



OPEN

Mass wasting susceptibility assessment of snow avalanches using machine learning models

Bahram Choubin¹, Moslem Borji², Farzaneh Sajedi Hosseini², Amirhosein Mosavi^{3,4}✉ & Adrienn A. Dineva^{5,6}✉

Snow avalanche is among the most harmful natural hazards with major socioeconomic and environmental destruction in the cold and mountainous regions. The devastating propagation and accumulation of the snow avalanche debris and mass wasting of surface rocks and vegetation particles threaten human life, transportation networks, built environments, ecosystems, and water resources. Susceptibility assessment of snow avalanche hazardous areas is of utmost importance for mitigation and development of land-use policies. This research evaluates the performance of the well-known machine learning methods, i.e., generalized additive model (GAM), multivariate adaptive regression spline (MARS), boosted regression trees (BRT), and support vector machine (SVM), in modeling the mass wasting hazard induced by snow avalanches. The key features are identified by the recursive feature elimination (RFE) method and used for the model calibration. The results indicated a good performance of the modeling process (Accuracy > 0.88, Kappa > 0.76, Precision > 0.84, Recall > 0.86, and AUC > 0.89), which the SVM model highlighted superior performance than others. Sensitivity analysis demonstrated that the topographic position index (TPI) and distance to stream (DTS) were the most important variables which had more contribution in producing the susceptibility maps.

Snow avalanche is among the most destructive natural hazards in the cold and mountainous regions with devastating socioeconomic and environmental impacts^{1–4}. The mobility, transportation, tourism, and the leisure industries of the snowy mountain regions are under the avalanche's uncertain threats. Damaging infrastructures, roads and railways obstruction, threatening human life and the built environments and settlements, and harming the water resources, ecosystems, and vegetations are associated with the propagation and deposition of snow avalanche debris^{5–11}. The devastating propagation of the snow avalanches may also contribute to the mass wasting of surface rocks and vegetation particles transported along the way and accumulated together with the snow avalanche debris¹². The mass wasting induced by snow avalanche and the deposition of such snow-rock-debris poses longer-lasting damages with more destructive effects^{4,13,14}. The global snow avalanche regime is increasingly reported to be changing within the past decade^{15–18}. The climate change is introduced as a significant contributor in raising the occurrence rate and irregularity and increasing the risk and devastation^{19–22}. Therefore, more than ever, the accurate spatial hazard modeling and susceptibility mapping of the avalanche slopes and hazardous locations are seen crucial for risk management, planning efficient mitigation and adaptation practices, and territorial land-use policies.

Modeling the avalanche triggering mechanisms is complicated^{23–26}. The complexity of avalanche models has been discussed in many studies^{27–31}. The snowpack, meteorology, terrain, and slope characteristics are the predominant contributing factors initiating the avalanche movement and propagation, and the debris deposition^{2,32,33}. Based on the interaction of these factors, the motion and run out of snow and eventually, the avalanche formation and propagation can be modeled^{2,32,34–36}.

Various numerical methods for modeling the avalanche flow dynamics^{37–41}, as well as statistical approaches for processing historical database information and climatological data sets^{15,42–45}, have been proposed to predict the hazard susceptibility mapping. The models have been enhanced with the involvement of recent advanced

¹Soil Conservation and Watershed Management Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, Iran. ²Reclamation of Arid and Mountainous Regions Department, Faculty of Natural Resources, University of Tehran, Karaj, Iran. ³Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam. ⁴Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam. ⁵Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam. ⁶Kalman Kando Faculty of Electrical Engineering, Obuda University, Budapest, Hungary. ✉email: amirhosein.mosavi@tdtu.edu.vn; adrienn@duytan.edu.vn

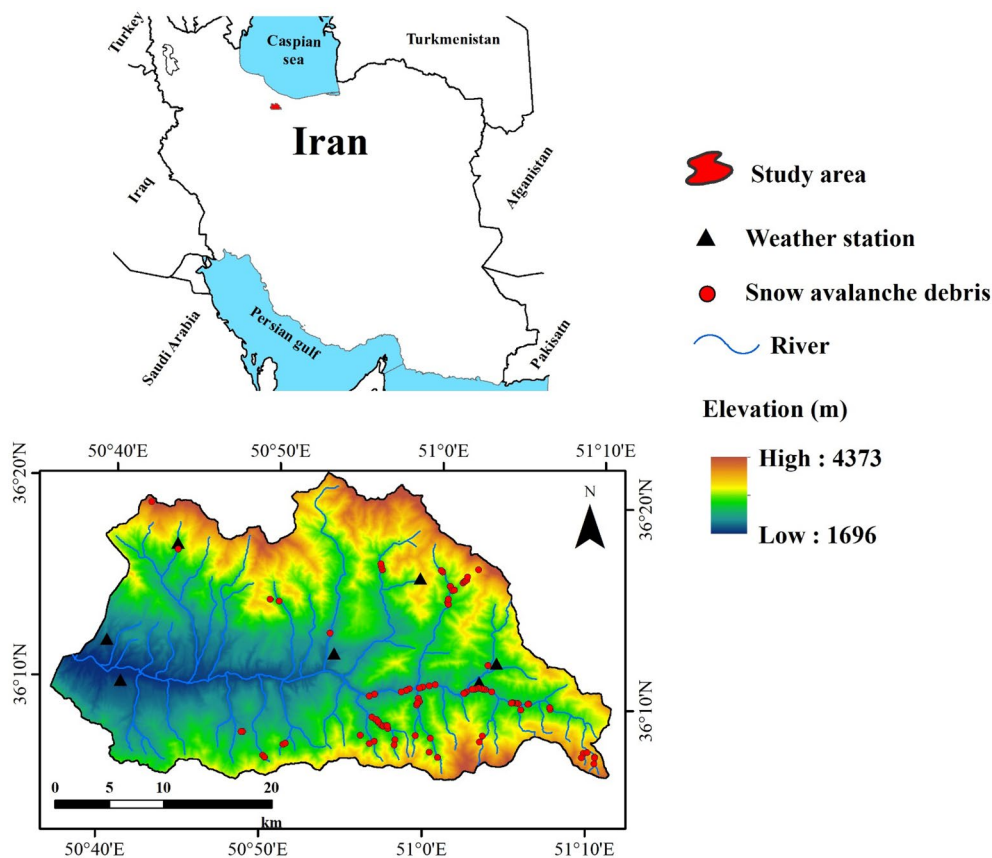


Figure 1. Location of the study area. The red circles denote locations of occurred snow avalanche debris. The black triangular indicate weather stations. The map was generated using ArcGIS Desktop 10.3. <https://desktop.arcgis.com/en>.

technologies of geographic information systems (GIS), remote sensing (RS), Satellite image processing, and artificial intelligence (AI) methods and applications^{2,29,46–49}. GIS is a powerful tool for the construction of terrain's geographical database to support building accurate prediction and decision-making models with high precision for terrain visualization^{50–54}. Also, RS contributes to data collection through remotely sensing the inaccessible mountainous areas, which are a real asset in replacing costly and slow ground data collection systems without disturbing the snow cover^{2,29,55,56}.

Machine learning (ML) has recently delivered fascinating results in advancing accurate models for susceptibility mapping of geohazards, e.g., earthquakes, landslides, earth fissures, and rockfalls^{57–62}. However, the application of machine learning methods in avalanche prediction has been minimal^{63,64}. Moreover, limited attention has been devoted to the hazard susceptibility of the mass wasting induced by snow avalanches which are accumulated as avalanche debris. As a response to these research gaps, this research's contribution is to model the hazard susceptibility of mass wasting by the snow avalanches in the mountainous region of Alborz Province, Iran, using well-known techniques that have not been used in this field so far.

This manuscript is structured in four sections. Section two presents the details of the study area, data, and the ML and the feature selection methods used for snow avalanche debris modeling. The results and discussion are described in section three. Section four presents the conclusions and discusses the ML methods' future direction for prediction of the snow avalanche mass wasting.

