LANDSLIDE DETECTION AND SUSCEPTIBILITY MAPPING USING LIDAR AND ARTIFICIAL NEURAL NETWORK MODELING: A CASE STUDY IN GLACIALLY DOMINATED CUYAHOGA RIVER VALLEY, OHIO

M. Kenneth Brown

A Thesis

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

December 2012

Committee:

Peter V. Gorsevski, Advisor

Charles M. Onasch

Xinyue Ye

ABSTRACT

Peter Gorsevski, Advisor

The purpose of this study was to detect shallow landslides using hillshade maps derived from Light Detection and Ranging (LiDAR)-based Digital Elevation Model (DEM) and validated by field inventory. The landslide susceptibility mapping used an Artificial Neural Network (ANN) approach and back propagation method that was tested in the northern portion of the Cuyahoga Valley National Park (CVNP) located in Northeast Ohio. The relationship between landslides and different predictor attributes extracted from the LiDAR-based-DEM such as slope, profile and plan curvatures, upslope drainage area, annual solar radiation, and wetness index was evaluated using a Geographic Information System (GIS) based investigation. The approach presented in this thesis required a training study area for the development of the susceptibility model and a validation study area to test the model. The results from the validation showed that within the very high susceptibility class, a total of 42 % of known landslides that were associated with 1.6% of total area were correctly predicted. On the other hand, the very low susceptibility class that represented 82 % of the total area was associated with 1 % of correctly predicted landslides. The results suggest that the majority of the known landslides occur within a small portion of the study area, which is consistent with field investigation and other studies. Sample probabilistic maps of landslide susceptibility potential and other products from this approach are summarized and presented for visualization which is intended to help park officials in effective management and planning.

This is dedicated to the loving memory of my mother, who has inspired me every day of my life. To my father who convince me to pursue geology, and my loving family who has supported me throughout this project. Lastly to my loving wife Sarah, who encouraged, supported, and sacrificed to make this project possible. I cannot truly thank you enough.

ACKNOWLEDGMENTS

Above all, I would like to thank my advisor, Dr. Peter Gorsevski, for his advice, persistence, dedication to me as an individual and as his student. I thank him for instilling in me that new technology should not deter us, but inspire us to accomplish new and exciting discoveries.

I want to thank Dr. Charles Onasch for enlightening me to the true power of investigating landslides and his advice to embark on this research. I wish to express my gratitude for our meetings, which provided different perspectives that furthered this research. Additionally, I would like to thank Dr. Xinyue Ye for his commitment and involvement to this research.

I am very grateful to Dr. Robert Vincent for the opportunity to work with the NOAA Lake Erie Water Quality Research Project as well as providing me with a volunteer opportunity within the Rotary International water exploration projects. These experiences have enabled me to see the world in a different way and I truly thank him.

I would like to thank Bill Butcher for his computer help which has truly transformed me within this field. I am also indebted to Gail Nader for her editorial advice within this project.

Finally, I would like to thank all my fellow graduate students and friends through which has made my graduate experience extraordinary.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Landslides are important geomorphic factors that reshape the landscape and transform local topography by redistributing materials far away from its source. Cruden (1991) defines the landslides as the "movement of a mass of rock, debris or earth down a slope" caused by the action of gravity. Sudden and rapid landslide events are often associated with fatalities, environmental degradation and different types of damage to businesses, buildings, roads, public utilities, and disruption to people and the environment that cost millions of dollars annually. For instance, in 1996, damage from a series of rotational slump slides in Hamilton County, Ohio exceeded $10 million and in 1996-98 landslides and mudslides initiated by massive floods in the eastern and the southeastern Ohio counties led to repair, rehabilitation, and maintenance costs which exceeded $34 million (State of Ohio, 2008). Other estimates of landslide losses, such as from the State of Ohio Hazard Mitigation Plan (SOHMP) crafted in 2008 in response to Federal Disaster Mitigation Act (DMA), surpass $22 million (State of Ohio, 2008).

In Ohio, the potential susceptibility and incidence of landslides varies across different regions, but unglaciated southeastern areas along the Ohio River and other major drainages across the state experience frequent and costly failures (Hansen, 1995). For instance, the City of Cincinnati has some of the highest cost per capita landslide damages within the United States due to road failures. Historic figures suggest more than $5 million was spent annually between 1973 and 1978 (Fleming and Taylor, 1980) and $7.2 million annually between 1988 and 1992 (Pohana, 1992). The common landslide types that occur across the state are: rock-falls, earth flows, and rotational slumps (Hansen, 1995; State of Ohio, 2008). While earth flows are the most common landslides across the state, in the Cuyahoga River valley between the cities of Cleveland and Akron rotational slumps occur mostly in the areas where the river has eroded into

the Pleistocene glaciolacustrine deposits leaving unstable valley walls (Nandi and Shakoor, 2009). Such landslides consist of a coherent mass of loosely consolidated materials or weakened rock layers that move short distances down a slope on a concave-upward surface. Such landslides produce a scarp at the head of the rupture and a series of step-down blocks, which lead to lateral spreading at the bottom of the slide called the toe (Varnes, 1978).

There are many factors that control the location and extent of landslides, including geological (weathered materials, permeability, earthquakes), physical (intense rainfall, steepened slopes, erosion) and anthropogenic (removal of vegetation, land-use change, population growth, urbanization, over-steepened upper part of the slope) (State of Ohio, 2008). Nandi and Shakoor (2006) have shown within the Cuyahoga River valley that steep slopes and erosion, proximity to streams, and soil types are the main factors in landslides within this region. Understanding such causes, controlling factors and answering specific questions of spatial and temporal patterns requires production of detailed landslide inventory, which is the first step in assessing landslide susceptibility models that can be used for decision and policy-making purposes.

The main purpose of inventory maps is to provide the baseline information on the types of landslides in the affected area, distribution and displacement caused by one or multiple events, and assessment of factors that influence the slope failure (Galli et al., 2008). Traditional inventorying methods include field observations, aerial photo interpretation (API), as well as using historical records of landslides for locating areas where landslide occur (Jones and Shakoor, 1989; Nandi and Shakoor, 2006; 2007; 2009). Some of the shortcomings of these traditional methods include lack of adequate aerial photo resolution required to map small landslides, obscuration of morphologic features by vegetation, and the time-consuming and difficult nature of detailed field mapping, especially when done in rugged terrain.

In recent years the airborne-derived products from Light Detection and Ranging (LiDAR) measurements, such as high-resolution digital elevation models (DEMs) and maps of landslides obtained from beneath dense vegetation, are becoming increasingly important for producing a detailed landslide inventory (Van Den Eeckhaut et al., 2006). LiDAR applications include the construction of DEMs (Liu, 2008), shaded relief maps (Haneberg, 2005), detection of historic landslides under forested area (Van Den Eeckhaut et al., 2007), creation of topographic contours (Schulz, 2004), multi-temporal DTM tracking of landsides (Dewitte et al., 2008; McKean and Roering, 2004; Glenn et al., 2006), hydrological modeling (Liu, 2008 ; Liu et al., 2005; Hopkinson et al., 2005; Hopkinson and Chasmer, 2008), landform and or soil classification (Liu, 2008; and Anderson and Croft, 2009; Bork and Su, 2007)**,** and understanding fine-scale landslides patterns (McKean and Roering, 2004; Glenn et al., 2006; Van Den Eeckhaut et al., 2007). The main advantage of LiDAR is that it allows for robust terrain mapping of landslide patterns through visual or quantitative analysis.

After the production of detailed landslide inventory maps is completed, landslide susceptibility mapping is applied to identify potential landslide areas and to determine where or when landslides are most likely to occur. Often, it is assumed that future landslides can be predicted by using relationships of past landslides and predisposing factors such as elevation, slope, plan and profile curvature, flow path length, wetness index, and specific catchment area. Such factors represent the primary and the secondary derivatives from LiDAR-derived DEMs, which reduce the high cost of detailed field data collection (Gomez and Kavzoglu, 2005; Lee et al., 2006; Melchiorre et al., 2008; Pradhan and Lee, 2010). Some of the methods and techniques which use the knowledge of past landslide events to predict future events include: heuristic approaches which use expert knowledge and decision rules (Van Westen, 2000; Barredo et al.,

2000; Ruff and Czurda, 2008), statistical models which link factor through spatial correlations associated with landslide occurrence (Dai et al., 2000; 2002; Lee and Min, 2001; Santacana et al, 2003; Suzen and Doyuran, 2003; Lee et al., 2004; Gorsevski et al., 2006; Nandi and Shakoor, 2006; Van Den Eeckhaut et al., 2006; Godt et al., 2008; Yilmaz and Keskin, 2009), and artificial intelligence approaches capable of learning complex behaviors with minimal human intervention and prior knowledge (Yilmaz, 2009; Pradhan et al., 2010; Pradhan and Lee, 2010).

The attractiveness of artificial intelligence techniques such as Artificial Neural Network (ANN) analysis is that it is independent of the statistical distribution of the data and there is no need for specific assumptions with the multivariate data distributions. ANN analysis allows for a non-linear relationship between the landslide and the main susceptibility factors. Some of the important capabilities of the ANNs models include the learning abilities, generalization and working with multiple numbers of variables, which are used to extract patterns and detect complex trends that are difficult to be noticed otherwise. ANNs simulate how the human brain processes a specific problem through learning algorithms that model knowledge and save this knowledge in weighted connections. The multi-layer perceptron (MLP) consists of a set of layers and nodes that are used in the back-propagation (BP) learning algorithm which is the most widely used algorithm for an ANN model. Such framework contains of an input layer, an output layer, and one or more hidden layers. The initial network starts with random weights that represent the inputs and expected outputs, but iterative procedure is used to optimize the weights and reduce the output errors (Lee et al., 2004; Gomez and Kavzoglu, 2005; Lee and Evangelista, 2006; Prodhan and Lee, 2007; Melchiorre et al., 2008; Nefeslioglu et al., 2008). ANNs have been used in other landslide susceptibility studies with great success for discerning susceptible and non susceptible areas (Lee et al., 2006; Yesilnacar and Topal, 2005; Melchiorre et al., 2008).

In this study, the ANN approach uses LiDAR-derived landslide and controlling factors using a case study of the Cuyahoga Valley National Park (CVNP), where the following three objectives were set: (1) to assess the landslides through LiDAR-based interpretation and field inventory; (2) to evaluate sets of different combinations of LiDAR-derived controlling factors in the development of an predictive ANN model; and (3) to evaluate the ANN predictive model using independent test datasets in the CVNP. The first objective will show how effective and proficient LiDAR is in identifying landslides within the hillshade map interpretation. Field investigations helped in verifying some landslide locations from the LiDAR-derived hillshade map interpretation inventory. The second objective will help explain the topographic characteristics of landslides within the study area. The most significant factors that control the presence of landslides will help in providing the basis for the susceptibility of the region. The most significant controlling factors will be used as the input variables for the ANN model. The third objective will help in determining how successful the proposed ANN model is at predicting landslides within this region by performing the ANN susceptibility model for the validation study area. These landslide locations predicted by the ANN model will allow park managers to manage landslide-prone areas and to update existing landslide susceptibility studies to provide a safe environment for the visitors of the national park.

2. MODELING APPROACH

2.1 Extracting Features from LiDAR

LiDAR is an active remote sensing technique, similar to radar, but uses laser light. The advantage of this high resolution cutting-edge technology is that can penetrate through vegetative cover to derive fine-scale features on the ground surface. The data collection and subsequent processing allows for quantitative measurements of landforms at sub-meter scale and extraction of hidden features under tree canopy with a high degree of accuracy (Van Den Eeckhaut et al., 2005, 2007; Razak et al., 2011). Commonly, techniques for feature extraction from LiDAR-derived DEMs use computed relief images that are interpreted like aerial photographs (Schultz, 2004; Glenn et al., 2006; Ardizzone et al., 2007; Fiorucci et al., 2011) or semi-automated or automated methods that classify or detect various types of landforms such as hillslope, glacial, fluvial, and coastal geomorphology (McKean and Roering, 2004; Ioannilli and Paregiani, 2008; Booth et al., 2009). Typical steps in the extraction of landslides and creation of inventory maps often is based on products such as LIDAR-derived hillshade, slope, and contour line maps in a GIS environment.

2.2 Artificial Neural Networks Model

An Artificial Neural Network (ANN) is a computational technique that was inspired by attempts to model the human central nervous system using human reasoning and problem-solving abilities (Bain, 1873; James, 1890; McCullock and Pitts, 1943; Werbos, 1975). The design of this computational technique is to emulate biological neural networks in a similar way as the human brain learns tasks that are difficult to simulate with other logical and/or analytical techniques. An ANN differs from other forms of computer intelligence because it is not rule-based and doesn't need a predefined knowledge base. ANN is trained to learn associative

patterns to recognize and generalize the relationship between a set of inputs and outputs. A well-trained ANN in some instances could be highly effective in discerning patterns that are difficult to detect otherwise. The ANN has been used in a number of applications such as image processing, pattern recognition, function approximation, optimization, forecasting, data retrieval, and landslide susceptibility (Egmont-Petersen et al., 2002; Lee et al., 2004; Ermini et al., 2005; Yesilnacar and Topal, 2005; Kanungo et al., 2006; Melchiorre et al., 2008).

