

## CLASSIFICATION OF TRANSLATIONAL LANDSLIDE ACTIVITY USING VEGETATION ANOMALIES INDICATOR (VAI) IN KUNDASANG, SABAH

Mohd Radhie Mohd Salleh<sup>1</sup>, Nur Huda Athirah Norhairi<sup>1</sup>, Zamri Ismail<sup>1</sup>, Muhammad Zulkarnain Abd Rahman<sup>1\*</sup>, Mohd Faisal Abdul Khanan<sup>1</sup>, Mohd Asraff Asmadi<sup>2</sup>, Khamarrul Azahari Razak<sup>2</sup>, Tze Huey Tam<sup>1</sup>, Mohamad Jahidi Osman<sup>1</sup>

<sup>1</sup> Department of Geoinformation, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Johor – mohdradhie.gis@gmail.com, hudathirah@gmail.com, zamriismail@utm.my, mdzulkarnain@utm.my, mdfaisal@utm.my, asraff65@gmail.com, thtam2@live.utm.my, mohamadjahidiosman@gmail.com

<sup>2</sup> UTM Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia, Kuala Lumpur - khamarrul.kl@utm.my

### Commission 4, WG 7

**KEY WORDS:** Landslide, Landslide Activity, Vegetation Anomalies Indicator, Remote Sensing, GIS, Support Vector Machine.

### ABSTRACT

This paper introduced a novel method of landslide activity mapping using vegetation anomalies indicators (VAIs) obtained from high resolution remotely sensed data. The study area was located in a tectonically active area of Kundasang, Sabah, Malaysia. High resolution remotely sensed data were used to assist manual landslide inventory process and production on VAIs. The inventory process identified 33, 139, and 31 of active, dormant, and relict landslides, respectively. Landslide inventory map were randomly divided into two groups for training (70%) and validation (30%) datasets. Overall, 7 group of VAIs were derived including (i) tree height irregularities; (ii) tree canopy gap; (iii) density of different layer of vegetation; (iv) vegetation type distribution; (v) vegetation indices (VIs); (vi) root strength index (RSI); and (vii) distribution of water-loving trees. The VAIs were used as the feature layer input of the classification process with landslide activity as the target results. The landslide activity of the study area was classified using support vector machine (SVM) approach. SVM parameter optimization was applied by using Grid Search (GS) and Genetic Algorithm (GA) techniques. The results showed that the overall accuracy of the validation dataset is between 61.4–86%, and kappa is between 0.335–0.769 for deep-seated translational landslide. SVM RBF-GS with 0.5m spatial resolution produced highest overall accuracy and kappa values. Also, the overall accuracy of the validation dataset for shallow translational is between 49.8–71.3%, and kappa is between 0.243–0.563 where SVM RBF-GS with 0.5m resolution recorded the best result. In conclusion, this study provides a novel framework in utilizing high resolution remote sensing to support labour intensive process of landslide inventory. The nature-based vegetation anomalies indicators have been proved to be reliable for landslide activity identification in Malaysia.

### 1. INTRODUCTION

Landslides are one of the frequent natural disasters in the world, causing substantial economic losses and casualties (Dehnavi et al., 2015, Gaidzik et al., 2017, Tiranti and Cremonini, 2019). About 7000 deaths and more than 1,000 million USD in economic losses have been documented based on reported landslide occurrences in Asia (2005 to 2016) (Sanderson and Sharma, 2016). Landslides impact on many aspects (socio-economic, ecological, etc.) of the affected inhabitants. This phenomenon is mainly caused by the urbanisation process, deforestation, and uncontrolled development (Scaioni et al., 2014, Schuster, 1996). An effective risk assessment framework is required, which involves susceptibility and hazard modelling, vulnerability assessment, and risk assessment (Van Westen et al., 2008). Landslide inventories are the initial and fundamental steps for susceptibility, hazard, and risk assessments which consider the rule that the past is the key to the future (Highland and Bobrowsky, 2008). In this case, future landslides will probably happen under similar conditions (Van Westen et al., 2008). Therefore, landslide-prone areas with a different state of activity need to be assessed for mitigation purposes.

However, previous studies have shown that landslide mapping in a vegetated area is very challenging due to the covering effect

of dense vegetation (Jaboyedoff et al., 2012), its widespread distribution in an undulating area (Tien Bui et al., 2018), and rapid vegetation growth (Mezaal et al., 2017, Pradhan et al., 2015) that will remove the landslide signature. Furthermore, the hilly and inaccessible area complicates the actual boundary identification of landslide. Conventionally, recognition of landslide type and activity conducted by an interpreter depends on various diagnostic features like morphology or shape, drainage, and structural conditions. For instance, an active landslide is commonly fresh, in which their morphological features, such as scarps and ridges, are readily identifiable due to gravitational movement and have not been substantially altered by weathering and erosion processes (Varnes, 1978). Besides, synthetic aperture radar (SAR) technology is primarily used to measure surface deformations, which creates time-series data of surface deformations, at single points (Berardino et al., 2002). However, the quality of data depends on various factors such as foreshortening, layover effects, atmosphere propagation effects, and vegetation decorrelation in forested terrain (Scaioni et al., 2014). Furthermore, the techniques of the conventional survey are also widely used for landslide activity mapping and analysis (Artese and Perrelli, 2018). In Malaysia, landslide analysis requires frequent site visits, real-time deformation monitoring, and expensive instrument such as electronic distance meter (EDM) that is installed in the landslide area for

\* Corresponding author

monitoring purposes (Tsai et al., 2012). This method, however, produces some hurdles, especially for active landslide zones in which enormous financial allocation and instruments can lead to costly damage once the landslide strikes. For this reason, the application of vegetation as one of the indicators of landslide activity may lead to an effective, natural, and inexpensive approach for landslide inventory and monitoring processes especially in tropical region.

Vegetation may have a beneficial impact on the stability of slope because the roots reinforce the layer by binding the soil. Previously, several VAI were used to determine the relationship between vegetation characteristics and landslide occurrences such as tree height irregularities, leaning trunk, tree-ring, etc. Razak et al. (2013) found that trees in landslide active areas are characterised by low vegetation, small crown, and irregular tree height. Tree trunks are often curved as a result of external factors including catastrophic events, availability of light, and soil creep (Zhang et al., 2016). The fact that the trees frequently grow with curved trunks on slopes where other evidence indicates there is an occurrence of creep, strongly suggests that their curvature is due to creep (geotropism) (Harker, 1996). Wang et al. (2016b) and Wang et al. (2016a) also demonstrated the application of Terrestrial Laser Scanning (TLS) in characterising tree stem curve. They found that most of the trees inside landslide (shallow) areas had curved stems due to the effects of soil movement. The utilisation of dendro-chronological or tree-ring analysis methods could enhance or improve the understanding of landslide activity and risk (Łuszczynska et al., 2017) since stem deformations affect the structure of wood and tree-rings (Stefanini, 2004). However, using tree-rings in studying the behaviour of landslides has its limitations; for example, some trees need to be cut down, which is impossible in some places such as forest nature reserves. Since very limited study conducted in tropical area, the primary objective of this study was to derive various VAIs from high resolution remotely sensed data i.e., satellite image and airborne LiDAR. The relationship between VAIs and the landslide activity can be further investigated in detail. In order to determine the status of landslide activity in the tropical area of Kundasang, Sabah, statistical and machine learning methods were implemented. Through this case study, the difficulties of landslide activity mapping over tropical vegetated area can be addressed. This can be useful to guide the in-situ landslide activity monitoring and labour-intensive manual image interpretation over large landslide zones in the tropical area.

