Geoscience Frontiers 12 (2021) 385–393

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

Research Paper

Modelling of shallow landslides with machine learning algorithms

Zhongqiang Liu *, Graham Gilbert, Jose Mauricio Cepeda, Asgeir Olaf Kydland Lysdahl, Luca Piciullo, Heidi Hefre, Suzanne Lacasse

Norwegian Geotechnical Institute (NGI), Sognsveien 72, 0855, Oslo, Norway

ARTICLE INFO

Keywords:
Shallow landslide
Spatial modelling
Machine learning
GIS

ABSTRACT

This paper introduces three machine learning (ML) algorithms, the 'ensemble' Random Forest (RF), the 'ensemble' Gradient Boosted Regression Tree (GBRT) and the MultiLayer Perceptron neural network (MLP) and applies them to the spatial modelling of shallow landslides near Kvam in Norway. In the development of the ML models, a total of 11 significant landslide controlling factors were selected. The controlling factors relate to the geomorphology, geology, geo-environment and anthropogenic effects: slope angle, aspect, plan curvature, profile curvature, flow accumulation, flow direction, distance to rivers, water content, saturation, rainfall and distance to roads. It is observed that slope angle was the most significant controlling factor in the ML analyses. The performance of the three ML models was evaluated quantitatively based on the Receiver Operating Characteristic (ROC) analysis. The results show that the 'ensemble' GBRT machine learning model yielded the most promising results for the spatial prediction of shallow landslides, with a 95% probability of landslide detection and 87% prediction efficiency.

1. Introduction

Precipitation-induced shallow landslides are common natural hazards of mountainous regions in Norway. They can cause massive casualties and significant losses and damage to property. An improved prediction of precipitation-induced landslides would help mitigate losses, reduce number of fatalities and significantly improve risk mitigation strategies for current and future climate scenarios.

Landslide prediction models can be categorized into two groups: numerical models and data-driven models. In the last years several physically-based numerical models have been proposed for simulating landslide initiation (e.g. Montgomery and Dietrich, 1994; Pack et al., 1999; Montrasio and Valentino, 2007; Baum et al., 2008; Liao et al., 2011). This type of models generally accounts for the spatial variability of the involved parameters (e.g. physical-mechanical parameters of the slope material, rainfall intensity), thus they appear most suitable for determining the occurrence of site-specific shallow landslides (Schilirò et al., 2016). However, the implementation of sufficiently representative and accurate numerical models can be time-consuming and expensive. As of now, numerical models are used to analyse the time-varying landslide susceptibility, but often the idealisation in the computed model can limit the representativeness of the case studies.

Compared with numerical models, the data-driven models are often more popular because of a perceived simpler process, often more accurate prediction and lower costs (Corominas et al., 2005; Zhou et al., 2018). Data-driven models, especially machine learning techniques, have been widely applied for "learning" the relationship among landslide occurrence and landslide related predictors. Such models have also achieved good performances.

A review of available approaches indicates that various machine learning algorithms have been applied for modelling landslides, such as support vector machines (Pourghasemi et al., 2013; Kumar et al., 2017), artificial neural network (Zare et al., 2013; Arnone et al., 2014), and decision tree (Wu et al., 2014). Recently, the performance of machine learning methods used separately has been shown to improve significantly when 'ensemble' techniques were used together, such as random forest (RF) (Hong et al., 2016; Pourghasemi and Kerle, 2016) and the gradient boosted regression tree (GBRT) (Dickson and Perry, 2016). Ensemble method is a machine learning technique that combines several base ML models to produce one optimal predictive model. Ensemble techniques utilize multiple learning algorithms to generate hybrid models, and can thus efficiently handle complex input to produce an output closer to the functions modelled (Pham et al., 2018).

The objective of the present study is to demonstrate that 'ensemble'

* Corresponding author.
E-mail address: zhongqiang.liu@ngi.no (Z. Liu).
Peer-review under responsibility of China University of Geosciences (Beijing).

https://doi.org/10.1016/j.gsf.2020.04.014
Received 25 November 2019; Received in revised form 7 March 2020; Accepted 22 April 2020
Available online 6 May 2020

