gaudio V, Pierri P, Wasowski J (2003) An approach to time-probabilistic evaluation of seismically induced landslide hazard. *Bull Seismol Soc Am* 93(2):557–569
- Del Gaudio V, Wasowski J (2004) Time probabilistic evaluation of seismically induced landslide hazard in Irpinia (Southern Italy). *Soil Dyn Earthq Eng* 24(12):915–928
- Delgado J, García-Tortosa FJ, Garrido J, Garrido A, Loffredo A, López-Casado C, Martín-Rojas I, Rodríguez-Peces MJ (2015) Seismically induced landslides by a low-magnitude earthquake: the Mw 4.7 Ossa De Montiel event (central Spain). *Eng Geol* 196:280–285
- DISS-Working-Group (2018) Database of Individual Seismogenic Sources (DISS), version 3.2.1: a compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. <http://diss.rm.ingv.it/diss/>. Istituto Nazionale di Geofisica e Vulcanologia
- Ermini L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. *Geomorphol* 66(1–4):327–343
- Esposito C, Martino S, Pallone F, Martini G, Romeo R (2016) A methodology for a comprehensive assessment of earthquake-induced landslide hazard, with an application to pilot sites in Central Italy
- Evans IS (1980) An integrated system of terrain analysis and slope mapping. *Zeitschrift für Geomorphologie. Supplementband Stuttgart* 36:274–295
- Fan X, Scaringi G, Korup O, West AJ, van Westen CJ, Tanyas H, Hovius N, Hales TC, Jibson RW, Allstadt KE et al (2019) Earthquake-induced chains of geologic hazards: patterns, mechanisms, and impacts. *Rev Geophys*
- Finke P, Montanarella L (2001) Basic principals of the manual of procedures (version 1.1) for the georeferenced soil database. *Options Méditerranéennes: Série B. Etudes Et Recherches* 34:49–65
- Fortunato C, Martino S, Prestininzi A, Romeo R (2012) New release of the Italian catalogue of earthquake-induced ground failures (CEDIT). *Italian Journal of Engineering Geology and Environment* 2:63–75
- Frattini P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. *Eng Geol* 111(1):62–72
- Gao Y, Wang S, Guan K, Wolanin A, You L, Ju W, Zhang Y (2020) The ability of sun-induced chlorophyll fluorescence from OCO-2 and MODIS-EVI to monitor spatial variations of soybean and maize yields in the Midwestern USA. *Remote Sens* 12(7):1111
- Goetz JN, Guthrie RH, Brenning A (2011) Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphol* 129(3–4):376–386
- Gomez H, Kavzoglu T (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin. *Venezuela Eng Geol* 78(1–2):11–27
- Graser A (2016) *Learning Qgis*. Packt Publishing Ltd

- Guisan A, Theurillat J, Zimmermann N (1999) SB Weiss, and AD Weiss, 1999: GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol* 143:107–122
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study. *Central Italy Geomorphology* 31(1):181–216
- Guzzetti F, Galli M, Reichenbach P, Ardizzone F, Cardinali M (2006) Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Nat Hazard* 6(1):115–131
- Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72(1–4):272–299
- Hassoun MH et al (1995) *Fundamentals of artificial neural networks*. MIT press
- Heerdegen RG, Beran MA (1982) Quantifying source areas through land surface curvature and shape. *J Hydrol* 57(3–4):359–373
- Herrera G, Mateos RM, García-Davalillo JC, Grandjean G, Poyiadji E, Maftai R, Filipciuc T-C, Auflič MJ, Jež J, Podolszki L et al (2018) Landslide databases in the Geological Surveys of Europe. *Landslides* 15(2):359–379
- Hosmer DW, Lemeshow S (2000) *Applied logistic regression*, 2nd edn. Wiley, New York
- Hsieh S-Y, Lee C-T (2011) Empirical estimation of the Newmark displacement from the Arias intensity and critical acceleration. *Eng Geol* 122(1–2):34–42