Data and methods

Description of the study area. Taleghan watershed is located in the Alborz Province, Iran. It is extended from latitudes 36° 05' to 36° 20' N and longitudes 50° 36' to 51° 11' E with an area about 947 km² (Fig. 1). The average annual temperature is about 11.4 °C and the average annual precipitation is about 520 mm. According to Hosseini⁶⁵, most of the precipitation during the cold period (i.e., December to March) is as snow, which its ratio to annual precipitation varies between 33 and 51% during different years. The geology includes about 50 lithological strata, which the oldest is Precambrian, and the recent alluvial deposits are from the Pleistocene. This watershed's lithology is mainly marly formations. The main soil orders of this region are Entisol and Inceptisol. Previous studies have shown sedimentation rates to be approximately 871 m³/km²/y; the preponderance of marly deposits accounts for a high erosion rate⁶⁶. Considering how important the Taleghan dam is for metropolitan Tehran's drinking water supply, the maintenance of the dam's reservoir and reduction of sediment yield are crucial. Based on the long-term monitoring of river sediment, the annual sediment yield of the Taleghan water-

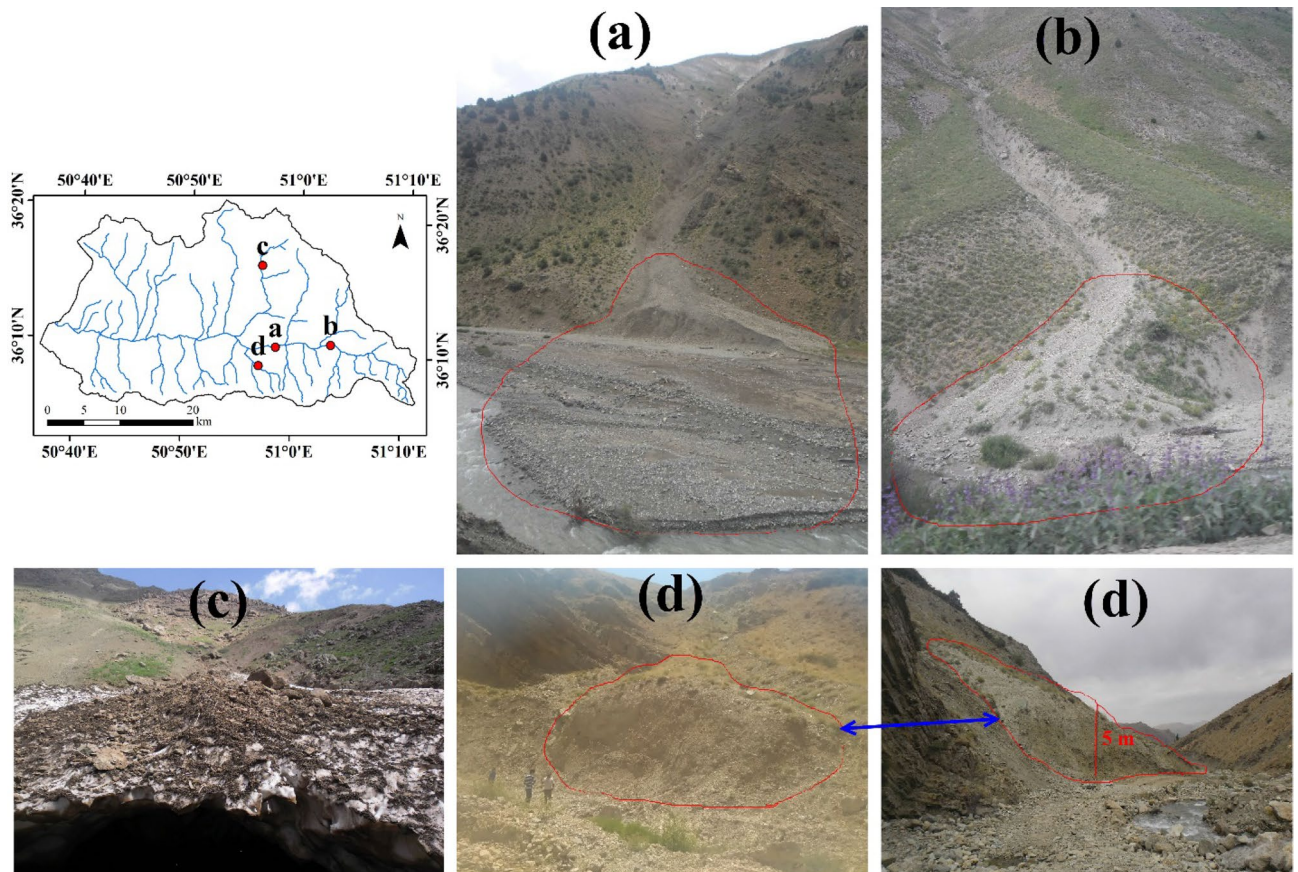


Figure 2. Some samples of snow avalanche debris which transferred the sediment into rivers in the region. The photographs were taken by Moslem Borji in 2019. The map was generated using ArcGIS Desktop 10.3, <https://desktop.arcgis.com/en>.

shed is about 10 tones/ha/y⁶⁷. The Taleghan watershed is the upstream of two dams: the Taleghan (completed in 2006) and the Sefid-rood dam (completed in 1963), which both of them suffer from high sedimentation. One of the main sources of the watershed's sediments is related to the snow avalanches debris, which deposits large volumes of sediment near the streams (Fig. 2). These deposits are exposed to being washed by the high flash floods. Therefore, it is important to know the hazardous snow avalanche debris areas to implement sediment-yield prevention initiatives.

Inventory map of snow avalanche debris. Each snow avalanche consists of the starting zone, track, and runout zone⁶³. After occurring a snow avalanche, the snow avalanche debris remains inside the snow runout (deposit) zone. As ice and snow melt, these materials become visible (e.g., Fig. 2c). Most snow avalanches' debris is very angular, enabling them to be differentiated from glacial, other drift-ice origins or rockfalls⁶⁸. In this research, by considering the abovementioned conditions, a number of 98 locations of occurred snow avalanche debris were recorded based on the repeated field surveys (Fig. 1). In first, the probable locations of snow avalanche debris were selected based on the topographic maps and then based on the Google Earth. Then, through field surveys, the location of the avalanches containing sedimentary deposits was recorded. By observing the material accumulated in the snow at the end of the snow avalanche passages in the spring thaw period, the cause of this debris (i.e., snow avalanche) has been confirmed. Some samples of the recorded snow avalanche debris are shown in Fig. 2.

Predictors used for snow avalanche debris modeling. Snow avalanche debris can apply the considerable erosive forces on soils and may scrape away and entrain soil into the deposition zone. The chemical composition of the soil material entrained by the snow avalanche debris is changed and potentially contributing to form the landforms⁶⁹. However, many factors affect the quantity of snow avalanche deposits and creating this natural hazard. These factors are including morphological features, climate, soil properties, vegetation cover, etc. In this study, for snow avalanche debris modeling, tried to consider the most related factors to the occurrence of snow avalanche debris:

Elevation: Snow accumulation and melt are related to elevation (Fig. 3a), but the relationship between elevation and snow avalanche debris is more complicated than what is perceived in the watershed⁷⁰. One of the

most important factors that cause more deposition at the snow avalanche debris's runout zone is the pressure of snow. Snow pressure destroys rocky outcrops, and top layers of soil located on the crossing avalanches pass. Although elevation is not an inherent property of snow pressure, it affects the relative density of snow. The elevation factor contributes to increasing the relative density of snow, which increases the creep of snow fragments. Between the elevation of 1500 to 3000 m, for every 100 m of elevation, a two percent rise in snow pressure is calculated⁷¹. Thus, there is a direct relationship between the snow avalanche's elevation and the sediment yield⁷¹.

Slope: High slopes (more than 50° or 120%) rarely contribute to snow avalanche debris, because they discharge continually during each new snowfall (Fig. 3b). Also, low slopes (less than 30° or 58%) do not have the required potential (force) to create avalanches. However, slopes 30°–50° are more likely to the occurrence of avalanches. Due to the numerous factors that affect the snow avalanche debris, even with an equal gradient, the likelihood of avalanches in different places will not be the same⁷¹.

Aspect: Geographic aspect (Fig. 3c) is of primary significance to avalanches. Different aspects have different influences on accumulating and melting of the snow because of the meteorological parameters, such as temperature, precipitation, wind velocity, and sunshine hours^{63,64,72}.

Curvature: Convex areas produce an increase in the rate of total motion of the snow cover downhill (a combination of creep and slide). In a suitable condition for avalanche releasing (presence of a stratum with low shear strength), the fracture will initially occur in these convex areas where there is tension force within the snow cover. So, the curvature map (Fig. 3d) can be an important layer in snow avalanche debris modeling.

Topographic Position Index (TPI): The general concept and application of the TPI is to define and determine the boundary of landforms in an accurate and non-descriptive manner such as heights, steep slopes, flat areas, valleys, etc. using a DEM automatically and rapidly⁷³. The TPI reflects the notion that the positive values of TPI represent areas that are higher than the surrounding points (hills), and the negative values represent regions with lower positions around them (valleys) (Fig. 3e). The values near zero indicate flat areas (where the gradient is close to zero) or areas with a steady slope⁷³. This layer indicates valleys, streams, open slopes, mid-alluvial valleys, and u form valleys, which affect the occurrence of snow avalanche debris.

Vector Ruggedness Measure (VRM): VRM (Fig. 3f) presents the roughness of the ground which, affects the movement of snow avalanche debris. For example, rocks and rock outcrops (i.e., irregular surfaces) may promote instability or stability of the snow avalanches⁶³.

