A Hybrid Machine-Learning Model to Map Glacier-Related Debris Flow Susceptibility along Gyirong Zangbo Watershed Under the Changing Climate

Chenchen Qiu[a,b], Lijun Su[b], Qiang Zou[b], Xueyu Geng[a*]

[a] *School of Engineering, University of Warwick, Coventry, UK*

[b] *Key Laboratory of Mountain Hazards and Earth Surface Process/Institute of Mountain Hazards and Environment (IMHE), Chinese Academy of Sciences (CAS), Chengdu, China*

Abstract: Gyirong serves as an important channel to Chine-Nepal Economic Corridor, which is also the only land route for China-Nepal trade since the 2015 earthquake. However, the Gyirong corridor suffers from glacier-related debris flow from every April to September because of the complex topographic features and the changing climate. Therefore, a susceptibility map in response to precipitation and temperature change is timely, not only to ensure the safe operation of this corridor, but also to provide decision-makers a guidance for hazard mitigation and environmental remediation. Conventional method is difficult to consider and link the meteorological factors (e.g. temperature and precipitation), topographies, ecological, geological conditions all together to produce the susceptibility map, as such, machine learning is utilised to conduct the analysis. Logistic Regression (LR) and Support Vector Machine (SVM) were firstly applied to evaluate their efficiency and effectiveness of the performance of producing the susceptibility map. In order to improve the fitting and prediction accuracy (ACC), genetic algorithm - support vector machine (GA-SVM) and certainty factor - genetic algorithm - support vector machine (CF-GA-SVM) were conducted based on the initial analysis results of receiver operating characteristics curve (ROC) and ACC. Through the analysis, it can be seen that over 61% of the

study areas have a high susceptibility to debris flow, requiring an intensive attention from the local government. To further optimise the computational time, when dealing with small amounts of sample data, SVM is more efficient than LR, but CF-GA-SVM can achieve the highest AUC (Area Under Curve) and ACC values, 0.945 and 0.800, respectively. Overall, CF-GA-SVM model presents a relatively high robustness according to sensitivity analysis.

Keywords: Susceptibility maps, glacier-related debris flow, optimised SVM, China-Nepal Economic Corridor, environmental change

1. Introduction

The China-Nepal Economic Corridor is one of the important components of ‘One Belt, One Road’. However, the unique geological conditions have caused high-frequency and large-scale glacier-related debris flows and landslides along the corridor, which induced river blockages (Zou et al., 2020), slope erosions (Gayen et al., 2019), destruction of aquatic biodiversity (Zabihi et al., 2018), damage to agricultural land and forest (Jakob et al., 2005), and even losses of human life.

Debris flow can be divided into various types according to different triggering factors, including rainfall (Paudel et al., 2020), glacier and snow melting and outburst of ice lakes (Chen et al., 2011; Ding et al., 2020), in which the glacier-related debris flow causes the most severe damage to the natural environment. This is because the volume of glacier-related debris flow could reach hundreds of thousands of cubic meters (Breien et al., 2008). There have been a lot of research in the past decades to investigate the trigger mechanism of glacier-related debris flow (Perutz, 1953; Jackson et al., 1989; Wikerson et al., 2003). However, due to the complex controlling factors relating to geological structures, topographical and ecological conditions, it is still a challenge task to predict when the glacier-related debris flow is going to occur (Takahashi, 2007; Erokhin et al., 2018). Furthermore, an increase in the frequency of extreme weather and climate events also increased the occurrences of the glacier-related debris flow (Turkington et al., 2016; Muñoz et al., 2016). Therefore, more attentions need to be paid to the susceptibility analysis of glacier-related debris flow in order to provide early warnings and mitigate disasters caused by the glacier-related debris flow.

In recent years, with the aids from the data science, empirical methods (Meng et al., 2016; Kang et al., 2018) and machine learning methods (Kadavi et al., 2018) are emerged to generate the susceptibility map, conduct risk evaluation and estimate vulnerability for debris flow. A single empirical method can reflect the hazard susceptibility to some extent, such as, analytic hierarchy process (Chen et al., 2015; Xue et al., 2019), information acquisition analysis and certainty factor (Shortliffe, 1975; Heckerman, 1986). However, it is difficult to generate a high accurate hazard susceptibility map on a regional scale (Yilmaz, 2010) because the single empirical method is still highly depending on experts' experiences. Machine learning methods, such as logistic regression (Wright, 1995) and support vector machine (Pisner et al., 2020), can significantly reduce the computational time when generating the susceptibility maps (Rahmati et al., 2017; Lin et al., 2017; Zhang et al., 2019). In general, Logistic Regression (LR) and Support Vector Machine (SVM) algorithms are the two widely recognised algorithms (Kalantar et al., 2018) with different susceptibility accuracies. SVM method usually is more suitable in dealing with problems with high dimensions, nonlinearly separable data and a small amount of data (Mojaddadi et al., 2017; Huang et al., 2018). The reason is that SVM only focuses on selected support vectors when producing the model (Tong et al., 2001; Suthaharan et al., 2016), and the kernel functions, such as linear, polynomial, radial basis function (rbf) and sigmoid function, can convert samples into a higher-dimensional feature space to achieve linearly separable. Whereas, LR considers global variables (Wright, 1995) especially when it needs to deal with a large amount of data. When there are not enough data, LR may be able to produce a good fitting curve but will be unable to meet the prediction requirement because of the involved noise data.

Overall, the rbf kernel has a better nonlinear mapping capacity and shows a higher predictive accuracy (Tehrany et al., 2015; Feizizadeh et al., 2017). While the previous studies mostly only concentrated on the final result of SVM instead of focusing on the optimisation of hyper parameters *C* and gamma, which control the development of fitting model of the predictive accuracy. The genetic algorithm (GA) is suitable for extracting optimal *C* and gamma from a wide range, 2^{-8} to 2^{8} at a higher rate and avoid possible local optimisation of particle swarm optimisation (PSO). Moreover, in order to guarantee the stability of the fitting model, certainty factor (CF) (Devkota et al., 2013; Wang et al., 2019; Arabameri et al., 2019) is introduced to improve the stability of the input data to benefit the model fitting.

Although a lot of studies have been conducted to map susceptibility of debris flows caused by intense rainfall in different regions (Xu et al., 2013; Cao et al., 2020), very few studies considered the susceptibility analysis of glacier-related debris flow under changing weather conditions. Therefore, to map susceptibility of glacier-related debris flow along Gyirong Zangbo watershed and reflect the impacts of glacier-related debris flow in the natural environment, LR and SVM were applied in this paper to generate susceptibility maps on a regional scale by considering topographic, ecological, geological and meteorological factors. The genetic algorithm (GA) was also utilised to extract two optimal parameters *C* and gamma, which influences the output of fitting model and prediction accuracy. The certainty factors (CF) were integrated with

GA-SVM to further improve the performance of GA-SVM. Model assessments, including prediction accuracy, sensitivity analysis and weights of factors, were also conducted for model validation and examine the model’s robustness. The results from this research can provide effective support and guidance for the mitigation of glacier-related debris flow.

2. Study area

The study area is along the Gyirong Zangbo river from the headwater to Gyirong town (Fig. 1a), located in the Southwestern part of Tibet with a total area of 2,788.145 km^2, most of which belong to the landform of alpine valleys, with an average altitude of 3650 m above the sea level. This river basin extends from 28°15′24″ N to 29°0′14″ N latitude and 84°56′46″ E to 85°40′56″ E longitude. Three main rivers originate from this area, including Donglin Zangbo, Gyirong Zangbo and Buri Gandaki River. The complex geological and geomorphological conditions in Gyirong form the second highest mountain (‘Shishapangma’, with an altitude of 8012m), several glacial lakes and primitive forest. The study area belongs to subtropical monsoon climate zone with an average annual temperature ranging from 10 to 18℃. The average precipitation is approximately 1000 mm per year with over 50.5% of rainfall happens between June and October. Although the Himalaya Mountain stops the warm air from Indian Ocean and benefit the natural environment in south part of Gyirong, it also provides the initiation conditions for glacier-related debris flow along Gyirong Zangbo watershed. Fig. 1b presented the affected areas by glacier-related debris flow along Gyirong Zangbo watershed.

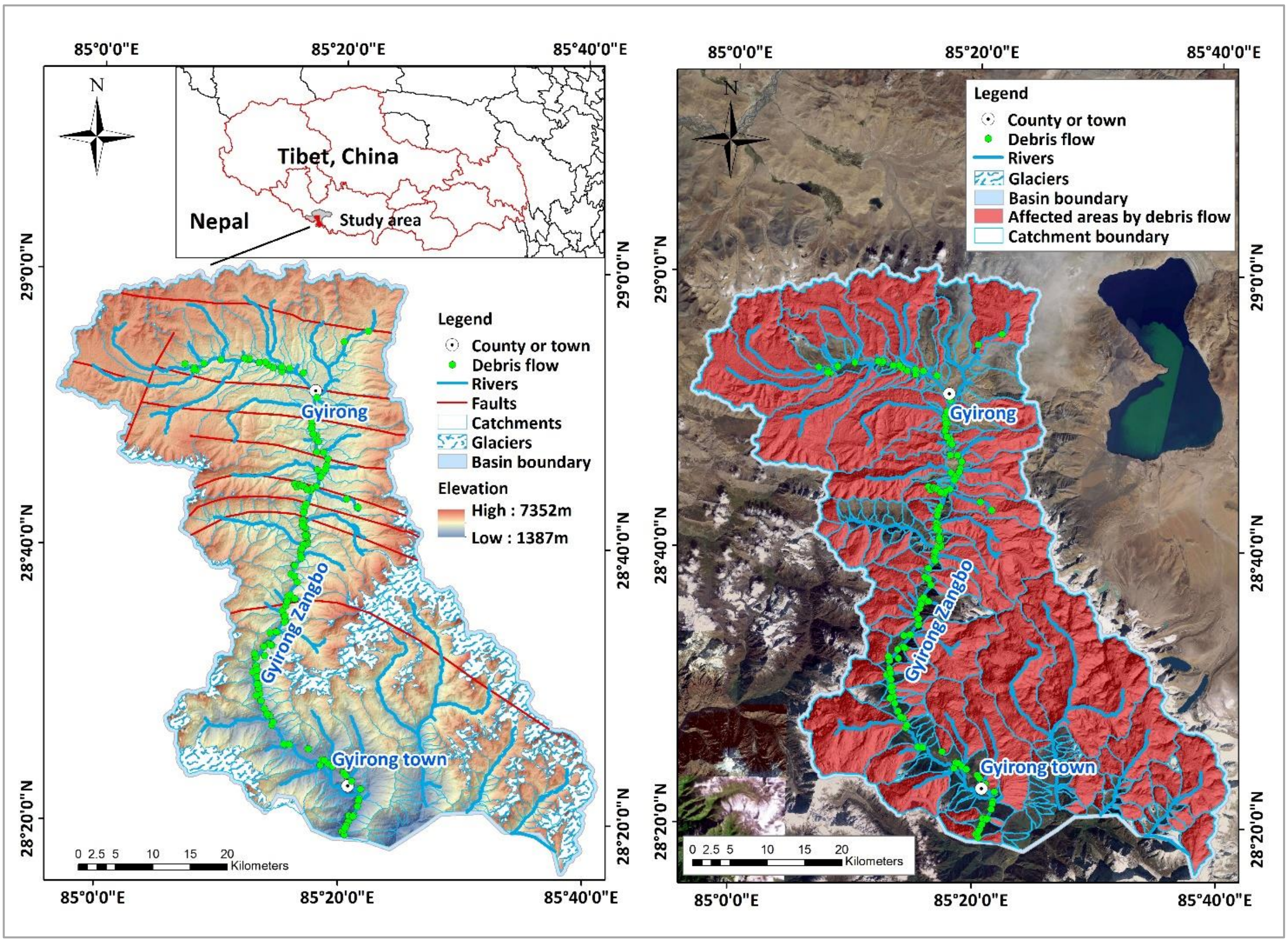

Figure 1. (a) Study area. (b) Affected areas by glacier-related debris flow. (The Digital Elevation Model (DEM) data was provided by the Geospatial Data Cloud, http://www.gscloud.cn)

3. Climate change in this area

Climate change has induced increased natural geohazards, such as debris flow, landslide, rock fall and flooding (Stoffel et al., 2014). From 1980 to 2015, there was an average annual temperature rising of 1.47℃ in the study area (Fig. 2), which was much higher than the global recorded temperature rising between 0.4℃ to 0.6℃ (Zou et al., 2020). The raised temperature triggered glacier melting, which formed streamflow with an average increase of 65 mm per 0.5 ℃ (Zhang et al., 2011). To make things even worse, a large volume of sediments on the slopes provides material source for the most common natural disaster, glacier-related debris flow with the contribution of the rainfall.