## 2. STUDY AREA

Kundasang is located in the Northwest of Ranau, Sabah (Figure 1). Generally, this area is well-known for its unstable hilly area. The study area covers 70.47 km<sup>2</sup> which is situated between 500 to 2000 m above mean sea level. The hilly terrain and ridges with high elevation and combinations of steep slopes were the consequence of violent tectonic activities in the past (Tating, 2006). The climate of the study area is determined by its proximity to the equator; it is humid and tropical, with temperatures ranging between 25°C and 35°C in lowland areas. Since the area is located within what is known as the ever wet zone, it receives at least 60 mm of rainfall per month with the annual rainfall ranges from 1920 mm to 3190 mm (average 2075 mm) (Tating et al., 2013). This area was selected to pursue the objective of this study covering the tropical environments and abundance of occurrences of landslides caused by natural and anthropogenic factors.

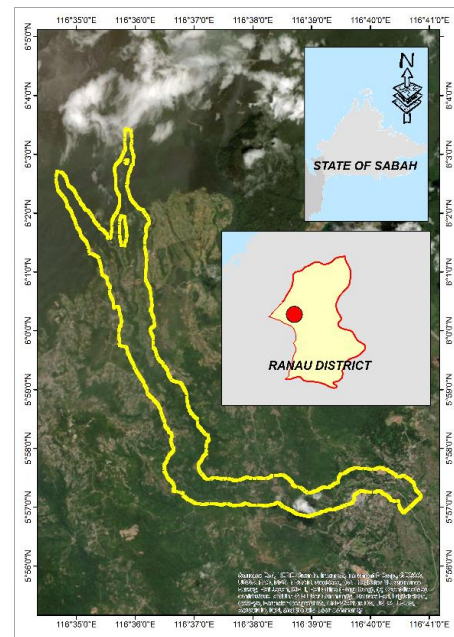


Figure 1. Location of study area

Landslides are increasing problem in Kundasang, as a result of natural or human interference to natural slopes (Sharir et al., 2017). This area has been identified as a high-risk area for major landslides due to ongoing movement characterized by ground tension cracks, sudden localized failures, and ground creep (Zainal Abidin et al., 2012). Rocks located beneath Kundasang vary in age and type, which is the rock starting from Paleocene-Eocene rocks to alluvial rock. Based on the record, Kundasang has an average annual soil translation of 0.5m, and approximately 70% of the 50 square kilometers surrounding Kundasang Town has been identified as a high-risk area (Morpi, 2011). On June 5, 2015, an earthquake measuring 6.0 Mw occurred in Sabah that had triggered the debris flow which disrupted roads, houses, and the vegetation along the channel. The earthquake was caused by movement on a Southwest-Northeast (SW-NE) trending normal fault and the epicenter was near Mount Kinabalu. The shaking caused massive landslides around the mountain (Tongkul, 2017). Therefore, it can be concluded that Kundasang area is located in the geohazard zone consist of complex geological structures involving chaotic geomaterials with highly jointed rock, fault in a zone of intense seismic activity (Komoo and Lim, 2003, Tjia, 2007).

## 3. MATERIALS AND METHOD

### 3.1 Overall Methodology

Overall, the first phase concentrates on the data acquisition and pre-processing of high resolution remotely sensed data, field data and ancillary data. The remotely sensed data, i.e., airborne LiDAR, aerial photos and high-resolution satellite images were geo-rectified into the same coordinate system. The point clouds obtained from the airborne LiDAR data were filtered to remove the non-ground points. The ground points were interpolated to produce Digital Terrain Model (DTM). Digital Surface Model (DSM) was produced by interpolating the non-ground points. The normalised point clouds were generated by normalising the original point clouds based on DTM. The field data included data collection of landslide inventory and vegetation characteristics. Intensive literature review was done on the vegetation anomalies that could be used to describe different landslide activities in tropical regions. The second stage focused

on the development of landslide inventory using manual interpretation of high resolution remotely sensed data. The landslide inventory map was validated using field-based landslide technique. The third stage emphasised on the development of VAIs using high resolution remote sensing data. The VAI can be grouped into 7 main groups with five different spatial resolution (i.e. 0.5m, 1m, 5m, 10m, and 20m): 1) tree height irregularities; 2) canopy gap; 3) density of different layer of vegetation; 4) vegetation type; 5) vegetation indices; 6) root strength index (RSI); and 7) distribution of water-loving trees. All the VAI maps and landslide inventory maps obtained from the manual interpretation were used in the fifth stage that aimed at classifying the landslide activities based on different landslide types using Support Vector machine approach. Finally, the classified activities of landslide were evaluated using several performance indicators.

### 3.2 Data acquisition and Pre-processing

**3.2.1 Airborne LiDAR Data:** Airborne LiDAR data were obtained by the Integrated Geospatial Innovations (IGI) Lite Mapper 6800-400 mounted onto a helicopter. The data capturing process was done over the steep debris flow terrain along the Mesilau River using multiple returns setting of laser scanning. The data were collected in 2014 for about two months after the debris flow that reached the Mesilau River. The LiDAR system utilised RIEGL LMS-Q680i laser scanner system that is capable of generating point clouds based on full waveform analysis. This scanner provides point clouds with the accuracy and precision of 20 mm, a maximum range of 3,000 m, and measurement up to 266,000 pulses per second on the ground. The final average point density ( $m^2$ ) of the airborne LiDAR data was 160 points per  $m^2$  Table 1 shows the complete specifications of the IGI Lite Mapper 6800-400 with RIEGL LMS-Q680i laser scanning system. The airborne LiDAR data were filtered using the Triangular Irregular Network (TIN) densification method by Axelsson (2000) to separate ground and non-ground points. The DTM of the study area was generated by interpolating the ground points to 0.5 m spatial resolution. The non-ground points were interpolated to produce the DSM. The Canopy height model (CHM) was produced by subtracting DTM from DSM. Finally, normalised point clouds were produced by subtracting the DTM value from the height information in each point.

