Conditioning Factors Determination for Landslide Susceptibility Mapping Using Support Vector Machine Learning

Bahareh Kalantar

Postdoctoral Researcher

- RIKEN Center for Advanced Intelligence Project, Goal-Oriented Technology Research Group, Disaster Resilience Science Team
- Individual Member of the International Society for Photogrammetry and Remote Sensing




Introduction

Landslides cause

- high fatality rates
- huge property losses

- Landslide initiation
- Landslide susceptibility
- Risk assessments

specifically looking at the contribution of individual conditioning variables (or factors).

geomorphical,

hydrological and

topographical

Landslide analysis can help

- to detect areas prone to landslides,
- to provide early warning for affected residents.

Identifying the appropriate conditioning factors is important specially when constructing a model to predict potential landslide area.

This research seeks to expand on previous works, and answer the following questions:

(1) Despite the existing pool of landslide factors, which of these factors best predict landslides susceptibility?

(2) What is the minimum number of factors to construct a model to come up with a consistent landslide potential map?

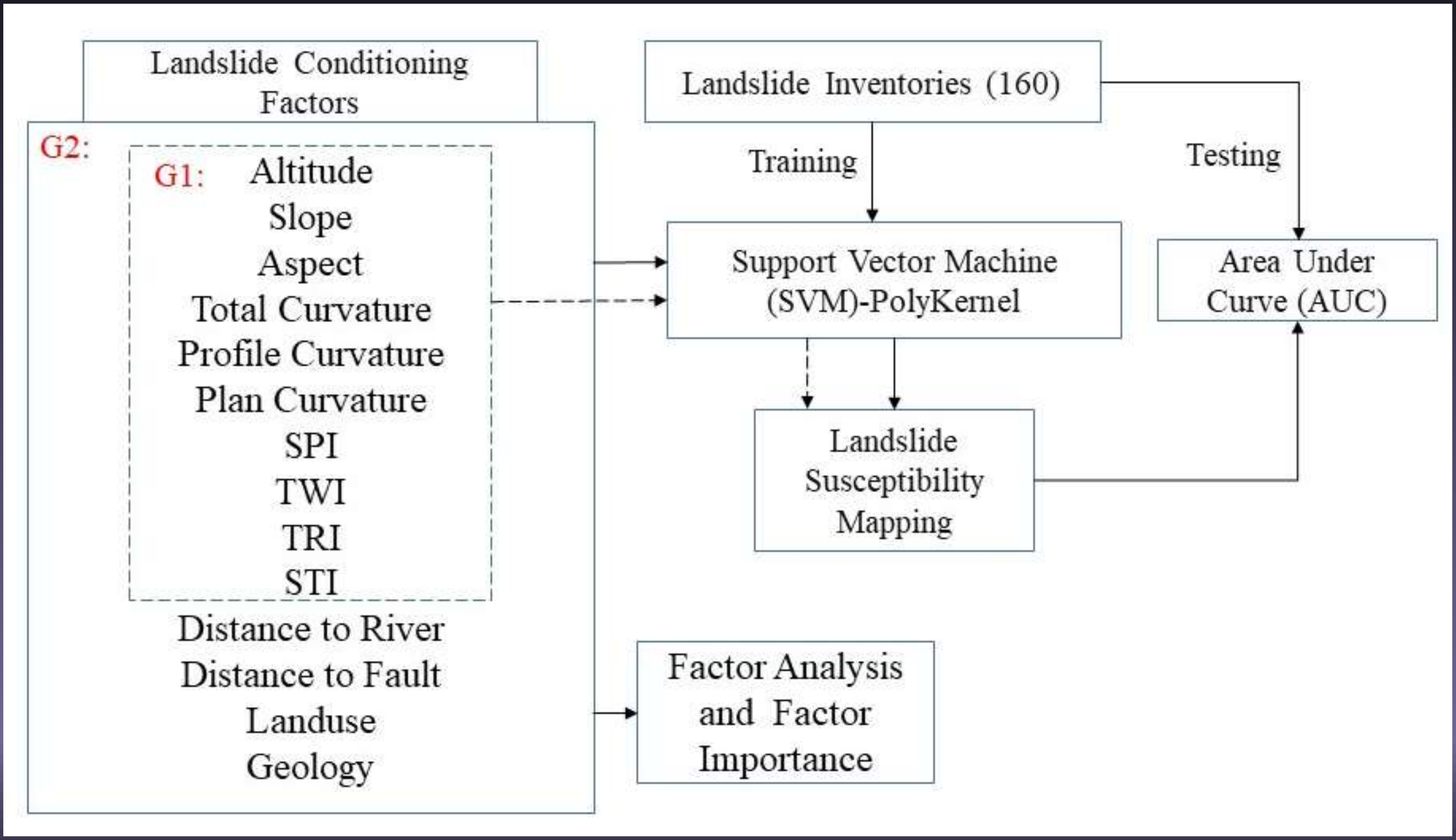
LANDSLIDE CONDITIONING VARIABLES

(1) Slope angle	**(8) Stream Power Index (SPI)**
(2) Slope aspect	(9) Topographic Roughness Index (TRI)
(3) Elevations	(10) Sediment Transport Index (STI)
(4) Total curvature	(11) Landuse-Landcover
(5) Profile curvature	(12) Geology
(6) Plan curvature	(13) Distance from rivers
(7) Topographic Wetness Index (TWI)	(14) Distance to fault

Objectives

- To determine whether or not adding selected factors will improve the prediction of landslide susceptibility.
- To evaluate the performance of the SVM model based on the selected group.

