Nicola Nocentini ⓘ · **Ascanio Rosi** ⓘ · **Luca Piciullo** ⓘ · **Zhongqiang Liu** ⓘ ·
Samuele Segoni ⓘ · **Riccardo Fanti** ⓘ

Regional-scale spatiotemporal landslide probability assessment through machine learning and potential applications for operational warning systems: a case study in Kvam (Norway)

Abstract The use of machine learning models for landslide susceptibility mapping is widespread but limited to spatial prediction. The potential of employing these techniques in spatiotemporal landslide forecasting remains largely unexplored. To address this gap, this study introduces an innovative dynamic (i.e., space–time-dependent) application of the random forest algorithm for evaluating landslide hazard (i.e., spatiotemporal probability of landslide occurrence). An area in Norway has been chosen as the case study because of the availability of a comprehensive, spatially, and temporally explicit rainfall-induced landslide inventory. The applied methodology is based on the inclusion of dynamic variables, such as cumulative rainfall, snowmelt, and their seasonal variability, as model inputs, together with traditional static parameters such as lithology and morphologic attributes. In this study, the variables' importance was assessed and used to interpret the model decisions and to verify that they align with the physical mechanism responsible for landslide triggering. The algorithm, once trained and tested against landslide and non-landslide data sampled over space and time, produced a model predictor that was subsequently applied to the entire study area at different times: before, during, and after specific landslide events. For each selected day, a specific and space–time-dependent landslide hazard map was generated, then validated against field data. This study overcomes the traditional static applications of machine learning and demonstrates the applicability of a novel model aimed at spatiotemporal landslide probability assessment, with perspectives of applications to early warning systems.

Keywords Landslides · Spatiotemporal prediction · Landslide hazard maps · Warning · Machine learning · Random forest

Introduction

Machine learning (ML) is a subset of artificial intelligence that focuses on developing algorithms that enable computers to learn from data, recognize patterns, and make decisions or predictions based on that learned information (Mitchell 1997; Bishop and Nasrabadi 2006; Hastie et al. 2001). It has gained popularity in various applications, including geological hazard, in particular for assessing landslide susceptibility maps (LSMs) (Catani et al. 2013; Reichenbach et al. 2018; Crawford et al. 2021; Tehrani et al. 2022; Merghadi et al., 2020; Pham et al. 2016).

The physical mechanism of landslide triggering is very complex and influenced by several geological, hydrological, climatic, and anthropogenic factors. Physically based models simulate the slope failure mechanism through rigorous mathematical equations but face difficulties in handling the spatial variability of geotechnical and hydrogeological soil properties over large areas, remaining applicable only at the slope scale (Vannocci et al. 2022; Alvioli and Baum 2016; Tran et al., 2018; Corominas et al. 2014). In contrast, most Landslide Early Warning Systems (LEWSs) are based on rainfall thresholds (Guzzetti et al. 2020), which are defined as a rainfall value beyond which landslides are expected to occur (Guzzetti et al. 2008; Segoni et al. 2018a; Piciullo et al. 2018). The strength of rainfall thresholds lies in their simplicity; in fact, they are typically based only on a single parameter, rainfall. Although physically based approaches are more accurate, rainfall thresholds are fast and sufficiently accurate for regional-scale predictions (Piciullo et al. 2018) and can easily be understood and implemented for operational warning purposes (Nocentini et al. 2023a). Among statistical models, ML algorithms can unravel the complexities behind landslide triggering over large areas by using vast datasets involving geological, geomorphological, meteorological, and other relevant variables. They can identify patterns, interplaying relationships, and non-linear trends among the data that escape traditional empirical models, such as rainfall thresholds (Wang et al. 2023). The ability of ML to process extensive data and discern intricate interactions has brought a paradigm shift in the field of landslide analysis (Reichenbach et al. 2018; Tehrani et al. 2022). Another advantage of the most sophisticated ML techniques is the possibility of estimating the importance that an input variable has in producing the model outcome. In this framework, there are various indices to estimate the variables' importance, such as the out-of-bag error (OOBE) and partial dependence plots (PDPs) (Friedman 2001). These indices are useful for exploring the logic of the algorithm and verifying its reliability (Nocentini et al. 2023b).

LSMs express only where landslides are likely to occur without considering the probability of occurrence over time (Fell et al. 2008). The extension of the ML application framework toward dynamic (i.e., space–time-dependent) predictions remains largely unexplored and limited to relatively few preliminary studies (Tehrani et al. 2022; Nocentini et al. 2023b; Mondini et al. 2023). Some attempts have been carried out to combine LSMs with rainfall thresholds to develop a hazard matrix, enabling the spatial and temporal definition of landslide occurrence (Segoni et al. 2018b; Park et al. 2019; Pecoraro and Calvello 2021; Palau et al. 2022). Other authors have applied various ML models for a single rainfall

event-based (Ng et al. 2021; Liu et al. 2021) or single earthquake event-based (Pyakurel et al. 2021; Dahal and Lombardo 2023) landslide inventory, using the associated dynamic variables, to backanalyze such events. An innovative methodology for the dynamic application of the Artificial Neural Networks algorithm was presented by Distefano et al. (2022). This approach aims to automatically identify intensity-duration rainfall thresholds by feeding the model with a spatially and temporally explicit landslide inventory (a landslide dataset in which the location and time of occurrence of each landslide are known). Nocentini et al. (2023b) also proposed an innovative approach for the dynamic random forest (RF) model run. The authors sampled landslide and non-landslide events over time and space and used static and dynamic variables (e.g., total rainfall over increasing time intervals) to feed the model. They identified short and intense rainfalls, and their seasonal variability, as the most influential parameters in triggering landslides. Similarly, Steger et al. (2023), using the generalized additive mixed model, concluded that the seasonal variability of rainfall and vegetation plays a crucial role in triggering landslides. To the best of our knowledge, only a few authors have trained ML models incorporating dynamic parameters to generate Landslide Hazard Maps (LHMs). These maps indicate the probability of landslide occurrence within a specific period of time and within a given area (Varnes and IAEG 1984). For instance, Stanley et al. (2021) selected non-landslide cells over both space and time, enabling the integration of snow water equivalent and soil moisture content data as dynamic input parameters for an eXtreme Gradient Boosting model, mapping the landslide hazard on a global scale. Li et al. (2022) followed a similar approach, focusing on analyzing the influence of dynamic input parameters (cumulative rainfall and vegetation indexes) on landslide triggering. These works are innovative breakthroughs compared to static LSMs and represent a promising starting point toward the objective of a real-time ML application for spatiotemporal landslide forecasting.

This study introduces an innovative and dynamic approach for generating LHMs using a random forest algorithm within MATLAB software code (MathWorks version R2023a, TreeBagger object of Statistics and Machine Learning Toolbox™). The study area encompasses the locality of Kvam in Norway, which experienced two major rainfall events in 2011 and 2013, resulting in more than 100 landslides (Schilirò et al. 2021; Liu et al. 2021). Following the methodology proposed by Nocentini et al. (2023b), the model was trained and tested using a spatially and temporally explicit landslide inventory including different time intervals of cumulative rainfall and snowmelt as dynamic factors. The ability of the algorithm in simulating the landslide triggering mechanism was verified by analyzing the variables' importance estimates through OOBE and PDPs. This study aims to delineate a procedure for the dynamic application of the RF model to generate LHMs easily implementable in a LEWS for real-time spatiotemporal landslide forecasting.

Study area

The study area occupies most of the Gudbrandsdalen Valley and includes the municipalities of Nord-Fron, Sel and Sor-Fron (Innlandet County, southeastern Norway), for a total of 2800 km². The town of Kvam is situated at the junction between Gudbrandsdalen Valley and Veikledalen Valley (Fig. 1a). In 2011 and 2013, Kvam experienced hundreds of landslides triggered by extreme precipitation (Heyerdahl and Høydal 2017; Solheim et al. 2022) (Fig. 1b). In particular, a rainfall event on 10/06/2011 with a daily amount of about 80 mm triggered several channelized debris flows, causing damages estimating in 100 million $. On the 23/05/2013, a rainfall event with about 87 mm of daily amount caused several shallow landslides and an estimated damage of about 170 million $ (Schilirò et al. 2021; Liu et al. 2021).

The area is covered by 80% of forests and 6% of agricultural lands, while anthropic lands, water bodies and bare rocks account for the remaining 14%. Gudbrandsdalen is a long U-shaped glacier valley incised by the Gudbrandsdalslågen River during the last glaciation throughout the Quaternary (Johnsen 2010). The valley is characterized by steep slopes and a floor mantled by glaciofluvial deposits (or till), locally covered by Holocene fluvial deposits (Letten and Blikra 2007), affected by numerous small gullies and erosion tracts that frequently evolve into landslides (Oguz et al. 2020). Kvam is built on a large alluvial fan at the head of Veikledalen. The bedrock is composed of metamorphic rocks, mainly amphibolite and green or gray shists (Ramberg et al., 2008; Slagstad et al. 2011).

The area has a subarctic climate with cold winters (minimum temperature about -20 °C) and higher temperatures during summers (maximum temperature about 20 °C) (Hanssen-Bauer et al. 2017). In Kvam, the annual precipitation amounts to around 500 mm, with approximately 40% of it falling as snow (Liu et al. 2021).

Landslide inventory

An initial landslide database was assembled by NGI (Norwegian Geotechnical Institute) conducting on-site surveys and using high-resolution aerial orthophotos captured in the weeks following the events on 10/06/2011, and 23/05/2013. Consequently, the database provides the exact dates and locations of each landslide. These events were triggered by the infiltration of intense, short-duration rainfall, resulting in the immediate erosion of unconsolidated glacial deposits and in the triggering of several debris flows, which remained confined into existing tributary rivers and streams (Heyerdahl and Høydal 2017; Liu et al. 2021; Schilirò et al. 2021).

The landslides database was extended by adding historical landslide data from Norwegian National Landslide Inventory (NNLI) (https://temakart.nve.no/tema/skredhendelser, last accessed on 10 September 2022). NNLI contains a total of more than 65,000 mass movements reported in Norway since 1959, including information about the uncertainties related to the date and location of each event (Krøgli et al. 2018; Herrera et al. 2018). To meet the levels of the spatial and temporal resolution of the explanatory variables (see the "Input variables" section), a filtering process was implemented to exclude landslide events with a spatial and temporal uncertainty higher than 1000 m and 1 day, respectively. NNLI includes various types of mass movements (rock falls, rock slides, snow avalanches, and debris slides are the most frequent), and among them, only rainfall-induced landslides were selected.

A total of 373 landslides were identified in the period 2010–2022 (Fig. 1a), of these 166 from the NNLI and 207 from the NGI datasets; 144 of these occurred on 10/06/2011 and 69 on 23/05/2013. The collected landslide inventory exhibits evident spatial and temporal bias because most of the landslides that it contains occurred during the two major events (June 2011 and May 2013) and are located mainly
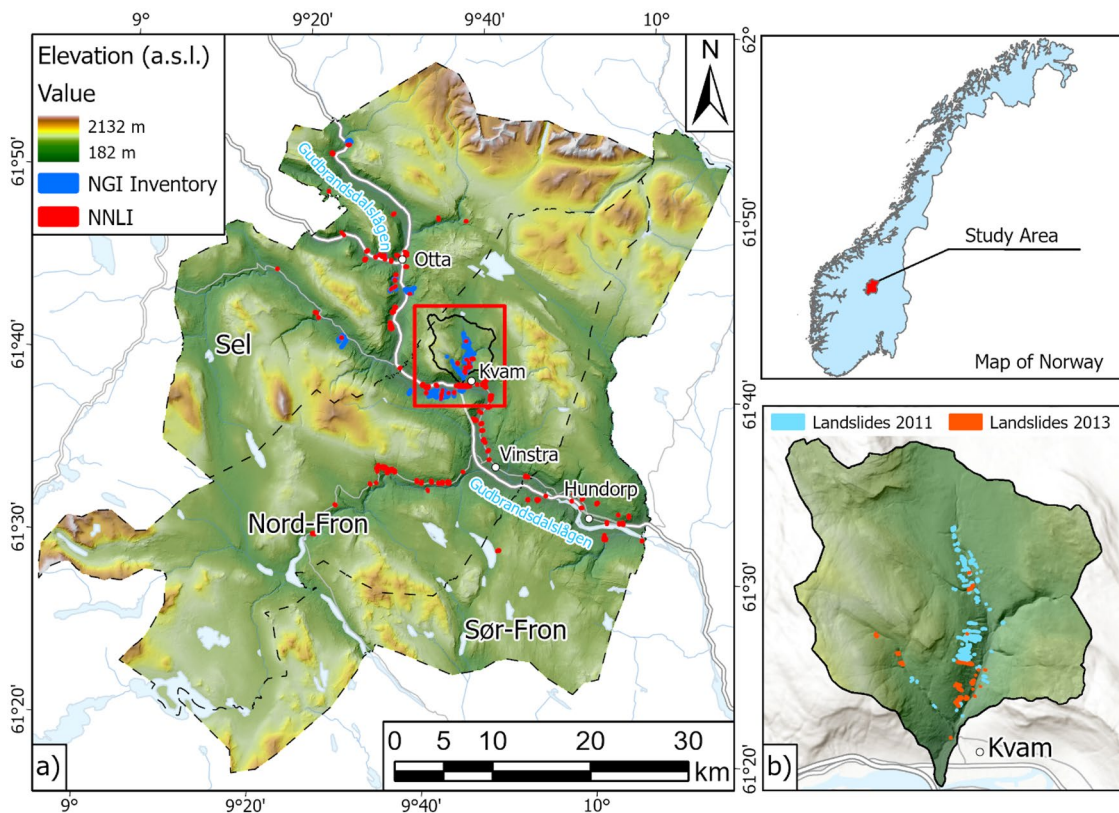
Fig. 1 Framing of the study area: **a** elevation, with NNLI (blue) and NGI inventory (red) overlaid, and a red box indicating the Kvam catchment area magnified in **b**, where 10/06/2011 (light blue) and 23/05/2013 (orange) landslide events are shown

in the same area (Kvam catchment area). Training a ML algorithm with such biased data could potentially affect the results, as it would assign greater importance to the rainfall events that occurred on those specific days and in that specific area. From a physical perspective, it is also justifiable for the model to prioritize these intense and spatially concentrated rainfall events that triggered hundreds of landslides in an already highly susceptible area. Consequently, to comprehensively analyze each case, two tests were conducted:

- Using the complete landslide inventory, thus considering landslides that occurred on the same day and in the same area as separate events (single landslides, SL), resulted in a total of 373 landslides;
- Grouping the landslide events that occurred on the same day and within the same 1 km pixel; thereby considering the Landslide Events (LE—a group of landslides occurred during the same rainfall event (Calvello and Piciullo 2016)), resulting in a total of 164 events.

