

Research papers

Snow avalanche hazard prediction using machine learning methods

Bahram Choubin^a, Moslem Borji^a, Amir Mosavi^{b,c}, Farzaneh Sajedi-Hosseini^a, Vijay P. Singh^d, Shahaboddin Shamshirband^{e,f,*}

^a Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural Resources, University of Tehran, Karaj, Iran^b School of the Built Environment, Oxford Brookes University, Oxford, UK^c Institute of Automation, Kando Kalman Faculty of Electrical Engineering, Obuda University, Budapest, Hungary^d Department of Biological & Agricultural Engineering, Texas A&M University, College Station, TX 77843-2117, USA^e Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Viet Nam^f Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief, with the assistance of Hamid Moradkhani, Associate Editor

Keywords:

Snow avalanche
Natural disasters
Hazard prediction
Machine learning models
Support vector machine (SVM)
Multivariate discriminant analysis (MDA)

ABSTRACT

Snow avalanches are among the most destructive natural hazards threatening human life, ecosystems, built structures, and landscapes in mountainous regions. The complexity of snow avalanche modelling has been discussed in many studies, but its modelling is not well-documented. Snow avalanche modeling in this study was done using three main categories of data, including avalanche occurrence locations, meteorological factors, and terrain characteristics. Two machine learning models, namely support vector machine (SVM) and multivariate discriminant analysis (MDA), were employed. A ratio of 70 to 30 of data was considered for calibrating and validating the models. Results indicated that both models had an excellent performance in snow avalanche modeling (area under curve, AUC > 90), although hits and misses analysis demonstrated the superior performance of MDA. Sensitivity analysis indicated that the topographic position index, slope, precipitation, and topographic wetness index were the most effective variables for modeling. A snow avalanche map indicated that the high snow avalanche hazard zone was mostly near the streams and was matched with hillsides around the water pathways. Findings of study can be helpful for land use planning, to control snow avalanche paths, and to prevent the probable hazards induced by it, and it can be a good reference for future studies on modeling snow avalanche hazards.

1. Introduction

Snow avalanches are among the most destructive natural hazards threatening human lives, ecosystems, built structures, and landscapes in mountainous regions. International Commission for Alpine Rescue (ICAR) (<http://www.alpine-rescue.org>) reports an increasing number of fatal accidents due to snow avalanches, with an annual average of 138, based on a 20-year survey in Alpine countries and North America. The actual number of deaths may be higher, as the database does not fully include the off-piste and backcountry incidents. The risk of snow avalanches is increasing worldwide due to climate change (Ballesteros-Cánovas et al., 2018). Among the climate change contributors, the climate warming is identified as one of the factors influencing the behavior, uncertainty, and increasing hazard frequencies of snow avalanches (Martin et al., 2001; Castebrunet et al., 2014). In some areas, global warming, mainly, causes the thinning and shortening of the duration of snow cover contributing to increased risk and irregularity

which increase the hazard (Gádeck et al., 2017). Therefore, the prediction of changes in snow cover and snow avalanche behavior projection is vital for administering many crucial societal issues concerning dynamic adaptation, risk mitigation, and resilience strategies (Komarov et al., 2017; Hovelsrud et al., 2018).

The Geographic Information Systems (GIS) have been identified as a powerful tool for snow avalanche terrain recognition, suitable for mountain risk and disaster management, with high capacity to enhance terrain visualization models (Jaedicke et al., 2014; Bourova et al., 2016). GIS has long enabled the web-based visualizing terrain software packages for snow avalanche hazard assessment (Stoffel et al., 2001; Haid et al., 2008; Lan et al., 2009). Furthermore, GIS is essential to enable the preparation of terrain's database for reliable prediction models (Noetzli et al., 2006; Kumar and Srivastava, 2018). For instance, the analytical hierarchy process method (AHP) can be coupled with GIS to advance models of prediction and snow avalanche susceptibility mapping (Wu and Chen, 2009; Pourghasemi, et al., 2016). Although

* Corresponding author at: Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Viet Nam.
E-mail address: shahaboddin.shamshirband@tdtu.edu.vn (S. Shamshirband).

these methods can detect snow avalanche hazards, however, the high amount of false alarm ratio is a drawback in using such models (Bellaire et al., 2017).

The operational verification of regional forecasts is also a popular approach which relies on high-quality field observations. Such local nowcasts can be used for hazard verification. They include substantial uncertainty which makes them unreliable (Techel and Schweizer, 2017). Gauthier et al. (2017) developed a logistic regression (LR) model for daily prediction of snow avalanches in the cold maritime climate located in northern Gaspésie, Québec, Canada. Their model was reported as an efficient tool to enhance daily decision making and for reducing risk for the mobility in the region.

The complexity of snow avalanche modelling has been discussed in a multitude of studies (Sharma and Ganju, 2000; Gruber and Bartelt, 2007; Valero et al., 2016). The terrain, snowpack, and meteorology are the three principal interacting factors motivating the snow avalanche release (Fredston et al., 1994). The interaction of these factors which are involved in motion and runoff of snow is referred to as avalanche formation (Schweizer et al., 2003). Additionally, the human factor has been identified as a major contributor in forming snow avalanches and risk thereof (Voiculescu, et al., 2016; Statham et al., 2018). Thus, advancement of novel technologies for prediction as well as warning systems with higher accuracies has been encouraged in many studies (Helbig et al., 2015; Van Herwijnen et al., 2016).

To the best of our knowledge, applications of machine learning (ML) algorithms to snow avalanche modelling have been very limited. Only limited studies applied artificial neural networks (ANN) as a promising approach for the prediction of avalanche. Singh and Ganju (2008) indicated the potential of ANN in detecting avalanche days in Indian Himalayas. Dekanová et al. (2018) proposed different architectures of ANNs and evaluated model performance and obtained promising results. However, modeling snow avalanches, especially their spatial prediction needs to be more widely addressed. In this context, the contribution of this paper is to develop novel models of SVM and MDA for the prediction of snow avalanches. Although using the SVM and MDA in modeling snow avalanches is novel, their applications for the modeling of other destructive natural hazards and extreme events have recently received great interest in the literature. For instance, the SVM has been widely used for the prediction of earthquakes (Hajikhodaverdikhan et al., 2018; Asim et al., 2018), floods (Sharifi Garmdareh et al., 2018; Bafithile and Li, 2019), drought (Mokhtarzad et al., 2017; Xu et al., 2018), and typhoons (Chang et al., 2018; Yang et al., 2018), as well as for landslide susceptibility analysis (Chen et al., 2017; Xie et al., 2017). The MDA has also shown great performance in the modeling of various hazards, such as groundwater contamination and floods (Sajedi-Hosseini et al., 2018b; Choubin et al., 2019). However, the current study aimed to evaluate performance of the SVM and MDA models for the prediction of snow avalanches. The organization of the rest of this paper is as follows. Section two discusses the study area. Section three presents the methodology, data used for snow avalanche modeling, the description of the ML models used, and building of the models. Section four presents the results and discussion. Section five draws conclusions of the study and presents the future direction of the ML methods in snow avalanche prediction.

