Unsupervised Classification of Earth Surface for Landslide Detection

Caitlin Tran[1], Omar E. Mora[1], Jessica Fayne[2]

[1]*California State Polytechnic University, Pomona, Department of Civil Engineering*
Pomona, CA 91768

[2]*University of Los Angeles, California, Department of Geography*
Los Angeles, CA 90095

Core Ideas

1. DEMs and unsupervised classification are used to predict areas that are susceptible to landslides
2. Feature extractors allowed for individual analysis of specific surface features
3. Unsupervised classification is performed on each surface feature using multiple number of clusters.
4. Clusters are gathered to create a landslide prone class and non-landslide prone class.
5. Comparison of documented landslide prone regions to algorithm mapped regions shows that algorithmic classification can accurately identify areas susceptible to landslides.

Abstract

Landslides are geological events in which masses of rock and soil slide down the slope of a mountain or hillside. They are influenced by topography, geology, weather and human activity, and can cause extensive damage to the environment and infrastructure, as well as delay transportation networks. Therefore, it is imperative to detect early-warning signs of landslide hazards as a means of prevention. Traditional landslide surveillance consists of field mapping, but the process is costly and time consuming. Modern landslide mapping uses Light Detection and Ranging (LiDAR) derived Digital Elevation Models (DEMs) and sophisticated algorithms to analyze surface roughness and extract spatial features and patterns of landslide and landslide-prone areas. In this study, a methodology based on k-means clustering and Gaussian Mixture Model (GMM) tested several feature extractors and employed an unsupervised classifier at the Carlyon Beach Peninsula in the state of Washington to attempt to distinguish between slide and non-slide terrain. When compared with the detailed, independently compiled landslide inventory map, our algorithms correctly classify up to 87% of the terrain in our study area. These results suggest that the proposed methods and LiDAR-derived DEMs can provide important surface information and be used as efficient tools for digital terrain analysis to create accurate landslide maps.

Key words: DEM; landslide; detection; feature extraction; LiDAR; k-means clustering; Gaussian Mixture Model;

Introduction

Landslides are a form of mass wasting that includes a wide range of ground movements, such as rock falls, deep failure of slopes, and shallow debris flows (Hala Adel Effat, 2014). They are caused by pre-conditional surface and/or sub-surface instability that occurs after slope changes, rainfall or land topography manipulation (Dalyot *et al*., 2008). Thus, landslides usually occur on steep slopes of mountains or hills (McKean and Roering, 2004). They are also influenced by natural factors and human activities such as topography, weather, mining, and construction. Landslides pose a global threat as they affect human conditions and extensively damage the

environment and infrastructure; they are natural disasters for highways, buildings, residential developments, bridges, and other human-inhabited areas (Mora *et al*., 2015). Therefore, it is imperative to prevent future risk posed by these events and accurate landslide mapping is essential in the mitigation of the effects caused by these hazardous events.

Traditional landslide mapping methods include aerial photography, field inspection and contour mapping (McKean and Roering, 2004; Glenn *et al*., 2006; Booth *et al*., 2009; Mora *et al*., 2015). However, despite the accuracy and precision of these surveillance methods, topographic maps produced by contour mapping do not have enough spatial resolution for small scale features and they cannot penetrate highly vegetated areas (Booth *et al*., 2009; Mora *et al*., 2015). As well, they have proven to be time consuming, costly, difficult to implement, subjective, and unable to detect small-scale failures (Booth *et al*., 2009; Mora *et al*., 2015).

These setbacks in traditional landslide mapping have made them less effective than modern mapping techniques. Newer landslide mapping methods, that utilize remote sensing, have shown incredible advancements in spatial resolution and accessibility in the past decade. Recent developments in topographic mapping techniques, such as laser scanners, have also made it easier to construct intricate models of the surface features that can help detect landslide susceptible areas (Tarolli *et al*., 2012). LiDAR measurement has improved from a resolution greater than 10 meters to less than 1 meter (Mora *et al*., 2015), thus allowing for the detection of finer spatial features. In particular, airborne LiDAR can penetrate vegetative cover, with filtering processes to reveal the bare earth terrain and map up to thousands of square meters (Booth *et al*., 2009; Shan and Toth, 2008). These technologies allow for the mapping of small failures in areas of slow mass movement and identification of landslide/non-landslide terrain (Jaboyedoff, 2012). They also provide great opportunities for studies about earth surfaces and landslide susceptibility (Shan and Toth, 2008; Jaboyedoff *et al*. 2012). The advancements in modern landslide mapping have proven to be highly accurate, cost effective, and accessible for the community.

