

# INTEGRATING INSAR INFORMATION AND SPATIAL-TEMPORAL FACTORS IN MACHINE LEARNING ANALYSIS FOR LANDSLIDE PREDICTION – A CASE STUDY FOR PROVINCIAL HIGHWAY 18 AREA IN TAIWAN

Y. K. Chen<sup>1</sup>, Y. T. Lin<sup>2</sup>, H. Y. Yen<sup>3</sup>, N. H. Chang<sup>4</sup>, H. M. Lin<sup>5</sup>, K. H. Yang<sup>6</sup>, C. S. Chen<sup>7</sup>, L. P. Wang<sup>8</sup>,  
H. K. Cheng<sup>9</sup>, H. H. Wu<sup>10</sup>, J. Y. Han<sup>11\*</sup>

<sup>1</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - d10521008@ntu.edu.tw

<sup>2</sup>National Center for Research on Earthquake Engineering, Taipei, Taiwan - cromating@gmail.com

<sup>3</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - hsiao.ying.yen@gmail.com

<sup>4</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - r08521109@ntu.edu.tw

<sup>5</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - f08521815@ntu.edu.tw

<sup>6</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - khyang@ntu.edu.tw

<sup>7</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - dchen@ntu.edu.tw

<sup>8</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - lpwang@ntu.edu.tw

<sup>9</sup>CECI Engineering Consultants, Inc., Taipei, Taiwan - tc561@ceci.com.tw

<sup>10</sup>CECI Engineering Consultants, Inc., Taipei, Taiwan - zephyr@ceci.com.tw

<sup>11</sup>Department of Civil Engineering, National Taiwan University, Taipei, Taiwan - jyhan@ntu.edu.tw

## Commission III, ICWG III/IV a

**KEYWORDS:** Landslide Prediction, InSAR, Spatial-Temporal Factors, Slope Unit, Spearman, Machine Learning.

## ABSTRACT:

Taiwan is located in subtropical monsoon area and Pacific Ring of Fire. Both the rate of crustal uplift and annual rainfall are among the highest in the world. Earthquakes and heavy rainfall have led to massive landslides and debris flow. Frequent disasters and the high rate of surface erosion have caused drastic changes in river topography and catchment areas, and, consequently, have impacted the safety of human lives. To mitigate the losses, better simulation and prediction of landslides are critical. Existing landslide prediction research works employed terrain, geology, rainfall, earthquakes and human activities as landslide triggering factors in the predicting model. In addition to aforementioned environmental conditions, this study would like to explore the use of SAR differential interferometry (InSAR) information to help observe characteristics of the slope movement behavior, which is also an important factor. Factors are analyzed and quantified on the basis of slope units. To confirm the applicability of selected factors to landslide, factors are firstly analyzed with Spearman correlation, and then those with higher correlations are incorporated into the prediction model. Machine learning based techniques are then employed to establish the prediction model. The experiment result demonstrates that InSAR information can improve the accuracy by more than 5% in landslide prediction.

## 1. INTRODUCTION

Slopes in Taiwan mountainous area are mostly geologically sensitive and fragmented, so when heavy rainfall or earthquake occur, it easily leads to landslides. The causes of landslide can be divided into two categories: potential conditions and triggering factors. The potential conditions are the natural conditions of slopes, namely topography, lithology, geological structure, and vegetation coverage; while the triggering factors are those including disasters, such as rainfall, earthquakes, and human activities. Existing studies mostly employ the factors from the above two categories into mechanic and statistical models to calculate landslide risk. However, due to the scarcity of data points and remote location, it is difficult to obtain large-scale surface displacement information. To solve this problem, this research explores the use of time series phase values measured by SAR interferometric techniques (InSAR) to obtain large-scale ground deformation information, i.e. the slope movement characteristics. Through correlation analysis and machine learning techniques, we further investigate the impact of spatial-temporal factors and InSAR information, and improve the accuracy and reliability of the landslide prediction research.