Methods

The RF algorithm, a nonparametric and multivariate ML method proposed by Breiman (2001), has been extensively applied to evaluate landslide susceptibility (Brenning 2005; Catani et al. 2013; Goetz et al. 2015). Its popularity is attributed to several advantages. Specifically, it can handle both numerical and categorical data without requiring assumptions regarding their statistical distribution. Additionally, RF allows to conduct

an analysis of the input variables' importance through OOBE and PDPs. RF requires dividing the input database into two subsets: the training dataset, which is used to train the algorithm, developing the model predictor; and the test dataset, for which the predictor is applied to evaluate its performance. The conventional procedure for defining traditional LSMs involves sampling input variables only over the space, from both landslide and non-landslide points. The resulting map represents the susceptibility to landslide, without temporal information, except for the assumption that its validity can be extended in the future as long as the input variables remain constant over time. Dynamic parameters, such as cumulative rainfall, cannot be directly used as model input parameters, because of their incompatibility with such static approach.

In this Chapter, the procedures for the dynamic application of the RF model (the "Dynamic random forest model" section), developed in three subsequent steps (the "Training and testing phases," "Variables' importance analysis," and "Dynamic random forest application" sections), and the method for validating the resulting LHMs (the "Validation procedure (DTVT)" section) are presented. Figure 2 provides a schematic summary of all the stages involved in the proposed method.

Dynamic random forest model

Training and testing phases

Building on the previous work of Nocentini et al. (2023b), the approach used in this study requires a spatially and temporally
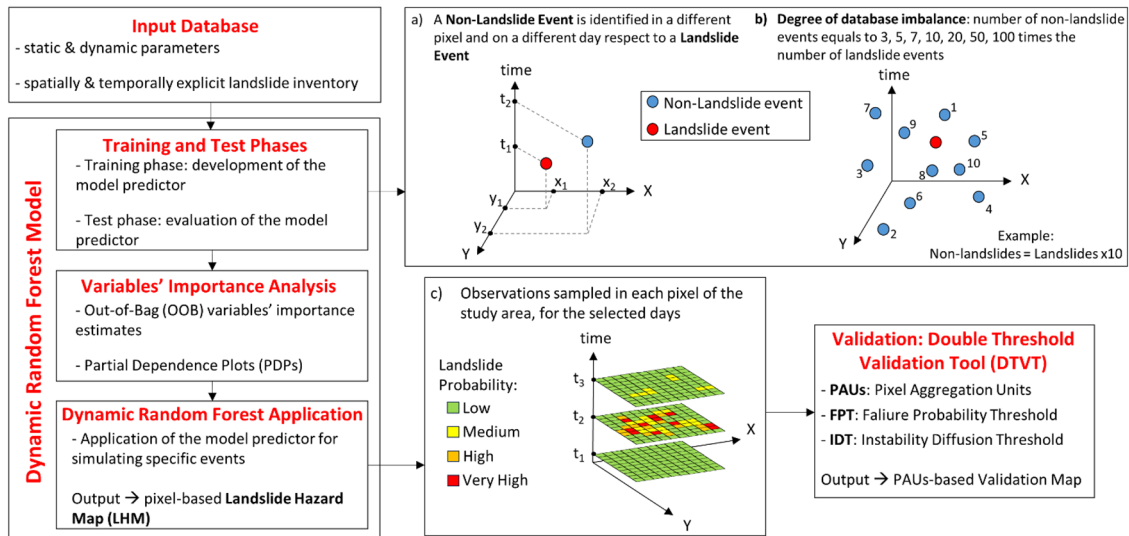
Fig. 2 Workflow of the proposed methodology with schematic examples of **a** the method of identification of non-landslide events for dynamic model training and testing phases; **b** the definition of various datasets based on the number of non-landslide events to calibrate the degree of database imbalance; and **c** the output produced by dynamic RF model application

explicit landslide inventory and involves defining non-landslide events in terms of location and date. Non-landslide events were identified through a random sampling of landslide-free pixels over both space and time (Fig. 2a). This method allows for the inclusion of dynamic parameters among the input variables to achieve spatiotemporal predictions.

To investigate the sensitivity of the results, the model was trained using different configurations based on the augmentation of non-landslide events (i.e., by randomly sampling an increasing number of non-landslide events over space and time). Therefore, we built a balanced dataset (1:1 ratio between landslide and non-landslide events) and several imbalanced datasets, with the number of non-landslide events equals to 3, 5, 7, 10, 20, 50, and 100 times the number of landslide events. The symbology $\times 1, \times 3, \times 5, \times 7, \times 10, \times 20, \times 50$, and $\times 100$ indicates the degree of database imbalance, representing the number of non-landslide events as a multiple of landslide events (Fig. 2b).

The dynamic RF model training phase was conducted by building 300 trees, as preliminary tests showed that this dimension ensures stable OOBE outputs. Each model configuration was run seven times to observe the variability of the results and to characterize the mean estimated variables' importance and its range of variation.

Variables' importance analysis

RF generates Bayesian trees by randomly sampling observations from the training dataset. This procedure consists in excluding one variable, called Out-Of-Bag (OOB) variable. Then, other trees are built using the same procedure, but randomly permuting the OOB variable. Therefore, the model calculates the OOBE for each tree, which expresses the potential error that would be committed if a given OOB variable would be excluded from the model (Catani et al. 2013). This technique is internally used by the RF algorithm to select the trees with the highest performance (lower OOBE). In our model

implementation, OOBE values are used to assess the importance of each variable, to estimate their predictive power, and to rank them based on their influence in the landslides triggering process (Liaw and Wiener 2002; Catani et al. 2013; Nocentini et al. 2023b).

PDPs are another powerful tool implemented in this study. PDPs are graphs depicting how a value or a class of values of a selected variable influences the model outcome. For instance, they can highlight direct or inverse correlations between the values of the explanatory variable and landslide probability, or discern more complex patterns (Friedman 2001; Friedman and Popescu 2008); thus, constituting a valuable tool for interpreting the results and determining whether the empirical outcomes of the algorithm are coherent with the current knowledge regarding the landslides triggering processes (Nocentini et al. 2023b). The partial dependence is calculated by measuring the change in predictions made by varying the values of a selected feature while keeping other features constant. We used the PDPs in various ways to analyze how input parameters influence landslide triggering as the degree of database imbalance increases and to explore the interactions among variables and landslide occurrences.

Dynamic random forest application

To generate LHMs and to demonstrate the applicability of the proposed dynamic RF model for real-time landslides forecasting, the model predictor produced during the training phase was subsequently applied across all pixels within the study area, but only for a specific date that occurred in the past. Specifically, this phase involves applying the logical reasoning learned by the model during the training phase to another dataset, called application dataset. This dataset includes observations sampled from all pixels, but concerning the dynamic variables, these observations are exclusively related to the selected date. This application produces a landslide prediction value for each pixel, similar to LSMs, with the difference that, in this case, the prediction is only valid for that selected date. Hence, the

resulting map indicates the spatiotemporal probability of landslide occurrence, namely LHM. The application was performed using the "predict" function on MATLAB, which provides the estimated posterior probability relative to the application dataset, computed starting from the model predictor (further explanations about the function can be found on the website: https://www.mathworks.com/help/stats/treebagger.predict.html?searchHighlight=predict%20for%20treebagger, last accessed on 21 March 2024).

In this study, we conducted a hindcast of the two major events that occurred on 10/06/2011 and 23/05/2013 (Fig. 2c). To evaluate the model performance in both critical and ordinary situations, the temporal domain of the simulation was extended from 3 days before to 3 days after the occurrence of these major events.

The results were a series of LHMs, each one valid for a specific day of analysis. This approach replicates the outcomes of a hypothetical nowcasting system that predicts landslide occurrence across the entire study area at daily time steps using dynamic and static input parameters.

Validation procedure (DTVT)

To validate the results, the procedure called DTVT (Double Threshold Validation Tool—Bulzinetti et al. 2021) was used in this study. This tool allows to validate the raw results of landslide hazard models by reaggregating the pixel-based LHMs over wider territorial units (henceforth called Pixel Aggregation Units—PAUs), comparing them with a landslide inventory and computing skill scores. The method is based on a double criterion to define a PAU as unstable: (i) Failure Probability Threshold (FPT): the probability above which a pixel is considered unstable; and (ii) Instability Diffusion Threshold (IDT): the percentage of unstable pixels that a PAU needs to justify the issuing of a warning.

In this study, the first-order catchments, extracted from the Norwegian national catchment database called REGINE (REGIster over NEdbørfelt i Norge—register of catchment areas in Norway) (https://www.nve.no/kart/kartdata/vassdragsdata/nedborfelt-regine/, last accessed on 10 September 2022) were used as PAUs. This choice is motivated by the possibility to partitioning the study area into hydro-geomorphological homogeneous polygons (Bell et al. 2014; Krøgli et al. 2018). Catchments with an area smaller than 5 km² were aggregated with adjacent polygons and only catchments fully within the study area and those containing landslides were employed. A total of 68 catchments were identified, ranging in size from a minimum of 5.20 km² to a maximum of 72.45 km².

FPT values between 0.5 and 0.8 and IDT values ranging from 1 to 20% were tested, and the configuration that correctly predicted all landslides with the minimum number of FA for both the 2011 and 2013 events was identified. In addition, this configuration allows to define a criterion for a territorial-based prediction output suitable for the implementation in a LEWS.

Input variables

Dynamic factors
Cumulative rainfall (CR_x [mm])

In the literature, a general agreement exists regarding the identification of two main landslides triggering factors: (i)

short-duration and high-intensity rainfalls, resulting in rapid infiltration of water into superficial and permeable soil layers, which rapidly increases water pressure and causes the triggering of shallow landslides; (ii) long-duration but low-intensity rainfalls, which are usually associated with the (re)activation of deep-seated landslides, which usually require longer infiltration times (Pereira et al. 2012; Giannecchini et al. 2012; Nocentini et al. 2023b). In this work, we defined 30 variables CR_x as the total amount of rain fallen in the past x days, where x ranges from 1 to 30, to account for the effect of short and intense rainfalls, mild but very prolonged rainfalls, and any possible intermediate conditions.

Rainfall data were obtained from the catalog provided by the Norwegian Meteorological Institute (MET Norway) for the period between 2010 and 2022 and with a spatial resolution of 1 km. The catalog consists of various sections, including the archive called "seNorge_2018" from which the daily rainfall maps were downloaded (https://threds.met.no/thredds/catalog/senorge/seNorge_2018/catalog.html, last accessed on 10 September 2022) (Lussana et al. 2018, 2019).

Cumulative snowmelt (CS_x [mm])
Similar to rainfall, the infiltration of snowmelt into the soil increases soil water content, thereby reducing shear strength and potentially triggering landslides. However, compared to rainfall, the snow melting process is generally slower and delivers to the soil an equivalent amount of water in longer times and at lower rates (Harr 1981; Ishikawa et al. 2015, 2016; Ishikawa and Miura 2011; Fu et al., 2018; Camera et al. 2021). Only heat waves can cause a rapid snow melting, resulting in a water supply comparable to that provided by a short-duration and intense rainfall (Gariano and Guzzetti 2016; Hanssen-Bauer et al. 2017; Dyrrdal et al., 2021). In this study, snowmelt (CS_x) was computed using time intervals ranging from 1 to 30 days (x corresponds to the specific interval selected). The dataset was provided by MET Norway for the period 2010–2022. Specifically, the Snow Water Equivalent (SWE) maps, available in the archive named "seNorge_snow," were extracted (https://thredds.met.no/thredds/catalog/senorge/seNorge_snow/catalog.html, last accessed on 10 September 2022) (Saloranta, 2014, 2016). SWE represents the amount of snowfall on the ground converted into millimeters of water. Considering that a decrease in SWE between two consecutive days is to be interpreted as snow melting, the daily cumulative snowmelt was derived from the difference in SWE between two consecutive days. In cases where the difference is positive, indicating snowfall, the snowmelt value was set equal to zero.