The use of 3D data such as DEMs in landslide mapping has improved the safety of infrastructures and reduced the hazard caused by landslides. To further analyze landslide occurrences, studies have concentrated not only on the slides, but on different aspects of the area such as water flow, rock, and sediment types, as these can affect the stability of the terrain (Leshchinsky *et al*., 2015, Mora *et al*., 2015; An *et al*., 2016). Using these newer technology-based methods has made it easier and more cost efficient to consider and analyze the different aspects of the terrain. Feature extraction is used on DEMs to identify and isolate landslides by examining the surface models (Mora *et al*., 2015; Jaboyedoff *et al*., 2012; Glenn *et al.,* 2006). Various feature extraction models have been used in landslide studies including slope, hillshade, aspect, and statistics (Shan and Toth, 2008). A few of these features have been tested through automated algorithms (Cheng *et al.,* 2013; Hölbing *et al*., 2015). For example, the statistics feature has been used in a very specific length scale to compute roughness in order to isolate the slide area (McKean and Roering, 2004). While such features are a key method to identifying landslide terrain, studies suggest that the type of features used to identify slides can be subjective to each site (Dou, *et al*., 2015). Different features are able to characterize and map landslide and non-landslide areas. Knowing the quantity of vegetation, in addition to the features already included within the site, is a crucial component that enables the use of these features with DEMs (Sarkar *et al*., 2008).

In this paper, two unsupervised classification methods are tested and compared to detect and map landslide and non-landslide terrain utilizing an airborne LiDAR-derived DEM. The unsupervised techniques based on k-means clustering and GMM employ feature extractors to evaluate the surface topography, specifically: slope, roughness, local topographic range and local topographic variability. The methods evaluated fuse feature extraction and unsupervised classification to map areas of potential landslide activity. To evaluate the performance of the proposed algorithms, a study area in the Carlyon Beach Peninsula in the state of Washington

was used for testing. The resulting landslide maps were compared with an independently compiled landslide inventory map to correctly identify up to 87% of the terrain in the study area. These results indicate that the proposed algorithms and LiDAR-derived DEMs provide important surface information for the detection of landslide surface features that can cause continuous damage to inhabited areas.

Study Area & Data

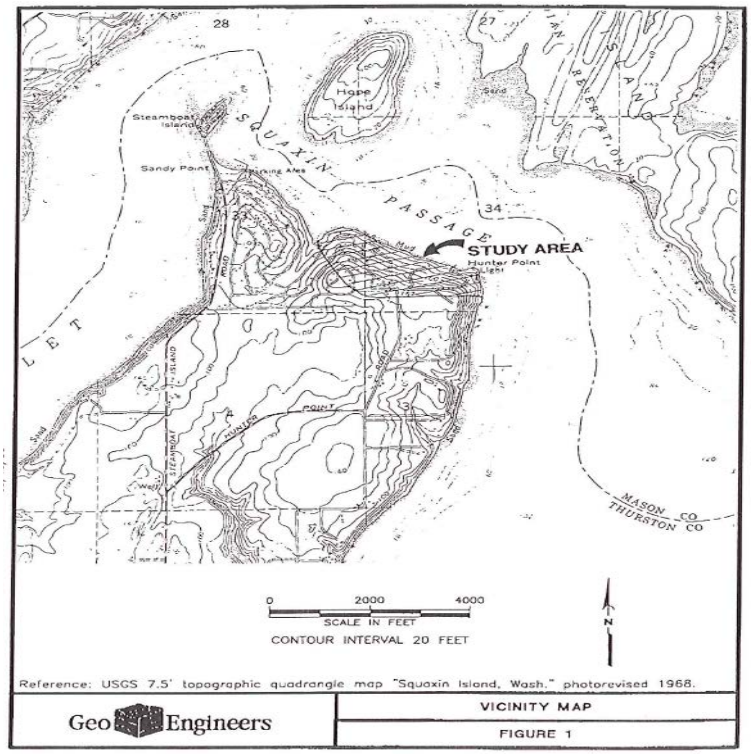

Figure 1. *Vicinity map of study area in Carlyon Beach, Washington.*

The study area is the Carlyon Beach/Hunter Point landslide in northwestern Thurston County, Washington (Approx. Latitude: N 47° 10' 46", Longitude: W 122° 56' 24"), see Figure 1. The general location resides at an elevation of 165 feet above mean sea level with slopes ranging from 7 to 20 degrees. Carlyon Beach has a sparsely developed amount of vegetation that consists of sub-mature, second growth coniferous trees, deciduous trees, and sword fern that play a crucial role in deeper-seated slope movement (GeoEngineers, 1999).

The landslide is located along the northern end of the Steamboat Island Peninsula and includes a portion of the private community of Carlyon Beach, as well as a number of rural residential dwellings along Northwest Hunter Point Road (GeoEngineers, 1999). Landslide movement was first detected in early 1999 from cracks in various residences and city infrastructures, distressed building foundations, and damaged subsurface utilities.

In addition, the area has experienced a continuous amount of movement due to the soil composition. Soil boring was performed in the area to analyze soil composition. Engineers concluded that the area resides on unstable soils including soft silt, stiff silt, and clay, which are heavily affected by weather, construction, and the environment itself (GeoEngineers, 1999).

Carlyon Beach failures can also be somewhat attributed to a shallow groundwater table that proves detrimental to slope stability when compounded with the weak overlying soil. The Carlyon Beach Peninsula experiences above average rainfall, between 3 and 65 percent above

the average (GeoEngineers Phase II, 1999), during the winter season, contributing an increased amount of ground water flow to the area and resulting in a higher landslide probability (GeoEngineers, 1999). About 26% of the Carlyon Beach site has deep-seated landslides (Booth *et al.*, 2009). Landslide conditions have caused considerable damage to hundreds of homes that have now been declared uninhabitable and have decreased in value. Human development of the area around the Carlyon Beach Peninsula has also removed some visible surface features of landslides, which may cause difficulty when determining potentially problematic areas (Booth *et al*., 2009).