## 2. STUDY AREA

The study area is selected the area along Provincial Highway 18, Chiayi, the central part in Taiwan (Figure 1). The streams and highways in here are all east-west. The study area is located in Fanlu Township and Alishan Township, Chiayi County. Covering an area of about 398.268 square kilometers. The main stream is the Bashang River. The terrain here is relatively rugged, as long as it encounters heavy rainfall events such as typhoons or plum rains, landslides usually occur.

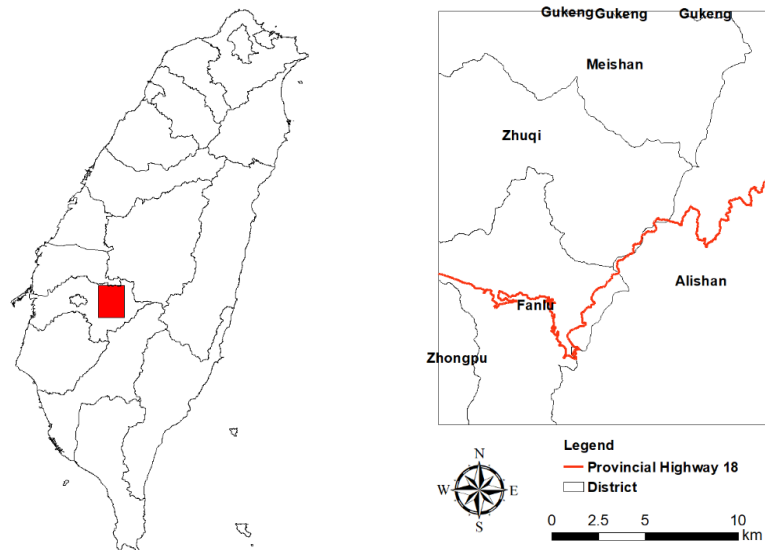
## 3. METHOD

### 3.1 Slope Unit

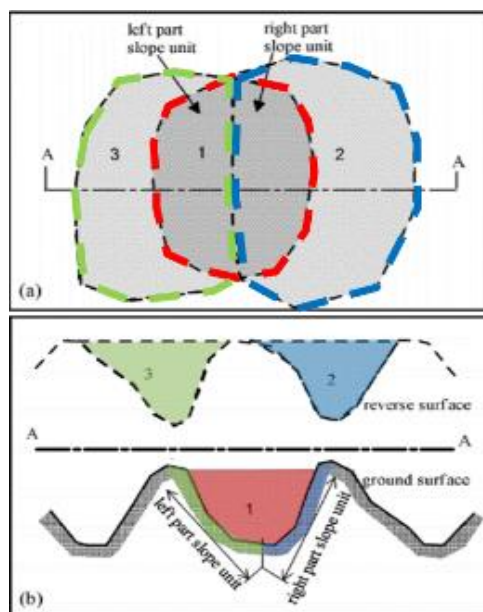
To preserve the complete topographic patterns, slope unit method is selected as the basic unit for analysis of various spatial-temporal factors. The parameters obtained by the slope unit are used to represent the entire unit, and it has the continuous meaning of a slope, which is a suitable method for landslide potential research. The slope unit is divided by the mountain ridge line and the watershed. Slope units processing is referred to the method proposed by Xie et al. (2004). The principle is to first find out and divide the digital elevation model (DEM) into several catchment areas. Second is to invert the DEM to get the reverse stream and mountain

ridge information, and then divide the reverse DEM into several catchment areas again to obtain the reverse catchment areas. Third is to combine the positive and negative catchment areas into a single file to cut out the slope unit (Fig 2). Since this result sometimes has unreasonable units,

manual editing is still needed to obtain slope units with a more homogeneous terrain information. In this study area, there are 3312 slope units, and each slope unit is range in 15 to 30 hectares.



**Figure 1.** The study area is the area along Provincial Highway 18 and located in Fanlu Township and Alishan Township, Chiayi County.



