Slope Failure Prediction Using Random Forest Machine Learning and LiDAR in an Eroded Folded Mountain Belt

Aaron E. Maxwell [1],*, Maneesh Sharma [2], James S. Kite [1], Kurt A. Donaldson [2], James A. Thompson [3], Matthew L. Bell [1] and Shannon M. Maynard [2]

[1] Department of Geology and Geography, West Virginia University, Morgantown, WV 26505, USA; steve.kite@mail.wvu.edu (J.S.K.); mlb0083@mix.wvu.edu (M.L.B.);

[2] West Virginia GIS Technical Center, Morgantown, WV 26505, USA; maneesh.sharma@mail.wvu.edu (M.S.); kurt.donaldson@mail.wvu.edu (K.A.D.); smmaynard@mail.wvu.edu (S.M.M.)

[3] Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV 26506, USA; james.thompson@mail.wvu.edu

* Correspondence: Aaron.Maxwell@mail.wvu.edu; Tel: +1-304-293-2026

Abstract: The probabilistic mapping of landslide occurrence at a high spatial resolution and over a large geographic extent is explored using random forests (RF) machine learning; light detection and ranging (LiDAR)-derived terrain variables; additional variables relating to lithology, soils, distance to roads and streams and cost distance to roads and streams; and training data interpreted from high spatial resolution LiDAR-derivatives. Using a large training set and all predictor variables, an area under the receiver operating characteristic (ROC) curve (AUC) of 0.946 is obtained. Our findings highlight the value of a large training dataset, the incorporation of a variety of terrain variables and the use of variable window sizes to characterize the landscape at different spatial scales. We also document important variables for mapping slope failures. Our results suggest that feature selection is not required to improve the RF modeling results and that incorporating multiple models using different pseudo absence samples is not necessary. From our findings and based on a review of prior studies, we make recommendations for high spatial resolution, large-area slope failure probabilistic mapping.

Keywords: slope failures; landslides; light detection and ranging; LiDAR; digital terrain analysis; machine learning; random forest; spatial predictive modeling

1. Introduction

Slope failures, including landslides, are estimated to cause 25 to 50 fatalities and more than 3 billion dollars in damage each year in the United States alone [1–4]. Based on a review of government statistics, aid agency reports and research papers, Petley [5] estimates that 32,322 global fatalities from non-seismically induced landslides occurred between 2004 and 2010. Further, although uncertainty still remains, it has been suggested that global climate change may result in alterations in the global and local frequency, intensity and distribution of failures [6–10]. Thus, there is a need to develop methods to monitor and predict slope failure occurrence that are appropriate for mapping large spatial extents using available geospatial data and mapped reference locations.

Accurate and consistent geospatial data are of great importance in mapping and predicting slope failures, as they represent key factors that may contribute to or inhibit slope stability, such as

geomorphologic, lithologic, soil and land use characteristics [11–14]. As early as the 1970s and 1980s, researchers were using geospatial data and statistical modeling to assess these hazards [11–14]. More recently, machine learning methods have been applied to mapping, predicting and modeling slope failures [15–25]. Generally, machine learning methods have been successfully applied to a wide range of predictive modeling and classification tasks in the geospatial sciences, which is at least partially attributed to their ability to model complex patterns and relationships from a variety of input data without distribution assumptions [26–28]. Further, the more recent development of deep learning methods and convolutional neural networks are further expanding our ability to map and model slope failure features, as recently demonstrated by Sameen et al. [29] and Wang et al. [30].

Development of light detection and ranging (LiDAR) technologies and their application to mapping bare earth terrains over large spatial extents with high spatial detail has improved our ability to quantify and map geomorphic features and processes [31–34]. In the United States, the federal government has implemented the 3D Elevation Program (3DEP) (https://www.usgs.gov/core-science-systems/ngp/3dep) with the goal of providing LiDAR coverage for the entire country, excluding Alaska, which will be collected using interferometric synthetic aperture radar (InSAR) data [35,36]. Further, the United States Geologic Survey (USGS) is currently curating a nation-wide landslide inventory with contributions from local, state and federal agencies (https://www.usgs.gov/natural-hazards/landslide-hazards) [37]. Given the risks that slope failures pose and these recent developments in availability of quality digital terrain data, landslide inventories and computational and machine learning methods, we argue that there is a need to develop methods that leverage these data and algorithms to map and predict slope failures over large spatial extents.

This research explores the mapping of slope failures throughout the entirety of the 10,765 km^2 Northern Appalachian Ridges and Valleys Major Land Resource Area (MLRA) within the state of West Virginia. We make use of terrain data derived from LiDAR, additional geospatial data and the position of mapped slope failure head scarps interpreted from LiDAR-derivatives. The objective of this research is to provide recommendations for large-area slope failure mapping from LiDAR and additional geospatial data as highlighted by our results and previous studies. We specifically address the following questions:

(1). Does combining multiple models using different sets of pseudo absence data improve slope failure prediction? The use of pseudo absence samples is an approach to generate negative (i.e., no slope failure) examples and is explained in more detail in the Methods section.

(2). Does incorporating additional variables representing lithology, soil characteristics and proximity to roads or streams improve the model in comparison to just using terrain variables?

(3). How does reducing the training sample size impact model performance?

(4). How does predictor variable feature selection impact model performance?

(5). What variables are most important for predicting slope failure occurrence?

(6). Does calculating terrain variables using multiple window sizes improve the prediction?

Modeling is conducted using the random forest (RF) algorithm to obtain a probabilistic output as opposed to a classification of slope failure extents. In this study, we define slope failures as the movement of a mass of rock, earth or debris down a slope [38]. Our goal is to predict the likelihood of slope failure occurrence broadly, regardless of material or movement type. Debris flows, lateral spread and slides, both translational and rotational, are predicted.

1.1. Mapping Slope Failures and Susceptibility

Remotely sensed data, other ancillary geospatial data and machine learning have already been applied to mapping the extent of slope failures or predicting susceptibility [12,15,17–24,39–51]. Optical data have been used to map the extent of failures that have a distinct spectral signature, such as debris flows and other events that expose bare earth material [51–53]. For example, Stumpf and Kerle [51] combined geographic object-based image analysis (GEOBIA), very high spatial resolution

imagery and the RF algorithm to map failures over multiple study areas. There focus was to assess the use of optical data for disaster response and the mapping of recent events [51].

Since many failures may not have a distinct spectral signature due to age, canopy cover or spectral confusion with other landscape features and because there is a need to assess future risk along with inventorying existing failures, it is common to attempt to map the likelihood of occurrence or the future susceptibility to failure [11,15,20,21,25,33,42,45,54,55]. For example, Trigilia et al. [54] compared logistic regression (LR) and RF for shallow landslide susceptibility mapping from a variety of terrain, lithologic and land use variables. Using terrain variables only, Goetz et al. [45] compared generalized additive models (GAM), generalize linear models (GLM), weights of evidence (WOE), support vector machines (SVM), RF and bootstrap aggregated classification trees (bundling) with penalized linear discrimination analysis (BPLDA) for generating susceptibility models. Duo et al. [25] noted the value of SVM for predicting earthquake and rainfall induced landslide susceptibility in comparison to four other methods. We expand upon such studies by exploring the application of LiDAR, a variety of predictor variables and RF machine learning over a large spatial extent, which is uncommon in the literature. Although this study focuses on probabilistic mapping using the RF traditional machine learning method, is should be noted that deep learning methods that rely on convolutional neural networks have been explored for slope failure mapping and predictive tasks in several recent studies [25,29,30,56–61].

1.2. Random Forest for Spatial Predictive Modeling

RF is a nonparametric, ensemble decision tree (DT) method capable of accepting continuous and categorical predictor variables to perform classification, regression and make probabilistic predictions [62]. DTs rely on recursive binary splits of the data based on learned decision rules to divide the data into more homogenous subsets [63]. Ensemble methods combine multiple decision trees to potentially improve upon the predictive performance of a single tree [27,64]. For RF specifically, each tree in the ensemble is trained using a subset of available training samples, selected using bootstrapping or random sampling with replacement. Additionally, instead of defining an optimal split using any variable, only a subset of the predictor variables will be available at each node or split. The goal is to decrease the correlation between trees in the ensemble by providing each with a different set of training data and input features, resulting in a large number of weak classifiers that, when combined, act as a strong model with the ability to generalize well and not overfit to the training examples [62]. RF has many positive attributes for predictive modeling including its ability to accept a variety of input predictor variables that may be correlated and/or scaled differently. Also, it is generally robust to complex feature space, can be trained quickly, can accept categorical predictor variables and can provide an assessment of variable importance based on the withheld or out-of-bag (OOB), data in each tree [27,62].

The RF algorithm has been applied to many mapping and spatial predictive modeling tasks with remotely sensed and/or geospatial data as input [15,27,51,52,54,65–68]. It has also been used extensively to obtain probabilistic predictions as opposed to classification products. For example, Evans and Cushman [65] assessed the algorithm for predicting conifer tree species occurrence. Maxwell et al. [69] assessed the algorithm for predicting the likelihood of palustrine wetland occurrence based on topographic variables, while Strager et al. [70] used the algorithm to predict the likelihood of surface mine expansion.

Many studies have already assessed RF for probabilistic mapping of landslide occurrence or susceptibility (for example, References [15,21,45,55,68]); Goetz et al. [45] document strong performance for RF when applied to slope failure susceptibility modeling, as it generally outperformed the other tested methods and was not negatively impacted by a high-dimensional feature space and highly correlated variables. Trigila et al. [54] note the convenience of incorporating categorical predictor variables without the need to generate dummy variables. Taalab et al. [21] document that the algorithm can be used for both probabilistic susceptibility mapping and differentiation of failure types. Thus, prior research generally suggests that RF is an appropriate algorithm for slope failure mapping and based on our previous experience and also a review of the

literature we suggest that it is an optimal method to investigate this mapping problem over large spatial extents with a variety of input variables. Given that many prior studies have noted the value of this algorithm for slope failure mapping and modeling and its many positive characteristics as described above, no additional methods were investigated. Further, as highlighted in our review of the literature, prior studies already offer a comparison of different machine learning methods for this mapping task; as a result, algorithm comparison was not pursued in this study.

1.3. LiDAR and Terrain Variables for Mapping Slope Failures

LiDAR is an active remote sensing method that relies on laser range finding to generate accurate horizontal and elevational coordinates at a high spatial resolution from a terrestrial, airborne or satellite platform. An emitted photon, generally in the visible or near infrared spectrum, can strike an object and a portion of the energy can reflect back to the sensor for detection. Further, a single laser pulse can also be divided into multiple returns, allowing for vegetation canopy penetration and the mapping of subcanopy terrain features, in contrast to other elevation mapping methods, such as InSAR. Other than laser range finding, LiDAR also relies on global positioning system (GPS) measurements to reference the point cloud to a geospatial datum and an inertial measurement unit (IMU), which measures the orientation and motion of the aircraft [71]. As highlighted in the review by Jaboyedoff et al. [32], LiDAR offers detailed terrain and geomorphic information for characterizing and detecting the topographic signature of slope failure; however, there are some limitations, such as the expense and time required to collect and post-process the data, the lack of world-wide open and freely available data, the absence of available historic LiDAR data due to the only recent development of these technologies for mapping large spatial extents, and variability in the data in regards to point density and collection conditions.