The DEM is the 3D data of the study area, which is used in the research for analysis and testing. Figure 2 shows the DEM in raster format, and Figure 3 is the inventory map. The LiDAR data was collected by the Puget Sound LiDAR Consortium in 2002. This DEM has been bare-earth filtered and has a point-spacing of 6.00 ft. or 1.83 m. The landslide inventory map was compiled in 2008 by M. Polenz, of the Washington State Department of Natural Resources, using a combination of the aforementioned DEMs in conjunction with aerial images and data collected on the ground. Deep-seated as well as surficial landslides are both accounted for in the map (Booth *et al*., 2009).

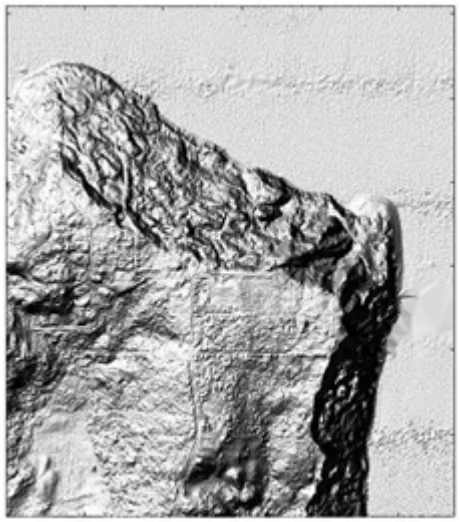

Figure 2. *Digital Elevation Model*

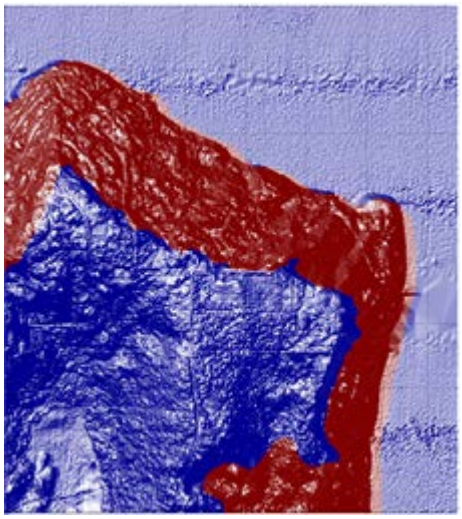

Figure 3. *Landslide Inventory Map*

Methodology

Landslide detection is conducted by applying individual feature extractors to a DEM of the study area that are able to identify patterns of landslide morphology. The features are extracted by applying a sliding window that analyzes the geomorphological features of the terrain, in particular, the surface topographic variability. Prior studies have shown that higher topographic variations in the terrain result in a rough surface as found in landslide terrain, whereas low variations result in a smooth surface found in stable terrain (McKean and Roering, 2004; Booth *et al.*, 2009, Mora *et al.*, 2015). Several feature extractors were tested; however, roughness, slope, local topographic range, and local topographic variability are discussed in detail based on the results obtained. After extracting the surface features an unsupervised classification, in particular k-means and GMM clustering, is applied to classify the surface features into different groups. An unsupervised classifier is applied to recognize topographic patterns and characteristics by categorizing smooth stable terrain as non-landslide and rough surface areas as landslide prone areas. Finally, a confusion matrix is used to evaluate the different methods' performance by comparing it to the independently complied landslide inventory map. The

methods are employed to detect the topographic expressions found in landslide terrain and map their location. The workflow of the methodology is described in Figure 4, which outlines the process for landslide detection and validation.

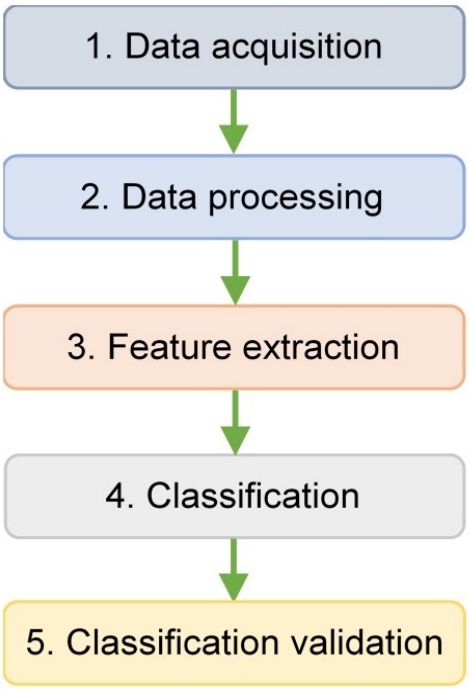

Figure 4. *Workflow of methodology*

Feature extraction is a type of dimensional reduction that efficiently represents specific parts of an image or DEM as a compact feature vector (Lee *et al.*, 1992). It performs a neighborhood operation where a specified algorithm visits each cell of the raster and calculates an allotted value within a selected neighborhood. Optimal landslide readings are identified after altering window sizes, neighborhood types, and feature extractors, as performed in our study, where several of each were evaluated. However, only the optimal results observed are shown.