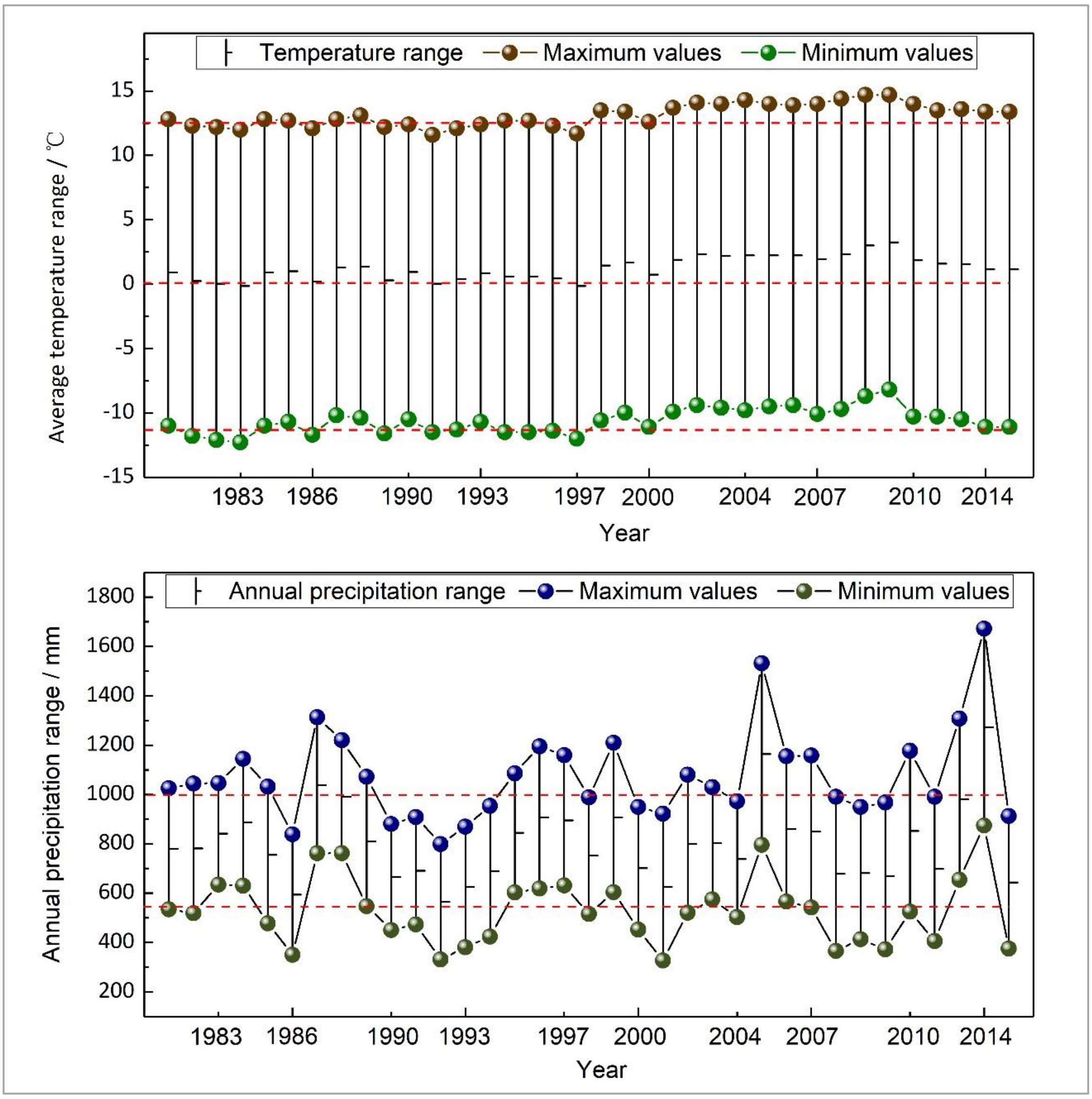

Figure. 2 Change of average annual temperature and annual precipitation in this area from 1980 to 2015

Table 1. Mean annual precipitation in five years internals

Time range	1980 -1984	1985 -1989	1990 -1994	1995 -1999	2000 -2004	2005 -2009	2010 -2015
Mean annual rainfall (mm)	822.13	837.14	646.91	860.88	733.09	846.55	852.58

In addition to the temperature rising in the past decades, annual precipitation also presented a noticeable change based on the curves in Fig.2. The raised temperature facilitated the evaporation of shallow soils and thus induced surface drying. However, the evaporated water will not return to the land immediately in the format of rainfall as the water holding capacity of air will be strengthen by the increased temperature with approximately 7% per 1 ℃ warming (Trenberth, 2011). Instead, the increased of water holding capacity of air leads to more frequent extreme weather events, such as storms, tropical cyclones (O'Gorman, 2015). These observed high intense rainfall events was consist with the precipitation change from 2000 to 2015 in Table 1. For example, the maximum precipitation occurred in 2014 with total amount of 1671.5 mm, followed by 1531.5 mm in 2005. It is inevitable that the climate change is leading to increased extreme weather conditions and associated natural geohazards. Therefore, in the past decades, not only the infrastructures in the regional residential areas were under high risks from these natural disasters induced by the rapid changes of the climates, but the well-developed vegetations were badly damaged. Moreover, the stable geological structures had also been disturbed which induced secondary geohazards with even more damages to the regional area. Through machine learning methods by taking meteorological factors into the consideration, a newly developed susceptibility map can provide a much better timely warning to the local residences to save lives and properties.

4. Data preparation

4.1 Modeling flow chart

The main steps in this study for the whole process in producing the susceptibility maps were showed in Fig.3. The debris flow inventory was determined through satellite images combine with field investigations. Satellite images provided the historical records of debris flow to identify the regions that suffered from debris flow in order for the field investigations to be carried out for a more detailed local investigation. The collected data were separated into a training set and a testing set to produce four fitting models. In the end, susceptibility maps were produced based on the constructed model. The results got from predictive accuracy (ACC) and receiver operating characteristic curve (ROC) were also evaluated and compared.

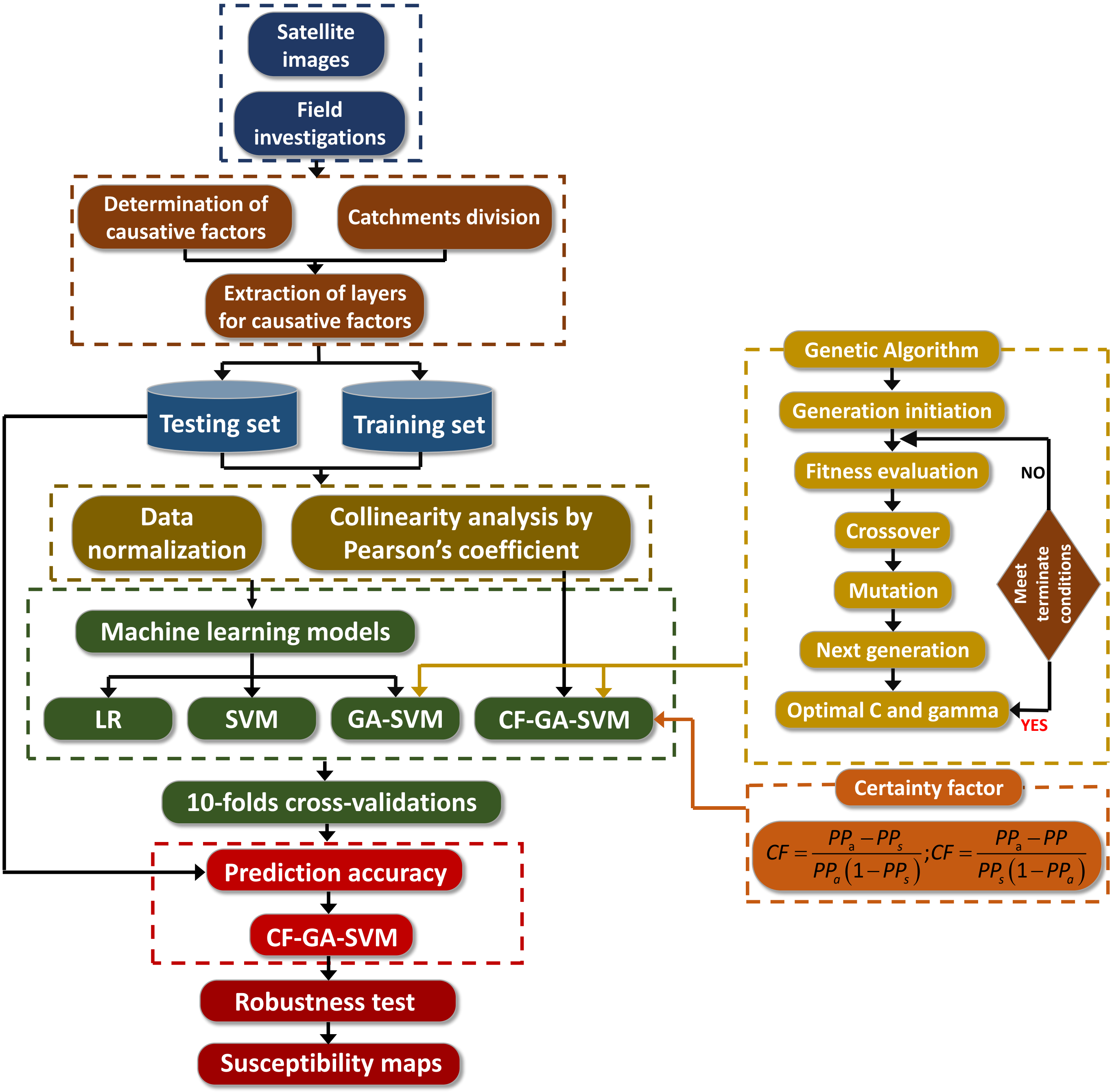

Figure 3. Modeling flow chart of this study

4.2 Determination of causative factors

There are many factors affecting the occurrence of glacier-related debris flow. Thus, the selection of causative factors is critical since it will directly influence the generated susceptibility map and also impact the computational time. Factors related to topographic, ecological, geological and meteorological conditions were selected based on the field investigations and past studies in Tibet (Chen et al., 2017; Liang et al., 2020). Slope, aspect, height difference, distance to stream and gully gradient were extracted from DEM with GIS. While NDVI, annual precipitation, average annual temperature, seismic intensity were based on the remote sensing data obtained from Resource and Environment Science and Data Center, Chinese Academy Sciences (http://www.resdc.cn). Similarly, lithology in geological group was acquired from the same database and field investigations.

Based on the determination of causative factors, susceptibility studies were required to be conducted through further classification for each factor. However, not all the factors can be classified quantificationally since both discrete data and continuous data were involved. The discrete data can be determined through field investigations and existing classification principles. However, the continuous data are difficult to decide since meteorological and geological conditions are varied in different regions in China. In order to decide the classifications of continuous data, the distribution of confirmed debris flow and relevant area ratio to total area were introduced to determine the classification principle for the continuous data. In addition, the classification of the slopes in this area was decided based on the study of Cui et al., (2014) and the reclassifications of gully gradient and seismic intensity were based on field investigations. The final classification principle (see Table 2) was decided based on the curves and past studies (Devkota et al., 2013; Chen et al., 2015).