2. Study area

The study area extends in the Karaj watershed, which is between latitudes of $35^{\circ} 52'$ to $36^{\circ} 10'$ north, and longitudes of $51^{\circ} 03'$ and $51^{\circ} 36'$ east (Fig. 1). The study area is about 844.97 km^2 , where the highest point is 4365 m a.s.l. located in the east and the lowest elevation is 1633 m a.s.l. located in the southwest. The Karaj dam (or Amir Kabir) is located at the outlet of the watershed. This watershed is most important, because it is the main source of drinking water for Tehran, irrigation water for Karaj plain, and also is the source of hydroelectric power production (Heidarnejad et al., 2006; Sakizadeh, 2015). Also,

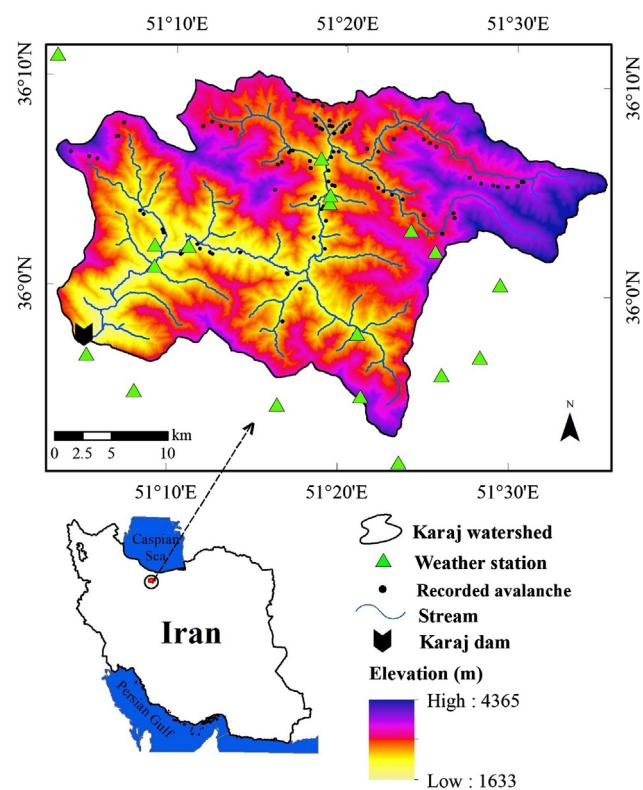


Fig. 1. Location of the study area.

Karaj-Chalous road, as one of the most important and busier Iranian roads, is located in that. The road and watershed present a connection of Tehran to the north, the beautiful nature with the numerous residential areas, and access to ski resort has made travelers boom this road. In recent years, with the boom of winter sports, this road also has significant traffic to reach the ski resort in the winter. The passage by road through the rocky and steep slopes that lack any preserving vegetation or supporting structures for controlling snow has led avalanches to descend toward the passengers on this road. In many cases, people are killed and vehicles fall into the valley. Also, the North-Tehran Freeway is one of the most important construction projects in Iran, in that 22 km of this freeway is located at the site of dangerous snow avalanches.

3. Methodology

3.1. Data used for snow avalanche modeling

Appropriate circumstances creating snow avalanches are affected by the interactions of three main factors of snowpack, weather, and terrain (Fredston et al., 1994; Delparte et al., 2008). The snowpack information is mostly not available, while weather conditions and terrain characteristics are most important to initiate a snow avalanche (Delparte, 2008). However, in this study according to the data availability, three main categories of data, including avalanche occurrence inventory map, meteorological factors, and terrain characteristics were used for snow avalanche modeling.

3.1.1. Snow avalanche occurrence inventory map

In order to prepare a snow avalanche occurrence inventory map, the locations of avalanche occurrences were collected through many field observations and surveys. In this study, at first, the locations of recorded points were approximately selected, based on different maps and digital layers (like slope and valleys) that affect snow avalanches. Then, through the Google Earth, these points were identified, and at the

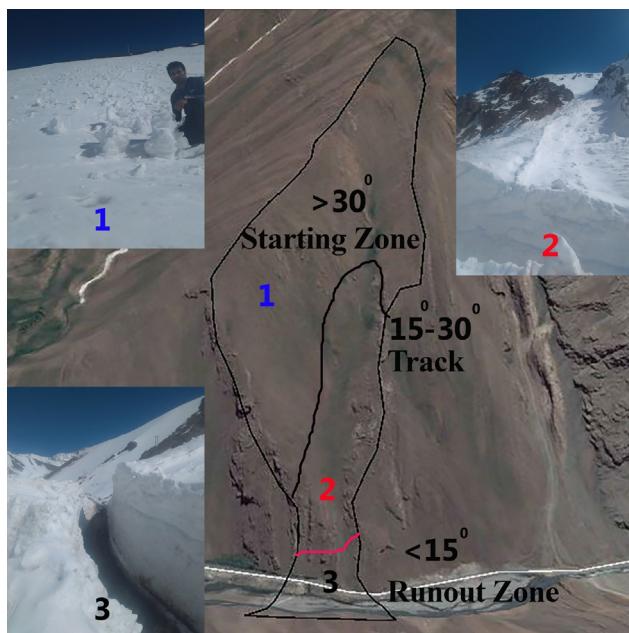


Fig. 2. A sample of snow avalanches that occurred in the study.

end, the locations of snow avalanches were confirmed by field surveys. Moreover, we surveyed different slopes, elevations, and lithological and morphometric structures to record snow occurrence locations. 91 locations were recorded during field surveys (Fig. 1). A sample of recorded snow avalanches in the study area with different zones is shown in Fig. 2.

3.1.2. Meteorological factors

Precipitation and temperature are the most important meteorological factors that affect snow accumulation and formation of weak layers within the snowpack. The amount of snowfall and the temperature gradient within the snowpack contribute to the overall stability of snowy areas. Meteorological factors considered in this study were precipitation (PCP) and temperature (Temp) from cold months (December to March) when snow falls (Fig. 3a and b). The PCP and Temp data recorded by the stations (Fig. 1) were obtained from Iranian Meteorological Organization as well as Iranian Water Resource Management Company (IWRMC). Although wind direction and exposing the slopes to wind are other important factors in snow accumulation and avalanche occurrence (Delparte, 2008) but were not available.

3.1.3. Terrain factors

Topography and terrain characteristics are the most important factors that affect snow avalanche activity (Naaim et al., 2004). Considering different terrain characteristics is the most important in snow avalanche modeling, for example, flat regions (which are distinct by slope or plan curvature) can be both safe and unsafe for the occurrence of avalanches. By elevation, they can be separated, as flat regions with low elevation are more hazardous areas than flat regions with high elevation. In this study we considered terrain factors, including elevation, slope, aspect, curvature, drainage density (DD), topographic wetness index (TWI), topographic position index (TPI), vector ruggedness measure (VRM), distance to stream (DTS), distance to fault (DTF), land use, and lithology (Fig. 3). An ASTER digital elevation map (DEM) with 10×10 m pixel size was used to extract most of the terrain factors.

Elevation (Fig. 3c) can affect precipitation, temperature, solar radiation, and windward, so it is a significant factor and can produce favorable conditions for avalanche formation (Smith and McClung, 1997; Delparte, 2008). Slope (Fig. 3d) is a necessary factor to create a

slide. Aspect (Fig. 3e) affects solar radiation received at the snowpack and influences the temperature gradient and snow surface warming (Bakermans and Jamieson, 2006). Curvature (Fig. 3f) can indicate deposition and refuge areas (Delparte, 2008). Drainage density indicates the density of streams per unit ground that can be important for avalanche occurrence. The drainage density was produced in ArcGIS software through line density tool (Fig. 3g).

TWI indicates the spatial variability of wetness and varies between 3.3 and 18.3 (Fig. 3h). TPI is the difference in elevation in each cell with the mean elevation of the surrounding cells (De Reu et al., 2013). It indicates the upper and lower parts of a landscape. Positive values show locations which are higher than their surroundings (ridges), while negative values indicate locations that are lower than neighborhood cells (valleys). TPI in the study area varies between -58.8 and 74.2 (Fig. 3i). VRM (Fig. 3j) indicates the roughness of the ground that is most important in the movement of snow avalanches. Irregular surfaces such as rocks and rock outcrops may promote instability or stability. In some locations mostly the bottom of starting zone may control the movement of snowpack, while by creating higher temperature gradients and slope upstream of the starting zone may foster a very locally weaker snowpack. In this study, TWI, TPI, and VRM were created in the SAGA GIS environment.