The typical architecture of an ANN model consists of topology, a learning paradigm, and a learning algorithm. Network topology refers to the organization, the connection of the nodes, and the flow of data and error information between the layers, while the learning paradigm often handles two broad categories of supervised and unsupervised network learning using different learning algorithms. The multilayered topology is the most common algorithm, where all processing units or the nodes are organized in three essential layers: input, hidden, and output. The input layer is used to pass predictive attribute data forward to the hidden layer which is connected to an output layer. Within this framework, the produced output is recursively compared with the desired output until error signal is minimized and computed output resembles the desired output. Figure 1 shows the architecture of the ANN model used within this study.

The most common feed-forward neural networks are constructed by multilayer perceptrons (MLP) and radial basis functions that use the back propagation (BP) algorithm (Lippman, 1987). In back propagation neural networks the data passes forward from input layers to output layers via the hidden layer(s). The initial output is produced for the input data and randomly assigned weights. The resulting outcome is compared with the desired output and error discrepancies are propagated backward through the network from the output to the input

nodes. This is an iterative backward propagation intended to adjust the synaptic strength of weights to ensure similarity between the computed and desired output.

3. MATERIAL AND METHODS

3.1 Study Area

The study area shown in Figure 2 is located within the CVNP, in northeast Ohio located along the lower Cuyahoga River, which is one of the most landslide-prone regions of the watershed (Nandi and Shakoor, 2009). According to the National Park Service, the park covers 133.6 km^2 with altitudes varying from 305 to 381 m above sea level. The annual precipitation in the area ranges from 927 to 1,027 mm, with changes driven by localized 'lake effect' precipitation. The maximum discharge for the Cuyahoga River occurs between March and May when snow melt runoff and rainfall amounts are the highest. The Cuyahoga River flows 35.4 km through the National Park and the drainage network is well developed with a significant surface runoff.

The Cuyahoga River Valley is a part of the glaciated Allegheny Plateau with a geologic history dominated by two significant periods of deposition (White, 1982, 1984; Ford, 1987; Nandi, 2007). The first period is from the Paleozoic time when the bedrock of the region was deposited. The bedrock consists of late Devonian, Mississippian, and Pennsylvanian siliciclastic sedimentary rocks. The late Devonian units were deposited within the shallow sea west of the Catskill Delta of Pennsylvanian; whereas the Pennsylvanian deposits resulted from a relative sea level rise, which deposited deltaic sediments as well as the formation of non-marine limestone, shale and coal deposits (Szabo, 1986; Nandi, 2007). There is a 270 million year unconformity (from Triassic to the Pliocene) present in between the two deposition periods which represents the sub aerial extent of Ohio during this time. The second period is from the Pleistocene time where overland glaciers deposited glacial sediments composed of clay, silt, sand, gravel, and boulders (Miller, 1983). Some of the characteristics of glacial sediments include extensive silt

and clay deposits of lacustrine origin. The Pleistocene deposits of the study area are deeply incised by fluvial processes that created steep valley walls that accelerate landsliding (Nandi, 2007; Nandi and Shakoor, 2009).

Figure 2 shows the LiDAR-derived hillshade map of the study areas that was divided into two smaller sections for the analysis: the training area (green outline) used for the prediction of landslides (18.6 km^2) and the test area (blue outline) used for the validation of landslides (20.9 km^2). The areas were chosen based on the similarity of the lithology. The training area consists of 31.15 % Berea Sandstone and Bedford **S**hale, 68.49% Ohio Shale and 0.36% Maxville Limestone, while the validation area consists of 30.23% Berea Sandstone and Bedford Shale, 69.2% Ohio Shale and 0.57% Maxville limestone. Because the lithology of the study area was derived from maps at a scale of 1:500,000 it lacks detail, and the lithology in this study is represented by similarity in the study units and it was not used in the ANN analysis.

3.2 LiDAR Datasets and Derivatives

The digital elevation model (DEM) is a grid-based three-dimensional representation of the terrain elevation and is a fundamental element of landslide inventories and spatial susceptibility analysis (Wilson and Gallant, 2000; Hengl and Reuter, 2009). The DEMs were constructed using publicly available LiDAR datasets from the Ohio Statewide Imagery Program (OSIP) (OGRIP, 2006). The OSIP datasets include LiDAR and high resolution imagery which were acquired in 2006 and 2007. The OSIP color imagery was produced at either 1-foot (0.3048 m) or 6-inch (0.152 m) resolution while the LiDAR dataset were produced with an average laser pulse spacing of 7 feet (2.1336 m) and with an accuracy of 1 foot (0.3048 m). One of the products is preprocessed bare-earth LiDAR-derived DEMs that has 2.5 ft (0.762 m) spatial resolution. The DEMs are available in ArcGIS and ASCII grid format which are organized in

tiles. The coverage of each tile is approximately an area of 2.3 km^2. In the study area the multiple tile covers were mosaicked to create training and validation DEM grid layers. The training study area used a total of eight DEMs whereas the validation study area used a total of nine DEMs.

While the DEM is often the primary derivative from the raw LiDAR data, terrain analysis using DEMs allow derivation of additional topographic attributes including: shaded relief maps (hillshading), elevation, slope, aspect, north, curvatures (profile, plan and mean curvatures), upslope drainage area, anisotropic coefficient of variation, annual solar insolation, sediment transport index, stream power index, elevation, general landform classification, and wetness index. Those attributes can be beneficial for both a landslide inventory and spatial prediction of landslide susceptibility (McKean and Roering, 2004). The effectiveness of LiDAR derived products, especially shaded relief maps is far more superior in landslides interpretation and improves landslide inventories compared with traditional aerial photography interpretation (Figure 3). Thus, the main advantage of LiDAR datasets and derivatives is the visualization of the terrain through multiple combinations of hillshades and other ancillary data sets and subsequent terrain analysis which are aimed at understanding landslide processes and development of susceptibility models.

3.3 Landslide Mapping

LiDAR-derived DEMs were used to identify landslides through high resolution ground models, such as hillshades, combined with slope maps and draped topographic contours. The principal dataset used in landslide mapping are the hillshades, which are calculated from DEMs with varying sun azimuths and sun directions. Such derivatives produce a 3-D pseudo image of the landscape that creates illumination and shadowing to emphasize and highlight geomorphic

features, such as landslides. In this study, hillshades were calculated from the DEMs with varying sun azimuths of 45°, 135°, 270°, and 315°, and a sun angle of 30° above the horizon. In addition, topographic contours produced at different distance intervals assisted in locating geomorphologic features, including: changes in slope, hummocky disturbed surfaces, convex and concave slope areas, midslope terraces, offset drainages, and potential slope changes at the scarp of a landslide.

The majority of the landslides that occur within the study area are rotational slumps that contain a main and/or multiple minor concave scarps, body of the slide with steepened sides occurring towards the toe where displaced material has been collected. Landslides were mapped as polygons that include all elements of a rotational slide (i.e., scarp, body, and toe). The interpretation of the mapped polygons relied on the ground-surface morphology characteristics and 3-D scenes of hillshades and imagery that were visually evaluated for the presence of landslides by systematic evaluation of areas through panning and zooming of the scenes. This interpretation procedure was repeated to produce an unbiased landslide inventory map that included the areas common to the two inventory trails. Within the training study area a total of 212 and 292 landslides where identified, whereas in the testing study area 208 and 297 landslides were identified within the first and second inventory trial. The final count had a total of 190 landslides, where 83 and 107 were recorded within the training and the validation study areas, respectively. The total area of landslides within the training and testing study areas are 0.139 km^2 and 0.168 km^2 or 0.74 % and 0.803% of the total area respectively. The statistics of the landslides occurrences are in Table 1.

3.4 Landslide Validation

The LiDAR-derived landslide inventory was validated by a field investigation. The field investigation incorporated the use of Global Positioning System (GPS) technology (Geo Explorer 6000 series Trimble GPS receiver) to travel to the exact location of each selected landslide. At each location the landslide was authenticated and the boundaries traced and described. The field investigation validated a total of 100 landslides that were randomly selected, where 43 and 57 landslides were identified from both the training and testing study areas.

The field investigation allowed for the comparison between the boundaries as determined in the field and those of proposed landslide locations identified within the hillshade map. Figure 4 shows how a given landslide is displayed within hillshade maps, aerial photographs, and field photographs to demonstrate the features within each dataset. The validation of landslides within the field compared to landslides present in the LiDAR landslide inventory allows for a generic accuracy value to be established for the inventory. From this procedure it was determine that the accuracy of the inventory is 90.7%, 78.9%, and 84% for the training, testing, and combined study areas respectively.

3.5 Terrain Attributes

The terrain or the predictor attributes were derived from LiDAR OSIP datasets and include elevation, slope, aspect, north, curvatures (profile, plan and mean curvatures), upslope drainage area, anisotropic coefficient of variation, annual solar insolation, sediment transport index, stream power index, elevation, general landform classification, and wetness index. Through multiple ANN modeling trials using different combinations and quantity of the

attributes above were narrowed down to the following six: slope, profile and plan curvatures, upslope drainage area, annual solar radiation, and wetness index (Hengl et al., 2003; Hengl and Reuter, 2009; Wilson and Gallant, 2000). The distribution of the six predictor attributes can be visualized in Figure 5 with respect to known landslide locations. A number of different studies have used the same six attributes and have shown to be significant in modeling landslide susceptibility (Ventura and Irvin, 2000; Gorsevski et al., 2003; Gorsevski et al., 2006; Gorsevski and Jankowski, 2008; Regmi et al., 2010a; Regmi et al., 2010b). In this study, those same predictor attributes were also used as inputs in the ANN susceptibility model. Each predictor attribute is discussed in detail bellow for a better understanding of the topography within the study area.

3.5.1 Slope

Slope is defined as the change in the vertical distance divided by the horizontal distance over a given region. Slope is one of the most important factors in controlling the stresses which act on the slope materials and determines how fluid will flow overland and within the subsurface. The amount and type of vegetation, precipitation, and soil water content over a given area are affected by the slope. Traditionally a high slope angle will result in more landsliding, which is significant when locating landslides. Within the study areas the slope ranges from 0 to 100% and the majority of high slope gradients are located within the valley walls which make up 8.65% and 16.09% of the training and testing study areas.

3.5.2 Upslope Drainage Area

The upslope drainage area is also referred to in the literature as the catchment area or flow accumulation which represents the total area that is upslope and drains in a particular catchment outlet. According to Hengl and Reuter (2009) catchment area is defined as:

$$CA = \sum_{i=1}^{n} A_i \qquad \text{Equation 1}$$

where A is the amount of runoff generated in each cell and *n* is the number of upslope cells within a given watershed. This value is summed for the surrounding cells, *n* which flow into an outlet down slope and allows for the total runoff to be calculated. Thus, the surrounding elevation and slope will contribute to the value for the catchment area. A larger value for catchment area will be located at lower elevations and slopes. Smaller values for catchment area will be for higher elevations and slopes (Tarboton, 1997; Hengl and Reuter, 2009). This parameter will also be used in the calculation of the wetness index. Within the study area the catchment area has an interquartile range of 25 m^2 to 110 m^2 within landslide-prone areas and 6 m^2 to 35 m^2 within the areas not prone to landslides.

3.5.3 Profile Curvature

Profile curvature is the amount of curvature parallel to the direction of the largest slope. Profile curvature is sometimes referred to as vertical curvature, where the amount of curvature is calculated directly down the slope. This predictor attribute explains how the fluid will accelerate or decelerate down a hillside based upon if the slope is concave, convex, or flat. The fluid will accelerate down a convex slope and decelerate down a concave slope. Overall, convex slopes will contribute to more erosion due to the acceleration of the fluid as it moves down the hillside.

The values for profile curvature will range from positive to negative, where a positive value represents a slope that is upwardly convex at the reference pixel and a negative value represents a slope that is upwardly concave at the reference pixel (Hengl and Reuter, 2009). A value at or near zero for profile curvature represents a uniform slope. The interquartile range of

the profile curvature is -8.3 to 0.52 for landslide-prone areas and -8.4 to 0.52 for non-landslide-prone areas.

3.5.4 Plan Curvature

Plan curvature represents the amount of curvature perpendicular to the direction of the largest slope. Plan curvature is sometimes referred to as horizontal curvature, where the curvature of a slope is calculated tangentially to a contour line. This predictor attribute explains how a fluid will converge or diverge as it moves down the slope. A concave plan curvature slope will contribute to more erosion because the fluid is focused and accumulates as it flows down slope, which erodes more readily compared to a convex plan curvature slope.

The values for plan curvature range from positive to negative, where a positive value represents a slope that is laterally convex at the reference pixel and a negative value represents a slope that is laterally concave at a reference pixel. A slope that has a plan curvature at or near zero will be laterally uniform. The interquartile range of the plan curvature is -2.7 to 1.9 for landslide-prone areas and -2.68 to 1.9 for the whole study areas.