Distance to Fault (DTF): The most important step in initiating the movement of thick snow fragments from the starting area is the formation of a weak layer fracture⁷⁴. Many natural and anthropogenic factors such as load variation due to skiers, explosion, aircraft noise, and distance from active faults can cause these weak layer fractures (Fig. 3g).

Distance to Road (DTR): This factor causes weak layers in the thick snow and increases the amount of debris deposition at the runout zone by moving vehicles on the road and generating heat and impact on local temperatures, as well as by the sound of beeps and vibrations (Fig. 3h).

Distance to Stream (DTS): The snow avalanche deposits in the study area are directly related to the distance from the stream (Fig. 3i). Sediments accumulated along the river are transported by water flow and as the toe of the avalanche is emptied. So, the slope of the passage increased, then snow avalanche debris is increased.

Drainage Density (DD): DD specifies the density of streams per unit ground, which is essential for occurring the snow avalanche debris occurrence⁷⁵. Increasing the drainage density will create snow passage channels, and avalanche falls (Fig. 3j).

Precipitation: Increasing the amount of precipitation increases the causes of the snow thickness in the area (Fig. 3k). Considering further essential factors to the avalanche formation, it is evidenced that snow-crossing discharge is increased by increasing precipitation. This would lead to increasing the debris.

Stream Power Index (SPI): The SPI (Fig. 3l) indicates the erosive power of running water, directly affecting the slope toe erosion and stream incision⁷⁶. So, it is the main factor that controls the slope erosion processes and snow avalanches.

Topographic Wetness Index (TWI): The TWI (Fig. 3m), is a steady wetness index, which is commonly used to quantify topographic control on hydrological processes⁷⁵. TWI is one of the most important factors that indicate the potential of runoff generation. In other words, the high values of TWI mean the high potential of runoff generation and vice versa. Therefore, in areas with high runoff, sediment deposits are transported downstream to different parts of the avalanche pass.

Land use: Land use has a complex effect on the occurrence of avalanches and resulting deposits. For example, if the height of grass and bushes are higher than snow in rangelands, then the rangeland has a decreasing role in avalanche fall⁷¹. For extracting the land use map in this study, satellite data from the Sentinel-2 mission was chosen for two main reasons: (1) its relatively high 10 m spatial resolution, and (2) its radiometry includes three vegetation red edge bands. These two characteristics make the Sentinel-2 data appealing for land use extraction. According to the land use map, the rangelands include the region's highest area (Fig. 3n).

Lithology: Lithology and rock units have an important role in the slope failure in mass movements such as avalanches, landslides, and debris flows⁷⁷. Rock unit type is important in heat absorption and transfers to snow cover. Dark igneous rocks have the highest thermal absorption. These rocks can provide a suitable base for fracturing snow masses by absorbing heat and melting snow in the lower layers and creating pores in the snow masses. Conversely, light rocks such as lime and sandstone and igneous rocks absorb less heat and therefore control the movement of snow masses. Most rock units in areas with snow avalanche risk include andesite trace, basalt trace, basanite, andesite, agglomerate, tuff, and pyroclastic (Fig. 3o).

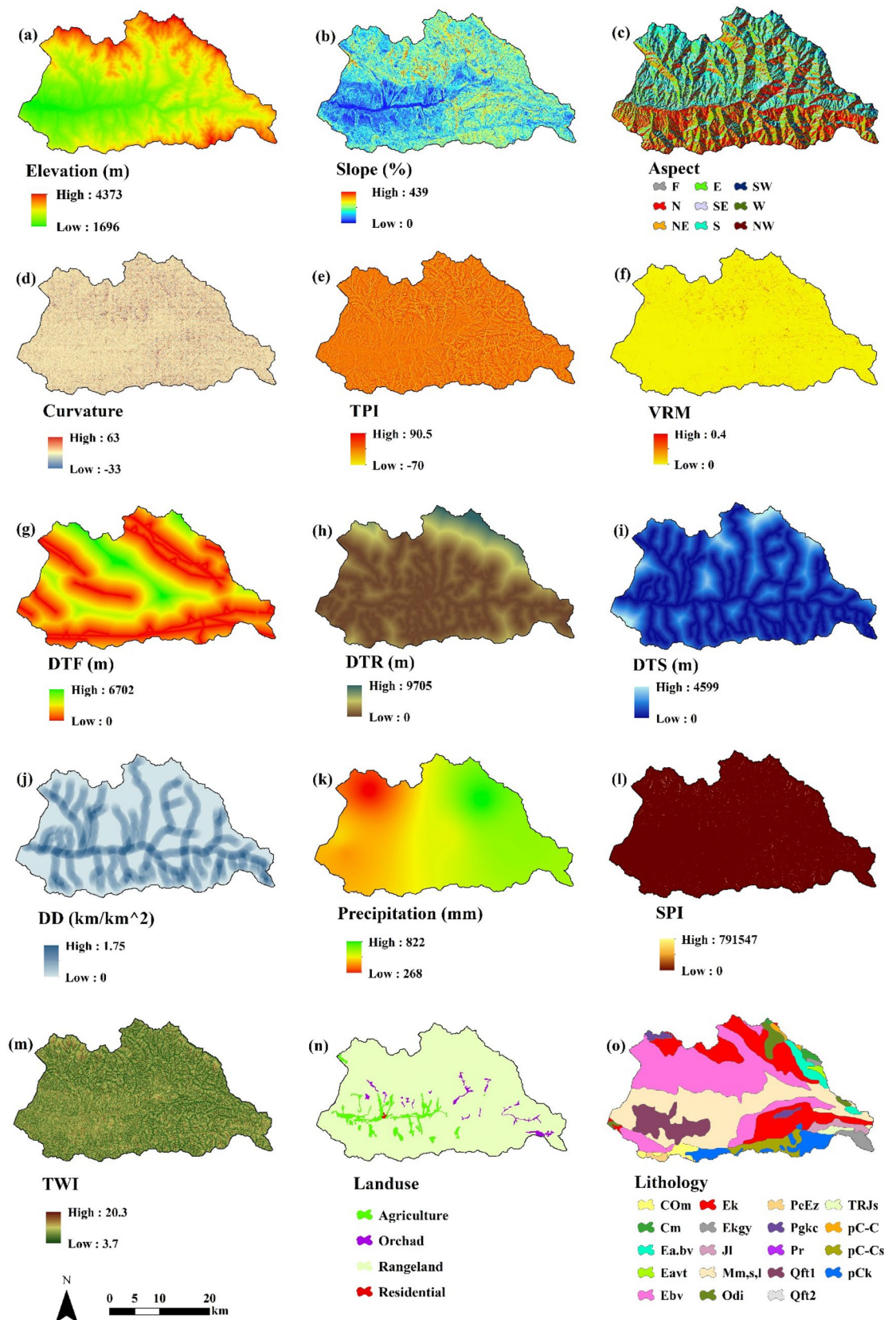


Figure 3. Snow avalanche debris conditioning factors: (a) elevation, (b) slope, (c) aspect, (d) curvature, (e) topographic position index (TPI), (f) vector ruggedness measure (VRM), (g) distance to fault (DTF), (h) distance to road (DTR), (i) distance to stream (DTS), (j) drainage density (DD), (k) precipitation, (l) stream power index (SPI), (m) topographic wetness index (TWI), (n) landuse, and (o) lithology. The maps were generated using ArcGIS Desktop 10.3, <https://desktop.arcgis.com/en>.

In this study, the ALOS PALSAR Digital Elevation Model (DEM) (<https://vertex.daac.asf.alaska.edu/>) with a pixel size of 12.5 m × 12.5 m was used to map the DEM derivatives (i.e., slope, aspect, curvature, TPI, VRM, SPI, and TWI). The spatial resolution of other input factors was resampled into an equal resolution of the DEM derivatives (i.e., 12.5 m × 12.5 m) using the Resample Tool in the ArcGIS environment. Also, the number of pixels and the extend of layers were controlled to be same for the spatial modeling.

Multicollinearity analysis. Multicollinearity expresses the linear relationship between two or more independent variables. It is a data problem that may cause serious difficulty with the reliability of the estimates during the modeling process⁷⁸. In this study, the Variance Inflation Factor (VIF) was used to assess the multicollinearity among the explanatory variables. Since the VIF values are low (VIF < 10) (Supplementary Table S1), there is not any multicollinearity among the predictor variables^{59,79}.

Feature selection using the recursive feature elimination (RFE). The RFE procedure is generally a recursive process that ranks features based on the value of their importance. The method tends to eliminate weak features that are not important when synthesized with other features, and therefore, leads to a considerable decrease in the redundant variables⁸⁰. The RFE method repeated many times and, in each time, the important features remained and the less important features are eliminated. The recursion is required because the relative significance of each feature can be essentially changed when assessing with a different subset of features in step-wise exclusion^{81,82}. In this study, the RFE method was implemented using the k-fold (tenfold) cross-validation methodology by the Caret R package⁸³.