Table 2. Causative factors for analysis of debris flow

No.	Causative factors	Reclassification groups	CF values	Groups
1	Slope	<6°	-0.758	Topographic Factors
		6-10°, ≥60°	-0.151	
		15-25°, 40-60°	0.712	
		25-40°	-0.133	
2	Aspect	<137.5	0.585	
		137.5-182.5	-0.382	
		182.5-227.5	-0.534	
		>227.5	0.670	
3	Height difference	<265	0.589	
		265-415m	-0.159	
		415-515m	-0.117	
		>515m	0.750	
4	Gully Gradient	<0.052	-0.702	
		0.052-0.105	-0.613	
		0.105-0.213	0.238	
		>0.213	0.697	
5	NDVI	<0.009	-0.377	Ecological Factor
		0.009-0.080	0.161	
		0.080-0.197	0.295	
		>0.197	-0.094	
6	Lithology	Solid rock	0.868	Geological Factors
		Soft rock	-0.168	
		Weak rock	-0.085	
		Soil and sand	/	
7	Fault distance (km)	<2.5	0.651	
		2.5-5.5	-0.214	
		5.5-9.5	-0.475	
		>9.5	-0.775	
8	Stream distance (km)	<2.25	0.848	
		2.25-4.25	0.573	
		4.25-5.75	-0.416	
		>5.75	-0.670	
9	Seismic intensity	<Ⅶ	-0.012	
		Ⅶ-Ⅸ	-0.258	
		>Ⅸ	/	
10	Annual precipitation (mm)	<662.5	0.732	Meteorological Factors
		662.5-712.5	0.398	
		712.5-787.5	-0.190	
		>787.5	-0.406	

11	Average annual temperature (℃)	<2.5	0.822
		2.5-6.5	-0.106
		6.5-9.5	-0.409
		≥9.5	-0.700

4.3 Collinearity analysis

Multicollinearity demonstrates the linear correlation between variables, which is, a certain independent variable can be indicated by other independent variables. Collinearity is a widespread phenomenon in the classification analysis and could cause unstable of the fitting model. Therefore, collinearity analysis is essential for causative factors to identify the degree of collinearity between factors. In this study, Person's Correlation Coefficient was applied to reflect the collinear relations among these factors through Python under the editing environment of Pycharm, in which 'math' package was utilised for the calculation. Slightly and moderate collinearity will not affect the fitting accuracy of model, but severe collinearity needs to be avoided. A correlation matrix was produced in this study (see Table 3). There was no correlation coefficient with the absolute value larger than 0.7 (Dormann et al., 2013) in Table 3, which meant all the variables can meet the demand of collinearity test.

Table 3. Parameter correlation matrix

Factors	Sl	HD	J	As	NDVI	Li	F	St	Ear	Pre	Temp
Sl	1	0.26	0.2	0.083	0.075	-0.07	0.37	0.67	-0.4	0.65	0.65
HD	0.26	1	0.057	-0.19	-0.27	0.073	-0.12	0.12	0.05	0.11	0.19
J	0.2	0.057	1	-0.19	-0.022	-0.14	0.24	0.22	-0.11	0.21	0.32
As	0.083	-0.19	-0.19	1	0.094	0.092	0.37	0.33	0.38	0.21	0.063
NDVI	0.075	-0.27	-0.022	0.094	1	0.092	0.15	0.053	-0.015	-0.095	0.14
Li	-0.07	0.073	-0.14	0.092	0.092	1	0.095	-0.003	0.11	-0.057	0.14
F	0.37	-0.12	0.24	0.37	0.15	0.095	1	0.6	-0.13	0.31	0.48
St	0.67	0.12	0.22	0.33	0.053	-0.003	0.6	1	-0.13	0.59	0.7
Ear	-0.4	0.05	-0.11	0.38	-0.015	0.11	-0.13	-0.13	1	-0.28	-0.24
Pre	0.65	0.11	0.21	0.21	-0.095	-0.057	0.31	0.59	-0.28	1	0.45
Temp	0.65	0.19	0.32	0.063	0.14	0.14	0.48	0.7	-0.24	0.45	1

('Sl' represents 'Slope', 'HD' is 'Height Difference', 'J' represents 'Gully Gradient', 'As' is 'Aspect', 'NDVI' is 'Normalised Difference Vegetation Index', 'Li' represents 'Lithology', 'F' is 'Fault Distance', 'St' is 'Stream Distance', 'Ear' represents 'Seismic Intensity', 'Pre' is 'Annual precipitation', 'Temp' is 'Average annual temperature')

4.4 Spatial distribution of causal factors

This study area was firstly divided into 164 catchments by using Digital Elevation Model (DEM) with a resolution of 30m, ranging from 0.1315 km^2 to 233.9289 km^2. The reason to use catchment analysis instead of girds is that debris flow usually includes accumulation area, flowing area and source zone and catchment analysis is much more reasonable and operational than grids to take topographic and geological conditions into consideration. Based on the satellite images and field investigations, a total number of 100 debris flow catchments were identified. The spatial distribution of causative factors was presented in Fig. 4.

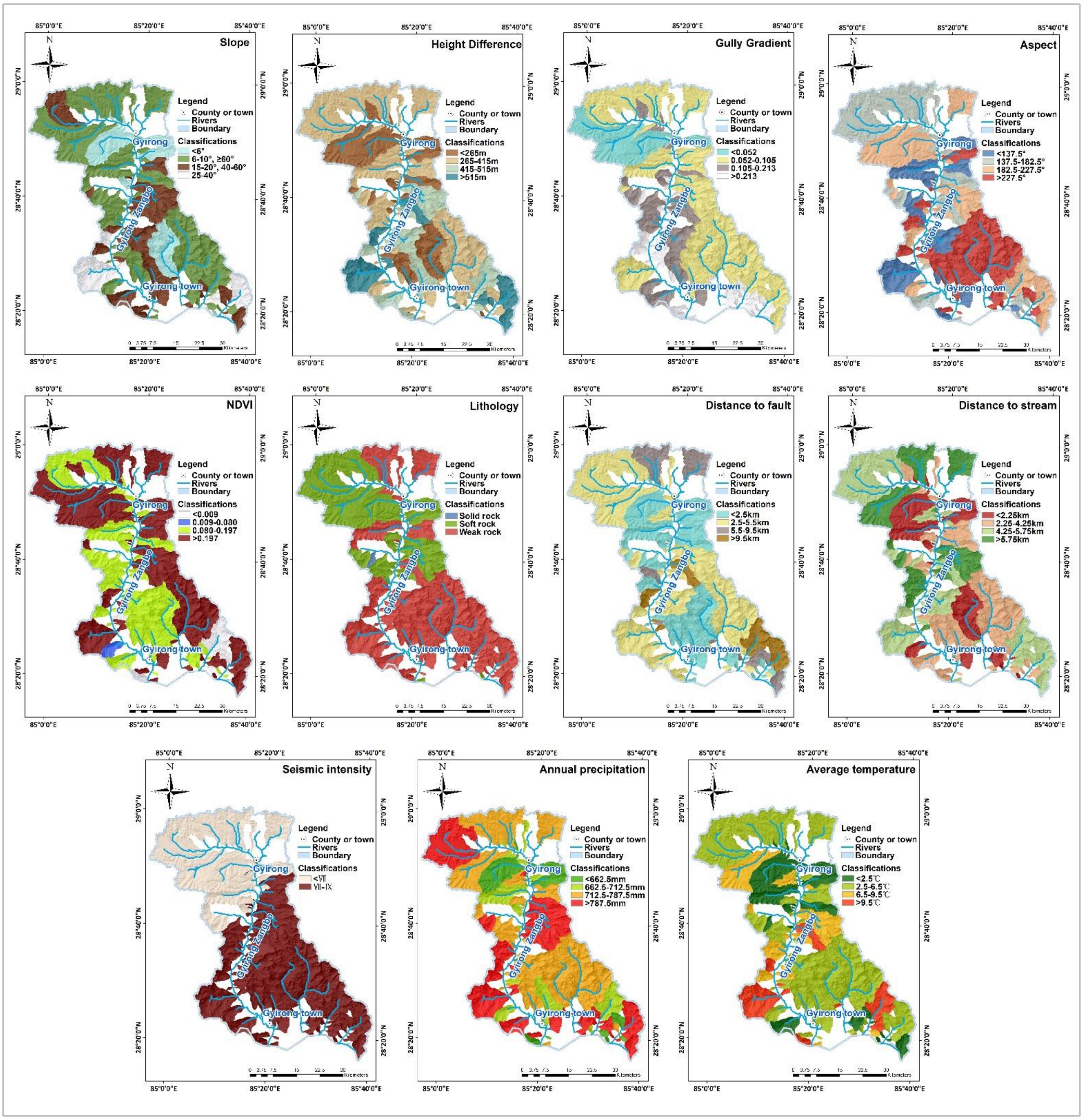

Figure 4. Maps showing causative factors in this study area with GIS

Slope gradient impact where the loose materials are going to settle and accumulate. A gully with steep slope could cause serious gravity erosion and increase soil erosion intensity due to the high flow kinetic energy. Height difference determines the potential energy of debris flow with the coupled effect of gully gradient. The south part of this area showed higher HD values than the north part based on Fig. 4, in which the catchment with highest HD was located in the southwestern part of this area, reaching 850m, followed by 659m and 650m in the southeastern part. The fourth highest HD value was 616m, which was close to the catchment with highest value. In consistent with the distribution of height difference, the catchment in south part presented steeper channels which provide higher potential energy for debris flow and also contributed to the increasing of magnitude of debris flow (Huang et al., 2020).

Aspect can affect the plant development because of the distributions of sunlight on slopes. A high vegetation coverage can benefit slope stability and decrease susceptibility of debris flow. The mechanical and hydrological effects provided by the vegetation roots, not only act as anchors to enhance the shear strength of soil, but also can decrease rainfall infiltration, moisture content of soil. NDVI is thus normally used as an index to describe the vegetation density based on remote sensing images:

$$NDVI = \frac{NIR - \mathrm{Red}}{NIR + \mathrm{Red}} \tag{1}$$

where NIR refers to the near-infrared band (0.85-0.88μm), and Red represents red-band (0.64-0.67μm). NDVI ranges from -1 to 1, in which negative value indicates the high reflective rate to visible light and positive value means the high vegetation coverage because of the high reflective to near-infrared. 0 value represents rocks or bare land.

Geological factors, such as lithology and distance to faults, reflect the development of geological structures, which further determines the magnitude of possible debris flow. The study area is mainly coved with weak and soft rocks, especially on the south part, which can benefit the generation of loose materials under the effects of faults movement. Based on the spatial distribution of distance to fault in Fig. 4, over 50% of catchments were distributed within a buffer of 5km with respect to seismic faults, most of which were distributed in the north part. Widespread compressed and deformed rocks were observed in north part of this area because of the fault movement during the field investigations. Another geological factor, seismic intensity, is a primary hazard which can directly trigger debris flow. This study area is all under the serve impacts of earthquake, such as the Gyirong earthquake (*Ms* = 5) in 2014 and Nepal earthquake (*Ms* = 8.1) in 2015. Distance to stream mainly focuses on the erosion of toe and riverbed since long-term fluvial erosion can cause slope instability, leading to landslide and triggering debris flow. There is no obvious spatial distribution pattern in this area when considering the factor, distance to stream. Apart from the geological factors which control the material supply of debris flow, meteorological factor, such as the annual precipitation, is the hydrodynamic condition for the formation of debris flow and also the most important triggering factor. According to the analysis result, the annual precipitation of 36% of this area was higher than 787.5mm, which was mainly distributed in the south part. The annual precipitation in 91% of area was higher than 600mm, which was the China annual precipitation value. Whilst another meteorological factor, average temperature, rises from north to south in this area with a decreasing of altitude. The lowest and the highest annual average temperatures were 0.13 ℃ in the north part and 12.91℃ in the south part, respectively. Most importantly, the average temperature can reflect the changes of glaciers since temperature rising can promote glacier melting and therefore lead to the increasing of water flow.