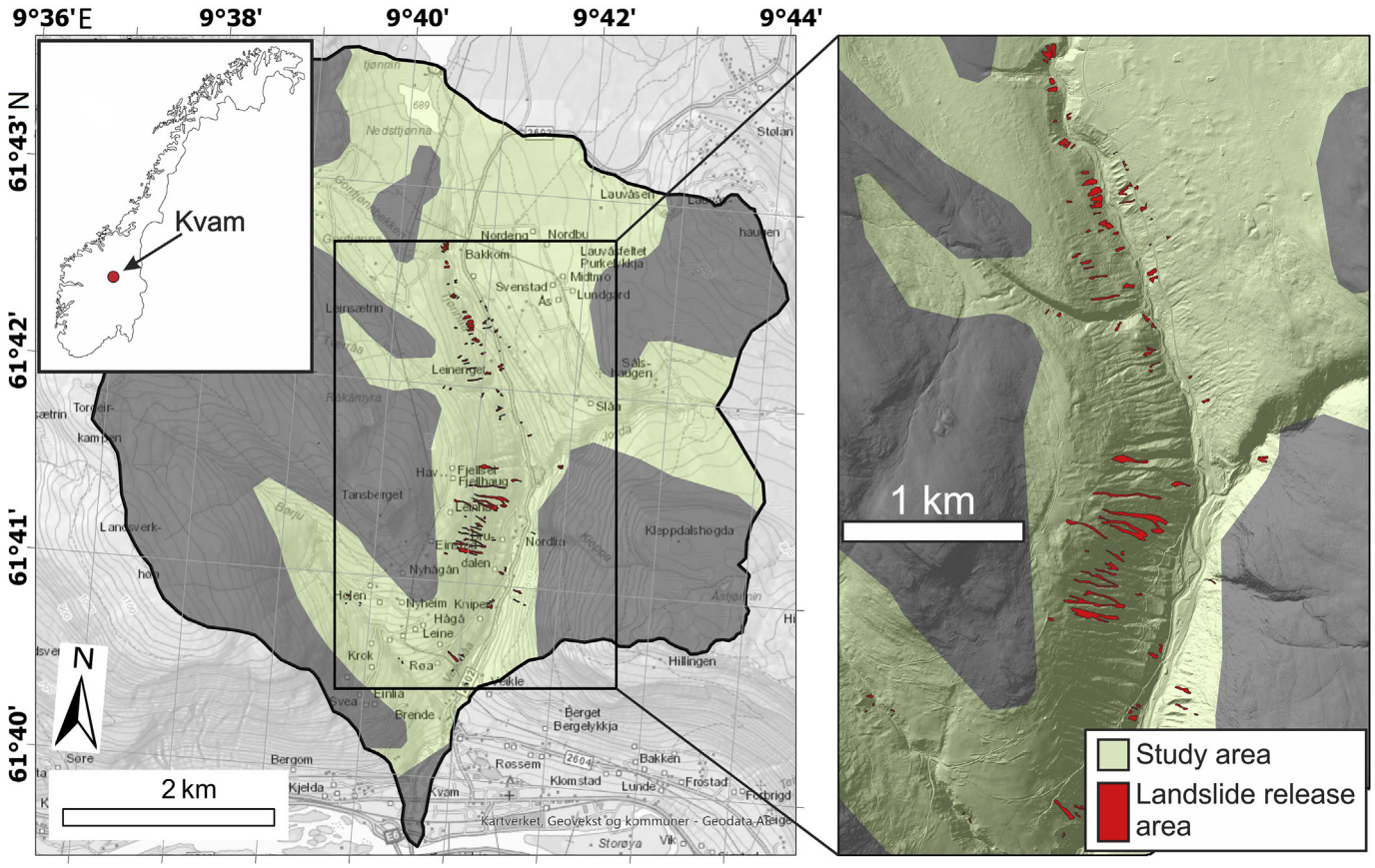

Fig. 1. Spatial extent of the landslide release in study area during 2011 rainfall.

machine learning techniques can be applied to a large, heterogeneous area containing a number of diverse landslide typologies in order to predict the spatial extent of shallow landslides. The following sections present the study area in the county of Innlandet in Norway, introduce the three well-known machine learning algorithms and describe the geomorphological, geological, geo-environmental and anthropogenic factors considered when modelling the spatial extent of landslides. The relative importance of the landslide conditioning factors is analysed and the accuracy of the ML models is compared. Data processing and analysis was done with ArcMap 10.2 and Python 3.7.3 software.

2. Study area

The study area is in the Veikledalen Valley (61°41.56′N; 9°41.65′E), a tributary to the Gudbrandsdalen Valley in Innlandet County in eastern Norway (Fig. 1). The village of Kvam is situated at the head of Veikledalen. The valley encompasses an area of about 32 km^2, whereof 69% (22 km^2) is forested by a temperate broad leaf and mixed coniferous forest. Agriculture land covers 7.2% (2.3 km^2) of the area, while bog and fell landscapes account for the remainder of the area. The area has a subarctic climate with cold winters and warm summers. The mean annual temperature in Kvam is about 3.3 °C (1981–2010). Mean diurnal temperatures typically range from −22 °C to 20 °C. Approximately 500 mm of precipitation occurs annually, whereof approximately 40% as snow.

The large-scale, regional geomorphology is characterized by U-shaped, glacier-incised valleys, which formed during repeated glaciation throughout the Quaternary. Within the study area, the elevation ranges from approximately 250–1190 m. The geology consists of green or grey phyllite, interlayered with grey sandstone, feldspathic sandstone and quartzite (Siedlecka et al., 1987). The unconsolidated deposits in the valley bottom are primarily glaciofluvial or fluvial in origin, with accumulations of colluvium at the base of over-steepened bedrock cliffs (Sletten and Blikra, 2007). Glacial deposits (or till), of variable thickness, blanket the slopes. In general, deposits are thinnest at higher elevations (typically less than 0.5 m) while sediment cover in the valley bottom can exceed 20 m. Locally, the till blanket is dissected by numerous small gullies and erosion tracts, and earlier landslides. The village of Kvam is built on a large alluvial fan at the head of Veikledalen.

2.1. Landslide inventory and event descriptions

The major natural hazards in the study area are soil slides, debris flows and floods, snow avalanches. Rockfalls are also serious geohazards in certain locations. The Norwegian landslide database indicates that past events were often hydrometeorological in origin, with the timing of release being related to periods of snowmelt or intense rainfall, frequently in combination. The event database for the investigation in this paper comprises landslide release zones and runout areas triggered by the intensive rainfall on 9–10 June 2011. The inventory was compiled from field survey observations in conjunction with high-resolution aerial photographs taken in the weeks following the landslide events. In total, 86 landslides were identified. This 2011 intensive rainfall resulted in a comprehensive landslide inventory and was selected for further investigation in this study. The location of the landslides recorded in the database is shown in Fig. 1.

Significant rainfall preceded the landslide event on 9–10 June 2011. The nearest meteorological station, located at Sjoa, 7 km WNW of Kvam, recorded 72 mm of rainfall between June 7 and 10, 2011, whereof 60 mm rainfall was reported for June 10. Air temperatures were also high during this time, leading to intense snowmelt in high-elevation areas. The shallow landslides triggered during this rainfall-snowmelt period resulted in the closure of local roads, the national (Rv3) and European (E6) highways, and closure of the railroad in Gudbrandsdalen. In total,

Table 1
Confusion matrix showing four potential outcomes of the classification (classifier).

		True class (observations)	
		Positives	Negatives
Hypothesized class (predictions or simulations)	"1" (landslide)	True Positives, *TP*	False Positives, *FP*
	"0" (no landslide)	False Negatives, *FN*	True Negatives, *TN*

270 persons were temporarily relocated and the damages were estimated to 800 million NOK (ca. 100 million USD). In the last decades proceeding year 2008, landslide activity in Gudbrandsdalen was comparatively modest. However, increased landslide occurrence was expected as a consequence of potential increased rainfall intensity and snowmelt (Jaedicke et al., 2008).

2.2. Preparation of the training and validation datasets

Spatial landslide mapping is considered to be a binary classification in which the landslide index is separated into two classes, i.e., landslide and non-landslide. Landslide pixels are assigned a value of "1", and non-landslide pixels are assigned a value of "0".

In this study, the area comprised 535,839 location points, of which 3399 were release areas. To avoid unbalanced data points between landslide and non-landslide, 6798 non-landslide points, i.e. twice the number of release areas, were randomly selected. In total, 10,197 location points were considered in the dataset. The 3399 landslide points were then randomly split into two parts: 70% of the data was used for training the ML landslide models and 30% were kept for model validation. The 6798 non-landslide points were also randomly split into the same ratio of 70/30. This ratio of training vs validation data has been used before for similar size of database (Tien Bui et al., 2016; Chen et al., 2018; Yang et al., 2019).