3.5.5 Annual Solar Radiation

Annual solar radiation represents the average solar radiation that converged at a given pixel within one year. This also takes into consideration the sun angle and what direction the solar radiation is arriving from when it comes in contact with the pixel, which is expressed in Kilowatt hours per square meter (Hengl and Reuter, 2009; Wilson and Gallant, 2000). This attribute is a secondary terrain attribute and is calculated by taking the product of the aspect and slope. For this reason aspect was not used within this study.

Annual solar radiation is important because the larger amount of solar radiation will contribute to a higher amount of evaporation from the soil. With a high value for solar radiation pore space will be available for the rain to accumulate within the soil. Lower solar radiation will contribute to less available pore spaces when a rain event take places and failure will have a higher probability to occur since saturated material are weaker due to pore fluid pressure effects. Furthermore, annual solar radiation controls the amount of vegetation which is present on the slope. The higher amount of solar radiation a slope receives, the more vegetation will be present on the slope leading to a more stable slope.

Within the study area there is a variability of the annual solar radiation depending on the orientation of the slope where the southerly facing slopes received relatively more solar radiation than the northerly facing slopes. The majority of the landslides occur within an annual solar radiation range of 50-60 Kilowatt hours per square meter. The level, non-oriented topography surrounding the valley contains little to no landslides and receives around 70 Kilowatt hours per square meter of annual solar radiation.

3.5.6 Wetness Index

The wetness index is also recognized as the topographic wetness index (TWI) or as the compound topographic index (CTI). The wetness index, which is a very important terrain attribute within landslide susceptibility studies and according to Hengl and Reuter (2009), is defined as:

$$\mathrm{TWI} = \ln \frac{CA}{tan\ \beta} \qquad \text{Equation 2}$$

where, CA is the catchment area and β is the slope. The natural log within this equation helps the outcome to fall within a normalized range. The TWI equation assumes that steady state

conditions occur spatially across the region, where a few assumptions need to be made in order for this equation to hold true. These assumptions include: the precipitation is constant over the region, consistent soil type and thickness, as well as no subsurface drainage network occurs within the region (Wilson and Gallant, 2000). These assumptions allow for both infiltration and transmissivity of the soil to be neglected within the calculation of TWI.

The values of the TWI are a positive non zero which increases as the catchment area increases and the slope angle decreases (Hengl and Reuter, 2009). This attribute in the past has been used to help describe how the topography affected the saturation of the land within a particular location. An increase in the wetness index increases the potential for landslide (Lee and Min, 2001; Gorsevski et al., 2006; Hengl and Reuter, 2009). Previous studies have shown that the higher the TWI values are associated with the landslide deposits compared to lower TWI values with non-landslide deposits. However, within this study area the interquartile range for the wetness index is between 4.1 to 5.83 for both landslide- and non-landslide-prone areas.

3.6 Cluster Analysis

Cluster analysis is the process whereby similarities in the data are grouped together (into smaller classes) so that the data which occurs within the same cluster are more similar to each other than the data which occurs in different clusters. This classification was used to group the six predictor attributes within the known landslide areas into three arbitrary chosen clusters to produce a landslide classification which depicted the main geomorphological elements of a landslide: the scarp, the body, and toe of the landslide. The classification used cluster centers with similar range of values for both the training and the testing study area.

The cluster centers for the six predictor attributes were used to understand similarities and dissimilarities of the predictor attributes with respect to the different landslide elements. The same cluster centers were used for both the training and testing study areas, where a direct comparison can take place for each predictor attribute. The cluster centers correspond to the average value for a given attribute within a particular class. The individual predictor attributes can be used to determine the topography of the given area. Knowing this, the data from Table 2 suggest that class C_1 is most likely to be the toe, C_2 is most likely to be the body and C_3 is most likely to be the scarp of a landslide. This determination was made by understanding the geomorphological characteristics of each element of the landslide. For example, C_3 has the steepest slope and C_1 has the gentlest slope for both training and validation study areas. C_3 also contains the lowest plan curvature and wetness index of all three classes. From this comparison, C_3 represents steep, dry concave areas. C_1 contains the highest solar radiation and plan and profile curvature indicating drier convex areas compared to C_2 and $_3$. C_2 contains a high catchment area, highly negative profile curvature, and moderately high wetness index, which represents a concave, high soil moisture region. These findings are consistent with all three elements of a landslide the scarp is the steepest concave part of the landslide; the body is the concave moderately steep, wet portion of the landslide and lastly; the toe is the convex, flatter dry region.

3.7 Artificial Neural Network Model

The type of ANN used within this study was a MLP with the BP learning algorithm. The MLP and BP were used to produce the landslide susceptibility of the study area. The BP algorithm may be regarded as a hard or soft classifier, but this study focused on the use of soft classification, to produce an output classification which is the direct result of multiple

classification maps, where each map is representing one class. The classification maps are combined together to show the degree of membership for the class being associated with a landslide. The degree of membership values range from 0 to 1, where the higher the degree of membership the more likely a landslide would occur within this area. Due to the use of fuzzy logic within the classification process the data may not always sum to 1.

Also, during the setup of the ANN model, the study area is normally separated into two similar, but distinct data sets, where a comparison of the model can be made. The two datasets are represented by the two study areas: the training and testing study areas (Gomez and Kavzoglu, 2005). As the name applies the training study area is where the model was produced and trained on and where the weights for the model were developed. The testing study area is where the developed weights from the training study area were applied to validate the performance of the model.

There are different parameters which can be changed within the MLP module, these include: the order or addition of different predictor attributes, the learning rate, momentum factor, automatic training, dynamic learning rate, root mean square (RMS) error, the amount of iterations, amount of training and testing pixels so that a ANN model can be produced with the least amount of error, and the highest accuracy between the training and testing portion of the model. Table 3 shows the effect that each parameter has on the modeling if the values are not optimal. From this table it is determined that the values for each parameter can limit the development, performance, and accuracy of the overall model (Basheer and Hajmeer, 2000). This was determined from multiple trials of modeling where the accuracy varied between the implementation of different parameters. For example, the amount of training and testing pixels per class affects the amount of data used from the given predictor attribute in the classification of

the study area. Depending upon the size of a given class, all or very little will be used within the classification and in turn affect the model.

Each parameter affects the model's ability to classify the data, for this reason there are certain recommendations for the initial value of each parameter. For instance, the recommendations for the initial values of the starting learning rate ranges between 0.1 and 0.2. This range of values will normally allow the training of the model to occur at a susceptible rate, but depending on the setup of the model as well as the input data, may require the value to occur outside of this range. The momentum factor also has a recommended range of 0.5-0.6 for initial modeling phases. This range help prevent oscillations that may occur within the training phase of the MLP ANN model. The other parameters discussed above contained the default initial values set by the MLP module and were adjusted accordingly throughout the modeling process.

After multiple trials with different parameters an ANN model was produced which best classified the training study area. The best model contained the predictor attributes in the following order: slope, annual solar radiation, upslope drainage area, plan and profile curvature, and wetness index. It was determined through trial and error that the training and testing pixel per class should be 500. The other parameters used within this model are described below:

Input layer node:	6
Output layer node:	2
Hidden layers:	1
Hidden layer nodes:	3
Starting learning rate:	0.005
Ending learning rate:	0.0001

Momentum factor:	0.05
Sigmoid function constant a:	1.0
Number of iterations:	10,000

The acceptable RMS value was set to stop at 0.001, and the number of iterations set at 10,000, which would be reached before the RMS value. For this reason different iteration values were examined, but higher values would produce over trained, lower accuracy models. The significance is that the model contains a low RMS, but that the cutoff RMS for the input data was set too low to be reached by this model. For this reason the termination iterations was set to stop modeling and not to overtrain the data. The best results from the ANN simulations yielded an RMS error of 0.2921 for the training, an RMS error of 0.3016 for the testing, with total accuracy rate of 87.93% for the final model. The model was further validated in the testing site using the weights developed in the training site.

4. RESULTS

4.1 Exploratory Data Analysis (EDA)

The field investigation of the LiDAR-derived landslides was followed by exploratory data analysis which allowed for each predictor attribute to be explored and analyzed so that a better understanding of the data can take place for both the landslide-prone areas as well as for stable areas within the training dataset. First, there were random points produced for the landslide and non-landslide localities within the training study area, which represented 1.33% and 0.005 % respectively. Figure 6 shows the univariate EDA of the predictor attributes used to visualize the distribution or range for each predictor attribute. The graphical representation is in the form of a probability density function (PDF) plot. The PDFs of each predictor attribute for landslide and non-landslide-prone areas allowed for a better understanding attribute distribution and statistical difference between the landslide and non-landslide populations. The EDA included the following attributes: elevation, slope, aspect, north, curvatures (profile, plan and mean curvatures), upslope drainage area, anisotropic coefficient of variation, annual solar insolation, sediment transport index, stream power index, general landform classification, and wetness index.

The PDFs for the final six attributes used as input variables within the final ANN model are shown in Figure 7. This distribution and spread between both of these populations can be observed within the PDF. The distributions are represented by a line, where the solid line within the figure shows landslide-prone areas, while the dashed line represents the non-landslide-prone areas. The distribution within the PDF for slope, upslope drainage area, and annual solar radiation are different with respect to landslide- and non-landslide prone areas. The slope contains the highest degree of dissimilarity between the two populations, while plan and profile

curvature as well as wetness index are very similar. For a more detailed example of the variability of the distributions take a closer look at the PDF of upslope drainage area; where the area for non-landslides contains a median value around 15 m^2, while the landslide prone areas contain a mean value of around 40 m^2. Whereas, the PDFs of wetness index, plan and profile curvature shows similar relationship for both landslide- and non-landslide prone areas. For example, the highest probability value which also represents the median for plan curvature is at or near 0, for non-landslide and landslide prone areas alike.

Another aspect of EDA was the use of Pearson Correlation analysis. Pearson correlation analysis is used to determine the correlation between two continuous variables. The values of this analysis can range from 0 to 1 and from -1 to 0, where a value of 0 represents no correlation and a value of 1 or -1 represents perfect correlation between the two variables. The Pearson correlation analysis for the final six attributes in Table 4 shows that the wetness index, slope and upslope drainage areas are highly correlated.

4.2 Artificial Neural Network Model Results

Figure 8 *(a)* and *(b)* shows the predicted or degree of membership of landslide susceptibility for a given pixel in both the training and validation study areas. The figure represents a visual representation of the degree of membership of landslide susceptibility for the study area. The values range from 0 to 1 where blue represents low susceptibility with a value at or near 0, and red represents high susceptibility with a value at or near 1. From this figure visual patterns of the susceptibility can be determined where the highest susceptibility regions occur within the valley sides within the smaller tributaries predominantly on the eastern side of the study areas. The lowest susceptibility regions occur along the valley top and bottoms where the slopes are minimal. To better understand which areas were highly susceptible compared to low

susceptible regions, within the training study area the membership image was overlaid by the known landslide scarps locations which were derived from the landslide inventory map and determined by the cluster analysis. An enlarged image of this relationship can be viewed in Figure 9 *(a)*, where the high susceptibility coincides with the scarp locations within the training study area. Figure 9 *(b)* shows the degree of membership for the testing study area and how well the known landslide scarps related to the high susceptible areas.

The variability of the high susceptible zones can be further explored within Figure 10, where the susceptibility is shown in a three-dimensional view. Figure 10 shows that the locations of the high susceptibility zones do not occur along all portions of the hillside, but do vary on what side of the valley and on the location on the hillside. From this figure it can be determined that the north, northwest and southwest sides of the valley contain the highest susceptibility, mostly constricted to the top of the hillside. However on the south and southeast sides of the valley the hillside is nearly absent of any high susceptibility and only when the valley migrates towards the southern direction the presence of high susceptibility regions along the top of the hillside occur within smaller adjacent tributaries. This variability with the susceptibility along the hillside directly correlates with the landslide scarps within Figure 9, which helps to determine visually that the ANN model is deciphering the scarps compared to non-landslide locations.

The classified landslide susceptibility is quantified in Figure 11, where (*a*) represents the goodness-of-fit from training and *(b)* validation study area. The degree of membership of landslide susceptibility was divided into five different classes which represent very low to very high susceptibility. Figure 11 *(a)* and *(b)* shows a light and dark gray bar graphs. The light gray bars represent the percentage of the total area for each membership class. Whereas, the dark

gray bars represents the percent of correctly predicted landslides within each membership class. Starting with Figure 11 *(a)* the very high susceptibility class or 0.8-1.0 correctly predicted 30 % with only representing 7% of the total area. However, the very low susceptibility class represented 78% of the total area with only correctly predicting 2% of landslides within the LiDAR inventory. Like *(a)*, *(b)* showed augmented results. Within the very high susceptibility class, the model correctly predicted 42% of known landslides represented 1.6% of total area. Furthermore, the very low susceptibility class represents 82% of the total area and contained only 1 percent of correctly predicted landslides. In conclusion, this figure demonstrated the majority of the known landslides occur within a small percentage of the study areas and the validation study area predicted more landslides within the high susceptibility class.

Figure 12 shows the relationship of the predictor attributes and the membership classes for the landslide scarps within the validation study area. This is shown as a series of box plots for every individual membership class. The box plots show the interquartile range for each attribute within a specific susceptibility class. The interquartile range shows the range from 25 to 75 % of the data, with the solid black line representing the median value for the predictor attribute. The main purpose of this figure is to better understand landslide susceptibility associated with predictor attributes. The relationship of each predictor attribute with respect to the individual susceptibility classes are described below.

