International Journal of Advanced Technology in Engineering and Science
Vol. No. 10, Issue No. 06, June 2022
www.ijates.com
ijates
ISSN 2348 - 7550

Risk Assessment of Landslides Using Fuzzy Techniques

[1]B. Rajya Lakshmi, [2]G. Krishna Swami, [3]G. Narendra,

[4]A. Brahma Reddy, [5]Jagadeesh Thati

[1,2,3,4]U.G. Student, Tirumala Engineering College, Narasaraopet, A.P

[5]Associate professor, Department of Electronics and Communication Engineering, Tirumala Engineering College , Jonnalagadda, Narasaraopet, Andhra Pradesh.

[1]busireddyrajyalakshmi@gmail.com, [2]gajavallikrishnaswamy@gmail.com, [3]gorantlanarendra0@gmail.com, [4]brahmareddyavula09@gmail.com,[5]jagadeeshthati@gmail.com

ABSTRACT

Landslides can result in enormous casualties and huge economic losses in mountainous regions. In order to mitigate landslide hazard effectively, new methodologies are required to develop a better understanding of landslide hazard. Data-driven risk assessment of landslide plays a vital role in preventing the incoming landslide occurrences. In this paper, we develop a human-centric framework using information granules to perform risk assessments of a group of landslides. The proposed methodology consists of (i) select density-based spatial clustering of applications with noise(DBSCAN) to sub clusters for land risk indication. The clustering outcomes are visualized via t-distributed stochastic neighbor embedding(t-SNE) in 2-D embedding space. (ii) the sub clusters produced by DBSCAN are computed for granular construction. (iii) interval-based information granules are constructed and measured via coverage, specificity and area under the coverage-specificity curve(AUC). The optimal information granules are constructed with two risk measures namely, Value-at-Risk(VaR) and Conditional-Value-at-Risk(CVaR). These measures are computed to interpret the rule-based information-granules with respect to the key attributes. The proposed methodology is capture the main essence of landslide pattern with higher interpretability and help to reduce the computing overhead.

INDEX TERMSDBSCAN, Landslide Risk, Information granule, VaR, CVaR

I INTRODUCTION

Landslide is an important geological hazard that causes damage to natural and social environment. The concept of landslide is dealt by many authors differently. Varnes and IAEG defined landslides as 'almost all varieties of mass movements on slope including some such as rock falls, topples and debris flow that involve little or no true sliding'. Brusden considered landslides as a unique form of mass transport and a process which do not require a transportation medium such as water, air or ice. Crozier defined landslides as 'the outward and downward gravitational movement of the earth material without the aid of running water as a transporting agent'. According to Hutchinson, 'A landslide in its strict sense is a relatively rapid mass wasting process that causes the down slope movement of mass of rock, debris or earth triggered by variety of external stimulus'. A recent definition by Courture R simply states that 'landslide is a movement of mass of soil (earth or debris) or rock down a slope'. This concept of landslide is more broaden with respect to the type of material that moves down slope.The majority of the landslide events are water-induced where the natural and human related water activities reduced the slope stability and induced slope failures [1]. Data-driven approaches to manage the landslides are becoming more popular in recent years as they necessarily characterize the landslides both internally and externally [2]. The external characteristics of landslides are often geomorphology related such as distance to rivers, distance to road, elevation, and slope. These external factors can be easily observed and measured via on-site investigation by geologists with domain knowledge. Meanwhile, measuring the internal characteristics is a more challenging task. The practitioners need to rely on point sensors, physical samples, soil samples, and many other in-depth characteristics. Both external and internal measures are often merged to construct the meta-dataset for the risk analytics of landslide occurrences. Based on the meta-dataset, many conventional data-driven analysis and recent approaches can be performed to indicate the potential risk of the underlying landslide.