The presence of streams or faults can accelerate the snowpack movement and create a snow avalanche. The areas near DTS and DTF must be more susceptible to the avalanche occurrence and were created using Euclidean distance tool in GIS software (Fig. 3k and l). Streams and fault layers were obtained from the IWRMC, and were used to calculate DTS and DTF maps.

Other land factors, such as land use and lithology, influence snowpacks. In this study, the land use map (1:50,000) and geological map (1:50,000) were obtained from the Forests, Range and Watershed Management Organization (FRWMO) of Iran, and were converted to raster format (Fig. 3m and n, respectively). Landuse, such as the presence of forest, can control avalanching. According to the land use map, rangelands cover most of the study area (Fig. 3m). Lithology and rock structures are the main factors responsible for the slope failure in natural hazards, such as landslides, debris flows, and snow avalanches (Liu et al., 2014). Different lithological units may change the susceptibility of the study area into an avalanche zone. The lithology map indicates the Ek unit, well bedded green tuff, and tuffaceous shale (KARAJ formation) of Eocene Cenozoic have the most area of the Karaj watershed (Fig. 3n).

3.2. Snow avalanche modeling

Snow avalanche modeling was done using the location of avalanche occurrence (as the dependent variable) and 14 predictor variables shown in Fig. 3 as independent variables and input data. We used two machine learning models, namely support vector machine (SVM) and multivariate discriminant analysis (MDA), for modeling snow avalanche. A ratio of 70 to 30 of data was considered for calibrating and validating the models. The models used are described as follows:

3.2.1. SVM

Support-vector networks have progressed through the work of Cortes and Vapnik (1995), and emerged as SVM suitable for two-group classification problems. The special properties of decision surface in SVM ensure high generalization ability of the learning machine which makes it efficient in dealing with non-separable training datasets (Drucker et al., 1997). The continued advancement of SVM with the implementation of structural risk minimization inductive principle has improved its generalization in a limited number of learning patterns which makes it useful for recognizing subtle patterns in complex datasets (Basak et al., 2007). In SVM the vectors represent the selection criteria as the best boundaries for data classification with more safety margin. Kecman (2001) provides a detailed description of model

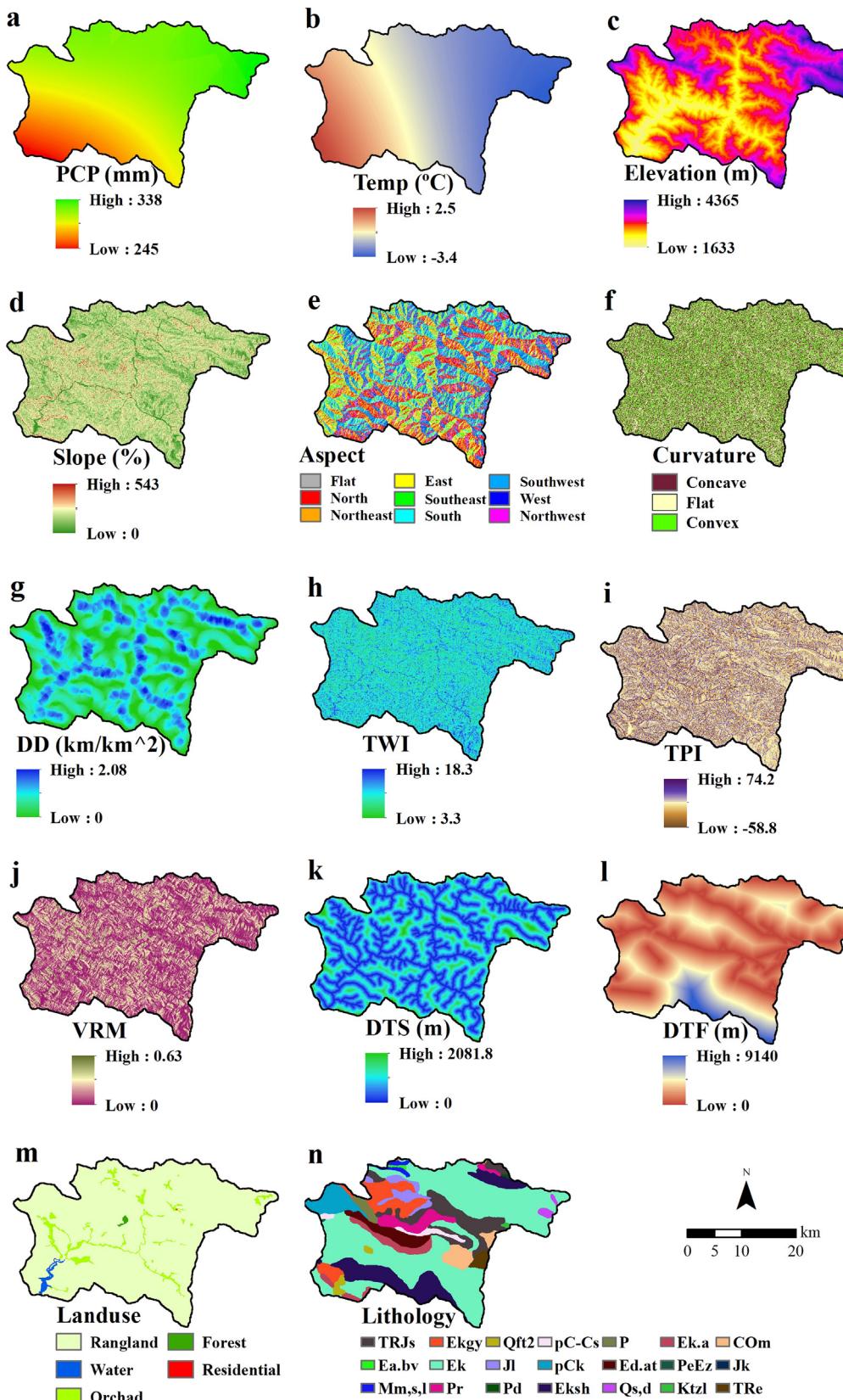


Fig. 3. Snow avalanche predictor variables: a) precipitation (PCP), b) temperature (Temp), c) elevation, d) slope, e) aspect, f) curvature, g) drainage density (DD), h) topographic wetness index (TWI), i) topographic position index (TPI), j) vector ruggedness measure (VRM), k) distance to stream (DTS), l) distance to fault (DTF), m) land use, and n) lithology.

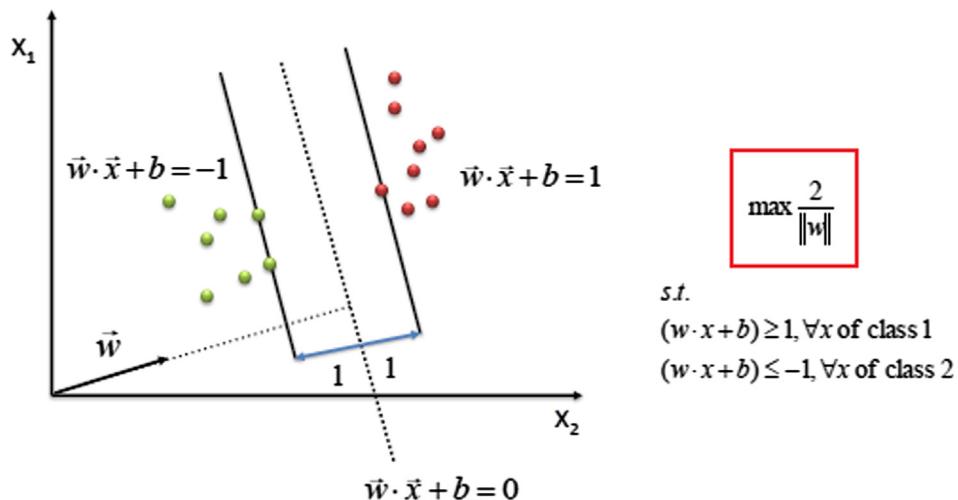


Fig. 4. Illustrating two hyper planes with a maximum margin, including the boundaries of classification and support vector samples located on the borders (Kecman, 2001).

building and implementation of SVM.