**3.2.2 Satellite Images:** Pleiades high resolution satellite image consisted of four spectral bands (blue, green, red, and NIR) with a spatial resolution of 2 m and 0.5 m for multispectral and panchromatic images, respectively. The multispectral images were spatially fused to the panchromatic image to produce high spatial resolution multispectral image. Both airborne LiDAR and Pleiades image were georeferenced to the same coordinate system.

**3.2.3 Landslide Inventory:** Landslide inventory was made based on the combination of remote sensing data interpretation and field observation. The remote sensing-based landslide inventory was done using several airborne LiDAR-derived datasets, i.e., topographic openness, hillshade, and colour composite. These datasets were generated using DTM and orthophoto with 7 cm spatial resolution. Topographic openness was used to observe the clear contrast between the planar surface and down-slope surface (Razak, 2014). It can be one of the landslide features that can be recognised through colour ramp observations.

Orthophoto image is capable of exhibiting any evidence of landslide occurrence either in a clear form of landslide polygon or other evidence leading to landslides, for example road crack, soil erosion, and bad surface or ponding in niches or back tilting area.

Next, the inventory process continued with landslide delineation. The delineation process was based on a scarp and accumulation area where the scarp area covered from the crown part until the head, while the accumulation area covered from the main body until toe. The signatures which related to morphology, vegetation, and drainage characteristics would be used in classifying landslide type and activity. For active landslides, the scarp and crown are obvious, and the accumulation zone is visible. The area is mostly characterised by less or almost no vegetation cover. A dormant landslide is an inactive landslide that may be reactivated due to its original causes or other causes. The scarp and body of the failure are still visible in the LiDAR image with low density vegetation and does not show any recent activity. A relict landslide is an inactive landslide, where the slope has been stabilised due to natural factors or mitigation measures. The scarp and body of the landslide area are not distinguishable through LiDAR image. Meanwhile the area is densely covered by vegetation or has been mitigated by any structural measures. The results of remote sensing-based landslide inventory were verified and corrected using the field-based inventory.

### 3.3 Parameterization of vegetation anomalies indicators (VAIs) using remotely sensed data

In this study, seven groups of VAIs were used in landslide activity modelling i.e., tree height irregularities, canopy gap, different layers of vegetation, vegetation type distribution, vegetation indices, root strength index, and distribution of water-loving trees. All the VAIs were derived from airborne LiDAR and high-resolution satellite images. The next section describes the detail explanation for each of the VAI.

**3.3.1 Tree Height Irregularities:** Tree height is defined as the perpendicular distance between the top and base of a tree (Verma et al., 2016). The height value varies across different stands and/or species. The distribution of the tree height significantly reflects the quality and quantity of tree stand and its future growth. Trees in landslide areas have relatively low height, small crown, and more irregularities (Mohd Salleh et al., 2019). Tree height irregularities were calculated using the standard deviation of tree height within a selected area size (i.e., a grid of 10 m). This calculation was applied to the LiDAR-derived CHM of the study area. High standard deviation indicates highly irregular tree height in a specific area.

$$\text{Tree height irregularities} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (1)$$

where  $x_i$  is individual height value of the pixel,  $\mu$  is mean height value within the 10 m grid.

**3.3.2 Canopy Gap:** Canopy gap is an empty area within forest canopies caused by natural disturbances where these areas can be filled by other trees (Muscolo et al., 2014). A strong relationship can be found between landslide and forest canopy gaps (Moos, 2014). The presence of landslides under the forested area is believed to be detectable by measuring the gaps in the area. Production of tree canopy gap involve several important steps including separation of forested areas from non-forested areas, individual tree crown delineation, and tree canopy gap generation. The separation of the forested and non-forested area was performed by manual identification from Pleiades images. The delineation of individual tree crowns was done from the CHM layer within the forested area using the inverse watershed segmentation method (Rahman and Gorte, 2009). The tree canopy layer was converted into grid format and the density of tree canopy gap was calculated within a grid of 10 m using Equation (2). An area with a low number of canopy pixel was considered as having high density of canopy gap.

$$\text{Canopy gap (\%)} = \frac{\sum \text{Pixel}_{\text{non-canopy}}}{\sum \text{Pixel}_{\text{Total area}}} \times 100 \quad (2)$$

**3.3.3 Density of Vegetation Layers:** The vegetation density layer at a certain height from the ground was measured using the density of vegetation points method (Rahman and Gorte, 2009). This method utilised point clouds obtained from the airborne LiDAR data and the density of the reflected laser pulses above a certain height from the ground. The vegetation layers were then classified into four classes i.e., low vegetation, young woody vegetation, matured woody vegetation, and old forest. The process started by calculating the height of non-ground point clouds also known as normalized point clouds. The normalized point clouds were then categorised based on height classification scheme and the vegetation density of each layer class was calculated with a final resolution of 0.5 m.

**3.3.4 Vegetation Type Distribution:** According to Malaysian Standard 1759 (MS 1759:2004), many categories of vegetation type can be classified. In the current study, four vegetation types were used, such as grass, secondary forest, primary forest, and agriculture. Landslide occurrence was higher in logging areas, secondary forest and settlement areas, and lower in primary forest areas (Shahabi and Hashim, 2015). This indicator was mapped from satellite imagery and LiDAR-CHM.

**3.3.5 Vegetation Indices:** Vegetation indices can be defined as the form of band ratios related to vegetation. Vegetation indices can be used in the identification of landslide. Six indices were used in this study, namely NDVI, DVI, SAVI, OSAVI, GDVI, and GNDVI. The vegetation indices were extracted from the Pleiades satellite images. These indices were derived from mathematical expressions based on the combination of visible light bands and Near Infrared (NIR) band as shown below:

$$\text{NDVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R})} \quad (3)$$

$$\text{DVI} = \text{NIR} - \text{R} \quad (4)$$

$$\text{SAVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R} + \text{L})} \times (1 + \text{L}) \quad (5)$$

$$\text{OSAVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R} + \text{L})} \times (1 + 0.16) \quad (6)$$

$$\text{GDVI} = \text{NIR} - \text{G} \quad (7)$$

$$\text{GNDVI} = \frac{(\text{NIR} - \text{G})}{(\text{NIR} + \text{G})} \quad (8)$$

**3.3.6 Root Strength Index:** Root strength is one of the factors of soil reinforcement (Abdi, 2018). Increasing the strength will increase the soil reinforcement. Changing the tree root strength affects slope stability. Root strength index (RSI) was derived based on the estimated tree height and tree density. The estimated tree height was obtained from the airborne LiDAR data, while tree density was defined as the number of trees in a 30 m grid. Equation (8) shows the product of tree height and the square root of estimated tree density in calculating the RSI.

$$\text{RSI} = H \times \sqrt{D} \quad (9)$$

where  $H$  is the tree height, and  $D$  is the estimated tree density.