overall workflow



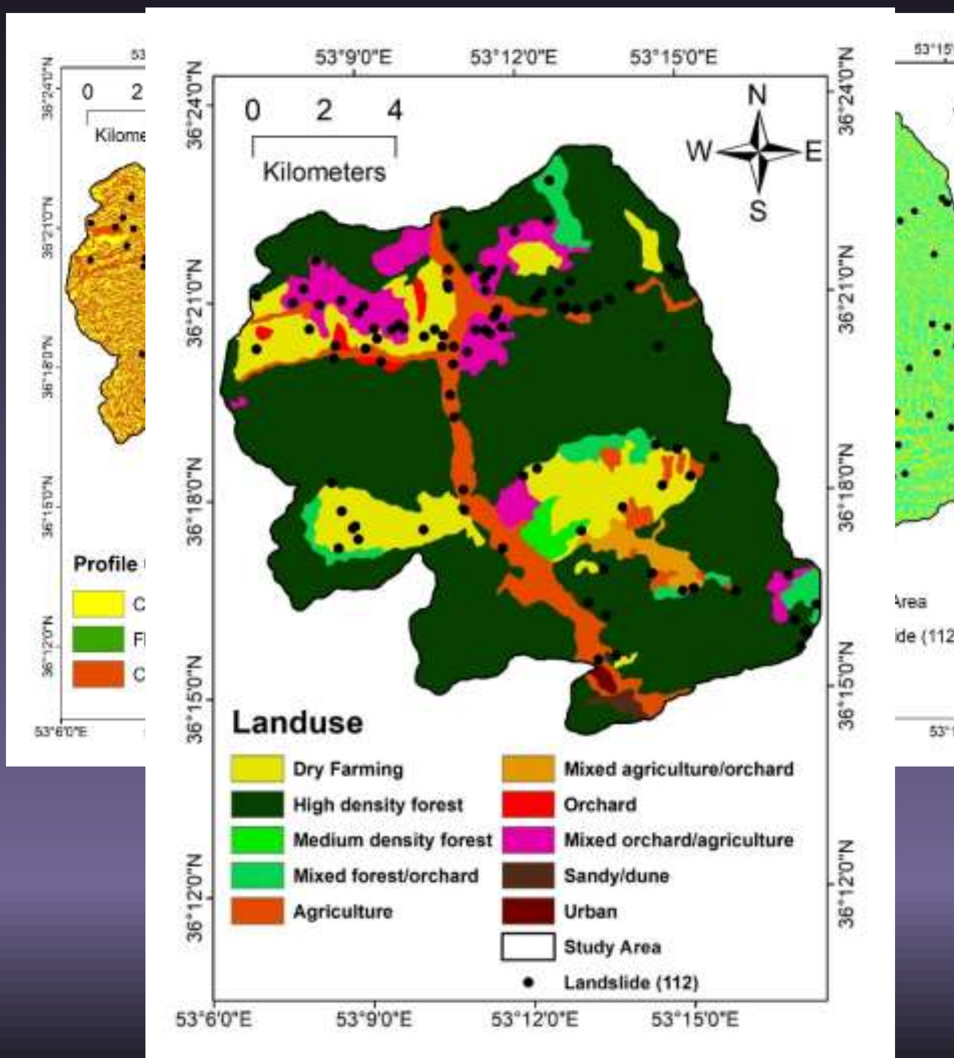
The principles of SVM

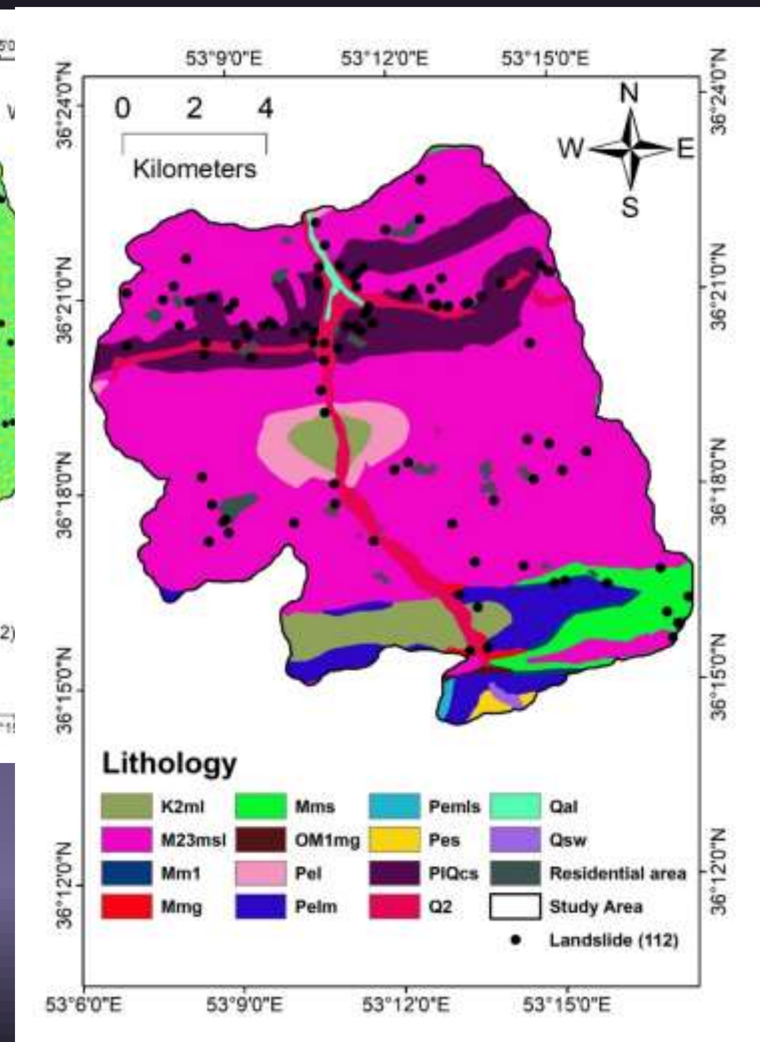
- Basically, the SVM tries to discover an optimal separating hyperplane that could effectively separate the input features of two classes with maximum margin.

$$\frac{y_i(w * x_i + b)}{\geq 1} - \delta_i$$

w is the coefficient vector that defines the hyperplane orientation in the feature space.
b is the offset of the hyperplane from the origin and
$\boldsymbol{\delta_i}$ the positive slack variables

Landslide conditioning factors





Factor analysis and factor importance

- variance-inflated factor (VIF)

$$VIF = \frac{1}{1 - R'^2}$$

where R' represent the multi correlation coefficient between individual feature and the other features in the model.