A variety of terrain variables can be calculated from raster-based digital terrain models (DTMs), which can be generated from LiDAR data; further, Goetz et al. [72], Goetz et al. [45] and Mahalingam et al. [73] all suggest that terrain variables are highly important in predicting landslide occurrence and susceptibility. Goetz et al. [72] suggest that empirical or trained models that incorporate terrain variables often outperform physical models that attempt to model slope failure susceptibility based on our understanding of the physical processes that produce them. Table 1 provides some example terrain variables that have been used in different slope failure mapping and modeling studies. Note that this is not an exhaustive list and is simply meant to provide some examples. Also, many of these studies incorporate additional, non-terrain variables, such as variables relating to lithology, soil characteristics or land use, that are not summarized here. A review of these papers suggests that several variables have consistently been used in slope failure studies, such as topographic slope [74], topographic aspect [74], topographic wetness index (TWI) [75] and measures of curvature, such as profile curvature (PrC)—which measures curvature in the direction parallel to maximum slope—and plan or planform curvature (PlC), which measures curvature in the direction perpendicular to the maximum slope [76,77]. However, it does not appear that a consistent or optimal set of terrain variables have been determined. Also, many of these variables are calculated based on a neighborhood or moving window analysis [69,76–78] and studies have not generally investigated the effect of altering this window size. Thus, there is a need for further investigation of terrain variables for mapping and predicting slope failures.

Table 1. Example of terrain variables used in slope failure mapping or susceptibility studies [1].

Study [2]	Terrain Variables	Year	Algorithm(s)
Lee et al.	Curvature (non-directional), Slope	2004	LR
Ayalew and Yamagishi	Aspect, Elevation, Slope	2005	LR
Lee	Aspect, Curvature (non-directional), Slope	2005	LR
Yao et al.	Aspect, Curvature (Profile), Elevation, Slope, TWI	2008	SVM
Yilmaz	Aspect, Elevation, Slope, SPI, TWI	2009	ANN, FR, LR
Marjanović et al.	Aspect, Curvature (Plan), Curvature (Profile), Elevation, Slope, Slope Length, TWI	2011	SVM

Ballabio and Sterlacchini	CBL, CI, Downslope Distance Gradient, Elevation, HLI, Internal Relief, MPI, Slope, SPI, TWI	2012	SVM
Tien et al.	Aspect, Relief Amplitude, Slope	2012	DT, NB, SVM
Catani et al.	Aspect, Curvature (Plan), Curvature (Profile), Curvature (Classified), FA, Second Derivative of Elevation, Slope, TWI	2013	RF
Pradhan	Curvature (Plan), Elevation, Slope, TWI	2013	ANN, DT, SVM
Goetz et al.	Aspect, Curvature (Plan), Curvature (Profile), CH, CI, Elevation, FA, Slope, TR, TWI	2015	BPLDA, GAM, GLM, RF, SVM, WOE
Trigila et al.	Aspect, Curvature (non-directional), Curvature (Plan), Curvature (Profile), FA, Slope, SPI	2015	LR, RF
Mahalingam et al.	Slope, SPI, SR, TR, TWI	2016	ANN, DA, FR, LR, SVM, WOE
Pourghasemi and Kerle	Aspect, CTMI, Curvature (Plan), Curvature (Profile), Elevation, Slope, TWI	2016	RF
Youssef et al.	Aspect, Curvature (Plan), Curvature (Profile), Elevation, Slope	2016	Boosted DT, DT, GLM, RF
Chen et al.	Aspect, Curvature (Profile), Curvature (Plan), Elevation, Slope, TWI	2017	ANN, Maxent, SVM, Ensemble of methods
Hong et al.	Aspect, Curvature (Plan), Curvature (Profile), Elevation, Slope, SPI, STI, TWI	2018	Boosted DT, DT, RF
Kim et al.	Aspect, Curvature (non-directional), Slope, SPI, TWI	2018	Boosted DT, RF
Taalab et al.	Aspect, CTMI, Curvature (non-directional), Curvature (Plan), Curvature (Profile), Landform, Slope, TWI	2018	RF
Dou et al.	Aspect, Curvature (Plan), Drainage Density, Elevation, Slope	2019	ANN, CF, InV, PLFR, SVM

[1] CBL = Channel Base Level, CH = Catchment Height, CI = Convergence Index, FA = Flow Accumulation, HLI = Heat Load Index, MPI = Morphological Protection Index, SPI = Stream Power Index, SR = Slope Roughness, STI = Sediment Transport Index, TR = Topographic Roughness, TWI = Topographic Wetness Index, ANN = Artificial Neural Network, BPLDA = Bootstrap Aggregated Classification Trees with Penalized Linear Discrimination Analysis, CF = Certainty Factors, DA = Discriminant Analysis, DT = Decision Trees, FR = Frequency Ratio, GAM = Generalized Adaptive Model, GLM = Generalized Linear Model, InV = Information Value, LR = Logistic Regression, Maxent = Maximum Entropy, NB = Naïve Bayes, PLFR = Probabilistic Likelihood-Frequency Ratio, RF = Random Forest, SVM = Support Vector Machine, WOE = Weights of Evidence

[2] Studies are cited from [15–17,19,20–25,41,42,45,46,48,49,54,55,68,73].

2. Methods

2.1. Study Area

The 10,765 km² study area extent is defined relative to the Northern Appalachian Ridges and Valleys MLRA within the state of West Virginia (Figure 1). MLRAs are defined based on common patterns of physiography, lithology, soil, climate, water resources and land use [79]. This paper highlights findings for this specific MLRA; however, it is part of a larger project to assess slope failure occurrence across the entire state of West Virginia [80], which is ongoing at the time of writing. We have chosen to produce separate models for each MLRA in the state on the assumption that patterns and variable importance may vary based on landscape conditions and that a single model for the entire state would be inappropriate.
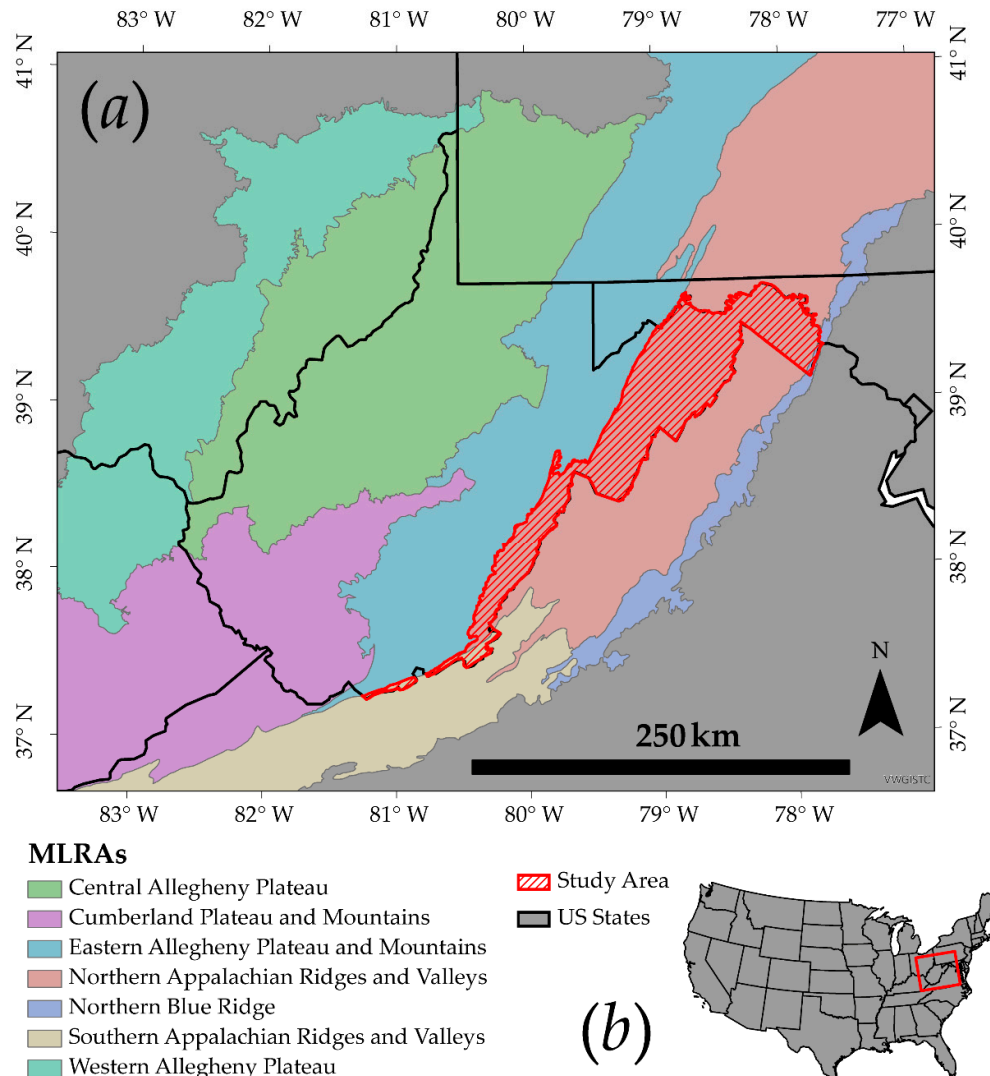
Figure 1. (**a**) Study area extent defined by Northern Appalachian Ridges and Valleys Major Land Resource Area (MLRA). (**b**) Location of study area within United States. MLRA boundaries were obtained from the Natural Resources Conservation Service (NRCS) of the United States Department of Agriculture (USDA).

The Northern Appalachian Ridges and Valleys MLRA is characterized by an eroded mountain belt of long, linear ridges and valleys, a pattern resulting from Paleozoic mountain building, folding and thrust faulting. Rock units vary in age from Precambrian to early Mississippian, with more recently formed units occurring in the western extent of the study area. The Precambrian exposure of metamorphosed basement rock occurs only in the extreme eastern extent of the study area in the Blue Ridge Mountains. The mapped extent also contains a portion of the Great Valley, which is relatively flat and dominated by Cambrian and Ordovician limestone, dolomite and shale. Within the folded mountain belt, valleys are commonly composed of shale and siltstone while ridges are supported by resistant sandstone and limestone [81–83]. Based on the 1:250,000 scale geologic map of West Virginia by Cardwell et al. [81], 30 geologic formations occur throughout the extent. The landscape is dominated by a trellis stream network with elevations ranging from 0 to 1400 m and an average elevation of 540 m. Mean annual temperature is near 0°C and yearly total precipitation is around 65 cm, though this can vary greatly based on elevation and topographic aspect; for example, east-facing slopes tend to receive less precipitation than west-facing slopes due to a rain shadow

effect. The region is dominated by oak-pine and oak-history forests, with large expanses of agriculture in the Great Valley [82].

2.2. Reference Data

Reference data were generated based on manual interpretation of hillshades and slopeshades produced at a 1 m spatial resolution. These raster-based representations of the terrain were derived from publicly available LiDAR data that cover the entire study area extent. These data were funded by the Federal Emergency Management Agency (FEMA) and are made freely available as part of the 3DEP program (https://www.usgs.gov/core-science-systems/ngp/3dep) and can also be obtained from the West Virginia GIS Technical Center and West Virginia View (http://data.wvgis.wvu.edu/elevation/). Hillshades are generated by modeling illumination over the landscape based on topography and the position of an illuminating source in the local sky [84] while slopeshades are created from a topographic slope model where dark shades represent steep slopes and light shades represent shallow slopes. In contrast to hillshades, slopeshades are illumination invariant, as they do not rely on modeling brightness relative to a specific illumination position [75,85].