The classification error reduces along with increasing boundary separation and the distance between two parallel hyper planes (see Fig. 4). The two-parallel hyperplanes on both sides of the maximum margin hyperplane classifies the data with linear decision boundary and can be represented, following Raghavendra and Deka (2014), as:

$$w \cdot x + b = 0 \quad (1)$$

where b , x , and w , represent the bias, points on the decision boundary; and n -dimensional weight vector orthogonal to the hyper plane, respectively.

$$w \cdot x_i + b = 1 \quad (2)$$

The boundary of the maximum margin is obtained through the minimum magnitude of weights and maximum margins of classes:

$$\max_{w,b} \min_{i=1 \dots L} \frac{|y_i(w \cdot x_i + b)|}{\|w\|} \quad (3)$$

Eq. (3) can further be presented as:

$$\min_{w,b} \frac{1}{2} \|w\|^2, y_i(w \cdot x_i + b) - 1 \geq 0 \quad i = 1, \dots, L \quad (4)$$

When the classes are separable through the Lagrange multiplier of λ_i the optimization problem according to Chao and Horng (2015) can be reconfigured as:

$$\begin{aligned} & \min_{\sum_{i=1}^L \lambda_i} \left[\sum_{i=1}^L \lambda_i - \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L \lambda_i y_i (x_j \cdot x_j) y_j \lambda_j \right] \\ & \lambda_i \geq 0 \quad i = 1, \dots, L \quad \sum_{i=1}^L \lambda_i y_i = 0 \end{aligned} \quad (5)$$

When the classes are non-separable and the linear decision boundaries overlap, the error increases, the optimization problem, following Demyanova et al. (2017), can be formulated as:

$$w = \sum_{i=1}^L \sum_{j=1}^L \lambda_i y_i x_i \quad (6)$$

$$\lambda_i (y_i(w \cdot x_i + b - 1)) = 0 \quad i = 1, \dots, L \quad (6)$$

$$\begin{aligned} & \max_{\sum_{i=1}^L \lambda_i} \left[\sum_{i=1}^L \lambda_i - \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L \lambda_i y_i (\varphi(x_i) \varphi(x_j)) y_j \lambda_j \right] \\ & C \gg \lambda_i \gg 0 \\ & \sum_{i=1}^L \lambda_i y_i = 0 \end{aligned} \quad (7)$$

Computation of non-separable cases is done through kernel functions where x is mapped into the high-dimensional space of φ . Linear kernel, polynomial kernel, Sigmoid kernel, and Radial basis kernel are presented, respectively, as follows:

$$K(x_i \cdot x_j) = \varphi(x_i) \varphi(x_j) \quad (8)$$

$$K(x, y) = (xy + 1)^p \quad (9)$$

$$K(x, y) = \tanh(kxy - \delta) \quad (10)$$

$$K(x, y) = e^{-\|x-y\|^2 / 2\sigma^2} \quad (11)$$

Due to the flexibility of radial basis kernel in dealing with different dimensionalities of data and its excellent generalization capability, it has been found suitable for the proposed snow avalanche modeling (Jiao et al., 2016; Karballaeenezadeh et al., 2019). The drawbacks to modeling with SVM are generally associated with its difficulty to capture important modeling variables (Mosavi et al., 2018; Choubin et al., 2019).

3.2.2. MDA

The principle of MDA was presented by Hair et al. (1998), and later advanced by Johnson and Wichern (2002) for the classification of observations of multiple independent variables. MDA builds efficient linear combinations of independent variables which are called discriminant functions for the classification of observations into predetermined nearest groups where the distance is calculated by the normal distribution of parameters while it is assumed that the correlation and variability among the parameters of each group are equal (Lombardo et al., 2006). Accordingly, a discriminant function can be described as:

$$Y = W_1 X_1 + W_2 X_2 + \dots + W_n X_n \quad (12)$$

where Y , W_i , and X_i represent the discriminant score, the discriminant weights, and the independent variables, respectively. MDA has shown an acceptable performance in critical modeling variables (Elith et al., 2008; Sajedi-Hosseini et al., 2018b; Choubin et al., 2019).

3.3. Sensitivity analysis

The relative importance of the variables was assessed using the Jackknife test (Efron, 1982). We excluded one of the snow avalanche predictor variables (in turn) from the datasets and modeling, then calculated the performance of modeling at each time. Two statistics of the decrease in area under the curve (DAUC; Eq. (1)) of the receiver

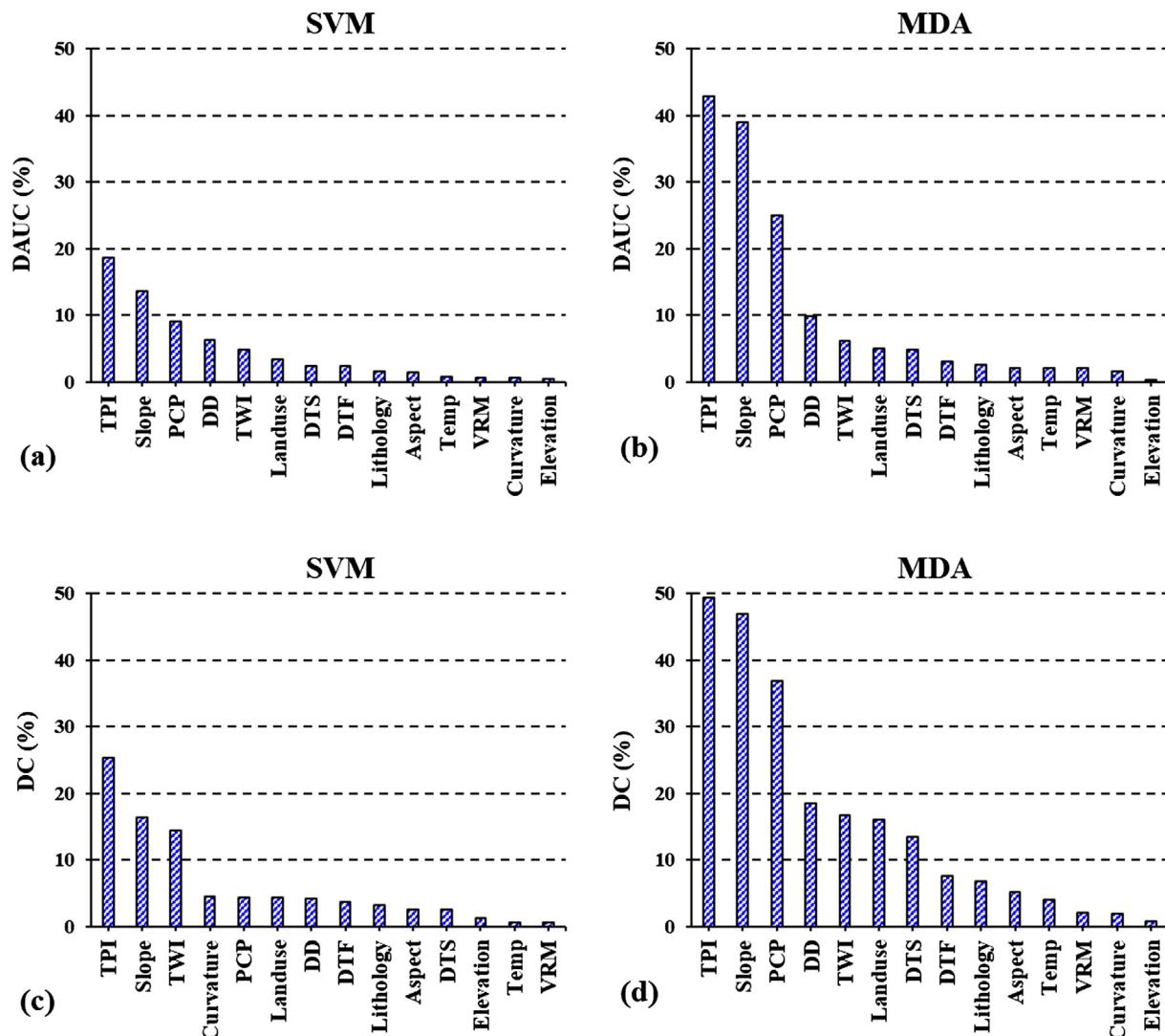


Fig. 5. Importance of variables in snow avalanche modeling: (a) and (b) decrease in AUC (%), (c) and (d) decrease in correlation (%).