The month of observation (Month [-])
The variable Month indicates the month of the year in which an observation (both landslide and non-landslide) was sampled. Therefore, it is a categorical variable containing 12 classes, one for each month, from January to December. It was used as an empirical proxy to capture the seasonal variability of rainfall, snowmelt, temperature, and vegetation; which in turn influence the humidity of the soil (Nocentini et al. 2023b). In fact, the seasonal variability of soil moisture is considered one of the most influential factors in triggering landslides (Gariano and Guzzetti 2016; Rosi et al. 2021; Piciullo et al. 2022), and it is widely recognized that an ordinary

amount of water (from rainfall or snowmelt) can more likely trigger landslides if it occurs when the soil is already saturated, compared to a dry soil (Gariano and Guzzetti 2016; Nocentini et al. 2023b).

Static factor

A Landslide Susceptibility Index (LSI) was chosen as a static input variable to summarize the effect of the main static predisposing factors involved in the landslides triggering process. A basic LSI was obtained using a traditional static application of the RF algorithm, selecting as input variables only the parameters most commonly used in literature (Reichenbach et al. 2018; Segoni et al. 2021; Lima et al. 2022; Liu et al. 2021) and those used in previous works in the same area (Liu et al. 2021), namely: slope, aspect, total curvature, flow accumulation, distance from the stream network, lithology (5 classes: amphibolite facies, granulite facies, green schist facies, magmatic intrusions, subgreen schist facies) and land cover (6 classes: forests and seminatural areas, wetlands, agricultural areas, artificial surfaces, glacier and perpetual snow, bare rocks). Additionally, as previously proposed by Luti et al. (2020), a random parameter with values between 0 and 1 was also used in a preliminary version of LSI, to detect if the model correctly recognizes it as influential and to check if some of the parameters have an explanatory power close to a randomly generated field of values. After confirming that the model recognizes the random variable as irrelevant to landslide triggering, this variable was discarded.

The resulting LSI was used as the only static factor in the subsequent dynamic assessment. This brings some advantages, which include: the reduction of the number of input parameters and consequently of the computational times (which, in perspective, would be a relevant issue for operational applications), and it helps to focus the analyses and interpretation of the results on the role played by the different dynamic factors.

LSI was created using 100m × 100m pixels as basic computation units. The landslides from the original NNLI with a spatial uncertainty higher than 100m were discarded; instead, since the susceptibility mapping is based on a spatial assessment without temporal information, landslides with inaccurate or uncertain temporal definition were kept. In addition, the whole NGI landslide inventory is considered, due to the very high spatial resolution. A total of 453 landslides were collected for the study area from 1959 to 2022 (246 from NNLI and 207 from NGI). Since most landslides from the NNLI are geolocated as points representing the location of the impact on human infrastructures, the most probable landslide bodies were recreated by using the same methodology described in Nocentini et al. (2023b). Applying the Watershed tool in Arc-GIS PRO, the polygons representing the catchment areas upslope of each landslide point were identified; then clipped with a 200 m buffer. The resulting polygons can be considered a good approximation of the unstable areas, including the pixels upstream the landslide points.

LSI was generated by applying the RF algorithm with 100 decision trees, after testing larger numbers and confirming that
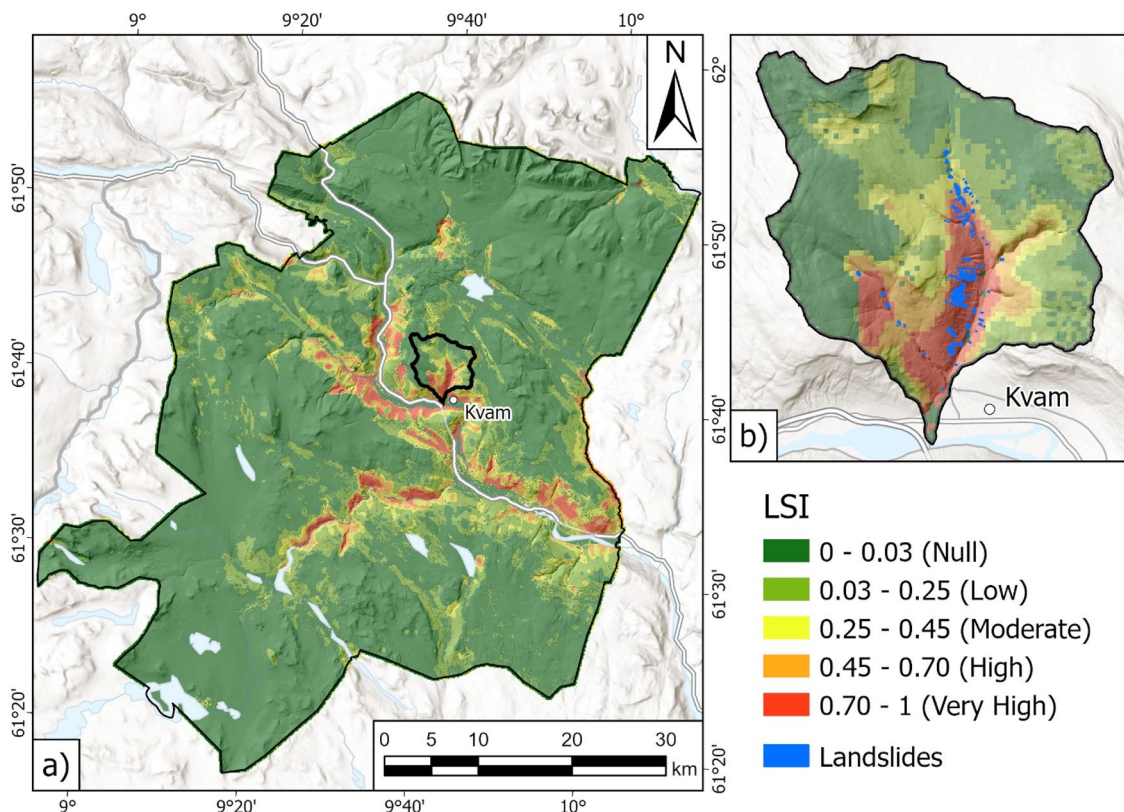

Fig. 3 a Map of LSI. **b** Magnification of LSI for the Kvam catchment area

they have no significant impact on model results. Four hundred fifty-three landslide polygons, encompassing a total of 18,826 pixels, and an equal number of randomly selected non-landslide pixels were used for model training and testing. The database was divided into two subsets: 70% for the training phase and 30% for the testing phase, both subsets having 50% of landslides and 50% of non-landslides points. The resulting predictor was then applied to all pixels in the study area to build the susceptibility map. Figure 3 shows LSI classified for graphic representation only using the natural breaks method (Jenks 1967). The resulting map achieved an AUC (Area Under the Receiver Operating Characteristic (ROC) Curve) value of 0.88, indicating high model performance. Among the static variables, the most important were slope, aspect, distance from the stream network, and land cover.

Results

Estimation of variables' importance

Due to the exceptional number of input parameters (63 in total, including the 30 cumulative rainfalls, the 30 cumulative snowmelts, LSI, Month, and a control parameter with random values between 0 and 1), the model was initially implemented in two separate preliminary phases. The first phase involved only cumulative rainfall, and the second phase only cumulative snowmelt. Both phases included Month, LSI, and the random variable and were tested on both SL- and LE-balanced databases. These tests were conducted to identify the periods of cumulation with the higher importance. The obtained histograms of OOB variables' importance estimates are presented in Fig. 4. The histograms illustrate the estimated importance of each variable, as averaged across seven model runs, with the maximum and minimum values identified by the whiskers.

The short-duration cumulative rainfalls show a greater importance compared to other rainfall parameters, with the highest peak for the daily rainfall. Conversely, longer cumulative periods show a lower influence in triggering the inventoried landslides. Additionally, the Month variable and LSI exhibit high importance when using both SL and LE database. Long-duration cumulative snowmelt displays a greater influence than the short-duration ones, albeit slightly.

For the subsequent training, test, and application phases of the dynamic RF model, it was decided to employ the most important variables, namely Month, LSI, CR_1, and CS_30. Additionally, other cumulative variables were included at weekly intervals to further analyze their contribution in triggering landslides across different degrees of database imbalance and through PDPs. Specifically, CR_7, CR_14, CR_21, and CR_30 were added to represent longer cumulative rainfall periods, while only CS_14 was included to represent medium-duration cumulative snowmelt, as short-duration snow melting demonstrates very low importance in triggering landslides. The random variable was excluded because the model recognizes it as insignificant for landslide prediction.

Figure 5 displays the average estimates of variables' importance obtained for the dynamic RF model for different degrees of database imbalance. These plots confirmed CR_1 as the most significant cumulative rainfall for each configuration, as expected, with an importance about double than the other cumulative rainfalls. The Month variable registers a high level of importance and LSI emerges as the most impactful variable in each case.
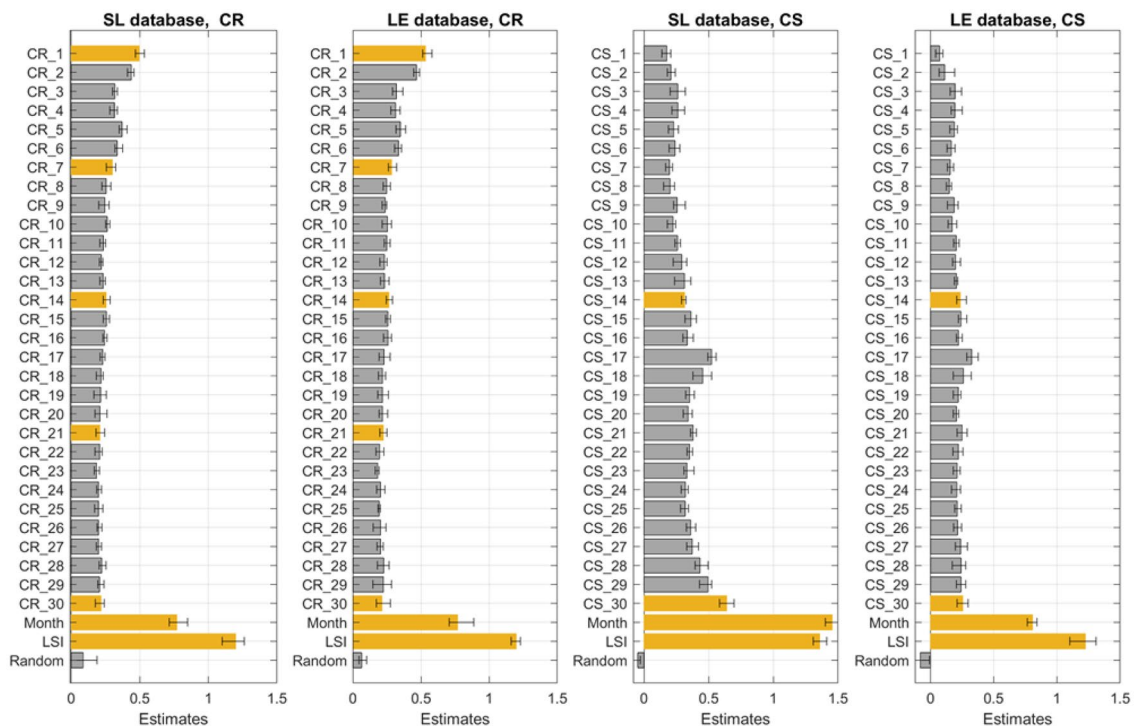

Fig. 4 OOB variables' importance estimates performed as preliminary analysis, separating cumulative rainfall (CR) to cumulative snowmelt (CS), for both SL- and LE-balanced databases. The variables chosen for the subsequent dynamic RF model runs are highlighted in orange
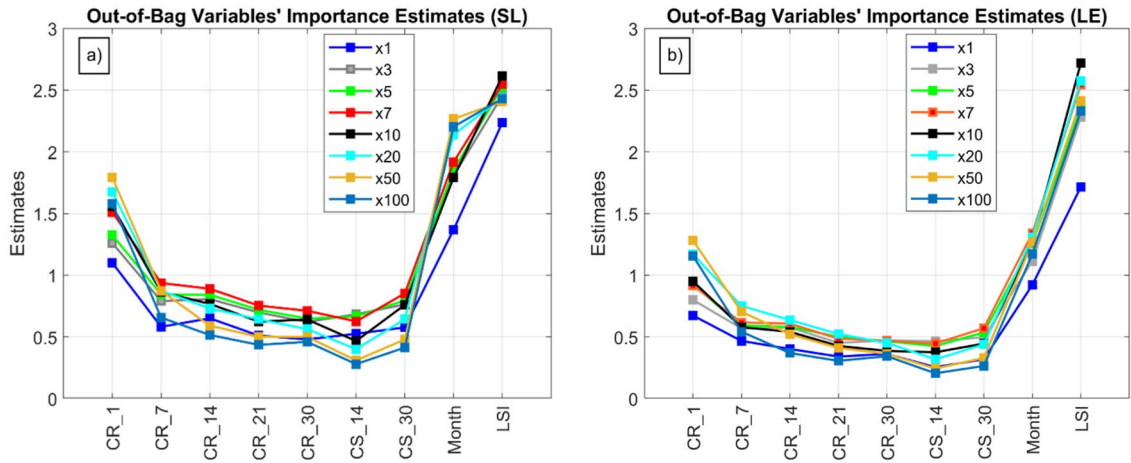
Fig. 5 Variability of the average OOB variables' importance estimates for the selected variables under different degrees of database imbalance; obtained using **a** SL database and **b** LE database

The effect of snowmelt on triggering the inventoried landslides appears to be relatively minor, and its importance is similar to those of antecedent rainfall, both using SL or LE database. Increasing the degree of imbalance does not significantly alter the correlation among variables, and the ranking of the variables' importance remain essentially the same.