3. Machine learning (ML) algorithms for landslide prediction

This paper compares the performance of three machine learning algorithms in predicting the triggering of landslides in Kvam, Norway: the random forest (RF), the gradient boosted regression tree (GBRT) and the multilayer perceptron (MLP). A recent comparison of the performance of 10 advanced machine learning techniques proved that ensemble methods, e.g. the RF and GBRT methods, showed the best performance for modelling landslide susceptibility (Pourghasemi and Rahmati, 2018; Dou et al., 2019; Park and Kim, 2019). An 'ensemble' method is a technique that creates and combines multiple models in order to improve the final result. Furthermore, the paper looks into the multilayer perceptron (MLP) neural networks that is another known method for landslide spatial prediction (Zare et al., 2013; Alkhasawneh et al., 2014).

3.1. Random forest (RF) algorithm

A random forest (RF) is an ensemble learning method for classification and regression consisting in the definition of multiple decision trees (Breiman, 2001). The RF method generates uncorrelated decision trees that operate as an ensemble. RF generates thousands of random decision trees to form a forest. Each tree is grown based on a re-sampling, using a classification and regression decision tree procedure with a random subset of variables selected at each node (Micheletti et al., 2014; Pourghasemi and Kerle, 2016; Rahmati et al., 2018; Zhang et al., 2019). Thus, the aim of the RF is to create a large number of uncorrelated decision tree models to produce more accurate predictions. The decisions on class affiliation ("landslide" or "non-landslide") and model construction are determined by the majority vote among all trees (Micheletti et al., 2014; Park and Kim, 2019). In order to make reliable predictions with the 'ensemble' model, at least two conditions should be verified: the selected features should have some predictive ability; the prediction of the different decision tree models need to be uncorrelated.

3.2. Gradient boosted regression tree (GBRT) algorithm

The gradient boosted regression tree (GBRT) is a model combining two techniques: boosting and regression. These techniques are combined to improve the model accuracy and decrease the variance. As the RF model, the GBRT defines different decision trees to improve the performance of the prediction taking a random subset of variables. While RF models use the bagging method, which means that each occurrence has an equal probability of being selected in subsequent samples, GBRT use the boosting method in which the input data are weighted in subsequent trees (De'Ath, 2007; Elith et al., 2008; Franklin, 2010). Basically, the subsequent trees are useful to better define the not-corrected observations of the previous trees. The idea behind the boosting is to improve upon the prediction of the previous trees.

In a GBRT technique, two parameters need to be specified: tree complexity (tc) and learning rate (lr). The first defines the number of splits of each tree. The second defines the number of trees to be built.

The GBRT technique was adopted in this study to compare with the RF technique. Other boosting methods, e.g. Adaboost, XGBoost (Zhang et al., 2020), could be also used for the prediction of landslide distribution.

3.3. Multilayer perceptron neural network (MLP) algorithm

The multilayer perceptron neural network (MLP) technique is perhaps the most popular and most widely used artificial neural networks (ANN). The perceptron is an algorithm for supervised learning of binary classifiers. The ANN-MLP technique consists of three or more layers: an input layer, an output layer and one or more hidden layers. The nodes of the input layer are linked to the hidden layers through connections that can have different weights as a function of their importance. The nodes of the hidden layers receive a value that is the sum of the input variables multiplied by the weights of their connections. The weights between the nodes are adjusted with a learning algorithm.

The weighted sums in each nodes of the hidden layer are passed into a nonlinear activation function. The outcome of the activation function is then passed on to the next hidden layer or the output layer (Lek and Guégan, 1999; Olden and Jackson, 2002; Franklin, 2010). The connections between the hidden layers and the output layer are also weighted. The value at the output nodes is the result of the weighted sum of the hidden nodes. The performance and the error of the model to predict the real outcomes are then evaluated. The ANN-MLP utilizes a supervised learning technique with backpropagation algorithm to reduce the difference between the actual output and the neural network output results.

3.4. Classifiers and performance indicators for the ML methods

To verify the performance of the ML models for the prediction of spatial landslide occurrence, statistical evaluation measures based on the

Table 2
Performance indicators.

Formula	Range	Optimal value
$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	[0,1]	1
$POD = \frac{TP}{TP + FN}$	[0,1]	1
$POFD = \frac{FP}{FP + TN}$	[0,1]	0
$EI = \frac{TP}{FP + TP + FN}$	[0,1]	1

Receiver Operating Characteristic (ROC) analysis were adopted in this study. The ROC analysis was first developed by engineers during World War II for detecting enemy objects in battlefields. A binary classifier system was used. Soon, the method was applied in many fields from medicine to natural hazards, and machine learning.

In ROC analysis, 'true' class instances (observations) are compared with hypothesised class instances (predictions or simulations) using a classification model or classifier (Hosmer and Lemeshow, 1989; Fawcett, 2006). The observations are classified in either positive or negative depending on whether they indicate occurrence or non-occurrence of, in this case, a landslide. Table 1 presents the four possible outcomes of this classification using a confusion matrix in a binary classification context, i.e. the class labels 1 and 0 (Wilks, 1995; Accadia et al., 2003; Cepeda et al., 2010; Staley et al., 2013). The orange and green cells in Table 1 indicate a correct classification, and the two hatched cells indicate an incorrect classification.

Several skill scores, or performance indicators, are proposed in the literature for the calculation of the performance indicators of the confusion matrix. Four indicators are used for the performance evaluation of the three ML methods investigated herein: *Accuracy* (*ACC*), *Probability of Detection* (*POD*), *Probability of False Detection* (*POFD*) and *Efficiency Index* (*EI*). The performance indicators are defined in Table 2, along with their range and the optimal (ideal) value for a 100% reliable prediction.

Accuracy (*ACC*) is the most used indicator to analyse the confusion matrix. It gives an overall evaluation of the number of correct predictions ($TN + TP$) over the total number of predictions. The value ranges from 0 or no correct predictions, to 1 where 100% of the predictions are correct.

Probability of Detection (*POD*), also called Sensitivity or Hit rate or True positive rate, represents the rate of occurred events positively predicted, where 0 means none of the occurred events were predicted, and 1 where all the events having occurred were predicted correctly.

Probability of False Detection (*POFD*) is the counterpart of *POD* since it measures the ratio of forecasted events to have occurred when no events have been observed. It ranges between 0 and 1.

The last indicator for the performance evaluation of the three ML methods is the *Efficiency Index* (*EI*), also called Threat score or Critical success index. It is the ratio of true positives (successfully predicted landslides) to the sum of the true positives and the unsuccessful predictions. The indicator evaluates the performance without considering *TN*, which number often lies an order of magnitude above the other classifiers in the confusion matrix (Table 1). In other words, the model does not get credit for the correctly predicted 'non-landslide' situations. Such indicator has been used for the validation of rainfall thresholds (Tiranti and Rabuffetti, 2010; Staley et al., 2013), the performance evaluation of landslide early warning systems (Calvello and Piciullo, 2016; Piciullo et al., 2017) and landside susceptibility mapping (Tien Bui et al., 2016; Pham et al., 2018). Not taking into account the *Efficiency Index* (*EI*) indicator can often lead to an overestimation of the performance of a model (Calvello and Piciullo, 2016; Piciullo et al., 2017).