operator characteristic (ROC) plot and a decrease in correlation (DC; Eq. (2)) were calculated after excluding the variables:

$$DAUC_i = \frac{[AUC_{all} - AUC_{i1}]}{AUC_{all}} \quad (13)$$

$$DC_i = \frac{[C_{all} - C_i]}{C_{all}} \quad (14)$$

where i is the variable, AUC_{all} and C_{all} are, respectively, the AUC and correlation values computed with the existence of all variables in the modeling process. AUC_i and C_i are the values of the decrease in AUC and correlation, respectively, after excluding the i th variable. $DAUC_i$ and DC_i are, respectively, the decreases in AUC and correlation when the i th variable has been excluded from modeling, and they were represented as a percentage.

3.4. Model evaluation

The model evaluation was done using the AUC statistic which is the most popular criterion for evaluating spatial modeling performance (Sajedi-Hosseini et al., 2018a). Moreover, we calculated other statistics, including accuracy, bias, probability of detection (POD; Panofsky and Brier, 1968), false alarm ratio (FAR; Burke et al., 1988), critical success index (CSI; Stanski et al., 1989), and Heidke skill score (HSS; Heidke, 1926) from hits and misses analysis. These indices are calculated using

the contingency table (Stanski et al., 1989; Johnson and Olsen, 1998), with a number of hits (H), false alarms (FA), misses (M), and correct negatives (CN).

Accuracy indicates the fraction of predictions that are correct (Eq. (3)). Accuracy is between 0 and 1 where 1 shows the perfect prediction. Bias (or frequency bias) is calculated by the ratio of the predicted occurrences divided by observed occurrences (Eq. (4)). The perfect score of bias is equal to 1. POD (or hit rate) is the ratio of the number of hits to the total number of observed occurrences (Eq. (5)). The perfect value of POD is equal to 1. FAR is the number of false alarms divided by the total number of predicted occurrences (Eq. (6)). FAR is between 0 and 1 that lower values are better. CSI (or threat ratio) is the ratio of successfully predicted occurrences to the number of observed occurrences plus the number of FAs (Eq. (7)). CSI changes between 0 and 1, where perfect prediction is equal to 1. Since some correct occurrence predictions can be random, HSS (Eq. (8)) removes random predictions and specifies the fraction of correct predictions (Sigaroodi et al., 2014). The HSS varies between -1 and 1 for dichotomous (yes/no) forecasts. Negative values show that the casual prediction is better, 0 means no skill, while the value of 1 indicates a perfect prediction.

$$\text{Accuracy} = \frac{H + CN}{H + CN + M + FA} \quad (15)$$

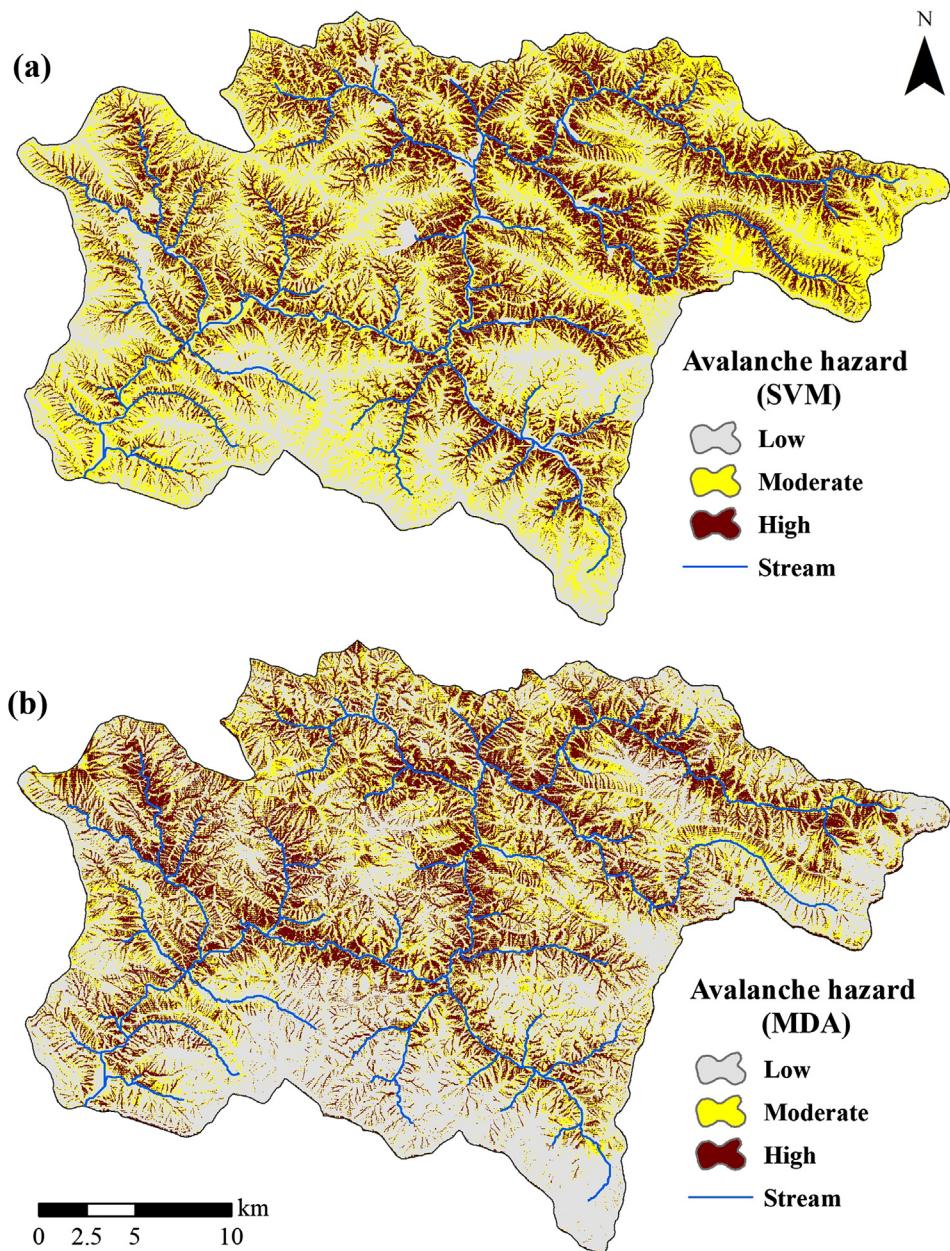


Fig. 6. Snow avalanche hazard map produced by SVM (a) and MDA (b) models.

Table 1
Snow avalanche hazard classes.