**3.3.7 Water-Loving Trees:** Water-loving trees, also known as hydrophytes, are usually found in wetlands of all sorts, either in or on the water, or where soils are flooded or saturated long enough to establish anaerobic conditions in the root zone (Cronk and Fennessy, 2016). Distribution of water-loving trees indicates the high presence of active landslides. In this study, the distribution of water-loving trees was produced by combining the aerial photographs, satellite imagery, CHM, and topographic wetness index (TWI) datasets, which were derived from the high resolution DTM over the study area. Water loving trees were characterised by their lower height value, i.e., 1 m to 3 m. The vegetated pixel area was classified automatically from satellite imagery. The height value of vegetation pixel was extracted from the CHM dataset and the pixel value of 1 m to 3 m was used for estimation of water loving trees. The presence of low vegetation and high TWI value for certain area indicates a high density of water-loving trees.

### 3.4 Support Vector Machine (SVM) for Landslide Activity Classification

SVM was initially developed to find a hyperplane that separates two classes optimally (i.e., landslides with a specific activity class and other landslides with activity class) (Vapnik, 2013) by maximising the margins of class boundaries for linearly separable cases]. The optimum hyperplane was derived from support vectors with the closest values to the classification margin. The classification of new data can be performed once the decision surface is acquired. In this study, given a training set of instance-label pairs  $(x_i, y_i)$ ,  $i = 1, \dots, m$  where  $x_i \in R^n$  and  $y_i \in \{1, -1\}$ ,  $x_i$  is a vector of input space that represent all the VAI such as tree height irregularities, canopy gap, and vegetation indices. Meanwhile,  $\{1, -1\}$  represents landslide and non-landslide pixels, which are categorised into different state of activities (i.e., active, dormant, and relict). In the SVM classification approach, several parameters needed to be configured such as kernel function, regularisation parameter  $C$ , gamma ( $\gamma$ ), and degree of polynomial ( $d$ ). The parameters are crucial to find a model that performs a good classification task. Therefore, the selection of parameter value of  $C$ ,  $\gamma$ , and  $d$  are defined by applying Grid Search (GS) and Genetic Algorithm (GA) techniques. The following pseudocode demonstrates the implementation of GS-SVM and GA-SVM techniques in SVM parameter optimization.

---

**Algorithm GS-SVM**

- 1 **Input:** A training set:  $T$ , Kernel parameter set:  $I$ , Penalty set:  $\{C_1, C_2, \dots, C_k\}$ , Gamma set:  $\{\gamma_1, \gamma_2, \dots, \gamma_k\}$ , Degree set:  $\{d_1, d_2, \dots, d_k\}$
- 2 where  $C_1 < C_2 < \dots < C_k$ ,  $0 < \eta < 1$
- 3 **Output:** Optimal parameters:  $\gamma^*$ ,  $C^*$  and the final model:  $M$
- 4 Initialize the best cross validation accuracy,  $CV = 0$
- 5 **for each** ( $y \in I$ ) **do**
- 6   **for**  $C$  in  $[2^{-5}, 2^{-3}, \dots, 2^5]$
- 7     **for**  $\gamma$  in  $[2^{-5}, 2^{-3}, \dots, 2^5]$
- 8       Train an SVM model,  $M$ , using  $\gamma$  and  $C$ , with  $T$
- 9        $SV_i$  = Support vectors of  $M$ ,
- 10        $CV_i$  = cross validation accuracy for  $M_i$
- 11       **if**  $CV_i > CV$  **then**
- 12           $CV = CV_i$ ;
- 13           $\gamma^* = \gamma$ ;
- 14           $C^* = C_i$ ;
- 15       **end if**
- 16     **end for**
- 17   **end for**
- 18 **end for**
- 19 Train an SVM model,  $M$  using  $\gamma^*$  and  $C^*$  with  $T$

---



---

**Algorithm GA-SVM**

- 1 **Input:** A training set:  $T$ ;  $t_{max}$ : maximal number of iterations;  $M_p$ : mutation probability;  $C_p$ : crossover;  $N_p$ : The size of the search population
- 2 **Output:** Optimal parameters:  $\gamma^*$ ,  $C^*$  and the final model:  $M$
- 3 **initialize**
- 4  $N_p$ : random sample individuals within the search size;
- 5  $t$ : current iteration
- 6  $r_1, r_2$ : randomly selected solutions
- 7 **while**  $t \leq t_{max}$  **do**;
- 8   **select** chromosomes by selection operation;
- 9   **for** chromosomes( $i$ ) in  $N_p$  **do**

---

- 10   **if**  $r_1 \leq C_p$  **then**
- 11     crossover operation;
- 12   **end if**
- 13   **if**  $r_2 \leq M_p$  **then**
- 14     mutation operation;
- 15   **end if**
- 16   **end for**
- 17 **update** population
- 18 **end while**
- 19 **return** optimal solution
- 20 Train an SVM model,  $M$  using  $\gamma^*$  and  $C^*$  with  $T$