Susceptibility modeling of snow avalanche debris. Four machine learning (ML) models, including generalized additive model (GAM), multivariate adaptive regression spline (MARS), boosted regression trees (BRT), and support vector machine (SVM), were used to model the snow avalanche debris. Dependent data were the location of the snow avalanche debris (Fig. 1) and the independent data were the effective environmental factors identified by the RFE method. Locations of 98 snow avalanche debris, as well as the equal number of non-occurrence points, are converted to a binary scale (respectively 1 and 0) and considered as dependent data. The k-fold cross-validation methodology (k = 10) was used to model calibration in the R software environment using the 'sdm' package⁸⁴. A ratio of 70% to 30% was considered respectively for calibrating and testing phases. Description of the applied ML models as follows:

The GAM model, developed by Hastie and Tibshirani⁸⁵, blends the properties of the generalized linear model (GLM) with additive models. Indeed, this model is a statistical model in which the linear relationships between the dependent variable and independent variables are replaced by non-linear smooths⁸⁶. On the contrary of the GLM, which the researcher needs to have proper knowledge of the correct functional form before the modeling, the GAM uses the additive approach in which the suitable functional form is selected based on the data⁸⁷. To spatial modeling with the GAM model, as well as the 'sdm' package, the 'mgcv'⁸⁸ and 'gam'⁸⁹ R packages are required.

The MARS model, introduced by Friedman⁹⁰, is a non-parametric statistical method which acts based on the divide and conquers strategy. This means that the calibration datasets are partitioned into separate splines (piecewise linear segments) of differing slopes (gradient)⁹¹. Indeed, it is known as an extension of linear models which automatically explores and models nonlinearities between the dependent and independent variables, without considering any assumptions about their relationships⁹². In this study, the 'earth'⁹³ package, as well as the 'sdm' package, was used to spatial modeling of snow avalanche debris by the MARS model.

The BRT model is a combination of boosting and decision tree methods. It frequently fits many decision trees to increase the model's accuracy, like the random forest (RF) model^{94,95}. The model takes the random subsets, from the complete dataset with the same number of data points, for new trees that are built. According to the boosting method used in this model, the model continuously tries to improve its accuracy in each new tree which is built^{96,97}. In this study, in addition to the 'sdm' package, the 'gbm'⁹⁸ package is required for spatial modeling using the BRT model.

The SVM model, introduced by Cortes and Vapnik⁹⁹, is a non-parametric statistical monitoring method, which, to solve problems it uses the structural-risk minimization principle along with the dimension theory of Vapnik Chervonenk¹⁰⁰⁻¹⁰². The mathematical function (kernel) transforms the data (inputs) into the required form and maps into a high dimensional feature space. The model creates a line (or a hyperplane) that splits the data into classes of 0 (non-occurrence) and 1 (occurrence). The SVM model with the kernel of radial basis function (RBF) was implemented using the 'kernlab'⁸² R package, which is a dependent package in spatial modeling by the 'sdm' package.

Modeling evaluation. Modeling results were evaluated using the several metrics which are used for evaluation of the dichotomous forecasts. These metrics are including the area under the curve (AUC) of receiver operator characteristic (ROC) plot, Accuracy (Eq. 1¹⁰³), Kappa (Eq. 2¹⁰⁴), Precision (Eq. 4¹⁰⁵), and Recall (Eq. 5¹⁰⁵) metrics which are calculated by the contingency table¹⁰⁶:

$$\text{Accuracy} = \frac{H + CN}{H + CN + M + FN} \quad (1)$$

$$\text{Kappa} = \frac{\text{Acc} - P_e}{1 - P_e} \quad (2)$$

Variable	Accuracy	Kappa
1	0.867	0.737
2	0.928	0.856
3	0.942	0.883
4	0.956	0.912
5	0.949	0.898
6	0.956	0.912
7	0.956	0.912
8	0.963	0.926
9	0.956	0.912
10	0.942	0.883
11	0.953	0.916
12	0.956	0.912
13	0.956	0.912
14	0.949	0.898
15	0.956	0.912

Table 1. Resampling performance for different number of variables using the tenfold cross-validation method.

$$P_e = \frac{(H + FA)(H + M) + (M + CN)(FA + CN)}{(H + FA + M + CN)^2} \quad (3)$$

$$\text{Precision} = \frac{H}{H + FA} \quad (4)$$

$$\text{Recall} = \frac{H}{H + M} \quad (5)$$

where H, FA, M, and CN are respectively the number of hits, false alarms, misses, and correct negatives^{63,106}. P_e represents the expected agreement between the modeled and forecasted values¹⁰⁴. Variations of the evaluation metrics are between 0 and 1 which 1 indicates perfect prediction¹⁰⁷.

Results and discussion

Feature selection results. Results of the recursive feature elimination (RFE) method indicated that the contribution of eight features, among 16, in snow avalanche debris modeling has better results (Accuracy = 0.963; Kappa = 0.926), according to the resampling performance using tenfold cross-validation method (Table 1). Variations of the accuracy with the different numbers of the features during the model runs are presented in Supplementary Figure S1. As can be seen, the mean accuracy (red plus in Supplementary Figure S1) with eight features is higher than the other number of the features. Therefore, based on the occurrence frequency of the features in the model (RFE) runs through the resampling process with a tenfold cross-validation method, eight variables with a higher occurrence frequency than others are selected as key features (Supplementary Figure S2). A higher occurrence frequency of a variable indicates its higher importance because the RFE is a backward select method¹⁰⁸, which the lower important variables are removed from modeling and higher important variables are reminded¹⁰⁹. Accordingly, variables of TPI, DTS, SPI, lithology, precipitation, TWI, DTR, and VRM with the contribution of 100%, 93%, 86%, 80%, 72%, 68%, 55%, and 53% in the model runs are selected as key features and used to snow avalanche debris modeling (Supplementary Figure S2).

Unlike the results of Mosavi et al.⁶⁴ in snow avalanche susceptibility mapping, variables of slope and aspect in this study were eliminated during the feature selection. The reason is that, unlike the snow avalanche which consists of three zones⁶³, the avalanche debris is located at the end of the snow avalanche passage. So, they mostly have a low slope and no aspect (i.e., flat) which cannot produce sufficient variance to be selected as important variables in snow avalanche debris modeling.

Snow avalanche debris modeling. Model calibration and validation were conducted using the key features identified by the RFE method. The AUC of ROC plots is presented in Fig. 4. As can be seen, according to the mean AUC across tenfold cross-validation methodology, the BRT and SVM models have close AUC values (respectively 0.970 and 0.964) and higher than the MARS (AUC = 0.935) and GAM (AUC = 0.892) models (Fig. 4).

For further understanding of the models' performance, the Accuracy, Kappa, Precision, and Recall values for the testing dataset are calculated (Table 2). Results of the model evaluation indicated that the SVM model had a higher Accuracy, Kappa, Precision, and Recall values (respectively equal to 0.91, 0.83, 0.90, and 0.93) rather than other models (Table 2). Other models (i.e., BRT, MARS, and GAM) had the same performance given the Accuracy and Kappa values (respectively equal to 0.88 and 0.76), while the MARS model in view of Precision

