A Spatial Ensemble Model for Rockfall Source Identification from High Resolution LiDAR Data and GIS

Ali Mutar Fanos[1], Biswajeet Pradhan [2,3]

[1]Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia, 43400, UPM, Serdang, Malaysia; Email. engalim87@gmail.com

[2]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, NSW 2007, Australia; Biswajeet.Pradhan@uts.edu.au

[3] Department of Energy and Mineral Resources Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea; biswajeet24@gmail.com

Corresponding author: Biswajeet Pradhan (e-mail: Biswajet.Pradhan@uts.edu.au; cwlee@kangwon.ac.kr).

ABSTRACT Rockfall source identification is the most challenging task in rockfall hazard and risk assessment. This difficulty rises in the areas where there is a presence of other types of landslide such as shallow landslide and debris flow. The aim of this research is to develop and test a hybrid model that can accurately identify the source areas. High-resolution light detection and ranging data (LiDAR) were employed to derive Digital Terrain Model (DTM) from which several conditioning factors were extracted. These conditioning factors were optimized utilizing Ant Colony Optimization (ACO). Different machine learning algorithms namely logistic regression (LR), random tree (RT), random forest (RF), support vector machine (SVM), and artificial neural network (ANN), in addition to their ensemble models (stacking, bagging, and voting), were examined. This based on the selected best subset of conditioning factors and the inventory dataset. Stacking LR-RT (the best fit model) was then utilized to produce the probabilities of different landslide types. On the other hand, Gaussian Mixture Model (GMM) was optimized and applied for automatically identifying the slope threshold of the occurrence of the different landslide types. In order to reduce the model sensitivity to the alteration in various conditioning factors and to improve the model computations performance, land use probability area was formed. The rockfall sources were identified by integrating the probability maps and the reclassified slope raster based on GMM results. The accuracy assessment reveals that the developed hybrid model can identify the probable rockfall regions with an accuracy of 0.95 based on validation dataset and 0.94 on training dataset. The slope thresholds calculated by GMM were found > 58°, 22° - 58°, and 9° - 22° for rockfall, shallow landslide, and debris flow, respectively. This indicates that the model can be generalized and replicated in different regions and the proposed method can be applied in various landslides studies.

INDEX TERMS Remote sensing; machine learning; rockfall; hybrid model; GIS; LiDAR.

I. INTRODUCTION

Rockfall is a natural phenomenon encompassing a downslope movement of a detached boulder or groups of boulders including free falling/flying, bouncing, sliding, and rolling [1]. Rockfall can result in serious damage to infrastructures and structures, even with small rock magnitude, consequently representing a pertinent risk for people and goods. Rockfall can be caused by rainfall, weathering, jointing, man-made or the combination of them [2].

Rockfall source identification is of essential significance for generating rockfall hazard maps [3]. Numerous approaches have been developed for the rockfall sources identification at local and regional scales [4], and to assess the resulting susceptibility [5], or hazard [6]. The availability of high-resolution Digital Terrain Models (DTM) allows the topography analysis at a regional scale with high levels of details. Nevertheless, the identification of probable rockfall sources at a regional scale is one of the major challenges in the assessment of rockfall hazard [7]. Rockfall source regions, that correspond to the detachment regions of rocks, are normally taken from evidences such as scree deposits at foot-slope and talus slope [8], historical and field inventory of fallen rocks. Therefore, a DTM-based geomorphometric methods become more appropriate for identification of probable rockfall sources [6]. The potential rockfall sources are identified using the slope angle distribution obtained from high-resolution DTM crossed with other information obtained from topographic and geological maps as GIS layers. The

distribution of slope angle can be decomposed in many Gaussian distributions that can be considered as morphological units characteristics [9].

The steep slope angle is one of the major elements necessary for rockfall initiation [10]. The processes of gravity-driven surfaces in mountainous areas are exceedingly associated with the topography steepness and the relief morphology, and thus reflects these instabilities [11]. The stability slope angles rely on the type of rock and the mechanical and geometrical properties of the discontinuities set [12]. Consequently, it can be considered that the distribution of slope angle reflects the relief type, such as plains, alluvial, glacial, etc., and the rocks mechanical properties. Therefore, various rock types and morphologies result in a range of slope angle values that are characteristic for a particular morphotectonic setting and expressed in the distribution of slope angle [13].

In the past decades landslide probability, hazard, and risk have been widely analyzed and explored. Nevertheless, the selection of optimized conditioning factors in such application remains a difficult mission [4]. The mapping of landslide probability employs topological, hydrological, geological, and environmental factors. Some researchers presume that the precision and accuracy of the produced probability maps increase by increasing the number of conditioning factors. In contrast, other researchers have proved that a small number of conditioning factors are adequate to generate landslide probability map with a reasonable accuracy [14].

Depending on the aim, there are numerous categories of data mining tasks. Classification is one of those categories. The main aim of classification is to gain knowledge of hidden patterns to make prediction about the class of some unknown data. For data classification, most of the standard machine learning algorithms can be utilized effectively for classification precision if class labels are equally scattered. However, such algorithms show less or poor learning performance in case of classifying the imbalanced dataset that have variation in the class labels [15]. On the other hand, some new methods were proposed for classification issues such as robust manifold matrix factorization [16] and deep learning [17]. In order to get the accuracy of classification algorithms, one or more algorithms can be combined and can get the reasonable accuracy. The process of combining the multiple algorithms is called ensembling [18]. The most popular ensemble models are voting, bagging, and stacking [19].

Despite the numerous attempts that have been conducted for rockfall source identification using photogrammetry or laser scanning data in local scale [20-22], and the few studies in regional scale [4], There is a main problem which is still not considered. This problem is when the study area contains other types of landslides that have nearly comparable geomorphometric characteristics such as shallow landslide, rockfall, and debris flow. Nevertheless, Fanos et al. [23] proposed a model for rockfall source identification using a single machine learning algorithm. However, their study did not take into account the optimization of the algorithm hyperparameters that can significantly influence the derived results. Moreover, they used a limited conditioning factors without optimization or examining the multicollinearity among these factors. In addition, the GMM was run without optimization and based on the inventory dataset not on the geomorphological units of the slope. Therefore, this paper aims to develop and test an ensemble model using high-resolution LiDAR data to solve such problem. This model is based on the combination of machine learning algorithms and Gaussian Mixture Model (GMM). The former is to produce the probability maps of different landslide types (rockfall, shallow landslide, and debris flow), and the latter is to determine the slope thresholds of different landslide types. Kinta Valley, Malaysia was selected to perform the proposed ensemble model.

II. STUDY AREA

Rockfall incidents are frequently occurred in Malaysia due to the high and steep terrain in addition to high density of rainfall. Kinta Valley which is situated in the eastern part of Malaysia (Fig. 1), is one of the most hardly hit districts in Malaysia which contributes to the economic and tourism sectors. Approximately, the Kinta district has an area of 1,950 km^2. Geographically, the area is located approximately between the northeast corner (101^0 8' 33", 4^0 39' 04") and the southwest corner (101^0 3' 32", 4^0 31' 31").

The geological setting of Kinta Valley and adjacent regions are limestone bedrock, granitic hills, and mine. Consequently, Kinta Valley and its surrounding areas have encountered several engineering geologic problems including rockfall, landslide, and debris flow. The limestone bedrocks rise over the alluvial plain in every region forming limestone hills with very steep slopes (sub- vertical to vertical). The study area consists of diverse lithology with a high coverage of igneous rocks. Such features are existing in the high-altitude areas on the east and west sides of Kinta Valley. The sedimentary (limestone) and metamorphic rocks (marble) and are widely present in Kinta Valley [24].

Kinta Valley receives high precipitation along the year with average annual rainfall of 321 mm. In addition, it experiences tropical climate with temperature ranging between 23 ^{0}C and 35 ^{0}C. The humidity in the study area is relatively high ~81.5% (Meteorological Service Department of Malaysia).

FIGURE 1. Study area.

III. USED DATA

This section demonstrates the dataset that were used to implement the proposed ensemble model. LiDAR data is the main dataset that were used to produce a high-resolution DTM and main conditioning factors. The LiDAR data were gathered in 2015 with a flight height of 1000 m and average point density of 8 pts./m² using an airborne laser scanning system (ALS). In addition, different GIS layers were used as conditioning factors. Inventory data were also utilized to train and validate the models.

A. ROCKFALL CONDITIONING FACTORS

Landslides and rockfalls are controlled by various conditioning factors and cannot be explained using individual conditioning factors [4]. Therefore, this study employs many conditioning factors for identifying the rockfall sources within Kinta Valley. These factors encompassed hydrological, morphological, geological, anthropogenic, and environmental factors (Fig. 2), which are commonly mentioned in the literature. Various conditioning factors were obtained from LiDAR dataset, satellite images, and government agencies data archives.