Each feature was mapped as a point at the interpreted initiation location of the failure, generally the head scarp. So, each feature was mapped as an individual point as opposed to multiple points or an areal extent due to the difficulty of accurately mapping the full extent of the failure consistently and to reduce spatial autocorrelation in the training data. This process was completed by two trained analysts under the supervision of a professional geomorphologist. Debris flows, lateral spread and slides were mapped and all locations were interpreted by one analyst then verified by the other. A total of 1,798 slope failures were generated using this method. Examples are provided in Figure 2.
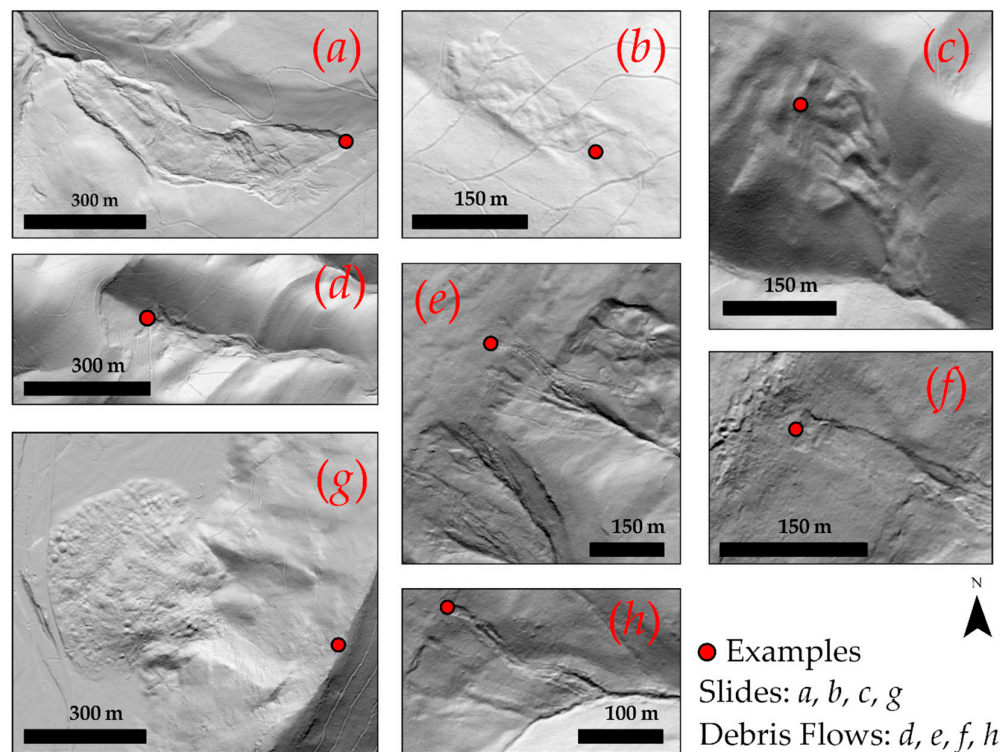
Figure 2. Example slope failure initiation points used to train the model. The hillshade was created from a 1 m resolution LiDAR-derived digital terrain model (DTM).

RF requires both presence and absence data to create a probabilistic prediction [62]. We generated pseudo absence data as 100,000 random points throughout the study area extent. Any random point that was within 30 m of a slope failure observation was removed. Also, additional slope failure data were obtained from the West Virginia Department of Transportation (WVDOT) and the West Virginia Geologic and Economic Survey (WVGES) and any random points that occurred within these mapped extents or within 30 m of them were removed. Given that a complete inventory of failures was generated, we argue that it is unlikely to randomly select a slope failure feature in the pseudo absence data. Similar methods were used by Strager et al. [70] for predicting future surface mine extents and Maxwell et al. [69] for predicting palustrine wetland occurrence.

2.3. Predictor Variables

Table 2 provides a list of all terrain predictor variables used in the study. A total of 43 variables are include, of which 32 represent terrain variables calculated from the LiDAR-derived DTM. All modeling was conducted at a 2 m spatial resolution. Most of the terrain variables were calculated within ArcGIS Pro 2 using built-in tools, such as the Slope Tool [86] or the ArcGIS Geomorphometry & Gradient Metrics Toolbox [87]. All curvature measures were produced using the Morphometric Features module from the open-source System for Automated Geoscientific Analysis (SAGA) software package [88–91]. Since many raster-based terrain calculations rely on local neighborhoods or moving windows to measure local patterns and compare a cell to its neighbors, the window size and shape can have an impact on the resulting measures and representation of the terrain [92,93]; therefore, we used multiple window sizes to calculate all terrain variables that rely on moving windows in order to capture patterns at multiple scales. Specifically, we used circular windows with radii of 7, 11 and 21 cells. These scales were decided upon based on measures of ridge-to-valley distance across the study area extent The curvature measures calculated in SAGA rely on second-order polynomials that can accept moving windows of variable size [89–91].

Table 2. Description of terrain variables used in study. Abbreviations defined in this table will be used throughout the paper.

Variable [1]	Abbreviation	Description [2]	Window Radius [3] (cells)
Slope Gradient	Slp	Gradient or rate of maximum change in Z as degrees of rise	1
Mean Slope Gradient	SlpMn	Slope averaged over a local window	7, 11, 21
Linear Aspect	LnAsp	Transform of topographic aspect to linear variable	1
Profile Curvature	PrC	Curvature parallel to direction of maximum slope	7, 11, 21
Plan Curvature	Plc	Curvature perpendicular to direction of maximum slope	7, 11, 21
Longitudinal Curvature	LnC	Profile curvature intersecting with the plane defined by the surface normal and maximum gradient direction	7, 11, 21
Cross-Sectional Curvature	CSC	Tangential curvature intersecting with the plane defined by the surface normal and a tangent to the contour - perpendicular to maximum gradient direction	7, 11, 21
Slope Position	SP	$Z - $ Mean Z	7, 11, 21
Topographic Roughness	TR	Square root of standard deviation of slope in local window	7, 11, 21
Topographic Dissection	TD	$\dfrac{Z - \text{Min Z}}{\text{Max Z} - \text{Min Z}}$	7, 11, 21
Surface Area Ratio	SAR	$\dfrac{\text{Cell Area}}{\text{Cosine}(slope * \pi * 180)}$	1

Surface Relief Ratio	SRR	$\dfrac{\text{Mean Z} - \text{Min Z}}{\text{Max Z} - \text{Min Z}}$	7, 11, 21
Site Exposure Index	SEI	Measure of exposure based on slope and aspect	1
Heat Load Index	HLI	Measure of solar insolation based on slope, aspect and latitude	1

[1] Variables are cited from [84,85,90,91,94–101].

[2] Max = maximum, Min = minimum, Z = elevation.

[3] A window radius of 1 is equivalent to a 3 by 3 cell window.

Other than terrain variables, we also calculated 11 additional features (Table 3). Distance to the nearest US, state and local road were calculated as three separate variables using the Euclidean Distance Tool in ArcGIS Pro 2 [86]. We also calculated the distance from all mapped streams using the same method. To further characterize these factors, we also calculated cost distance to roads and streams by weighting the distance relative to topographic slope, since failures may arise along steep slopes resulting from road construction or stream incision. This was accomplished using the Cost Distance Tool in ArcGIS Pro 2 [86].

Table 3. Additional predictor variables. Abbreviations defined in this table will be used throughout the paper.

Variable	Abbreviation	Description
Distance to Roads (US, state and local)	USD, StD, LoD	Euclidean distance to nearest US, state or local road
Cost Distance to US Roads (US, state and local)	USC, StC, LoC	Euclidean distance to nearest US, state or local road weighted by slope
Distance from Streams	StrmD	Distance from mapped streams
Cost Distance from Streams	StrmC	Distance from mapped streams weighted by slope
Geomorphic Presentation	Lith	Classification of rock formations based on geomorphic presentation
Dominant Soil Parent Material	DSPM	Dominant parent material of soil
Soil Drainage Class	SDC	Drainage class of soil

The three remaining variables are categorical and represent lithological and soil characteristics. A professional geomorphologist categorized all geologic formations in the extent based on their geomorphic presentation as defined in Table 4. We did not include the individual formations as a predictor variable due to the large number of categories. To characterize the soils, a soil scientist augmented the Soil Survey Geographic Database (SSURGO) [102] to derive measures of dominant soil parent material (DSPM) and drainage class. The derived categories are also listed in Table 4. Other than just providing additional information for the predictive modeling, including these variables also allowed us to incorporate expert knowledge into the prediction.

Table 4. Defined classes for lithologic and soil predictor variables.

Geomorphic Presentation (Lith)	Dominant Soil Parent Material (DSPM)	Soil Drainage Class (SDC)
Low relief carbonates	Colluvium	Excessively drained
Low relief mudrock	Disturbed areas	Somewhat excessively drained
Major ridge formers	Lacustrine	Well drained
Moderate or variable quartzose ridge formers	Marl	Moderately well drained
Moderate relief clastic rocks	Mine regolith	Somewhat poorly drained

Other	Old alluvium	Poorly drained
Shaley units with interbedded sandstone	Recent alluvium	Very poorly drained
	Residuum, acid clastic	
	Residuum, calcareous clastic	
Variable low ridge or hill forming carbonates with chert or sandstone	Residuum, Limestone	
	Residuum, metamorphic/igneous	
	Water	

As highlighted in the literature review above, a consistent set of variables has not been determined for slope failure likelihood prediction tasks. The variables in this study were selected based on suggestions from the literature, initial experimentation, data availability and expert knowledge. Although a larger number of features could be evaluated, we argue that the variables generated for this study offer a detailed representation of the geomorphic, soil and lithologic characteristics of the terrain even given data availability and additional limitations.

All raster-based variables were then extracted at the mapped slope failure and pseudo absence point locations using the Extract Multi Values to Points Tool in ArcGIS Pro 2 [86] in order to generate tables from which to extract training and validation data.

2.4. RF Modeling and Validation

The randomForest package [103] within the open-source data analysis software R [104] was used to generate and validate the RF models. Of the 1798 available slope failure samples, 1500 were randomly selected to use for training while the remaining 298 were withheld for validation. As one goal of this study is to assess whether incorporating a variety of pseudo absence samples can improve the model performance and also to avoid model bias resulting from an imbalanced training sample, we paired the 1500 training samples with five sets of non-overlapping pseudo absence samples using random sampling without replacement, resulting in five training sets containing all 1500 slope failure samples and a different set of pseudo absence samples. This process resulted in five training datasets, each with 3000 samples and a validation dataset with 596 samples, all of which contain an equal number of samples in the presence and absence class. A model was then trained using each training set and 500 trees, as this was found to be adequate to stabilize the results. The mtry parameter, which defines the number of variables available for splitting at each node in the multiple decision trees, was optimized using 5-fold cross validation and 10 values were tested. Hyperparameter optimization was performed separately for each model. All five models were then combined into a single model containing 1500 trees. In order to compare models using less variables or training samples, models were also generated using feature and training sample subsets.

Variable importance measures produced by RF have been shown to be biased if variables are highly correlated [105,106]. As demonstrated in Figure 3, which compares correlation between a subset of the terrain variables based on Spearman's rho [107], variable correlation is an issue in this study. Further, calculating the same measure at different window sizes result in sets of highly correlated variables; for example, Spearman's rho between the three SP measures were all above 0.80. So, we used a measure of variable importance based upon conditional random forests that takes into account correlation in the importance calculation as implemented in the R party package [105,106]. In order to explore the impact of feature space reduction, we used a feature selection method from the rfUtilities R package [65], which selects variables using RF-based variable importance estimates.
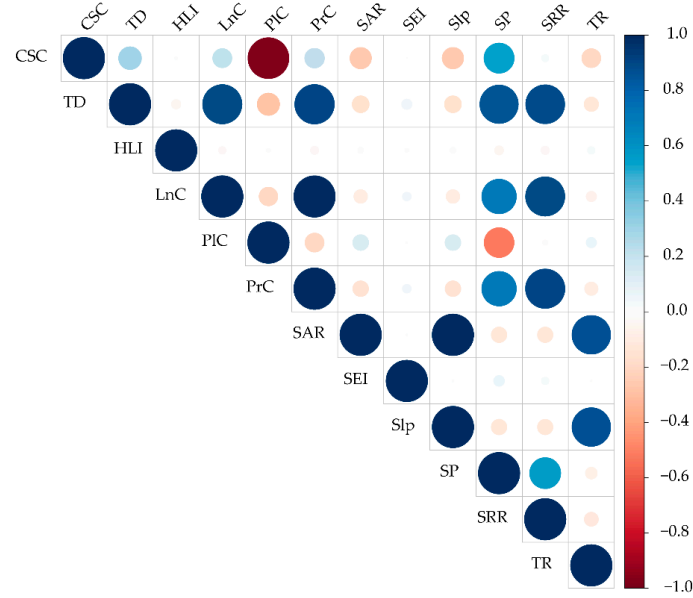
Figure 3. Correlation between a subset of terrain variables. Cross sectional curvature (CSC), topographic dissection (Diss), longitudinal curvature (LnC), planform curvature (PLC), slope position (SP), surface relief ratio (SRR) and terrain roughness (TR) were calculated using a 7-cell radius circular window. Both the size and color of the circle symbol represent the magnitude of correlation based on Spearman's rho.