However, despite many data-driven techniques aiming to assess the landslide risks using meta-data, the interpretability of such results is often insufficient and vague for readers. As a consequence, the reproducibility of those techniques can be only limited to a small dataset but can hardly be applied universally to the research problem with sufficient level of interpretability. For conventional models, Kirschbaum et al. [3] proposed a meta-heuristic fuzzy overlay model to create a regional susceptibility map and evaluated the performance using receiver operating characteristic (ROC) curves. Van et al. [4] integrated the analytical hierarchy process and weighted linear combination to construct the risk assessment model for landslide occurrences. Nefeslioglu et al. [5] improved the vanilla analytical hierarchy process and evaluated the landslide susceptibility. Ahmed. [6] performed comparative analysis among various state-of-the-art approaches such as Artificial Hierarchy Process (AHP), Weighted Linear Combination (WLC), and Ordered Weighted Average (OWA) to assess landslide risks.

For the time being, the more recent approaches using machine-learning and artificial intelligence are becoming the mainstream of predicting landslide risks. They often describe the highly nonlinear relationships between internal/external factors and landslide risks and can formulate the problem as classification or regression problem quantitatively. Althuwaynee et al. [7] initially performed multivariate analysis of landslide risk assessment using decision trees (DT) and logistic regression (LR) models. Huang et al. [8] modelled the landslide susceptibility using support vector machine (SVM) with a case study analysis in Nantian area of China. Youssef et al. [9] evaluated the landslide risk using random forest (RF) model to classify the landslide risk levels including low, moderate, high and extremely high. Eleven landslide conditioning factors were prepared in the study which contains both internal and external factors. Gorsevski et al. [10] performed landslide risk assessment with a case study in the Cuyahoga Valley National Park, Ohio using artificial neural network (ANN) integrated with lidar data. The ANN algorithms provided superior performance in construct nonlinear mapping between the landslides and predictor attributes and the prediction accuracy outperforms on average.

Based on the discussed outlined above, a data-driven granular computation framework is proposed in this study to assess the landslide risks. First, the density-based spatial clustering of applications with noise (DBSCAN) algorithm is adopted to perform the clustering analysis to cluster the landslide dataset into subgroups. Second, a collection of prototypes within each cluster are selected as the representatives of each sub-cluster and the interval-based information granules are constructed accordingly. The neighboring data points are computed with respect to the density to discover the “thinnest” cluster which is the optimality with respect to coverage and specificity. In addition, to measure the classification performance of landslide risks using these constructed interval-based granular, three metrices including coverage, specificity, and area under the coverage specificity curve (AUC) are computed. Comparative analysis is also performed against the benchmarking clustering algorithms such as fuzzy C-means (FCM), k-mean and k-medoids. The two statistical measures namely Value-at-Risk and Conditional-Value-at-Risk are computed to interpret the generated rule-based information-granules. The main contribution of this paper is as follows:

• This research firstly introduced the concept of granular computation into the field of landslide risk assessment. None of the related work has been discussed in this field yet.

• Second, the proposed approach provided explicit decision boundaries with respect to attributes which illustrates the decision-making process of the landslide risk assessment.

To know the proposed approach clearly, the paper is organized as follows. Section II deals with methods utilized and Section III deals with results, Section IV deals with conclusion of the research.

II METHODOLOGY

1. DBSCAN ALGORITHMDBSCAN clustering is a super useful clustering algorithm for unsupervised learning problems. DBSCAN is a density-based clustering algorithm that works on the assumption that clusters are dense regions in space separated by regions of lower density.It groups ‘densely grouped’ data points into a single cluster. It can identify clusters in large spatial datasets by looking at the local density of the data points. The most exciting feature of DBSCAN clustering is that it is robust to outliers. It also does not require the number of clusters to be told beforehand, unlike K-Means, where we have to specify the number of centroids.

DBSCAN requires only two parameters: epsilon and minPoints. Epsilon i s the radius of the circle to be created around each data point to check the density and minPoints is the minimum number of data points required inside that circle for that data point to be classified as a Core point.