The morphological factors (elevation, aspect, slope, and curvature) were extracted from LiDAR and GIS spatial analysis tools. Furthermore, four hydrological factors included Stream Power Index (SPI), Sediment Transport Index (STI), Topographic Roughness Index (TRI) and Topographic Wetness Index (TWI). Other factors such as flow length, distance to stream, distances to roads and lineaments were included. Geological factors such as lineaments are main triggering for landslides. The anthropogenic factors included land use. The land use map was produced by classifying a high-resolution aerial photo (0.1 m) with supervised SVM method with an accuracy of (91%). Field surveys were conducted to verify the land use map. A total number of twelve classes were specified as shown in Figure 2. The lithology of Kinta Valley is predominantly limestone / marble. Other types include sandstone and granite are also presence in the study area.

Finally, areas with low vegetation density are more exposed to erosion and increase instability than highly vegetated regions. The low vegetation coverage can simply result in landslides formation. In the current research, vegetation density was utilized as a conditioning factor for mapping landslide and rockfalls probabilities.

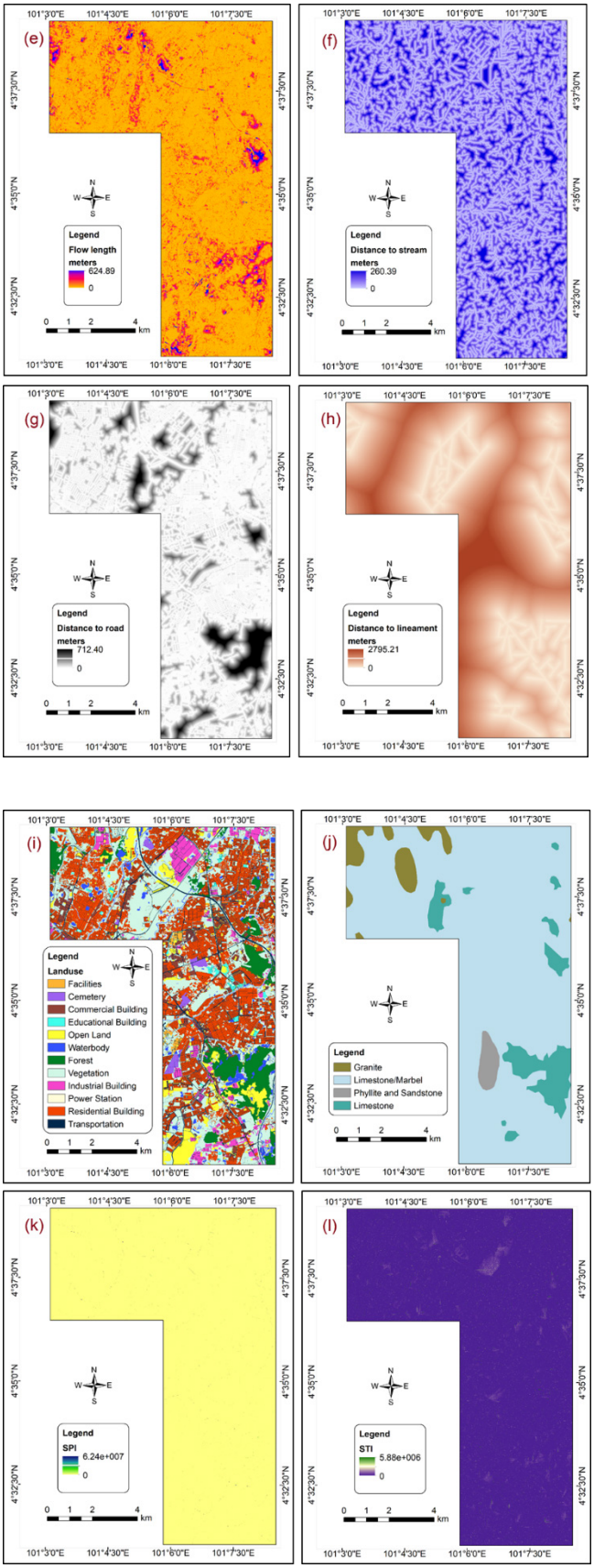


FIGURE 2. Landslide conditioning factors.

B. INVENTORY DATASET

The landslide inventory was obtained from different sources involving historical records, field surveys, and remote sensing. SPOT fused images and aerial photo (0.1 m) were utilized for the visual identification of landslides in the selected area. Field surveys were performed to map the landslides can be occurred beneath vegetation or in regions that cannot be detected from the satellite images. Multiple in-situ campaigns were performed utilizing a GNSS device. A total number of 179 samples with their related characteristics were prepared for assessment (Figure 1). The inventory dataset include three types of landslide namely rockfall, shallow landslides, and debris flow. The inventory was split into two subsets (training (70%) and testing (30%)) ensuring each group has all the landslides types [25].

IV. METHODOLOGY

This section presents an overview of the proposed hybrid model to identify the potential rockfall sources using high-resolution LiDAR dataset. The model is based on an ensemble stacking LR-RT and GMM methods. The details of this model and its implementation and validation methods are given in the following sub-sections. The ensemble models were implemented through Python, whereas the GMM and ACO were implemented using MATLAB (2017).

A. OVERALL METHODOLOGY

The developing steps of the proposed model are illustrated in Figure 3, which include four major steps. In the first stage, field and remote sensing dataset were obtained. ALS dataset, landslides inventory, and a GIS dataset that contains topographical, geological, rainfall, and vegetation were prepared. The acquired datasets were pre-processed in the second step. In addition, the conditioning factors were optimized using ACO. The ACO was run with different subsets of the conditioning factors to find the best subset for each landslides types. Three various machine learning algorithms (RF, SVM, and LR) were employed to validate and compare the obtained results. The third stage was the core-processing module in the proposed method, which consisted of developing the stacking LR - RT and GMM models. The stacking LR - RT model was developed to produce a probability of landslide, rockfall, and debris flow occurrences in the area using the inventory dataset and the best subset of the conditioning factors. The stacking LR - RT model was employed after performing an elaborate comparative study with other individual ensemble models such as LR, RT, RF, SVM, and ANN and their ensembles (bagging, voting, and stacking). The selection of the base models hyperparameters is significantly affect the performance a model. In the current study, grid search method was used to select these parameters. The best model was selected based on the standard accuracy metrics such as overall accuracy, Receiver Operating Characteristics (ROC) curves, and Prediction Rate Curve (PRC). Then the best fit

ensemble model (stacking LR – RT) was trained based on the inventory dataset with the best subset of the conditioning factors and thus the landslide probability was generated for each type. On the other hand, The GMM is developed to determine the distributions of slope angle and deriving the geomorphological units in the selected area. The GMM was first optimized to find the best number of Gaussian components. Consequently, GMM was run using the slope dataset of the selected study areas to determine the slope thresholds of each landslides types automatically. The final step composed of mapping, validations, and model comparisons with other models. Consequently, the produced landslides probabilities were integrated with the reclassified slope based on the obtained thresholds from GMM and then land-use probability area to detect the potential landslide, debris flow, and rockfall source areas.

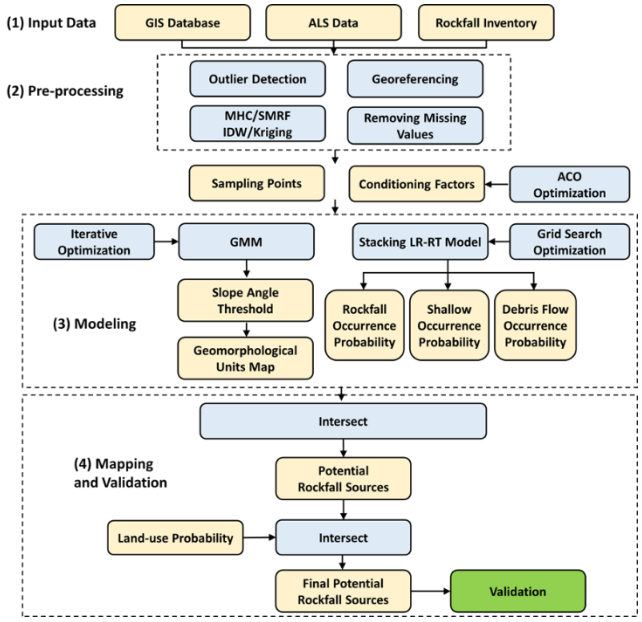


FIGURE 3. The workflow of the proposed integrated model for detecting potential rockfall sources.

B. DTM GENERATION

Filtering process is fundamental step to derive an accurate Digital Terrain Model (DTM). This because of the raw LiDAR data contain ground and up-ground points. Several filtering methods are existed for generating DTM, and this study employs two various algorithms: morphology-based filtering and multi-scale comparison. The former is suitable for terrain with small features and steep terrains whereas the later appropriate for urban area [26]. Simple Morphological Filter (SMRF) is a morphology-based method proposed by Pingel et al. [27]. The multi-resolution hierarchical classification (MHC) is a multi-scale comparison method developed by Chen et al [28]. Generally, the advantage of the multiscale comparison filtering methods is deriving an accurate DTM in urban regions with different man-made natural objects [26]. Both SMRF and MHC were employed in this research to filter the point clouds of different regions of the study area.