**Figure 2.** Slope units processing. (Xie et al., 2004).

### 3.2 Spatial-Temporal Factors

Common spatial-temporal factors can be categorized into three parts according to their characteristics, i.e. geomorphology, location and geology. Geomorphic factors include elevation, slope angle, aspect, topographic roughness, curvature and vegetation coverage (NDVI); location factors represent the influence of the distance of factors that disturb the slope unit, including river distance, road distance and fault distance; geological factors represent the influence caused by regional geological conditions, such as sensitive area index, fold and downslope index; Triggering factors are

factors which can trigger landslides, such as rainfall, earthquake and deforestation. In this study, monthly accumulated rainfall information has been applied.

### 3.3 Spearman Correlation Coefficient

The main purpose of correlation analysis is to understand the statistical dependence between two variables. To discover the applicability of the selected factors in this study, a correlation analysis of these factors is applied to understand the correlation between each factor and landslide. Considering that the distribution of each factors is different, and may not be able to fit the strict assumptions of the statistical method, such as the normal distribution, this study adopts Spearman Correlation Coefficient, which is a nonparametric measure of rank correlation. The data are only required to be converted into accordance with ascending or descending order, which can be applied to the description of the linear relationship of variables. The definition of Spearman Correlation Coefficient is as follows:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

where  $r_s$  = Spearman Correlation Coefficient  
 $d_i$  = the difference between the two ranks  
 $n$  = the number of observations

### 3.4 InSAR Information Collection

This study collects the InSAR information from Sentinel-1, ESA, for five years, from 2016 to 2020. The deformation information from InSAR has been calculated in time interval: 4 months. InSAR technique is to obtain SAR images by using the repeated satellite orbits in the same area at different times, and obtains surface deformation information from the phase value. To explore the rationality of the ground displacement

from InSAR, the displacement velocity will be converted from the satellite line of sight ( $V_{LOS}$ ) to slope ( $V_{slope}$ ) direction. After converting the line of sight to slope, there are 600  $V_{slope}$  points in the study area, which means only 18% slope units have InSAR information. According to Aslan et al. (2020), the speed of deformation displacement in the direction of line of sight of SAR satellite can be converted into the direction of the slope. Figure 3 is the schematic of the conversion between  $V_{slope}$  and  $V_{LOS}$ , and the equation of conversion is as follows:

$$V_{slope} = \frac{V_{LOS}}{C} \quad (2)$$

Where the coefficient C represents the proportion of the real surface displacement velocity ( $V_{slope}$ ) along the slope when projected to the LOS direction, which can be calculated by

the geometric relationship between line of sight direction and slope units. The equation of the coefficient C is as follows:

$$C = N \cdot [\cos(S) \cdot \sin(A - 90^\circ)] + E \cdot [-\cos(S) \cdot \cos(A - 90^\circ)] + H \cdot \sin(S) \quad (3)$$

$$N = \cos(90^\circ - \alpha) \cdot \cos(180^\circ - \gamma) \quad (4)$$

$$E = \cos(90^\circ - \alpha) \cdot \cos(270^\circ - \gamma) \quad (5)$$

$$H = \cos(\alpha) \quad (6)$$

where A = slope direction of the slope unit  
S = slope angle of the slope unit  
 $\alpha$  = incident angle of line of sight direction  
 $\gamma$  = azimuth angle of line of sight direction