Roughness
Determining the greatest difference between the center pixel and its neighborhood generates a value to describe the roughness of the surface terrain. This computation utilizes the following formula and Eq. (1), $R = Max(Z_{ij} - Z_{11})$, where $i = 0\text{-}2$, $j = 0\text{-}2$ (Mora *et al*., 2015).

$$\begin{matrix} Z_{02} & Z_{12} & Z_{22} \\ Z_{01} & Z_{11} & Z_{21} \\ Z_{00} & Z_{10} & Z_{20} \end{matrix} \qquad \text{Eq (1)}$$

Slope
The slope of the surface is defined as the greatest rate of change among the center cell and the adjacent cells. Calculating the value of the sharpest grade of a DEM can evaluate the slope. This can be done using Eq. (1), and the formula, $S_{D8} = \max\left[\frac{Z_{ij}-Z_{11}}{h\varphi(ij)}\right]$, where, $i = 0\text{-}2$, $j = 0\text{-}2$. Use the value of f $(ij) = 1$ for the cardinal directions (north, south, east and west) and f$(ij)=\sqrt{2}$ for diagonally adjacent cells (Mora *et al*., 2015).

Local Topographic Range
The local topographic range describes the difference between the highest and lowest elevations of a local neighborhood. This is essentially the range of elevations for the particular section of the surface, which can be evaluated using Eq (1), $Range = Max(Z_{ij}) - Min(Z_{ij})$, where $i = 0\text{-}2$, $j = 0\text{-}2$.

Local Topographic Variability
Local topographic variability assesses the standard deviation of local neighborhoods and is applied to the data obtained from the feature extractors of roughness, slope, and local topographic range detailed above. In addition, local topographic variability is applied to the pure surface elevation data. The data obtained from applying the local topographic variability feature extractor is used in our classification.

K-means Clustering
In this classification, any individual feature is compared to each discrete cluster to determine which one it is closest to. A map of all features in the DEM, classified by which cluster each feature is most likely to belong, is produced (Seber, 2008). This was tested using a class setting from two through five. This sub-feature extractor employs Map Algebra to perform the following equation:

$$Z = \frac{(\boldsymbol{X-oldmin}) \times (\boldsymbol{newmax-newmin})}{(\boldsymbol{oldmax-oldmin})} + \boldsymbol{newmin} \qquad \text{Eq (2)}$$

Where:
Z is the output raster with new data ranges.
X is the input raster.
Oldmin is the minimum value of the input raster.
Oldmax is the maximum value of the input raster.
Newmin is the desired minimum value for the output raster.
Newmax is the desired maximum value of the output raster.

In other words, the k-means function will study the original data and look for similarities, then separate the data into different classes. This means that data in each class has similarities but they are distinct from other classes.

Gaussian Mixture Model Clustering

A second common model used for data classification is a GMM. GMMs cluster data based on assigning each data point to a cluster that will maximize the statistical probability that a hypothesis is true, relative to the data set. Clusters for GMM are defined as Gaussian distributions that are centered on their barycenter's. GMM clustering is more flexible because it has the option of hard or soft/fuzzy clustering. Hard clustering is when one data point is assigned to exactly one cluster. Whereas soft/fuzzy clustering is where data point is assigned a score for each cluster. The score corresponds with the strength of a data point to a cluster. GMM employs model clustering as a mixture of multivariate normal density components (McLachlan, 2000). This was tested using a class setting from two through five. This sub-feature extractor employs Map Algebra to perform the following equation:

$$P(\theta|x) = \sum_{i=1}^{K} \widetilde{\Phi}_i N(\widetilde{\mu_\iota}, \widetilde{\sum_\iota}) \qquad \text{Eq. (3)}$$

Where:
$P(\theta|x)$ is the posterior distribution.
$\widetilde{\Phi}_i$ is the prior probability of observation associated with component *i.*
N is the number of observations.
$\widetilde{\mu_\iota}$ is the mean of the *i*th vector component.
$\widetilde{\sum_\iota}$ is the covariance matrices of the *i*th vector component.

Accuracy Assessment (Confusion Matrix)

The accuracy assessment method used is the confusion matrix, which will provide an evaluation of the unsupervised methods tested. The confusion matrix provides assessment methods, specifically: accuracy (AC), true positive (TP), false positive (FP), true negative (TN), false negative (FN), and precision (P). AC is the overall performance of the unsupervised clustering method, in other words, how often does the algorithm correctly identify landslide and non-landslide terrain. The positive/negative portion of the labels refer to the identification of the terrain: positive for landslide terrain and negative for non-landslide terrain. The true/false nomenclature indicates whether the clustering correctly identified the terrain: true for correct identification and false for incorrect identification. For example, FP means that an area was incorrectly identified as landslide terrain. In other words, it was false that the terrain exhibited landslide surface features, and thus in reality a non-landslide area. FP is known as a type I error, while FN is known as a type II error. A type II error is more severe as it neglects to identify a landslide, this may potentially lead to severe damage to structures built on top under the assumption that it was safe from any sliding. The confusion matrix is used to compare the percentage of matching terrain between the algorithm mapped areas and the inventory map. The algorithm mapped terrain in this step would be the results from the k-means and GMM clustering techniques. However, the unsupervised classification methods were tested by applying several classes to each surface feature extracted, thus each resulting class needs to be designated landslide or non-landslide manually.