Once the point clouds were filtered by removing the non-ground points, both Kriging and IDW were tested based on cross-validation method. IDW was found as an appropriate interpolation method due to the high density of the point cloud [29]. Consequently, IDW interpolation method was used to generate the DTM based on the remaining points with a resolution of 0.5 m.

C. OPTIMIZATION OF THE CONDITIONING FACTORS

The variances between the factors characteristics should be assessed to generate landslide probability maps that utilizes different conditioning factors. The conditioning factors characteristics differ from area to another, thus, the first step in producing probability map is to evaluate the significance of each factor. The conditioning factors building is a hard mission and no particular rules exists to define the number of the conditioning factors that adequate for a particular probability assessment. Moreover, no frameworks exist for the chosen of conditioning factors. Such factors are normally selected based on the experts' opinions [30].

In this research two conditioning factors datasets were built within the environment of GIS. The first dataset were obtained from the airborne LiDAR data encompassing of eight landslide conditioning factors: altitude, aspect, slope, curvature, sediment transport index (STI), topographic roughness index (TRI), topographic wetness index (TWI), and stream power index (SPI). The second dataset was the addition of other conditioning factors: environmental and geological factors of, land use/cover (LULC), distance from road and river, soil, and geology. Seventeen conditioning factors were used for landslide probability mapping.

The analysis of multicollinearity is a significant step in landslide probability. The presence of a near-linear relation among factors can produce a division-by-zero problem during regression calculations. Such problem can result in abortion of the calculations and inexact relationship; division by a very small quantity still deforms the result. Consequently, it is significant to analyze the landslide conditioning factors before probability mapping. The collinear (dependent) factors can be identified through multicollinearity analysis by examining a correlation matrix built via computing R^2. Several quantitative approaches are existing for multicollinearity detection, such as examination the eigenvalue in a correlation matrix, evaluation the variance inflation factor (VIF), and pairwise scatter plots. In the current research, the VIF values were calculated to detect the multicollinearity for each conditioning factor. Moreover, communality similar to R^2 was computed for each factor.

After multicollinearity was performed, Ant Colony Optimization (ACO) with three different types of machine learning algorithms (RF, SVM, and LR), was tested and compared. Consequently, the best fit method was utilized to identify the best subset of landslides conditioning factors. The

area under curve (AUC) and overall accuracy were used to assess the obtained results.

In last decade ACO has attracted an increasing number of researchers, and its applications have developed remarkably [31]. ACO was efficiently employed in several remote sensing applications, such as optimal attribute subset selection [14], image segmentation [32], selection of parameters [33], and object derivation [34]. The advantages of ACO are, first, it makes a probabilistic decision in terms of local heuristic information and artificial pheromone trails and consequently permits the examination of a huge solutions number than greedy heuristic [35]. Second, it assists the evaporation of pheromone trail, that is a process which result in decreasing of pheromone trail intensity along time. The evaporation of pheromone aids to prevent the fast algorithm convergence toward a suboptimal area [36]. Third, in the rule exploring context, an ACO algorithm can implement a robust, flexible search for an optimal terms combination (logical conditions) including the predictor factors values [37]. Moreover, ACO provides better and more reliable results than genetic algorithm (GA) and the common sequence method (CSM) [38-39]. The overall workflow of ACO-based attribute selection is presented in Figure. 4.

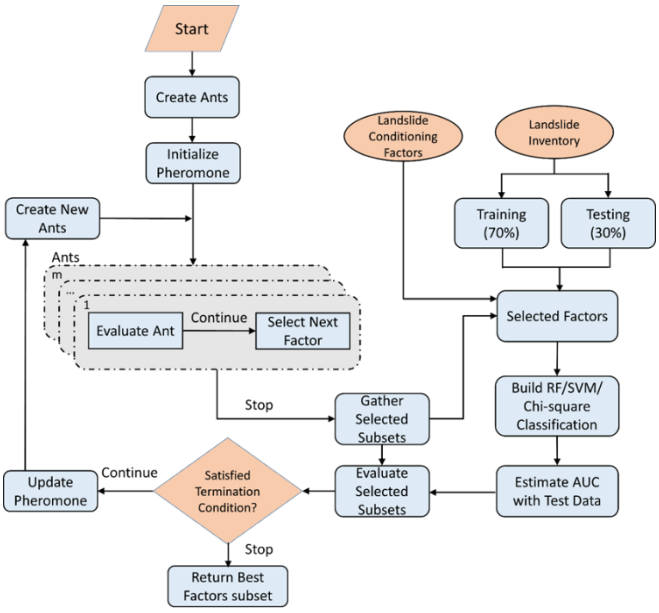


FIGURE 4. **The workflow of ACO-based factors selection.**

Several conditioning factors are available for landslide probability mapping. However, not all the factors are useful, therefore, these factors should be evaluated, and only important (related) factors must be chosen to derive better landslide probability result. Consequently, ACO method was employed in this research to choose important factors in the probability of landslide. A total of 17 conditioning factors (Table 1), were primary chosen according to the condition of the study area and literature. Afterward, the best factors subset with their attributes was selected via ACO, that can attain the highest potential classification result. Different machine learning algorithms (RF, SVM, and LR) were utilized for the evaluation process. The conditioning factors were split into two different groups (70%) training and (30%) testing. In order to perform ACO factors selection, the ACO parameters values should be set according to previous works and a preliminary analysis and [40].

TABLE I
THE SELECTED LANDSLIDE CONDITIONING FACTORS

Attribute ID	Attribute name	Attribute ID	Attribute name
1	Altitude	10	Distance to Road
2	Slope	11	Distance to Lineament
3	Aspect	12	Distance to River
4	Curvature	13	SPI
5	Rainfall	14	STI
6	Landuse	15	TRI
7	Geology	16	TWI
8	Soil	17	Vegetation Density
9	Flow Length		

D. ENSEMBLE MODEL

Ensemble models are methods that combine multiple base models to create a more robust model that can produce improved results. Ensemble models are often more accurate than single models [41-43]. There are several ensemble methods such as stacking, boosting, and voting. Stacking is defined as a way of combining multiple machine learning models [45]. It usually combines models of different types of classifiers. The basic steps of stacking are: divide the training dataset into two separate groups, train many base learners using the first group, test the base learners using the second group, and utilizing the prediction from the previous step as an input, and a proper response as the output, train a higher level learner. Boosting is an algorithm that can be presented as a model averaging approach. It basically developed for classification. However, it also can be used for regression applications. First, a weak classifier is created then a sequence of models are constructed iteratively. Every model trained using the dataset in which points misclassified via the former model are assigned more weight. Eventually, the sequence models are weighted based on their success and the results are gathered through voting thus producing a final model. Boosting means applying a weak classifier, running it multiple times on the training data and then allowing the learned classifier to vote. On the other hand, voting ensembles create multiple models (typically of different algorithms), and simple statistics (such as computing the majority and mean) are utilized to combine prediction.

In this study, several base models such as LR, RT, RF, SVM, ANN were used. In addition, three ensemble methods, stacking, boosting, and voting was investigated. The grid

search method was employed in order to optimize the base models. The best model is then decided according to AUC and overall accuracy. A grid search is a standard search method for selecting sub-optimal hyperparameters of a machine learning / statistical model. Suppose there are *k* parameters and each of them has c_i values, then the number of search possibilities (*P*) is:

$$P = \prod_{i=1}^{k} c_i \qquad (1)$$

E. GAUSSIAN MIXTURE MODEL (GMM)

Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities [45]. GMM is frequently utilized as a parametric model of the probabilities distribution of features or continuous measurements in biometric systems. The parameters of GMM are derived from training dataset utilizing the Maximum A Posteriori (MAP) estimation from a well-trained prior model or the iterative Expectation-Maximization (EM) algorithm.

A GMM is a weighted sum of M component Gaussian densities as given by the equation,

$$p(X/\lambda) = \sum_{i=1}^{M} w_i g(x/\mu_i, \sum_i) \qquad (2)$$

In which x is a D-dimensional continuous-valued data vector (i.e. features or measurements), $w_i, i = 1, \ldots . M$, are the mixture weights, and $g(X/\mu_i, \sum_i), i = 1, \ldots, M,$ are the component Gaussian densities. Each component density is a D-variate Gaussian function of the form,

$$g(X/\mu_i, \sum_i) = \frac{1}{(2\pi)^{D/2}|\sum_i|^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu_i)' \sum_i^{-1} (X - \mu_i)\right\} \qquad (3)$$

With mean vector μ_i and covariance matrix $\sum_i$ The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$.