- Huang F, Yin K, Huang J, Gui L, Wang P (2017) Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Eng Geol* 223:11–22
- Jacek S (1997) Landform characterization with geographic information systems. *Photogramm Eng Remote Sens* 63(2):183–191
- Jasiewicz J, Stepinski TF (2013) Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology* 182:147–156
- Jenness J (2006) Topographic position index (tpi jen. avx) extension for ArcView 3. x, v. a. Jenness Enterprises
- Jibson RW (2007) Regression models for estimating coseismic landslide displacement. *Eng Geol* 91(2–4):209–218
- Jibson RW, Harp EL, Michael JA (2000) A method for producing digital probabilistic seismic landslide hazard maps. *Eng Geol* 58(3–4):271–289
- Keefer DK, Wartman J, Ochoa CN, Rodriguez-Marek A, Wiczorek GF (2006) Landslides caused by the M 7.6 Tecoman, Mexico earthquake of January 21, 2003. *Eng Geol* 86(2–3):183–197
- Lima P, Steger S, Glade T, Tilch N, Schwarz L, Kociu A (2017) Landslide susceptibility mapping at national scale: a first attempt for Austria. In: *Workshop on World Landslide Forum*, pp 943–951
- Liu Z, Shao J, Xu W, Chen H, Shi C (2014) Comparison on landslide nonlinear displacement analysis and prediction with computational intelligence approaches. *Landslides* 11(5):889–896
- Loche M, Alvioli M, Marchesini I, Bakka H, Lombardo L (2022a) Landslide susceptibility maps of Italy: lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. *Earth Sci Rev* 104125
- Loche M, Scaringi G, Yunus AP, Catani F, Tanyas H, Frodella W, Fan X, Lombardo L (2022) Surface temperature controls the pattern of post-earthquake landslide activity. *Sci Rep* 12(1):988
- Lombardo L, Bakka H, Tanyas H, van Westen C, Mai PM, Huser R (2019) Geostatistical modeling to capture seismic-shaking patterns from earthquake-induced landslides. *J Geophys Res Earth Surf* 124(7):1958–1980
- Lombardo L, Cama M, Marcker, M. and Rotigliano, E. (2014) A test of transferability for landslides susceptibility models under extreme climatic events: application to the Messina 2009 disaster. *Nat Hazards* 74(3):1951–1989
- Lombardo L, Mai PM (2018) Presenting logistic regression-based landslide susceptibility results. *Eng Geol* 244:14–24
- Lombardo L, Opitz T, Ardizzone F, Guzzetti F, Huser R (2020a) Space-time landslide predictive modelling. *Earth Sci Rev* 103318
- Lombardo L, Saia S, Schillaci C, Mai PM, Huser R (2018) Modeling soil organic carbon with Quantile Regression: dissecting predictors' effects on carbon stocks. *Geoderma* 318:148–159
- Lombardo L, Tanyas H (2020) Chrono-validation of near-real-time landslide susceptibility models via plug-in statistical simulations. *Eng Geol* 278:105818
- Lombardo L, Tanyas H, Nicu IC (2020b) Spatial modeling of multi-hazard threat to cultural heritage sites. *Eng Geol* 105776
- Lupiano V, Machado GE, Molina LP, Crisci GM, Di Gregorio S (2018) Simulations of flow-like landslides invading urban areas: a cellular automata approach with SCIDDICA. *Nat Comput* 17(3):553–568
- Mantovani M, Bossi G, Marcato G, Schenato L, Tedesco G, Titti G, Pasuto A (2019) New perspectives in landslide displacement detection using sentinel-1 datasets. *Remote Sensing* 11(18):2135
- Marjanović M, Kovačević M, Bajat B, Voženilek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. *Eng Geol* 123(3):225–234
- Martino S (2016) Earthquake-induced reactivation of landslides: recent advances and future perspectives. In: *Earthquakes and their impact on society*, pp 291–322. Springer