Within the validation study area the slope contains a trend where high susceptibility is associated with steep slopes and very low susceptibility is associated with gentle slopes. The steep slopes are produced by the down cutting of the Cuyahoga River creating the valley within the recent geologic past. Modern stream processes have created steepened slopes along the bank of the Cuyahoga River and its tributaries. The wetness index contains an opposite trend, where

the wettest areas occur within the very low susceptibility class and the relatively drier areas occurs at the high susceptibility class. This relationship occurs due to the scarps of the landslides that are located at the steepened slopes where available soil moisture is removed in the down slope direction concluding a rain event. The upslope drainage area has similar trend to the wetness index, where the very low susceptibility class contains the relatively higher values and trends downward towards the very high susceptibility class. However, the annual solar radiation has a trend where the low susceptibility class contains smaller values, which represents a region containing less sunlight and more soil moisture. The trend increases towards the direction of the moderately susceptible class, where more sunlight and less soil moisture occurs at this location. The trend decreases again towards the high susceptibility class, where more soil moisture occurs and results in a greater probability of failure to occur in that location. Lastly, both plan and profile curvatures have very similar values close to zero to slightly negative for all susceptibility classes.

5. DISCUSSION

Data collection and data processing are important aspects of this research. The first step is the production of a hillshade map from processed LiDAR data. The hillshade maps were used to highlight different sides of the valley so that a landslide could be identified. The majority of the landslides where identified using sun azimuths of 45° and 315° and sun angle of 30°. However, the use of only four oriented hillshade maps could have placed an emphasis on a select number of the landslides present. To improve this, a systematic approach should be taken, where a hillshade map would be produced for an azimuth of every 10-20^{o}. This will insure that the majority of landslides are incorporated within the landslide inventory. The field investigation was used to check the accuracy of the LiDAR landslide inventory map. However, the field investigation occurred during the peak vegetation time period in July and August, which hindered the deciphering of landslides in the field. The use of GPS technology was used to help locate the exact landslide from the inventory map, but the methods used during the first trial of the field investigation placed a bias on the landslide boundaries. A solution to improve this problem would be to conduct the field investigation during non-growing season, where obscure landslide boundaries can be detected in the field.

With further investigation of the close-up three-dimensional landslide susceptibility image within Figure 10, it can be clearly identified that there is a preferred orientation to the landslide susceptibility. The landslide susceptibility orientation is based on consistently located which side of the valley as well as the position along the valley wall. The highest landslide susceptibility is associated with north, northwest and southwest directed valleys as well as at the top of the valley walls and adjacent to the bank of the river, where the river processes produced over-steepened slopes. These preferred orientations are associated with weather pattern

migration and the location of high percent slope within the region. The position within the Cuyahoga River Valley provides an exposure of weather patterns from the west and northwest through the valley, bringing a wide variety of precipitation including lake effect snow patterns (Nandi and Shakoor, 2009; Goddard, 1998). However, Nandi and Shakoor (2009) showed that the concentration of landslides with respect to amount of precipitation was found not to be as significant as the erodability of the soils, which was outside of the scope of this study. Furthermore, the high susceptibility associated along the rim of the valley is consistent with Nandi and Shakoor (2007), where the highest percentage of landslides occurred around a 35 ° or 70 % slope and these slope values are consistently near the top of the valley wall as well as adjacent to the erosional banks of the Cuyahoga River and the surrounding tributaries.

The use of LiDAR data provides an exceptional tool for deciphering landslide processes within the study area and around the world. LiDAR provides incredible benefits to the investigator compared to API, remote sensing and other landslide detection methods. LiDAR produces high resolution DEM and hillshade maps, which can be processed to remove vegetation and manmade objects. LiDAR enables the user to view the earth surface under forested areas like the Cuyahoga River Valley. LiDAR is also cost efficient with respect to field investigations in terms of pinpointing landslide locations prior to a field investigation so that a field check can confirm the presence of a landslide. Furthermore, LiDAR improves landslide inventory maps where unknown, remote, large, older landslides can be incorporated. Schulz (2007) showed that four times more landslides were deciphered with the use of LiDAR compared with API and field investigations alone. Likewise, Van Den Eeckhaut, et al. (2007) showed that the use of experts analyzed LiDAR derived hillshade maps were equivalent to field investigations. Their study showed that the use of multiple experts investigating the same study area can produce a high

quality landslide inventory map, even under forested areas. For these reasons LiDAR was the ultimate tool for deciphering landslides within the Cuyahoga River Valley.

Although there are numerous benefits LiDAR data provides to decipher landslide properties there are also some limitations. LiDAR data requires tremendous computer efficiency and memory which limits the size of study areas. Also the Schulz (2007) study found that the smallest landslide detected by their LiDAR survey was a little over 20 meters across, which is consistent with this study. Within older landslides, the lateral and the toe boundaries were especially difficult to distinguish compared to younger landslides. However, the scarp of the landslides was easily detected within both young and older landslides and provides an indicator of the presence of an older landslide. Even though there are limitations to LiDAR data it still provides the best avenue to detect landslides under forested areas, which dominates this study area.

6. CONCLUSION AND FUTURE WORK

Accurate prediction of landslide susceptibility is difficult due to the complex processes associated with landslides. To address this problem this study demonstrates the application of an ANN modeling method to predict landslide susceptibility using training and validation study areas within the northern portion of CVNP. The production of the landslide inventory map was the first step in the modeling of landslide susceptibility, which featured datasets derived from LiDAR. The LiDAR-derived hillshade maps improve the recognition and identification of landslides under highly forested areas and improved the quality and quantity of landslides identified within the inventory map compared with API alone. Furthermore, the LiDAR dataset allowed for the production of the topographic attributes which were narrowed down through the ANN modeling process to the following six: slope, annual solar radiation, upslope drainage area, wetness index, and plan and profile curvature. These LiDAR derivatives are a cost effective method for the identification of landslides compared with traditional API, remote sensing, and other landslide detection methods.

The production of the final model showed that the landslide susceptibility of this region can be determined using this type of modeling, where such large datasets are incorporated. The final ANN model classified the study area into a continuous degree of membership output which is between 0 and 1. This range was then reclassified into five distinct classes which help to better understand the landslide susceptibility of the study area. This can be examined by looking at the high susceptibility class within the validation study area within Figure 11*(b)*, where 42 % of the landslides were correctly identified in only 1.6 % of the total area within the validation study area. Finally, the ANN model has shown that regions where steep, lower solar radiated,

slightly concave hillsides have a higher potential of a landslide to occur now or in the future than gently sloping, sunny, convex hillsides.

There are some main aspects of the future work for this project. The first aspect should be to produce a higher accuracy field investigation, where the landslide boundary within the landslide inventory map is collected independently of the field investigation. The use of GPS technology allows for the landslide boundary in the field to be collected and compared with the LiDAR-derived landslide boundary within GIS software. Pixels from both boundaries can be compared to produce a better understanding of how accurate the landslide inventory is. Additionally, the use of different experts to decipher landslides within the LiDAR-derived hillshade maps, as well as the use of high resolution imagery, and image fusion techniques can also play a vital role in improving the quality of the landslide inventory map.

The field investigation should take place during the non-growing season, which allows for the landslide features to be easily detected and the boundaries clearly marked so the direct comparison can be made. A temporal landslide susceptibility study should take place comparing this study with a future LiDAR survey completed by the state of Ohio or from private funding. This will allow park official to monitor the changes in the landslide susceptibility of the study area and help with discriminating how fast processes are changing within the CVNP. Lastly, this model should be expanded to include the whole CVNP to produce a regional scaled landslide susceptibility model.

REFERENCES

Anderson, K., Croft, H., 2009. Remote sensing of soil surface properties. Progress in Physical Geography 33, 4, 457-473.

Ardizzone, F., Cardinali, M., Galli, M., Guzzetti, F., Reichenbach, P., 2007. Identification and mapping of recent rainfall-induced landslides using elevation data collected by airborne LiDAR. Natural Hazards Earth System Sciences, 7, 637-650.

Bain, A., 1873. Mind and Body: Theories of their relation, D. Appleton and Company, New York.

Barredo, J.I., Benavides, A., Hervas, J., Van Western, C.J., 2000. Comparing heuristic landslide hazards assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. International Journal of Applied Earth Observation & Geoinformation, 2, 1, 10-22.

Basheer, A., Hajmeer, M., 2000. Artificial neural networks: Fundamentals, computing, design, and application. Journal of Microbiological Methods 43, 3-31.

Booth, A.M., Roering, J.J., Perron, J.T., 2009. Automated landslide mapping using spectral analysis and high resolution topographic data: Puget Sound lowlands, Washington and Portland Hills, Oregon. Geomorphology, 109, 132-147.

Bork, E.W., Su, J.G., 2007. Integrating LiDAR data and multispectral imagery from enhanced classification of rangeland vegetation: A meta analysis. Remote Sensing of Environment 111, 11-24.

Cruden, D.M., 1991. A simple definition of a landslide. Bulletin of the International Association of Engineering Geology 43, 27-29.

Dai, F.C., Lee, C.F., Li, J., Xu, Z.W., 2000. Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environmental Geology, 40, 3, 381-391.

Dai, F.C., Lee, C.F., Ngai, Y.Y., 2002. Landslide risk assessment and management: An overview. Engineering Geology, 64, 65-87.

Dewitte, O., Jasselatte, J.C., Cornet, Y., Van Den Eeckhaut, M., Collignon, A., Poesen, J., Demoulin, A., 2008. Tracking landslide displacements by multi-temporal DTMs: A combined aerial stereophtogrammetric and LiDAR approach in western Belgium. Engineering Geology, 99, 11-22.

Egmont–Petersen, M., de Ridder, D., Handels, H., 2002. Image processing with neural networks–a review. Pattern Recognition, 35, 2279-2301.

Ermini, L., Catani, F., Casagli, N., 2005. Artificial neural networks applied to landslide susceptibility assessment. Geomorphology, 66, 327-343.

Fiorucci, F., Cardinali, M., Carla, R., Rossi, M., Mondini, A.C., Santurri, L., Ardizzone, F., Guzzetti, F., 2011. Seasonal landslide mapping and estimation of landslide mobilization rates using aerial and satellite images. Geomorphology, 129, 59-70.

Fleming, R.W., Taylor, F.A., 1980. Estimating the cost of landslide damage in the United States. United States Geological Survey Circular 832, 1-21.

Ford, J.P., 1987. Glacial and surficial geology of Cuyahoga County, Ohio. Ohio Department of Natural Resources, Report of Investigations 134, 1-29.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008. Comparing landslide inventory maps. Geomorphology, 94, 261-289.

Glenn, N.F., Streutker, D.R., Chadwick, D.J., Thackray, G.D., Dorsch, S.J., 2006. Analysis of LiDAR–derived topographic information for characterizing and differentiating landslide morphology and activity. Geomorphology, 73, 131-148.

Goddard, D., 1998. Dick Goddard's Weather Guide and Almanac for Northeast Ohio. Gray and Company, Cleveland.

Godt, J.W., Baum, R.L., Savage, W.Z., Salciarini, D., Schulz, W.H., Harp, E.L., 2008. Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS Framework. Engineering Geology, 102, 214-226.

Gomez, H., Kavzoglu, T., 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Engineering Geology 78, 11-27.

Gorsevski, P.V., Jankowski, P., Gessler, P.E., Jankowski, P., 2003. Integrating a fuzzy k-means classification and a bayesian approach for spatial prediction of landslide hazard. Journal of Geographical Systems, 5, 223-251.

Gorsevski, P.V., Gessler, P.E., Foltz, R.B., Elliot, W.J., 2006. Spatial Prediction of landslides hazard using logistic regression and ROC analysis. Transactions in GIS, 10, 3, 395-415.

Gorsevski, P.V., Jankowski, P., 2008. Discerning landslide susceptibility using rough sets. Computers, Environment and Urban Systems 32, 53-65.

Haneberg, W.C., 2005. The ins and outs of airborne LiDAR: An introduction for practicing engineering geologist. AEG News 48, 16-19.

Hansen, M.C., 1995. Landslides in Ohio. Geofacts, Ohio Department of Natural Resources, Division of Geological Survey, No. 8. Accessed 14 April 2011, http://www.dnr.state.oh.us/Portals/10/pdf/GeoFacts/geof08.pdf

Hengl, T., Gruber, S., Shrestha., D.P., 2003. Digital terrain analysis in ILWIS: Lecture notes and user guide. International Institute for Geoinformation Science and Earth Observations (ITC), Enschede.

Hengl, T., Reuter, H.I., 2009. Geomorphometry: Concepts, Software, Applications, Developments in Soil Science 33. Elsevier, Hungary.

Hopkinson, C., Chasmer, L.E., Sass, G., Creed, I.F., Sitar, M., Kalbfeisch, W., Treitz, P., 2005. Vegetation class dependent errors in LiDAR ground elevation and canopy height estimates in a boreal wetland environment. Canadian Journal of Remote Sensing, 31, 2, 191-206.