Since our product is a probabilistic prediction as opposed to a classification, models are assessed and compared using receiver operating characteristic (ROC) curves and the area under the ROC curve (AUC) measure as implemented in the pROC R package [108–110]. An ROC curve plots the true positive rate against the false positive rate at various thresholds. The AUC measure is the area under the ROC curve and is equivalent to the probability that the classifier will rank a randomly chosen positive (true) record higher than a randomly chosen negative (false) record. Generally, values over 0.9 indicate excellent prediction rates [108–110]. To statistically compare models, we also made use of Delong's test for two ROC curves, which provides a *p*-value for statistical comparison of ROC curves [109,111]. Note that a balanced validation sample was used in this study, as ROC curves have been shown to be misleading when applied to imbalanced datasets [112].

To further assess the classification results, we calculated overall accuracy and the Kappa statistic using a 0.5 probability threshold. We also calculated precision, recall, specificity and the F1 score relative to the slope failure class using the number of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) withheld validation samples. Precision represents the portion of the predicted slope failures that were slope failures while recall represents the ratio of correctly predicted slope failures relative to the total number of slope failures. Specificity represents the proportion of not slope failure locations that are correctly identified as not slope failure. The F1 score is the harmonic mean of precision and recall [112]. The equations for these metrics are provided below in equations 1 through 4. Lastly, to provide an additional measure of performance that does not rely on selecting a threshold, we also calculated the area under the precision-recall curve (AUC (PRC)) using the PPROC package in R [29,113,114].

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{3}$$

$$\text{F1 Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{4}$$

In order to make predictions across the full spatial extent and at each 2 m cell location, the trained model was applied to the raster-based predictor variables using a combination of R and Python scripts. Since all predictor variables across the full study area extent gridded at a 2 m cell size sum to several terabytes of data, it was not possible to generate the prediction across the entire extent at once. Instead, predictions were made over 858 4-by-4 km tiles with a 100 m overlap to avoid data gaps. Also, terrain variables were derived for each tile prior to performing the prediction then subsequently deleted, which allowed for the model to be generated without excessive storage requirements. Once all tiles were processed, they were merged to generate a continuous probabilistic prediction across the entire study area extent.

3. Results

3.1. Impact of Combining Multiple Models

AUC calculated from the withheld validation samples for each separate model and the combined model are provided in Table 5. AUC varied by only 0.006 between all the models; further, based on Delong's test statistical difference between pairs of models was only observed between Model 5 and the combined model (*p*-value = 0.021). This generally suggests that providing a wide variety of pseudo absence examples to train multiple models did not improve the classification performance; thus, the sampling scheme used here was not necessary to stabilize the prediction. This result if further support by the AUC (PRC) metric and all threshold-based metrics, as metrics were similar for all models and the combined model.

Table 5. Results for each model and the combined model. Lower and upper bounds represent a 95% confidence interval calculated using 2000 stratified bootstrap replicates. X indicates statistically significantly different ROC curves at a 95% confidence level.

	Model 1	Model 2	Model 3	Model 4	Model 5	Combined
AUC	0.945	0.942	0.942	0.946	0.940	0.946
AUC Lower	0.928	0.925	0.926	0.925	0.924	0.930
AUC Upper	0.962	0.958	0.959	0.959	0.957	0.962
AUC (PRC)	0.945	0.943	0.945	0.944	0.942	0.949
Kappa	0.748	0.748	0.738	0.738	0.715	0.742
Overall Accuracy	87.4%	87.4%	86.9%	86.9%	85.7%	87.1%
Precision	0.839	0.833	0.831	0.831	0.816	0.834
Recall	0.926	0.936	0.926	0.926	0.923	0.926
Specificity	0.822	0.812	0.812	0.812	0.792	0.815
F1 Score	0.880	0.882	0.876	0.876	0.866	0.878
Model 1						
Model 2						
Model 3						
Model 4						
Model 5						X
Combined						

Figure 4 represents the distribution of predicted probabilities for the withheld validation data using a kernel density function. In support of the 0.946 AUC value obtained for the combined model, this plot suggests a strong separation between slope failure samples and random pseudo absence data. The median probability for the slope failure locations is 0.84 while the median probability for the pseudo absence points is 0.15. Of the slope failure points, 92.6% have a predicted probability higher than 0.50 while only 18.5% of the pseudo absence data have a probability higher than 0.50. Using a probability threshold of 0.5, the overall accuracy for predicting the validation data is 87.1% and the Kappa statistic is 0.742. For the slope failure class specifically, precision is 0.834 and recall is 0.926. The resulting prediction across the entire mapped extent and some example areas at a larger scale are provided in Figure 5. Red areas are those that are predicted to have a high likelihood of slope failure occurrence while green areas are predicted as having a low likelihood. Based on a visual

inspection, the figure suggests a strong relationship between predicted occurrence and topographic slope and incision.
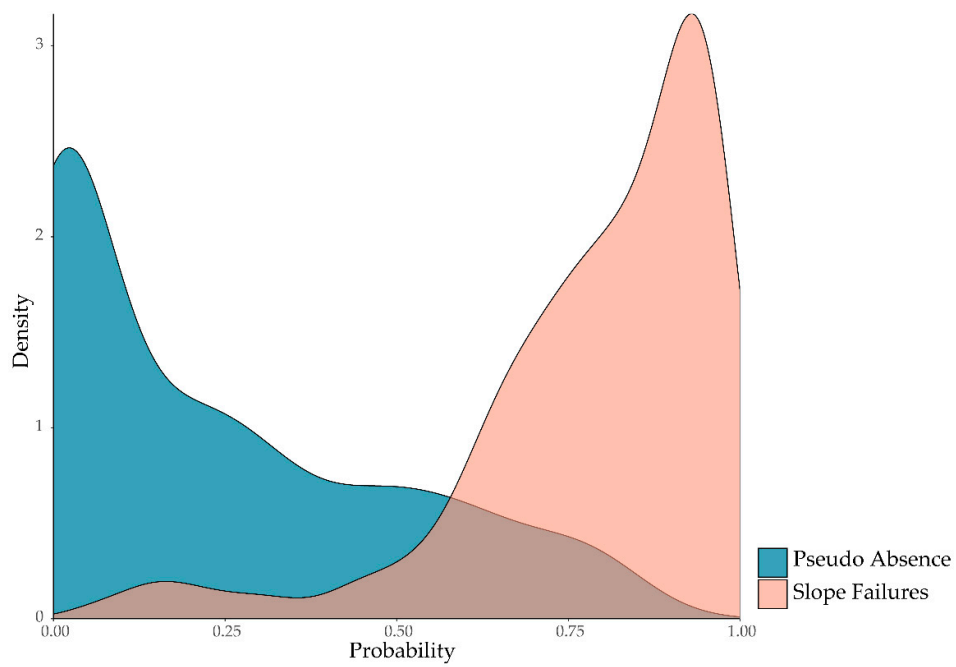
Figure 4. Kernel density plot of predicted probabilities for the validation samples.
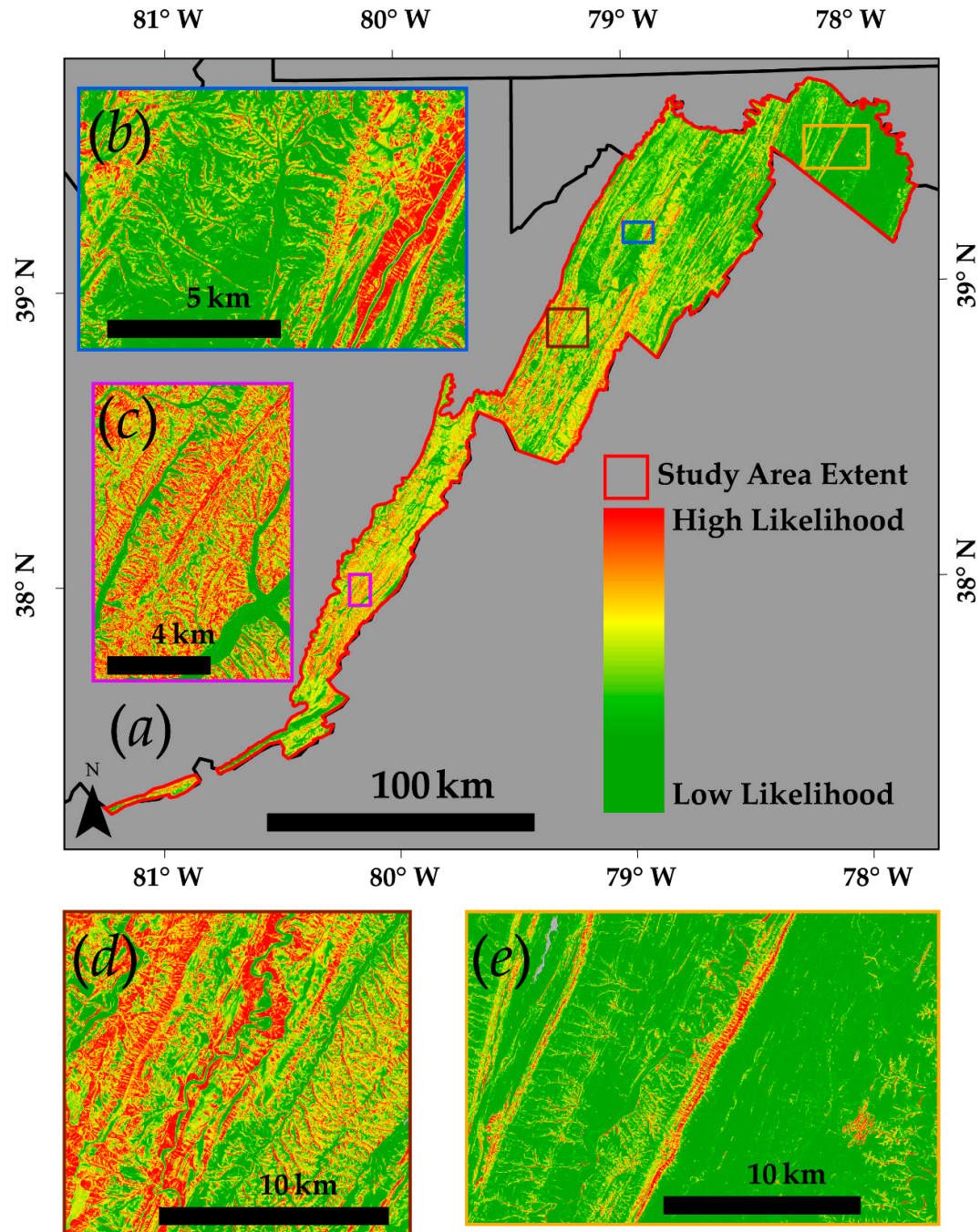
Figure 5. Resulting slope failure likelihood model created using the five combined models for (**a**) the entire mapped extent and (**b–e**) more detailed examples at larger scales.