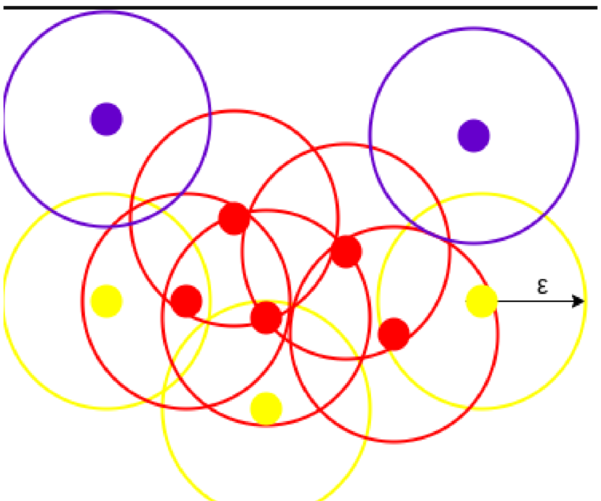


Figure 1: Clusters created by DBSCAN with minpoints=3

One can observe three different points as a part of DBSCAN clustering i) core points ii) boarder points iii) noise points. From the above figure core points are represented by red color, boarder points by yellow and noise points by purple color. The time-complexity for running a DBSCAN algorithm is O(n^2).

2. VISUALIZATION USING T-SNE

T- Distributed Stochastic Neighbor Embedding visualizes high-dimensional data by giving each data point a location in a two or three-dimensional map. To visualize multi-dimensional dataset, linear compression techniques such as Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) are being widely used in previous literature. They often well perform with low-dimensional dataset with fair quality of preserving the original data structure. However, one major drawback of these conventional methods is that they often fail to generate high-quality low-dimensional representation of the original high-dimensional dataset[10]. To address this issue, t-distributed Stochastic Neighbor Embedding (t-SNE) proposed by Van der Maaten & Hinton [11] offered a more reliable solution.

T-SNE is much easier to optimize, and produces significantly better visualizations

International Journal of Advanced Technology in Engineering and Science
Vol. No. 10, Issue No. 06, June 2022
www.ijates.com
ijates
ISSN 2348 - 7550

by reducing the tendency to crowd points together in the center of the map. T-SNE is better than existing techniques at creating a single map that reveals structure at many different scales. This is particularly important for high-dimensional data that lie on several different, but related, low-dimensional manifolds, such as images of objects from multiple classes seen from multiple viewpoints. For visualizing the structure of very large data sets, we show how t-SNE can use random walks on neighborhood graphs to allow the implicit structure of all of the data to influence the way in which a subset of the data is displayed. We illustrate the performance of t-SNE on a wide variety of data sets and compare it with many other non-parametric visualization techniques, including Sammon mapping, Isomap, and Locally Linear Embedding. The visualizations produced by t-SNE are significantly better than those produced by the other techniques on almost all of the data sets. The visualization obtained by using T-SNE is shown in the below figure

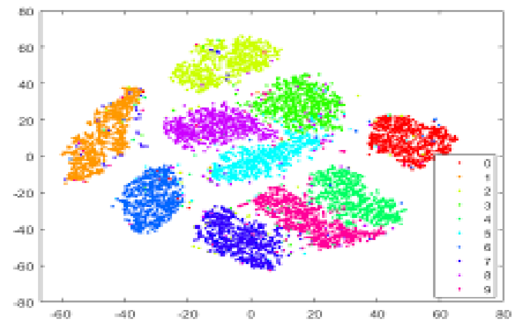


Figure 2: Visualization using T-SNE

3. AUC VS ROC CURVES

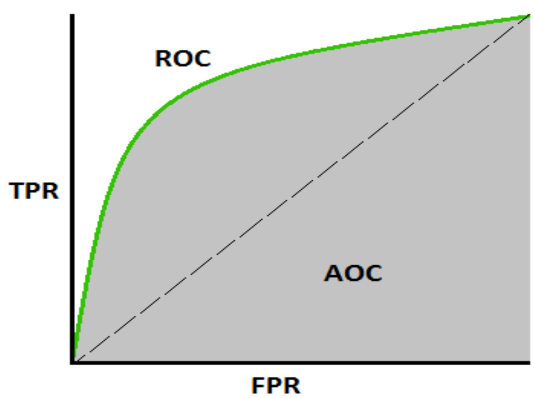
The area under the coverage-specificity curve (AUC) is regarded as the global evaluation criteria for the information granules constructed. AUC - ROC curve is a performance measurement for the classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. AUC measures the entire two-dimensional area underneath the entire ROC curve. AUC provides an aggregate measure of performance across all possible classification thresholds.