Hopkinson, C., Chasmer, L., 2008. Testing LiDAR models of fractional cover across multiple forest ecozones. Remote Sensing of Environment 113, 275-288.

Ioannilli, M., Paregiani, A., 2008. Automated unsupervised geomorphometric classification of Earth surface for landslide susceptibility assessment. In Proceeding of the International Conference on Computational Science and its Applications part 1, 268-283.

James, W., 1890. The Principles of Psychology. H. Holt and Company, New York.

Jones, M.L., Shakoor, A., 1989. Some landslide hazards in northern Summit County, Ohio. Bulletin of Association of Engineering Geologist 26, 3, 335-368.

Kanungo, D.P., Arora, M.K., Sarkar, S., Gupta, R.P., 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Engineering Geology, 85, 3-4, 347-366.

Lee, S., Evangelista, D.G., 2006. Earthquake–induced landslide–susceptibility mapping using an artificial neural network. Natural Hazards and Earth System Sciences 6, 687-695.

Lee, S., Min, K., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology 40, 1095-1113.

Lee, S., Ryu, J., Won, J., Park, H., 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Engineering Geology, 71, 289-302.

Lee, S., Ryu, J.H., Lee, M.J., Won, J.S., 2006. The application of artificial neural networks to landslide susceptibility mapping at Janghung, Korea. Mathematical Geology, 38, 2, 199-220.

Lippman, R.P., 1987. An introduction to computing with neural nets. Acoustical Speech and Signal Processing Magazine 4, 4-22.

Liu,X., Peterson, J., Zhang, Z., 2005. High–resolution DEM generation from LiDAR data for water resource management. In Proceedings of International Congress on Modeling and Simulation MODSIM05: Advances and Applications for Management and Decision Making, December 2005, Melbourne, Australia, 1402-1408.

Liu, X., 2008. Airborne LiDAR for DEM generation: Some critical issues. Progress in Physical Geography 32, 31-49.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 115-133.

McKean, J., Roering, J., 2004. Objective landslide detection and surface morphology mapping using high–resolution airborne laser altimetry. Geomorphology 57, 331-351.

Melchiorre, C., Matteucci, M., Azzoni, A., Zanchi, A., 2008. Artificial neural networks and cluster analysis in landslide susceptibility zonation. Geomorphology, 94, 379-400.

Miller, B.B., 1983. Late Quaternary fluvial history of the lower Cuyahoga River in northeastern Ohio. Journal of Great Lakes Research 9, 1, 97-105.

Nandi, A., 2007. Factors affecting the frequency and distribution of landslides in selected counties in northeast Ohio. PH.D Dissertation, Kent State University, Kent, Ohio, USA, 1-33.

Nandi, A., Shakoor, A., 2006. Preparation of a landslide, susceptibility map of Summit County, Ohio, USA, using numerical models. The Geological Society of London 660, 1-11.

Nandi, A., Shakoor, A., 2007. Application of logistic regression model for slope instability prediction in Cuyahoga River watershed, Ohio, USA. GeoRisk 2, 1, 16-27.

Nandi, A., Shakoor, A., 2009. A GIS–based landslide susceptibility evaluation using bivariate and multivariate statistical analysis. Engineering Geology 110, 11-20.

Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Engineering Geology, 97, 3-4, 171-191.

Ohio Geographically Referenced Information Program (OGRIP), 2006. Ohio Statewide Imagery Program, Ohio Office of Information Technology. Accessed November–December 28 2010, http://ogrip.oit.ohio.gov/

Pohana, R.E., 1992. Landslide remediation and prevention by the city of Cincinnati. Boulder, Colorado, Geological Society of America Abstracts with Programs 24, A204.

Pradhan, B., Lee, S., 2007. Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis using an artificial neural network model. Earth Science Frontiers 14, 6, 143-152.

Pradhan, B., Lee, S., 2010. Regional landslide susceptibility analysis using back propagating neural network model at Cameron Highland, Malaysia. Landslides 7, 13-30.

Pradhan, B., Lee, S., Buchroithner, M.F., 2010. A GIS back propagation neural network model and its cross–application and validation for landslide susceptibility analysis. Computers, Environment and Urban Systems, 34, 216-235.

Razak, K.A., Straatsma, M.W., Van Westen, C.J., Malet, J.P., De Jong, S.M., 2011. Airborne laser scanning of forested landslide characterization: Terrain model quality and visualization. Geomorphology, 126, 1-2, 186-200.

Regmi, N.R., Giardino, J.R., Vitek, J.D., 2010a. Assessing susceptibility to landslides: Using models to understand observed changes in slopes. Geomorphology, 122, 25-38.

Regmi, N.R., Giardino, J.R., Vitek, J.D., 2010b. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology, 115, 172-187.

Ruff, M., Czurda, K., 2008. Landslide susceptibility analysis with a heuristic approach in the Eastern Alps (Vorarlberg, Austria). Geomorphology 94, 314-324.

Santacana, N., Baeza, B., Corominas, J., De Paz, A., Marturin, J., 2003. A GIS–based multivariate statistical analysis for shallow landslides susceptibility mapping in La Pobla de Lillet Area (Eastern Pyrenees, Spain). Natural Hazards, 30, 281-295.

Schulz, W.H., 2004. Landslides mapped using LiDAR imagery, Seattle, Washington. United States Geological Survey Open–File Report 2004-1396, 1-11.

Schulz, W.H., 2007. Landslides susceptibility revealed by LiDAR imagery and historical records, Seattle, Washington. Engineering Geology 89, 67-87.

State of Ohio, 2008. State of Ohio Hazard Mitigation Plan, Section 2.5 Landslides, 126-137.Accessed November 2010, http://ohiosharpp.ema.state.oh.us/OhioSHARPP/Documents/SHMP/2-HIRA/SOHMP_Sec_2_5.pdf

Suzen, M.L., Doyuran, V., 2003. Data driven bivariate landslide susceptibility assessment using geographic information systems: A method and application to Asarsuyu catchment, Turkey. Engineering Geology 71, 303-321.

Szabo, J.P., 1986. Wisconsinan stratigraphy of the Cuyahoga Valley in the Erie Basin, northeastern Ohio. Canadian Journal of Earth Science 24, 279-290.

Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Research 33, 2, 309-319.

Van Den Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Nyssen, J., Moeyersons, J., Van Beek, L.P.H., 2005. The effectiveness of hillshade maps and expert knowledge in mapping old deep–seated landslides. Geomorphology, 67, 351-363.

Van Den Eeckaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L., 2006. Prediction of landslide susceptibility using rare events logistic regression: A case–study in the Flemish Ardennes (Belgium). Geomorphology, 76, 292-410.

Van Den Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Nyssen, J., Moeyersons, J., Van Beek, L.P.H., Vandekerckhove, L., 2007. Use of LiDAR–derived images for mapping old landslides under forest. Earth Surface Processes and Landforms, 32, 754-769.

Van Westen, C. J., 2000. The modeling of landslide hazards using GIS. Surveys in Geophysics 21, 241-255.

Varnes, D.J., 1978. Slope movement types and processes, in: Schuster, R.L., Krizek, R.J., (Eds.), Landslides, Analysis and Control, Special Report 176. Transportation Research Board, Washington, pp. 11-33.

Ventura, S.J., Irvin, B.J., 2000. Automated landform classification methods for soil–landscape studies, in: Wilson, J.P., Gallant, J.C., (Eds.), Terrain Analysis Principles and Applications. John Wiley and Sons, New York, pp. 267-294.

Werbos, P.J., 1975. Beyond regression: New tools for prediction and analysis in the behavior sciences. PH.D Dissertation, Harvard University, Cambridge, Massachusetts, USA.

White, G.W., 1982. Glacial geology of northeast Ohio. Ohio Department of Natural Resources Bulletin 68, 1-75.

White, G.W., 1984. Glacial geology of Summit County, Ohio. Ohio Department of Natural Resources Report of Investigations 123, 1-25.

Wilson, J.P., Gallant, J.C., 2000. Terrain Analysis Principles and Applications. Wiley and Sons, New York.

Yesilnacar, E., Topal, T., 2005. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek Region (Turkey). Engineering Geology 79, 251-266.

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat, Turkey). Computers and Geosciences 35, 1125-1138.

Yilmaz, I., Keskin, I., 2009. GIS based statistical and physical approaches to landslide susceptibility mapping (Sebinkarahisar, Turkey). Bulletin of Engineer Geology and Environment 68, 459-471.

APPENDIX A: FIGURES

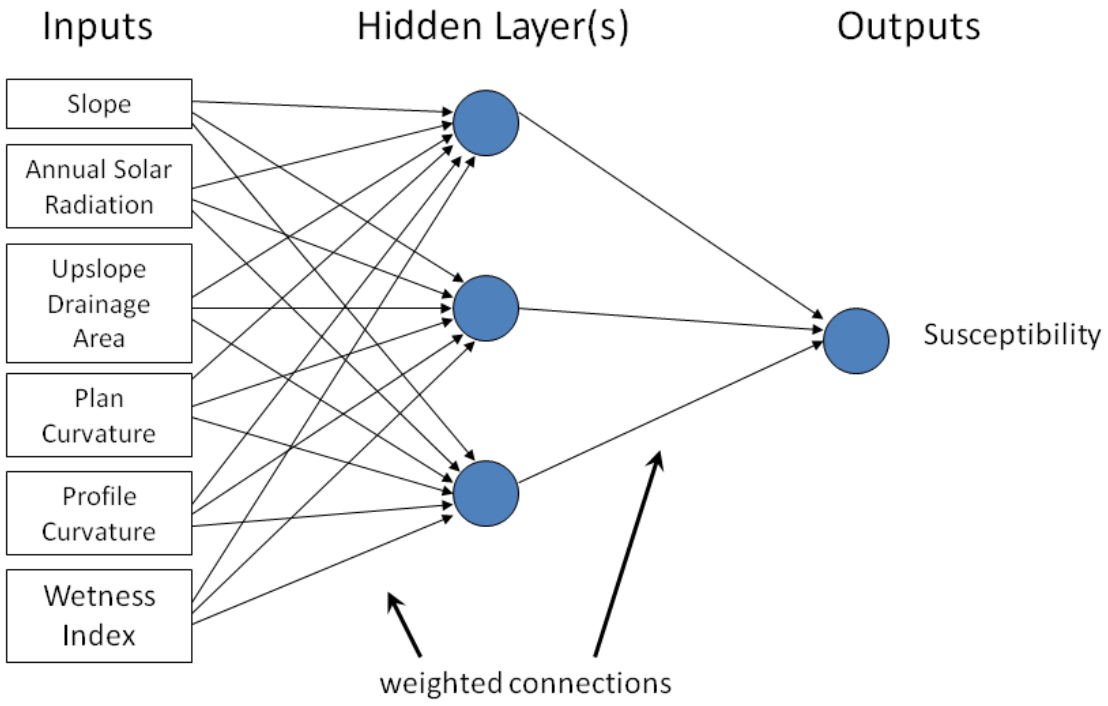

Fig 1. The Artificial Neural Network architecture used within this study.

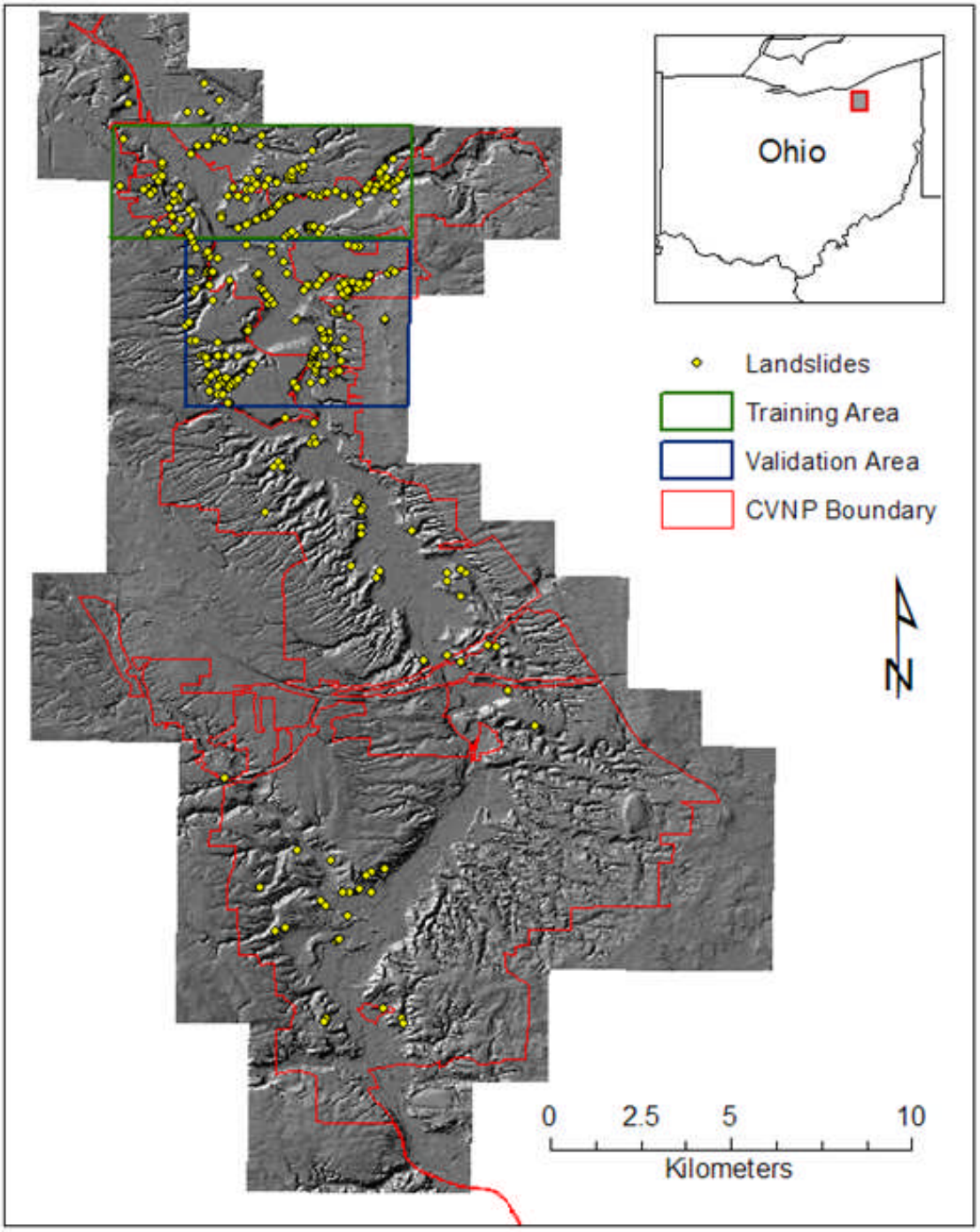

Fig 2. Distribution of training and validation landslides over the CVNP.