The complete Gaussian mixture model is parameterized by the mixture weights, covariance matrices, and mean vectors from all component densities. These parameters are collectively represented through the notation,

$$\lambda = \{w_i, \mu_i, \sum_i\} \quad i = 1, \ldots, M. \qquad (4)$$

Many variants are existing on the GMM illustrated in Equation (3). The covariance matrices, $\sum_i$, can be constrained or full rank to be diagonal. In addition, parameters can be tied or shared among the components of Gaussian, such as having a frequent covariance matrix for all components. The selection of model setting (parameter tying, diagonal or full covariance matrices, and components number) is normally identified via the magnitude of dataset available for the estimation of the GMM parameters and how the GMM is utilized in a specified biometric application.

The terrain morphology displays the slope angle characteristics that can be straight linked to the geomorphic processes implicated in the stability of slopes. The rockfall sources are frequently exist in the steepest morphological units [45]. Consequently, Loye et al. [47] have constructed a DEM-based geomorphometric method to identify these morphological units and thus rockfall sources. Using this method, the distribution of slope angle is decomposed in many Gaussian distributions that can be considered morphological units characteristics such as rock cliff, steep slope, moderate steep, foot slope, and plain. When the slope angle exceeds a certain threshold, the terrain is considered a probable source of rockfall which in turn is determined when the Gaussian distribution of the morphological units rock cliff become dominant over the steep slopes units [47].

In many applications, the components number (k), and proper covariance structure, $\sum_i$, are unknown. In order to tune a GMM comparison of information criteria has to be perform. Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are the most common information criteria. AIC and BIC take the negative log likelihood, optimized, and then penalize it with the parameters number in the model (the complexity of the model). Nevertheless, the AIC penalizes for complexity less severely than BIC. Thus, the BIC tends to select simpler model that may underfit, whereas AIC tends to select more complex model that may overfit. Therefore, it is better to look at AIC and BIC criteria to decide the best model. The best fit model is with lower AIC or BIC values.

The GMM was trained on slope dataset extracted from three different areas within the area of interest. Iterative search based on AIC and BIC values was performed to select the hyperparameters of the model. The output of the GMM model was thresholds of slope angle (MUs), which allows detecting the probable sources of various landslide types automatically.

F. ACCURACY ASSESSMENT

The proposed ensemble model was validated based on the success and prediction rate curves. These curves show the cumulative frequency graph and explain the known landslides percentage that fall into probability level ranks [48]. The success rate was generated from the landslides utilized for training whereas the prediction rate was generated using the landslides utilized for validation. Moreover, AUC was used to identify the accuracy of the probability maps qualitatively in which greater AUC means higher accuracy attained [42].
Field verification was also carried out to validate the identified rockfall source areas. Several locations that distributed over the whole study area, were selected randomly from the produced map and then in-situ validation was performed.

V. RESULTS AND DISCUSSION

This section presents the main results and discusses the major findings obtained from this work. First, a summary statistics of modeling data and its pre-processing is given. Then, the results of stacking LR - RF model including the probability maps, optimization results, and comparisons with other methods are explained. After that, the results of GMM are presented. Finally, the validation and field verification were discussed.

A. SUMMARY STATISTICS

The inventory data had 179 sampling points. The slope angles in the landslide data samples ranged from 19^0 to 49^0 with an average slope angle of 34.75^0 (std. = 13.43^0). On the other hand, the slope angles in the rockfall data samples had a minimum of 47^0 and a maximum of 76.39^0. The average slope angle was 65.86^0, and the standard deviation was 11.34^0. For debris flow the slope angles ranged from 7° to 25° and the average slope angle was 15.13°.

In addition, since the sampling points (landslide, debris flow, and rockfall) are subjected to strong correlations in different conditioning factors, the multicollinearity of the factors was analyzed using the Variance Inflated Factor (VIF) method. Table 2 shows the VIF values calculated among the conditioning factors in the landslide, rockfall, and debris flow data samples. According to the previous works [25], a VIF value of greater than four is considered high collinear. Thus, the corresponding factors should be removed from further analysis. The highest VIF value is 3.107 for slope factor in the landslide data samples, 3.272 for STI factor in the rockfall data samples, and 3.155 for altitude factor in the debris flow data samples. As a result, none of the factors was removed.

TABLE II
THE VIF VALUES FOR THE CONDITIONING FACTORS IN LANDSLIDE, ROCKFALL, AND DEBRIS FLOW DATA SAMPLES

Factor	Landslide		Rockfall		Debris Flow	
	Multiple R2	VIF	Multiple R2	VIF	Multiple R2	VIF
Vegetation density	0.417	1.72	0.464	1.87	0.435	1.77
Soil	0.375	1.60	0.331	1.50	0.352	1.54
Landuse	0.454	1.83	0.478	1.92	0.435	1.77
Geology	0.455	1.83	0.410	1.69	0.481	1.93
Rainfall	0.308	1.45	0.268	1.37	0.285	1.40
Aspect	0.290	1.41	0.331	1.49	0.302	1.43
Curvature	0.264	1.36	0.201	1.25	0.247	1.33
Altitude	0.662	2.96	0.671	3.04	0.683	3.16
Flow length	0.412	1.70	0.449	1.82	0.403	1.68
Distance to road	0.491	1.96	0.416	1.71	0.479	1.92
Distance to lineament	0.605	2.54	0.650	2.86	0.597	2.48
Slope	0.678	3.11	0.685	3.18	0.635	2.74
SPI	0.547	2.21	0.623	2.65	0.574	2.35
STI	0.610	2.56	0.694	3.27	0.634	2.73
Distance to stream	0.304	1.44	0.330	1.49	0.317	1.46
TRI	0.379	1.61	0.430	1.75	0.359	1.56
TWI	0.302	1.43	0.333	1.50	0.317	1.46

B. THE RESULT OF THE OPTIMIZED CONDITIONING FACTORS

ACO was used to choose the optimum conditioning factors subset for shallow landslide, rockfall, and debris flow probability. The ACO approach was performed in MATLAB R2016b. The preparation of testing and training datasets were conducted in Excel sheets and GIS. Scaling process was performed for all used factors during the data processing stage. This to minimize computation complexity and to prevent factors with high numerical ranges from dominating those with low numerical ranges. Equation (5) was used to linearly scale each factor to [0, 1]:

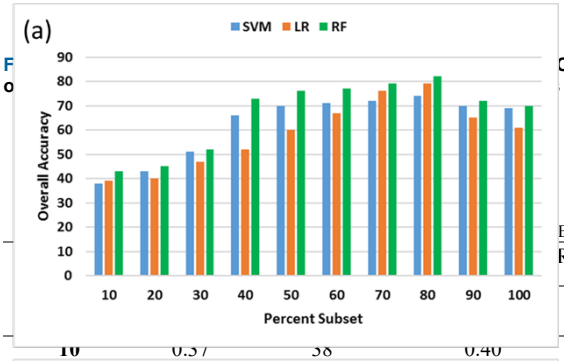
$$\bar{x} = \frac{x - min_i}{max_i - min_i} \qquad (5)$$

In which x and $\bar{x}$ are the original and scaled values, respectively, and min_i is the minimum value of factor i and max_i is the maximum values of factor i.

With the training and testing datasets, many percent subsets of factors were examined with the RF, SVM, and LR algorithms using the area under curve and overall accuracy.
In this study, the implementation of ACO approach has the flexibility to set the factors number needed to be chosen. Ten experiments were performed to choose the best subset starting from 10% to 100% of the 17 selected factors. 100 iterations were conducted in each experiment and then the best subset was chosen. Figure 5 shows the best factors subsets selected by ACO. It also illustrates the classification performance according to the overall accuracy. The highest overall accuracy was observed in experiment 8, 7, and 9 for shallow landslide, rockfall, and debris flow, respectively. In which 14, 12, and 15 factors were selected by the algorithm for shallow landslide, rockfall, and debris flow, respectively. Accordingly, this subset was considered the best factors subset for the dataset and the study area.

The selected factors in each experiment for shallow landslide, rockfall, and debris flow are shown in Table 3, Table 4, and Table 5, respectively, and the factors names are presented in Table 1.

The result indicates that utilizing a big number of conditioning factors does not necessarily increase the accuracy of the calcification. The experiments illustrate that the accuracy of classification deteriorated after experiment 8, 7, and 9 for shallow landslide, rockfall, and debris flow. The ideal number of factors among 17 available factors is 14, 12, and 15 for shallow landslide, rockfall, and debris flow, respectively. Careful analysis of Table 3 indicates that many factors, such as slope, distance to road, and distance to river are strongly affect the shallow landslide probability. In addition, RF outperforms LR and SVM, respectively. However, with the three evaluators the best subset was in experiment 8 with 80 % of conditioning factors (14 factors). Table 4 shows the best subset of factors for rockfall probability. Slope, distance to lineament and TRI, are found highly important for rockfall probability mapping. The best subset was observed in experiment 7 with 70 % of conditioning factors (12 factors). The highest overall accuracy and AUC were recorded with RF evaluator which are 86 and 0.89, respectively. The best factors subset for debris flow is shown in Table 5. Factors such as distance to road, TRI, and vegetation density are found strongly influence the probability of debris flow. Nevertheless, the best subset is in experiment 9 with 90 % of the conditioning factors (15 factors). LR outperforms RF and SVM with overall accuracy of 80. However, based on AUC, RF outperforms LR and SVM, respectively.