Partial dependence plots

Figure 6 and Fig. 7 show the PDPs of the selected variables for different degrees of database imbalance, averaged for the seven model runs, using SL or LE database respectively. The figures highlight that the importance of the selected variables decreases as the degree of imbalance increases. This is an expected outcome, which is not specifically related to the model, but it is an intrinsic characteristic of PDPs: they represent the marginal effect of a given value of a given parameter on the model prediction, averaged over the total number of instances in the training dataset (see Molnar (2020) for further mathematical explanations). As a consequence, the relative score of each variable decreases if the degree of imbalance increases, because the total number of instances in the dataset increases accordingly (Friedman 2001; Molnar 2020; Nocentini et al. 2023b). For our purposes, when observing a PDP of a given variable, the graph is interpreted in relative terms, without focusing on the scores reported on the y axis and considering the shape depicted by the curves over different degrees of imbalance. Conversely, when comparing plots of different variables, emphasis is placed on the difference in importance estimate among curves of the same degree of imbalance.
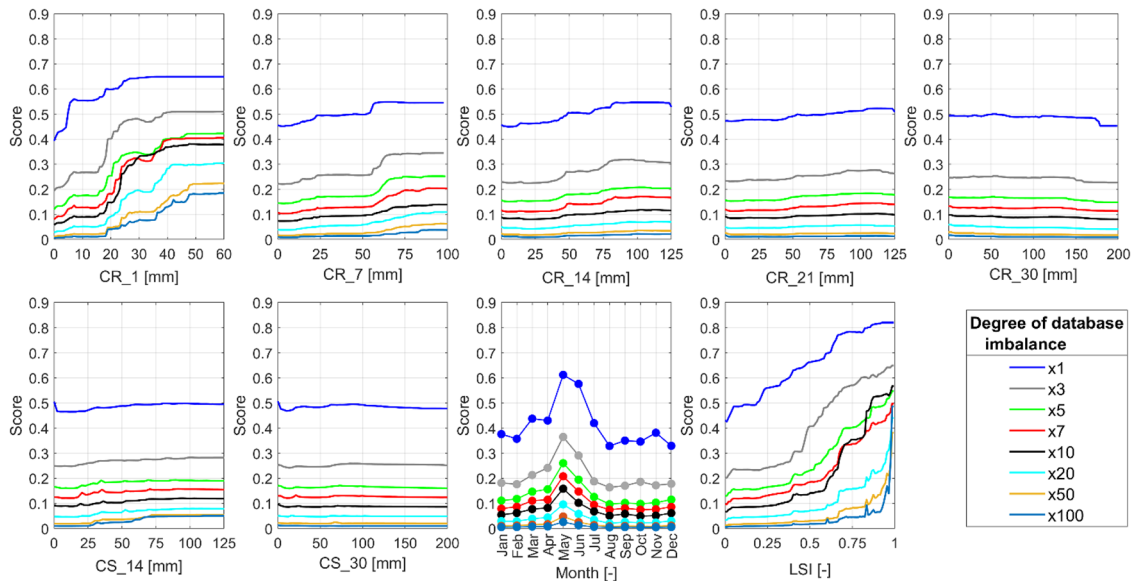

Fig. 6 Mean partial dependence of the selected variables for different degrees of database imbalance by using a SL database
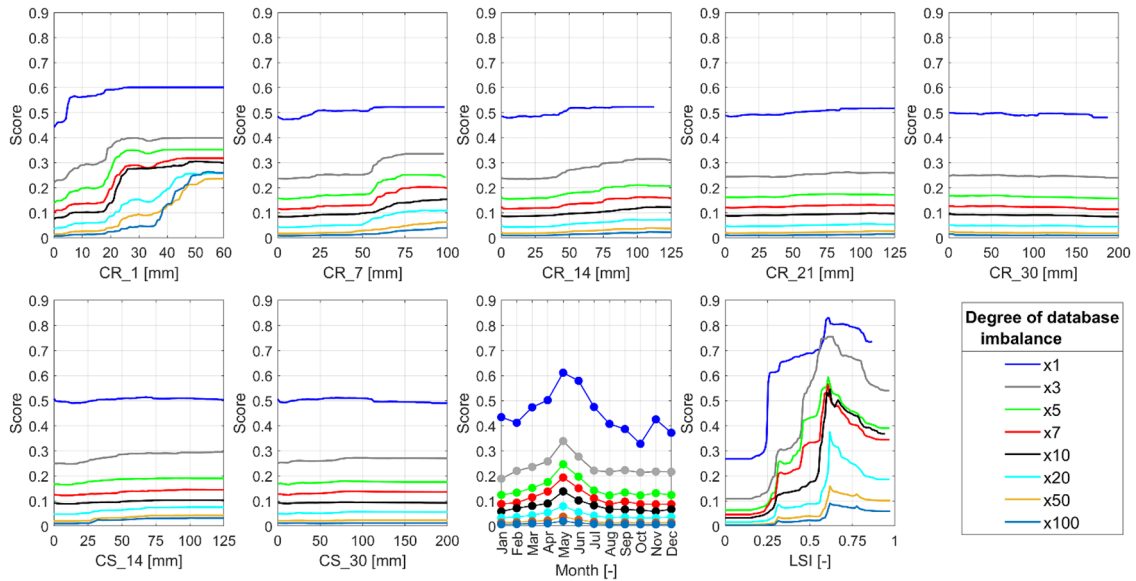
Fig. 7 Mean partial dependence of the selected variables for different degrees of database imbalance by using a LE database

Looking at Figs. 6 and Fig. 7, and comparing PDPs among different variables, CR_1 shows a growing trend, both in case of SL and LE. This outcome aligns with the physical mechanism of shallow landslide triggering, suggesting a higher influence for more intense rainfalls, up to a threshold value beyond which landslides are expected to occur. A similar trend, albeit less marked, was observed for CR_7, CR_14, and CR_21, showing that antecedent rainfall played a minor but yet noticeable role in landslides triggering. Again, this outcome was considered a further proof of the model adherence to the physics of the triggering mechanism, as shallow landslides are typically triggered by short and intense rainfall, but antecedent rainfall may play a not negligible predisposing role (Ponziani et al. 2012; Kim et al. 2021).

For CR_30, CS_14 and CS_30, PDPs are flat, indicating that these parameters do not play a key role in triggering the inventoried landslides. Indeed, several studies relate their influence in triggering shallow landslides on the permeability of soils, with less permeable soils more sensitive to the accumulation of water over extended periods (Glade et al. 2000). Therefore, it is plausible that the physical properties of the soil of this study area determine that the predisposing effect of cumulative rainfall and snowmelt is almost completely lost over long period due to its draining capacity. The Month variable displays a complex behavior, with the main positive peaks of importance during May and June, two of the months with the highest amount of rainfall and snowmelt (see Fig. 15). LSI exhibits a monotonous rising trend using SL, where the peak of importance is observed for LSI values tending toward 1. This is obviously in accordance with physical evidence, as this static variable, even if it does not bear temporal information, helps the dynamic model to detect the places where landslides should be expected. However, the PDP obtained for LSI using LE shows an anomalous relationship: the peak is reached for LSI values equal to 0.6, then the relationship remains flat, or it decreases. This outcome is probably an artifact due to the identification method of the LE database: all landslides that occurred on the same day and

in the same 1 km pixel (spatial resolution of the meteorological data) were considered a single event and an average value of LSI was considered. This reduces the spatial accuracy of the dynamic model, leading to unexpected results.

Looking at different degrees of imbalance, there is a noticeable change in trend for CR_1. Increasing the degree of database imbalance results in a clear shift of the peak of importance of CR_1 toward more intense rainfall values. These findings align with the observations made by Nocentini et al. (2023b). This trend can be observed in greater detail through the normalization of PDPs. By converting partial dependence values into a 0–1 range, it becomes feasible to compare them across different degrees of imbalance within a plot of the same variable. Considering that PDPs highlighted the incoherence of LE results and the significance of antecedent rainfall and snowmelt remains substantially flat, the normalization focused only on CR_1, Month, and LSI and only on the SL database. Figure 8 shows the normalized partial dependence plots (nPDPs), where the normalized partial dependence values (x_n) were obtained by applying the following equation:

$$x_n = \frac{(x - x_{min})}{(x_{max} - x_{min})}$$

where x, x_{min}, and x_{max} represent respectively the current, minimum, and maximum values of importance reported in the classic PDPs.

Looking at Fig. 8, CR_1 registers the greatest increase in importance (from 0 to 0.65) at 5mm of rainfall when using a balanced database. By increasing the number of non-landslide events seven times, the greatest increase is observed for about 20mm of rainfall (up to 0.75). With a further increase in the degree of imbalance, PDPs largely remain consistent, but it becomes progressively evident that daily rainfalls lower than 20mm do not affect landslide triggering. Instead, once that 20mm's threshold is crossed, the impact of CR_1 in triggering landslides rapidly increases. LSI also experiences a considerable change: as the number of
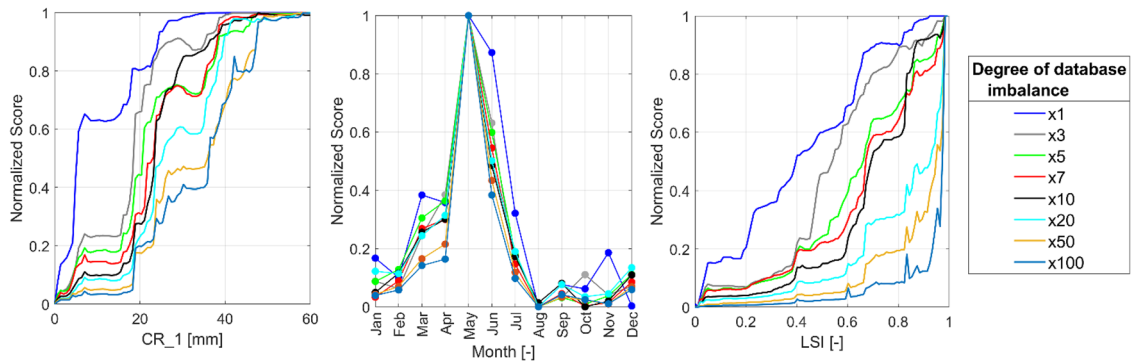
Fig. 8 nPDPs obtained for CR_1, Month, and LSI variables, for different degrees of SL database imbalance

non-landslide events increases, the importance of LSI values less than 0.5 decreases. This change in nPDPs' trend comes from the augmentation of non-landslide events: as the database imbalance grows, the model is fed with more data, which allows for a better representation of the real situation, where non-landslide instances are usually orders of magnitude more abundant than landslide instances. Consequently, the RF algorithm can obtain a better calibration, establishing a more suitable threshold to distinguish between ordinary and triggering rainfall events.

RF allows to generate the PDPs of interactions between two variables (iPDPs), which illustrates how they interact to affect model outcomes. The most important cumulative rainfall is CR_1, so it was decided to analyze its interactions with the other variables. The iPDPs obtained using the configurations × 1 (Fig. 9) and × 7 (Fig. 10) were taken as an example to illustrate the differences between balanced and imbalanced database.

It results that low-intensity CR_1 (< 5mm) has always a minimal impact on landslide triggering, regardless on the amount of

antecedent rainfall. Conversely, higher CR_1 values consistently show a high importance, even with low antecedent rainfall. Similar results are obtained when CR_1 is combined with snowmelt. Instead, the influence of CR_1 varies with Month. During wet seasons (May and June), even moderate daily rainfall has a significant impact. Conversely, during dry seasons, the influence of CR_1 decreases, even with heavy rainfall. LSI also controls the influence of CR_1: in low-susceptibility areas, an intense rainfall is less influential than where susceptibility is high. In addition, moving from the × 1 to × 7 configuration, a decrease in importance is observed for CR_1 values between 5 and 20mm, confirming the behavior already observed with the nPDPs as the number of non-landslide events increases.