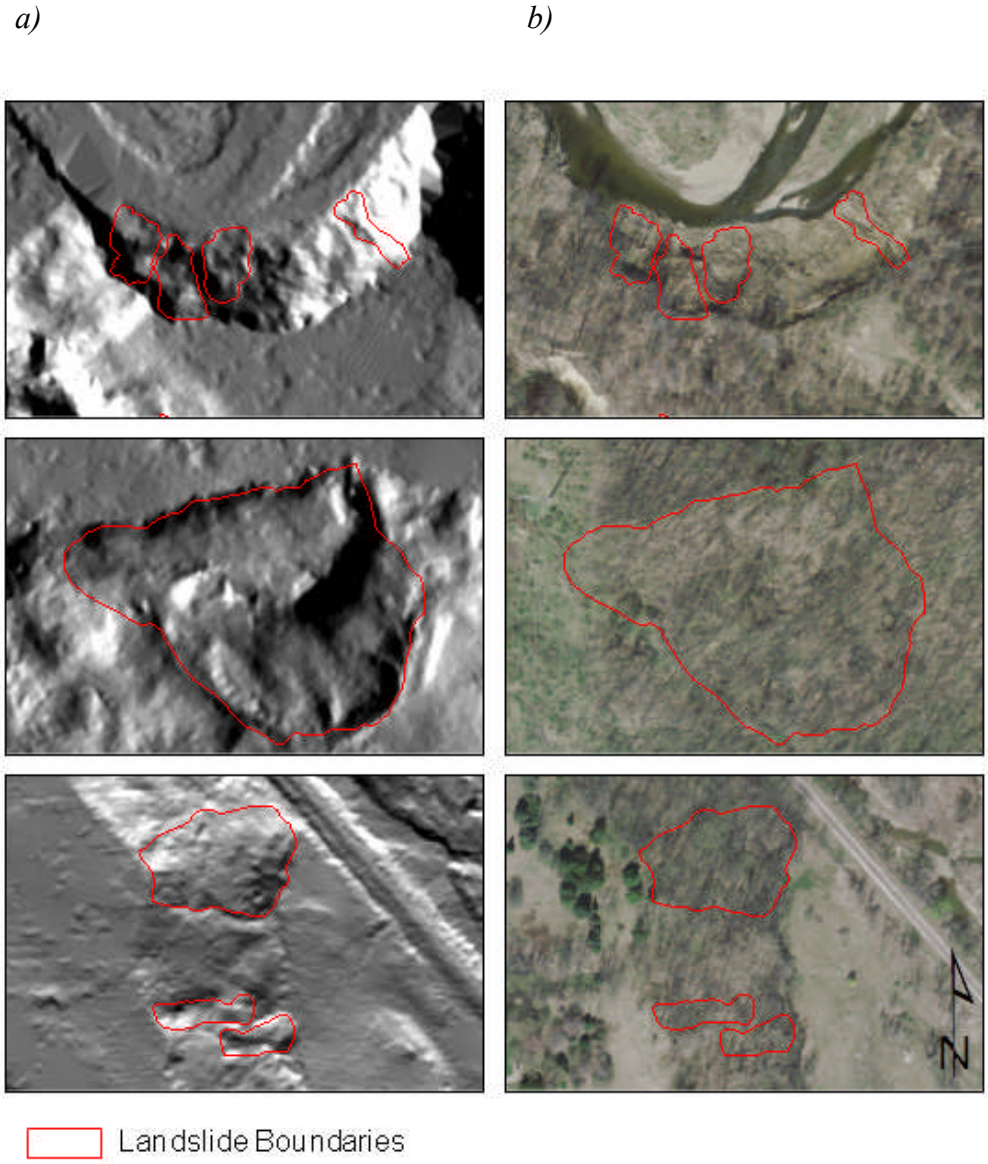

Fig 3. Landslide inventory from *(a)* LiDAR derived hillshade maps and *(b)* aerial photographs.

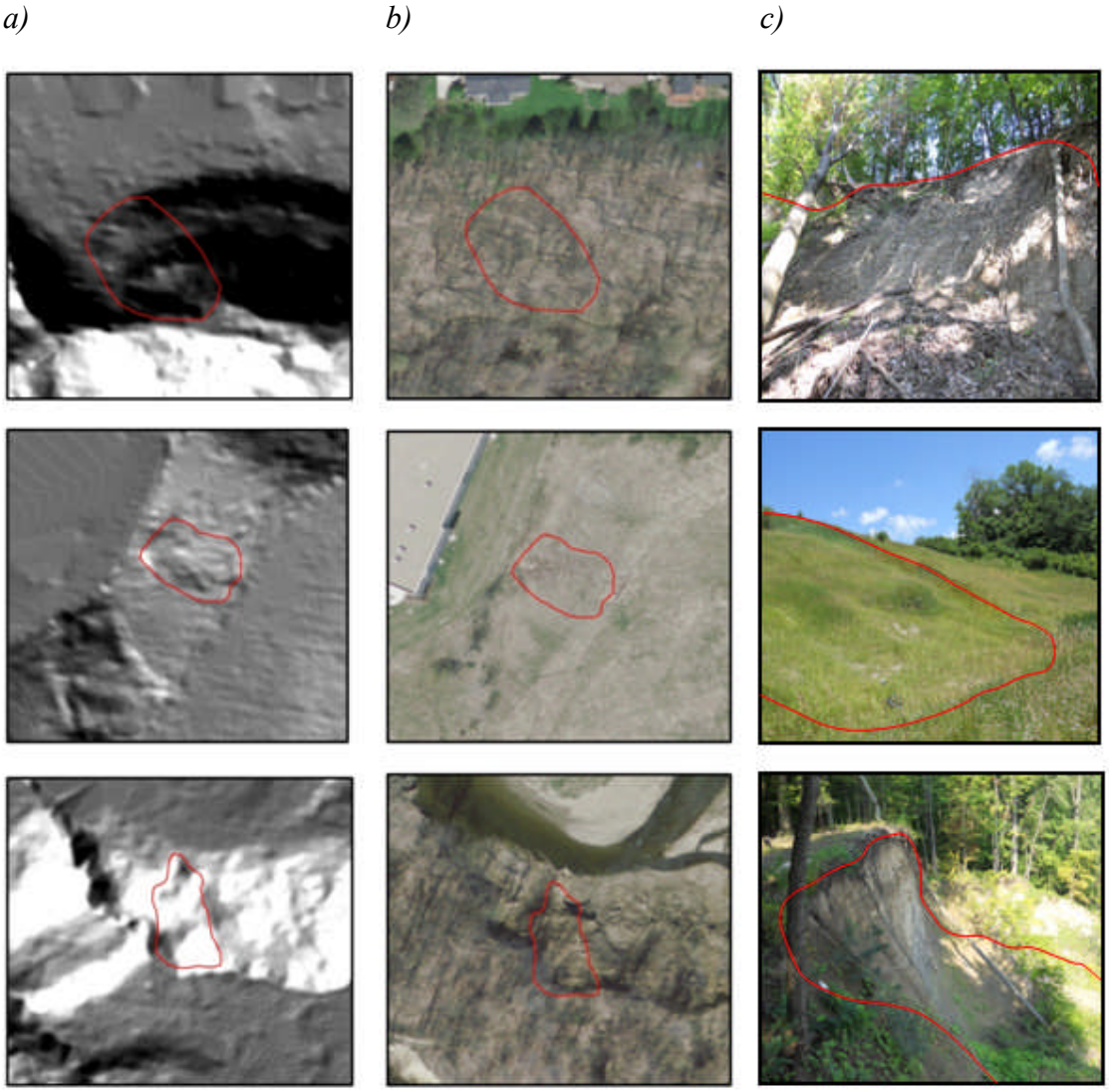

Fig 4. Displaying *(a)* LiDAR derived hillshade maps *(b)* with high resolution aerial photographs and *(c)* photographs taken in the field for a given landslide.

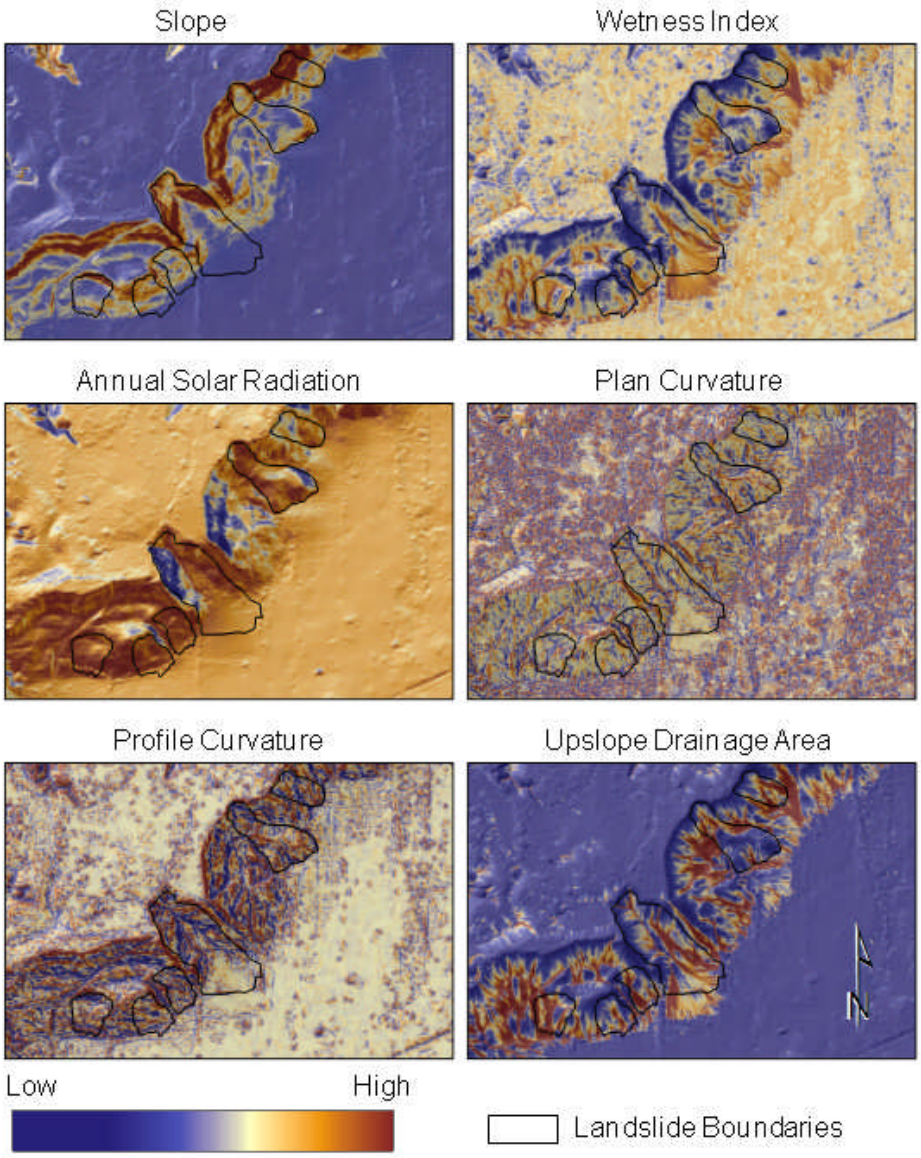

Fig 5. Predictor attributes derived from LiDAR data.