Hazard class	Area (km ²)		Area (%)	
	SVM	MDA	SVM	MDA
Low	289.86	421.84	34.30	49.92
Moderate	344.51	204.79	40.77	24.23
High	210.60	218.34	24.93	25.85

$$CSI = \frac{H}{H+M+FA} \quad (19)$$

$$HSS = \frac{2[(H \times CN) - (FA \times M)]}{[(H + M)(M + CN) + (H + FA)(FA + CN)]} \quad (20)$$

4. Results and discussion

4.1. Analysis of variable importance

Sensitivity analysis of the SVM model indicated that the most important variables for modeling were TPI (DAUC = 18.6%, DC = 25.22%), slope (DAUC = 13.6%, DC = 16.4%), PCP (DAUC = 9.0%, DC = 4.3%), and TWI (DAUC = 4.8%, DC = 14.38%), while drainage density (DAUC = 4.2%, DC = 6.3%), curvature (DC = 4.6%), and landuse (DAUC = 3.3%, DC = 4.3%) indicated the moderate importance due to the decrease in AUC or in correlation (Fig. 5a and c). Also, sensitivity analysis of the MDA model identified

$$Bias = \frac{H + FA}{H+M} \quad (16)$$

$$POD = \frac{H}{H+M} \quad (17)$$

$$FAR = \frac{FA}{H+FA} \quad (18)$$

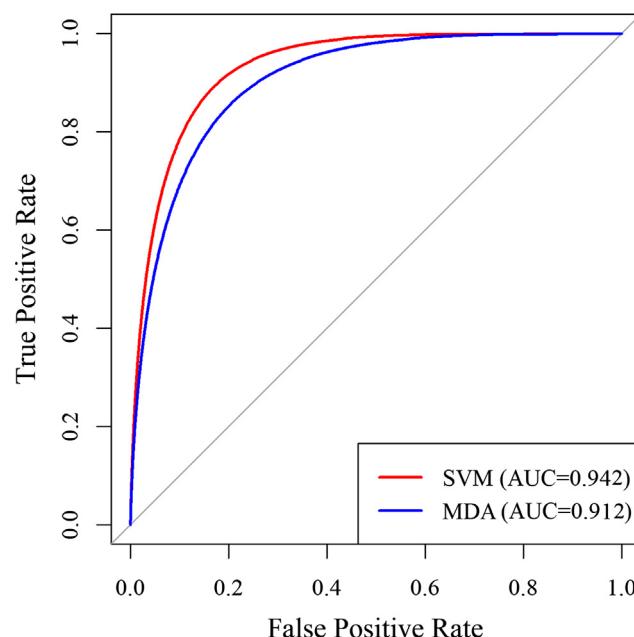


Fig. 7. The area under the curve (AUC) of models for testing datasets.

Table 2
Statistics for model evaluation.

Statistic	Training dataset		Testing dataset	
	SVM	MDA	SVM	MDA
Accuracy	0.90	0.87	0.83	0.85
Bias	1.11	1.16	1.11	1.07
POD	0.95	0.95	0.89	0.89
FAR	0.14	0.18	0.20	0.17
CSI	0.82	0.79	0.73	0.75
HSS	0.80	0.75	0.67	0.70

that TPI (DAUC = 42.8%, DC = 49.3%), slope (DAUC = 39.0%, DC = 46.0%), PCP (DAUC = 25.0%, DC = 36.8%), and drainage density (DAUC = 9.8%, DC = 18.5%) were the main factors influencing snow avalanche occurrence (Fig. 5b and d). Moreover, TWI (DAUC = 6.0%, DC = 16.7%) and landuse (DAUC = 5.0%, DC = 16.0%) were of moderate importance due to the decrease in AUC or in correlation (Fig. 5b and d).

Due to high complexity of snow avalanche processes, understanding the manner of governing parameters is not yet fully manifested (Singh and Ganju, 2008). However, the terrain information, such as slope, affects snow stability, the characteristics of snow cover, and avalanche danger (Bakermans and Jamieson, 2006; Dekanová et al., 2018). Sensitivity analysis indicated that elevation was mostly an unimportant variable (Fig. 5). The recorded snow avalanches in this study indicated the elevation change from 1911 to 3396 m, so the high elevation difference among snow avalanches that occurred can be a reason that elevation is an unimportant variable.

Comparison of the results from the Jackknife indicated that the performance of the MDA model was more sensitive than the SVM model, as excluding one variable reduced the performance of MDA (Fig. 5).

4.2. Snow avalanche hazard mapping

The SVM and MDA models were employed for snow avalanche hazard mapping with the predictors and avalanche occurrence locations. The model prediction output was classified into three classes of low, moderate, and high hazard zones, according to the natural breaks

(Jenks) classification scheme.

Fig. 6 indicates the snow avalanche hazard maps produced by the SVM and MDA models. As can be seen, the high hazard zone is mostly near the streams and matches with hillsides around the water pathways.

The MDA model indicated the most significant class of hazards was the low class (421.84 km^2 ; about 50% of the study area), while the greatest area in the SVM was the moderate class (344.51 km^2 ; about 41% of the study area) (Fig. 6; Table 1). Predicted maps indicated low hazard class in the SVM model was about 34% of the study area, while the MDA model covered about 50% of the study area. Approximately 41% and 24% of the study area belonged to the moderate class, respectively, for the SVM and MDA models. The high hazard classes predicted by the models were approximately equal, as SVM indicated 24.93% and MDA showed 25.85% of the study area (Fig. 6; Table 1).

4.3. Model evaluation and comparison

The AUC statistic from the ROC indicated that models had excellent performance (> 90%, Yesilnacar, 2005; Choubin et al., 2019), as the values of AUC were equal to 0.912 and 0.942, respectively, for the MDA and SVM models for the testing periods (Fig. 7). In addition to AUC, we investigated hits and misses with the calculation of accuracy, Bias, POD, FAR, CSI, and HSS (Table 2).

From Table 2, the models indicated good accuracy, equal to 0.83 for SVM and 0.85 for MDA. The values of the Bias in models were close and equal to 1.11 and 1.07, respectively, for SVM and MDA in testing datasets. Results indicated that the probability of snow avalanche detection in the models was the same in both training (POD = 95%) and testing (POD = 0.89%) phases. Although SVM had a lower FAR (FAR = 0.14) in the training phase, MDA indicated lower FAR (FAR = 0.17) than SVM (FAR = 0.20) in the testing phase. According to the CSI and HSS indices, although the difference in performance of the models was not high, the MDA indicated higher values (CSI = 0.75 and HSS = 0.70) than did SVM (CSI = 0.73 and HSS = 0.67) in the testing phase (Table 2).

Due to the lack of any related previous studies on the SVM and MDA models used for snow avalanche modeling, a direct comparison of the models with previous studies was not possible in this regard. However, in other natural hazard assessment fields, applications of SVM and MDA have been recently compared in the literature. For instance, Sajedi-Hosseini et al. (2018b) indicated that the SVM model had a superior performance to the MDA (respectively accuracy was equal to 81% and 87%) for the prediction of nitrate groundwater contamination. In another study, like our results, Choubin et al. (2019) confirmed approximately the same performance of MDA and SVM (respectively, with accuracy 89% and 88%), which applied for flood susceptibility prediction in the Khiyav-Chai watershed in Iran.

5. Conclusion

This study modeled snow avalanche hazard using two machine learning methods. Results from the models indicated that the two models had an excellent performance in snow avalanche modeling (AUC > 90). Sensitivity analysis indicated that the topographic position index, slope, precipitation, and topographic wetness index were the most effective variables for modeling. A snow avalanche hazard map indicated that the high hazard zone was mostly near the stream and matched with hillsides around the water pathway. The recorded locations of snow avalanches can be a challenging issue to study snow avalanche hazard, because a snow avalanche includes starting, track, and runout zones for which the topographic information of each zone (like slope, TPI, DTS, and curvature) is the difference. It seems that applying different zones (as the location of occurring avalanches) can create different hazard mapping results. In this study, we had recorded the locations of starting zones and used them in snow avalanche

modeling. Therefore, comparison of hazard maps concerning different snow avalanche zones used in modeling can clarify this issue and help improve snow avalanche modeling. Adding other important information, such as snowpack data, the frequency of snowfall, and wind direction, which were not available in the study area, can improve snow avalanche modeling. The snowpack information, such as the volume of snowpack, snow stability, and the presence of weak layers, are effective for snow avalanche movement. Also, wind direction is an important factor in snow avalanche occurrence and causes deposition of snow on the leeward side and scouring of snow on the windward side. Our suggestions can be used in future studies for better modeling snow avalanche hazards. Our findings can be helpful for land-use planning, controlling avalanche paths, and preventing probable hazards, and can be a good reference for spatial modeling of snow avalanche hazard for future studies.