3.2. Removing Variable Groups

Table 6 provides comparisons for models using subsets of the predictor variables while the ROC curves are visualized in Figure 6. These models were created using five combined models with different pseudo absence data, as described above. At the 95% confidence level statistical difference was noted between all experiments other than those using all the variables and just the terrain variables (*p*-value = 0.479). Further, using all variables provided only a 0.002 improvement in AUC

and a 0.003 improvement in AUC (PRC) in comparison to only using the terrain variables. The model using all variables provided the best performance while models trained with only the ancillary data provided the poorest performance, which highlights the value of including terrain variables. Further, overall accuracy was lower than 80% and the Kappa statistic was lower than 0.60 for all models that did not incorporate terrain variables. The lithologic, soil, distance and cost distance variables were not able to provide a statistically comparable performance to the results obtained using only the terrain data and combining these variables with the terrain data did not statistically improve the model performance. As previously noted by Duo et al. [25], LiDAR-derived terrain variables are valuable for slope failure predictive modeling tasks.

Table 6. Comparison for different feature spaces. Lower and upper bounds represent a 95% confidence interval calculated using 2000 stratified bootstrap replicates. X indicates statistically significantly different receiver operating characteristic (ROC) curves at a 95% confidence level.

	Soil/ Lithology	Roads/ Streams	All Except Terrain	Just Terrain	All Variables
Number of Variables	3	8	11	32	43
AUC	0.677	0.830	0.856	0.944	0.946
AUC Lower	0.656	0.807	0.834	0.927	0.930
AUC Upper	0.698	0.853	0.878	0.961	0.962
AUC (PRC)	0.661	0.791	0.838	0.946	0.949
Kappa	0.218	0.527	0.560	0.732	0.742
Overall Accuracy	60.9%	76.3%	78.0%	86.6%	87.1%
Precision	0.572	0.728	0.738	0.830	0.834
Recall	0.862	0.842	0.869	0.919	0.926
Specificity	0.356	0.685	0.691	0.812	0.815
F1 Score	0.688	0.781	0.798	0.873	0.878
Soil/Lithology		X	X	X	X
Roads/Streams			X	X	X
All Except Terrain				X	X
Just Terrain					
All Variables					
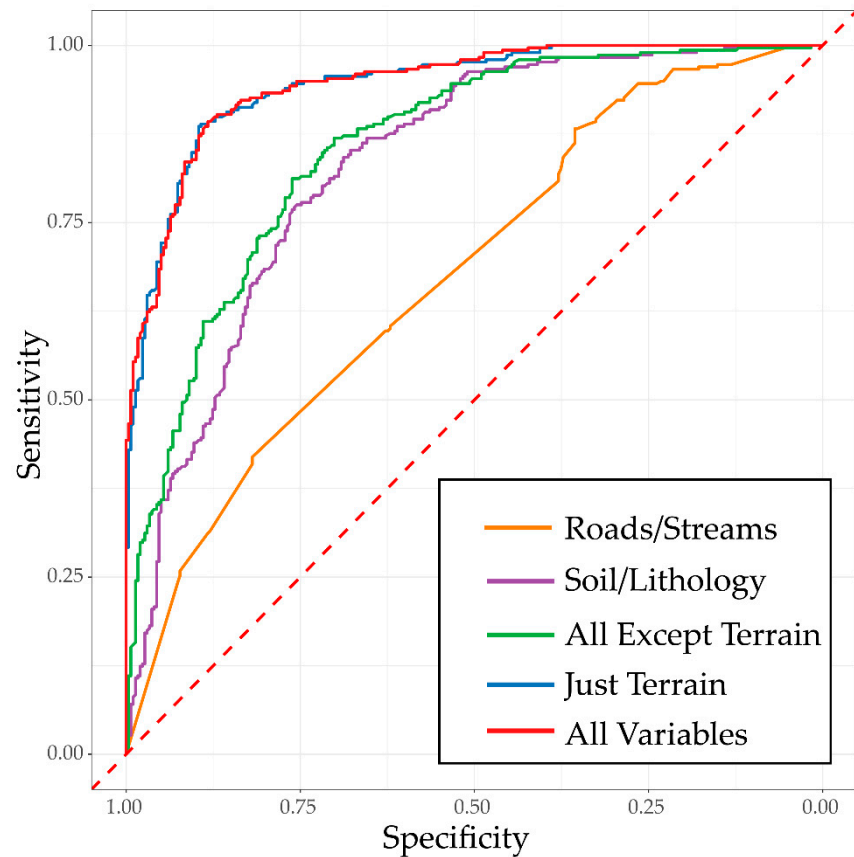
Figure 6. ROC curve comparison for different feature spaces.

3.3. Impact of Sample Size

Figures 7 and 8 summarize the impact of sample size on model performance. Note that the sample size is the number of samples for each class not the overall number of samples. In Figure 7 red stars indicate statistical difference at the 95% confidence level between the model and the model trained with 1500 samples per class. All models performed statistically significantly poorer than the model with 1500 samples other than the models trained with 500 and 1250 samples, although the *p*-value when using 500 samples was 0.084, just larger than the 0.05 threshold. Further, an increase in performance is noted up to the model with 1500 samples, though the largest changes occur between smaller sample sizes. AUC values larger than 0.900 are observed until the sample size is reduced to fewer than 75 samples per class. Figure 8 shows patterns similar to those in Figure 7; improvement in performance metrics is observed as sample size increases, with the largest improvement at lower sample sizes. This suggests that increased sample size can improve the results; however, this benefit diminishes as sample size increases. Further, this highlights the value of developing large slope failure inventories to support model generation.
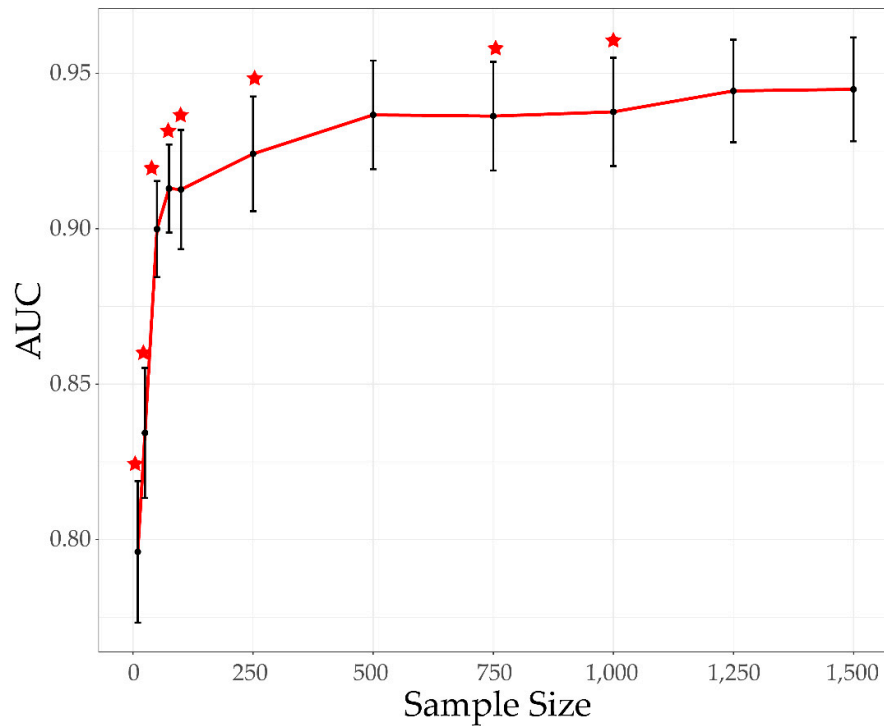
Figure 7. Area under receiver operating characteristic curve (AUC) comparison at different sample sizes. Error bars represent a 95% confidence interval calculated using 2000 stratified bootstrap replicates. Stars indicate statistical difference in ROC curves in comparison to using 1500 samples per class. The sample size represents the number of samples for each class as opposed to the overall sample size.
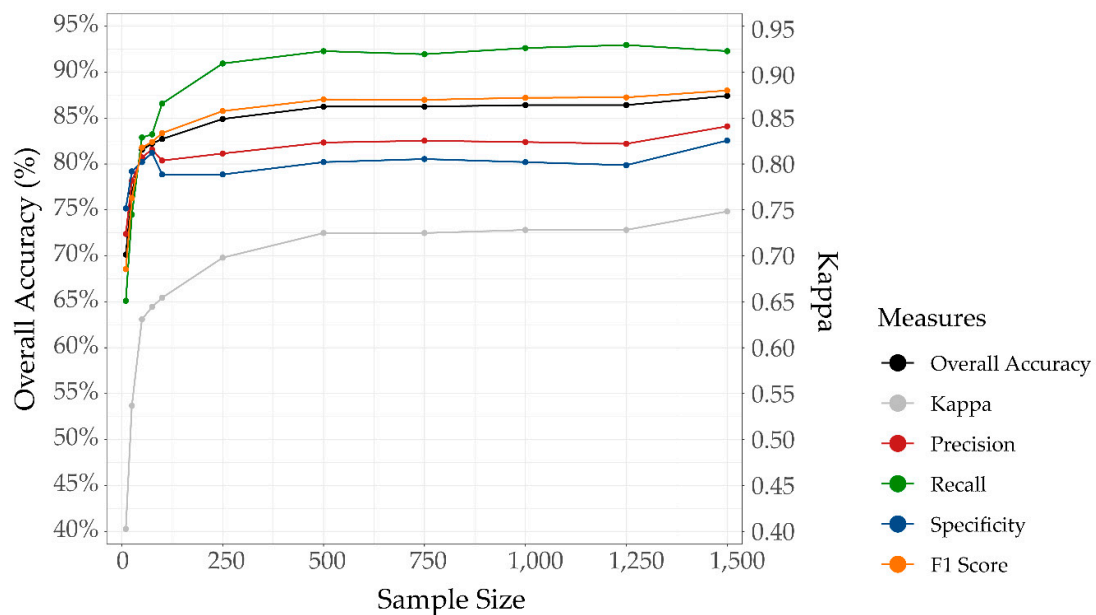
Figure 8. Model comparison at different sample sizes using multiple classification assessment metrics. The sample size represents the number of samples for each class as opposed to the overall sample size.