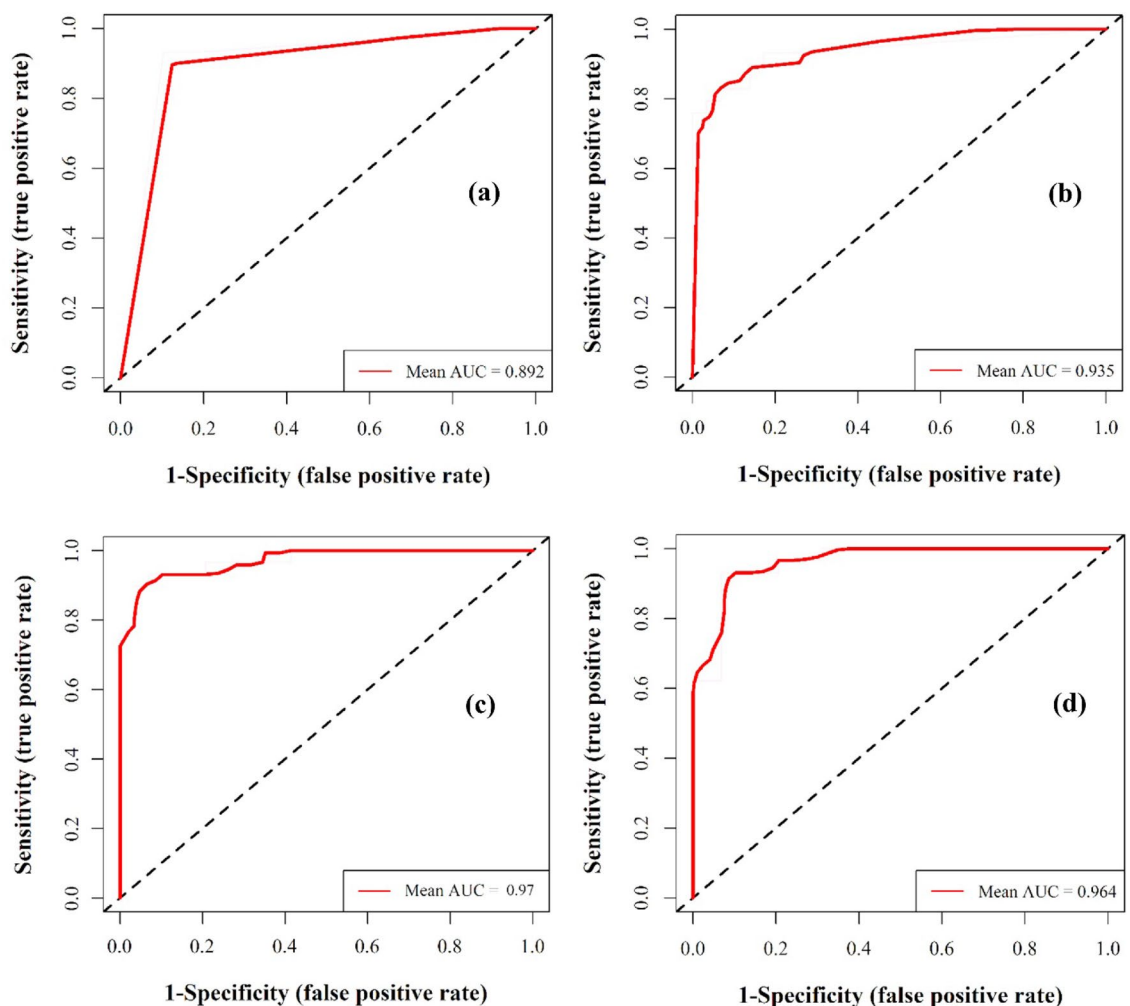


Figure 4. The area under curve (AUC) values: (a) GAM, (b) MARS, (c) BRT, and (d) SVM. Presented AUCs are mean AUC across tenfold cross-validation methodology.

Statistic	GAM	MARS	BRT	SVM
Accuracy	0.88	0.88	0.88	0.91
Kappa	0.76	0.76	0.76	0.83
Precision	0.87	0.89	0.84	0.90
Recall	0.90	0.86	0.93	0.93

Table 2. Snow avalanche debris modeling performance. A higher value of the statistics indicates a higher performance of the modeling.

(equal to 0.89) and the BRT model in view of Recall were the best ones, after the SVM model (Table 2). However, the SVM model has different characteristics that make it successful in classification studies such as (i) applying kernel functions to solve the problems of linear separating classes resulting from multiple independent input variables^{110,111}, (ii) maximizing the boundaries between classes for achieving an optimal separating hyperplane¹¹⁰, and (iii) excluding the outliers to avoid over-fitting of the model^{111,112}.

Spatial prediction of the snow avalanche debris susceptibility. After model calibration and evaluation, the spatial prediction of the snow avalanche debris susceptibility was done using the calibrated models and the pixels' value of the predictors for the whole region. A natural break (Jenks) classification method was used to classify the predicted values into three low, moderate, and high classes. This method is based on the inherent classification in data which distinguishes classes based on the minimum difference within groups and the maximum difference between groups. Also, in other related studies such as snow avalanche hazard mapping (e.g., Choubin et al.⁶³; Mosavi et al.⁶⁴) this method has been used.

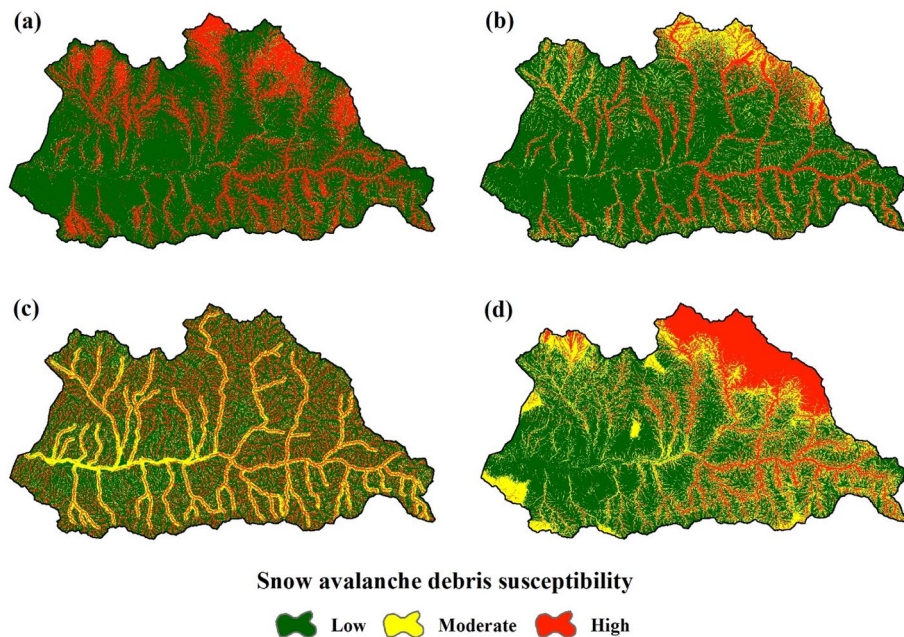


Figure 5. Snow avalanche debris susceptibility maps predicted by the GAM (a), MARS (b), BRT (c), and SVM (d). Low, moderate, and high susceptibility areas are shown in green, yellow, and red colors, respectively. The maps were generated using ArcGIS Desktop 10.3, <https://desktop.arcgis.com/en>.

Class	GAM	MARS	BRT	SVM
Low	16.5	25.5	35.7	12.1
Moderate	48.6	44.6	27.7	29.4
High	34.9	29.9	36.6	58.5

Table 3. The susceptibility classes area (%) predicted by the models.

The predicted susceptibility maps by the models with a pixel size of $12.5 \text{ m} \times 12.5 \text{ m}$ are shown in Fig. 5. The susceptibility classes area (%) predicted by the models are presented in Table 3. The GAM and MARS models have the most area of the region in the moderate class (respectively 48.6% and 44.6% of the study area), while the most area of the region in the BRT and SVM models are related to the high susceptibility class (respectively 36.6% and 58.5% of the study area).

The results of susceptibility maps are different in each model (Fig. 5). The main reason refers to the different structures of the models which result in different outputs¹⁰⁷. Also, this can be more clarified by the importance analysis of the variables according to the decrease in AUC (DAUC) values after excluding the parameters from the modeling process (Fig. 6). For the GAM model, variables of TPI (DAUC = 51.7%), DTS (DAUC = 30.5%), and SPI (DAUC = 29.9%) have more contribution to the modeling (Fig. 6) and the predicted map (Fig. 5a) is more match with them. In the MARS model, three main factors are respectively DTS (DAUC = 41.9%), TPI (DAUC = 30.7%), and SPI (DAUC = 12.8%). In this model, since the DTS is more important than the other variables, the susceptibility map (Fig. 5b) is more line with the DTS. Regarding the BRT model, the TPI (DAUC = 55.9%) is the main factor (Fig. 6). So, the high susceptible areas in this model (Fig. 5c) are matched with high values of TPI (i.e., heights) (Fig. 3). According to the SVM result, TPI and DTS are the most important variables with close importance (respectively DAUC equal to 17.1% and 18.1%) (Fig. 6). In this model, the high susceptibility areas are situated in mountainous regions of the northeast with high TPIs, that some glaciers of the Alam Kuh area are located in these areas. Also, the areas near to streams show a high susceptibility that this decreases from east to west (Fig. 5d). The reason for this can be because of decreasing the elevation and precipitation towards the west (Fig. 3).

Generally, the sensitivity analysis indicated that the most important variables were TPI and DTS with different importance in different models. Also, SPI was another variable that had a significant role in the GAM and MARS models (Fig. 6). Although limited attention has been devoted to the hazard susceptibility of the mass wasting induced by snow avalanche, in other nearby fields such as snow avalanche modeling, the importance of the TPI variable has been proved by Choubin et al.⁶³ and Mosavi et al.⁶⁴.