Results and Discussion

For each geomorphological feature extracted, two to five classes are evaluated by applying GMM and k-means clustering. However, only a sample based on each technique were selected

and shown. After clustering the data from the feature extractors, an evaluation process of selecting which group (landslide or non-landslide) the various classes pertain to is performed manually. Figure 5 is a demonstration of the process, it begins with the features being extracted (A), followed by the clusterization (B), then the manual designation of each cluster to a landslide or non-landslide group (C), and finally the process ends with a comparison between algorithm mapped landslide locations and the inventory mapped landslides (D). Figure 5 demonstrates the results and process based on the GMM clustering method after applying the local topographic range feature extractor and clustering the data into five different groups. In Figure 5D the locations marked in red signify correctly mapped areas (TP and TN), while the blue defines incorrectly mapped locations (FP and FN). The confusion matrix results for Figure 5 are shown in Table 1. The evaluation process was performed similarly for all feature extractors (roughness, slope, local topographic range, local topographic variability) that were evaluated, where two through five clusters were examined by applying both the GMM and k-means clustering methods. Shown in Figure 6 are the results obtained from the roughness feature extractor and by clusterizing the data into 3 classes and by applying the k-means clustering method. In general, the process is the same for both algorithms, other than the clusterizing method.

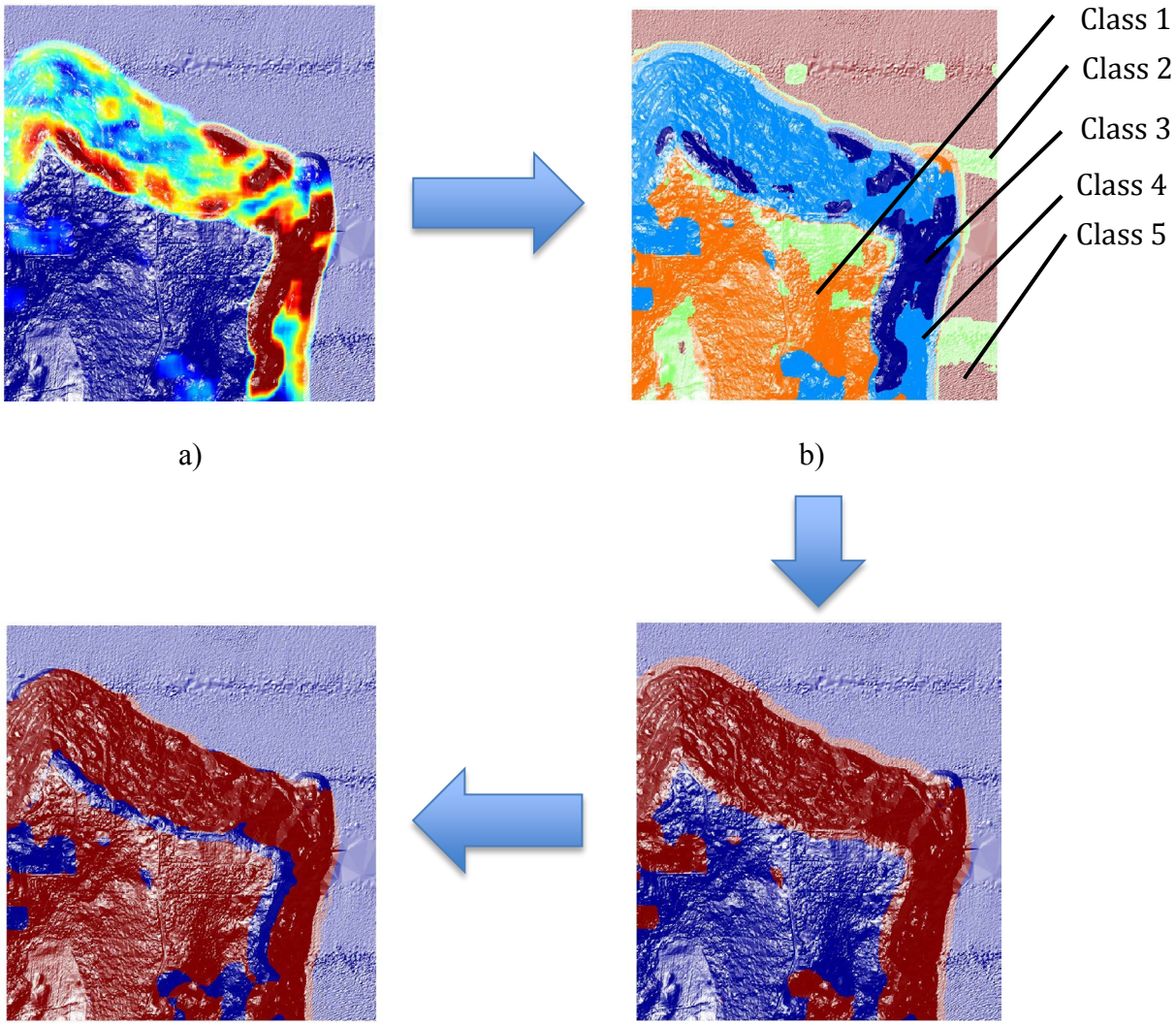

Figure 5. *GMM clustering by applying 5 classes and the local topographic range feature extractor. a) Represents the results from the feature extractor. b) GMM clustering results. c) Re-generated from b) where red presents landslide and blue represents non-landslide. d) Comparison map between c) and inventory map where red means matched and blue means not matched.*