4. Landslide conditioning and triggering factors

Eight static landslide conditioning factors and three time-dependent triggering factors were selected for input into the machine learning models: slope angle, aspect, plan curvature, profile curvature, flow accumulation, flow direction, distance to rivers, total water content, saturation, rainfall and distance to roads, illustrated in 11 vignettes in Fig. 2. There is no fixed list of relevant input parameters. Conditioning and triggering factors are generally selected based on the type, scale of the analysis and study area characteristics. The parameters selected for this investigation were chosen based on local conditions, data availability and the parameter relevant for a geo-mechanical analysis of a landslide. Maps of the landslide controlling factors were converted into raster format with a spatial resolution of 5 m × 5 m prior to sampling for the machine learning models.

4.1. Surficial geology

A regional map of the Quaternary sediment cover (1:250,000) was used to constrain the study area. Four main Quaternary sediment types are present within the catchment: till (thick cover), till (thin cover), fluvial deposits, and peat (organic material). The sediment cover is used as a proxy to the geotechnical units and hence the mechanical properties of the material on the most critical failure planes for landslide initiation. The sediment type in the area selected is till with varying thickness up to 3 m.

4.2. Terrain parameters

A Digital Elevation Model (DEM) derived from LiDAR-data with 1 m spatial resolution was used to derive the terrain parameters. To improve the efficiency of the data processing, the 1-m DEM was resampled to reduce the spatial resolution to 5 m. Slope angle, aspect, curvature (both plan and profile), flow accumulation and flow direction were then derived from the 5-m DEM. The terrain slope (Fig. 2a) is used as a proxy for the inclination of the failure surface, which is an adequate assumption for infinite slopes. Such as approach is commonly used in shallow landslide analysis like those observed in the inventory in Kvam. The other terrain parameters (aspect, curvature, flow direction and flow accumulation) account for the landforms influencing landslide susceptibility. Aspect (Fig. 2b) identifies the downslope direction which affects moisture on slopes due to rainfall. Plan curvature (Fig. 2c) and profile curvature (Fig. 2d) are perpendicular to and parallel to the direction of the maximum slope, respectively, which help understand more accurately the flow across a surface. Flow accumulation (Fig. 2e) and flow direction (Fig. 2f) determine accumulated flow into each cell and the flow direction from each cell to its downslope neighbour (Jenson and Domingue, 1988).

4.3. Hydrometeorological parameters

The distance to brooks and rivers (Fig. 2g) was calculated as the minimum horizontal distance to the nearest water course. This parameter was taken as conditioning factor (i.e., not time-dependant) as no consideration was given to seasonal variations of the water courses. This parameter accounts for the influence of runoff and erosion in the initiation of landslides along water courses.

The water content of the soil column, percentage of soil saturation, and rainfall were all taken as time-dependent parameters. The water content of the soil column (Fig. 2h) and the percentage of soil saturation (Fig. 2i) were daily values modelled and interpolated on a 1 km^2 grid. The rainfall (Fig. 2j) was obtained from hourly meteorological radar measurements operated by the Norwegian Meteorological Office. These three parameters were the only triggering factors considered in the analyses. Each has an influence on the pore water pressure conditions (and variations) in the ground, particularly on the critical failure planes where the potential landslides are initiated.