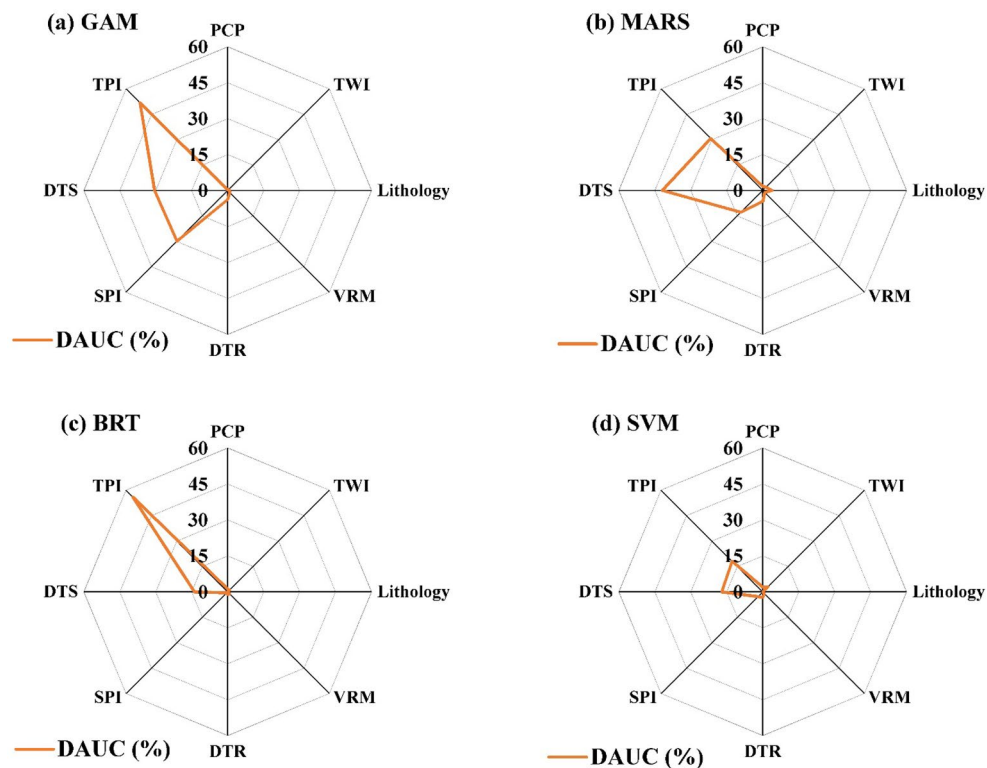


Figure 6. Importance of the variables based on the decrease in AUC (DAUC). A higher DAUC indicates a higher importance.

Conclusions

Mass wasting induced by snow avalanche is among the major natural hazards in the cold and mountainous regions. Current research tried to model this natural hazard, for the first time, by considering the related environmental factors. Key features in this study were TPI, DTS, SPI, lithology, precipitation, TWI, DTR, and VRM which were identified by the recursive feature elimination (RFE) method (with an Accuracy equal to 0.963 and Kappa equal to 0.926). Modeling results with key predictors produced good results (Accuracy > 0.88 and AUC > 0.89). However, model comparison highlighted a superior performance of the SVM model (i.e., Accuracy = 0.91, Kappa = 0.83, Precision = 0.90, and Recall = 0.93) rather than the BRT, MARS, and GAM models. Results from the sensitivity analysis indicated that TPI and DTS variables were the most important factors that had more contribution in producing the susceptibility maps. Lack of the soil (or near-surface) temperature data in the study area was the main limitation of the research. Near-surface temperature can affect the avalanche movement which stimulates depleting the snow from hillsides and bringing the rock materials by itself into the runout zones (down the hillsides). Although this data is not available in the watershed area, its existence can help future studies in modeling the snow avalanche debris on the hillside scale. However, the maps produced by this study can help the watershed managers and land use policymakers to identify and protect the vulnerable areas.

Received: 5 June 2020; Accepted: 15 October 2020

Published online: 27 October 2020

References

- Gusain, H. S., Mishra, V. D. & Singh, D. K. Study of a snow avalanche accident along Chowkibal–Tangdhar road in Kupwara district, Jammu and Kashmir, India. *Curr. Sci.* **115**, 969–972 (2018).
- Singh, D. K., Mishra, V. D., Gusain, H. S., Gupta, N. & Singh, A. K. Geo-spatial modeling for automated demarcation of snow avalanche hazard areas using Landsat-8 satellite images and in situ data. *J. Indian Soc. Remote Sens.* **47**, 513–526 (2019).
- Stethem, C. *et al.* Snow avalanche hazard in Canada—a review. *Nat. Hazards* **28**, 487–515 (2003).
- Wesselink, D. S., Malnes, E., Eckerstorfer, M. & Lindenberg, R. C. Automatic detection of snow avalanche debris in central Svalbard using C-band SAR data. *Polar Res.* **36**, 1333236 (2017).
- Bertrand, D., Naaim, M. & Brun, M. Physical vulnerability of reinforced concrete buildings impacted by snow avalanches. *Nat. Hazards Earth Syst. Sci.* **10**, 1531–1545 (2010).
- Casteller, A. *et al.* Assessing the interaction between mountain forests and snow avalanches at Nevados de Chillán, Chile and its implications for ecosystem-based disaster risk reduction. *Nat. Hazards Earth Syst. Sci.* **18**, 1173–1186 (2018).
- De Biagi, V., Chiaia, B. & Frigo, B. Impact of snow avalanche on buildings: forces estimation from structural back-analyses. *Eng. Struct.* **92**, 15–28 (2015).
- Feistl, T. *et al.* Forest damage and snow avalanche flow regime. *Nat. Hazards Earth Syst. Sci.* **15**, 1275–1288 (2015).