Dynamic landslide probability mapping

Figure 11 shows the boxplots of the landslide probability values obtained through the procedure described in the "Dynamic random
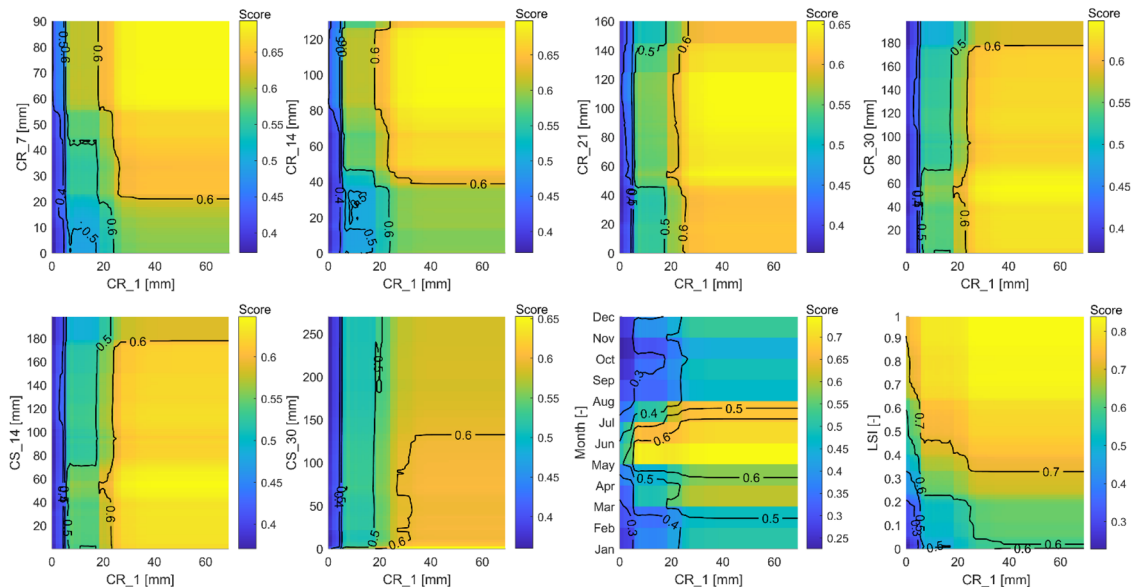

Fig. 9 iPDPs between CR_1 and the other input variables used for the dynamic RF model run through a balanced database
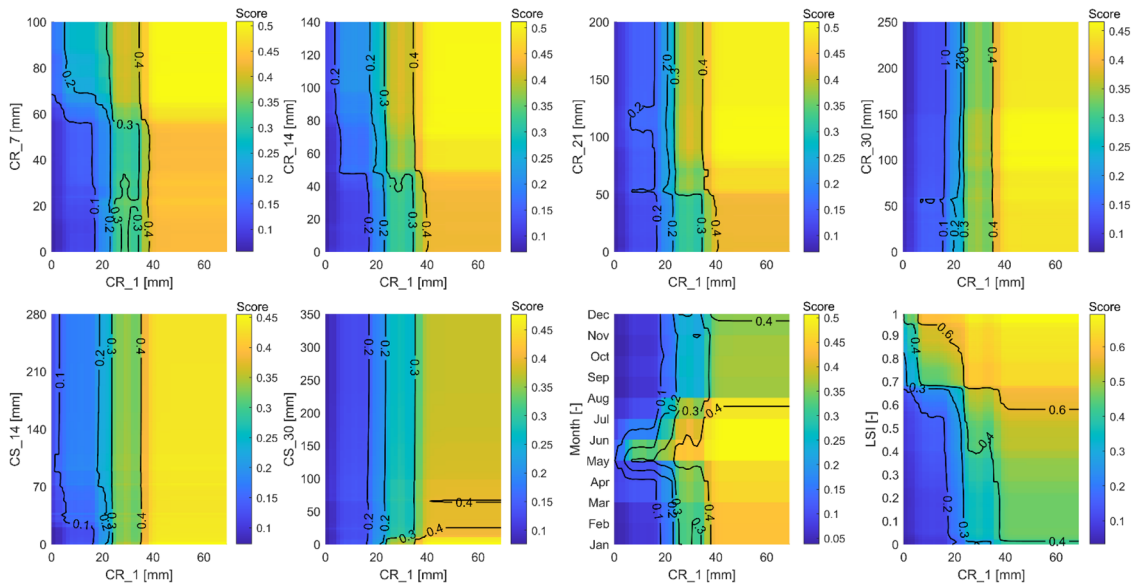
Fig. 10 iPDPs between CR_1 and the other input variables used for the dynamic RF model run through a 7 times imbalanced database

forest application" section, for 3 days before, on the exact day, and 3 days after the major events of 2011 and 2013, for different degrees of imbalance. Each boxplot evaluates the dynamic change in the overall instability of the area and offers a preliminary assessment of the temporal component of the predictions provided by the dynamic RF model. The whiskers of the boxplots extend to the minimum and maximum values of the maps, while the boxes contain the data

between the 25th and 75th percentile. The black lines within them correspond to the mean values.

The graphics demonstrate the high dynamicity of the model outputs. In fact, the simulations of 10/06/2011 and 23/05/2013, which experienced the major landslide events, produced mean probability values greater than those of the other simulated events, without landslides. This indicates the model's positive response to the
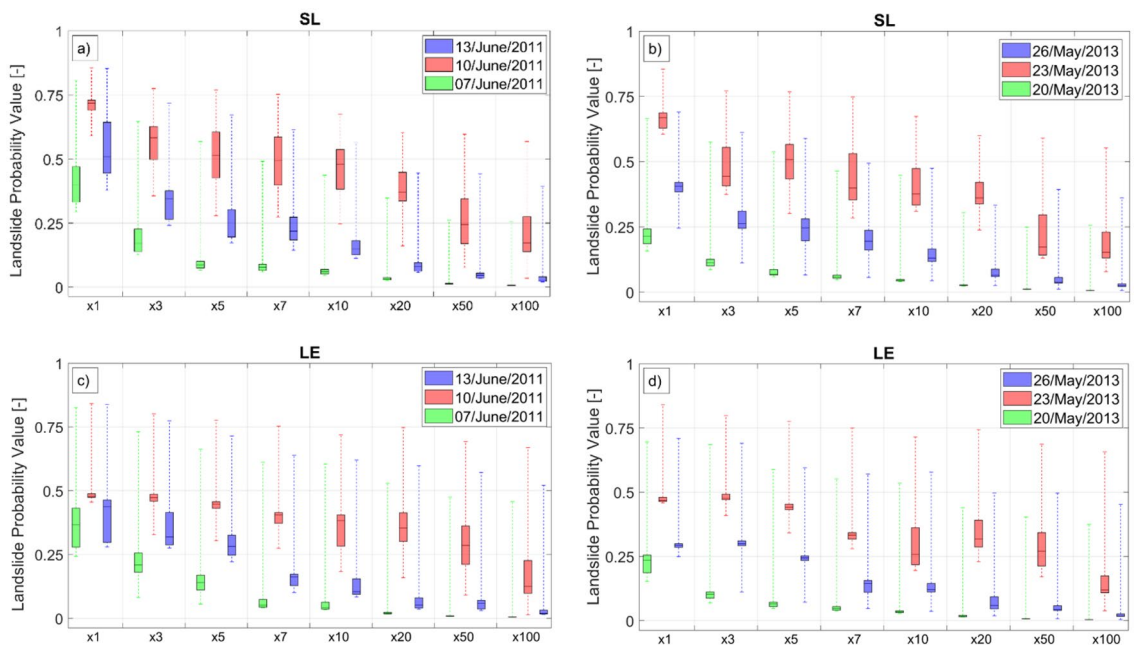

Fig. 11 Boxplot of the LHMs obtained for **a** 2011 event and SL database; **b** 2013 event and SL database; **c** 2011 event and LE database; and **d** 2013 event and LE database
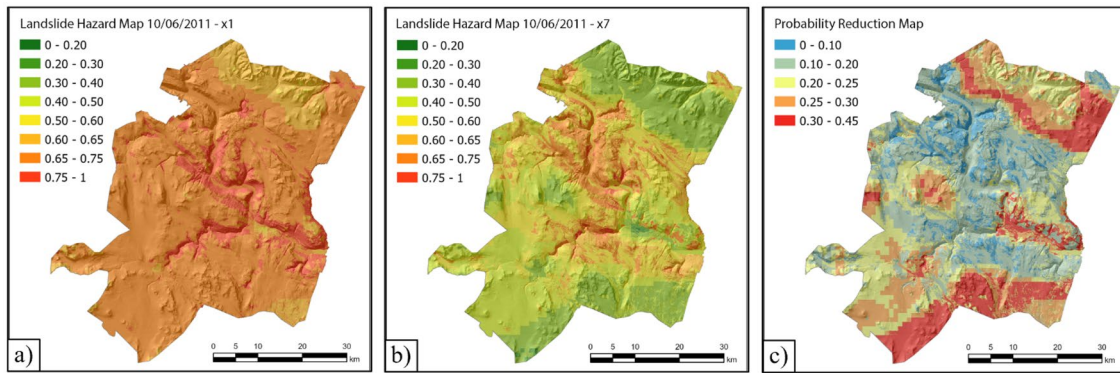
Fig. 12 LHMs of the 10/06/2011 event obtained with **a** SL-balanced database and **b** SL database imbalanced 7 times. In **c** is illustrated the probability reduction map obtained by the difference between the former and the latter map

dynamic input provided by meteorological variables, which are the only input data that vary on the three simulated dates. For the simulation conducted 3 days before, the maximum probability value remains below 0.5 using an imbalanced database from the degree × 7 onwards, as expected for an event without landslides. The LHMs obtained for 3 days before, always produces probability values lower than the ones obtained for the simulation 3 days after the major events. This result can be explained by the significant rainfall that occurred during the major events, which saturates the soil on slopes for days afterwards. By incorporating long-duration cumulative rainfall, the model accounts for these conditions, resulting in

higher probability values even days after a critical rainfall due to the persistence of saturated slopes. Using LE, the boxes of the major events remain below 0.5 for each degree of imbalance for both 2011 and 2013. Therefore, the prediction based on LE again shows a lower effectiveness than SL.

By excessively increasing the degree of imbalance, the mean probability values decrease significantly. This is due to the mechanism previously identified with the PDPs. A drastic increase in the number of non-landslide events causes an excessive raising of the thresholds used by the model to discriminate between stability and instability. This outcome could potentially reduce the predictive

Table 1 Performance of the pixel-based LHMs, calculated by setting a probability threshold equal to 0.5

Landslide inventory	Degree of imbalance	Performance indicators	07/06/2011	10/06/2011	13/06/2011	20/05/2013	23/05/2013	26/05/2013
SL	×1	TP	0	115	0	0	63	0
		FN	0	0	0	0	0	0
		FP	39,866	279,088	141,198	21,136	279,140	29,572
		TN	239,337	0	138,005	258,067	0	249,631
	×7	TP	0	113	0	0	63	0
		FN	0	2	0	0	0	0
		FP	0	135,232	7006	0	86,163	215
		TN	279,203	143,856	272,197	279,203	192,977	278,988
LE	×1	TP	0	98	0	0	45	0
		FN	0	17	0	0	18	0
		FP	26,380	33,054	29,619	26,324	26,986	26,298
		TN	252,823	246,034	249,584	252,879	252,157	252,905
	×7	TP	0	91	0	0	14	0
		FN	0	24	0	0	49	0
		FP	1535	22,769	15,201	534	10,197	8559
		TN	277,668	256,319	264,002	278,669	268,946	270,644

capability of LHMs, because for higher thresholds, a higher number of FNs is expected.

The LHMs obtained for the 2011 event using either the balanced or the ×7 imbalanced SL database were taken as examples to highlight the effects of increasing the degree of imbalance, and the probability reduction between the former and the latter is shown in Fig. 12. With the ×1 configuration, the probability is consistently high throughout the study area, whereas with the ×7 configuration, a more gradual variation in probability is observed. This trend is also more consistent with LSI (Fig. 3). In fact, the probability reduction map indicates a decrease in landslide probability of up to 40% in regions outside the Gudbrandsdalen Valley, which are landslide-free and exhibit very low susceptibility values. While inside the valley itself, where susceptibility is maximum, the probability values remain largely constant. Therefore, through an increase in database imbalance, a smaller number of FPs is expected due to the decrease in landslide triggering probability in areas without landslides.

Validation

Table 1 presents the performance of pixel-based LHMs in terms of True Positive (TP), True Negative (TN), False Positives (FP), and False Negative (FN), calculated by considering a probability threshold equal to 0.5. To illustrate the differences between a balanced and an imbalanced database, the results from the ×1 and ×7 configurations are taken as examples and presented herein. With a SL-balanced database, a high number of FPs was observed both 3

days before and 3 days after the major events, which instead exhibited the highest FPs counts with no TNs recorded, but also no FNs. The use of LE resulted in a substantial increase in FNs counts for both major events. In the case of an imbalanced SL database ×7, the number of FPs significantly decreased for each simulation, but there was also a slight increase in the number of FNs only for the 10/06/2011 event (2 FNs), an increase that was more pronounced for both major events when using LE.