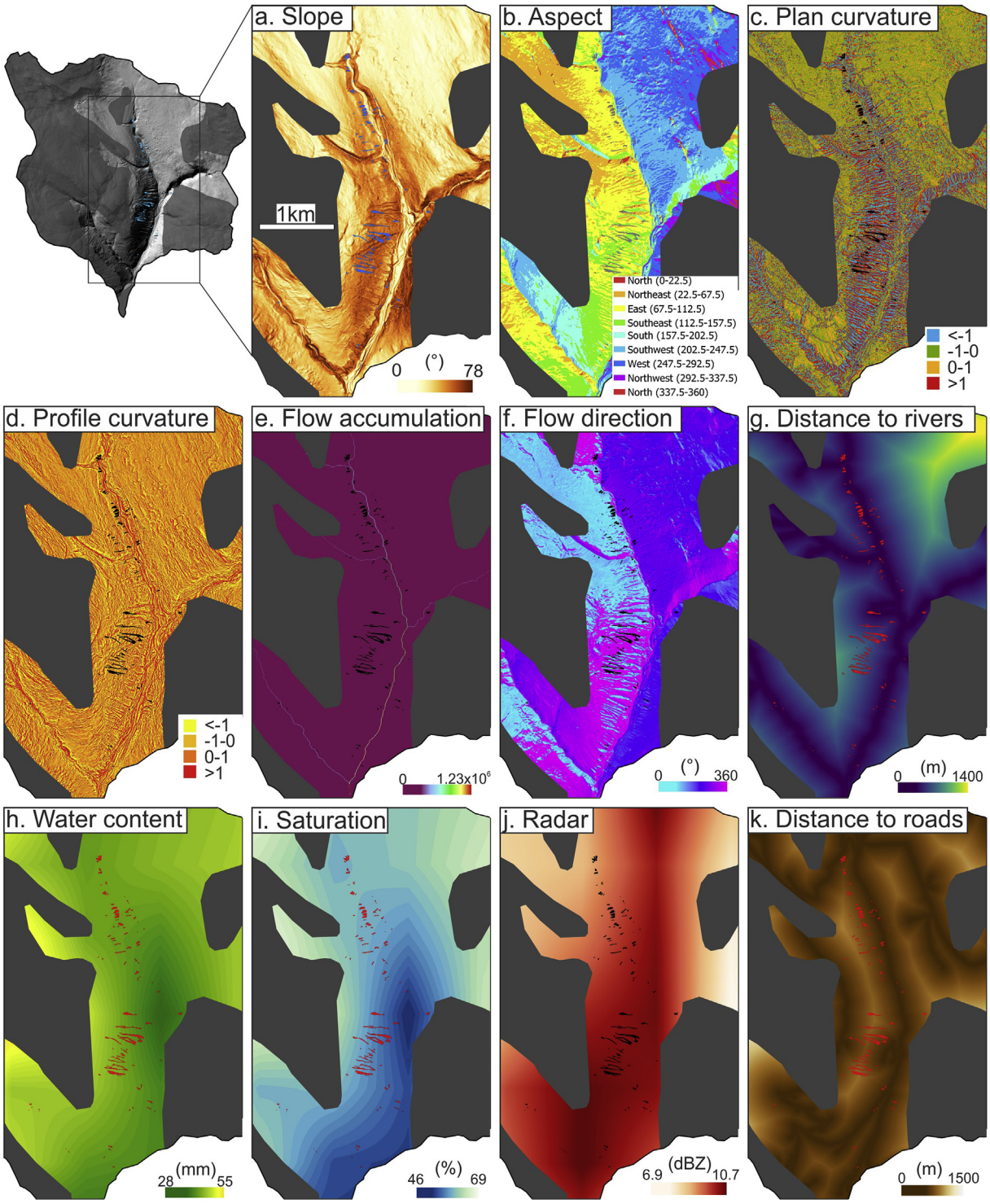

Fig. 2. Illustration of selected landslide conditioning and triggering factors in study area: (a) slope angle map, (b) aspect map, (c) plan curvature map, (d) profile curvature map, (e) flow accumulation map, (f) flow direction map, (g) distance to rivers map, (h) water content map, (i) saturation map, (j) rainfall from radar map and (k) distance to roads map.

4.4. Anthropogenic parameters

Distance to roads was the anthropogenic parameter involved in the analyses. It was calculated as the minimum distance to the nearest road. The distance to roads (Fig. 2k) can be considered as a proxy for the effects of the presence or absence of road elements. These effects on slope stability can be either improvement or reduction of safety (e.g., a fill at the toe or at the top of a slope, respectively), and it certainly increases the level of risk (higher consequence).

4.5. Selection of controlling factors

As part of the analysis, it is essential to remove irrelevant or less relevant factors in order to improve the efficiency and performance of landslide models. There are several techniques to quantify the predictive capacity of controlling factors, e.g. Information Gain Ratio (Tien Bui

Table 3
Importance (*VI*) of each landslide conditioning and triggering factor.

Class	Conditioning and triggering factor	Variable importance indicator, *VI*
Conditioning factors	Slope angle	0.29
	Plan curvature	0.1
	Aspect	0.08
	Distance to rivers	0.06
	Distance to roads	0.06
	Flow accumulation	0.05
	Flow direction	0.04
	Profile curvature	0.03
Triggering factors	Degree of saturation	0.11
	Rainfall	0.11
	Water content	0.07

et al., 2016), Least Support Vector Machine (Pham et al., 2018) and Gini information gain (Quinlan, 1993). The Gini information gain method was adopted in this study to assess the relative importance of each factor.

The evaluation and comparison of the predictive capability of the 11 landslide conditioning and triggering factors with the Gini information gain method using the RF algorithm indicate that all 11 factors contribute to the learning of the landslide prediction model, although the degree of each contribution differs, as indicated by the variable importance indicator *VI* in Table 3. The results of the Gini information gain analysis show that the slope has the highest importance (0.29), followed by plan curvature (0.1), aspect (0.08), distance to rivers (0.06), distance to roads (0.06), flow accumulation (0.05), flow direction (0.04), and profile curvature (0.03). Saturation (0.11) and rainfall (0.11) are the second and third most significant factors, and they are external triggering factors in the study area. Water content (0.07) is another major factor. It can be seen that 11 landslide controlling factors have positive values for landslide occurrences ($VI > 0$), therefore, all these factors were selected for generating datasets for training and testing the models. It should be also noted that the factors with lower importance (e.g. less than 0.05) could be neglected, and would probably not affect the prediction by the ML models.

A landslide occurrence can be described as the interaction of conditioning factors and triggering factors. The slope angle is a conditioning factor to the occurrence of landslides (i.e., in a flat area landslide occurrence is rare). Given a certain amount of rainfall and soil with homogenous characteristics, the higher or steeper the slope, the higher the probability of a landslide event. Moreover, the occurrence of landslides is not exclusively a matter of rainfall, but rather a combination of several conditioning and triggering. As shown in Table 3, the contribution of those three triggering factors (saturation, rainfall and soil water content) combined (0.11, 0.11, 0.07) have the same weight with the slope angle (0.29).

5. Results of analyses and comparisons of three ML models

5.1. Establishment of landslide models

With the selected 11 controlling factors, the grid search method was used to search the optimal parameters for each of the ML models. The search process can be described as follows: plausible values for each parameter are combined, and then all the combination values are used in the training of the model. The performance metric of each combination is typically measured by cross-validation on the training set (Hsu et al., 2010). When all the parameter combinations have been tried, an optimal parameter combination with the best performance is returned automatically. The optimal number of decision trees, the maximum depth of the tree and the minimum number of samples required to be at a leaf node with the RF algorithm were 48, 25 and 1, respectively. The optimal number of decision trees, the maximum depth of the tree, and the learning rate with the GBRT algorithm were 104, 9 and 0.18, respectively. The optimal number of neurons in the hidden layer and the regularization parameter for the MLP algorithm were 56 and 1. The solver for weight optimization for the MLP algorithm was 'lbfgs' that is an optimizer in the family of quasi-Newton methods.

Table 4
Model performance using the training dataset.

Parameter	ML model		
	RF	GBRT	MLP
True positive	2379	2379	2282
True negative	4758	4758	4540
False positive	0	0	218
False negative	1	0	97
ACC (%)	99.99	100	95.6
POD (%)	99.96	100	96
POFD (%)	0	0	4.6
EI (%)	99.96	100	87.9

Using the 11 controlling factors and the optimal model parameters, the RF, GBRT and MLP models were built using the training dataset with a total of 4758 non-landslide points and 2379 landslide points. Table 4 lists the parameters evaluating the performance of each model. All three models were quite successful in terms of accuracy (*ACC*), but significant differences were observed in terms of efficiency (*EI*). The GBRT has the highest performance with an accuracy of 100%, followed by RF (99.99%) and MLP (95.6%). A *POD* value of 1 indicates that the GBRT model correctly classifies pixels in the landslide class in all of the cases.

Overall, Table 4 shows that the predicted landslide and non-landslide points fitted well with the occurrence of landslides in the establishment of the model. The model was then used to predict spatial distribution of landslides.

5.2. Model validation

The ROC curve analysis results of different landslide models using the validation dataset are shown in Fig. 3. The results show that the GBRT and RF models have the highest AUC value of 0.99, followed by the MLP model (0.97), and the TRIGRS model (0.87). Table 5 presents the predictions of landslide occurrence (and non-occurrence) for each of the RF, GBRT and MLP model, using the validation dataset with a total of 2040 non-landslide points and 1020 landslide points. Fig. 4 illustrates the predictions of landslide occurrence with the RF, GBRT and MLP models. From Table 5, it is observed that GBRT has the highest performance with an accuracy of 95.4%, followed by RF (94.7%) and MLP (92.1%). Its efficiency is 87.3%.