4.5 Methodology

In this paper, a hybrid machine learning model, CF-GA-SVM, was used to generate susceptibility maps to achieve timely and accurate warning of happening of glacier-related debris flow. After that, three models, LR, SVM and GA-SVM, were introduced to evaluate the performance of CF-GA-SVM model based on AUC and ACC values. The basic steps for generation of susceptibility maps by using CF-GA-SVM were presented as follows:

(1) Data preparation

Firstly, eleven causative factors were decided based on satellite images and field investigations, in which topographical, ecological and geological factors were used to demonstrate the potential materials supplied by landslide within catchments and meteorological factors was considered to reflect the triggering effect of waterflow caused by rainfall and glacier melting. Then, ArcGIS was used to analyse the spatial distribution of causal factors based on DEM (resolution of 30m).

(2) Calculation of *CF* values

CF values can be expected to improve the uncertainty and heterogeneity of input data (Pourghasemi et al., 2012a) in step (1) in this paper because the initiation of settled and accumulated materials by waterflow and mix to form debris flow is difficult to be clearly clarified. *CF* values were introduced to simplify the theses processes and initially present sensitivity of factors on happening of debris flow. It can be used to conduct sensibility calculations of the causative factors based on the equations as follows:

$$CF=\begin{cases}\dfrac{PP_{\mathrm{a}}\text{-}PP_{\mathrm{s}}}{PP_{\mathrm{a}}\left(1\text{-}PP_{\mathrm{s}}\right)} & PP_{\mathrm{a}}\geq PP_{\mathrm{s}}\\[2ex] \dfrac{PP_{\mathrm{a}}\text{-}PP_{\mathrm{s}}}{PP_{\mathrm{s}}\left(1\text{-}PP_{\mathrm{a}}\right)} & PP_{\mathrm{a}}<PP_{\mathrm{s}}\end{cases} \tag{2}$$

where PP_{a} represents the conditional probability of reclassification group *a* (Table 1) within a certain causative factor. PP_{a} can be obtained according to the ratio of number of hazards in group *a* to total area of group *a* (for debris flow, PP_{a} represents the ratio of debris flow area in group *a* to total area of group *a*). PP_s is the prior probability of the whole research region, which can be calculated based on the ratio of debris flow area to total area. The *CF* value varies from -1 to 1, in which $CF>0$ indicates the high occurring probability of debris flow, and $CF<0$ suggests the low occurring probability of debris flow. While the occurring probability cannot be identified if *CF* value is close to 0. The calculated *CF* values were presented in Table 1.

(3) Generation of optimal *C* and *g*

The calculated *CF* values in step (2) was substituted into genetic algorithm (GA) to generate optimal *C* and *g*. The main steps of GA were presented as:

a) Generation initialization: Decide the population quantity, the number of iterations, length of chromosome, and encode the parameters of SVM.

b) Fitness evaluation: Define a fitness function in order to evaluate the fitness of each individual.

c) Selection: The best individual with highly fitness will be chosen randomly by using roulette wheel mechanism.

d) Crossover: This step controls the production of new generations based on a certain crossover probability.

e) Mutation: This process changes the genes randomly according to mutation threshold value so that the diversity of individuals could be maintained, leading to better generations.

f) Output: Produce new generations and calculate the optimal *C* and *g*.

(4) Model training and validation:

The optimal *C* and *g* in step (3) were used to generate radial basis function and then develop improved SVM model, named GA-SVM. After that, the input data (*CF* values in step (2)) was randomly separated into training set and testing set by ratio of 7:3. The CF-GA-SVM model was further developed based on the training set, and then testing set was applied to test the prediction accuracy of this LR model. It should be noted that this is a widely used approach to evaluate the performance of model by resampling the original data when lacking independent testing dataset. Three strategies are usually and effectively used for model selection, accuracy assessment and hyperparameters tuning, including K-fold cross-validation, Monte-Carlo K-fold cross-validation and bootstrap resampling (Molinaro et al., 2005). In this paper, the training dataset was internally resampled using 10-folds cross-validations. Indeed, three purposes of 10-folds cross-validations in our studies were expected to achieve:

a) Collet effective information from the training dataset as much as possible.

b) Aoid local optimisation and overfitting.

c) Model validation, assessment and comparison.

Based on the purposes of this method, the basic process of K-fold cross-validation can be described as: the initial sample was firstly assigned to *K* partitions with equal size, in which one of the partitions was labeled as testing set and the other *K*-1 partitions were used for model training. Each of the partitions was labeled as testing set once and the average results was exported as the final output after *K* times cross-validation.

(5) Generation of susceptibility maps

The optimised SVM model (CF-GA-SVM) was finally used to map susceptibility along Gyirong Zangbo and divided this area into four different susceptibility levels, including very low, low, high and very high.

(6) Model assessment

Three models, LR, SVM and GA-SVM, were introduced to evaluate the performance of CF-GA-SVM model in mapping susceptibility of glacier-related debris flow based on receiver operating characteristics curve (ROC) and prediction accuracy (ACC). Finally, sensitivity was conducted to decide the robustness of this model, and the weight of causative factors was also calculated to evaluate the contribution rate of each factor. Additionally, the mathematical details of SVM and LR were presented in appendix A.

5 Result analysis

5.1 Glacier-related debris flow susceptibility maps

The DEM with a resolution of 30m in Gyirong was adopted to produce susceptibility maps of glacier-related debris flow. The target area was divided into 164 catchments with the area ranging from 0.13 km^2 to 233.93 km^2 containing 11 causative factors. The susceptibility levels for debris flow were classified into four categories (very low, low, high and very high) in this study by using LR, GA-SVM and CF-GA-SVM machine learning models based on remote sensing and field investigation data. Table 4 presented the percentage of susceptibility levels by four models, in which the regression model of LR was showed in equation (3) and (4).

$$P=\frac{e^{z}}{1+e^{z}} \tag{3}$$

$$Z = 0.155 - 2.230I_{Sl} + 2.244I_{HD} - 3.222I_{J} - 2.682I_{As} + 1.777I_{NDVI} + 9.942I_{Li} - 0.731I_{F} - 2.612I_{St} + 7.239I_{Ear} - 1.539I_{Pre} + 2.163I_{Temp}$$

(4)

Table 4. Percentage of susceptibility levels by four models

Models	LR	SVM	GA-SVM	CF-GA-SVM
Very low	30.5%	31.7%	33.5%	36.0%
Low	15.9%	14.0%	10.4%	3.1%
High	23.8%	12.8%	38.4%	19.5%
Very high	29.9%	41.5%	17.7%	41.5%

The predicted probability of LR in Table 4 ranged from 0.03481 to 0.99989, in which 0.0348 indicated the very low probability and 0.9998 meant the very high probability. Susceptibility levels were reclassified into four categories based on natural break point method which is based on clustering analysis. The results in Table 4 indicated that more than half of the study areas were on high susceptibility level, in which the extreme high level regions were mostly distributed in north part with a relatively gentle topographic fluctuation and small annual precipitation (Fig. 4). While the impacts of glaciers on happening of debris flow were not reflected by LR-generated susceptibility map. The susceptibility of SVM ranged from 0.08332 to 0.96, in which 0.08332 - 0.29426 belonged to very low level. The high susceptibility, ranging from 0.81332 to 0.96, showed same ratio with that of CF-GA-SVM, in which the ratio of very high susceptibility increased in the south part when compared with the LR-generated map due to the well-developed glaciers and sharp topographic fluctuation. Furthermore, the ratio of high and very high levels in the susceptibility map generated by GA-SVM increased when compared with LR and SVM. This was mostly due to the optimal C and g which improved the training model. The evolutive hyper parameters of GA-SVM and CF-GA-SVM were presented in Table 5..

Table 5. Parameters for SVM models

Models	Penalty factor (C)	Gamma (G)	Kernel function	Decision-function
SVM	Default	Default	rbf	ovr
GA-SVM	46.0423	2.3838	rbf	ovr
CF-GA-SVM	2.3667	4.0839	rbf	ovr

At last, the susceptibility of CF-GA-SVM ranged from 0.15603 to 0.87755, in which values between 0.4623 to 0.75274 belonged to high susceptibility and values between 0.74274 to 0.87755 were decided at a very high level. The results showed that the susceptibility level in large catchments (>100km^2) were generally high and very high, especially the large catchments in the south part, which might be due to the existence of glaciers and abundant rainfall (Fig.4). Therefore, the glacier-related debris flow in this area is the main type which caused severe damage to infrastructures and casualties. To make things even worse that the rising temperature will further stimulate glacier degradation and generate much more loose materials for potential glacier-related debris flow. To be more specific, the endlessly glacier melting flow as well as rainfall would constantly cause lateral erosion of gully, leading to slope instability and deposit settlement in the gully. The catchments in the north part that are highly prone to debris flow are mostly due to the crushed rocks (See Fig.1), snow melting (High altitude, see Fig.1) and rainfall (See Fig.4). faults movement (one main fault and eight secondary faults were identified in this area) caused compressional deformation of rocks, which destroyed the stability of the rock structure and thus generated large amounts of loose materials from both sides of gully (Tiranti et al., 2016). The developed gully under the effect of water erosion can contain much more materials and lay hidden risk for large-scale glacier-related debris flow. As a result, glacier-related debris flow will occur when the settled rocks and materials are initiated under the effect of water flow, sweeping the plants away and destroying farmland. Therefore, the glacier-related debris flow can not only cause losses of properties and human lives, but also destroy the original ecological environment.