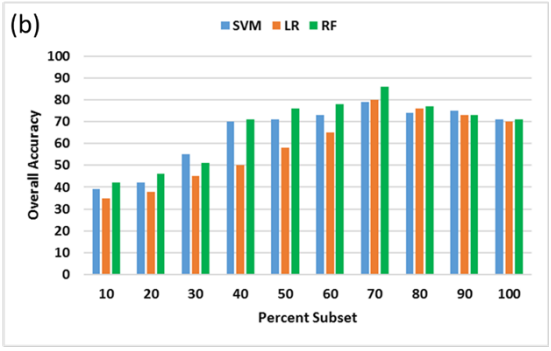


TABLE III
ED ON ACO FOR EACH PERCENT SUBSET FOR SHALLOW LANDSLIDE

R		RF		Selected Factors
Overall Accuracy	AUC	Overall Accuracy		
39	0.41	43		2 10
40	0.47	45		12 2 10
47	0.55	52		6 2 10 12
52	0.71	73		3 15 10 2 12 17 6
60	0.75	76		1 2 8 15 10 9 6 12
67	0.79	77		7 11 2 17 10 6 8 12 1 15
76	0.83	79		2 10 6 8 17 12 16 7 1 10 5 15
79	0.87	82		4 10 12 17 2 15 3 7 11 14 6 5 16 8
65	0.73	72		10 2 4 12 9 11 3 15 17 5 8 6 13 3 7
61	0.71	70		10 7 15 2 12 8 17 13 6 4 1 3 5 11 14 9 16

TABLE IV
S BASED ON ACO FOR EACH PERCENT SUBSET FOR ROCKFALL

Percent Subset	SVM		LR		RF		Selected Factors
	AUC	Overall Accuracy	AUC	Overall Accuracy	AUC	Overall Accuracy	
10	0.36	39	0.33	35	0.41	42	2 11
20	0.41	42	0.37	38	0.49	46	15 11 1
30	0.52	55	0.43	45	0.53	51	2 11 7 1
40	0.67	70	0.51	50	0.69	71	1 11 15 2 4 17 2
50	0.73	71	0.59	58	0.75	76	4 2 8 15 12 9 6 11
60	0.75	73	0.63	65	0.79	78	7 11 13 17 2 6 8 10 1 15
70	0.83	79	0.84	80	0.89	86	11 3 6 8 17 12 2 7 1 10 5 15
80	0.73	74	0.78	76	0.75	77	4 7 13 17 2 15 1 10 11 14 6 8 16 5
90	0.76	75	0.72	73	0.75	73	12 1 4 7 9 11 2 15 17 5 8 6 10 3 13
100	0.73	71	0.74	70	0.72	71	2 7 15 10 11 8 5 13 6 4 1 3 17 12 16 9 14

TABLE V
SELECTED CONDITIONING FACTORS BASED ON ACO FOR EACH PERCENT SUBSET FOR DEBRIS FLOW

Percent Subset	SVM		LR		RF		Selected Factors
	AUC	Overall Accuracy	AUC	Overall Accuracy	AUC	Overall Accuracy	
10	0.34	36	0.38	37	0.44	43	10 17
20	0.40	39	0.40	41	0.49	47	15 10 17
30	0.55	53	0.48	47	0.55	54	10 15 3 17
40	0.67	61	0.55	53	0.67	61	12 11 3 15 10 17 2
50	0.71	65	0.60	61	0.73	68	17 4 2 15 12 9 3 11
60	0.73	67	0.63	65	0.79	71	7 11 13 17 2 6 8 10 1 15
70	0.70	66	0.74	76	0.84	73	12 3 10 8 17 15 2 4 1 11 7 16
80	0.73	69	0.76	78	0.75	77	10 15 12 17 11 4 2 1 13 14 6 8 16 5
90	0.75	72	0.78	80	0.85	79	12 3 15 7 13 11 2 10 17 5 8 4 10 1 9
100	0.71	68	0.73	73	0.71	74	15 3 2 10 11 17 5 13 12 4 1 9 17 6 16 7 8

C. The Results of Stacking LR-RT

The produced probability maps of shallow landslide, rockfall, and debris flow based on stacking LR - RF model are illustrated in Figure 6. Figure 6a shows the probability map of shallow landslide occurrence, the probability maps of rockfall and debris flow occurrences are shown in Figure 6b and Figure 6c, respectively. The stacking LR - RT model predicted the occurrence probability, ranging from 0 (no potential of landslide/rockfall/debris flow occurrence) to 1 (very high potential of landslide/rockfall/debris flow occurrence).

However, to simplify the interpretation of the maps, the probability values were reclassified into five classes using quantile method. The classes were very low, low, moderate, high, and very high. According to the generated probability maps Gua Tambun, Gunung Rapat, and Gunung Lang areas are highly potential for landslides occurrences. However, the highest probability density was observed in rockfall probability map following by shallow landslide and debris flow probability maps, respectively.

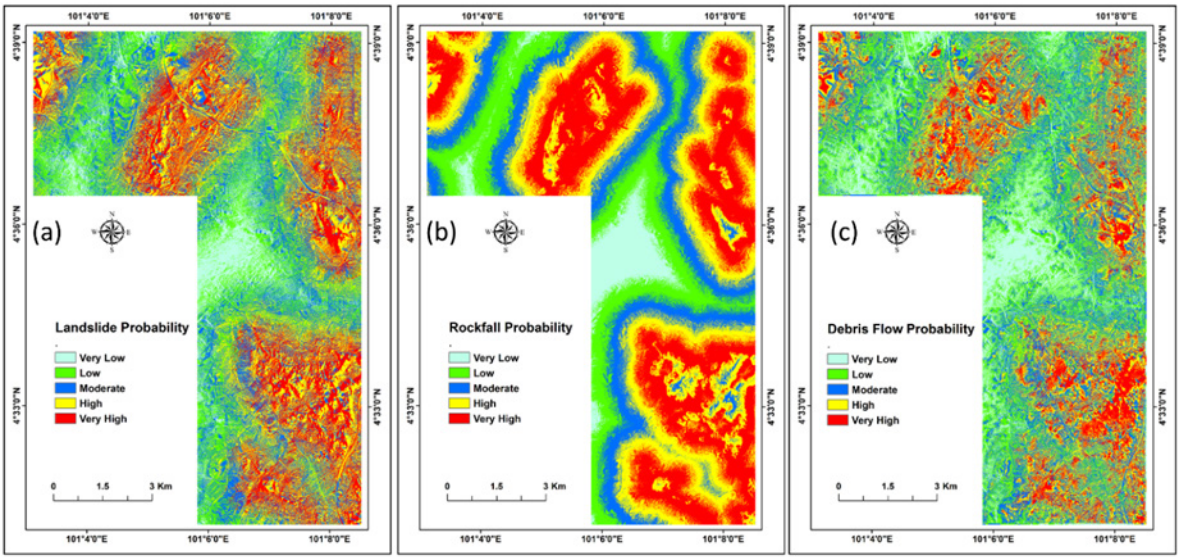


FIGURE 6. **Probability maps based on stacking LR – RT model of (a) shallow landslide, (b) rockfall, and (c) debris flow.**

D. THE RESULT OF GAUSSIAN MIXTURE MODEL (GMM)

This section presents the distributions of slope angle and thresholds calculated through the optimized GMM model. The optimal k value was determined to be 5 for all the three areas (Gunung Rapat, Gua Tambun, and Gunung Lang) (Figure 7). The average number of iterations and regularization value were 500 and 0.01, respectively. The means and standard deviations of five slope angle distributions (MU) are also computed. For Gunung Lang area, GMM has determined 5 MU values, 1.78, 6.07, 16.04, 41.25, and 63.54, for five geomorphological units, plains, foot slopes, moderately steep slopes, steep slopes, and cliffs. The plain unit indicates low slope angles corresponding to the fluvial and fluvioglacial deposits. Foot slope is gentle slope angles featuring the lower part of the hillslope characterized by colluvial fans, debris flow, and landslide deposits. Moderately steep slopes and steep slopes are the units those containing deposits and rocky outcrops covered with vegetation. On the other hand, cliffs are very steep slopes, which correspond to rocky outcrops. The slope angles distributions for Gua Tambun area were 1.92, 5.18, 13.35, 37.60, and 63.78. On the other hand, the MU values for Gunung Rapat area were 1.46, 6.23, 16.43, 43.21, and 66.31. The debris flow threshold identified when moderate steep slope dominates foot-slope. Whereas, shallow landslide determined when steep slope dominates moderate steep slope. On the other hand, rockfall source area located when cliff dominates steep slope. In Gunung Lang area, the slope thresholds were determined as 10° – 23°, 23° – 59°, and > 59° for debris flow, shallow landslide, and rockfall, respectively. Whereas, in Gua Tambun area, the slope thresholds were found 9° – 19°, 19° – 57°, and > 57° for debris flow, shallow landslide, and rockfall, respectively. On the other hand, the slope thresholds were calculated as 9° – 24°, 24° – 58°, and > 58° for debris flow, shallow landslide, and rockfall, respectively. Therefore, the used slope thresholds are 9° – 22°, 22° – 58°, and > 58° for debris flow, shallow landslide, and rockfall, respectively.