The last column in Table 5 shows the performance of the results obtained by using the physically-based numerical model TRIGRS (Baum et al., 2008). TRIGRS is a hydro-mechanical model that couples infiltration (vertical flow only) and a limit-equilibrium model using an infinite slope assumption. The model allows modelling both unsaturated and saturated conditions (i.e., as in an advancing wetting front). TRIGRS is suited for shallow soil slides triggered by rainfall events with relatively short durations. The output of the model is the factor of safety, *FS*, for each cell in the domain. The factor of safety at a depth *Z* is calculated according to:

$$FS(Z,t) = \frac{\tan\varphi'}{\tan\delta} + \frac{c' - \psi(Z,t)\gamma_w \tan\varphi'}{\gamma Z \sin\delta\cos\delta} \quad (1)$$

where $\psi(Z, t)$ is the pressure head response, δ is the slope angle, c' is the effective cohesion of the soil, ϕ' is the effective friction angle of the soil, γ_w is the unit weight of groundwater and γ is the total unit weight of soil.

The distribution of *FS* in the study area is shown in the upper right panel in Fig. 4.

5.3. Discussion

In the development of the ML models (learning process) with the three different algorithms, the three approaches resulted in accuracy and probability of detection above 95%, and probability of false detection

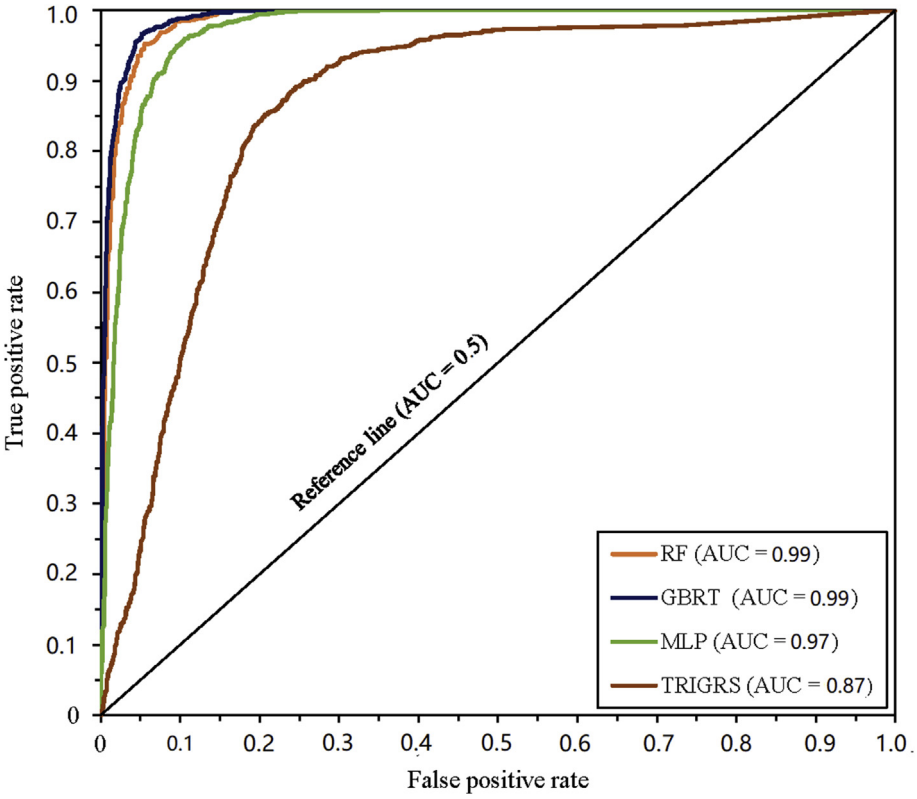

Fig. 3. The ROC curve analysis for four landslide models using the testing dataset.

less than 5%. The methods differed in terms of efficiency, where the GBRT gave an efficiency of 100%, the RF 99.96% and the MLP 88%.

For prediction of landslide occurrence with the remainder of the dataset (validation process), the three approaches again resulted in accuracies above 90%. This accuracy was much better than the prediction accuracy of the geomechanical model TRIGRS where the accuracy was about 82%. The probability of detection and the probability of non-detection were close to the values obtained under the learning process. However, the efficiency was considerably reduced for the three ML models. However, the efficiency of all three ML models was better than that of the geomechanical model TRIGRS.

Globally, the GBRT model gave the best results, and all three ML techniques gave a greatly improved prediction of landslide occurrence than the geomechanical TRIGRS model.

A possible refinement of the three ML methods could be to redo similar analyses at a few other sites to try to distinguish whether specific controlling factors are the reason for the difference in the accuracy and efficiency of the three ML models.

Table 5
Model prediction using the validation dataset.

Parameter	Model			
	RF	GBRT	MLP	TRIGRS
True positive	954	968	925	854
True negative	1944	1951	1893	1643
False positive	96	89	147	166
False negative	66	52	95	397
ACC (%)	94.7	95.4	92.1	81.6
POD (%)	93.5	94.9	90.7	68.3
POFD (%)	4.7	4.4	7.2	9.2
EI (%)	85.5	87.3	79.3	60.3

6. Conclusions and lessons learned

The prediction of the spatial distribution of shallow landslides is probably one of the most difficult tasks in landslide hazard and risk assessments. The reliability of a spatial landslide prediction model depends on the approach used. Physical models (e.g. centrifuge and model tests etc) are limited because of high costs and limitations on the number of tests that can be run, the material used for testing and the number of locations that can be tested, and by the small scale of most of the tests run (at least so far). Numerical models can more easily model varying conditions, but often the idealisation in the computed model can limit the representativeness of the case studies. In most cases, the numerical models do not include three-dimensional effects. Using observations from *in situ*, enhanced by machine learning techniques, presents a third option.

Two 'ensemble' machine learning methods, the Random Forest and the Gradient Boosted Regression Tree, and one neural network method, the MultiLayer Perceptron method, were tested for the prediction of the spatial distribution of landslides in the Kvam region in Norway. Eleven landslide controlling factors were considered: slope angle, aspect, plan curvature, profile curvature, flow accumulation, flow direction, distance to rivers, water content, saturation, rainfall and distance to roads.

The performance of the three ML models was evaluated quantitatively based on the Receiver Operating Characteristic analysis. The results show that machine learning techniques can be applied to a large, heterogeneous dataset containing the characteristics of diverse landslide typologies to successfully predict the spatial extent of shallow landslides.