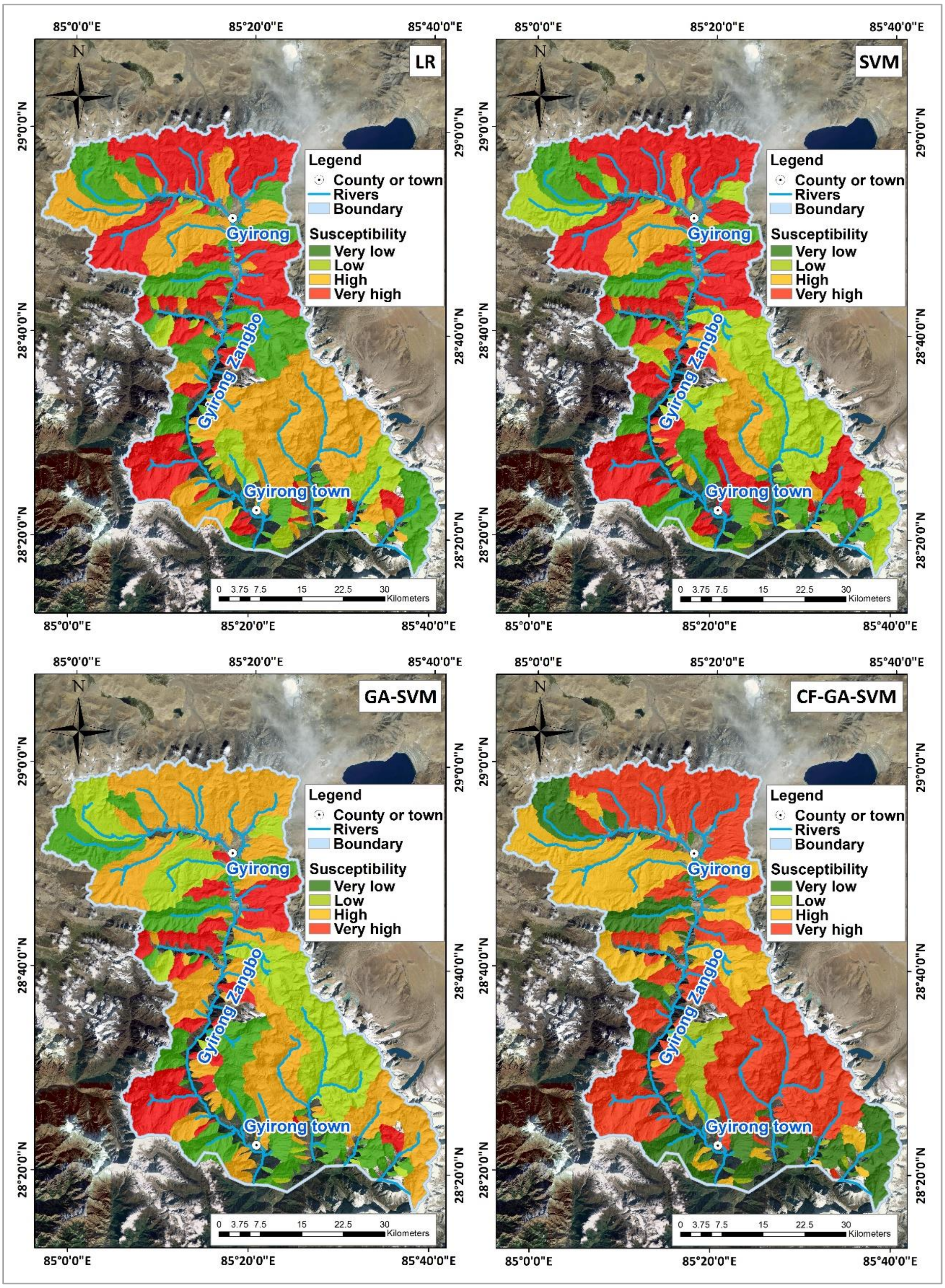

Figure 5. Glacier-related debris flow susceptibility maps produced by four models

5.2 Model assessment

(1) Prediction accuracy

The accuracy of the classification of the classifiers and the prediction accuracy were presented in Table 6. The CF-GA-SVM model showed the highest AUC value, reaching to 0.945. And the AUC of GA-SVM was lower than CF-GA-SVM, which was 0.888. SVM represented the third highest AUC value, 0.856, which was higher than that of LR method, 0.776. In addition, the CF-GA-SVM method showed the highest prediction accuracy, which was mainly due to the application of CF values and therefore avoid noise data. The integration of certainty factors with GA-SVM not only can enhance sensibility of causative factors, but also benefit the analysis of susceptibility but with a longer duration. SVM performed better than LR since SVM was more suitable for analysis of small set of samples (Huang et al., 2018), but LR method could not provide a comprehensive interpretation for environmental factors, especially for the important area with limited catchments division. GA-SVM with optimal parameters *C* and gamma showed higher precision than SVM. This was because appropriate penalty factor *C* and gamma could optimise fitting model to avoid over-fitting so that the generalization ability of model can be further strengthened.

Table 6. Obtained AUC and ACC values for four models

Models	AUC	ACC	Time duration / s
LR	0.776	0.733	3.68
SVM	0.856	0.750	2.78
GA-SVM	0.888	0.766	8.12
CF-GA-SVM	0.945	0.800	8.25

As indicated in Table 6 that CF-GA-SVM model performed best in model fitting and susceptibility prediction. More than half of the areas in this region were under high risk to glacier-related debris flow, reaching to 61%. Therefore, further distribution analysis of predicted debris flow by using CF-GA-SVM was conducted to depict the various state that causative factors might be in through the presentation of radar maps in Fig. 6.

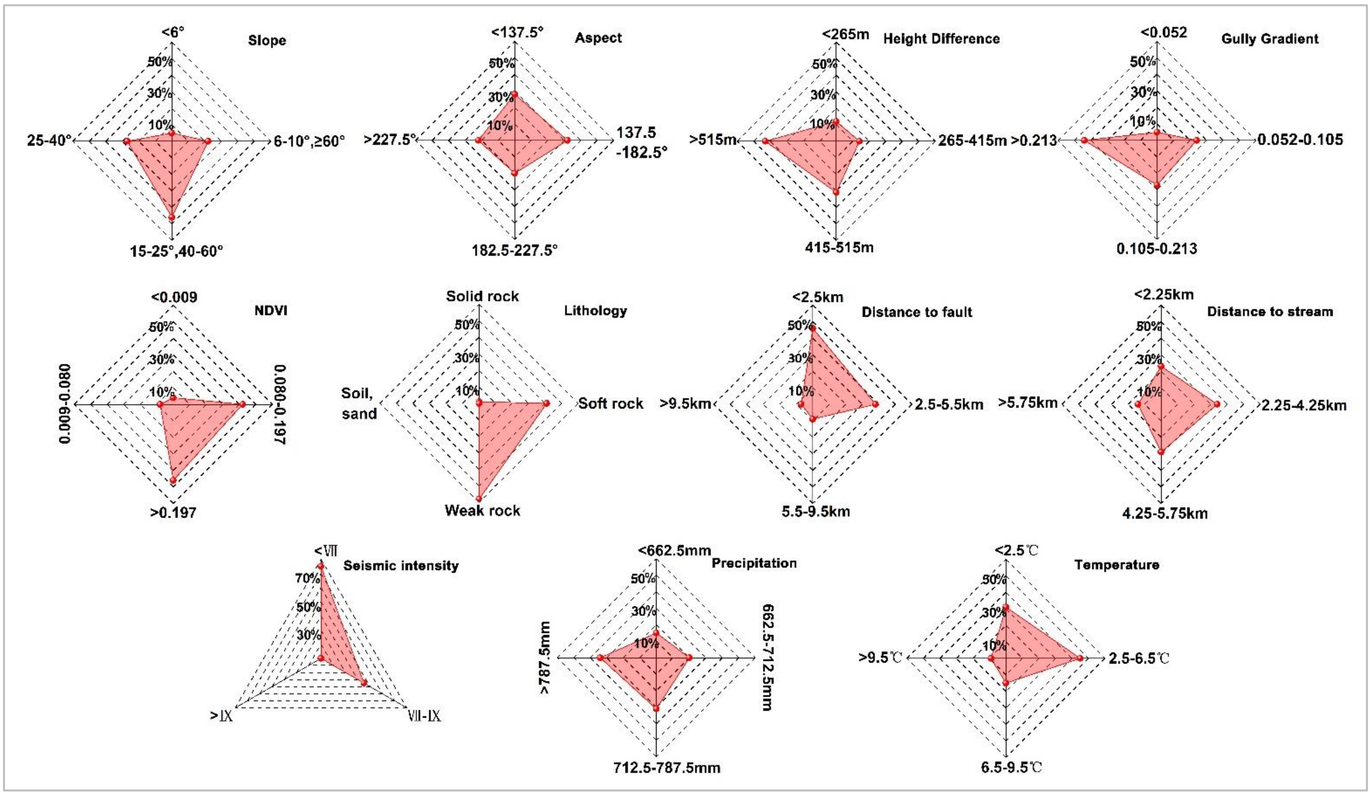

Figure 6. Distribution of debris flow in different classes of factors

From the maps in Fig. 6, we can observe the distribution characteristics of glacier-related debris flow in this area. Firstly, the catchments prone to glacier-related debris flow were mainly distributed in the areas with slopes ranging from 6 to 10° and 15 to 25°, which indicated that gentle slopes can contribute to the settlement and accumulation of loose materials as material source for glacier-related debris flow. These materials will be initiated under the stimulation of high-intense short-term rainfalls or long-term moderate rainfalls. Aspect indicated that approximately 32% of high risk areas were distributed at the aspect of South and Southwest, which was, 137.5°<As<182.5°. The other two topographic factors, *HD* and *J*, served as the indicators to reflect the potential power of debris flow through estimation of glacier-related debris flow volume and peak discharge. 43% of catchments that were distributed in areas with *HD* > 515m and *J*>0.213 were at a high risk to glacier-related debris flow. In addition to the topographic factors, the only ecological factor, *NDVI*, showed that glacier-related debris flow was mostly distributed in areas with better vegetation coverages. This is due to the special geographic condition in Himalaya mountains mentioned in the introduction section. For the geological factors, the high susceptibility areas to glacier-related debris flow were mainly distributed in the weak rock areas, which might be caused by the impacts from the faults and earthquakes. The exposed rocks during our investigations were observed to present severe deformation in this area due to the fault movements and seismic shaking in the past and therefore caused a large quantity of broken rocks on both sides of the roads along Gyirong Zangbo. The last geological factor, *St*, illustrated that the mean distance to Gyirong Zangbo for 57% of high risk areas was involved in the buffer zone with a radius of 4.25km. Finally, the two meteorological factors, *Pre* and *Temp*, serving as the excitation conditions, showed that 34% of high risk areas were identified with *Pre* > 787.5mm, and 45% of debris flow-prone areas with a temperature ranging from 2.5-6.5℃.

(2) Sensitivity

Based on the results in Table 5, the CF-GA-SVM model performed better than the other three models in model fitting and prediction accuracy. Therefore, further investigations were conducted here to investigate the robustness of the optimised SVM model. One factor was excluded from model development at a time. AUC and ACC (Table 7) were then calculated to conduct sensitivity analysis. The analysis results showed that all of the selected factors posed positive impacts in model fitting since there was no AUC value larger than 0.945 in Table 7. Whilst the factor, Precipitation, representing the lowest AUC value can indicate its importance in generating appropriate CF-GA-SVM model. In accordance with the ACC value that precipitation also presented a low value. The northward warm air from Indian Ocean reaches Gyirong area but is blocked by Himalaya mountains with the mountain peaks almost into the cloud. As a result, a lot of rainfalls pour in this area, although benefited the plant growth, triggered frequent debris flow, especially in the north part of this study area. In conclusion, the produced CF-GA-SVM model can adapt to a wide range of selected factors with a strong robustness performance.

Table 7. AUC and ACC values of model without excluded factors

Excluded factors	AUC	ACC
Annual precipitation	0.871	0.750
Aspect	0.908	0.733
Distance to Stream	0.910	0.766
NDVI	0.917	0.750
Slope	0.917	0.766
Height Difference	0.927	0.783
Gully Gradient	0.932	0.733
Distance to Fault	0.934	0.733
Seismic Intensity	0.940	0.775
Lithology	0.940	0.717
Average annual temperature	0.941	0.766
Pre + Temperature	0.917	0.700
No factor excluded	0.945	0.800

5.3 Weight of factors

Sensitivity analysis can reflect the robustness of the model and describe the importance of selected factors that influence the model fitting and the prediction accuracies on different levels. However, detailed weight of each factor in this model should be further investigated through factors analysis for further model opmisation. Table 8 presented the weight of all the factors. It can be seen that gully gradient ranked the first place with biggest weight value, followed by distance to stream. The values of these two factors were all larger than 0.1, which indicated the importance of gully gradient and distance to stream in the generation of susceptibility maps. Next to the gully gradient and distance to stream, the other four factors, geology-related factor, distance to fault, edaphic-related factors, aspect and *NDVI*, weather-related factor, average annual temperature, also presented high weight values, ranging from 0.0917079 to 0.0983527. This was consistent with the weight ranking of factors in (Zhang et al., 2019) in which aspect, *NDVI* occupied high importance places among the 15 causative factors. Finally, the rest factors, precipitation, slope, seismic intensity, lithology and height difference showed relatively low weight values, in which the calculated weight of precipitation did not achieve the expected rank. This might be due to the coupled effect of the glacier melting and rainfall in this area that triggered debris flow, underplaying the weight of precipitation. Therefore, studies of debris flow in this area should be further divided into rainfall-triggered and glacier-related debris flow in future studies to better reflect the triggering conditions.