The Receiver Operator Characteristic (ROC) curve is an evaluation metric for binary classification problems. It is a probability curve that plots the TPR against FPR at various threshold values and essentially separates the 'signal'

International Journal of Advanced Technology in Engineering and Science
Vol. No. 10, Issue No. 06, June 2022
www.ijates.com
ijates
ISSN 2348 - 7550

from the 'noise'. The Area Under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve.

The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.



4. MEASUREMENT METRICES

The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. It is used to obtain accuracy, error rate, precision, recall, f-measure etc.

III EXPERIMENTAL RESULTS

1. CLUSTERING ANALYSIS

Here the dataset taken is divided in to training and testing data. The training data is always the same however, the testing data is sampled continuously at different sampling rate. Clusters formed by varying radius are given in the below figures.

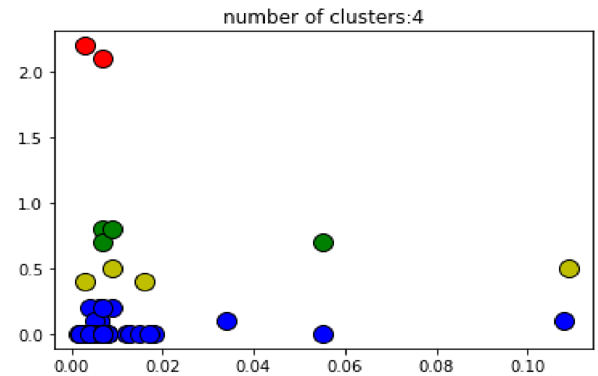


Figure 3: DBSCAN Clustering with ε=0.2

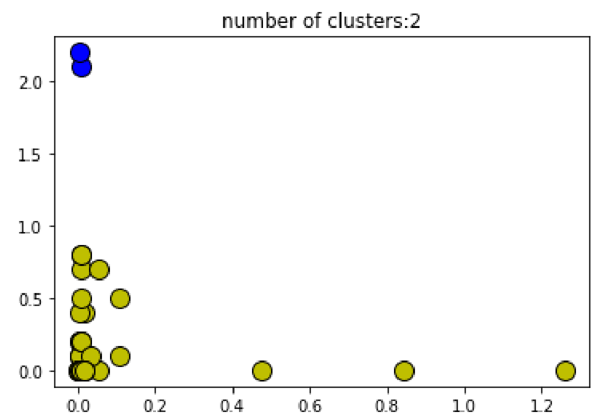


Figure 4: DBSCAN Clustering with ε=0.5

2. VISUALIZATION IN THE EMBEDDING SPACE

Here we input the testing data set into the pre-trained clusters and visualize them using the T-SNE Algorithm.

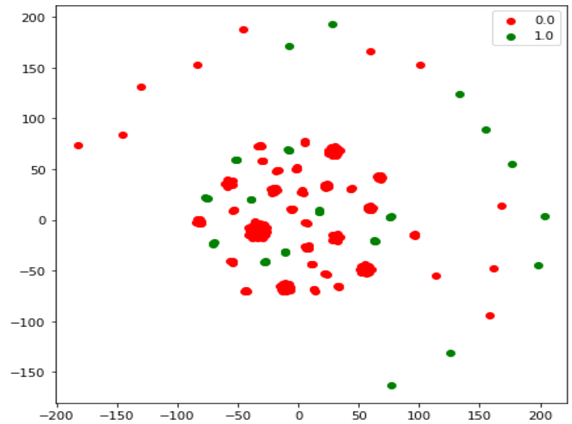


Figure 5: T-SNE Visualization of DBSCAN Clusters

3. AUC VS ROC CURVES

The area under coverage specificity curve obtained for the taken data set is obtained as shown in the figure below.