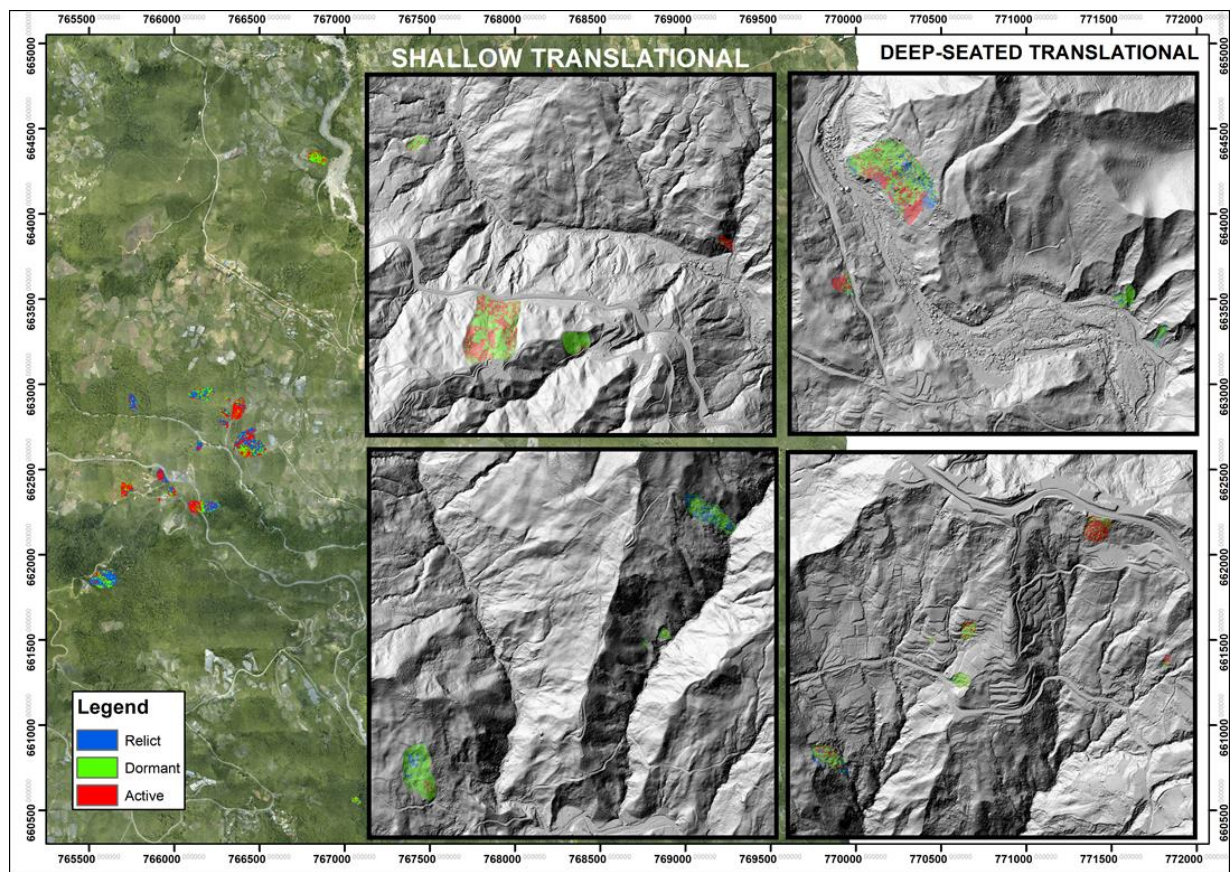
---

### 3.5 Accuracy Assessment

The measures of overall accuracy (OA) and kappa coefficient (Kappa) were used to assess the performance of the SVM algorithm in classifying the landslide activity based on VAIs. OA is the percentage of correctly classified samples. Kappa is a widely-used and powerful multivariate technique for accuracy assessment. Kappa coefficient is a measure of overall statistical agreement of a confusion matrix. Therefore, it provides a more rigorous assessment of classification accuracy.

## 4. RESULTS AND DISCUSSION

The results of this study cover two landslide depths: (i) deep-seated, and (ii) shallow translational landslides. Overall, 54 deep-seated translational landslide was successfully delineated while 149 for shallow translational. Figure 2 illustrates the distribution of some deep-seated and shallow translational landslides with specific activity over the study area.