Table 8. Weight of the causative factors

Factors	Weight	Ranking
Distance to Stream	0.1001002	2
Slope	0.0882511	8
Temperature	0.0917079	6
Distance to Fault	0.0983527	3
Precipitation	0.0877273	7
Seismic Intensity	0.0854788	9
Aspect	0.0967044	4
Groove Gradient	0.1005165	1
Lithology	0.0821672	10
NDVI	0.0930783	5
Height Difference	0.0759155	11

6 Discussion

The studies in this paper mapped the susceptibility of glacier-related debris flow along

Gyirong Zangbo river and also provided detailed analysis of the warning precision, factor sensitivity and weight of factors to further optimised the model. The triggering conditions of glacier-related debris flow in this study area were far more complicated than other regions (Tang et al., 2009; Nikolopoulos et al., 2014; Marra et al., 2014; Nikolopoulos et al., 2015; Bel et la., 2017). In order to further investigate the impacts from the complex triggering environment for subsequent studies of risk assessment (Lugon et al., 2010), the distribution characteristics of a total of 100 debris flow with the consideration of catchment areas, happening frequency and size were presented in Fig. 7.

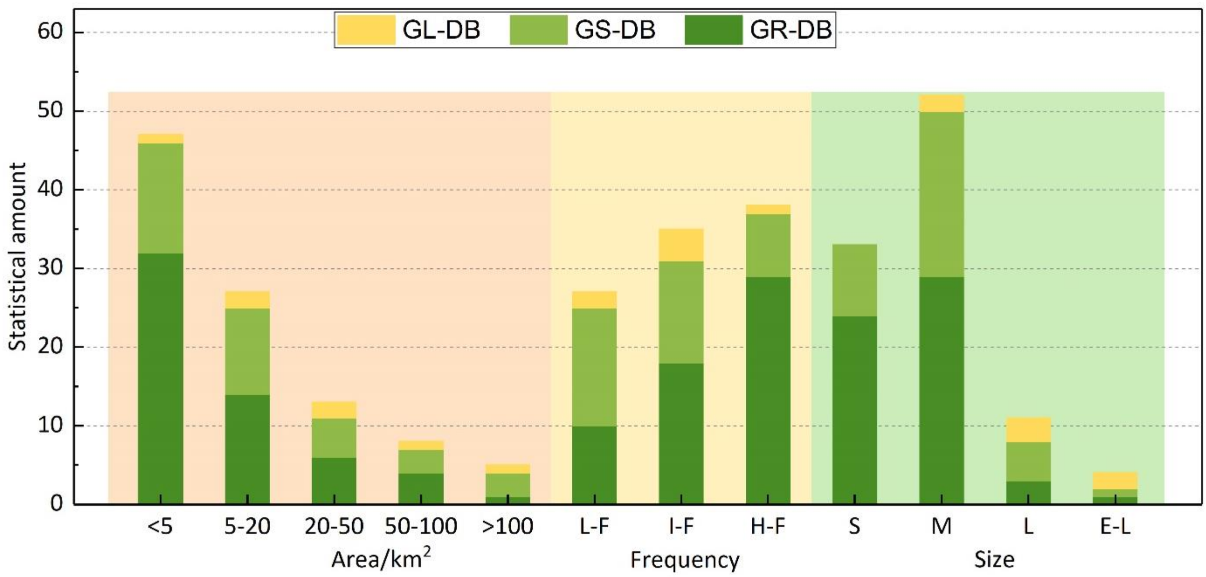

Figure 7. Relationships between four types of debris flow and different classifications of catchment area, debris flow frequency and debris flow volume

('GL-DB', 'GS-DB', 'GR-DB' represents 'Debris Flow induced by break of Glacial Lake', 'Glacier and Snow melting Debris Flow', 'Glacier Rainfall Debris Flow' respectively. 'L-F', 'I-F', 'H-F' represents 'Low-Frequency', 'Intermediate-Frequency' and 'High-Frequency' respectively. 'S', 'M', 'L', 'E-L' represents 'Small', 'Medium', 'Large' and 'Extremely Large' respectively)

The areas of debris flow catchments were classified into five levels, <5km^2, 5-20km^2, 20-

50km^2, 50-100km^2 and >100km^2. As indicated in Fig.7, the number of GR-DB was 57, in which there were 32 of debris flows with areas smaller than 5km^2 and 14 of debris flows with areas ranging from 5 to 20km^2. While there was only one identified debris flow with area larger than 100km^2. Additionally, a total of 36 debris flows were found to be triggered by the coupled effects of glacier and snow melting, in which the number of areas smaller than 5km^2 is 14. Whilst there were 11 of debris flow with areas ranging from 5 to 20km^2 and 5 of debris flow with areas ranging from 20 to 50km^2. Moreover, the same number of debris flows, 3, is found in both areas larger than 100km^2 and areas ranging from 50 to 100km^2. Finally, GL-DB presented a nearly even distribution with 1, 2, 2, 1, 1, debris flow observed in <5km^2, 5-20km^2, 20-50km^2, 50-100km^2 and >100km^2, respectively. Overall, 47% of debris flows were found within an area smaller than 5km^2, followed by areas ranging from 5 to 20 km^2 and 20 to 50km^2, 27% and 13%, respectively. The smallest proportion belonged to areas larger than 100km^2, 5%.

The happening frequency of debris flow in Tibet was affected by climate change, seismic activity and accumulation period of loose materials (Beniston et al., 2004; Fuhrer et al., 2006; Zhao et al., 2020). Therefore, the studies of frequency for glacier-related debris flows are essential for effective mitigation and prevention measures. Based on the investigation results relating to accumulation situations in channels and alluvial fan, three frequency levels were decided qualitatively to classify the total of 100 debris flow into L-F, I-F and H-F. As indicated in Fig. 8, L-F indicated that no obvious new material was found in the catchment mouth and the accumulation fan was covered by bryophyte. I-F indicated that there were no new debris flow materials observed in the accumulation fan and a few bryophytes can be seen in partial regions of alluvial fan. H-F indicated that obviously new materials can be found in accumulation fan.

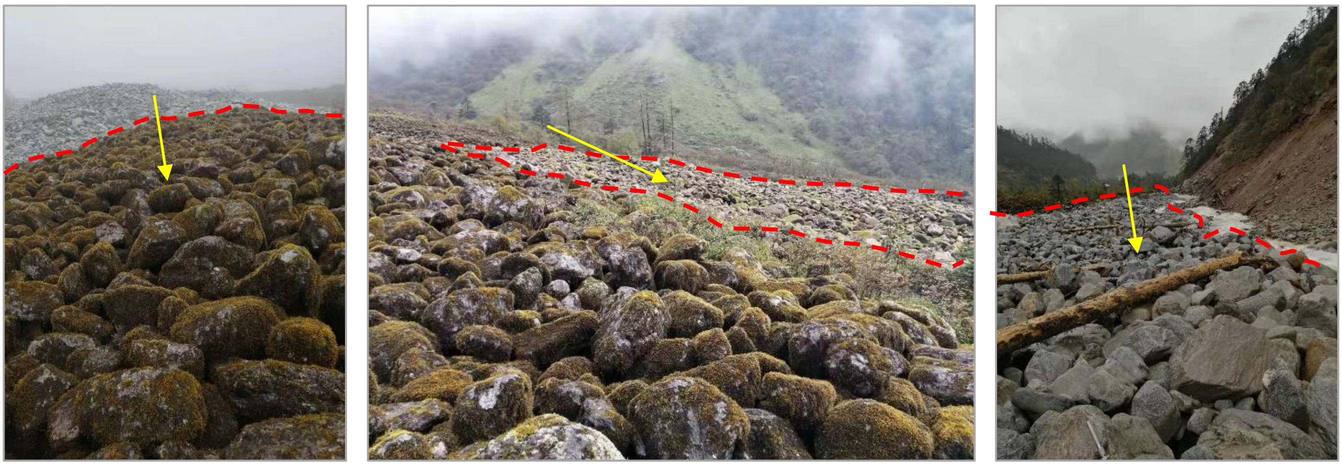

Figure 8. Classification of frequency for happening of debris flow. (a)L-F. (b) I-F. (c) H-F.

As illustrated in Fig.7, 36% of debris flows were occurred frequently followed by I-F and L-F with 35, 27 percentage, respectively. This meant that most parts of this area were under high risk to debris flow which was also consistent with the susceptibility analysis in § 5.1. Additionally, GR-DB with H-F presented the highest ratio, 29%, followed by GR-DB with I-F, GS-DB with L-F and GS-DB with I-F, 18%, 15%, 13%, respectively. GL-DB with H-F presented the lowest one, only 1%.

Susceptibility analysis can provide warning for local resident about the potential risks for properties and their lives, and volume calculation can further benefit the decision-makers in measurements setting (Rickenmann, 1999). The volume Q_c of the identified 100 debris flow can be estimated by using following equations (5), (6), (7) (Cui et al., 2013) based on the length, width and depth of accumulation fan:

$$Q_t = 152.97 Q_c^{1.266} \tag{5}$$

$$Q_c = (1+\phi) Q_B D_U \tag{6}$$

$$\phi = (\gamma_c - \gamma_w) / (\gamma_s - \gamma_c) \tag{7}$$

where Q_t (m^3) represents the volume of a debris flow, Q_c (m^3/s) is the peak discharge of a debris flow (Fei and Shu, 2004), Q_B represents the peak discharge of flood (m^3/s), D_U is the blockage coefficient, γ_c (kg/m^3), γ_w (kg/m^3), γ_s (kg/m^3) represent the density of debris flow, water density and solid materials density, respectively.

The classification principle in Table 9 for debris flow size is based on (Jakob, 2005). According to investigation results, 4 levels were further recategorised:

Table 9. Size classification for debris flow

Size class	S	M	L	E-L
Q_c (m^3)	$<10^4$	10^4-10^5	10^5-10^6	$>10^6$

The categorisation of the 100 debris flow was presented in the right part of Fig.7. 33% of debris flow (100 in total) were S size, in which 73% was GR-DB and 27% was GS-DB. Then M size presented the highest ratio, reaching to 52 percent. The GR-DB in this class occupied 56%, followed by GS-DB, 40% and GL-DB, 4%. Moreover, a total of 11 and 4 were observed in the L and E-L class, respectively. The GS-DB in the L class presented the highest one, 5 debris flows were found. Both GR-DB and GL-DB in L class were 3.

Overall, the classification of glacier-related debris flow types aimed at the improvement of recognition of debris flow in this area. The catchments with large areas which were larger than 20km^2 can contain larger amounts of materials (Zhang, 2016) and therefore will lead to large-scale debris flows (Xu, 1988) in a high probability. They were mostly triggered by glacier melting and rainfall or outburst of ice lake. The hybrid model is well applicable for GR-DB since temperature and precipitation factors can reflect the triggering conditions and provide accurate susceptibility evaluation.

However, the limitations appeared in happening warning of GL-DB because the triggering condition of GL-DB depended on the capacity of ice lake and the shear strength of terminal moraine (Allen et al., 2016). Debris flow would occur when the flow accumulation exceeds the capacity of glacial lake, or the water seepage causes failure of terminal moraine. In order to achieve an accurate susceptibility warning, the geometric dimension of glacial lake needs to be considered as a causative factor. The calculations of water flow generated by glacier melting and rainfall should be conducted to estimate the possible date of peak capacity of the glacial lake based on the environmental changes. Moreover, the physical and mechanical properties of moraine around the glacial lake should be obtained to estimate the shear strength of terminal moraine. Overall, an accurate warning of debris flow caused by the outburst of glacial lake is more difficult to achieve than the rainfall-triggered debris flow and other types of glacier-related debris flow. The difficulties came from the complexity in the determinations of the controlling factors and the uncertainties in predicting the outburst of glacial lake. Temperature and precipitation are not direct controlling factors which is the limitation in the current hybrid model to conduct an accurate susceptibility analysis for catchments under the risk of outburst of glacial lake. Therefore, catchment with a glacial lake inside required additional attentions. Overall, the proposed model can achieve susceptibility warning with relatively high accuracy for debris flows triggered by glacier and snow melting and rainfall. In general, the classifications of areas, frequency and runout volume can benefit the development of prevention measurements and help to further identify the possible limitations of proposed model.