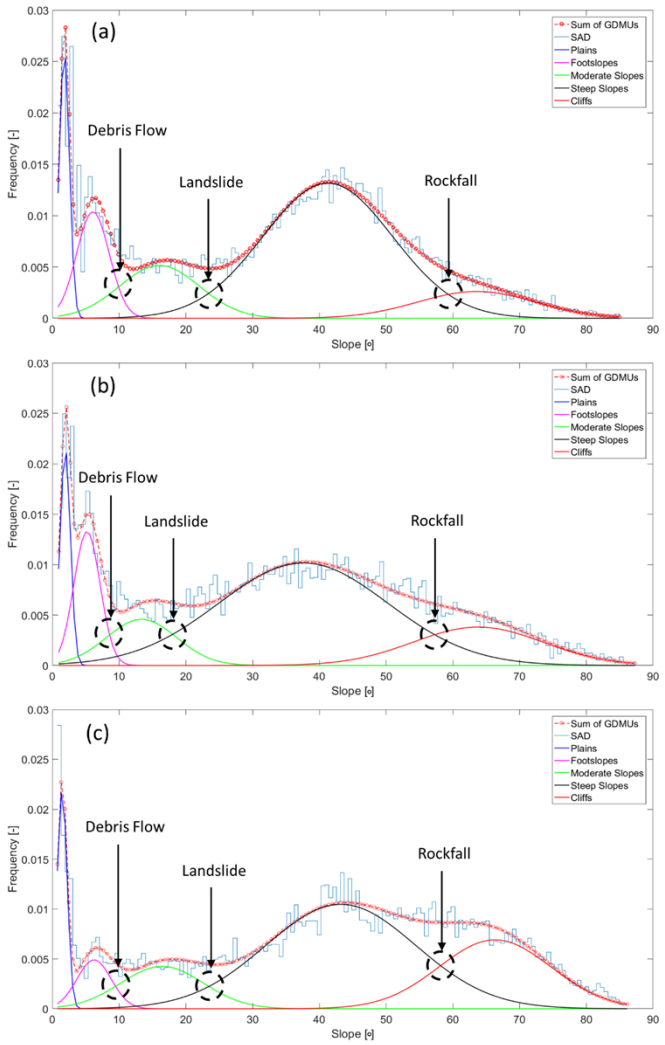


FIGURE 7. Slope distribution and thresholds based on GMM for (a) Gua Tambun, (b) Gunung Rapat, and (c) Gunung Lang.

E. LAND-USE PROBABILITY AREA

Land-use probability area is a raster encompasses of two classes that represent the regions which are probable or not probable to encounter landslides. This raster is formed based on the land-use map. The land-use factor is classified into two classes by integrating residential building, river, transportation, other building, a cemetery, and water body in one class and the other class contains forest, mixed vegetation, and open land. The advantage of this binary raster is reducing the model sensitivity to the spatial variation in the conditioning factors and the computing time. In average, 23% of Kinta Valley is prone to landslides occurrences.

F. RESULT OF ROCKFALL SOURCES REGIONS

The intersections of slope angles thresholds determined by GMM and probability values resulted in the potential landslide and rockfall source areas. This section presents the results of detecting potential landslide and rockfall source areas. The potential sources of landslides were detected by intersecting the slope angles threshold (23^0) and the probability values estimated through stacking LR – RT model.

The identification of probable rockfall sources was performed by crossing the obtained threshold of slope angle (> 58^0) and the probability values of the sacking LR - RT. Figure 8 demonstrates the result of potential rockfall sources in the study areas. The inventories of rockfalls are mainly situated in the regions those detected as probable source areas proving that the proposed model is eligible to identify the sources of both rockfalls and landslides in the selected areas.

FIGURE

G. TH odel for

DC) and

prediction (PRC) rate curves were employed [46]. The ROC is generated based on the training landslide dataset while PRC was produced utilizing the validation landslide dataset. Moreover, the area under curve (AUC) is used to identify the accuracy of the probability maps [48-50].

Table 6 illustrates the comparison among the various machine learning algorithms individually based on ROC, PRC, and overall accuracy. RF outperforms ANN, LR, SVM, and RT, respectively. The highest accuracy (ROC = 0.89, PRC = 0.86, and overall accuracy = 85.62), was achieved by RF with rockfall data samples. However, all algorithms achieved the highest accuracy with rockfall data samples in comparison with shallow landslide and debris flow. On the other hand, LR, RT, and ANN performed better with debris flow data samples than shallow landslide. Whereas, RF and SVM performed better with shallow landslide data samples than debris flow. The comparison of bagging models is shown in Table 7. RF outperformed the other machine learning algorithms. Nevertheless, RF in bagging model was slightly improved in comparison with individual RF this is because RF itself is ensemble model. On the other hand, the bagging models of the other algorithms showed better improvement than the individual implementation of the same algorithms. Table 8 demonstrates the comparison of voting models. Nine voting ensemble models were tested based on the selected machine learning algorithms using rockfall, shallow landslide, and debris flow data samples. The accuracy of some algorithms was improved by adding another algorithm while some the accuracy of some algorithms was decreased by adding another algorithm. For instance, the accuracy of LR was improved by adding RT, RF, and ANN, whereas the accuracy of RF was decreased by adding SVM and ANN. Table 9 illustrates the comparison of different stacking models. The stacking LR – RT model outperformed the other tested models with all landslides data samples (shallow landslide, rockfall, and debris flow). The highest accuracy that achieved using stacking LR – RT model with rockfall data samples is 0.94, 0.95, and 92.89 of ROC, PRC, and overall accuracy. However, the model achieved good accuracy with shallow landslide and debris flow of 0.86, 0.88, and 86.46, and 0.84, 0.81, and 83.95 of ROC, PRC, and overall accuracy, respectively. In comparison with some other tested models in the literature, the stacking LR - RT model achieved better accuracy than a hybrid Bagging based Support Vector Machines (BSVM) model [51], Decision tree (DT)-based CHi-squared Automatic Interaction Detection (CHAID) model [52], a hybrid approach of Reduced Error Pruning Trees (REPT) and Random Subspace Ensemble (RSS) [53], and a hybrid fuzzy weight of evidence model [54]. Consequently, the stacking LR – RT model was employed to produce the probability maps of shallow landslide, rockfall, and debris flow.

Since the stacking LR – RT model performs well on both datasets (training and validation), the proposed ensemble model is thus considered effective for landslides probability mapping. In addition, as the model achieved a good accuracy on validation landslides, it is expected to perform well in other regions that have similar or nearly similar condition as the tested area.

TABLE VI
COMPARISON BETWEEN THE INDIVIDUAL ALGORITHMS FOR DIFFERENT LANDSLIDE TYPES

Algorithm	Rockfall			Shallow Landslide			Debris Flow		
	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy
LR	0.84	0.82	79.98	0.79	0.75	78.55	0.78	0.75	79.63
RT	0.81	0.79	77.24	0.73	0.72	67.96	0.73	0.72	68.96
RF	0.89	0.86	85.62	0.87	0.88	81.61	0.85	0.81	78.61
SVM	0.83	0.81	78.86	0.77	0.72	73.99	0.75	0.73	71.83
ANN	0.86	0.84	82.93	0.68	0.65	64.08	0.69	0.67	66.08

TABLE VII
COMPARISON OF BAGGING MODELS BASED ON THE SELECTED ALGORITHMS

Algorithm	Rockfall			Shallow Landslide			Debris Flow		
	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy
LR	0.89	0.87	83.74	0.80	0.82	79.79	0.83	0.81	80.79
RT	0.87	0.86	84.06	0.83	0.80	76.59	0.81	0.79	75.02
RF	0.92	0.91	86.99	0.89	0.87	82.93	0.86	0.88	82.61
SVM	0.91	0.88	84.55	0.76	0.75	76.85	0.76	0.75	74.59
ANN	0.89	0.90	85.37	0.73	0.74	68.25	0.73	0.74	71.25