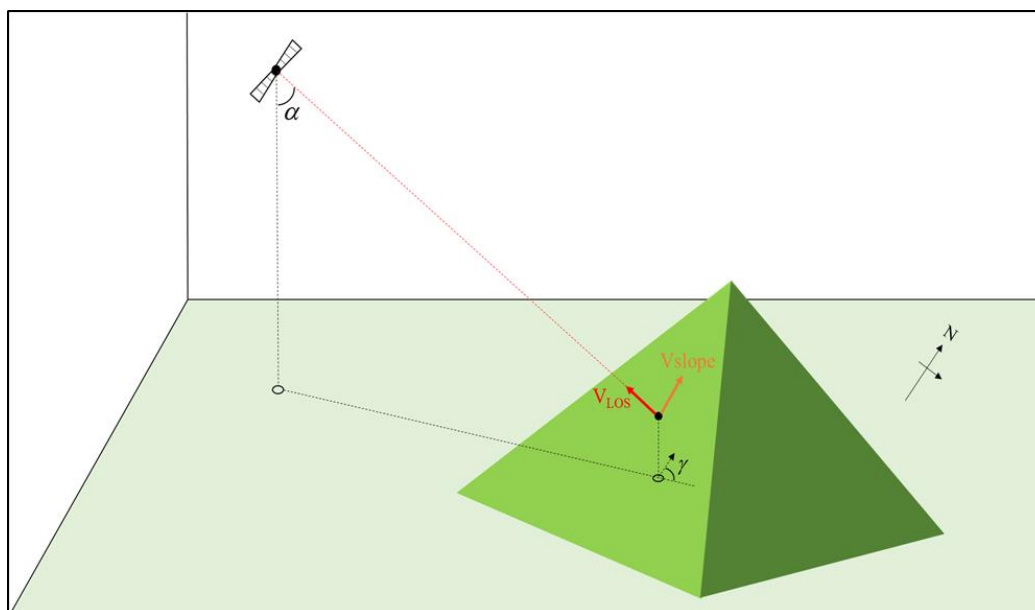


Figure 3. The schematic of the conversion between  $V_{slope}$  and  $V_{LOS}$ .

### 3.5 Machine Learning Prediction

Machine learning techniques are commonly applied in establishment of landslide prediction model. In this study, Random Forest algorithm has been applied for landslide prediction. Random forest is a classifier that contains multiple decision trees, and add random selection of training data to greatly improve the final calculation result. The equation of Random Forest algorithm is as follows:

$$P(c | v_i(x)) = \frac{P(c|v_i(x))}{\sum_{i=1}^n P(c_i|V_i(x))} \quad (7)$$

$$g_c(x) = \frac{l}{t} \sum_{i=1}^t \hat{P}(c | v_i(x)) \quad (8)$$

where P = probability  
c = category  
v = node  
l = number of categorise  
t = number of decision trees  
 $g_c$  = the average probability for category c

For classification, the output of Random Forest is the class selected by most trees. Random Forests correct for decision

trees' habit of overfitting to their training set. This study uses 14 spatial-temporal factors as import data, of which 5 years from 2016 to 2020 are used as training samples and then predict in March 2020 and September 2020 whether each slope unit will occur landslides. Finally, calculate and compare the prediction accuracy between entire slope units and slope units with InSAR information.

### 3.6 Landslide Ground Truth Map

The landslide inventory announced by Forestry Bureau is once a year, and the latest landslide inventory was announced in 2017. The time scale in years does not meet the needs in this study, thus, SPOT images have been downloaded for making the landslide ground truth map in this study. The SPOT image has been supervised classification to preliminarily classify into categories such as bare soil, vegetation, water body, and cloud cover. Since the spectral characteristics of river channel and settlements are similar to landslides, they were classified into the bare soil category in the preliminary classification, however, river channel and settlements are not landslides. Thus, manual editing is carried out to separate the preliminary category bare soil into real landslide areas, river channels, settlements, and vegetation regeneration areas.

#### 4. RESULTS AND DISCUSSION

The spatial-temporal factors in this study are all quantified and analyzed based on slope units and show in Figure 4. From the significance test (Table 1), the significant factors related to the landslide in the study area are slope, topographic roughness, curvature, distance to river, distance to fault, downslope index, fold and sensitive area index. The results of correlation analysis demonstrate that the correlation between factors and landslide are mostly weak to modest, which means the impact of one single factor is not obvious to landslide.

There are 3312 slope units in the study area, of which 14 spatial-temporal factors are used as the import data in Random Forest landslide prediction model. The prediction

result (Table 2) shows that in March 2020, the overall accuracy is 88.1%; in September 2020 forecast, the overall accuracy is 88.9%. For 600 slope units with  $V_{slope}$  points (about 18% of the total units), the overall accuracy for landslide prediction in March 2020 is 94.2%; the overall accuracy for landslide prediction in September 2020 is 94.8%. Both are significantly higher than the accuracy of the entire slope unit by more than 5%. In the study area, 213 landslide units (about 6% of the total units) were increased in March 2020, and 4 landslide units (about 0.1% of the total units) were increased in September 2020. For increased landslides prediction (Table 2), the overall accuracy in March 2020 and in September 2020 can reach 97.2% and 100% respectively. The results demonstrate that the machine learning model has enough capability to predict increased landslides.