9. Meseşan, F., Gavrilă, I. G. & Pop, O. T. Calculating snow-avalanche return period from tree-ring data. *Nat. Hazards* **94**, 1081–1098 (2018).
10. Šilhán, K., Kluzová, O. & Tichavský, R. The on field differentiation of snow avalanche- and debris flow-induced scars in trees as a fundament for improving dendrogeomorphic sampling strategy (case study from the Great Cold Valley in High Tatra Mountains). *Cold Reg. Sci. Technol.* **158**, 1–9 (2019).
11. Strapazzon, G. & Brugger, H. On-site treatment of snow avalanche victims: from bench to mountainside. *High Alt. Med. Biol.* **19**, 307–315 (2018).
12. Bell, I., Gardner, J. & De Scally, F. An estimate of snow avalanche debris transport, Kaghan Valley, Himalaya, Pakistan. *Arct. Alp. Res.* **22**, 317–321. <https://doi.org/10.1080/00040851.1990.12002795> (1990).
13. Eckerstorfer, M. & Malnes, E. Manual detection of snow avalanche debris using high-resolution Radarsat-2 SAR images. *Cold Reg. Sci. Technol.* **120**, 205–218 (2015).
14. Vickers, H., Eckerstorfer, M., Malnes, E., Larsen, Y. & Hindberg, H. A method for automated snow avalanche debris detection through use of synthetic aperture radar (SAR) imaging. *Earth Sp. Sci.* **3**, 446–462 (2016).
15. Podolskiy, E. A., Izumi, K., Suchkov, V. E. & Eckert, N. Physical and societal statistics for a century of snow-avalanche hazards on Sakhalin and the Kuril Islands (1910–2010). *J. Glaciol.* **60**, 409–430 (2014).
16. Naaim, M. *et al.* Impact du réchauffement climatique sur l'activité avalancheuse et multiplication des avalanches humides dans les Alpes françaises. *Houille Blanche* <https://doi.org/10.1051/lhb/2016055> (2016).
17. Oleinikov, A. D. & Volodicheva, N. A. Recent trends of snow avalanche regime in the Central Caucasus (Elbrus region as an example). *Led i Sneg* **59**, 191–200 (2019).
18. Sinickas, A., Jamieson, B. & Maes, M. A. Snow avalanches in western Canada: investigating change in occurrence rates and implications for risk assessment and mitigation. *Struct. Infrastruct. Eng.* **12**, 490–498 (2016).
19. Ballesteros-Cánovas, J. A., Trappmann, D., Madrigal-González, J., Eckert, N. & Stoffel, M. Climate warming enhances snow avalanche risk in the Western Himalayas. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3410–3415 (2018).
20. Coe, J. A., Bessette-Kirton, E. K. & Geertsema, M. Increasing rock-avalanche size and mobility in Glacier Bay National Park and Preserve, Alaska detected from 1984 to 2016 Landsat imagery. *Landslides* **15**, 393–407 (2018).
21. Gądek, B. *et al.* Snow avalanche activity in Żleb Żandarmerii in a time of climate change (Tatra Mts., Poland). *CATENA* **158**, 201–212 (2017).
22. Laute, K. & Beylich, A. A. Potential effects of climate change on future snow avalanche activity in western Norway deduced from meteorological data. *Geogr. Ann. Ser. A Phys. Geogr.* **100**, 163–184 (2018).
23. Gaume, J., van Herwijnen, A., Gast, T., Teran, J. & Jiang, C. Investigating the release and flow of snow avalanches at the slope-scale using a unified model based on the material point method. *Cold Reg. Sci. Technol.* **168**, 102847 (2019).
24. Martin, J. P. & Germain, D. Dendrogeomorphic reconstruction of snow avalanche regime and triggering weather conditions: a classification tree model approach. *Prog. Phys. Geogr.* **40**, 527–548 (2016).
25. McClung, D. M. Effects of triggering mechanism on snow avalanche slope angles and slab depths from field data. *Nat. Hazards* **69**, 1721–1731 (2013).
26. Pérez-Guillén, C. *et al.* Evaluation of a snow avalanche possibly triggered by a local earthquake at Vallée de la Sionne, Switzerland. *Cold Reg. Sci. Technol.* **108**, 149–162 (2014).
27. Bellaire, S., van Herwijnen, A., Mitterer, C. & Schweizer, J. On forecasting wet-snow avalanche activity using simulated snow cover data. *Cold Reg. Sci. Technol.* **144**, 28–38 (2017).
28. Côté, K., Madore, J. B. & Langlois, A. Uncertainties in the SNOWPACK multilayer snow model for a Canadian avalanche context: sensitivity to climatic forcing data. *Phys. Geogr.* **38**, 124–142 (2017).
29. Kumar, S., Srivastava, P. K. & Snehamani, S. Geospatial modelling and mapping of snow avalanche susceptibility. *J. Indian Soc. Remote Sens.* **46**, 109–119 (2018).
30. Valero, C. V., Wever, N., Christen, M. & Bartelt, P. Modeling the influence of snow cover temperature and water content on wet-snow avalanche runout. *Nat. Hazards Earth Syst. Sci.* **18**, 869–887 (2018).
31. Voiculescu, M. Patterns of the dynamics of human-triggered snow avalanches at the Făgăraş massif (Southern Carpathians), Romanian Carpathians. *Area* **46**, 328–336 (2014).
32. Reuter, B., Proksch, M., Löwe, H., Van Herwijnen, A. & Schweizer, J. Comparing measurements of snow mechanical properties relevant for slab avalanche release. *J. Glaciol.* **65**, 55–67 (2019).
33. Wever, N., Vera Valero, C. & Techel, F. Coupled snow cover and avalanche dynamics simulations to evaluate wet snow avalanche activity. *J. Geophys. Res. Earth Surf.* **123**, 1772–1796 (2018).
34. Gauthier, F., Germain, D. & Héту, B. Logistic models as a forecasting tool for snow avalanches in a cold maritime climate: northern Gaspésie, Québec, Canada. *Nat. Hazards* **89**, 201–232 (2017).
35. Heck, M., Hammer, C., Van Herwijnen, A., Schweizer, J. & Fäh, D. Automatic detection of snow avalanches in continuous seismic data using hidden Markov models. *Nat. Hazards Earth Syst. Sci.* **18**, 383–396 (2018).
36. Helbig, N., Van Herwijnen, A. & Jonas, T. Forecasting wet-snow avalanche probability in mountainous terrain. *Cold Reg. Sci. Technol.* **120**, 219–226 (2015).
37. Benjamin, J. *et al.* Transferability of a calibrated numerical model of rock avalanche run-out: application to 20 rock avalanches on the Nuussuaq Peninsula, West Greenland. *Earth Surf. Process. Landforms* **43**, 3057–3073 (2018).
38. Bühler, Y. *et al.* Automated identification of potential snow avalanche release areas based on digital elevation models. *Nat. Hazards Earth Syst. Sci.* **13**, 1321–1335 (2013).
39. De' Michieli Vitturi, M., Esposti Ongaro, T., Lari, G. & Aravena, A. IMEX-SfloW2D 1.0: a depth-averaged numerical flow model for pyroclastic avalanches. *Geosci. Model Dev.* **12**, 581–595 (2019).
40. Gaume, J., Chambon, G., Eckert, N. & Naaim, M. Influence of weak-layer heterogeneity on snow slab avalanche release: application to the evaluation of avalanche release depths. *J. Glaciol.* **59**, 423–437 (2013).
41. Rauter, M., Kofler, A., Huber, A. & Fellin, W. FaSavageHutterFOAM 1.0: depth-integrated simulation of dense snow avalanches on natural terrain with OpenFOAM. *Geosci. Model Dev.* **11**, 2923–2939 (2018).
42. Germain, D. A statistical framework for tree-ring reconstruction of high-magnitude mass movements: case study of snow avalanches in eastern Canada. *Geogr. Ann. Ser. A Phys. Geogr.* **98**, 303–311 (2016).
43. Lavigne, A., Eckert, N., Bel, L. & Parent, E. Adding expert contributions to the spatiotemporal modelling of avalanche activity under different climatic influences. *J. R. Stat. Soc. Ser. C Appl. Stat.* **64**, 651–671 (2015).
44. Rheinberger, C. M. Learning from the past: statistical performance measures for avalanche warning services. *Nat. Hazards* **65**, 1519–1533 (2013).
45. Sinickas, A. & Jamieson, B. Comparing methods for estimating β points for use in statistical snow avalanche runout models. *Cold Reg. Sci. Technol.* **104–105**, 23–32 (2014).
46. Bergua, S. B., Piedrabuena, M. Á. P. & Alfonso, J. L. M. Snow avalanche susceptibility in the eastern hillside of the aramo range (Asturian central massif, cantabrian mountains, nw Spain). *J. Maps* **14**, 373–381 (2018).
47. Hancock, H., Prokop, A., Eckerstorfer, M. & Hendriks, J. Combining high spatial resolution snow mapping and meteorological analyses to improve forecasting of destructive avalanches in Longyearbyen, Svalbard. *Cold Reg. Sci. Technol.* **154**, 120–132 (2018).
48. Ni, S., Ma, C., Yang, H. & Zhang, Y. Spatial distribution and susceptibility analysis of avalanche, landslide and debris flow in Beijing mountain region. *J. Beijing For. Univ.* **40**, 81–91 (2018).