The optimal DTVT configuration that ensures zero FNs and minimizes the number of FPs for both the 10/06/2011 and 23/05/2013 events was identified for FPT = 0.65, IDT = 5%, and the SL database imbalanced by a factor of 7. Table 2 illustrates the performance of the reaggregated maps obtained by applying DTVT using both a balanced and a ×7 imbalanced database.

Concerning the optimal DTVT configuration identified for SL, moving from a balanced to a ×7 imbalanced database, every landslide was still correctly predicted, and the number of FPs was reduced by 24 in the case of the 2011 event and by 36 for the 2013 event. Instead, for simulations 3 days before and 3 days after the major events, the FPs were reduced to zero, and only TNs were recorded.

To visualize the dynamic outcomes of the model, Figs. 13 and Fig. 14 show the results of the hindcasts of the 2011 and the 2013 events, respectively; in both cases the ×7 SL database was used. The maps generated 3 days before and 3 days after the major events exhibit a low landslide probability, with peak values of respectively 0.49 and 0.61 for the 2011 event, and 0.46 and 0.50 for the 2013 event, mainly concentrated around the Gudbrandsdalen Valley, where the susceptibility is very high. The validation phase of these

Table 2 Performance of the reaggregated LHMs obtained by applying DTVT by setting FPT = 0.65 and IDT = 5%

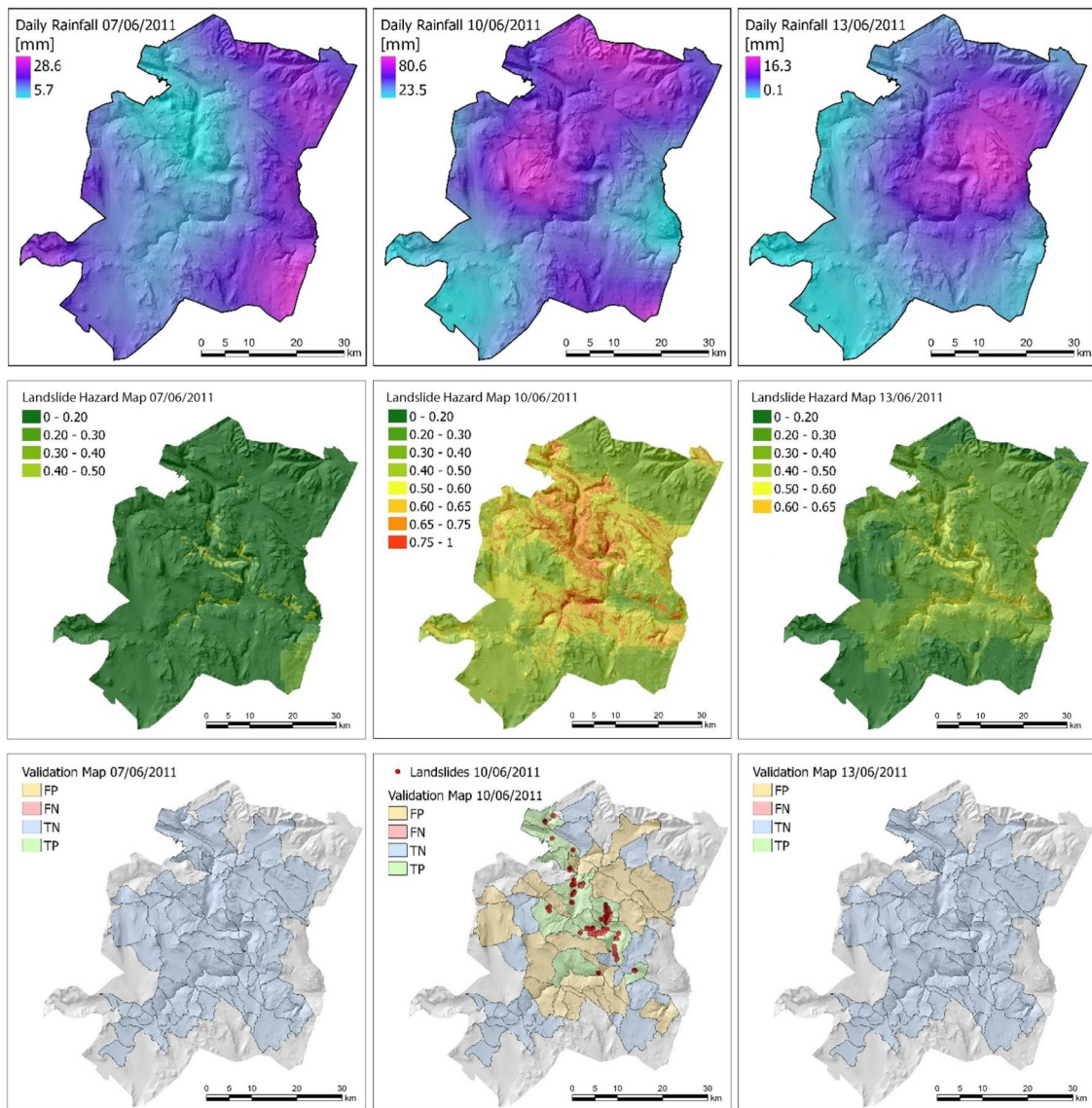| Landslide inventory	Degree of imbalance	Performance indicators	07/06/2011	10/06/2011	13/06/2011	20/05/2013	23/05/2013	26/05/2013
SL	×1	TP	0	14	0	0	10	0
		FN	0	0	0	0	0	0
		FP	24	54	45	0	53	14
		TN	44	0	23	68	5	54
	×7	TP	0	14	0	0	10	0
		FN	0	0	0	0	0	0
		FP	0	30	0	0	17	0
		TN	68	24	68	68	41	68
LE	×1	TP	0	10	0	0	7	0
		FN	0	4	0	0	3	0
		FP	23	23	33	1	26	20
		TN	45	31	35	67	32	48
	×7	TP	0	10	0	0	7	0
		FN	0	4	0	0	3	0
		FP	0	12	0	0	14	0
		TN	68	42	68	68	44	68

Fig. 13 Daily cumulative rainfall, pixel-based LHMs, and validation maps for each simulated event in June 2011

maps results in only TNs, as expected on days when less intense rainfall occurred, and no landslides were reported. The LHM for 10/06/2011 shows a high landslide probability, particularly in the Gudbrandsdalen Valley, with a maximum of 0.76. However, the map also reveals areas with lower probability despite high daily rainfall. This discrepancy arises because these areas also exhibit a low LSI. Conversely, regions with high probability are characterized by both high daily rainfall and susceptibility, aligning with the iPDPs shown in Fig. 10. The related validation map recorded each PAU hit by landslides as TP, but 30 FPs were issued. For the event on 23/05/2013, the maximum probability value is 0.75, observed once again in the valley. The validation results in a 100% rate of correct predictions and 17 FPs, thus obtaining better results than for the 2011 event.

Moving from pixel-based to PAUs-based validation, the model performance improved, as the number of FNs resulted equal to 0 for both events when using SL database imbalanced seven times.

Discussion

The variables' importance results shows that daily cumulative rainfall is more important than antecedent rainfall and snowmelt; in line with what was expected for the inventoried landslides, mainly shallow landslides, that are primarily influenced by short-duration and intense rainfall. These results are also consistent with the interpretations made by Heyerdahl and Høydal (2017), who observed that daily rainfall was the direct triggering factor for the 2011 and 2013 events. However, antecedent snowmelt results in being more influential in triggering landslides compared to short-duration snowmelt, as expected, due to the lower soil water supply provided by daily snowmelt. The influence of long-duration cumulative snowmelt is also comparable to that of antecedent rainfall, which has similar infiltration rates.

Among the dynamic variables, the Month shows the highest importance, and the PDPs associate the months of May and June
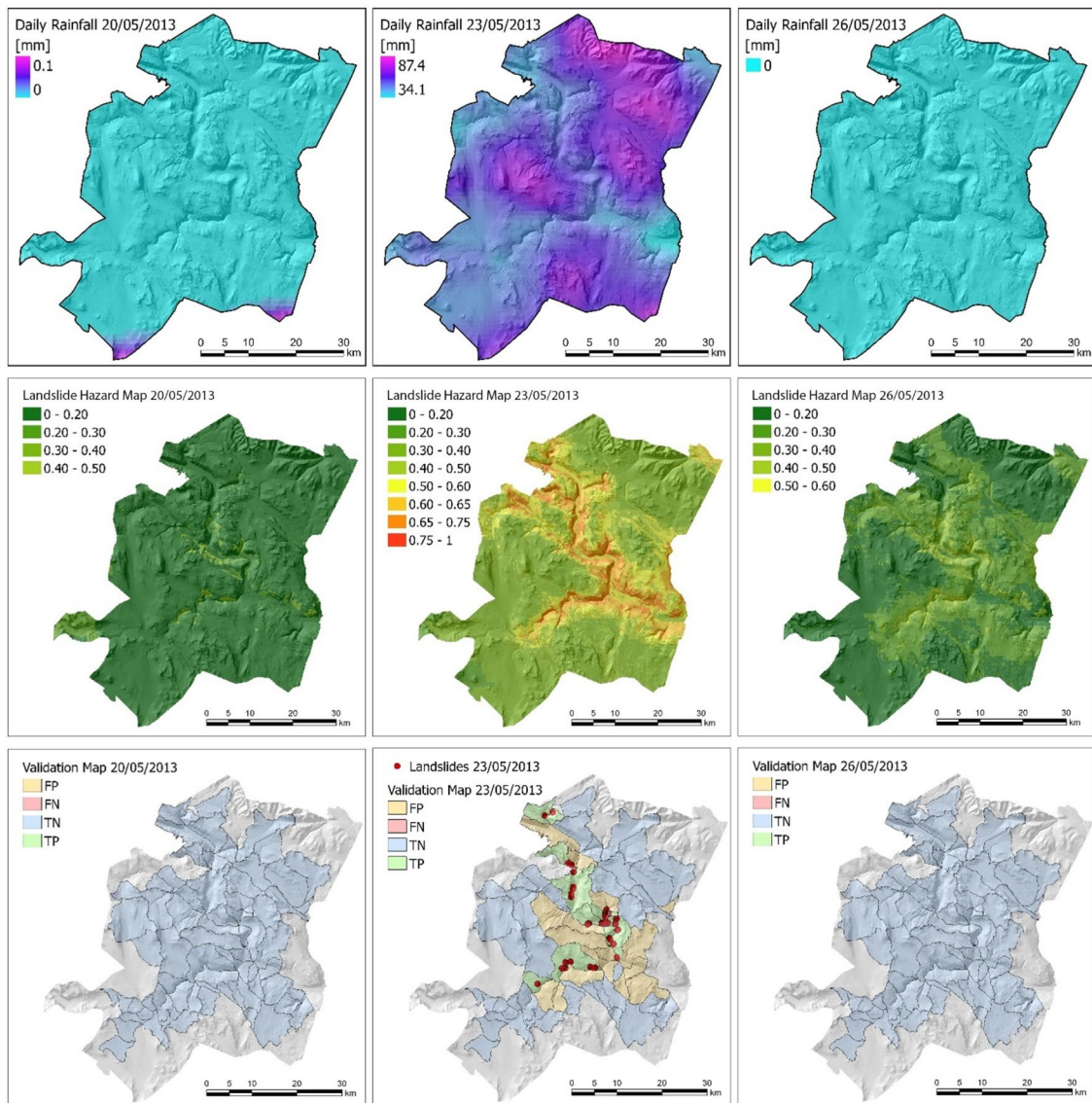
Fig. 14 Daily cumulative rainfall, pixel-based LHMs, and validation maps for each simulated event in May 2013

with the greatest influence on landslide triggering. The occurrence of two exceptional events in June 2011 and May 2013 certainly has a significant impact on these results, considering the use of a data-driven model (Steger et al. 2021), but the Month variable also shows a correlation with the seasonal variability of rainfall, snowmelt, and temperature. In fact, as shown in Fig. 15, in May and June, the degree of soil moisture is very high due to the coupling effect of snowmelt and rainfall. July and August are the rainiest months, but they are also the months with the highest temperatures. The combined effects of abundant rainfall and high temperatures results in vegetation growth. The presence of healthy and lush vegetation in July and August has a stabilizing effect on the slopes, particularly by reducing soil moisture through evapotranspiration (Löbmann et al. 2020; Capobianco et al. 2021; Masi et al. 2021). Therefore, the use of the Month variable results an effective method to represent in an empirical way the soil moisture variability over seasons, in

turn correlated to seasonal variability of snowmelt, rainfall, temperature, and vegetation.