7 Conclusion

LR and SVM algorithms are widely used machine learning methods for binary classification when producing debris flow susceptibility maps. The results in this study showed that four models performed well in predicting the happening of glacier-related debris flow in Gyirong (along the

China-Nepal Economical Corridor) where debris flow frequently occurred. In specifically, SVM algorithm not only showed superiority in solving small samples when compared with LR method, but also provided a higher accurate prediction accuracy. In order to optimise SVM to a achieve higher predicti*on accuracy (ACC = 0.766 for GA-SVM, ACC = 0.75 for single SVM), GA-SVM mode*l was developed with the integration of optimal parameters *C* and *gamma* by using genetic algorithm which can benefit the model fitting (*AUC=0.888 for GA-SVM model, AUC=0.856 for single SVM*) to avoid overfitting and help to produce appropriate model with higher generalization ability. Based on the optimised GA-SVM model, further improvement was considered to integrate *CF* values to optimise the analysis to better reflect the susceptible of causative factors to the occurrence of glacier-related debris flow. As a result, CF-GA-SVM model showed the highest prediction accuracy (ACC = 0.8) and best fitting curves (AUC = 0.945) among the four models. Finally, sensitivity analysis was conducted to investigate the robustness of the CF-GA-SVM model. The results showed that the CF-GA-SVM model can achieve compatibility of various factors combination, the increasing of causal factors could improve the model fitting and prediction accuracy. Therefore, CF-GA-SVM model is suitable for generation of susceptibility maps in a areas with small samples. However, further investigations need to be conducted to test if this model is applicative in a large area.

Appropriate susceptibility map with high accuracy is important for debris flow mitigation. The produced susceptibility map for glacier-related debris flow in this study with the consideration of environmental changes can be an effective supplement of disaster warning for residents and also guidance for ecological remediation. Debris flow can pose direct impacts on environmental damage, disruption of water supply systems and devaluation of fisheries (Jakob et al., 2005).

Therefore, susceptibility warning with high accuracy is indispensable for areas where debris flow takes place frequently. The proposed machine learning model can adapt to the real-time changes of meteorologic factors and achieve accurate disaster warning. Nevertheless, attentions still need to be paid when applying this model for mapping susceptibility because performance may vary in different study regions. Overall, the study area in this paper is a significant component of China-Panel Economical Corridor, which plays an important role in connecting trade between China and Nepal, and even to other South Asian Countries. The presented results can provide scientific support to the local government to reduce losses of human lives and properties.

Acknowledgements:

This work was financially supported by the Key Laboratory of Mountain Hazards and Earth Surface Processes, Chinese Academy of Sciences; the National Nature Science Foundation of China [grant numbers 51978533]; European Union’s Horizon 2020 research and innovation program Marie Skłodowska–Curie Actions Research and Innovation Staff Exchange (RISE) under grant agreement [grant number 778360]; the International Partnership Program of Chinese Academy of Sciences [131551KYSB20180042]; the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) [grant number 2019QZKK0902]; the Natural Science Foundation of Shandong Province [grant number ZR2016EEM40].

References

Allen, S.K., Rastner, P., Arora, M., et al., 2016. Lake outburst and debris flow disaster at Kedarnath, June 2013: hydrometeorological triggering and topographic predisposition. Landslides. 13(6), 1479-1491. https://doi.org/10.1007/s10346-015-0584-3.

Arabameri, A., Pradhan, B., Rezaei, K., 2019. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. Journal of environmental management. 232, 928-942. https://doi.org/10.1016/j.jenvman.2018.11.110.

Bel, C., Liébault, F., Navratil, O., et al., 2017. Rainfall control of debris-flow triggering in the Réal Torrent, Southern French Prealps. Geomorphology. 291, 17-32. https://doi.org/10.1016/j.geomorph.2016.04.004.

Beniston, M., Stephenson, D.B., 2004. Extreme climatic events and their evolution under changing climatic conditions. Global and Planetary. 44(1-4), 1-9. https://doi.org/10.1016/j.gloplacha.2004.06.001.

Breien, H., De Blasio, F. V., Elverhøi, A., et al., 2008. Erosion and morphology of a debris flow caused by a glacial lake outburst flood, Western Norway. Landslides. 5(3), 271-280. https://doi.org/10.1007/s10346-008-0118-3.

Cama, M., Lombardo, L., Conoscenti, C., et al., 2017. Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). Geomorphology. 288, 52-65. https://doi.org/10.1016/j.geomorph.2017.03.025.

Cao, J., Zhang, Z., Du, J., et al., 2020. Multi-geohazards susceptibility mapping based on machine learning—a case study in Jiuzhaigou, China. Natural Hazards. 102(3), 851-871. https://doi.org/10.1007/s11069-020-03927-8.

Chen, J., Li, Y., Zhou, W., et al., 2017. Debris-flow susceptibility assessment model and its application in semiarid mountainous areas of the Southeastern Tibetan Plateau. Natural

Hazards Review. 18(2), 05016005. https://doi.org/10.1061/(asce)nh.1527-6996.0000229.

Chen, X.C., Chen, H., You, Y., et al., 2015. Susceptibility assessment of debris flows using the analytic hierarchy process method - A case study in Subao river valley, China. Journal of Rock Mechanics and Geotechnical Engineering. 7(4), 404-410. https://doi.org/10.1016/j.jrmge.2015.04.003.

Chen, N.S., Zhou, H.B., Hu, G.S., 2011. Development of rules of debris flow under the influence of climate change in Nyingchi. Advances in Climate Change Research. 7(6), 412-417.

Cui, P, et al., Atlas of Mountain Hazards and Soil Erosion in the Upper Yangtze. Beijing, China Science Publishing, 2014.

Cui, P., Xiang, L., Zou, Q., 2013. Risk assessment of highways affected by debris flow in Wenchuan earthquake area. Journal of Mountain Science. 10(2), 173-189. https://doi.org/10.1007/s11629-013-2575-y.

Devkota, K.C., Regmi, A.D., Pourghasemi, H. R., et al., 2013. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. Natural Hazards. 65(1), 135-165. https://doi.org/10.1007/s11069-012-0347-6.

Ding, M., Gao, Z., Huang, T., et al. 2020. The hazard assessment of glacial lake debris flow: A case study on Dongcuoqu, Luolong County, Tibet. In IOP Conference Series: Earth and Environmental Science. IOP Publishing.

Dormann, C.F., Elith, J., Bacher, S., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 36, 27-46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Elkadiri, R., Sultan, M., Youssef, A.M., et al., 2014. A remote sensing-based approach for debris-flow susceptibility assessment using artificial neural networks and logistic regression modeling. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7(12), 4818-4835. https://doi.org/10.1109/jstars.2014.2337273.

Erokhin, S. A., Zaginaev, V. V., Meleshko, A. A., et al. 2018. Debris flows triggered from non-stationary glacier lake outbursts: the case of the Teztor Lake complex (Northern Tian Shan, Kyrgyzstan). Landslides. 15(1), 83-98. https://doi.org/10.1007/s10346-017-0862-3.

Feizizadeh, B., Roodposhti, M.S., Blaschke, T., et al., 2017. Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping. Arabian Journal of Geosciences. 10(5), 1-13. https://doi.org/10.1007/s12517-017-2918-z.

Fei, X.J., Shu, A.P., 2004. Movement Mechanism and Disaster Control for Debris Flow. Beijing: Tsinghua University Press. 114-119. (in Chinese).

Fuhrer, J., Beniston, M., Fischlin, A., et al., 2006. Climate risks and their impact on agriculture and forests in Switzerland. Climate Variability, Predictability and Climate Risks. 79, 79-102. https://doi.org/10.1007/978-1-4020-5714-4_5.

Gayen, A., Pourghasemi, H.R., Saha, S., et al., 2019. Gully erosion susceptibility assessment and management of hazard prone areas in India using different machine learning algorithms. Science of the Total Environment. 668, 124-138. https://doi.org/10.1016/j.scitotenv.2019.02.436.

Heckerman D., 1986. Probabilistic interpretations for mycin's certainty factors. Machine Intelligence & Pattern Recognition. 4, 167-196.

Heckmann, T., Gegg, K., Gegg, A., et al., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Natural Hazards and Earth System Sciences. 14(2), 259-278. https://doi.org/10.5194/nhess-14-259-2014.

Huang, J., Hales, T.C., Huang, R.Q., et al., 2020. A hybrid machine-learning model to estimate potential debris-flow volumes. Geomorphology. 367, 107333. https://doi.org/10.1016/j.geomorph.2020.107333.

Huang, Y., Zhao, L. 2018. Review on landslide susceptibility mapping using support vector machines. Catena. 165, 520-529. https://doi.org/10.1016/j.catena.2018.03.003.

Jakob, M., 2005. A size classification for debris flows. Engineering Geology. 79(3-4), 151-161. https://doi.org/10.1016/j.enggeo.2005.01.006.

Jakob, M., Hungr, O., Jakob, D. M., 2005. Debris-flow hazards and related phenomena (Vol. 739). Berlin: Springer.

Jackson, L.E., Hungr, Jr.O., Gardner, J.S., 1989. Cathedral mountain debris flow, Canada. International Association of Engineering Geology. 35-53. https://doi.org/10.1007/BF02590340.

Jomelli, V., Pavlova, I., Eckert, N., et al., 2015. A new hierarchical Bayesian approach to analyse environmental and climatic influences on debris flow occurrence. Geomorphology. 250, 407-421. https://doi.org/10.1016/j.geomorph.2015.05.022.

Kadavi, P.R., Lee, C.W., Lee, S., 2018. Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping. Remote Sensing. 10(8), 1252. https://doi.org/10.3390/rs10081252.

Kang, S.H., Lee, S.R., 2018. Debris flow susceptibility assessment based on an empirical approach in the central region of South Korea. Geomorphology. 308, 1-12.

https://doi.org/10.1016/j.geomorph.2018.01.025

Lameski, P., Zdravevski, E., Mingov, R., 2015. SVM Parameter Tuning with Grid Search and Its Impact on Reduction of Model Over-fitting. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. 9437, 464-474. https://doi.org/10.1007/978-3-319-25783-9_41.

Lee, S., Sambath, T., 2006. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Environmental Geology. 50(6), 847-855. https://doi.org/10.1007/s00254-006-0256-7.

Li, W.W., Xing, X.X., Liu, F., et al., 2014. Optimizing the Hyper-parameters for SVM by Combining Evolution Strategies with a Grid Search. International Journal of Multimedia and Ubiquitous Engineering. 9(11), 181-188. https://doi.org/10.1007/978-3-540-37256-1_87.

Liang, Z., Wang, C. M., Zhang, Z. M., 2020. A comparison of statistical and machine learning methods for debris flow susceptibility mapping. Stochastic Environmental Research and Risk Assessment. 34(11), 1887-1907. https://doi.org/10.1007/s00477-020-01851-8.

Lin, G.F., Chang, M.J., Huang, Y.C., et al., 2017. Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression. Engineering Geology. 224, 62-74. https://doi.org/10.1016/j.enggeo.2017.05.009.

Liu, R., Liu, E., Yang, J., et al., 2006. Optimizing the Hyper-parameters for SVM by Combining Evolution Strategies with a Grid Search. Intelligent Control and Automation. 344, 712-721. https://doi.org/10.1007/978-3-540-37256-1_87.

Lombardo, L., Cama, M., Conoscenti, C., et al., 2015. Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in Messina (Sicily, southern Italy). Natural Hazards. 79(3), 1621-1648. https://doi.org/10.1007/s11069-015-1915-3.