In the current study, factors with a VIF greater than **5 or 10** were identified as the high correlation and should be removed.

Factor analysis and factor importance

- Pearson's correlation coefficients method

$$r_{xy} = \sum_{i=1}^{n} \frac{X_i - \bar{X}}{\sqrt{\sum_{k=1}^{n}(X_i - \bar{X})^2}} \frac{Y_i - \bar{Y}}{\sqrt{\sum_{k=1}^{n}(Y_i - \bar{Y})^2}}$$

where X_i and Y_i are the values of X and Y for the ith individual.

A high level of colinearity is identified when the Pearson's correlation coefficient is greater than 0.7.

Factor analysis and factor importance

- Cohen's kappa index

$$K = \frac{P_{obs} - P_{exp}}{1 - P_{exp}}$$

P_{obs} denotes the correctly classified proportion of landslide and non-landslide pixels.

P_{exp} indicates the proportion of pixels expected to show agreement, on the basis of chance.

Validation

- The area under the receiver operating characteristic curve (AUC) by evaluation the prediction and success rates was looked at to evaluate the performance of both SVMs.

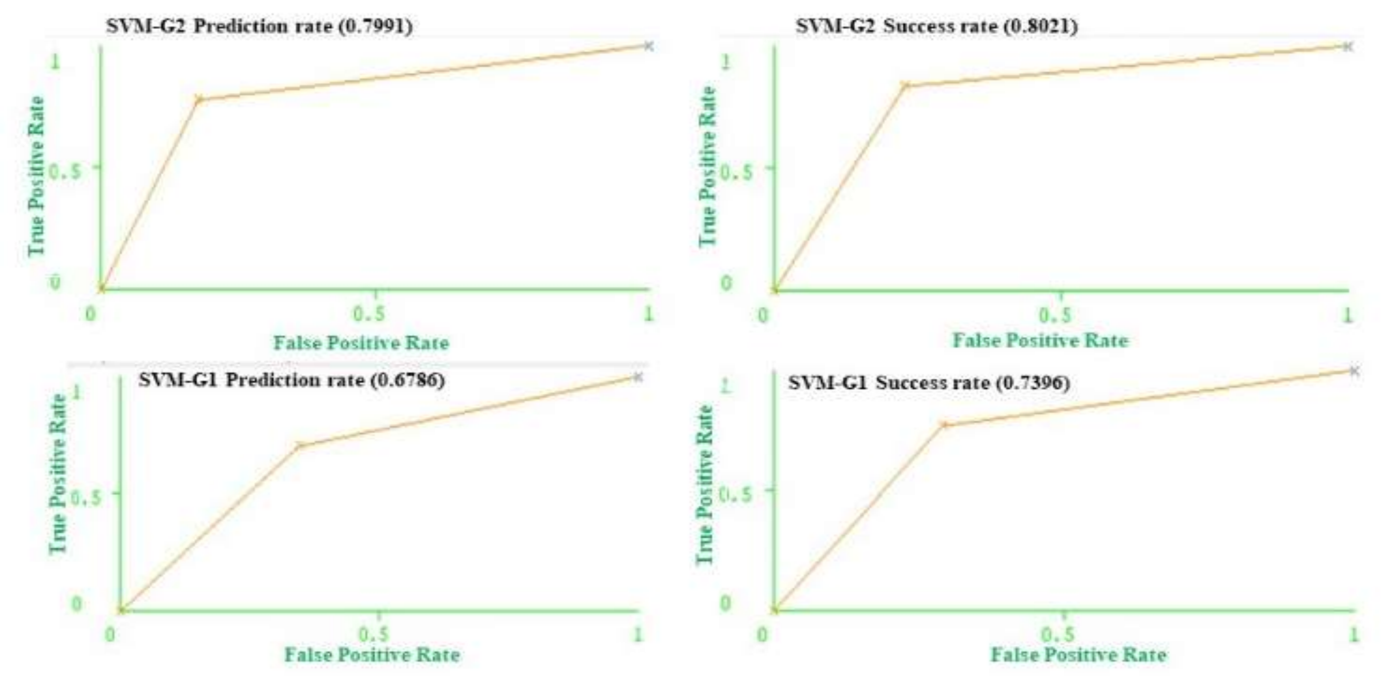
Values from

- 0.5-0.6 indicates poor,
- 0.6-0.7 average
- 0.7-0.8 as good
- 0.8-0.9 means very good
- 0.9-1 is exceptional (or excellent)

Results

ACCURACY OF THE SVM MODEL FOR BOTH G1 AND G2 DATASETS.