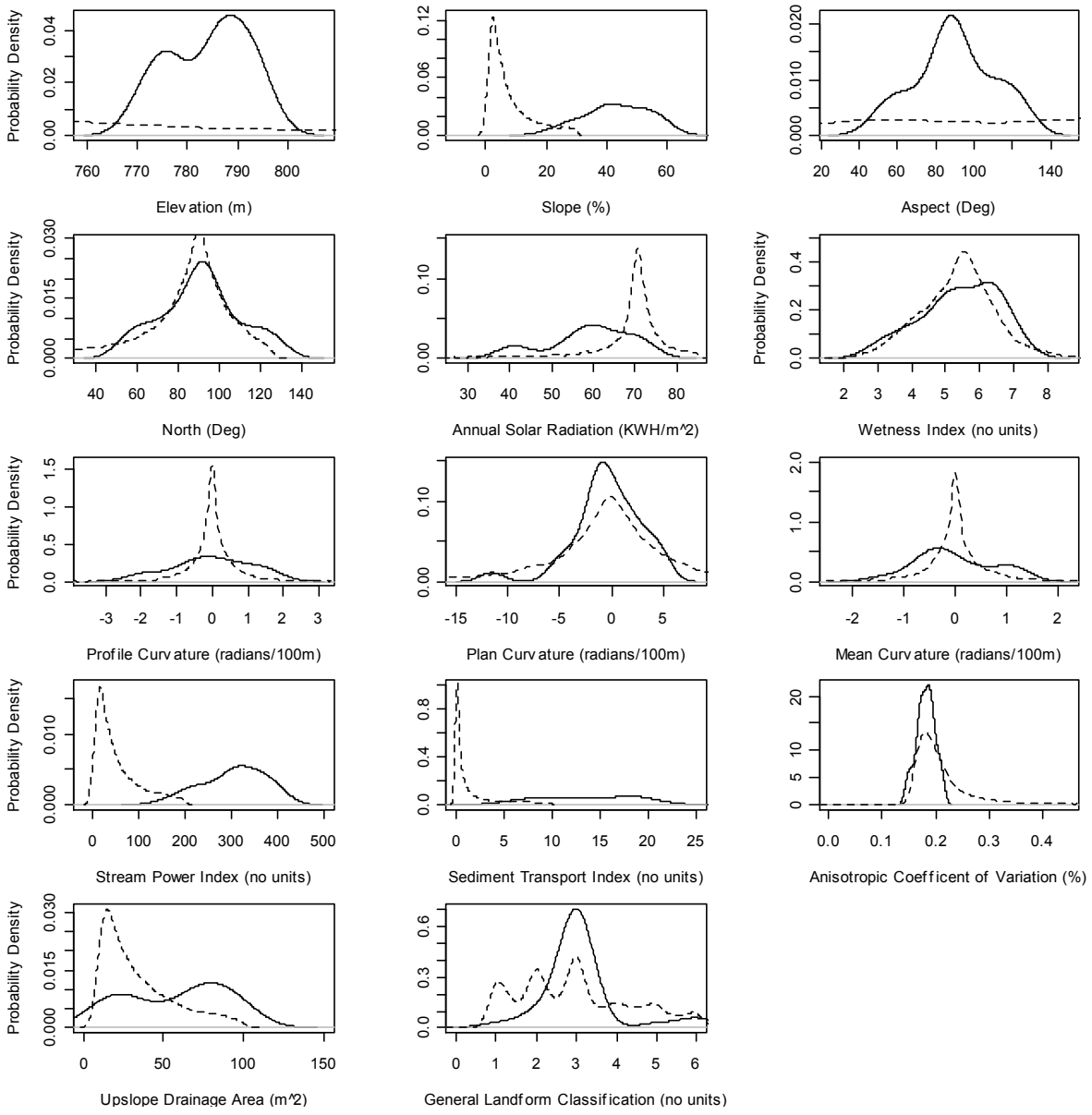

Fig 6. Showing the 14 topographic attributes used during the ANN modeling process. The solid line represents the landslide-prone areas and the dashed line represents the non-landslide-prone areas.

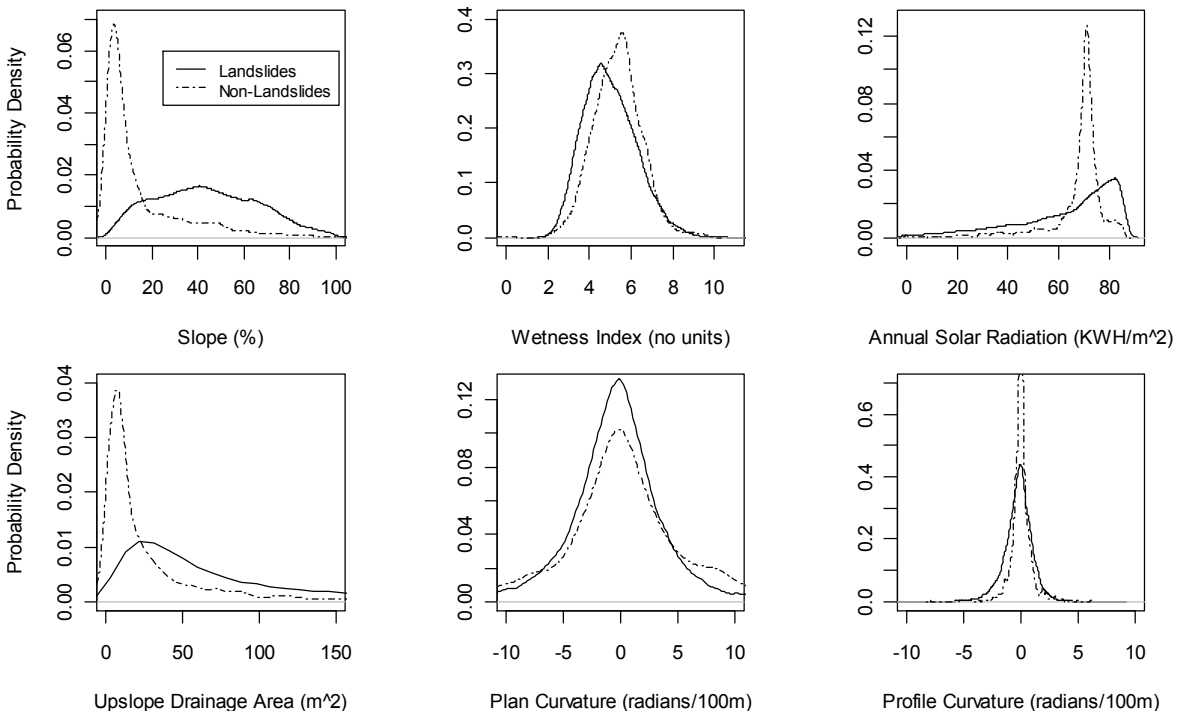

Fig 7. Probability density functions of landslides and non-landslide areas associated with individual topographic attributes.

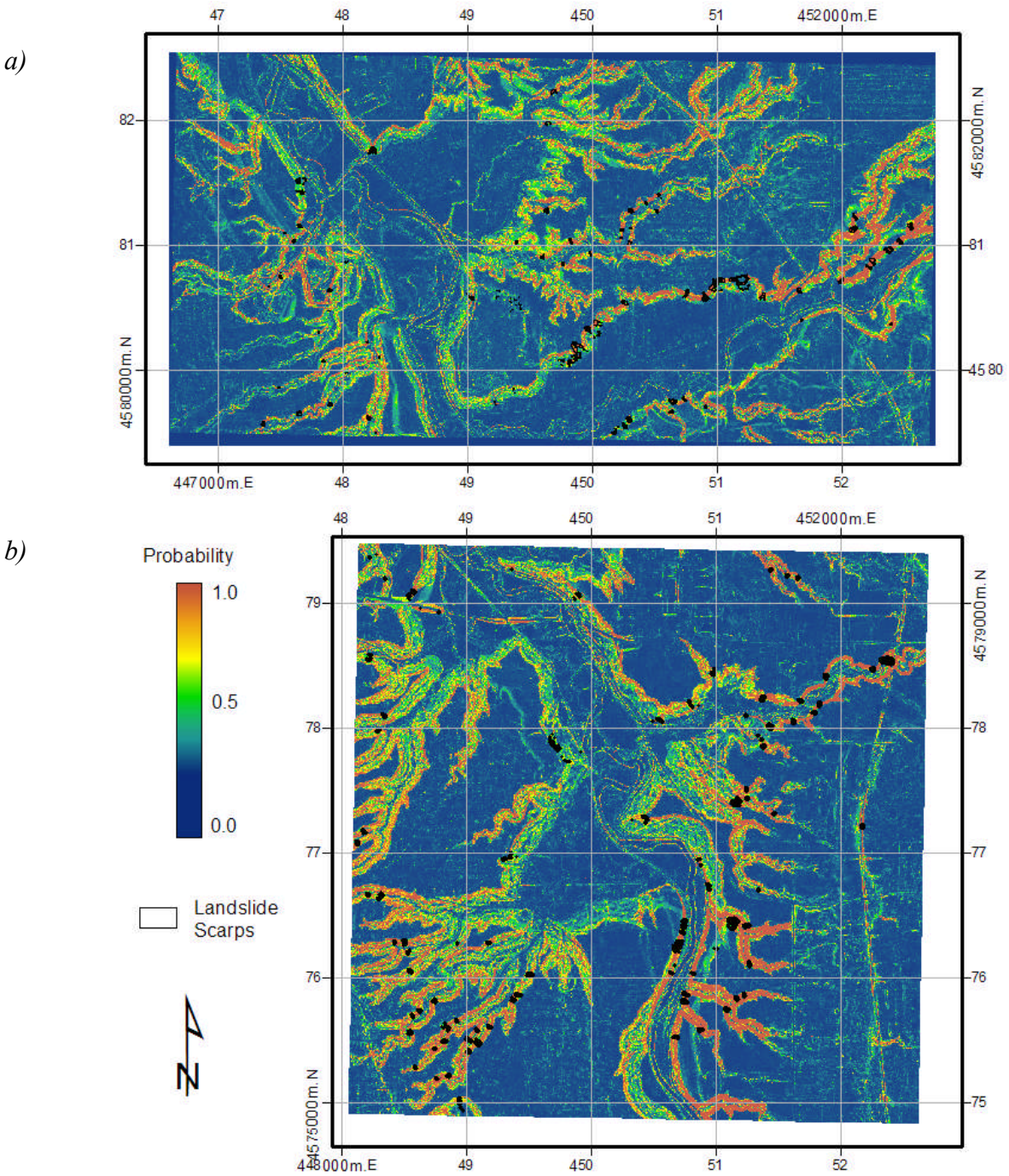

Fig 8. Probabilities of landslide susceptibility derived by ANN where *a)* is the training area used for the development of the ANN weights; and *b)* is the validation area used for testing the ANN weights from the training area.

a)

b)

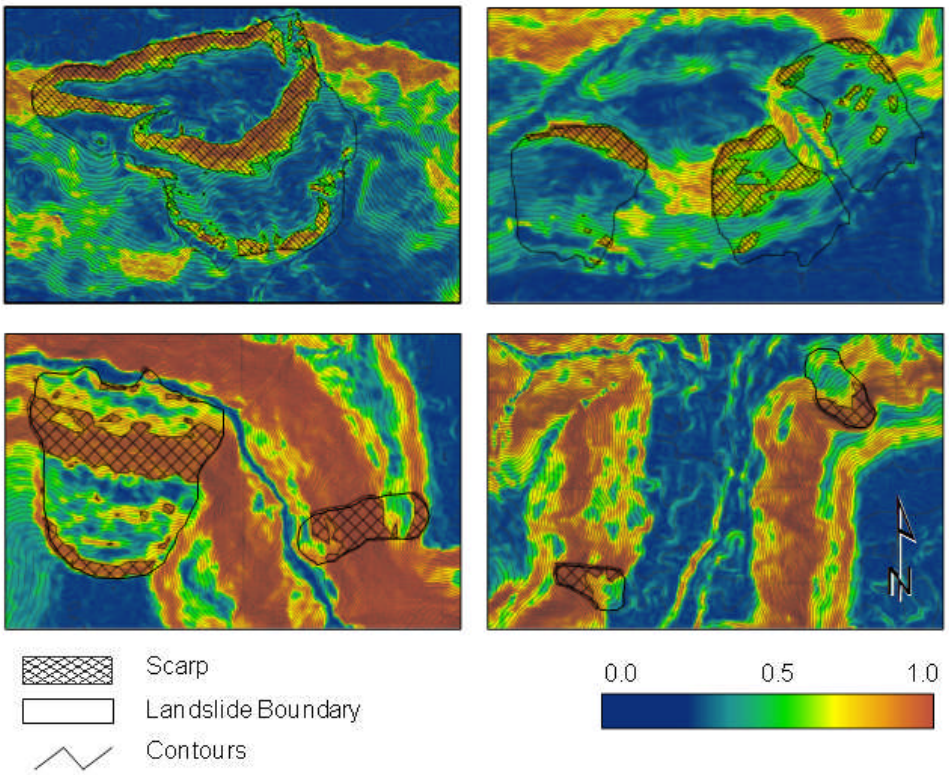

Fig 9. A zoomed in image of the ANN produced landslide susceptibility. Examples from the training and testing study areas are labeled *a)* and *b)* respectively.