Declaration of Competing Interest

None.

Acknowledgments

The authors express their gratitude to Iranian Meteorological Organization, Iranian Water Resource Management Company (TAMAB), and the Forests, Range and Watershed Management Organization (FRWMO) of Iran for providing the climatic data and maps. Also, the authors would like to appreciate the Editor and three anonymous reviewers for their constructive criticism and comments.

References

- Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F., 2018. Earthquake prediction model using support vector regressor and hybrid neural networks. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0199004>.
- Bafithile, T.M., Li, Z., 2019. Applicability of e-Support Vector Machine and Artificial Neural Network for Flood Forecasting in Humid, Semi-Humid and Semi-Arid Basins in China. *Water* 11, 85. <https://doi.org/10.3390/w11010085>.
- Bakermans, L., Jamieson, B., 2006. Measuring near-surface snow temperature changes over terrain. In: Proceedings ISSW, pp. 377–386.
- Ballesteros-Cánovas, J.A., Trappmann, D., Madrigal-González, J., Eckert, N., Stoffel, M., 2018. Climate warming enhances snow avalanche risk in the Western Himalayas. *Proc. Natl. Acad. Sci.* 115 (13), 3410–3415.
- Basak, D., Pal, S., Patranabis, D.C., 2007. Support vector regression. *Neural Information Processing-Lett. Rev.* 11 (10), 203–224.
- Bellaire, S., van Herwijnen, A., Mitterer, C., Schweizer, J., 2017. On forecasting wet-snow avalanche activity using simulated snow cover data. *Cold Reg. Sci. Technol.* 144, 28–38.
- Bourova, E., Maldonado, E., Leroy, J.B., Alouani, R., Eckert, N., Bonnefoy-Demongeot, M., Deschartres, M., 2016. A new web-based system to improve the monitoring of snow avalanche hazard in France. *Nat. Hazards Earth Syst. Sci.* 16 (5), 1205–1216.
- Burke, D.S., Brundage, J.F., Redfield, R.R., Damato, J.J., Schable, C.A., Putman, P., Visintine, R., Kim, H.I., 1988. Measurement of the false positive rate in a screening program for human immunodeficiency virus infections. *N. Engl. J. Med.* 319 (15), 961–964.
- Castebrunet, H., Eckert, N., Giraud, G., Durand, Y., Morin, S., 2014. Projected changes of snow conditions and avalanche activity in a warming climate: the French Alps over the 2020–2050 and 2070–2100 periods. *The Cryosphere* 8 (5), 1673–1697.
- Chang, M.-J., Chang, H.-K., Chen, Y.-C., Lin, G.-F., Chen, P.-A., Lai, J.-S., Tan, Y.-C., 2018. A Support Vector Machine Forecasting Model for Typhoon Flood Indentation Mapping and Early Flood Warning Systems. *Water* 10, 1734. <https://doi.org/10.3390/w10121734>.
- Chao, C.F., Horng, M.H., 2015. The construction of support vector machine classifier using the firefly algorithm. *Comput. Intell. Neurosci.* 2015 (2). <https://doi.org/10.1155/2015/212719>.
- Chen, W., Pourghasemi, H.R., Panahi, M., Kornejad, A., Wang, J., Xie, X., Cao, S., 2017. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* 297, 69–85. <https://doi.org/10.1016/j.geomorph.2017.09.007>.
- Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019. An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci. Total Environ.* 651, 2087–2096.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learning* 20 (3), 273–297.
- De Reu, J., Bourgeois, J., Bats, M., Zwervvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P., Van Meirvenne, M., 2013. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* 186, 39–49.
- Dekanová, M., Duchoň, F., Dekan, M., Kyžek, F., Biskupič, M., 2018. In: May. Avalanche forecasting using neural network. IEEE, pp. 1–5.
- Delporte, D., 2008. Avalanche Terrain Modeling in Glacier National Park. University of Calgary, Calgary, Canada.
- Delporte, D., Jamieson, B., Waters, N., 2008. Statistical runout modeling of snow avalanches using GIS in Glacier National Park Canada. *Cold Reg. Sci. Technol.* 54 (3), 183–192.
- Demyanova, Y., Pani, T., Veith, H., Zuleger, F., 2017. Empirical software metrics for benchmarking of verification tools. *Formal Methods in System Design* 50 (2–3), 289–316.
- Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J., Vapnik, V., 1997. Support vector regression machines. In: Advances in neural information processing systems (pp. 155–161).
- Efron, B., 1982. The Jackknife, the Bootstrap, and Other Resampling Plans 38 SIAM.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77 (4), 802–813.
- Fredston, J., Fesler, D., Tremper, B., 1994. The human factor—Lessons for avalanche education. In: Proc. 1994 International Snow Science Workshop, pp. 473–487.
- Gądek, B., Kaczka, R.J., Rączkowska, Z., Rojan, E., Casteller, A., Bebi, P., 2017. Snow avalanche activity in *Zięb Żandarmierii* in a time of climate change (Tatra Mts., Poland). *Catena* 158, 201–212.
- Gauthier, F., Germain, D., Hétu, B., 2017. Logistic models as a forecasting tool for snow avalanches in a cold maritime climate: northern Gaspésie, Québec Canada. *Nat. Hazards* 89 (1), 201–232.
- Gruber, U., Bartelt, P., 2007. Snow avalanche hazard modelling of large areas using shallow water numerical methods and GIS. *Environ. Modell. Software* 22 (10), 1472–1481.
- Haid, E., Kiechle, G., Göll, N., Soutschek, M., 2008. Evaluation of a web-based and mobile ski touring application for GPS-enabled smartphones. *Information and communication technologies in tourism 2008*, 313–323.
- Hair Junior, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate data analysis. New Jersey.
- Hajikhodaverdikhan, P., Nazari, M., Mohsenizadeh, M., Shamshirband, S., Chau, K., 2018. Earthquake prediction with meteorological data by particle filter-based support vector regression. *Eng. Appl. Comput. Fluid Mech.* 12, 679–688. <https://doi.org/10.1080/19942060.2018.1512010>.
- Heidarnejad, M., Golmaee, S.H., Mosaeedi, A., Ahmadi, M.Z., 2006. Estimation of sediment volume in Karaj Dam Reservoir (Iran) by hydrometry method and a comparison with hydrography method. *Lake Reservoir Manage.* 22 (3), 233–239.
- Heidke, P., 1926. Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdiest. *Geogr. Ann.* 8 (4), 301–349.
- Helbig, N., Van Herwijnen, A., Jonas, T., 2015. Forecasting wet-snow avalanche probability in mountainous terrain. *Cold Reg. Sci. Technol.* 120, 219–226.
- Hovelsrud, G.K., Karlsson, M., Olsen, J., 2018. Prepared and flexible: Local adaptation strategies for avalanche risk. *Cogent Social Sciences* 4 (1), 1460899.
- Jaedicke, C., Syre, E., Sverdrup-Thygeson, K., 2014. GIS-aided avalanche warning in Norway. *Comput. Geosci.* 66, 31–39.
- Jiao, G., Guo, T., Ding, Y., 2016. A new hybrid forecasting approach applied to hydrological data: a case study on precipitation in northwestern China. *Water* 8 (9), 367.
- Johnson, L.E., Olsen, B.G., 1998. Assessment of quantitative precipitation forecasts. *Weather Forecasting* 13 (1), 75–83.
- Johnson, R.A., Wichern, D.W., 2002. Applied multivariate statistical analysis (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.
- Karballaezadeh, N., Mohammadzadeh, S.D., Shamshirband, S., Hajikhodaverdikhan, P., Mosavi, A., Chau, K.W., 2019. Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan-Firuzkuh road). *Eng. Appl. Comput. Fluid Mech.* 13 (1), 188–198.
- Kecman, V., 2001. Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. MIT press.
- Komarov, A., Seliverstov, Y., Sokratov, S., Glazovskaya, T., Turchaninova, A., 2017. Avalanche risk assessment in Russia. *EGU General Assembly Conference Abstracts* 19, 9104.
- Kumar, S., Srivastava, P.K., 2018. Geospatial Modelling and Mapping of Snow Avalanche Susceptibility. *J. Indian Soc. Remote Sens.* 46 (1), 109–119.
- Lan, H.X., Martin, C.D., Froese, C.R., Kim, T.H., Morgan, A.J., Chao, D., Chowdhury, S., 2009. A web-based GIS for managing and assessing landslide data for the town of Peace River, Canada. *Nat. Hazards Earth Syst. Sci.* 9 (4), 1433–1443.
- Liu, H.X., Xu, Q., Li, Y.R., 2014. Effect of lithology and structure on seismic response of steep slope in a shaking table test. *J. Mountain Sci.* 11 (2), 371–383.
- Lombardo, F., Obach, R.S., DiCapua, F.M., Bakken, G.A., Lu, J., Potter, D.M., Gao, F., Miller, M.D., Zhang, Y., 2006. A hybrid mixture discriminant analysis – random forest computational model for the prediction of volume of distribution of drugs in human. *J. Med. Chem.* 49 (7), 2262–2267.
- Martin, E., Giraud, G., Lejeune, Y., Boudart, G., 2001. Impact of a climate change on avalanche hazard. *Ann. Glaciol.* 32, 163–167.
- Mokhtarzad, M., Eskandari, F., Jamshidi Vanjani, N., Arabasadi, A., 2017. Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environ. Earth Sci.* 76, 729. <https://doi.org/10.1007/s12665-017-7064-0>.
- Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: Literature review. *Water* 10 (11), 1536.
- Naaim, M., Naaim-Bouvet, F., Faug, T., Bouchet, A., 2004. Dense snow avalanche modeling: flow, erosion, deposition and obstacle effects. *Cold Reg. Sci. Technol.* 39 (2–3), 193–204.
- Noetzli, J., Huggel, C., Hoelzle, M., Haeberli, W., 2006. GIS-based modelling of rock-ice