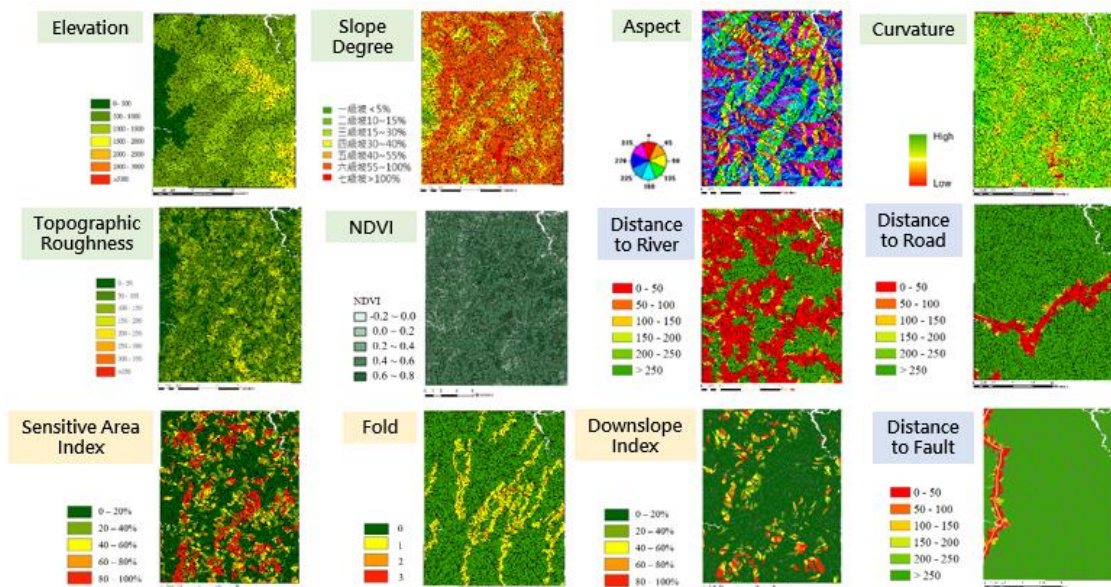


Figure 4. Spatial-temporal factors in slope units.

Table 1. Spearman correlation coefficient test between selected factors and landslide.

	Factors	Correlation Coefficient	Level of Significance	Relevance
Geomorphic Factors	Slope Angle	0.18	0.00	Modestly correlated
	Elevation	0.01	0.126	Weakly correlated
	Topographic Roughness	0.14	0.00	Modestly correlated
	Curvature	-0.21	0.00	Modestly correlated
Location Factors	Distance to River	-0.12	0.00	Modestly correlated
	Distance to Road	0.01	0.16	Weakly correlated
	Distance to Fault	0.15	0.00	Modestly correlated
Geological Factors	Downslope Index	0.04	0.00	Weakly correlated
	Fold	0.04	0.00	Weakly correlated
	Sensitive Area Index	0.28	0.00	Modestly correlated
InSAR Factors	$V_{slope}$	-0.06	0.00	Weakly correlated

Note: The lower the level of significance level, the better the factor significance, and the level of significance < 0.05 means significant.

Table 2. The overall accuracy for landslide prediction result.

Prediction Time	Total Slope Units	Slope Units with InSAR V <sub>slope</sub> Points	Slope Units with Newly- Increased Landslides
March 2020	88.1 %	94.2 %	97.2
September 2020	88.9 %	94.8 %	100

## 5. CONCLUSION

In this study, the area along Provincial Highway 18, Chiayi was chosen to be the study area. The significant factors related to the landslide are slope, topographic roughness, curvature, distance to river, distance to fault, downslope index, fold and sensitive area index. The weak to modest correlation results demonstrate that the impact of one single factor is not obvious to landslide. For landslide prediction, Random Forest algorithm is performed by applying the 5-year (2016-2020) spatial-temporal factors to the prediction of landslides in March and September 2020, and the accuracy can be higher than 88%. Research results demonstrate that after applying InSAR information, the capability of the prediction model has been improved, which is more than 5% higher than slope units without InSAR points. For predicting newly increased landslides, the prediction accuracy can even exceed 95%. In this study, the analysis and prediction method of slope displacement is proposed to realize its feasibility by integrating common spatial factors (geomorphology, location and geology) and InSAR temporal observation data.