- Martino S, Antonielli B, Bozzano F, Caprari P, Discenza M, Esposito C, Fiorucci M, Iannucci R, Marmoni G, Schilir'ò L (2020a) Landslides triggered after the 16 August 2018 Mw 5.1 Molise earthquake (Italy) by a combination of intense rainfalls and seismic shaking. *Landslides* 1–14
- Martino S, Battaglia S, Delgado J, Esposito C, Martini G, Missori C (2018) Probabilistic approach to provide scenarios of earthquake-induced slope failures (PARSIFAL) applied to the Alcoy Basin (south Spain). *Geosci* 8(2):57
- Martino S, Bozzano F, Caporossi P, D'angiò D, Della Seta M, Esposito C, Fantini A, Fiorucci M, Giannini L, Iannucci R et al (2019) Impact of landslides on transportation routes during the 2016–2017 Central Italy seismic sequence. *Landslides* 16(6):1221–1241
- Martino S, Bozzano F, Paolo C, Danilo D, Della Seta M, Carlo E, Andrea F, Matteo F, Giannini LM, Roberto I et al (2017) Ground effects triggered by the 24th August 2016, Mw 6.0 Amatrice (Italy) earthquake. surveys and inventoring to update the CEDIT catalogue. *Geografia Fisica e Dinamica Quaternaria* 40(1):77–95
- Martino S, Caprari P, Fiorucci M, Marmoni G (2020b) The CEDIT Catalogue: from inventoring of earthquake-induced ground effects to analysis of scenario. *Mem Descr Carta Geol D'it* 107:441–450
- Martino S, Prestininzi A, Romeo R (2014) Earthquake-induced ground failures in Italy from a reviewed database. *Nat Hazard* 14(4):799
- McElroy TS, Jach A (2019) Testing collinearity of vector time series. *Economet J* 22(2):97–116
- Ohlmacher GC (2007) Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Eng Geol* 91(2):117–134
- Patel V, Sotiropoulos F (1997) Longitudinal curvature effects in turbulent boundary layers. *Prog Aerosp Sci* 33(1–2):1–70
- Petley D (2012) Global patterns of loss of life from landslides. *Geology* 40(10):927–930
- Prestininzi A, Romeo R (2000) Earthquake-induced ground failures in Italy. *Eng Geol* 58(3–4):387–397
- Putin E, Mamoshina P, Aliper A, Korzinkin M, Moskalev A, Kolosov A, Ostrovskiy A, Cantor C, Vijg J, Zhavoronkov A (2016) Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging (albany NY)* 8(5):1021
- Rahmati O, Kornejady A, Samadi M, Deo RC, Conoscenti C, Lombardo L, Dayal K, Taghizadeh-Mehrjardi R, Pourghasemi HR, Kumar S et al (2019) PMT: New analytical framework for automated



- evaluation of geo-environmental modelling approaches. *Sci Total Environ* 664:296–311
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60–91
- Riley SJ, DeGloria SD, Elliot R (1999) Index that quantifies topographic heterogeneity. *Int J Sci* 5(1–4):23–27
- Romeo R (2000) Seismically induced landslide displacements: a predictive model. *Eng Geol* 58(3–4):337–351
- Rossi M, Reichenbach P (2016) LAND-SE: a software for statistically based landslide susceptibility zonation, version 1.0. *Geosci Model Dev* 9(10):3533
- Sassa K (1996) Prediction of earthquake induced landslides. In: *Landslides*, pp 115–132
- Schlögel R, Marchesini I, Alvioli M, Reichenbach P, Rossi M, Malet JP (2018) Optimizing landslide susceptibility zonation: effects of DEM spatial resolution and slope unit delineation on logistic regression models. *Geomorphol* 301:10–20
- Shrestha S, Kang T-S (2019) Assessment of seismically-induced landslide susceptibility after the 2015 Gorkha earthquake. *Nepal Bulletin of Engineering Geology and the Environment* 78(3):1829–1842
- Song Y, Gong J, Gao S, Wang D, Cui T, Li Y, Wei B (2012) Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan, China. *Comput Geosci* 42:189–199
- Steger S, Schmaltz E, Glade T (2020) The (f) utility to account for pre-failure topography in data-driven landslide susceptibility modelling. *Geomorphology* 354:107041