TABLE VIII
COMPARISON OF VOTING MODELS BASED ON THE SELECTED ALGORITHMS

Algorithm	Rockfall			Shallow Landslide			Debris Flow		
	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy

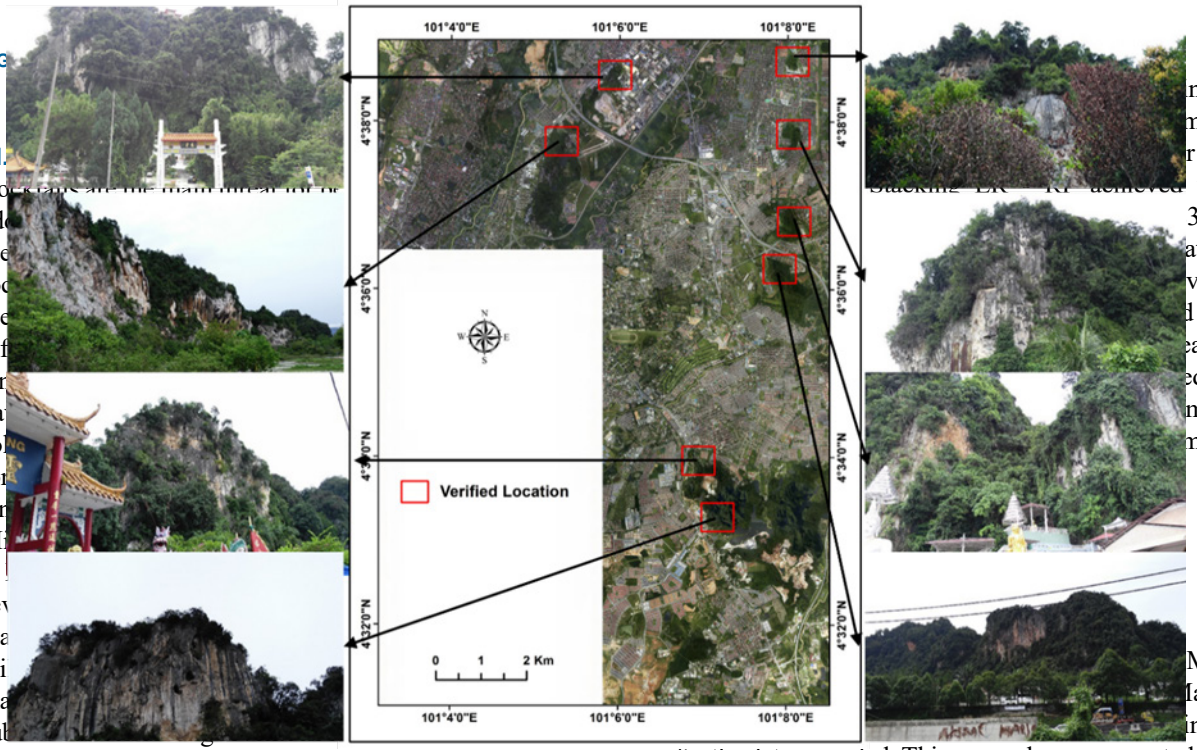
LR - RT	0.89	0.91	88.19	0.83	0.80	83.15	0.79	0.82	82.97
LR - RF	0.89	0.87	86.37	0.81	0.76	80.55	0.81	0.76	79.55
LR - SVM	0.82	0.83	78.86	0.79	0.80	76.99	0.79	0.80	78.99
LR - ANN	0.86	0.82	86.62	0.71	0.65	70.87	0.71	0.65	71.87
RT - SVM	0.82	0.79	81.30	0.65	0.63	64.08	0.64	0.60	66.08
RT – ANN	0.88	0.85	87.06	0.67	0.61	65.05	0.65	0.60	64.05
RF – SVM	0.79	0.77	78.86	0.67	0.62	66.99	0.67	0.62	68.99
RF – ANN	0.84	0.81	84.55	0.65	0.62	65.05	0.65	0.60	67.05
SVM-ANN	0.80	0.76	78.86	0.69	0.65	66.99	0.67	0.62	64.99

TABLE IX
COMPARISON OF STACKING MODELS BASED ON THE SELECTED ALGORITHMS

Algorithm	**Rockfall**			**Shallow Landslide**			**Debris Flow**		
	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy	ROC Area	PRC Area	AO Accuracy
LR - RT	0.94	0.95	92.89	0.86	0.88	86.64	0.84	0.81	83.95
LR - RF	0.94	0.92	91.87	0.84	0.86	81.17	0.83	0.80	81.93
LR - SVM	0.90	0.90	87.19	0.83	0.82	78.99	0.80	0.81	80.97
LR - ANN	0.92	0.90	85.12	0.83	0.82	77.96	0.71	0.73	74.75
RT - SVM	0.90	0.92	88.19	0.81	0.79	76.99	0.72	0.69	71.08
RT – ANN	0.90	0.89	85.12	0.74	0.72	70.19	0.69	0.67	68.94
RF – SVM	0.82	0.81	80.42	0.77	0.78	72.14	0.73	0.69	71.78
RF – ANN	0.85	0.86	81.86	0.72	0.71	71.84	0.65	0.60	69.93
SVM-ANN	0.87	0.84	80.49	0.80	0.78	75.05	0.67	0.62	67.87

H. FIELD VERIFICATION

In-situ survey was carried out to verify the produced map of rockfall sources areas. Several locations were selected randomly that distributed on the entire study area. Figure 9 shows the selected locations and the in-situ photos. All locations were found reasonably prone to rockfall incidents. Discontinuities and fractures are clear in the selected locations especially along the cliff face. In addition, some inventory fallen rocks were observed in various locations close to the foot-slope. According to different interviews that performed in the study area, these areas were encountered rockfall incidents in different periods of time.



FIG

ned the best
mble models
r voting and
the highest
3.95% with
ata samples,
ved a good
d to perform
early similar
ed model can
mance of the
mproved by

Mineral and
Malaysia, the
ng Malaysia

the used factors), were selected for shallow landslide probability based on the highest accuracy of 0.87 and 82% for AUC and overall accuracy, respectively. On the other hand, twelve of the best subset of conditioning factors were selected for the producing of rockfall probability based on the highest overall accuracy (86%) and area under curve (AUC = 0.89). For the generating of debris flow probability map, fifteen on the best conditioning factors were chosen based on the highest accuracy of 0.85 and 79% for AUC and overall accuracy.

The GMM model was integrated with stacking LR – RT model to automatically determining the slope thresholds and detecting the high probability of landslide/rockfall/debris flow occurrences combined. The GMM was precisely defined the distributions of slope angle and identify the thresholds of slope angle that permitted the source identfication of debris flow (at 9° – 22°), shallow landslide (at 22^0 - 58°), and rockfall (at >

for the data provided. This research was supported by the UTS under grant number 323930, 321740.2232335 and 321740.2232357

REFERENCES

[1] D. J. Varnes, “Slope movement types and processes,” Special report, 176, pp.11-33, 1978.

[2] M. K. Ansari, M. Ahmad, R. Singh, T. N. Singh, "2D and 3D rockfall hazard analysis and protection measures for Saptashrungi Gad Temple, Vani, Nashik, Maharashtra–A case study," J GEOL SOC INDIA, vol. 91, no. 1, pp. 47-56, 2018.

[3] A. M. Fanos, B. Pradhan, "Laser scanning systems and techniques in rockfall source identification and risk assessment: a critical review," Earth Syst Environ, pp. 1-20, 2018.

[4] K. Messenzehl, H. Meyer, J. C. Otto, T. Hoffmann, R. Dikau, "Regional-scale controls on the spatial activity of rockfalls (Turtmann valley, Swiss Alps)—a multivariate modeling approach," Geomorphology, vol. 287, pp. 29-45, 2017.

[5] B, Matasci *et al*., "Impacts of fracturing patterns on the rockfall susceptibility and erosion rate of stratified limestone," Geomorphology, vol. 241, pp. 83-97, 2015.

[6] A. M. Fanos, B. Pradhan "Multi-scenario rockfall hazard assessment using LiDAR data and GIS," Geotech Geol Eng, vol. 34, no. 5, pp. 1375-1393, 2016.

[7] B. Pradhan, A. M. Fanos, "Rockfall hazard assessment: an overview," In Laser Scanning Applications in Landslide Assessment, 2017, pp. 299-322. Springer, Cham.

[8] P. Frattini, G. Crosta, A. Carrara, F. Agliardi. "Assessment of rockfall susceptibility by integrating statistical and physically-based approaches," Geomorphology, vol. 94, no. 3-4, pp. 419-437, 2008.

[9] C. Corona, D. Trappmann, M. Stoffel, "Parameterization of rockfall source areas and magnitudes with ecological recorders: when disturbances in trees serve the calibration and validation of simulation runs," Geomorphology, vol. 202, pp. 33-42, 2013.