For the three methods investigated, the random selection of training and validation points apparently produced a training dataset that could be very easily mapped onto the validation data. The three ML landslide models, using the same training dataset, showed a satisfactory performance. The results show that the 'ensemble' GBRT machine learning

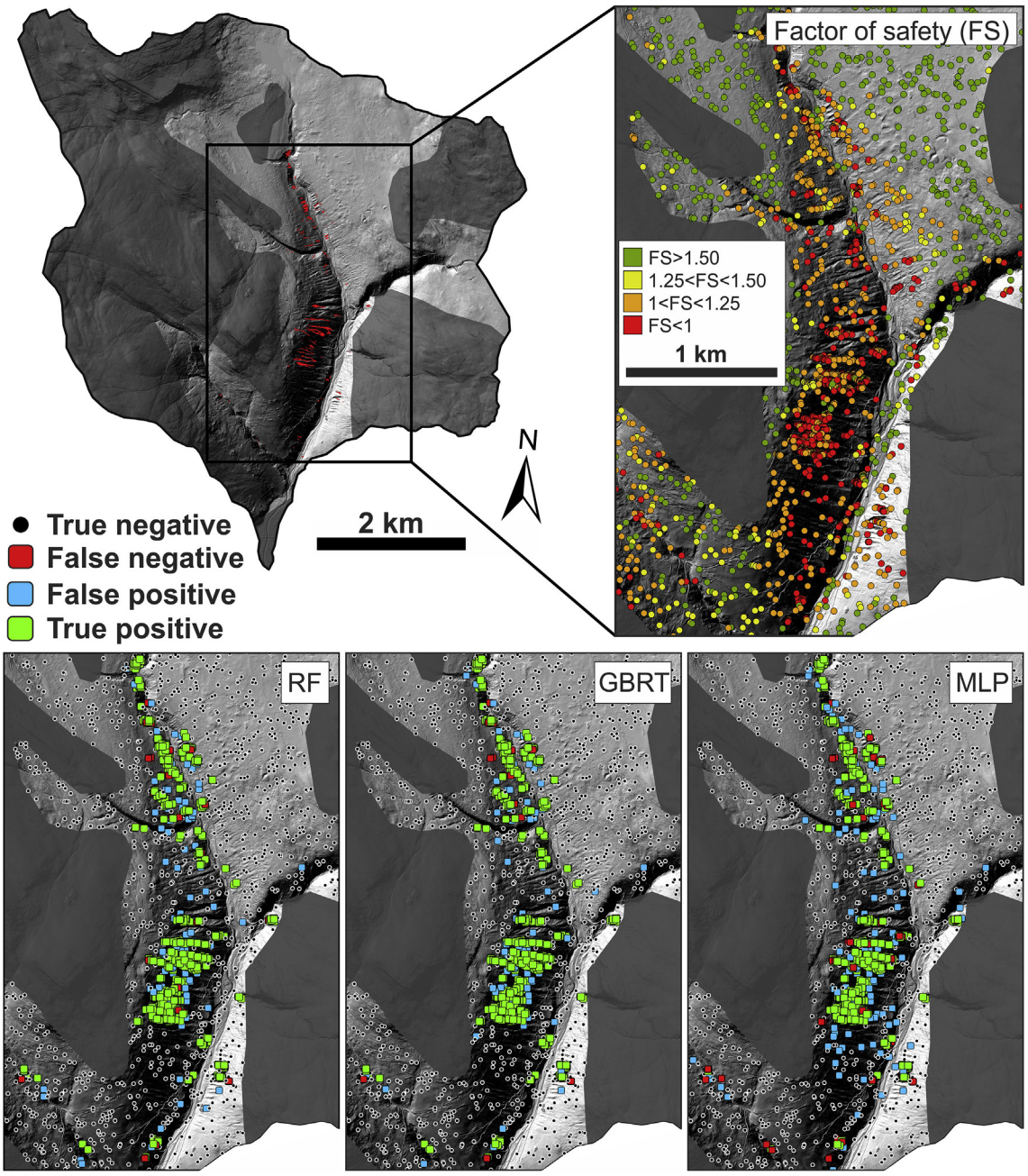

Fig. 4. Predictions of landslide occurrence from the RF, GBRT and MLP models: black circles (True negative) represent correct predictions of non-landslide locations, red rectangles (False negative) represent wrong predictions of actual landslide locations, blue rectangles (False positive) represent wrong predictions of non-landslide locations, and green rectangles (True positive) represent correct predictions of actual landslide locations. The distribution of safety factors in the upper right panel was computed by TRIGRS.

model yielded the most promising results for the spatial prediction of shallow landslides, with a 95% probability of landslide detection and 87% prediction efficiency compared to the actual landslides recorded in the Kvam area during a rainy period in 2011.

The selection of the controlling conditioning and triggering factors is a key requirement for successfully predicting landslide occurrence. In general, topography, geology, hydrology, geomorphology, meteorological events and anthropogenic effects are the most important controlling factors to include in the modelling of landslides. A significance indicator was calculated for the 11 controlling factors selected. The slope angle proved to be the most significant conditioning factor and the three triggering factors, i.e. saturation, rainfall and soil water content, were also important factors.

However, the results are convincing enough to conclude that the three ML algorithms, especially the Gradient Boosted Regression Tree and Random Forest algorithms, are suitable to predict the spatial distribution of landslides over a large area.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge NGI's financial support for this study. The funding comes in from The Research Council of Norway. The work did not receive any grant from other public funding agencies nor from the commercial or not-for-profit sectors.

References

Accadia, C., Mariani, S., Casaioli, M., Lavagnini, A., Speranza, A., 2003. Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbour average method on high-resolution verification grids. Weather Forecast. 18, 918–932.

Alkhasawneh, M.S., Ngah, U.K., Tay, L.T., Isa, N.A.M., 2014. Determination of importance for comprehensive topographic factors on landslide hazard mapping using artificial neural network. Environ. Earth Sci. 72 (3), 787–799.

Arnone, E., Francipane, A., Noto, L.V., Scarbaci, A., La Loggia, G., 2014. Strategies investigation in using artificial neural network for landslide susceptibility mapping: application to a Sicilian catchment. J. Hydroinf. 16 (2), 502–515.

Baum, R.L., Savage, W.Z., Godt, J.W., 2008. In: TRIGRS — a Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis, Version 2.0. U.S. Geological Survey. Open-File Report 2008-1159, p. 75.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Calvello, M., Piciullo, L., 2016. Assessing the performance of regional landslide early warning models: the EDuMaP method. Nat. Hazards Earth Syst. Sci. 16, 103–122.