3.4. Feature Reduction and Feature Importance

Figures 9 and 10 show how model performance varies with feature selection. In comparison to the model using all predictor variables, statistical significance in AUC is observed when only variables in the upper 2.5 percentile of importance or less were used (*p*-value = 0.036). Model performance stabilizes once roughly the upper 10th percentile of variables is included. In contrast to the sample size results explored above, this generally suggests that the model is not negatively impacted by substantial feature reduction. Also, feature selection does not improve the modeling results, as the highest AUC is obtained when all variables are included. Additional metrics, which are shown in Figure 10, further support this observation. Similar results were noted by Maxwell et al. [115] for general land cover mapping and RF has generally been shown to be robust to complex and large feature spaces [27,62]. Practically, this suggests that feature selection may not need to be undertaken to improve the predictive performance of the model. However, feature selection could be used to reduce the number of variables that must be produced following a pilot study to assess which variables are most important. This could be particularly useful if large extents are to be mapped.
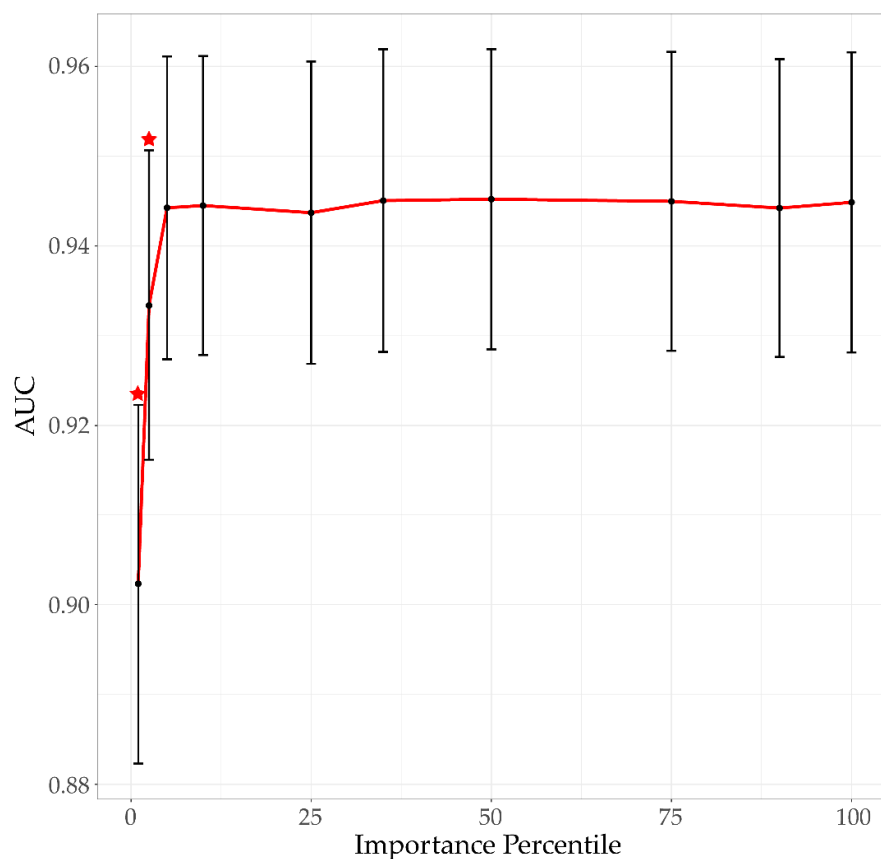
Figure 9. AUC comparison using feature selection. Error bars represent a 95% confidence interval calculated using 2000 stratified bootstrap replicates. Stars indicate statistical difference in ROC curves in comparison to using the entire feature set.
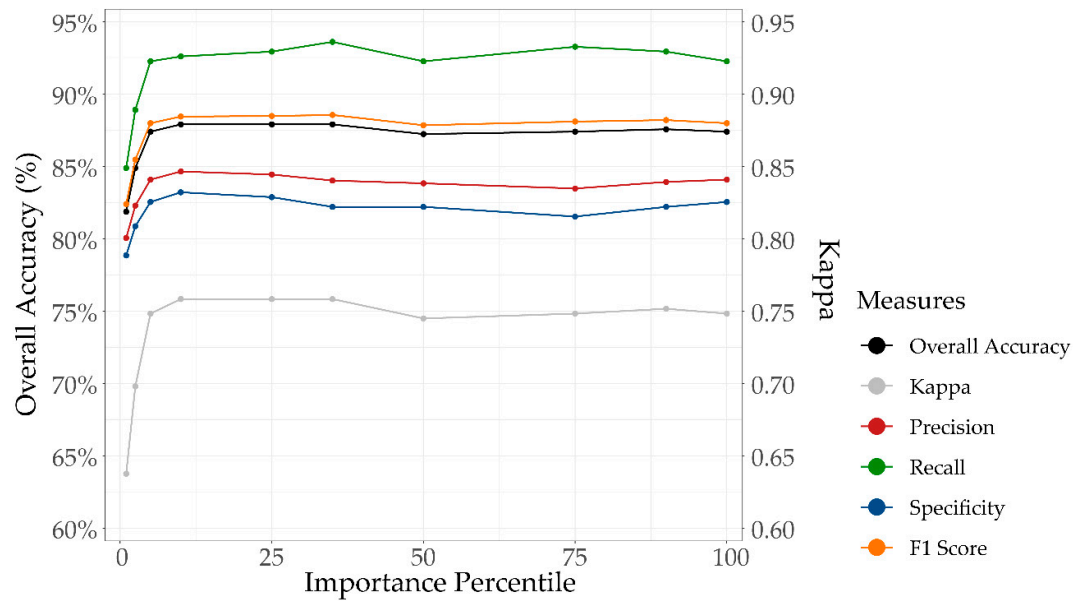
Figure 10. Model comparison when incorporating feature selection using multiple classification assessment metrics.

Figure 11 summarizes the variable importance results obtained using conditional variable importance. The most important five variables in the model are topographic slope (Slp), surface area ratio (SAR), cross-sectional curvature (CSC), surface relief ratio (SRR) and plan curvature (PlC). Specifically, the most important CSC, SRR and PlC variables are those calculated using a 7-cell radius circular window. All variables calculated using a 7-cell radius circular window are found to be more important than their counterparts calculated using an 11-cell or 21-cell radius window, suggesting the importance of characterizing local terrain conditions. Generally, terrain features show high importance in the model. Other than distance to US roads and cost distance from streams, the lithologic, soil, distance and cost distance variables are found to be of comparatively low importance. This makes sense, as adding these variables does not statistically improve the model performance in comparison to only using the terrain variables as discussed above. It should be noted that variable importance is not consistent given the large standard deviation displayed here with error bars calculated by replicating the experiment five times using different training sample subsets.
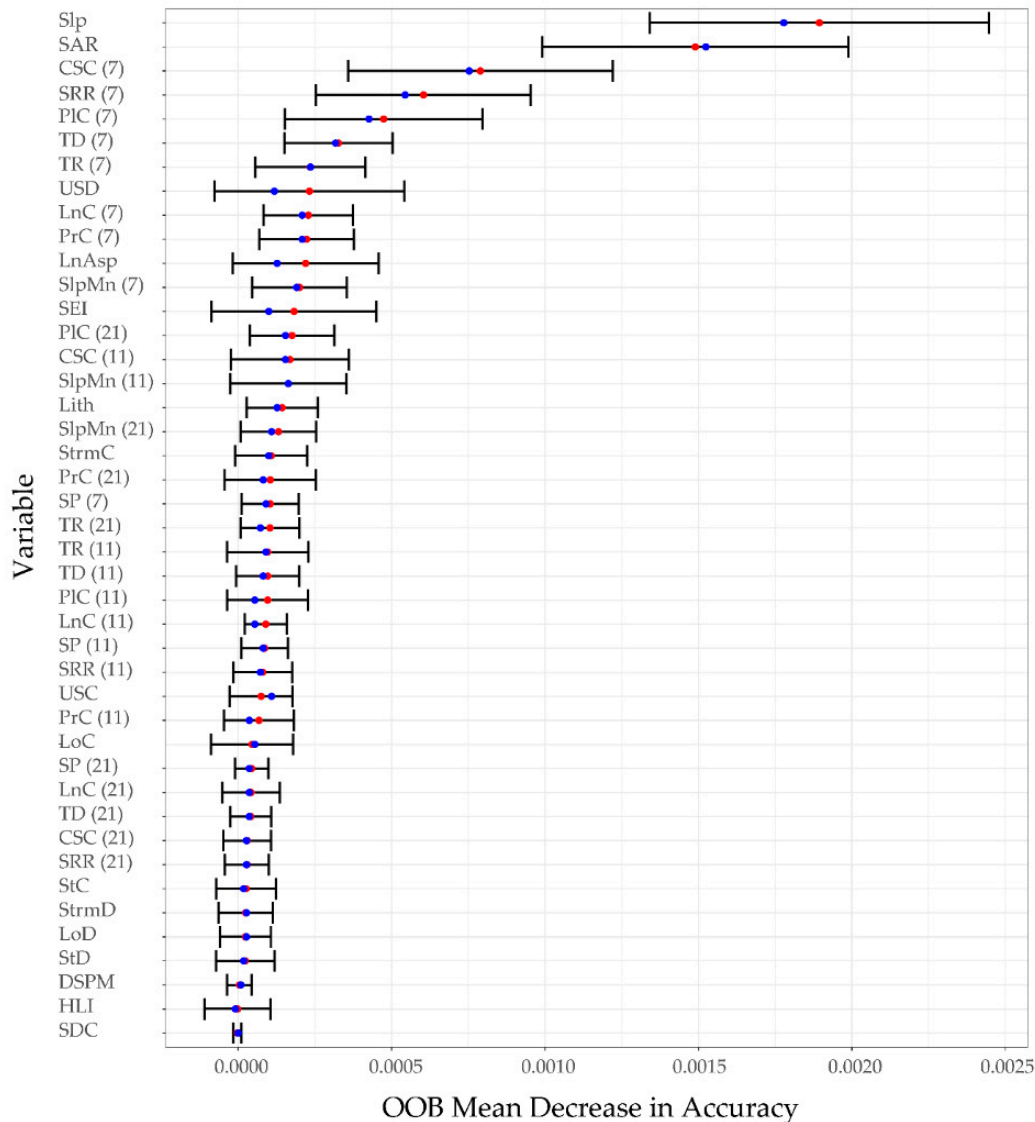
Figure 11. Variable importance calculated using conditional variable importance. Five replicates were used to obtain the standard deviation, as represented here with error bars. The red point represents the mean while the blue point shows the median.

This study confirms the importance of some variables noted as valuable in prior studies. For example, Goetz et al. [45] note the value of Slp, TR and PlC and Trigila et al. [54] document the importance of Slp, aspect and PlC. Interestingly, other studies contradict our results and the results of Goetz et al.[45] and Trigila et al. [54]. For example, Taalab et al. [21] and Pourghasemi and Kerle [68] both document low importance of PrC and PlC for landslide susceptibility mapping using RF. Some prior studies suggest the value of including non-terrain variables; for example, Trigila et al. [54] document the value of including lithology while Taalob et al. [21] highlight the value of distance from stream. As note in prior studies (for example, Goetz [45]), importance of variables may vary based on the characteristics of the study area, mapped failures and the modeling methods being used. This again highlights the value of assessing a variety of variables for predicting landslide occurrence perhaps using a pilot study. Additional studies to compare importance assessment methods and the value of variables between different study area extents is needed.

3.5. Effect of Variable Window Sizes

The results in Table 7 were generated for models using only the terrain variables. Models were produced using all the terrain variables that were not calculated using different window sizes along with the variables calculated at the window size of interest. The model that incorporated variables calculated at only a window size of 21-cells was statistically less accurate in regard to AUC than the model using the variables calculated at all window sizes (*p*-value = 0.001) while the models using only 7-cell (*p*-value = 0.337) and 11-cell (*p*-value = 0.078) windows were not statistically different from the model using all window sizes. The 7-cell window model statistically outperformed the 21-cell model (*p*-value = 0.337) but not the 11-cell model (*p*-value = 0.398), again highlighting the value of using smaller window sizes in this study. The additional metrics generally support these observations. These results generally suggest that there is value in incorporating terrain measures at multiple scales.

Table 7. Comparison of different window sizes. Lower and upper bounds represent a 95% confidence interval calculated using 2000 stratified bootstrap replicates. X indicates statistically significantly different ROC curves at a 95% confidence level.

	7	11	21	All Sizes
AUC	0.941	0.937	0.922	0.944
AUC Upper	0.924	0.920	0.904	0.927
AUC Lower	0.958	0.955	0.941	0.961
AUC (PRC)	0.942	0.940	0.922	0.947
Kappa	0.721	0.735	0.688	0.735
Overall Accuracy	86.1%	86.7%	84.4%	86.7%
Precision	0.821	0.837	0.812	0.831
Recall	0.923	0.913	0.896	0.923
Specificity	0.799	0.822	0.792	0.812
F1 Score	0.869	0.873	0.852	0.874
7			X	
11			X	
21				X
All Sizes				

4. Discussion and Recommendations

Given recent developments in the availability of landslide inventories and high spatial resolution LiDAR data over broad spatial extents, we argue that there is a need to develop methodologies for predicting the likelihood of slope failure occurrence using these data. Thus, the primary objective of this study is to provide recommendations for producing these large-area slope failure mapping products based on our findings and prior studies.