- Tacchia D, Masella G, Pannuti V, Vitale V (2005) La nuova Carta Geologica d'Italia scala 1:1,000,000. In: *Atti della 9 Conferenza Nazionale ASITA*, volume 15, p 18
- Tanyaş H, van Westen C, Allstadt K, Nowicki AJM, Görüm T, Jibson R, Godt J, Sato H, Schmitt R, Marc O, Hovius N (2017) Presentation and analysis of a worldwide database of earthquake-induced landslide inventories. *J Geophys Res Earth Surf* 122(10):1991–2015
- Tanyaş H, Görüm T, Kirschbaum D, Lombardo L (2022) Could road constructions be more hazardous than an earthquake in terms of mass movement? *Nat Hazards* 112(1):639–663
- Tanyaş H, Lombardo L (2020) Completeness index for earthquake-induced landslide inventories. *Eng Geol* 264:105331
- Tanyaş H, Rossi M, Alvioli M, van Westen CJ, Marchesini I (2019) A global slope unit-based method for the near real-time prediction of earthquake-induced landslides. *Geomorphology* 327:126–146
- Titti G, Napoli GN, Conoscenti C, Lombardo L (2022a) Cloud-based interactive susceptibility modeling of gully erosion in google earth engine. *Int J Appl Earth Obs Geoinf* 115:103089
- Titti G, Sarretta A, Lombardo L, Crema S, Pasuto A, Borgatti L (2022b) Mapping susceptibility with open-source tools: a new plugin for QGIS. *Front Earth Sci* 229
- Titti G, van Westen C, Borgatti L, Pasuto A, Lombardo L (2021) When enough is really enough? on the minimum number of landslides to build reliable susceptibility models. *Geosciences* 11(11):469
- Trigila A, Frattini P, Casagli N, Catani F, Crosta G, Esposito C, Iadanza C, Lagomarsino D, Mugnozza GS, Segoni S et al (2013) Landslide susceptibility mapping at national scale: the Italian case study. In: *Landslide science and practice*, pp 287–295. Springer
- Umar Z, Pradhan B, Ahmad A, Jebur MN, Tehrany MS (2014) Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *CATENA* 118:124–135
- Van Westen C (2002) Use of weights of evidence modeling for landslide susceptibility mapping. *International Institute for Geoinformation Science and Earth Observation: Enschede, The Netherlands* p 21
- Van Westen C, Van Asch TW, Soeters R (2006) Landslide hazard and risk zonation—why is it still so difficult? *Bull Eng Geol Env* 65(2):167–184
- Varnes and the IAEG Commission on Landslides and Other Mass-Movements (1984) *Landslide hazard zonation: a review of principles and practice*. Natural Hazards, Series. Paris: United Nations Economic, Scientific and cultural organization. UNESCO 3:63
- Wang Y, Fang Z, Hong H (2019) Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Sci Total Environ* 666:975–993
- Wasowski J, Del Gaudio V (2000) Evaluating seismically induced mass movement hazard in Caramanico Terme (Italy). *Eng Geol* 58(3–4):291–311
- Wood J (1996) The geomorphological characterisation of digital elevation models. Ph.D. thesis, University of Leicester
- Zevenbergen LW, Thorne CR (1987) Quantitative analysis of land surface topography. *Earth Surf Proc Land* 12(1):47–56
- Zhou C, Yin K, Cao Y, Ahmed B (2016) Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. *Eng Geol* 204:108–120
- Zhou C, Yin K, Cao Y, Ahmed B, Li Y, Catani F, Pourghasemi HR (2018) Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China. *Comput Geosci* 112:23–37
- Zhou S, Fang L (2015) Support vector machine modeling of earthquake-induced landslides susceptibility in central part of Sichuan province. *China Geoenvironmental Disasters* 2(1):2
- Zhu X, Xu Q, Tang M, Nie W, Ma S, Xu Z (2017) Comparison of two optimized machine learning models for predicting displacement of rainfall-induced landslide: a case study in Sichuan Province, China. *Eng Geol* 218:213–222

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.