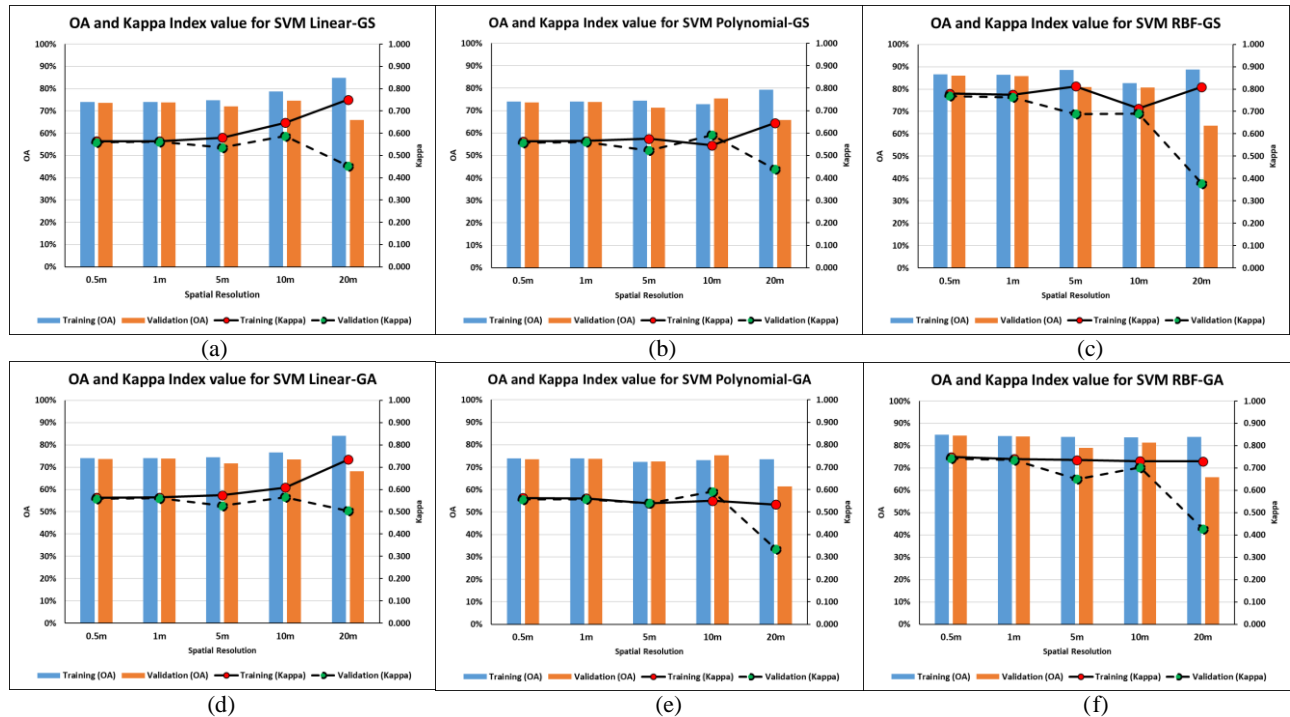


**Figure 2.** Distribution of translational landslide with classified activity from the modelling results

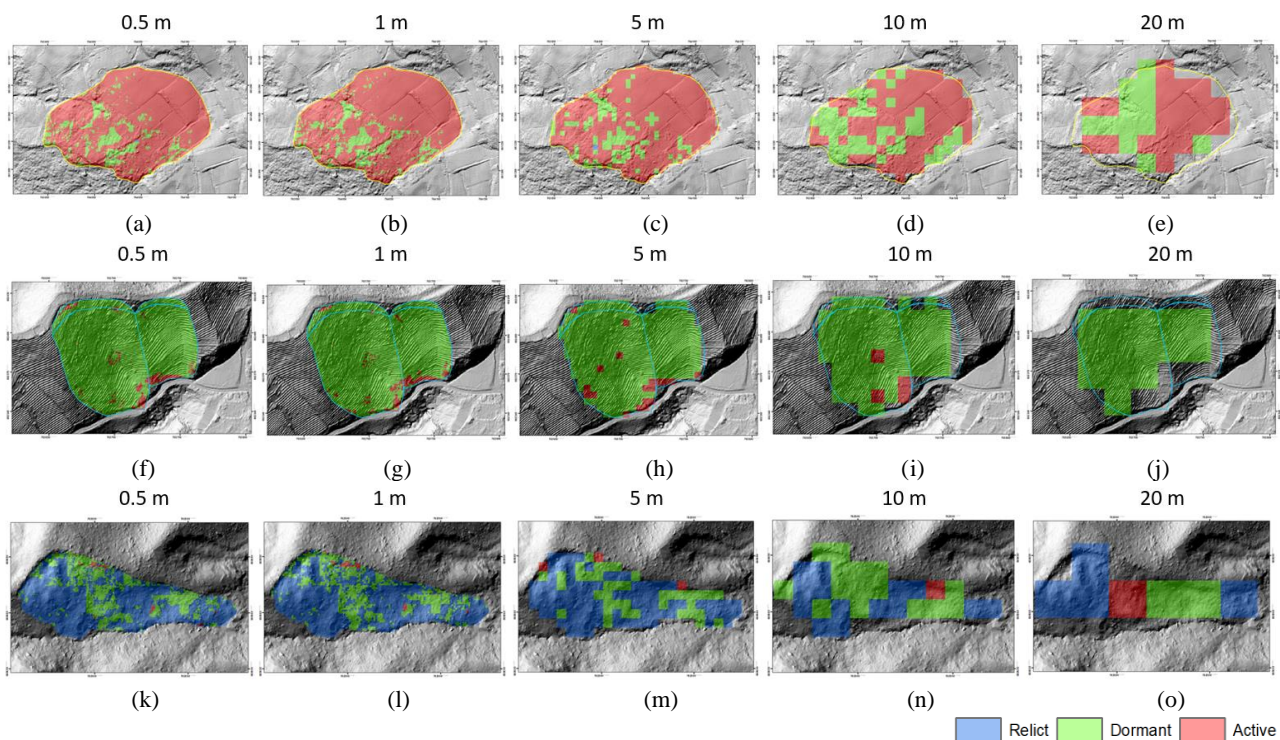
#### 4.1 Deep-Seated Translational Landslide

Figure 3 shows the overall accuracy and kappa coefficients of the classified activity for deep-seated translational and the comparison between different spatial resolution. The overall accuracy of the training dataset is between 72.3–88.7%, kappa is

0.534–0.812, and the overall accuracy of the validation dataset is between 61.4–86%, and kappa is between 0.335–0.769. Sample visualisation comparison of classified landslide with different spatial resolution can be found in Figure 4.



**Figure 3.** Overall accuracy and kappa coefficient of classified landslide activity for deep seated, translational landslide with different spatial resolution using (a) SVM Linear-GS, (b) SVM Polynomial-GS, (c) SVM RBF-GS, (d) SVM Linear-GA, (e) SVM Polynomial-GA, and (f) SVM RBF-GA

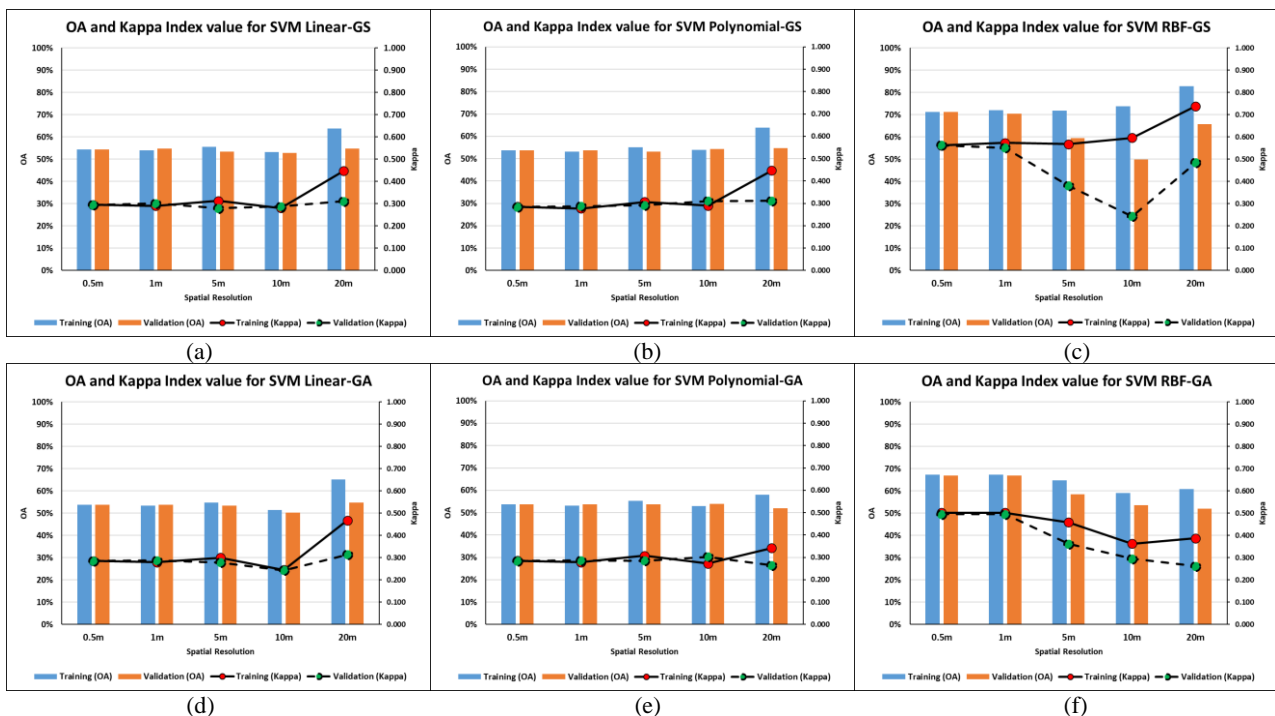


**Figure 4.** Sample of classified landslide activity for deep-seated translational landslide with different spatial resolution using SVM RBF-GS method where, (a) – (e) for active, (f) – (j) for dormant, and (k) – (o) for relict landslides.