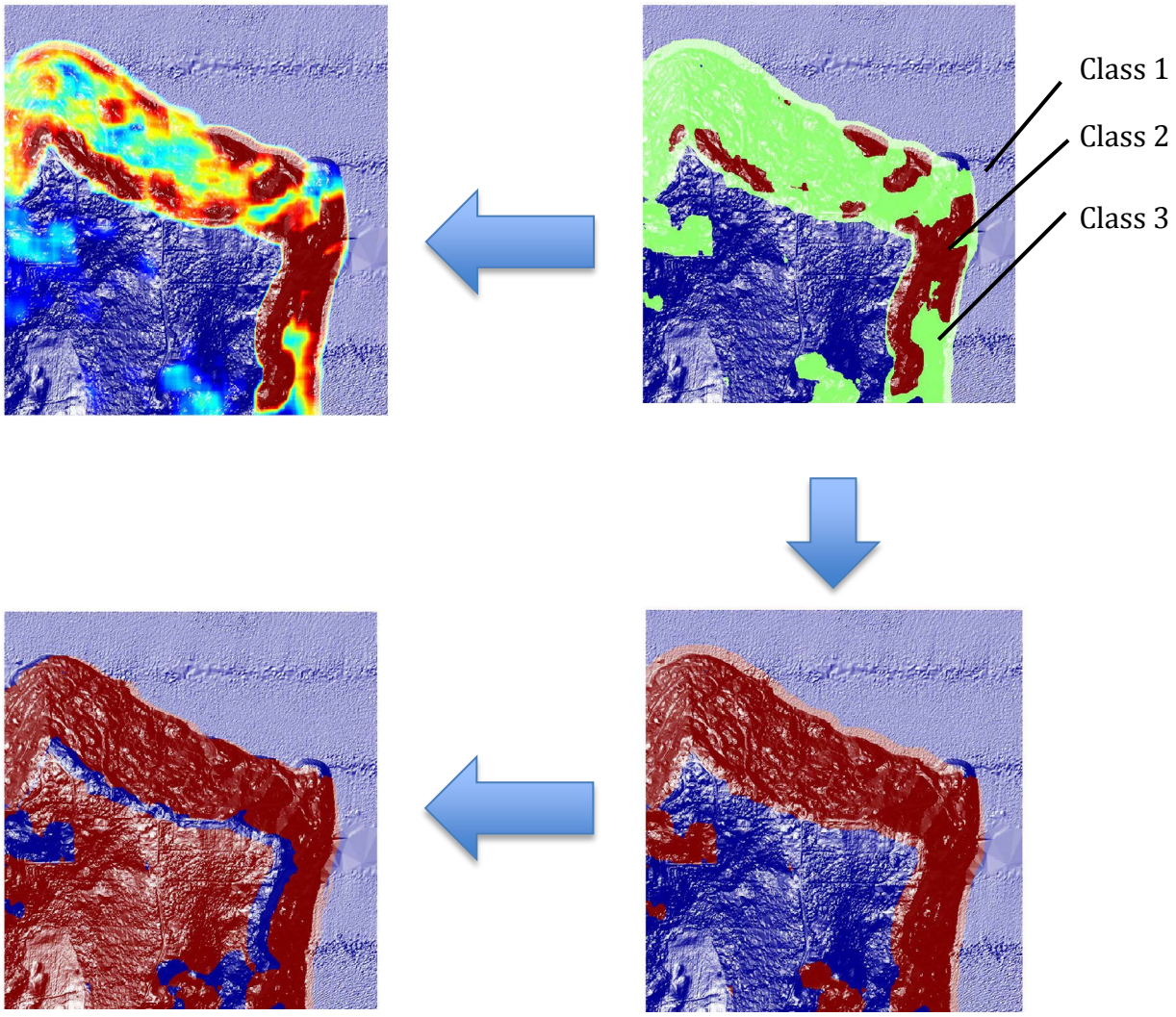

Figure 6. *K-means clustering by applying 3 classes and the roughness feature extractor. a) Represents the results from the feature extractor. b) K-means clustering results. c) Re-generated from b) where red presents landslide and blue represents non-landslide. d) Comparison map between c) and inventory map where red means matched and blue means not matched.*