	Training points%	Testing points%
G1	68%	74%
G2	80%	81%



The Estimated Variance Information Factor (VIF) for Landslide Conditioning Factors.

No	Conditioning factors	VIF
1	Aspect	1.011966
2	TWI	1.33363
3	TRI	9.315751
4	SPI	7.677249
5	STI	8.555234
6	Geology	1.070003
7	Landuse	1.024453
8	Plan Curvature	4.33E+13
9	Profile Curvature	9.01E+13
10	Total Curvature	1.88E+14
11	Slope	7.029521
12	Distance to Fault	1.013054
13	Distance to River	1.012054
14	Altitude	3.521458

Pearson Correlations Between Landslide Conditioning Factors

Conditioning factors	Aspect	TWI	TRI	SPI	STI	Geology	Landuse	Plan	Profile	Total	Slope	Fault	River	Altitude
Aspect	1.00													
TWI	-0.01	1.00												
TRI	-0.01	-0.34	1.00											
SPI	0.03	0.42	-0.05	1.00										
STI	0.03	0.42	0.08	**0.95**	1.00									
Geology	0.10	0.09	-0.33	-0.06	-0.08	1.00								
Landuse	-0.15	0.13	-0.19	-0.03	-0.04	0.13	1.00							
Plan	-0.01	-0.49	0.01	-0.25	-0.36	0.04	0.03	1.00						
Profile	-0.01	0.23	0.04	0.05	0.12	0.03	-0.06	-0.42	1.00					
Total	0.00	-0.40	-0.02	-0.16	-0.26	0.00	0.05	**0.77**	**-0.90**	1.00				
Slope	-0.01	-0.36	**0.81**	-0.02	0.14	-0.13	-0.12	-0.01	0.06	-0.05	1.00			
Fault	-0.08	0.07	-0.16	0.02	0.00	0.21	0.01	-0.02	-0.03	0.01	-0.09	1.00		
River	0.04	-0.13	-0.06	-0.11	-0.11	0.06	0.15	0.14	-0.08	0.12	0.01	0.19	1.00	
Altitude	0.05	-0.08	0.55	-0.06	-0.05	-0.43	-0.24	0.04	-0.01	0.03	0.03	-0.14	-0.14	1.00

Cohen's Kappa Index for the SVM Technique of Landslide Susceptibility by Removing One Conditioning Factor.

G1		G2	
Landslide conditioning factors	CKI	Landslide conditioning factors	CKI
Without altitude	0.34	Without altitude	0.6941
Without slope	0.28	Without slope	0.5923
Without total curvature	0.30	Without Total curvature	0.6536
Without profile curvature	0.30	Without Profile curvature	0.6533
Without plan curvature	0.32	Without plan curvature	0.6536
Without aspect	0.32	Without aspect	0.6941
Without SPI	0.38	Without SPI	0.6334
Without TWI	0.28	Without TWI	0.6739
Without TRI	0.30	Without TRI	0.6122
Without STI	0.38	Without STI	0.6331
		Without fault	0.6122
		Without River	0.6334
		Without LULC	0.6530
		Without geology	0.6739

Conclusion

- Conditioning factors such as geology, landuse, distance to river, and distance to fault to the DEM-derived dataset, provided better accuracy.

- SVM-G2 has higher accuracy (Testing 81% Training 80%) to compare to SVM-G1 (Testing points: 74%, Training points: 68%).

- High correlation between SPI and STI, total curvature and profile curvature, slope and TRI, as well as between plan curvature and total curvature.

- Slope is the most significant factors between both dataset (G1 and G2) followed by TWI, TRI and distance to fault for landlside susceptibility modeling.


Thank you