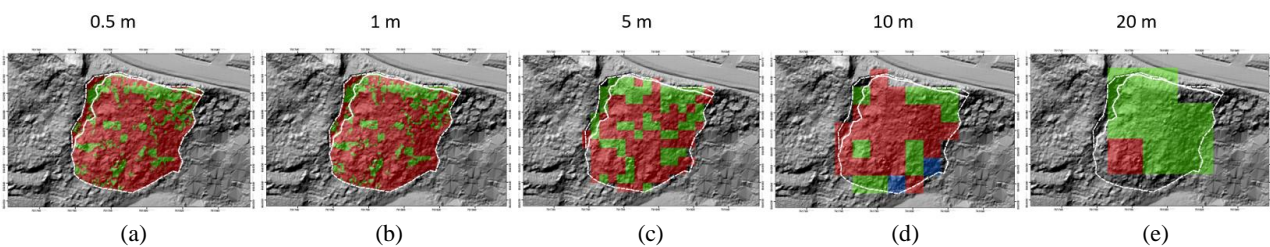
Based on the results, SVM RBF-GS with 0.5m spatial resolution produced highest overall accuracy (86%) and kappa coefficient (0.76) values for validation dataset while SVM Polynomial-GA produced the lowest overall accuracy (61.4%) and kappa coefficient (0.335) values with 20m spatial resolution. The difference of 24.6% indicates the effect of spatial resolution, where decreasing the cell size producing higher accuracy of classified data. However, there is no significant increase of overall accuracy for spatial resolution in between 0.5m to 10m since small increasing values were recorded. The lowest overall accuracy recorded by all the methods under 20m resolution, particularly for validation dataset indicates that coarse resolution tend to generalize the pixel value in the landslide location. For active landslide, combination of active and dormant classified pixel were observed from the results especially for 10m and above, meanwhile most of dormant landslide was correctly classified for all spatial resolution. For relict landslide, combinations of dormant and relict pixels were clearly observed since it share similar vegetation characteristics.

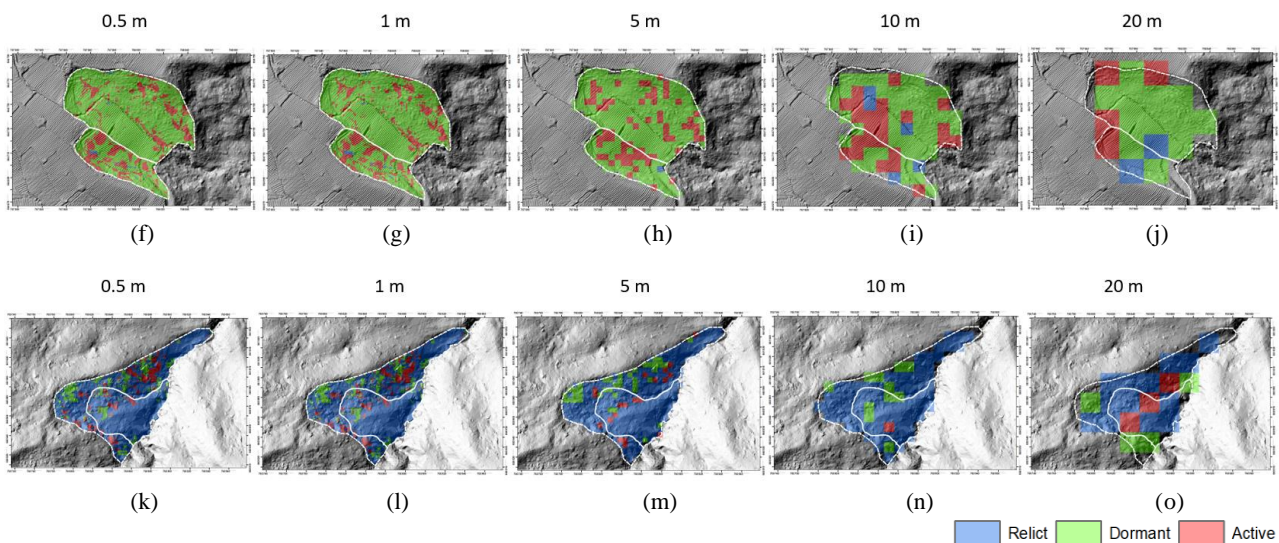
#### 4.2 Shallow Translational Landslide

For shallow translational landslide, the overall accuracy of the training dataset is between 51.4–82.8%, kappa is 0.245–0.738, and the overall accuracy of the validation dataset is between 49.8–71.3%, and kappa is between 0.243–0.563. As shown in Figure 5, SVM RBF-GS with 0.5m spatial resolution recorded highest overall accuracy (71.3%) and kappa coefficient (0.563) values on validation dataset while SVM RBF-GS also with 10m resolution obtained the lowest overall accuracy (49.8%) and kappa (0.243) values. Similar to deep-seated category where the difference of 21.5% indicates higher spatial resolution give high reliable results. This can be seen in Figure 6 where high spatial resolution produced more accurate classified pixel. For active landslide, combination of active and dormant pixel were found which similar to dormant landslide. Meanwhile, combinations of dormant and relict pixels were clearly observed since it share similar vegetation characteristics.



**Figure 5.** Overall accuracy and kappa coefficient of classified landslide activity for shallow, translational landslide with different spatial resolution using (a) SVM Linear-GS, (b) SVM Polynomial-GS, (c) SVM RBF-GS, (d) SVM Linear-GA, (e) SVM Polynomial-GA, and (f) SVM RBF-GA





**Figure 6.** Sample of classified landslide activity for shallow translational landslide with different spatial resolution using SVM RBF-GS method where, (a) – (e) for active, (f) – (j) for dormant, and (k) – (o) for relict landslides

## 5. CONCLUSION

Accurate landslide activity map is necessary to ensure a complete and good quality of landslide inventory process. This consequently increases the reliability of landslide susceptibility, hazard, and risk mapping based on the geospatial approach. Previous studies concentrated on utilising morphological drainage pattern interpretations for landslide activity mapping. In this study, VAI obtained from remotely sensed data proved to be very useful as a bio-indicator for landslide activity determination. The classification process revealed that VAIs are significant in determining landslide activity as they give reliable

## ACKNOWLEDGEMENTS

This research was funded by the Universiti Teknologi Malaysia, Ministry of Higher Education of Malaysia, FRGS grant vot number R.J130000.7852.5F396, Zamalah/ Institutional Scholarship, and Minerals and Geoscience Department Malaysia.

## REFERENCES

Abdi, E. (2018). Root tensile force and resistance of several tree and shrub species of Hyrcanian forest, Iran. *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, 39(2), 255-270.

Artese, S., & Perrelli, M. (2018). Monitoring a landslide with high accuracy by total station: a DTM-based model to correct for the atmospheric effects. *Geosciences*, 8(2), 46.

Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. *International archives of photogrammetry and remote sensing*, 33(4), 110-117.

Berardino, P., Fornaro, G., Lanari, R., & Sansosti, E. (2002). A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on Geoscience and remote sensing*, 40(11), 2375-2383.

results. The ability of machine learning algorithms to automate in analysing the datasets (i.e., VAI) and identify the relationship between variables increases the effectiveness of landslide activity classification. The effect of spatial resolution on the classified results can be seen for all activities. High resolution data give more accurate results rather than low resolution. Therefore, the resulting landslide activity maps may provide powerful tools for landslide hazard assessment and could assist decision-makers during site selection and planning processes. This product could also be used as a basis for landslide inventory mapping, especially for a forested area.

Cronk, J. K., & Fennessy, M. S. (2016). *Wetland plants: biology and ecology*: CRC press.