Table 1. *Confusion Matrix for GMM*

FEATURE	NO. CLUSTERS	AC	TP	FP	TN	FN	P
ROUGHNESS	2	83.70%	95.35%	25.34%	74.66%	4.65%	74.48%
	3	82.82%	95.58%	27.08%	72.92%	4.42%	73.25%
	4	86.84%	91.86%	17.06%	82.94%	8.14%	80.68%
	5	86.92%	91.43%	16.58%	83.42%	8.57%	81.05%
SLOPE	2	43.69%	100.0%	100.0%	0.00%	0.00%	43.69%
	3	84.36%	81.75%	13.61%	86.39%	18.25%	82.33%
	4	85.63%	90.98%	18.53%	81.47%	9.02%	79.21%
	5	85.81%	88.69%	16.42%	83.58%	11.31%	80.73%
LOCAL TOPOGRAPHIC RANGE	2	82.82%	95.79%	27.23%	72.77%	4.21%	73.18%
	3	85.63%	94.30%	21.09%	78.91%	5.70%	77.62%
	4	86.30%	93.27%	19.11%	80.89%	6.73%	79.11%
	5	86.50%	92.69%	18.30%	81.70%	7.31%	79.71%
LOCAL TOPOGRAPHIC VARIABILITY	2	43.69%	100.0%	100.0%	0.00%	0.00%	43.69%
	3	78.79%	90.49%	30.28%	69.72%	9.51%	69.86%
	4	84.89%	77.15%	9.11%	88.31%	22.85%	86.79%
	5	84.71%	80.06%	11.69%	87.56%	19.94%	84.16%

Table 2. *Confusion Matrix for k-means clustering*

FEATURE	NO. CLUSTERS	AC	TP	FP	TN	FN	P
ROUGHNESS	2	87.19%	87.91%	13.37%	86.62%	12.09%	83.60%
	3	86.28%	93.70%	19.47%	80.53%	6.30%	78.87%
	4	87.09%	86.81%	12.69%	87.31%	13.19%	84.14%
	5	86.98%	91.16%	16.26%	83.74%	8.84%	81.30%
SLOPE	2	54.54%	99.42%	80.27%	19.73%	0.58%	49.00%
	3	83.80%	79.98%	13.24%	86.76%	20.02%	82.41%
	4	85.45%	91.55%	19.28%	80.72%	8.45%	78.65%
	5	85.47%	86.05%	14.94%	85.06%	13.95%	81.72%
LOCAL TOPOGRAPHIC RANGE	2	87.20%	87.96%	13.37%	86.61%	12.04%	83.59%
	3	86.00%	93.89%	20.13%	79.87%	6.11%	78.35%
	4	87.22%	88.22%	13.55%	76.45%	11.78%	83.47%
	5	86.70%	91.92%	19.36%	82.64%	8.08%	80.43%
LOCAL TOPOGRAPHIC VARIABILITY	2	84.12%	81.07%	13.51%	86.49%	18.93%	82.31%
	3	72.27%	96.29%	46.38%	53.62%	3.42%	61.69%
	4	84.02%	73.63%	7.92%	92.08%	26.38%	87.83%
	5	85.03%	78.93%	10.24%	89.76%	21.07%	85.68%

Tables 1 and 2 display the confusion matrix results from the feature extractors tested with their corresponding classes. In Table 1 are the results obtained by applying the GMM clustering and in Table 2 are those based on k-means clustering. Two important factors analyzed from the confusion matrix are AC and FN. AC is the overall matching percentage between algorithm mapped locations and the landslide inventory map. Thus, a higher AC indicates that most

locations were mapped correctly. However, FN should be minimized due to the fact that it defines falsely identified landslide areas as non-landslide. FN areas can be hazardous as it may define an area as stable, when in reality it is not. FN values are inversely related to TP, by minimizing FN it ensures a higher TP value. When analyzing the data, it is important to consider all parts of the data, not just one specific value. For example, the best results do not necessarily reflect the lowest possible FN values. In our case, some of the data shows that the AC for the lowest FN values are lower than acceptable. The results may have been impacted by an over classification that resulted in a low FN; however, the AC classification performance is lower than desired. Therefore, an overall balance must be found amongst the confusion matrix results.

In Table 2 the k-means clustering method demonstrates that the highest AC value for the slope feature extractor is 85.47% with an FN value of 13.95%. The slope feature with four clusters, yielded an AC of 85.45% and a FN value of 8.45%. It did comparatively better, and performed similarly as the roughness, local topographic variability and local topographic range features. When comparing AC values of different number of classes within each feature extractor, results were generally within 2% of each other. The only two exceptions to this are the local topographic variability where AC values ranged from approximately 72% to 85% and slope from approximately 55% to 85%. Two of the best performing results were the local topographic range and roughness feature extractors with 3 clusters. The k-means method with the roughness feature extractor with three clusters achieved an AC of 86.28% and an FN of 6.30%. The FN value was 0.19% higher than that found through the local topographic range filter with three clusters, which on its own, is not significantly different. The AC value was 0.28% more than local topographic range, in this instance it is slightly more accurate, however, in general the results are similar.

When comparing TP and FN values, GMM performed differently as k-means clustering. Using GMM clustering, approximately 63% of all values had a FN value equal to or less than 10% compared to the 31% that passed the same criteria using k-means clustering. The opposite is true with AC values. 50% of values found through k-means analysis were over 86% accurate, compared to the 25% of GMM values with the same cut-off value. This data shows that GMM clustering classified landslides better than k-means clustering, but was not done as accurately, therefore, there was an over classification.

The roughness and local topographic range feature extractors performed similarly amongst all clusters in the GMM analysis. The FN values were all under 9% and had AC values in the 80% range. These results were generally better than those found through slope and local topographic variability. For local topographic range and roughness, clusters of four and five had an AC of at least 86% while clusters of 2 and 3 were slightly lower, closer to approximately 82% on average. The FN values were 2 to 3% higher in the larger number of clusters than in the lower numbers. Again, the difference in AC gives a slightly higher FN value in these results by not overclassifying. From the GMM analysis, one of the top performing results was from the range filter with 5 classes having an AC of 86.50% and an FN of 7.31%.