[10] L. Losasso, M. Jaboyedoff, F. Sdao, "Potential rock fall source areas identification and rock fall propagation in the province of Potenza territory using an empirically distributed approach," Landslides, vol. 14, no. 5, pp. 1593-1602, 2017.

[11] E. Hoek, J. D. Bray, Rock slope engineering. CRC Press, 2014.

[12] D. R. Montgomery, M. T. Brandon, "Topographic controls on erosion rates in tectonically active mountain ranges," Earth Planet. Sci. Lett, vol. 201, no. 3-4, pp. 481-489, 2002.

[13] A. N. Strahler, "Statistical analysis in geomorphic research," J. Geol., vol. 62, no. 1, pp. 1-25, 1954.

[14] M. N. Jebur, B. Pradhan, M. S. Tehrany, "Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale," REMOTE SENS ENVIRON, vol. 152. pp. 150-165, 2014.

[15] Y Gu *et al*., "A novel MKL model of integrating LiDAR data and MSI for urban area classification." IEEE Trans Geosci Remote Sens, vol. 53, no. 10, pp. 5312-5326, 2015.

[16] L. Zhang et al., "Hyperspectral Image Unsupervised Classification by Robust Manifold Matrix Factorization". *Information Sciences*, 2019.

[17] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *Geosci. Remote Sens.* ***Mag***, vol. *5, no.* 4, pp.8-36, 2017.

[18] S. P. Potharaju, M. Sreedevi, "Ensembled rule based classification algorithms for predicting imbalanced kidney disease data," J Eng Sci Technol Rev, vol. 9, no. 5, pp. 201-207, 2016.

[19] M. Akour, I. Alsmadi, I. Alazzam, "Software fault proneness prediction: a comparative study between bagging, boosting, and stacking ensemble and base learner methods," IJDATS, vol. 9, no. 1, pp. 1-16, 2017.

[20] A. M. Fanos, "Assessment of multi-scenario rockfall hazard based on mechanical parameters using high-resolution airborne laser scanning data and GIS in a tropical area," Environ. Earth Sci., vol. 75, no. 15, p. 1129, 2016.

[21] B. Pradhan, A. M. Fanos, "Application of LiDAR in rockfall hazard assessment in tropical region," In Laser Scanning Applications in Landslide Assessment, 2017, pp. 323-359. Springer, Cham.

[22] R. Muzzillo, L. Losasso, F. Sdao, "Rockfall Source Areas Assessment in an Area of the Pollino National Park (Southern Italy)," In International Conference on Computational Science and Its Applications, 2018, pp. 366-379. Springer, Cham.

[23] A. M. Fanos, "A hybrid model using machine learning methods and GIS for potential rockfall source identification from airborne laser scanning data," Landslides, vol. 15, no. 9, pp. 1833-1850, 2018.

[24] B. Pradhan *et al*., "Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS." NAT HAZARDS, vol. 73, no. 2, pp. 1019-1042, 2014.

[25] H. Hong *et al*., "Spatial prediction of rotational landslide using geographically weighted regression, logistic regression, and support vector machine models in Xing Guo area (China)," GEOMAT NAT HAZ RISK, vol. 8, no. 2, pp. 1997-2022, 2017.

[26] Z. Chen, B. Gao, B. Devereux, "State-of-the-art: DTM generation using airborne LIDAR data," Sensors, vol. 17, no. 1, p. 150, 2017.

[27] T. J. Pingel, K. C. Clarke, W. A. McBride, "An improved simple morphological filter for the terrain classification of airborne LIDAR data," ISPRS J Photogramm Remote Sens, vol. 77, pp. 21-30, 2013.

[28] C. Chen, Y. Li, W. Li, H. Dai, "A multiresolution hierarchical classification algorithm for filtering airborne LiDAR data," ISPRS J Photogramm Remote Sens, vol. 82, pp. 1-9, 2013.

[29] J. Li, A. D. Heap, "Spatial interpolation methods applied in the environmental sciences: A review," Environ Model Softw, vol. 53, pp. 173-189, 2014.

[30] W. M. Abdulwahid, B. Pradhan, "Landslide vulnerability and risk assessment for multi-hazard scenarios using airborne laser scanning data (LiDAR)," Landslides vol. 14, no. 3, pp. 1057-1076, 2017.

[31] J. Shang *et al*., "A Review of Ant Colony Optimization Based Methods for Detecting Epistatic Interactions," IEEE Access, 2019.

[32] L. Y. Cao, L. Z. Xia, "An ant colony optimization approach for SAR image segmentation," In 2007 International Conference on Wavelet Analysis and Pattern Recognition, vol. 1, pp. 296-300. IEEE, 2007.

[33] H. B. Alwan, K. R. Ku-Mahamud, "Optimizing support vector machine parameters using continuous ant colony optimization," In 2012 7th International Conference on Computing and Convergence Technology (ICCCT), pp. 164-169. IEEE, 2012.

[34] L. Li, J. Wang, "SAR image ship detection based on ant colony optimization," In 2012 5th International Congress on Image and Signal Processing, pp. 1100-1103. IEEE, 2012.

[35] J. Gottlieb, M. Puchta, C. Solnon, "A study of greedy, local search, and ant colony optimization approaches for car sequencing problems" In Workshops on Applications of Evolutionary Computation, pp. 246-257. Springer, Berlin, Heidelberg, 2003.

[36] M. Dorigo, T. Stützle, "The ant colony optimization metaheuristic: Algorithms, applications, and advances," In Handbook of metaheuristics, pp. 250-285. Springer, Boston, MA, 2003.

[37] R. S. Parpinelli, "Data mining with an ant colony optimization algorithm." IEEE Trans. Evol. Comput., vol .6, no. 4, pp. 321-332, 2002.

[38] R. Putha, L. Quadrifoglio, E. Zechman, "Comparing ant colony optimization and genetic algorithm approaches for solving traffic signal coordination under oversaturation conditions," COMPUT-AIDED CIV INF, vol. *27* no. 1, pp.14-28, 2012.

[39] Z. Jiang et al., "Comparing an ant colony algorithm with a genetic algorithm for replugging tour planning of seedling transplanter," *COMPUT ELECTRON* ***AGR***, vol. *113*, pp.225-233, 2015.

[40] D. N. Kumar, M. J. Reddy, "Ant colony optimization for multi-purpose reservoir operation," WATER RESOUR MANAG, vol. 20, no. 6, pp. 879-898, 2006.

[41] A. M. Youssef, "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, vol. 13, no. 5, pp. 839-856, 2016.

[42] B. T. Pham *et al*., "A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India)," Environ Model Softw, vol. 84, pp. 240-250, 2016.

[43] L. Xiao, Y. Zhang, G. Peng, "Landslide susceptibility assessment using integrated deep learning algorithm along the china-nepal highway," Sensors, vol. 18, no. 12, p. 4436, 2018.

[44] Z. Ouyang *et al*., "Multi-view stacking ensemble for power consumption anomaly detection in the context of industrial internet of things," IEEE Access, vol. 6, pp. 9623-9631, 2018.

[45] Y. Chen, T. T. Georgiou, A. Tannenbaum, "Optimal transport for Gaussian mixture models," IEEE Access, vol. 7, pp. 6269-6278, 2019.

[46] C. Michoud *et al*., "Rockfall hazard and risk assessments along roads at a regional scale: example in Swiss Alps," Nat. Hazards Earth Syst. Sci., vol. 12, no. 3, 2012.

[47] A. Loye, M. Jaboyedoff, A. Pedrazzini, "Identification of potential rockfall source areas at a regional scale using a DEM-based geomorphometric analysis," Nat. Hazards Earth Syst. Sci , vol. 9, no. 5, pp. 1643-1653, 2009.

[48] N. Intarawichian, S. Dasananda, "Frequency ratio model based landslide susceptibility mapping in lower Mae Chaem watershed,

Northern Thailand," Environ. Earth Sci., 64, no. 8, pp. 2271-2285, 2011.
[49] J. Mathew, V. K. Jha, G. S. Rawat, "Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method," Landslides 6, no. 1, pp 17-26, 2009.
[50] H. R. Pourghasemi *et al*., "Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran," J. Earth Syst. Sci ,vol. 122, no. 2, pp. 349-369, 2013.
[51] B. T. Pham, D. T. Bui, I, Prakash, "Bagging based Support Vector Machines for spatial prediction of landslides," Environ Earth Sci,, vol. *77, no* 4, pp.146, 2018.
[52] O. F. Althuwaynee et al.. "A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping," *Landslides*, vol. *11, no* 6, pp.1063-1078, 2014.
[53] B. T. Pham, I. Prakash, "A novel hybrid model of bagging-based naïve bayes trees for landslide susceptibility assessment," *B ENG GEOL ENVIRON,* pp.1-15, 2017.
[54] H, Hong et al., "A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China, " *Geomorphology, vol. 290*, pp.1-16, 2017.