- avalanches from Alpine permafrost areas. *Comput. Geosci.* 10 (2), 161–178.
- Panofsky, H.A., Brier, G.W., 1968. Some applications of statistics to meteorology.
- Pourghasemi, H.R., Beheshtirad, M., Pradhan, B., 2016. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. *Geomatics, Nat. Haz. Risk* 7 (2), 861–885.
- Raghavendra, S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput.* 19, 372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>.
- Sajedi-Hosseini, F., Choubin, B., Solaimani, K., Cerdà, A., Kavian, A., 2018a. Spatial prediction of soil erosion susceptibility using FANP: application of the Fuzzy DEMATEL approach. *Land Degrad. Dev.* 29, 3092–3103.
- Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., Pradhan, B., 2018b. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci. Total Environ.* 644, 954–962.
- Sakizadeh, M., 2015. Assessment the performance of classification methods in water quality studies, a case study in Karaj River. *Environ. Monit. Assess.* 187 (9), 573.
- Schweizer, Jürg, Bruce Jamieson, J., Schneebeli, Martin, 2003. Snow avalanche formation. *Rev. Geophys.* 41, 4.
- Sharifi Garmdareh, E., Vafakhah, M., Eslamian, S.S., 2018. Regional flood frequency analysis using support vector regression in arid and semi-arid regions of Iran. *Hydrolog. Sci. J.* 63, 426–440. <https://doi.org/10.1080/02626667.2018.1432056>.
- Sharma, S.S., Ganju, A., 2000. Complexities of avalanche forecasting in Western Himalaya—an overview. *Cold Reg. Sci. Technol.* 31 (2), 95–102.
- Sigaroodi, S.K., Chen, Q., Ebrahimi, S., Nazari, A., Choobin, B., 2014. Long-term precipitation forecast for drought relief using atmospheric circulation factors: a study on the Maharloo Basin in Iran. *Hydrol. Earth Syst. Sci.* 18 (5), 1995–2006.
- Singh, A., Ganju, A., 2008. Artificial Neural Networks for Snow Avalanche Forecasting in Indian Himalaya. 12th Int Conf. Int. Assoc. Comput. Methods Adv. Geomech.
- Smith, M.J., McClung, D.M., 1997. Avalanche frequency and terrain characteristics at Rogers' pass, British Columbia Canada. *J. Glaciol.* 43 (143), 165–171.
- Stanski, H.R., Wilson, L.J., Burrows, W.R., 1989. Survey of common verification methods in meteorology.
- Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., Kelly, J., 2018. A conceptual model of avalanche hazard. *Nat. Hazards* 90 (2), 663–691.
- Stoffel, A., Brabec, B. and Stoeckli, U., 2001. GIS Applications at the Swiss Federal Institute for Snow and Avalanche Research. In: Proceedings of the 2001 ESRI International User Conference, San Diego.
- Techel, F., Schweizer, J., 2017. On using local avalanche danger level estimates for regional forecast verification. *Cold Reg. Sci. Technol.* 144, 52–62.
- Valero, C.V., Wever, N., Bühler, Y., Stoffel, L., Margreth, S., Bartelt, P., 2016. Modelling wet snow avalanche runoff to assess road safety at a high-altitude mine in the central Andes. *Nat. Hazards Earth Syst. Sci.* 16 (11), 2303.
- Van Herwijnen, A., Heck, M., Schweizer, J., 2016. Forecasting snow avalanches using avalanche activity data obtained through seismic monitoring. *Cold Reg. Sci. Technol.* 132, 68–80.
- Voiculescu, M., Ardelean, F., Török-Oance, M., Milian, N., 2016. Topographical factors, meteorological variables and human factors in the control of the main snow avalanche events in the Făgăraș Massif (Southern Carpathians-Romanian Carpathians): case studies. *Geographia Polonica* 89 (1), 47–64.
- Wu, C.H., Chen, S.C., 2009. Determining landslide susceptibility in Central Taiwan from rainfall and six site factors using the analytical hierarchy process method. *Geomorphology* 112 (3–4), 190–204.
- Xie, Z., Chen, G., Meng, X., Zhang, Y., Qiao, L., Tan, L., 2017. A comparative study of landslide susceptibility mapping using weight of evidence, logistic regression and support vector machine and evaluated by SBAS-InSAR monitoring: Zhouqu to Wudu segment in Bailong River Basin, China. *Environ. Earth Sci.* 76, 313. <https://doi.org/10.1007/s12665-017-6640-7>.
- Xu, L., Chen, N., Zhang, X., 2018. A comparison of large-scale climate signals and the North American Multi-Model Ensemble (NMME) for drought prediction in China. *J. Hydrol.* 557, 378–390. <https://doi.org/10.1016/j.jhydrol.2017.12.044>.
- Yang, T.-C., Yu, P.-S., Lin, K.-H., Kuo, C.-M., Tseng, H.-W., 2018. Predictor selection method for the construction of support vector machine (SVM)-based typhoon rainfall forecasting models using a non-dominated sorting genetic algorithm. *Meteorol. Appl.* 25, 510–522. <https://doi.org/10.1002/met.1717>.
- Yesilnacar, E.K., 2005. The application of computational intelligence to landslide susceptibility mapping in Turkey. University of Melbourne, Department, p. 200.