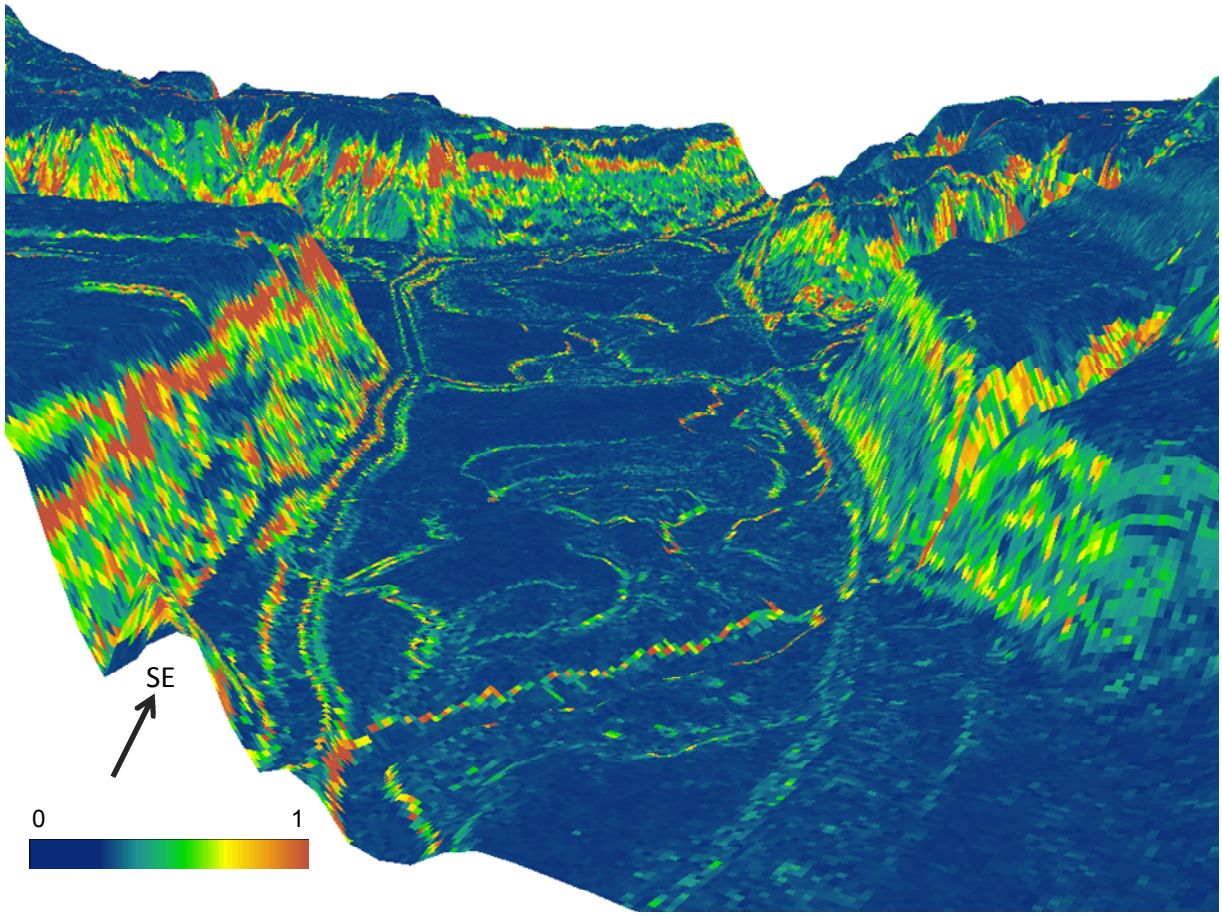

Fig 10. A close-up of landslide susceptibility associated with the training area.

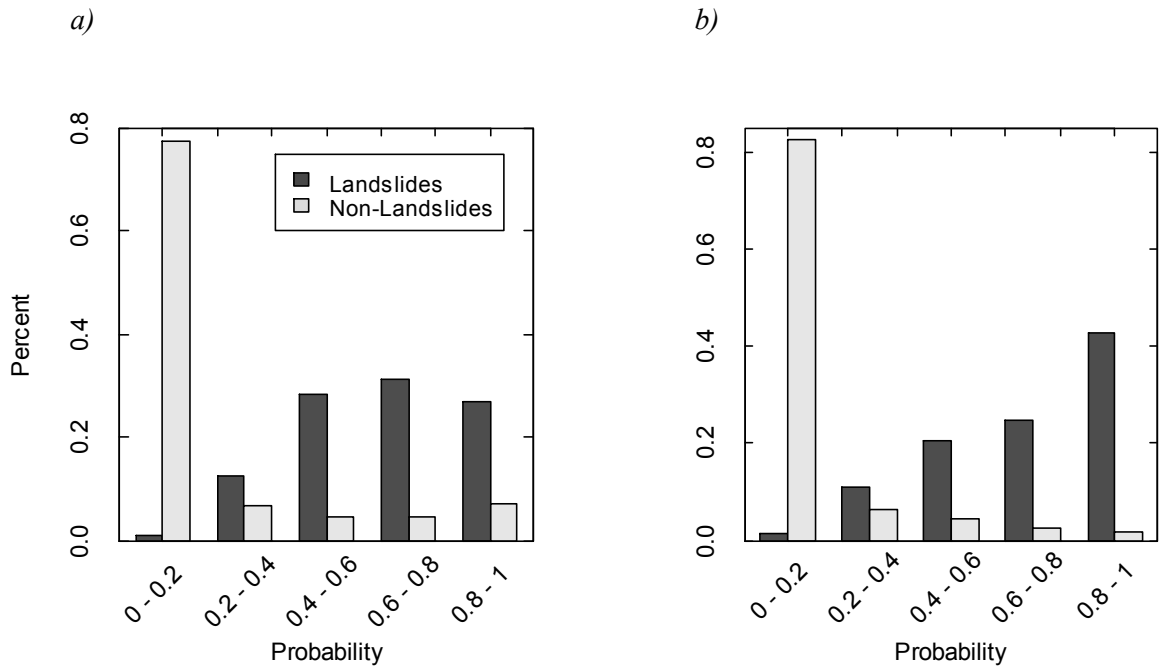

Fig 11. Proportion of correctly predicted landslides and assigned area to each probability class *a)* training area and *b)* validation area

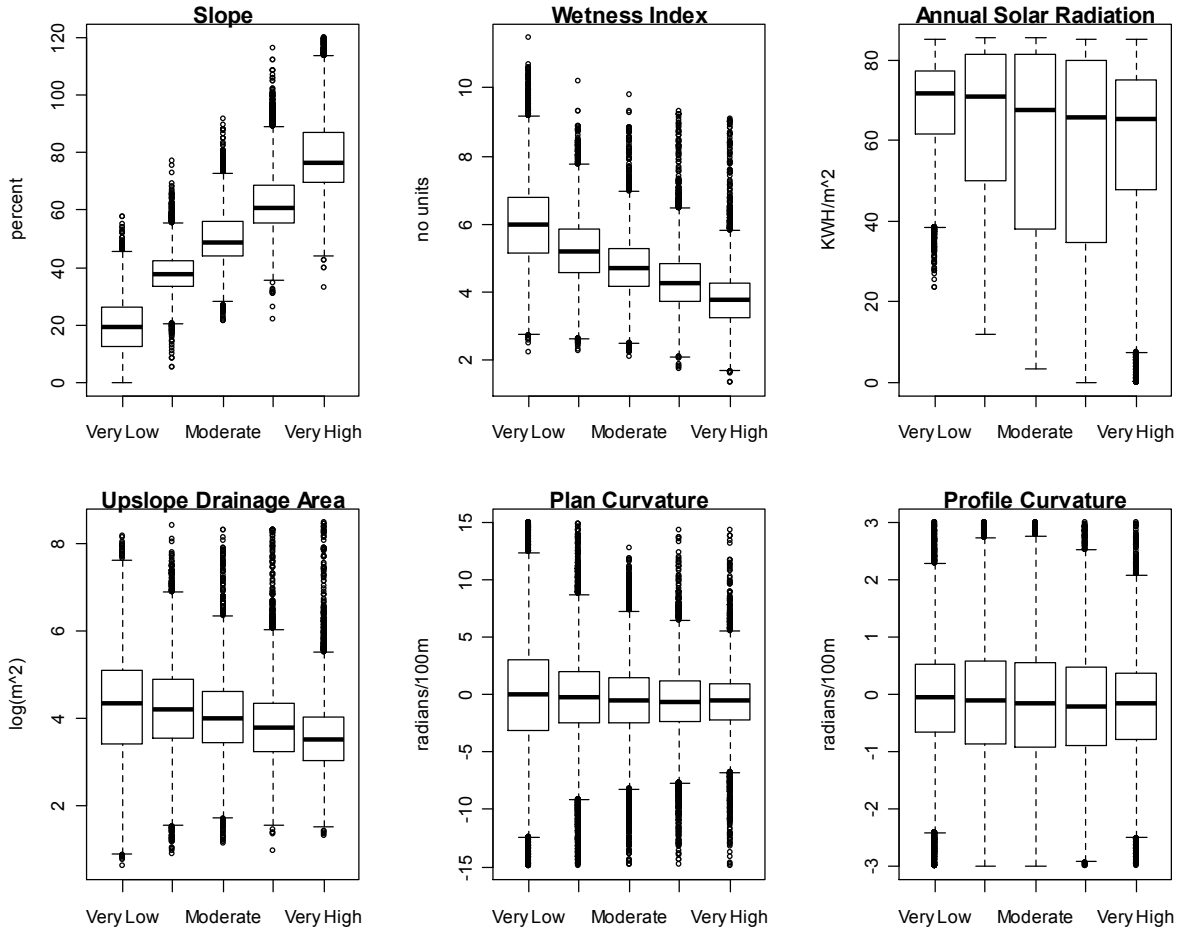

Fig 12. Box plots of modeled landslide susceptibility scarps vs. predictor attributes for the validation area.

APPENDIX B: TABLES

Table 1. Landslides statistics from the training and testing study areas.

	Minimum Landslide Area (ha)	Maximum Landslide Area (ha)	Mean Landslide Area (ha)	Total Landslide Area (ha)	Percentage of Total Area (%)
Training	4.03×10^{-3}	1.61	1.68×10^{-1}	13.95	0.75
Testing	6.94×10^{-3}	2.35	1.57×10^{-1}	16.81	0.80
Combined	4.03×10^{-3}	2.35	1.61×10^{-1}	30.82	0.78

Table 2. Clusters centers for the training and the validation study areas (C_3 is the scarp based on the cluster center values of the highest slope, concave, dry region; compared to C_1 which is the toe of the landslide. The toe contains cluster center values which represent a shallow slop, convex, wet reagion.C_2 represent the body of the landslides and contains cluster center values of a region dominated by relatively steep slope, concave, high soil moisture).

	Training			**Validation**		
Input data	**C*1***	**C*2***	**C*3***	**C*1***	**C*2***	**C*3***
Slope	17.67	40.85	74.52	23.03	35.49	65.86
Solar Radiation	70.77	63.67	48.48	71.02	68.13	54.61
Catchment	109.77	115.22	58.57	52.6	284.53	54.13
Plane Curvature	0.4	-0.79	-0.91	1	-2.93	-0.44
Profile Curvature	-0.04	-0.23	-0.22	-0.01	-0.62	-0.19
Wetness Index	5.86	5.2	4.11	5.22	6.48	4.17

Table 3. Effect of extreme values of ANN model parameters (Basheer and Hajmeer, 2000).

ANN Model Parameter	If Too High	If Too Low
Number of Training Pixels per Class	Slower training, higher likelihood of over training to take place	ANN model not capable of classifying the data, faster training, lower accuracy
Number of Testing Pixels per Class	Ability to validate ANN classification capability	insufficient validation of ANN classification capability
Number of Hidden Layers (1 to 3 layers)	More complex models, slower training, more likelihood of overfitting	One hidden layer provides an simple model , less likely of overfitting, faster training
Number of Hidden Layer Nodes	Increases network complexity, slows training, causes overfitting to occur.	Unable to classify complex data, produces a model with lower accuracy, but faster training
Learning Rate	Unstable model where no learning occurs, weights fluctuate widely	Slowing the rate of learning and training of the model and poor results
Momentum Factor	Speeds up training, increase risk of exceed the solution while training and, causes unstable learning	Increases the risk of entrapment in a local minimum in the error surface and, cause slower training
Number of iterations	Overfitting occurs in unseen data due to memorization of training data.	Produces an ANN which is unable to classify the data.

Table 4. The result of the Pearson Correlation Analysis for all predictor attributes in landslide prone areas.

	SLOPE	Wetness Index	Annual Solar Radiation	Upslope Drainage Area	Plan Curvature	Profile Curvature
SLOPE	-1					
Wetness Index	-0.7493	1				
Annual Solar Radiation	-0.3555	0.6337	-1			
Upslope Drainage Area	-0.6620	0.9358	-0.4669	-1		
Plan Curvature	-0.1181	-0.2426	-0.0464	-0.2484	1	
Profile Curvature	-0.0960	-0.1583	0.1677	-0.0308	0.3840	1