The variable that shows the highest importance is LSI, demonstrating once again that the model can correctly simulate the physical mechanism of landslides triggering, as they mainly occur in slopes that, due to geological and geomorphological characteristics, are already subject to this type of instability. Using the LE database, the partial dependency of LSI shows a decreasing trend (Fig. 7), thus assigning progressively lesser importance to susceptibility values greater than 0.6. This anomalous behavior is not consistent with the physical process investigated, as higher landslide probability is expected in areas with high susceptibility. The use of the LE database was aimed at the attempt to remove the bias contained in the landslide inventory, but its use, in turn, introduces an incorrect interpretation of the LSI variable, due to the method of incorporating multiple landslide events, which forces sampling an
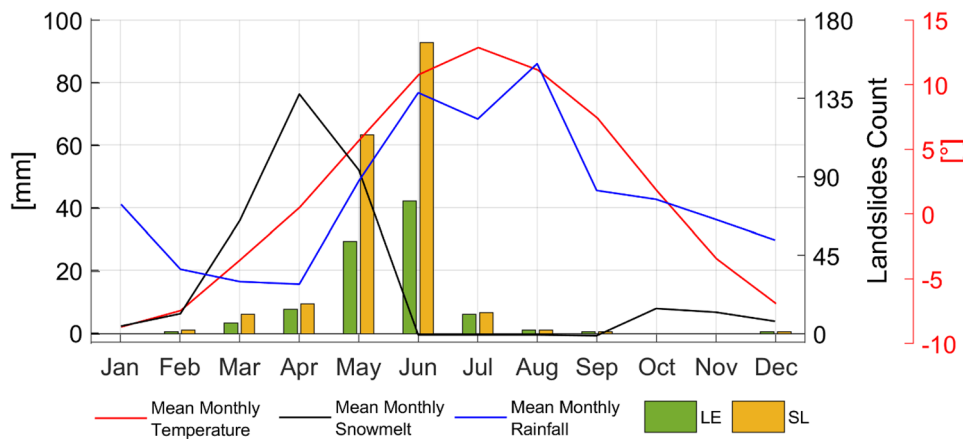
Fig. 15 Values of monthly rainfall, snowmelt, and temperature (from 2010 to 2022) averaged for the entire study area, with overlayed the monthly landslide frequency for both SL and LE databases

average LSI value. The use of PDPs confirmed this anomaly, thereby suggesting to exclude an approach based on LE database for further applications of the model.

Increasing the degree of database imbalance toward non-landslide events contributes to better represent the real situation, in which non-landslide events occur more frequently than landslide events. Therefore, increasing the number of observations allows feeding the model with a more representative database of the actual rainfall frequency distribution. However, since incompleteness is one of the most critical issues of most landslide inventories, the possibility of unreported landslides should always take into account, as they could be wrongly identified as non-landslide events. Hence, this imbalance should be carefully evaluated in future applications. For these reasons, PDPs are very helpful for identifying the correct degree of database imbalance. In fact, an excessive increase in the number of non-landslide events tends to nullify the importance of certain classes of values, supporting only a close range of them. For example,

shows that a database imbalanced by a factor of 100 assigns very low importance to values of CR_1 below 20mm and assigns high importance (above 0.6) only to values greater than 40mm. Similarly, it almost nullifies the importance of LSI values lower than 0.9. In these cases, the contribution provided by some classes in producing the model's predictions is lost, resulting in outcomes not consistent with the physical mechanism of landslide triggering.

The use of the proposed dynamic RF model application to hindcast past events shows promising results. The generated LHMs show clear patterns, which are consistent with the ground truth, both from a spatial and temporal perspective. Concerning the simulations performed before and after the major events, when no particularly intense rainfall occurred, and no land-slides were reported, very low landslide probability values were correctly observed. In contrast, for the simulations of the major event on 10/06/2011 and 23/05/2013, where heavy rainfall and several landslides were recorded, the model produced high landslide probability values, as expected. Moreover, 3 days after the major events, the maximum probability values remained high, because the model takes into account antecedent rainfall,

albeit with less importance than other variables. Hence, post-event maps maintain a relatively high probability of occurrence instead of dropping immediately after the end of the rain. This outcome can be considered precautionary and in accordance with certain civil protection procedures, for which the transition from a high emergency level to an ordinary condition is usually gradual.

The validation of the maps obtained using the DTVT has shown satisfactory results. The optimal configuration identified, using a × 7 imbalanced SL database with FPT = 0.65 and IDT = 5%, allowed to maximize the TPs, reducing to zero the number of FNs and obtaining a relatively low number of FPs. This configuration has been verified for both simulated major events; hence, considering it valid also for future applications, it represents the basis of a hypothetical LEWS based on probability maps produced using ML. Moreover, for the simulations before and after the major events, the validation maps show only TNs, as expected for ordinary criticality events. Using LE, it was not possible to identify a configuration with zero FNs recorded: even with FPT = 0.50 and IDT = 1%, and both with a balanced and a × 7 imbalanced data-base, 4 and 3 FNs were obtained for the 10/06/2010 and 23/05/2013 events respectively. This outcome further demonstrates that a forecasting system based on LE is less effective than one based on SL. The use of DTVT not only proved extremely useful for a quick and accurate validation of the maps, but also reaggregating pixels into PAUs allowed to neglect the uncertainty related to input data and helped in identifying a criticality level suitable for issuing warnings. This represents a further step forward in producing an instrument that is more useful for civil protection purposes and easily implementable in a LEWS. The FNs generated are probably due to limitations in the landslide database; in fact, some land-slides might have occurred in remote areas during the two major events but were not detected as they did not cause any damage.

Conclusion

In this study, we introduce a method based on the use of the RF algorithm that integrates dynamic variables, cumulative rainfall, and snowmelt, for spatiotemporal landslide prediction. The RF

model is calibrated and evaluated using a spatially and temporally explicit landslide inventory; aiming to produce LHMs for a region in Norway affected by two major landslide events in June 2011 and May 2013. Below are listed the main results obtained:

- The reliability of the model was verified using OOBE and PDPs: these indexes allowed to evaluate to which extent the results of the ML were in accordance with the physical processes governing the documented landslide phenomena.
- Daily rainfall is the primary trigger for documented landslides, while antecedent rainfall and snowmelt result secondary, but not negligible. LSI's significance underscores geological and geomorphological factors. Seasonal variability, particularly before summer when soil moisture is high, also plays a pivotal role in landslide initiation.
- Increasing database imbalance (through the augmentation of non-landslide events), if well-calibrated, provides a more realistic representation of events but needs to be carefully managed to avoid an increase in FNs.
- The procedure for defining LHMs was conceived, demonstrating the applicability of RF for spatiotemporal landslide forecasting.
- The use of DTVT results effective, allowing for the transition from pixel-based to catchment-based validation, enhancing model performance and suggesting potential applications for operational warning systems.

The innovative methodology proposed in this study offers a versatile framework suitable for different contexts for both spatial and temporal landslides forecasting. In these cases, recalibrating the model considering the site-specific characteristics is advisable, particularly when considering different landslide typologies and alternative dynamic variables.

Hence, it will be central for future applications to explore the effects of soil moisture, vegetation, and their seasonal variability on landslide triggering with specific approaches, in order to confirm the results achieved. In addition, for future implementation in LEWSs, the efforts will be focused on the calibration of different levels of warning, which could be based on FPT and IDT values.

Author contribution
NN conceived the work, performed the analysis, contributed to the interpretation of the results, and drafted the manuscript. AR conceived the work, contributed to the interpretation of the results, and participated in writing the manuscript. LP conceived the work, contributed to the interpretation of the results, and participated in writing the manuscript. ZL contributed to data collection, result interpretation, and manuscript writing. SS conceived the work, contributed to the interpretation of the results, and participated in writing the manuscript. RF supervised the work.

Data availability
The raw data supporting the conclusion of this article will be made available by the authors upon request.

Declarations

Competing interests   The authors declare no competing interests.

References

Alvioli M, Baum RL (2016) Parallelization of the TRIGRS Model for Rainfall-Induced Landslides Using the Message Passing Interface. https://doi.org/10.1016/j.envsoft.2016.04.002

Bell R, Cepeda J, Devoli G (2014) Landslide susceptibility modeling at catchment level for improvement of the landslide early warning system in Norway.Conference Proceedings of the World Landslide Forum. 3

Bishop CM, Nasrabadi NM (2006) Pattern recognition and machine learning, vol 4. Springer

Breiman L (2001) Random Forests, pp 5–32

Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation. Natural Hazards and Earth System Science 5. https://doi.org/10.5194/nhess-5-853-2005

Bulzinetti MA, Segoni S, Pappafico G, Masi EB, Rossi G, Tofani V (2021) A tool for the automatic aggregation and validation of the results of physically based distributed slope stability models. Water (Switzerland) 13. https://doi.org/10.3390/w13172313

Calvello M, Piciullo L (2016) Assessing the performance of regional landslide early warning models: the EDuMaP method. Nat Hazards Earth Syst. Sci. 16(1):103–122. https://doi.org/10.5194/nhess-16-103-2016

Camera CAS, Bajni G, Corno I, Raffa M, Stevenazzi S, Apuani T (2021) Introducing intense rainfall and snowmelt variables to implement a process-related non-stationary shallow landslide susceptibility analysis, Science of the Total Environment, 786, https://doi.org/10.1016/j.scitotenv.2021.147360

Capobianco V, Robinson K, Kalsnes B, Ekeheien C, Høydal Ø (2021) Hydro-mechanical effects of several riparian vegetation combinations on the streambank stability — a benchmark case in southeastern Norway. Sustainability (Switzerland) 13. https://doi.org/10.3390/su13074046

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

Corominas J, van Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K, Winter MG, Pastor M, Ferlisi S, Tofani V, Hervás J, Smith JT (2014)

Recommendations for the quantitative analysis of landslide risk. Bull Eng Geol Env 73:209–263. https://doi.org/10.1007/s10064-013-0538-8

Crawford MM, Dortch JM, Koch HJ, Killen AA, Zhu J, Zhu Y, Bryson LS, Haneberg WC (2021) Using landslide-inventory for a combined bagged-trees and logistic regression approach to landslide susceptibility in eastern Kentucky. United States, Quarterly Journal of Engineering Geology and Hydrogeology. https://doi.org/10.1144/qjegh2020-177

Dahal A, Lombardo L (2023) Explainable artificial intelligence in geoscience: a glimpse into the future of landslide susceptibility modeling. Comput Geosci 176. https://doi.org/10.1016/j.cageo.2023.105364

Distefano P, Peres DJ, Scandura P, Cancelliere A (2022) Brief communication: introducing rainfall thresholds for landslide triggering based on artificial neural networks. Nat Hazard 22:1151–1157. https://doi.org/10.5194/nhess-22-1151-2022

Dyrrdal AV, Jonas O, Erika M, Karsten AN, Piia P, Svetlana A, Søren T, Eirik F, Lennart W (2021) Observed changes in heavy daily precipitation over the Nordic-Baltic region. https://doi.org/10.1016/j.ejrh.2021.100965

Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Eng Geol 102:85–98. https://doi.org/10.1016/j.enggeo.2008.03.022

Friedman J (2001) Greedy function approximation: a gradient boosting machine. Ann Statistics 29:1189–1232. https://doi.org/10.1214/aos/1013203451

Friedman J, Popescu BE (2008) Predictive learning via rule ensembles. The Annals of Applied Statistics 2(3):916–954

Fu Q, Hou R, Li L, Wang M, Yan J (2018) The functions of soil water and heat transfer to the environment and associated response mechanisms under different snow cover conditions, Geoderma, Volume 325. ISSN 325:9–17. https://doi.org/10.1016/j.geoderma.2018.03.022. (ISSN 0016–7061)

Gariano SL, Guzzetti F (2016) Landslides in a Changing Climate. https://doi.org/10.1016/j.earscirev.2016.08.011,1November

Giannecchini R, Galanti Y, D'Amato Avanzi G (2012) Critical rainfall thresholds for triggering shallow landslides in the Serchio River Valley (Tuscany, Italy). Natural Hazards and Earth System Science 12:829–842. https://doi.org/10.5194/nhess-12-829-2012

Glade T, Crozier M, Smith P (2000) Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "antecedent daily rainfall model." Pure Appl Geophys 157:1059–1079. https://doi.org/10.1007/s000240050017

Goetz JN, Brenning A, Petschkoc H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 81:1–11

Guzzetti F, Gariano SL, Peruccacci S, Brunetti MT, Marchesini I, Rossi M, Melillo M (2020) Geographical landslide early warning systems. Earth Sci Rev 200. https://doi.org/10.1016/j.earscirev.2019.102973

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2008) The rainfall intensity-duration control of shallow landslides and debris flows an update. Landslides 5. https://doi.org/10.1007/s10346-007-0112-1

Hanssen-Bauer I, Førland EJ, Haddeland I, Hisdal H, Lawrence D, Mayer S, Nesje A, Nilsen JEØ, Sandven S, Sandø AB, Sorteberg A, Ådlandsvik B (2017) Climate in Norway 2100 - a knowledge base for climate adaption. Nor Cent Clim Serv 1(2017):49. ISSN 2387-3027

Harr RD (1981) Some characteristics and consequences of snowmelt during rainfall in western Oregon. J Hydrol 53(3–4):277–304

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. New York Springer-Verlag. https://doi.org/10.1007/978-0-387-21606-5

Herrera G, Mateos RM, García-Davalillo JC, Grandjean G, Poyiadji E, Maftei R, Filipciuc TC, Jemec Auflič M, Jež J, Podolszki L, Trigila A, Iadanza C, Raetzo H, Kociu A, Przyłucka M, Kułak M, Sheehy M, Pellicer XM, McKeown C, Ryan G, Kopačková V, Frei M, Kuhn D, Hermanns RL, Koulermou N, Smith CA, Engdahl M, Buxó P, Gonzalez M, Dashwood C, Reeves H, Cigna F, Lik P, Pauditš P, Mikulėnas V, Demir V, Raha M, Quental L, Sandić C, Fusi B, Jensen OA (2018) Landslide databases in the Geological Surveys of Europe. Landslides 15:359–379. https://doi.org/10.1007/s10346-017-0902-z