In order to alleviate the impact of training data class imbalance and to provide the algorithm with many examples of pseudo absence data, we produced five separate models then combined the results into one model, which is one benefit of using RF. However, we found that this was not necessary since the combined model did not outperform the separate models based on a variety of metrics. Thus, providing the model with one set of pseudo absence data was adequate; however, we had to produce a complete inventory of slope failures across the study area extent to minimize the chance of randomly selecting a slope failure as an absence location. If a complete mapping cannot be completed, we would suggest that a manual interpretation of the random pseudo absence points be performed in order to avoid any false negative samples.

Generally, incorporating measures of lithology, soils and proximity to roads and streams did not statistically improve the model in comparison to just using the LiDAR-derived terrain variables. This is an encouraging finding, as this may alleviate the need to produce variables from a wide variety of input data that may be of different quality and scale. For example, we used lithology data from a 1:250,000 scale geologic map in this study, which is much coarser than the available LiDAR data and was a limitation. Similar boundary uncertainties are an issue in the SSURGO soil data.

Reducing the sample size tended to decrease the model accuracy; however, AUC remained above 0.90 with only 75 samples per class. Further, the largest improvements for a variety of metrics was observed at smaller sample sizes. When adding additional samples past 75, performance metrics increased at a slower rate but improvement was documented. So, we would suggest that developing a large training dataset is of great importance for obtaining quality predictions and is worth the investment in resources. As noted above, we used a point feature at the head scarp to represent each slope failure feature in the training and validation datasets. A review of the literature suggests that there is not a consistent method used to represent slope failure features when generating likelihood or susceptibility models; some studies use points (for example, References [19,45,48,72]) while other use polygons (for example, References [15,73]). Thus, there is a need for further investigation of the impact of sample selection and feature representation methods in slope failure modeling.

In contrast to sample size, our results suggest that RF is not heavily impacted by feature selection. The best performance was obtained using all variables; however, results were not statistically different when using all variables vs. the top 10th percentile of variables. Even though variable selection may not be necessary, it may still be desirable as a means to reduce the model complexity and the need to produce a large set of variables over a large spatial extent. A pilot study over a smaller extent or multiple smaller extents could be used to determine appropriate variables.

The most important five variables in the model were topographic slope (Slp), surface area ratio (SAR), cross-sectional curvature (CSC), surface relief ratio (SRR) and plan curvature (PlC). Generally, we also document that variables calculated using a 7-cell radius moving window showed greater importance than their counterparts calculated using 11- or 21-cells, which suggests the need to measure local conditions. However, including the measures at multiple scales did improve the model, so we suggest using multiple window sizes for calculating terrain variables that rely on moving windows. More work is required to assess the impact of window size and to determine optimal scales at which to produce these variables. The optimal terrain variables may be case specific and may depend on the characteristics of the slides and the landscape. We recommend experimenting with a variety of variables, perhaps as a pilot study.

In a risk management context, these findings generally suggest that LiDAR data are of great value in mapping slope failures and producing likelihood models since they allow for the interpretation of slope failure locations for producing inventories and training data for models. Further, as highlighted in this study and prior studies, a variety of terrain variables can be generated from LiDAR that are valuable for predicting slope failure occurrence. Once these models are generated, occurrence and risk can be summarized relative to aggregating units, such as property boundaries, to generalize the model and provide valuable information to regulators and land owners.

5. Conclusions

Slope failure and landslide mapping is an important application of geospatial data due to the threats to property and life that they pose. With the development of slope failure inventories and high spatial resolution LiDAR data over large spatial extents, there is a need to develop consistent methods for mapping and predicting these features. This study specifically highlights the value of large and quality training datasets along with a characterization of the terrain using a variety of terrain variables calculated at different scales. In the United States specifically, we argue for the adoption of consistent methods to make use of landslide inventories, such as those currently being curated by the USGS and LiDAR data, such as the 3DEP products, to consistently generate products over large spatial extents.

Author Contributions: conceptualization, A.E.M., J.S.K. and M.S.; methodology, A.E.M., J.S.K., M.S. and J.A.T.; validation, M.L.B. and S.M.M..; formal analysis, A.E.M.; writing—original draft preparation, A.E.M.; writing—review and editing, A.E.M., M.L.B., K.A.D., J.S.K., S.M.M., M.S. and J.A.T.; data curation, M.L.B., K.A.D., S.M.M. and M.S.; supervision, K.AD. and M.S.; project administration, M.S. and K.A.D..; funding acquisition, M.S. and K.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this research has been provided by the Federal Emergency Management Agency's (FEMA's) Hazard Mitigation Grant Program and the WV Division of Homeland Security and Emergency Management (WVDHSEM) under project number FEMA-4273-DR-WV-0031 and WVU-10023420.1.1007860AR.

Acknowledgments: We would like to acknowledge the Federal Emergency Management Agency (FEMA) and the West Virginia Division of Homeland Security and Emergency Management (WVDHSEM) for funding the risk assessment project under the Hazard Mitigation Grant Program. We would like to acknowledge initial encouragement and support for this landslide risk assessment study for West Virginia by State Hazard Mitigation Officer Brian Penix. Ray Perry, Flood Plain Manager, is acknowledged for his help in identifying landslide locations in Logan County, West Virginia. Elizabeth Hanwell, undergraduate intern at the West Virginia GIS Technical Center, provided valuable help during various phases of this study. We would also like to think 3 anonymous reviewers whose suggestions and comments strengthened the work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. USGS Fact Sheet 2004-3072: Landslide Types and Processes. Available online: https://pubs.usgs.gov/fs/2004/3072/ (accessed on 11 November 2019).
2. Landslide Hazards. Available online: https://www.usgs.gov/natural-hazards/landslide-hazards (accessed on 11 November 2019).
3. Landslides 101. Available online: https://www.usgs.gov/natural-hazards/landslide-hazards/science/landslides-101?qt-science_center_objects=0#qt-science_center_objects (accessed on 7 November 2019).
4. Highland, L.M.; Bobrowsky, P. *The Landslide Handbook—A Guide to Understanding Landslides*; Circular; U.S. Geological Survey: Reston, VA, USA, 2008; p. 147.
5. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930.
6. Chiang, S.-H.; Chang, K.-T. The potential impact of climate change on typhoon-triggered landslides in Taiwan, 2010–2099. *Geomorphology* **2011**, *133*, 143–151.
7. Collison, A.; Wade, S.; Griffiths, J.; Dehn, M. Modelling the impact of predicted climate change on landslide frequency and magnitude in SE England. *Eng. Geol.* **2000**, *55*, 205–218.
8. Crozier, M.J. Deciphering the effect of climate change on landslide activity: A review. *Geomorphology* **2010**, *124*, 260–267.
9. Dixon, N.; Brook, E. Impact of predicted climate change on landslide reactivation: Case study of Mam Tor, UK. *Landslides* **2007**, *4*, 137–147.
10. Jakob, M.; Lambert, S. Climate change effects on landslides along the southwest coast of British Columbia. *Geomorphology* **2009**, *107*, 275–284.
11. Van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial data for landslide susceptibility, hazard and vulnerability assessment: An overview. *Eng. Geol.* **2008**, *102*, 112–131.
12. Carrara, A.; Cardinali, M.; Detti, R.; Guzzetti, F.; Pasqui, V.; Reichenbach, P. GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf. Process. Landf.* **1991**, *16*, 427–445.
13. Carrara, A.; Sorriso-Valvo, M.; Reali, C. Analysis of landslide form and incidence by statistical techniques, Southern Italy. *CATENA* **1982**, *9*, 35–62.
14. Carrara, A. Multivariate models for landslide hazard evaluation. *Math. Geol.* **1983**, *15*, 403–426.
15. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831.
16. Chen, W.; Pourghasemi, H.R.; Kornejady, A.; Zhang, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt and SVM machine learning techniques. *Geoderma* **2017**, *305*, 314–327.
17. Hong, H.; Liu, J.; Bui, D.T.; Pradhan, B.; Acharya, T.D.; Pham, B.T.; Zhu, A.-X.; Chen, W.; Ahmad, B.B. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *CATENA* **2018**, *163*, 399–413.
18. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *CATENA* **2018**, *165*, 520–529.
19. Kim, J.-C.; Lee, S.; Jung, H.-S.; Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* **2018**, *33*, 1000–1015.
20. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženílek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234.

21. Taalab, K.; Cheng, T.; Zhang, Y. Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* **2018**, *2*, 159–178.

22. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree and Naïve Bayes Models. *Math. Probl. Eng.* **2012**, *2012*, 974638.

23. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582.

24. Yilmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138.

25. Dou, J.; Yunus, A.P.; Tien Bui, D.; Sahana, M.; Chen, C.-W.; Zhu, Z.; Wang, W.; Thai Pham, B. Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. *Remote Sens.* **2019**, *11*, 638.

26. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sens.* **2015**, *7*, 16398–16421.

27. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817.

28. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10.

29. Sameen, M.I.; Pradhan, B.; Lee, S. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *CATENA* **2020**, *186*, 104249.

30. Wang, Y.; Wang, X.; Jian, J. Remote Sensing Landslide Recognition Based on Convolutional Neural Network. Available online: https://www.hindawi.com/journals/mpe/2019/8389368/ (accessed on 24 January 2020).

31. Höfle, B.; Rutzinger, M. Topographic airborne LiDAR in geomorphology: A technological perspective. *Z. Geomorpholog. Suppl. Issues* **2011**, *55*, 1–29.

32. Jaboyedoff, M.; Oppikofer, T.; Abellán, A.; Derron, M.-H.; Loye, A.; Metzger, R.; Pedrazzini, A. Use of LIDAR in landslide investigations: A review. *Nat. Hazards* **2012**, *61*, 5–28.

33. Passalacqua, P.; Belmont, P.; Staley, D.M.; Simley, J.D.; Arrowsmith, J.R.; Bode, C.A.; Crosby, C.; DeLong, S.B.; Glenn, N.F.; Kelly, S.A.; et al. Analyzing high resolution topography for advancing the understanding of mass and energy transfer through landscapes: A review. *Earth-Sci. Rev.* **2015**, *148*, 174–193.

34. Migoń, P.; Kasprzak, M.; Traczyk, A. How high-resolution DEM based on airborne LiDAR helped to reinterpret landforms: Examples from the Sudetes, SW Poland. *Landf. Anal.* **2013**, *22*, doi:10.12657/landfana.022.007.

35. Stoker, J.M.; Abdullah, Q.A.; Nayegandhi, A.; Winehouse, J. Evaluation of Single Photon and Geiger Mode Lidar for the 3D Elevation Program. *Remote Sens.* **2016**, *8*, 767.

36. Arundel, S.T.; Phillips, L.A.; Lowe, A.J.; Bobinmyer, J.; Mantey, K.S.; Dunn, C.A.; Constance, E.W.; Usery, E.L. Preparing *The National Map* for the 3D Elevation Program—Products, process and research. *Cartogr. Geogr. Inf. Sci.* **2015**, *42*, 40–53.

37. Kirschbaum, D.B.; Adler, R.; Hong, Y.; Hill, S.; Lerner-Lam, A. A global landslide catalog for hazard applications: Method, results and limitations. *Nat. Hazards* **2010**, *52*, 561–575.

38. Cruden, D.M. A simple definition of a landslide. *Bull. Int. Assoc. Eng. Geol.* **1991**, *43*, 27–29.

39. Nichols, J.; Wong, M.S. Satellite remote sensing for detailed landslide inventories using change detection and image fusion. *Int. J. Remote Sens.* **2005**, *26*, 1913–1926.

40. Lee, S.; Choi, J.; Min, K. Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. *Int. J. Remote Sens.* **2004**, *25*, 2037–2052.

41. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31.

42. Ballabio, C.; Sterlacchini, S. Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy. *Math. Geosci* **2012**, *44*, 47–70.

43. Casagli, N.; Cigna, F.; Bianchini, S.; Hölbling, D.; Füreder, P.; Righini, G.; Del Conte, S.; Friedl, B.; Schneiderbauer, S.; Iasio, C.; et al. Landslide mapping and monitoring by using radar and optical remote sensing: Examples from the EC-FP7 project SAFER. *Remote Sens. Appl. Soc. Environ.* **2016**, *4*, 92–108.

44. Colkesen, I.; Sahin, E.K.; Kavzoglu, T. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J. Afr. Earth Sci.* **2016**, *118*, 53–64.

45. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11.

46. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491.

47. Lu, P.; Stumpf, A.; Kerle, N.; Casagli, N. Object-Oriented Change Detection for Landslide Rapid Mapping. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 701–705.

48. Lee, S.; Ryu, J.-H.; Won, J.-S.; Park, H.-J. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Eng. Geol.* **2004**, *71*, 289–302.

49. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365.

50. Sarkar, S.; Kanungo, D.P. An Integrated Approach for Landslide Susceptibility Mapping Using Remote Sensing and GIS. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 617–625.

51. Stumpf, A.; Kerle, N. Object-oriented mapping of landslides using Random Forests. *Remote Sens. Environ.* **2011**, *115*, 2564–2577.

52. Stumpf, A.; Kerle, N. Combining Random Forests and object-oriented analysis for landslide mapping from very high resolution imagery. *Procedia Environ. Sci.* **2011**, *3*, 123–129.

53. Liu, Y.; Wu, L. Geological Disaster Recognition on Optical Remote Sensing Images Using Deep Learning. *Procedia Comput. Sci.* **2016**, *91*, 566–575.

54. Trigila, A.; Iadanza, C.; Esposito, C.; Scarascia-Mugnozza, G. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology* **2015**, *249*, 119–136.

55. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2016**, *13*, 839–856.

56. Ghorbanzadeh, O.; Meena, S.R.; Blaschke, T.; Aryal, J. UAV-Based Slope Failure Detection Using Deep-Learning Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 2046.

57. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sens.* **2019**, *11*, 196.

58. Jin, K.P.; Yao, L.K.; Cheng, Q.G.; Xing, A.G. Seismic landslides hazard zoning based on the modified Newmark model: A case study from the Lushan earthquake, China. *Nat. Hazards* **2019**, *99*, 493–509.

59. Lei, T.; Zhang, Q.; Xue, D.; Chen, T.; Meng, H.; Nandi, A.K. End-to-end Change Detection Using a Symmetric Fully Convolutional Network for Landslide Mapping. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3027–3031.

60. Lei, T.; Zhang, Y.; Lv, Z.; Li, S.; Liu, S.; Nandi, A.K. Landslide Inventory Mapping From Bitemporal Images Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 982–986.

61. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91.

62. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.

63. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565.

64. Ghimire, B.; Rogan, J.; Galiano, V.R.; Panday, P.; Neeti, N. An Evaluation of Bagging, Boosting and Random Forests for Land-Cover Classification in Cape Cod, Massachusetts, USA. *GISci. Remote Sens.* **2012**, *49*, 623–643.

65. Evans, J.S.; Cushman, S.A. Gradient modeling of conifer species using random forests. *Landsc. Ecol.* **2009**, *24*, 673–683.

66. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forest classification of multisource remote sensing and geographic data. In Proceedings of the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; Volume 2, pp. 1049–1052.

67. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300.

68. Pourghasemi, H.R.; Kerle, N. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environ. Earth Sci.* **2016**, *75*, 185.

69. Maxwell, A.E.; Warner, T.A.; Strager, M.P. Predicting Palustrine Wetland Probability Using Random Forest Machine Learning and Digital Elevation Data-Derived Terrain Variables. Available online: https://www.ingentaconnect.com/content/asprs/pers/2016/00000082/00000006/art00016 (accessed on 12 November 2019).

70. Strager, M.P.; Strager, J.M.; Evans, J.S.; Dunscomb, J.K.; Kreps, B.J.; Maxwell, A.E. Combining a Spatial Model and Demand Forecasts to Map Future Surface Coal Mining in Appalachia. *PLoS ONE* **2015**, *10*, e0128813.

71. Lillesand, T.; Kiefer, R.W.; Chipman, J. *Remote Sensing and Image Interpretation*, 7th ed.; Wiley: Hoboken, NJ, USA, 2015. Available online: https://www.wiley.com/en-us/Remote+Sensing+and+Image+Interpretation%2C+7th+Edition-p-9781118343289 (accessed on 28 October 2019).

72. Goetz, J.N.; Guthrie, R.H.; Brenning, A. Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology* **2011**, *129*, 376–386.

73. Mahalingam, R.; Olsen, M.J.; O'Banion, M.S. Evaluation of landslide susceptibility mapping techniques using lidar-derived conditioning factors (Oregon case study). *Geomat. Nat. Hazards Risk* **2016**, *7*, 1884–1907.

74. Huisman, O. Principles of Geographic Information Systems—An Introductory Textbook. Available online: https://webapps.itc.utwente.nl/librarywww/papers_2009/general/principlesgis.pdf (accessed on 28 January 2020).

75. Gessler, P.E.; Moore, I.D.; McKenzie, N.J.; Ryan, P.J. Soil-landscape modelling and spatial prediction of soil attributes. *Int. J. Geogr. Inf. Syst.* **1995**, *9*, 421–432.

76. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30.

77. Zevenbergen, L.W.; Thorne, C.R. Quantitative analysis of land surface topography. *Earth Surf. Process. Landf.* **1987**, *12*, 47–56.

78. Maxwell, A.E.; Warner, T.A. Is high spatial resolution DEM data necessary for mapping palustrine wetlands? *Int. J. Remote Sens.* **2019**, *40*, 118–137.

79. Nusser, S.M.; Goebel, J.J. The National Resources Inventory: A long-term multi-resource monitoring programme. *Environ. Ecol. Stat.* **1997**, *4*, 181–204.

80. Landslide Susceptibility Pilot Study_BerkleyCounty_20160408.pdf.

81. Cardwell, D.H.; Erwin, R.B.; Woodward, H.P. *Geologic map of West Virginia*; West Virginia Geological and Economic Survey: Morgantown, WV, USA, 1968.

82. Strausbaugh, P.D.; Core, E.L. *Flora of West Virginia*; West Virginia University Bulletin: Morgantown, WV, USA, 1952.

83. WVGES: WV Physiographic Provinces. Available online: https://www.wvgs.wvnet.edu/www/maps/pprovinces.htm (accessed on 14 November 2019).

84. Chang, K.-T. Geographic Information System. In *International Encyclopedia of Geography*; American Cancer Society: Atlanta, GA, USA, 2017; pp. 1–9, ISBN 978-1-118-78635-2.

85. Reed, M. *How Will Anthropogenic Valley Fills in Appalachian Headwaters Erode?* MS, West Virginia University Libraries: Morgantown, WV, USA, 2018.

86. *ArcGIS*, version Pro 2.2; software for technical computation; ESRI: Redlands, CA, USA, 2018.

87. ArcGIS Gradient Metrics Toolbox. Available online: https://evansmurphy.wixsite.com/evansspatial/arcgis-gradient-metrics-toolbox (accessed on 14 November 2019).

88. SAGA—System for Automated Geoscientific Analyses. Available online: http://www.saga-gis.org/en/index.html (accessed on 14 November 2019).

89. Module Morphometric Features/SAGA-GIS Module Library Documentation (v2.2.5). Available online: http://www.saga-gis.org/saga_tool_doc/2.2.5/ta_morphometry_23.html (accessed on 14 November 2019).

90. Wood, J. Chapter 14 Geomorphometry in LandSerf. In *Developments in Soil Science*; Hengl, T., Reuter, H.I., Eds.; Geomorphometry; Elsevier: Amsterdam, The Netherlands, 2009; Volume 33, pp. 333–349.

91. Wood, J. The geomorphological characterisation of Digital Elevation Models. Ph.D. Thesis, University of Leicester, Leicester, UK, 1996.

92. Albani, M.; Klinkenberg, B.; Andison, D.W.; Kimmins, J.P. The choice of window size in approximating topographic surfaces from Digital Elevation Models. *Int. J. Geogr. Inf. Sci.* **2004**, *18*, 577–593.

93. Hengl, T.; Gruber, S.; Shrestha, D.P. Reduction of errors in digital terrain parameters used in soil-landscape modelling. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 97–112.

94. Franklin, S.E. Geomorphometric processing of digital elevation models. *Comput. Geosci.* **1987**, *13*, 603–609.

95. Wilson, J.P.; Gallant, J.C. *Terrain Analysis: Principles and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2000; ISBN 978-0-471-32188-0.

96. Moreno, M.; Levachkine, S.; Torres, M.; Quintero, R. Geomorphometric Analysis of Raster Image Data to detect Terrain Ruggedness and Drainage Density. In *Progress in Pattern Recognition, Speech and Image Analysis*; Sanfeliu, A., Ruiz-Shulcloper, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 643–650.

97. Evans, I.S.; Minár, J. A classification of geomorphometric variables. *Int. Geom-orphometry* **2011**, *4*, 105–108.

98. Jenness, J.S. Calculating Landscape Surface Area from Digital Elevation Models. *Wildl. Soc. Bull.* **2004**, *32*, 829–839.

99. Pike, R.J.; Wilson, S.E. Elevation-Relief Ratio, Hypsometric Integral and Geomorphic Area-Altitude Analysis. *GSA Bull.* **1971**, *82*, 1079–1084.

100. Balice, R.G.; Miller, J.D.; Oswald, B.P.; Edminster, C.; Yool, S.R. *Forest Surveys and Wildfire Assessment in the Los Alamos Region, 1998–1999*; Los Alamos National Lab: New Mexico, NM, USA, 2000.

101. McCune, B.; Keon, D. Equations for potential annual direct incident radiation and heat load. *J. Veg. Sci.* **2002**, *13*, 603–606.

102. SSURGO|NRCS Soils. Available online: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/office/ssr12/tr/?cid=nrcs142p2_010596 (accessed on 14 November 2019).

103. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 6.

104. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.

105. Strobl, C.; Hothorn, T.; Zeileis, A. Party on! A new, conditional variable importance measure available in the party package. *R J.* **2009**, *1*, 14–17.

106. Conditional Variable Importance for Random Forests|BMC Bioinformatics|Full Text. Available online: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-307 (accessed on 15 November 2019).

107. Spearman, C. The proof and measurement of association between two things. *Int. J. Epidemiol.* **2010**, *39*, 1137–1150.

108. Beck, J.R.; Shultz, E.K. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch. Pathol. Lab. Med.* **1986**, *110*, 13–20.

109. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77.

110. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. | Radiology. Available online: https://pubs.rsna.org/doi/abs/10.1148/radiology.148.3.6878708 (accessed on 15 November 2019).

111. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* **1988**, *44*, 837–845.

112. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432.

113. Grau, J.; Grosse, I.; Keilwagen, J. PRROC: Computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics* **2015**, *31*, 2595–2597.

114. Keilwagen, J.; Grosse, I.; Grau, J. Area under Precision-Recall Curves for Weighted and Unweighted Data. *PLoS ONE* **2014**, *9*, e92209.

115. Maxwell, A.E.; Strager, M.P.; Warner, T.A.; Ramezan, C.A.; Morgan, A.N.; Pauley, C.E. Large-Area, High Spatial Resolution Land Cover Mapping Using Random Forests, GEOBIA and NAIP Orthophotography: Findings and Recommendations. *Remote Sens.* **2019**, *11*, 1409.