49. Techel, F. *et al.* Spatial consistency and bias in avalanche forecasts—a case study in the European Alps. *Nat. Hazards Earth Syst. Sci.* **18**, 2697–2716 (2018).
50. Aydin, A. & Eker, R. GIS-based snow avalanche hazard mapping: Bayburt-asagi dere catchment case. *J. Environ. Biol.* **38**, 937–943 (2017).
51. Covășnianu, A. *et al.* Mapping snow avalanche risk using GIS technique and 3D modeling. Case study-ceahlau national park. *Rom. Rep. Phys.* **56**, 476–483 (2011).
52. Kumar, S., Srivastava, P. K. & Snehamani, S. GIS-based MCDA–AHP modelling for avalanche susceptibility mapping of Nubra valley region, Indian Himalaya. *Geocarto Int.* **32**, 1254–1267 (2017).
53. Santos González, J., Redondo Vega, J. M., Gómez Villar, A. & González Gutiérrez, R. B. Snow avalanches in the Alto Sil (West of the Cantabrian Mountains, Spain). *Cuad. Investig. Geogr.* **36**, 7–26 (2010).
54. Suk, P. & Klimánek, M. Creation of the snow avalanche susceptibility map of the krkonoše mountains using gis. *Acta Univ. Agric. Silvic. Mendelianae Brun.* **59**, 237–246 (2011).
55. Abermann, J., Eckerstorfer, M., Malnes, E. & Hansen, B. U. A large wet snow avalanche cycle in West Greenland quantified using remote sensing and in situ observations. *Nat. Hazards* **97**, 517–534 (2019).
56. Meseşan, F., Man, T. C., Pop, O. T. & Gavrilă, I. G. Reconstructing snow-avalanche extent using remote sensing and dendrogeomorphology in Parâng Mountains. *Cold Reg. Sci. Technol.* **157**, 97–109 (2019).
57. Abedini, M. *et al.* A novel hybrid approach of Bayesian logistic regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* **34**, 1427–1457 (2019).
58. Bragagnolo, L., da Silva, R. V. & Grzybowski, J. M. V. Artificial neural network ensembles applied to the mapping of landslide susceptibility. *CATENA* **184**, 104240 (2020).
59. Choubin, B. *et al.* Earth fissure hazard prediction using machine learning models. *Environ. Res.* **179**, 108770 (2019).
60. Fanos, A. M. & Pradhan, B. A novel hybrid machine learning-based model for rockfall source identification in presence of other landslide types using LiDAR and GIS. *Earth Syst. Environ.* **3**, 491–506 (2019).
61. Hulbert, C. *et al.* Similarity of fast and slow earthquakes illuminated by machine learning. *Nat. Geosci.* **12**, 69–74 (2019).
62. Ochoa Gutiérrez, L. H., Vargas Jiménez, C. A. & Niño Vásquez, L. F. Fast estimation of earthquake arrival azimuth using a single seismological station and machine learning techniques. *Earth Sci. Res. J.* **23**, 103–109 (2019).
63. Choubin, B. *et al.* Snow avalanche hazard prediction using machine learning methods. *J. Hydrol.* **577**, 123929 (2019).
64. Mosavi, A. *et al.* Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping. *IEEE Access* **8**, 145968–145983 (2020).
65. Hosseini, M. Effect of land use change on water balance and suspended sediment yield of Taleghan catchment, Iran. (Universiti Putra Malaysia, 2010).
66. Bayat, R., Ghermez Cheshmeh, B. & Refahi, H. Investigating on the role of canopy cover of sediment yield, a case study in Taleghan watershed. *J. Watershed Eng. Manag.* **3**, 187–195 (2012).
67. Afshar, S., Saghafian, B. & Vosoughifar, H. R. Source apportionment of the sediments entering dam using lithological and mineralogical studies. *Glob. J. Environ. Sci. Manag.* **2**, 265–274 (2016).
68. Luckman, B. H. Drop stones resulting from snow-avalanche deposition on lake ice. *J. Glaciol.* **14** <https://www.cambridge.org/core> (1975).
69. Freppaz, M. *et al.* Soil erosion caused by snow avalanches: a case study in the Aosta Valley (NW Italy). *Arct. Antarct. Alp. Res.* **42**, 412–421 (2010).
70. Whitaker, A., Alila, Y., Beckers, J. & Toews, D. Evaluating peak flow sensitivity to clear-cutting in different elevation bands of a snowmelt-dominated mountainous catchment. *Water Resour. Res.* **38**, 11-1-11-17 (2002).
71. FAO. Guidelines: land evaluation for irrigated agriculture. *FAO Soils Bull.* **55**, 231 (1985).
72. Choubin, B. *et al.* Spatiotemporal dynamics assessment of snow cover to infer snowline elevation mobility in the mountainous regions. *Cold Reg. Sci. Technol.* <https://doi.org/10.1016/j.coldregions.2019.102870> (2019).
73. Jenness, J. Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x, v. 1.2. Jenness Enterprises. Available at: <https://www.jennessent.com/arcview/tpi.htm>. (2006).
74. Gauthier, D. & Jamieson, B. On the sustainability and arrest of weak layer fracture in whumpfs and avalanches. *Proc. ISSW* **2010**(1), 224–231 (2010).
75. Sajedi-Hosseini, F., Choubin, B., Solaimani, K., Cerdà, A. & Kavian, A. Spatial prediction of soil erosion susceptibility using a fuzzy analytical network process: application of the fuzzy decision making trial and evaluation laboratory approach. *L. Degrad. Dev.* **29**, 3092–3103 (2018).
76. Alilou, H. *et al.* Evaluation of watershed health using Fuzzy-ANP approach considering geo-environmental and topo-hydrological criteria. *J. Environ. Manag.* **232**, 22–36 (2019).
77. Liu, H. X., Xu, Q. & Li, Y. R. Effect of lithology and structure on seismic response of steep slope in a shaking table test. *J. Mt. Sci.* **11**, 371–383 (2014).
78. Alin, A. Multicollinearity. *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 370–374 (2010).
79. Rafiei Sardooi, E. *et al.* Applying the remotely sensed data to identify homogeneous regions of watersheds using a pixel-based classification approach. *Appl. Geogr.* **111**, 102071 (2019).
80. Chen, X. W. & Jeong, J. C. Enhanced recursive feature elimination. In *Proceedings—6th International Conference on Machine Learning and Applications, ICMLA 2007*. <https://doi.org/10.1109/ICMLA.2007.44> (2007).
81. Vapnik, V. N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10**. <https://ieeexplore.ieee.org/abstract/document/788640/> (1999).
82. Karatzoglou, A., Smola, A., Hornik, K. & Karatzoglou, M. Package ‘kernlab’. Technical report (2018).
83. Kuhn, M. Package ‘caret’—classification and regression training. *CRAN Repos.* **216**. <https://doi.org/10.1887/0750303123/b365c43> (2019).
84. Naimi, B. & Araújo, M. B. Sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography (Cop.)* **39**, 368–375 (2016).
85. Lewis-Beck, M., Bryman, A. & Futing Liao, T. Generalized additive models. In *The SA GE Encyclopedia of Social Science Research Methods* (eds Lewis-Beck, M. S. *et al.*) (Sage, Thousand Oaks, 2012). <https://doi.org/10.4135/9781412950589.n368>.
86. Jones, K. & Almond, S. Moving out of the linear rut: the possibilities of generalized additive models. *Trans. Inst. Br. Geogr.* **17**, 434 (1992).
87. Jones, K. & Wrigley, N. Generalized additive models, graphical diagnostics, and logistic regression. *Geogr. Anal.* **27**, 1–18 (1995).
88. Wood, S. N. mgcv: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL. v 1.8-15. *stat.ethz.ch* (2016).
89. Hastie, T. J. Generalized additive models. In *Statistical Models in S*. <https://doi.org/10.1201/9780203738535> (2017).
90. Friedman, J. H. Multivariate adaptive regression splines. *Ann. Stat.* <https://doi.org/10.1214/aos/1176347963> (1991).
91. Zhang, W. & Goh, A. T. C. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci. Front.* **7**, 45–52 (2016).
92. Kisi, O., Choubin, B., Deo, R. C. & Yaseen, Z. M. Incorporating synoptic-scale climate signals for streamflow modelling over the Mediterranean region using machine learning models. *Hydrol. Sci. J.* **64**, 1240–1252 (2019).

93. Milborrow, S. Derived from mda: Mars by T. Hastie and R. Tibshirani. earth: Multivariate adaptive regression splines. R package version 4.4.3. (2015).
94. Schapire, R. E. The boosting approach to machine learning: an overview. https://doi.org/10.1007/978-0-387-21579-2_9 (2003).
95. De'ath, G. Boosted trees for ecological modeling and prediction. *Ecology* **88**, 243–251 (2007).
96. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
97. Sajedi-Hosseini, F. *et al.* A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci. Tot. Environ.* **644**, 954–962 (2018).
98. Greg Ridgeway. Package 'gbm' (2013).
99. Cortes, C., Vapnik, V. & Saitta, L. Support-vector networks editor. *Mach. Learn.* https://doi.org/10.1007/978-0-387-21579-2_9 (1995).
100. Vapnik, V. N. & Mukherjee, S. Support vector method for multivariate density estimation. *Adv. Neural. Inf. Process. Syst.* **9**, 659–665 (2000).
101. Mountrakis, G., Im, J. & Ogole, C. Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* **66**, 247–259 (2011).
102. Rahmati, O. *et al.* Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. *Sci. Tot. Environ.* **688**, 855–866 (2019).
103. Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–75 (1986).
104. Viera, A. J. & Garrett, J. M. Understanding interobserver agreement: the kappa statistic. *Fam. Med.* **37**, 360–363 (2005).
105. Panofsky, H. & Brier, G. Some applications of statistics to meteorology. 224 (1958).
106. Johnson, L. E. & Olsen, B. G. Assessment of quantitative precipitation forecasts. *Weather Forecast.* **13**, 75–83 (1998).
107. Choubin, B. *et al.* Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Sci. Tot. Environ.* **701**, 134474 (2020).
108. Kuhn, M. & Johnson, K. *Applied predictive modeling.* <https://doi.org/10.1007/978-1-4614-6849-3> (2013).
109. Chen, W., Yeo, C. K., Lau, C. T. & Lee, B. S. Real-time twitter content polluter detection based on direct features. In *2015 IEEE 2nd International Conference on Information Science Security ICISSEC 2015.* <https://doi.org/10.1109/ICISSEC.2015.7371027> (2016).
110. Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **2**, 121–167 (1998).
111. Bachofer, F., Quénéhervé, G., Märker, M. & Hochschild, V. Comparison of SVM and boosted regression trees for the delineation of lacustrine sediments using multispectral ASTER data and topographic indices in the lake Manyara Basin. *Photogramm. Fernerkundung Geoinf.* **2015**, 81–94 (2015).
112. Foody, G. M. & Mathur, A. A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans. Geosci. Remote Sens.* **42**. <https://ieeexplore.ieee.org/abstract/document/1304900/> (2004).

Acknowledgements

We thank the support of Alexander von Humboldt Foundation.

Author contributions

B.C. and M.B. conceived of the presented idea. M.B. and F.S.H. collected the required dataset. B.C., A.M. and A.A.D. developed the theory. B.C. and F.S.H. performed the computations and carried out the simulations. A.M. and A.A.D. verified the analytical methods and contributed to the interpretation of the results. B.C. supervised the project. M.B. and F.S.H. designed the figures and tables. All authors wrote the paper, discussed the results, and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-75476-w>.

Correspondence and requests for materials should be addressed to A.M. or A.A.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020