## 6. REFERENCE

Amrani, M., Bey, A., Amamra, A., 2021: New SAR target recognition based on YOLO and very deep multi-canonical correlation analysis. *INTERNATIONAL JOURNAL OF REMOTE SENSING*, 2021.1953719. doi.org/10.1080/01431161.2021.1953719.

Bui, T., Lee, P., Lum, K., Loh, C., Tan, K., 2020: Deep Learning for Landslide Recognition in Satellite Architecture. *IEEE Access*, VOLUME 8, 2020. doi: 10.1109/ACCESS.2020.3014305.

Carla, T., Intrieri, E., Raspini, F., Bardi, F., Farina, P., Ferretti, A., Colombo, D., Novali, F., Casagli, N., 2019: Perspectives On The Prediction Of Catastrophic Slope Failures From Satellite InSAR. *SCIENTIFIC REPORTS*, (2019) 9:14137. doi: 10.5194/egusphere-egu2020-13158.

Casagli, N., Cigna, F., Bianchini, S., Hölbling, D., Füreder, P., Righini, G., Del Conte, S., Friedl, B., Schneiderbauer, S., Iasio, C., Vlcko, J., Greif, V., Proske, H., Granica, K., Falco, S., Lozzi, S., Mora, O., Arnaud, A., Novali, F., Bianchi, M., 2016: Landslide mapping and monitoring by using radar and optical remote sensing: Examples from the EC-FP7 project SAFER. *Remote Sensing Applications: Society and Environment*, 4, 92-108. doi.org/10.1016/j.rsase.2016.07.001.

Chae, B., Park, H., Catani, F., Simoni, A., Berti, M., 2017: Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal*, Vol. 21, No. 6, p. 1033-1070, December 2017. doi.org/10.1007/s12303-017-0034-4.

Chang, P.-L., Zhang, J., Tang, Y.-S., Tang, L., Lin., P.-F., Langston, C., Kaney, B., Chen, C.-R., Howard, K., 2021: An Operational Multi-Radar Multi-Sensor QPE System in Taiwan, *Bulletin of the American Meteorological Society*, 102(3), E555–E577. doi.org/10.1175/BAMS-D-20-0043.1.

Del Ventisette, C., Righini, G., Moretti, S., Casagli, N., 2014: Multitemporal landslides inventory map updating using spaceborne SAR analysis. *International Journal of Applied Earth Observation and Geoinformation*, 30, 238-246. doi.org/10.1016/j.jag.2014.02.008.

Haile, A. T., Rientjes, T. H. M., Habib, E., Jetten, V., and Gebremichael, M., 2011: Rain event properties at the source of the Blue Nile River, *Hydrol. Earth Syst. Sci.*, 15, 1023–1034. doi.org/10.5194/hess-15-1023-2011, 2011.

Lin, S., Lin, C., Gasselt, S., 2021: Processing Framework for Landslide Detection Based on Synthetic Aperture Radar (SAR) Intensity-Image Analysis. *remote sensing*, 2021, 13, 644. doi.org/10.3390/rs13040644.

Lin, Y., Chen, Y., Yang, K., Chen, C., Han, J., 2021: Integrating InSAR Observables and Multiple Geological Factors for Landslide Susceptibility Assessment. *Appl. Sci.*, 2021, 11, 7289. doi.org/10.3390/app11167289.

Luo, S., Sarabandi, K., Tong, L., Pierce, L., 2016: Landslide prediction using soil moisture estimation derived from polarimetric Radarsat-2 data and SRTM. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. doi: 10.1109/IGARSS.2016.7730403.