Cepeda, J., Chávez, J.A., Martínez, C.C., 2010. Procedure for the selection of runout model parameters from landslide back-analyses: application to the Metropolitan Area of San Salvador, El Salvador. Landslides 7 (2), 105–116.

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Tien Bui, D., Duan, Z., Li, S., Zhu, A.X., 2018. GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. Catena 164, 135–149.

Corominas, J., Moya, J., Ledesma, A., Lloret, A., Gili, J.A., 2005. Prediction of ground displacements and velocities from groundwater level changes at the Vallcebre landslide (Eastern Pyrenees, Spain). Landslides 2 (2), 83–96.

De'Ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88 (1), 243–251.

Dickson, M.E., Perry, G.L.W., 2016. Identifying the controls on coastal cliff landslides using machine-learning approaches. Environ. Model. Software 76, 117–127.

Dou, J., Yunus, A.P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.-W., Khosravi, K., Yang, Y., Pham, B.T., 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci. Total Environ. 662, 332–346.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.

Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, London, UK.

Hong, H., Pourghasemi, H.R., Pourtaghi, Z.S., 2016. Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. Geomorphology 259, 105–118.

Hosmer, D.W., Lemeshow, S., 1989. Applied Logistic Regression. Wiley, New York.

Hsu, C., Chang, C., Lin, C., 2010. A practical guide to support vector classification. National Taiwan University. Technical Report.

Jaedicke, C., Solheim, A., Blikra, L.H., Stalsberg, K., Sorteberg, A., Aaheim, A., Kronholm, K., Vikhamar-Schuler, D., Isaksen, K., Sletten, K., Kristensen, K., Barstad, I., Melchiorre, C., Høydal, Ø.A., Mestl, H., 2008. Spatial and temporal variations of Norwegian geohazards in a changing climate, the GeoExtreme Project. Nat. Hazards Earth Syst. Sci. 8, 893–904.

Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. Photogramm. Eng. Rem. Sens. 54 (11), 1593–1600.

Kumar, D., Thakur, M., Dubey, C.S., Shukla, D.P., 2017. Landslide susceptibility mapping & prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India. Geomorphology 295, 115–125.

Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Model. 120 (2), 65–73.

Liao, Z., Hong, Y., Kirschbaum, D., Liu, C., 2011. Assessment of shallow landslides from Hurricane Mitch in central America using a physically based model. Environ. Earth Sci 66, 1697–1705.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., Kanevski, M., 2014. Machine learning feature selection methods for landslide susceptibility mapping. Math. Geosci. 46 (1), 33–57.

Montgomery, D.R., Dietrich, W.E., 1994. A physically based model for the topographic control on shallow landsliding. Water Resour. Res. 30, 1153–1171.

Montrasio, L., Valentino, R., 2007. Experimental analysis and modelling of shallow landslides. Landslides 4, 291–296.

Olden, J.D., Jackson, D.A., 2002. Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling 154 (1), 135–150.

Pack, R.T., Tarboton, D.G., Goodwin, C.G., 1999. Sinmap - a stability index approach to terrain stability hazard mapping. Users Manual [online]. Available at https://digitalcommons.usu.edu/cee_facpub/16/. (Accessed 10 October 2019).

Park, S., Kim, J., 2019. Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. Appl. Sci. 9, 942.

Pham, B.T., Prakash, I., Tien Bui, D., 2018. Spatial prediction of landslides using a hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. Geomorphology 303, 256–270.

Piciullo, L., Gariano, S.L., Melillo, M., Brunetti, M.T., Peruccacci, S., Guzzetti, F., Calvello, M., 2017. Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. Landslides 14, 995–1008.

Pourghasemi, H.R., Kerle, N., 2016. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. Environ. Earth Sci 75 (3), 1–17.

Pourghasemi, H.R., Jirandeh, A.G., Pradhan, B., Xu, C., Gokceoglu, C., 2013. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. Journal of Earth System Science 122 (2), 349–369.

Pourghasemia, H.R., Rahmatib, O., 2018. Prediction of the landslide susceptibility: which algorithm, which precision? Catena 162, 177–192.

Quinlan, J.R., 1993. C4.5 Programs for Machine Learning. Morgan Kaufmann Publishers.

Rahmati, O., Naghibi, S.A., Shahabi, H., Tien Bui, D., Pradhan, B., Azareh, A., Melesse, A.M., 2018. Groundwater spring potential modelling: comprising the capability and robustness of three different modeling approaches. J. Hydrol 565, 248–261.

Schilirò, L., Montrasio, L., Scarascia Mugnozza, G., 2016. Prediction of shallow landslide occurrence: validation of a physically-based approach through a real case study. Sci. Total Environ. 569–570, 134–144.

Siedlecka, A., Nystuen, J.P., Englund, J.O., Hossack, J., 1987. Lillehammer-berggrunnskart, 1:250.000. Norges Geologiske Undersøkelse: Trondheim, Norway (in Norwegian).

Sletten, K., Blikra, L.H., 2007. Holocene colluvial (debris-flow and water-flow) processes in eastern Norway: stratigraphy, chronology and palaeoenvironmental implications. J. Quat. Sci. 22, 619–635.

Staley, D.M., Kean, J.W., Cannon, S.H., Schmidt, K.M., Laber, J.L., 2013. Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. Landslides 10, 547–562.

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13, 361–378.

Tiranti, D., Rabuffetti, D., 2019. Estimation of rainfall thresholds triggering shallow landslides for an operational warning system implementation. Landslides 7, 471–481.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, p. 467.

Wu, X., Ren, F., Niu, R., 2014. Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. Environ. Earth Sci. 71 (11), 4725–4738.

Yang, B.B., Yin, K.L., Lacasse, S., Liu, Z.Q., 2019. Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides 16, 677–694.

Zare, M., Pourghasemi, H.R., Vafakhah, M., Pradhan, B., 2013. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arabian Journal of Geosciences 6 (8), 2873–2888.

Zhang, W., Wu, C., Li, Y., Wang, L., Samui, P., July 2020. Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. Georisk 11 (8), 2211–2225. https://doi.org/10.1080/17499518.2019.1674340.

Zhang, W., Zhang, R., Wu, C., Goh, A., Lacasse, S., Liu, Z., Liu, H., July 2020. State-of-the-art review of soft computing applications in underground excavations. Geos. Front. 11 (4), 1095–1106. https://doi.org/10.1016/j.gsf.2019.12.003.

Zhou, C., Yin, K., Cao, Y., Intrieri, E., Ahmed, B., Catani, F., 2018. Displacement prediction of step-like landslide by applying a novel kernel extreme learning machine method. Landslides 15, 2211–2225.