Dehnavi, A., Aghdam, I. N., Pradhan, B., & Varzandeh, M. H. M. (2015). A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena*, 135, 122-148.

Gaidzik, K., Ramírez-Herrera, M. T., Bunn, M., Leshchinsky, B. A., Olsen, M., & Regmi, N. R. (2017). Landslide manual and automated inventories, and susceptibility mapping using LIDAR in the forested mountains of Guerrero, Mexico. *Geomatics, natural hazards and risk*, 8(2), 1054-1079.

Harker, R. I. (1996). Curved tree trunks: indicators of soil creep and other phenomena. *The Journal of Geology*, 104(3), 351-358.

Highland, L., & Bobrowsky, P. T. (2008). *The landslide handbook: a guide to understanding landslides*: US Geological Survey Reston.

Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R., & Pedrazzini, A. (2012). Use of LIDAR in landslide investigations: a review. *Natural hazards*, 61(1), 5-28.

Komoo, I., & Lim, C. (2003). Kompleks Gelinciran Tanah Kudasang: Pemetaan terperinci di kawasan Sekolah Menengah



- Kebangsaan Kundasang (IN MALAY), The Kundasang Landslides Complex: A detailed mapping of the Kundasang National Secondary School area.
- Łuszczynska, K., Wistuba, M., & Malik, I. (2017). Dendrochronology as a source of data for landslide activity maps—an example from Beskid Żywiecki Mountains (Western Carpathians, Poland). *Environmental & Socio-economic Studies*, 5(3), 40-46.
- Mezaal, M. R., Pradhan, B., Sameen, M. I., Mohd Shafri, H. Z., & Yusoff, Z. M. (2017). Optimized neural architecture for automatic landslide detection from high-resolution airborne laser scanning data. *Applied Sciences*, 7(7), 730.
- Mohd Salleh, M., Ismail, Z., Mohd Ariff, S., Abd Rahman, M. Z., Abdul Khanan, M., Asmadi, M. A., & Razak, K. A. (2019). Spatial Prediction Models for Landslide Activity Mapping Using Vegetation Anomalies. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Moos, C. (2014). *How does forest structure affect landslide susceptibility? Statistical prediction models for shallow landslides integrating forest structure*. Master Thesis. ETH Zurich,
- Morpi, M. (2011). Proposal to check landslides in Kundasang. *Borneo Post*. Retrieved from <https://www.theborneopost.com/2011/05/07/proposal-to-check-landslides-in-kundasang/>
- Musco, A., Bagnato, S., Sidari, M., & Mercurio, R. (2014). A review of the roles of forest canopy gaps. *Journal of Forestry Research*, 25(4), 725-736.
- Pradhan, B., Jebur, M. N., Shafri, H. Z. M., & Tehrani, M. S. (2015). Data fusion technique using wavelet transform and Taguchi methods for automatic landslide detection from airborne laser scanning data and quickbird satellite imagery. *IEEE Transactions on Geoscience and remote sensing*, 54(3), 1610-1622.
- Rahman, M., & Gorte, B. (2009). *Tree crown delineation from high resolution airborne lidar based on densities of high points*. Paper presented at the Proceedings ISPRS Workshop Laserscanning 2009, September 1-2, France, IAPRS, XXXVIII (3/W8), 2009.
- Razak, K. A. (2014). *Airborne laser scanning for forested landslides investigation in temperate and tropical environments*. (Doctor of Philosophy). University of Twente, Netherlands.
- Razak, K. A., Bucksch, A., Damen, M., van Westen, C., Straatsma, M., & de Jong, S. (2013). Characterizing tree growth anomaly induced by landslides using LiDAR. In *Landslide science and practice* (pp. 235-241): Springer.
- Sanderson, D., & Sharma, A. (2016). IFRC World Disasters Report, Resilience: Saving lives today, investing for tomorrow. *International Federation of Red Cross and Red Crescent Societies (IFRC)*.
- Scaioni, M., Longoni, L., Melillo, V., & Papini, M. (2014). Remote sensing for landslide investigations: an overview of recent achievements and perspectives. *Remote Sensing*, 6(10), 9600-9652.
- Schuster, R. L. (1996). *Landslides: Investigation and Mitigation. Chapter 2-Socioeconomic Significance of Landslides*.
- Shahabi, H., & Hashim, M. (2015). Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. *Scientific reports*, 5(1), 1-15.
- Sharir, K., Roslee, R., Ern, L. K., & Simon, N. (2017). Landslide factors and susceptibility mapping on natural and artificial slopes in Kundasang, Sabah. *Sains Malaysiana*, 46(9), 1531-1540.
- Stefanini, M. C. (2004). Spatio-temporal analysis of a complex landslide in the Northern Apennines (Italy) by means of dendrochronology. *Geomorphology*, 63(3-4), 191-202.
- Tating, F., Hack, H., & Jetten, V. (2013). *Weathering and deterioration as quantitative factors in slope design in humid tropical areas: case study Northern Kota Kinabalu, Sabah, Malaysia*. *Engineering in Exotic Environments*, (33), 1).
- Tien Bui, D., Shahabi, H., Shirzadi, A., Chapi, K., Alizadeh, M., Chen, W., . . . Hong, H. (2018). Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia. *Remote Sensing*, 10(10), 1527.
- Tiranti, D., & Cremonini, R. (2019). Landslide Hazard in a Changing Environment. *Frontiers in Earth Science*, 7, 3.
- Tjia, H. (2007). Kundasang (Sabah) at the intersection of regional fault zones of Quaternary age.
- Tongkul, F. (2017). Active tectonics in Sabah—seismicity and active faults. In (Vol. 64, pp. 27-36): Bulletin of the Geological Society of Malaysia.
- Tsai, Z.-X., You, G. J.-Y., Lee, H.-Y., & Chiu, Y.-J. (2012). Use of a total station to monitor post-failure sediment yields in landslide sites of the Shihmen reservoir watershed, Taiwan. *Geomorphology*, 139, 438-451.
- Van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering geology*, 102(3-4), 112-131.
- Vapnik, V. (2013). *The nature of statistical learning theory*: Springer science & business media.
- Varnes, D. J. (1978). Slope movement types and processes. *Special report*, 176, 11-33.
- Verma, N. K., Lamb, D. W., Reid, N., & Wilson, B. (2016). Comparison of canopy volume measurements of scattered eucalypt farm trees derived from high spatial resolution imagery and LiDAR. *Remote Sensing*, 8(5), 388.

Wang, D., Hollaus, M., Puttonen, E., & Pfeifer, N. (2016). Automatic and self-adaptive stem reconstruction in landslide-affected forests. *Remote Sensing*, 8(12), 974.

Wang, D., Hollaus, M., Schmaltz, E., Wieser, M., Reifeltshammer, D., & Pfeifer, N. (2016). Tree stem shapes derived from TLS data as an indicator for