Mondini, A., Santangelo, M., Rocchetti, M., Rossetto, E., Manconi, A., Monserrat, O., 2019: Sentinel-1 SAR Amplitude Imagery for Rapid Landslide Detection. *remote sensing*, 2019, 11, 760. doi.org/10.3390/rs11070760.

Ochoa-Rodríguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Krolle, S., Murlà-Tuyls, D., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P. and ten Veldhuis, J. A. E., 2015: Impact of spatial and temporal resolution of rainfall inputs on semi distributed urban drainage modelling outputs: A multi-catchment investigation, *J. Hydrol.*, 531, 389-407. doi.org/10.1016/j.jhydrol.2015.05.035.

Ochoa-Rodríguez, S., Wang, L. P., Willems, P., & Onof, C., 2019: A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resources Research*, 55(8), 6356–6391. doi.org/10.1029/2018WR023332.

Pourghasemi, H., Rahmati, O., 2017: Prediction of the landslide susceptibility: Which algorithm, which precision?

*Catena*, 162 (2018) 177–192.  
doi.org/10.1016/j.catena.2017.11.022.

Pradhan, B., Lee, S., 2010: Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling Biswajeet Pradhan. *Environmental Modelling & Software*, 25, 747-759. doi.org/10.1016/j.envsoft.2009.10.016.

Shibayama, T., Yamaguchi, Y., Yamada, H., 2015: Polarimetric Scattering Properties of Landslides in Forested Areas and the Dependence on the Local Incidence Angle. *Remote Sens.* 2015, 7, 15424-15442. doi.org/10.3390/rs71115424.

Shimada, M., Watanabe, M., Kawano, N., Ohki, M., Motooka, T., 2013: Detecting Mountainous Landslides by SAR Polarimetry : A Comparative Study Using Pi-SAR-L2 and X-band SARs. Trans. *JSASS Aerospace Tech. Japan*, Vol. 12, No. 12, pp. Pn\_9-Pn\_15, 2014. doi.org/10.2322/tastj.12.Pn\_9.

Tay, C. J., Yun, S., Chin, S., Bhardwaj, A., Jung, J., Hill, E., M., 2020: Rapid flood and damage mapping using synthetic aperture radar in response to Typhoon Hagibis, Japan. *Scientific Data*, 2020, 7:100. doi.org/10.6084/m9.figshare.11968443.

Tsou, C.Y., Feng, Z.Y., Chigira, M., 2011: Catastrophic landslide induced by Typhoon Morakot, Shialin, Taiwan. *Geomorphology*, 127, 166-178. doi.org/10.1016/j.geomorph.2010.12.013.

Utomo, D., Chen, S., Hsiung, P., 2019: Landslide Prediction with Model Switching. *applied sciences*, 2019, 9, 1839. doi.org/10.3390/app9091839.

Wang, L.-P., Ochoa-Rodríguez, S., Simões, N. E., Onof, C. and Maksimović, Č., 2013: Radar-raingauge data combination techniques: a revision and analysis of their suitability for urban hydrology, *Water Sci. Technol.*, 68(4), 737-747. doi.org/10.2166/wst.2013.300.

Wang, L.-P., Ochoa-Rodríguez, S., Pina, R. D., Pessemier, M., Kroll, S., Van Assel, J., Willems, P., Onof, C., 2015: Enhancement of radar rainfall estimates for urban hydrology through optical flow temporal interpolation and Bayesian gauge-based adjustment, *J. Hydrol.*, 531, 408–426. doi.org/10.1016/j.jhydrol.2015.05.049.

Yu, C., Chen, J., 2020: Landslide Susceptibility Mapping Using the Slope Unit for Southeastern Helong City, Jilin Province, China: A Comparison of ANN and SVM. *Symmetry*, 2020, 12, 1047. doi.org/10.3390/sym12061047.

Zhang, L., Dong, R., Yuan, S., Li, W., Zheng, J., Fu, H., 2021: Making Low-Resolution Satellite Images Reborn: A Deep Learning Approach for Super-Resolution Building Extraction. *Remote Sens.* 2021, 13, 2872. doi.org/10.3390/rs13152872.