Heyerdahl H, Høydal ØA (2017) Geomorphology and susceptibility to rainfall triggered landslides in Gudbrandsdalen Valley. Norway, in

Advancing Culture of Living with Landslides, Springer International Publishing 267–279. https://doi.org/10.1007/978-3-319-53485-5_31

Ishikawa T, Miura S (2011) Influence of freeze-thaw action on deformation-strength characteristics and particle crushability of volcanic coarse-grained soils. Soils Found 51(5):785–799

Ishikawa T, Tokoro T, Seiichi M (2015) Geohazard at volcanic soil slope in cold regions and its influencing factors. Japanese Geotechnical Society Special Publication 1(1):1–20

Ishikawa T, Kijiya I, Tokoro T, Sato M (2016) Application of coupled thermo-hydro-mechanical analysis to frost-heave behavior of earth structures. Japanese Geotechnical Society Special Publication 2(13):531–536

Jenks GF (1967) The data model concept in statistical mapping. International Yearbook of Cartography 7:186–190

Johnsen T (2010) Late Quaternary ice sheet history and dynamics in central and southern Scandinavia. Stockholm University, Diss. Department of Physical Geography and Quaternary Geology

Kim SW, Chun KW, Kim M et al (2021) Effect of antecedent rainfall conditions and their variations on shallow landslide-triggering rainfall thresholds in South Korea. Landslides 18:569–582. https://doi.org/10.1007/s10346-020-01505-4

Krøgli IK, Devoli G, Colleuille H, Boje S, Sund M, Engen IK (2018) The Norwegian forecasting and warning service for rainfall- and snowmelt-induced landslides. Nat Hazard 18:1427–1450. https://doi.org/10.5194/nhess-18-1427-2018

Letten K, and Blikra LH (2007) Holocene colluvial (debris-flow and waterflow) processes in eastern Norway stratigraphy, chronology and palaeoenvironmental implications. J Quaternary Sci Vol. 22 pp. 619–635. ISSN 0267–8179

Li B, Liu K, Wang M, He Q, Jiang Z, Zhu W, Qiao N (2022) Global dynamic rainfall-induced landslide susceptibility mapping using machine learning. Remote Sens (Basel) 14. https://doi.org/10.3390/rs14225795

Liaw A, Wiener M (2002) Classification and regression by random forest. R News 2:18–22

Lima P, Steger S, Glade T, Murillo-García FG (2022) Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility. J Mt Sci 19:1670–1698. https://doi.org/10.1007/s11629-021-7254-9

Liu Z, Gilbert G, Cepeda JM, Lysdahl AOK, Piciullo L, Hefre H, Lacasse S (2021) Modelling of shallow landslides with machine learning algorithms. Geosci Front 12:385–393. https://doi.org/10.1016/j.gsf.2020.04.014

Löbmann MT, Geitner C, Wellstein C, and Zerbe S (2020) The influence of herbaceous vegetation on slope stability – a review, https://doi.org/10.1016/j.earscirev.2020.103328

Lussana C, Saloranta T, Skaugen T, Magnusson J, Einar Tveito O, Andersen J (2018) SeNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. Earth Syst Sci Data 10:235–249. https://doi.org/10.5194/essd-10-235-2018

Lussana C, Einar Tveito O, Dobler A, Tunheim K (2019) SeNorge_2018, daily precipitation, and temperature datasets over Norway. Earth Syst Sci Data 11:1531–1551. https://doi.org/10.5194/essd-11-1531-2019

Luti T, Segoni S, Catani F, Munafò M, Casagli N (2020) Integration of remotely sensed soil sealing data in landslide susceptibility mapping. Remote Sens (Basel) 12. https://doi.org/10.3390/RS12091486

Masi EB, Segoni S, Tofani V (2021) Root Reinforcement in Slope Stability Models: a Review. https://doi.org/10.3390/geosciences11050212,1May

Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B (2020) Machine Learning Methods for Landslide Susceptibility Studies: a Comparative Overview of Algorithm Performance. https://doi.org/10.1016/j.earscirev.2020.103225,1August

Mitchell T (1997) Introduction to machine learning. Mach Learn 7:2–5

Molnar C (2020) Interpretable machine learning. [Online]. Available: https://christophm.github.io/interpretable-ml-book/

Mondini AC, Guzzetti F, Melillo M (2023) Deep learning forecast of rainfall-induced shallow landslides. Nat Commun 14. https://doi.org/10.1038/s41467-023-38135-y

Ng CWW, Yang B, Liu ZQ, Kwan JSH, Chen L (2021) Spatiotemporal modelling of rainfall-induced landslides using machine learning. Landslides 18:2499–2514. https://doi.org/10.1007/s10346-021-01662-0

Nocentini N, Medici C, Barbadori F, Gatto A, Franceschini R, del Soldato M, Rosi A, Segoni S (2023a) Optimization of rainfall thresholds for landslide early warning through false alarm reduction and a multi-source validation. Landslides 21:557–571. https://doi.org/10.1007/s10346-023-02176-7

Nocentini N, Rosi A, Segoni S, Fanti R (2023b) Towards landslide space-time forecasting through machine learning: the influence of rainfall parameters and model setting. Front Earth Sci (Lausanne) 11. https://doi.org/10.3389/feart.2023.1152130

Palau RM, Berenguer M, Hürlimann M, Sempere-Torres D (2022) Application of a fuzzy verification framework for the evaluation of a regional-scale landslide early warning system during the January 2020 Gloria storm in Catalonia (NE Spain). Landslides 19:1599–1616. https://doi.org/10.1007/s10346-022-01854-2

Park JY, Lee SR, Lee DH, Kim YT, Lee JS (2019) A regional-scale landslide early warning methodology applying statistical and physically based approaches in sequence. Eng Geol 260. https://doi.org/10.1016/j.enggeo.2019.105193

Pecoraro G, and Calvello M (2021) Definition and first application of a probabilistic warning model for rainfall-induced landslides, 181–187, https://doi.org/10.1007/978-3-030-60311-3_20

Pereira S, Zêzere JL, Bateira C (2012) Technical Note: Assessing predictive capacity and conditional independence of landslide predisposing factors for shallow landslide susceptibility models. Natural Hazards and Earth System Science 12:979–988. https://doi.org/10.5194/nhess-12-979-2012

Pham BT, Pradhan B, Tien Bui D, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). Environ Model Softw 84:240–250. https://doi.org/10.1016/j.envsoft.2016.07.005

Piciullo L, Calvello M, Cepeda JM (2018). Territorial Early Warning Systems for Rainfall-Induced Landslides. https://doi.org/10.1016/j.earscirev.2018.02.013,1April

Piciullo L, Capobianco V, Heyerdahl H (2022) A first step towards a IoT-based local early warning system for an unsaturated slope in Norway. Nat Hazards 114:3377–3407. https://doi.org/10.1007/s11069-022-05524-3

Ponziani F, Pandolfo C, Stelluti M, Berni N, Brocca L, Moramarco T (2012) Assessment of rainfall thresholds and soil moisture modeling for operational hydrogeological risk prevention in the Umbria region (central Italy). Landslides 9:229–237

Pyakurel A, Dahal BK, Gautam D (2021) Does machine learning adequately predict earthquake induced landslides? Soil Dyn Earthq Eng, Volume 171, 107994, ISSN 0267–7261, https://doi.org/10.1016/j.soildyn.2023.107994

Ramberg IB, Bryhni I, Norrvedr A, Rangnes K (2008) The making of a land - geology of Norway. Trondheim Norsk Geologisk Forening, 6xx p

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018). A Review of Statistically-Based Landslide Susceptibility Models. https://doi.org/10.1016/j.earscirev.2018.03.001,1May

Rosi A, Segoni S, Canavesi V, Monni A, Gallucci A, Casagli N (2021) Definition of 3D rainfall thresholds to increase operative landslide early warning system performances. Landslides 18:1045–1057. https://doi.org/10.1007/s10346-020-01523-2

Saloranta TM (2016) Operational snow mapping with simplified data assimilation using the seNorge snow model. J Hydrol 538:314–325

Saloranta TM, New version (v.1.1.1) of the seNorge snow model and the snow maps for Norway. NVE Report 6–2014, 2014

Schilirò L, Cepeda J, Devoli G, Piciullo L (2021) Regional analyses of rainfall-induced landslide initiation in upper Gudbrandsdalen (South-Eastern Norway) using TRIGRS model. Geosciences (basel). https://doi.org/10.3390/geosciences1101

Segoni S, Piciullo L, Gariano SL (2018a). A Review of the Recent Literature on Rainfall Thresholds for Landslide Occurrence. https://doi.org/10.1007/s10346-018-0966-4

Segoni S, Tofani V, Rosi A, Catani F, Casagli N (2018b) Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. Front Earth Sci (Lausanne) 6. https://doi.org/10.3389/feart.2018.00085

Segoni S, Nocentini N, Rosi A, Luti T, fico GP, Munafò M, Casagli N, Catani F (2021) New explanatory variables to improve landslide susceptibility mapping: testing the effectiveness of soil sealing information and multi-criteria geological parameterization. Italian Journal of Engineering Geology and Environment 209–220. https://doi.org/10.4408/IJEGE.2021-01.S-19

Slagstad T, Davidsen B, Daly JS (2011) Age and composition of crystalline basement rocks on the Norwegian continental margin: offshore extension and continuity of the Caledonian-Appalachian orogenic belt. J Geol Soc 168:1167–1185

Solheim A, Kalsnes B, Strout J, Piciullo L, Heyerdahl H, Eidsvig U, Lohne J (2022) Landslide risk reduction through close partnership between research, industry, and public entities in Norway pilots and case studies. Front Earth Sci (Lausanne) 10. https://doi.org/10.3389/feart.2022.855506

Stanley TA, Kirschbaum DB, Benz G, Emberson RA, Amatya PM, Medwedeff W, Clark MK (2021) Data-driven landslide nowcasting at the global scale. Front Earth Sci (Lausanne) 9. https://doi.org/10.3389/feart.2021.640043

Steger S, Mair V, Kofler C, Pittore M, Zebisch M, Schneiderbauer S (2021) Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling – benefits of exploring landslide data collection effects. Sci Total Environ 776: Article 145935. https://doi.org/10.1016/j.scitotenv.2021.145935

Steger S, Moreno M, Crespi A, Zellner PJ, Gariano SL, Brunetti MT, Melillo M, Peruccacci S, Marra F, Kohrs R, Goetz J, Mair V, Pittore M (2023) Deciphering seasonal effects of triggering and preparatory precipitation for improved shallow landslide prediction using generalized additive mixed models. Nat Hazard 23:1483–1506. https://doi.org/10.5194/nhess-23-1483-2023

Tehrani FS, Calvello M, Liu Z, Zhang L, Lacasse S (2022) Machine learning and landslide studies: recent advances and applications. Nat Hazards. https://doi.org/10.1007/s11069-022-05423-7

Tran TV, Alvioli M, Lee G, An HU (2018) Three-dimensional, time-dependent modelling of rainfall-induced landslides over a digital landscape a case study. Landslides 15:1071–1084. https://doi.org/10.1007/s10346-017-0931-7

Vannocci P, Segoni S, Masi EB, Cardi F, Nocentini N, Rosi A, Bicocchi G, D'Ambrosio M, Nocentini M, Lombardi L, Tofani V, Casagli N, Catani F (2022) Towards a national-scale dataset of geotechnical and hydrological soil parameters for shallow landslide modeling. Data (Basel) 7:37. https://doi.org/10.3390/data7030037

Varnes DJ, IAEG (International Association of Engineering Geology) (1984) Landslide hazard zonation: a review of principles and practice. UNESCO, Paris

Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, Chandak P, Liu S, Van Katwyk P, Deac A, Anandkumar A, Bergen K, Gomes CP, Ho S, Kohli P, Lasenby J, Leskovec J, Liu TY, Manrai A, Marks D, Ramsundar B, Song L, Sun J, Tang J, Veličković P, Welling M, Zhang L, Coley CW, Bengio Y, Zitnik M (2023) Scientific Discovery in the Age of Artificial Intelligence. https://doi.org/10.1038/s41586-023-06221-2,3August

Nicola Nocentini (✉) · Ascanio Rosi
Department of Geosciences, University of Padua, Via G. Gradenigo 6, 35131 Padua, Italy
Email: nicola.nocentini@unipd.it

Nicola Nocentini · Samuele Segoni · Riccardo Fanti
Department of Earth Sciences, University of Florence, Via La Pira 4, 50121 Florence, Italy

Luca Piciullo · Zhongqiang Liu
Department of Natural Hazards, Norwegian Geotechnical Institute (NGI), Sandakerveien 140, 0484 Oslo, Norway

Luca Piciullo
Department of Built Environment, Oslo Metropolitan University (OsloMet), Pilestredet 35, 0166 Oslo, Norway