Conclusion

Landslides in the Carlyon Beach Peninsula have caused a major impact on human life, infrastructure and economy. As a result, landslide-mapping technology has undergone innovation and improvement within the last few decades to help mitigate the damages caused by such geological disasters. Traditional mapping methods such as aerial photographic analysis and field inspection are still employed internationally to detect landslide prone areas, but these methods have proven to be time consuming, costly, and unable to detect small scale failures. Meanwhile, modern technologies employ the use of DEMs and automated algorithms that are able to successfully detect landslide and non-landslide terrain with highly accurate results in a cost-effective manner. The methods performed feature extraction on the DEM of the study area by identifying stable terrain with smooth features, and terrain with rough surfaces as landslide

prone. The classification results from k-means and GMM clustering, are able to attain an accuracy up to 87% when compared to the landslide inventory map. As a result, modern techniques are able to detect landslide terrain using DEMs in a time-effective manner that is affordable and accessible. Future recommendations include the use of a larger dataset and fusing several feature extractors to perform the classification. This may help determine if the results can be improved and if the techniques can be applied to other regions.

Acknowledges

The authors wish to acknowledge Joy Sellman and Alaeddin Slayyeh for their contributions to this project and to the Civil Engineering Department of California State Polytechnic University, Pomona for their support.

References

An, Hyunuk, et al. "Survey of spatial and temporal landslide prediction methods and techniques." *Korean Journal of Agricultural Science* 43 (2016): 507-521.

Booth, Adam M., Josh J. Roering, and J. Taylor Perron. "Automated landslide mapping using spectral analysis and high-resolution topographic data: Puget Sound lowlands, Washington, and Portland Hills, Oregon." *Geomorphology*109.3 (2009): 132-147.

Cheng, Gong, et al. "Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA." *International Journal of Remote Sensing* 34.1 (2013): 45-59.

Dalyot, Sagi, Eran Keinan, and Yerach Doytsher. "Landslide morphology analysis model based on LiDAR and topographic dataset comparison." *Surveying and Land Information Science* 68.3 (2008): 155-170.

Dou, Jie, et al. "Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan." *PloS one* 10.7 (2015): e0133262.Effat, Hala Adel, and Mohamed Nagib Hegazy. "Mapping landslide susceptibility using satellite data and spatial multicriteria evaluation: the case of Helwan District, Cairo." *Applied Geomatics* 6.4 (2014): 215-228.

Effat, Hala Adel, and Mohamed Nagib Hegazy. "Mapping landslide susceptibility using satellite data and spatial multicriteria evaluation: the case of Helwan District, Cairo." *Applied Geomatics* 6.4 (2014): 215-228.

GeoEngineers, *Report, Phase I- Reconnaissance Evaluation, Carlyon Beach/ Hunter Beach Landslide*. Tech. no. 731800100. Washington: n.p., 1999. Print.

GeoEngineers, *Report, Phase II Geotechnical Study, Carlyon Beach/ Hunter Beach Landslide Phase II- Reconnaissance Evaluation*. Tech. no. 731800100. Washington: n.p., 1999. Print.

Glenn, Nancy F., et al. "Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity." *Geomorphology* 73.1 (2006): 131-148.

Hölbling, Daniel, Barbara Friedl, and Clemens Eisank. "An object-based approach for semi-automated landslide change detection and attribution of changes to landslide classes in northern Taiwan." *Earth Science Informatics*8.2 (2015): 327-335.

Jaboyedoff, Michel, et al. "Use of LIDAR in landslide investigations: a review." *Natural hazards* 61.1 (2012): 5-28.

Lee, Jay, Peter K. Snyder, and Peter F. Fisher. "Modeling the effect of data errors on feature extraction from digital elevation models." *Photogrammetric Engineering and Remote Sensing* 58.10 (1992): 1461-1467.

Leshchinsky, Ben A., Michael J. Olsen, and Burak F. Tanyu. "Contour Connection Method for automated identification and classification of landslide deposits." *Computers & Geosciences* 74 (2015): 27-38.

McLachlan, Geoffrey, and David Peel. "Mixtures of factor analyzers." Finite Mixture Models (2000): 238-256.

Mora, Omar E., et al. "Small landslide susceptibility and hazard assessment based on airborne lidar data." *Photogrammetric Engineering & Remote Sensing* 81.3 (2015): 239-247.

McKean, J., and J. Roering. "Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry." *Geomorphology* 57.3 (2004): 331-351.

Sarkar, Shantanu, et al. "GIS based spatial data analysis for landslide susceptibility mapping." *Journal of Mountain Science* 5.1 (2008): 52-62.

Seber, G. A. F. "Cluster analysis." *Multivariate observations* (2008): 347-394.

Shan, Jie, and Charles K. Toth, eds. *Topographic laser ranging and scanning: principles and processing*. CRC press, 2008.

Tarolli, Paolo, Giulia Sofia, and Giancarlo Dalla Fontana. “Geomorphic Features Extraction from High-resolution Topography: Landslide Crowns and Bank Erosion.” *Natural Hazards*, 61.1 (2012): 65-83.