Lugon, R., Stoffel, M., 2010. Rock-glacier dynamics and magnitude–frequency relations of debris flow in a high-elevation watershed: Ritigraben, Swiss Alps. Global and Planetary Change. 73(3-4), 202-210. https://doi.org/10.1016/j.gloplacha.2010.06.004.

Marra, F., Destro, E., Nikolopoulos, E.I., 2014. Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. Geomorphology. 221, 286-297. https://doi.org/10.1016/j.geomorph.2014.06.015.

Meng, Q., Miao, F., Zhen, J., et al., 2016. GIS-based landslide susceptibility mapping with logistic regression, analytical hierarchy process, and combined fuzzy and support vector machine methods: a case study from Wolong Giant Panda Natural Reserve, China. Bulletin of Engineering Geology and the Environment. 75, 923-944. https://doi.org/10.1007/s10064-015-0786-x.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing. 42(8), 1778-1790. https://doi.org/10.1109/TGRS.2004.831865.

Mojaddadi, H., Pradhan, B., Nampak, H., et al., 2017. Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. Geomatics, Natural Hazards and Risk. 8(2), 1080-1102. https://doi.org/10.1080/19475705.2017.1294113.

Molinaro., A.M., Simon., R., Pfeiffer., R.M. 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics. 21(15), 3301-3307.

Muñoz, R., Gonzales, C., Price, K., et al. 2016. Managing glacier related risks disaster in the Chucchún catchment, Cordillera Blanca, Peru. In Climate Change Adaptation Strategies–An Upstream-downstream Perspective. Springer, Cham, 59-78. https://doi.org/10.1007/978-3-319-40773-9_4.

Nikolopoulos, E.I., Crema, S., Marchi, L., et al., 2014. Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. Geomorphology. 221, 286-297. https://doi.org/10.1016/j.geomorph.2014.06.015.

Nikolopoulos, E.I., Borga, M., Creutin, J.D., 2015. Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. Geomorphology. 243, 40-50. https://doi.org/10.1016/j.geomorph.2015.04.028.

O'Gorman, P. A., 2015. Precipitation extremes under climate change. Current climate change reports. 1(2), 49-59. https://doi.org/10.1007/s40641-015-0009-3.

Paudel, B., Fall, M., Daneshfar, B. 2020. GIS-based assessment of debris flow hazards in Kulekhani Watershed, Nepal. Natural Hazards. 101, 143-172. https://doi.org/10.1007/s11069-020-03867-3

Petrakov, D.A., Krylenko, I.V., Chernomorets, S.S., et al., 2007. Debris flow hazard of glacial lakes in the Central Caucasus. In: Chen CL, Major O (eds) Debris-flow hazards mitigation: mechanics, prediction, and assessment. Millpress, Rotterdam. 703-714.

Perov, V., Chernomorets, S., Budarine, O., et al., 2017. Debris flow hazards for mountain regions of Russia: regional features and key events. Natural Hazards. 88(1), 199-235. https://doi.org/10.1007/s11069-017-2841-3.

Pisner, D. A., Schnyer, D. M. 2020. Support vector machine. In Machine Learning. Academic

Press. 101-121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., et al., 2012(a). Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arabian Journal Geosciences. 6, 2351-2365. https://doi.org/10.1007/s12517-012-0532-7.

Perutz, M.F., 1953. The flow of glacier. Nature. November 21, No. 4396.

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., et al.,2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology. 298,118-137. https://doi.org/10.1016/j.geomorph.2017.09.006.

Rickenmann, D., 1999. Empirical relationships for debris flows. Natural hazards. 19(1), 47-77. https://doi.org/10.1023/A:1008064220727.

Shortliffe, E.H., 1975. A model of inexact reasoning in medicine. Mathematical Biosciences. 23(3), 351-379. https://doi.org/10.1016/0025-5564(75)90047-4.

Stoffel, M., Mendlik, T., Schneuwly-Bollschweiler, M., et al., 2014. Possible impacts of climate change on debris-flow activity in the Swiss Alps. Climatic Change. 122(1), 141-155. https://doi.org/10.1007/s10584-013-0993-z.

Stoffel, M., Tiranti, D., Huggel, C., 2014. Climate change impacts on mass movements - Case studies from the European Alps. Science of Total Environment. 493, 1255-1266. https://doi.org/10.1016/j.scitotenv.2014.02.102.

Suthaharan, S. Support vector machine. In Machine learning models and algorithms for big data classification (pp. 207-235). Springer, Boston, MA, 2016. https://doi.org/10.1007/978-1-4899-7641-3_9.

Takahashi, T., 2007. Debris flow: mechanics, prediction and countermeasures. Taylor & Francis.

Tang, C., Zhu, J., Li, W. L., 2009. Rainfall-triggered debris flows following the Wenchuan earthquake. Bulletin of Engineering Geology and the Environment. 68(2), 187-194. https://doi.org/10.1007/s10064-009-0201-6.

Tehrany, M.S., Pradhan, B., Mansor, S., et al., 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena. 125, 91-101. https://doi.org/10.1016/j.catena.2014.10.017.

Tiranti, D., Cavalli, M., Crema, S., et al., 2016. Semi-quantitative method for the assessment of debris supply from slopes to river in ungauged catchments. Science of the Total Environment. 554, 337-348. https://doi.org/10.1016/j.scitotenv.2016.02.150.

Tong, S., Koller, D., 2001. Support vector machine active learning with applications to text classification. Journal of machine learning research. 2(Nov), 45-66.

Trenberth, K.E., 2011. Changes in precipitation with climate change. Climate Research. 47, 123-138. https://doi.org/10.3354/cr00953.

Turkington, T., Remaître, A., Ettema, J., et al., 2016. Assessing debris flow activity in a changing climate. Climatic Change. 137(1), 293-305. https://doi.org/10.1007/s10584-016-1657-6.

Wang, Q., Guo, Y., Li, W., He, J., et al., 2019. Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor. Geomatics, Natural Hazards and Risk. 10(1), 820-835. https://doi.org/10.1080/19475705.2018.1549111.

Wang, Z.S., Yao, X., Sun, J.Z., et al., 2014. An overview on research development of glacier-related debris flow. Journal of Engineering Geology. 22(s1), 459-465. (in Chinese)

Wilkerson, F.D., Schmid, G.L., 2003. Debris flows in Glacier National Park, Montana: geomorphology and hazards. Geomorphology. 55(1-4), 317-328. https://doi.org/10.1016/S0169-555X(03)00147-8.

Wright, R. E., 1995. Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics. American Psychological Association. 217-244.

Xiong, K., Adhikari, B.R., Stamatopoulos, C.A., et al., 2020. Comparison of Different Machine Learning Methods for Debris Flow Susceptibility Mapping: A Case Study in the Sichuan Province, China. Remote Sensing. 12(2), 295. https://doi.org/10.3390/rs12020295.

Xue, R., Wang, C., Liu, M., et al. 2019. A new method for soil health assessment based on Analytic Hierarchy Process and meta-analysis. Science of The Total Environment. 650, 2771-2777. https://doi.org/10.1016/j.scitotenv.2018.10.049.

Xu, D., 1988. Characteristics of debris flow caused by outburst of glacial lake in Boqu river, Xizang, China, 1981. GeoJournal. 17(4), 569-580. https://doi.org/10.1007/BF00209443.

Xu, C., Dai, F.C., Xu, X.W., et al., 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. Geomorphology. 145-146, 70-80. https://doi.org/10.1016/j.geomorph.2011.12.040.

Xu, W., Yu, W., Jing, S., et al., 2013. Debris flow susceptibility assessment by GIS and information value model in a large-scale region, Sichuan Province (China). Natural hazards. 65(3), 1379-1392. https://doi.org/10.1007/s11069-012-0414-z.

Zabihi, M., Mirchooli, F., Motevalli, A., et al., 2018. Spatial modelling of gully erosion in Mazandaran Province, northern Iran. Catena. 161, 1-13. https://doi.org/10.1016/j.catena.2017.10.010.

Zhang, Y.H., Ge, T.T., Tian, W., et al., 2019. Debris Flow Susceptibility Mapping Using Machine-Learning Techniques in Shigatse Area, China. Remote Sensing. 11(23), 2801. https://doi.org/10.3390/rs11232801.

Zhang, M., Ren, Q., Wei, X., et al., 2011. Climate change, glacier melting and streamflow in the Niyang River Basin, Southeast Tibet, China. Ecohydrology. 4(2), 288-298. https://doi.org/10.1002/eco.206.

Zhao, Y., Meng, X., Qi, T.J., et al., 2020. AI-based identification of low-frequency debris flow catchments in the Bailong River basin, China. Geomorphology. 359, 107125. https://doi.org/10.1016/j.geomorph.2020.107125.

Zhou, Z.H., 2016. Machine learning. Beijing, Tsinghua University Press.

Zou, Q., Cui, P., Jiang, H., et al., 2020. Analysis of regional river blocking by debris flows in response to climate change. Science of the Total Environment. 741, 140262. https://doi.org/10.1016/j.scitotenv.2020.140262.

Appendix A

1. Mathematical details of SVM

Support Vector Machine is a supervised and popular classification algorithm in recent years when mapping hazard susceptibility. The main purpose of SVM is to find hyperplane of linearly distributed vectors and separate them into two sets with a maximum gap. For nonlinear distributed vectors, kernel functions will be used to convert them into higher feature space to achieve linearly separability (Melgani et al., 2004; Xu et al., 2012). Therefore, SVM was applied to generate susceptibility map without optimisation of hyper parameters, *C* and gamma. The vectors that are closest to the hyperplane are called support vectors. The optimal hyperplane can be decided based on the following convex quadratic equations for linearly separable vectors (Zhou, 2016):

$$\begin{cases} \min\limits_{\omega,b} \frac{1}{2}\|\omega\|^2 \\ \text{s.t.} y_i\left(\omega^T x_i + b\right) \geq 1, i = 1,2,...,m \end{cases} \tag{8}$$

where ω represents the normal vector of hyperplane; *b* presents the distance between hyperplane and origin. In order to obtain the optimal solution of Eq. (4), dual problem is introduced by using Lagrangian Multiplier method:

$$\begin{cases} \max\limits_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s.t.} \sum_{i=1}^{m} \alpha_i y_i = 0. \alpha_i \geq 0, i = 1,2,...,m \end{cases} \tag{9}$$

where α_i is Lagrangian multiplier. If the vectors are nonlinear distributed, kernel functions should be introduced to simplify the calculation of dual problem in order to avoid complex calculations in high feature space:

$$\begin{cases} \max \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa\left(x_i, x_j\right) \\ \alpha \\ \text{s.t.} \sum_{i=1}^{m} \alpha_i y_i = 0. \alpha_i \geq 0, i = 1,2,\ldots,m \end{cases} \tag{10}$$

The calculated results in original feature space can be expressed as:

$$f\left(x\right) = \omega^T \phi\left(x\right) + b = \sum_{i=1}^{m} \alpha_i y_i \kappa\left(x, x_i\right) + b$$

(11)

where $\kappa\left(x, x_i\right)$ is the kernel function, *m* is the number of causative factors.

2. Introduction of LR

Logistic regression is a multivariate analysis model (Lee et al., 2006) which predicts the absence or presence for an event through the input of a series of variables. Therefore, it has been extensively used for debris flow susceptibility assessment (Elkadiri et al., 2014; Heckmann et al., 2014; Lombardo et al., 2015; Cama et al., 2017). LR is expressed as a linear equation:

$$\log(y) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_i x_i \tag{12}$$

where *y* is the dependent variable, and $y = \frac{p}{1-p}$ Occurrence probability occurrence *P* can be estimated as follows:

$$P = \frac{\exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_i x_i)}{1 + \exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_i x_i)} \tag{13}$$

where α_0 is a constant, α_1, α_2, $\alpha_3 \ldots \alpha_i$ is the i_{th} regression coefficient, x_i is the i_{th} explanatory variable, which represents causative factors. Therefore, the occurrence probability of debris flow can be estimated based on this formula.