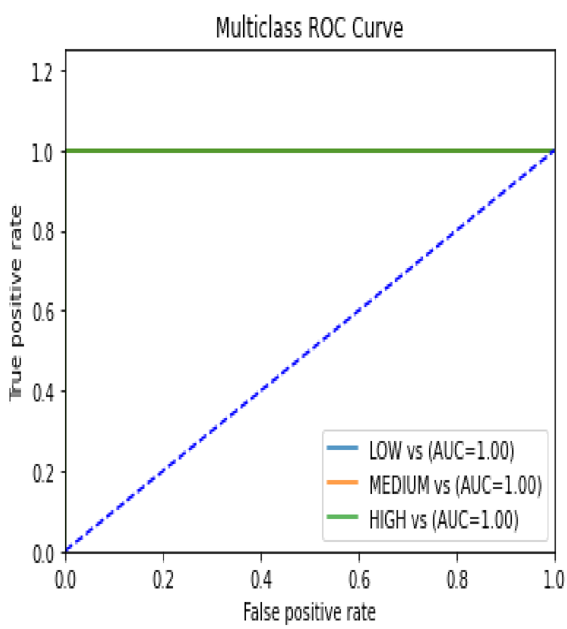


Figure 6: Coverage area with risk levels

4. MEASUREMENT METRICES

This parameter helps in giving the information about the areas which are highly accessible to risk and less accessible to risk.

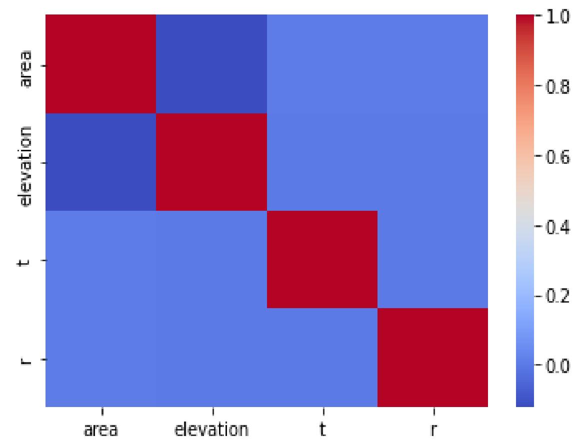
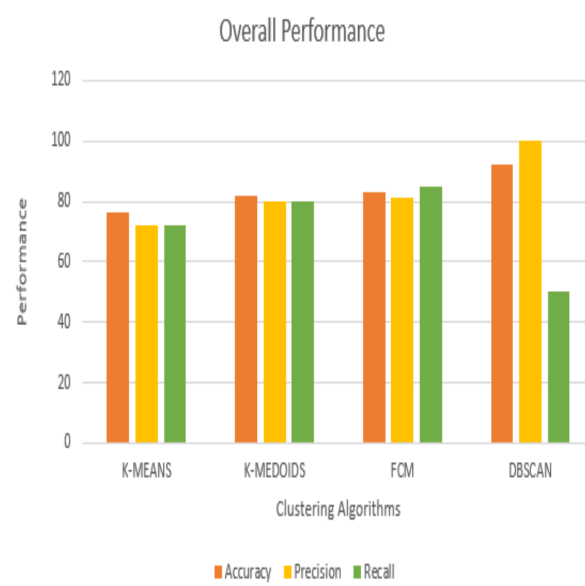


Figure 6: Figure representing areas prone to danger

5. COMPARITIVE ANALYSIS



IV CONCLUSION

This system proposed a novel granular computation approach to discover the fundamental structure within the landslide risk dataset. The DBSCAN firstly clustered the multiple landslide data points into sub-clusters for generating rule based granules. Graphs are also generated using t-SNE to project the landslide dataset in sub clusters into 2- dimensional embedding space for data visualization. Then, we construct information-granules by computing prototypes for each cluster and then set up decision boundaries by intervals.The rule-based information-granules are constructed to interpret the decision-making process for risk classification. We also optimize the granular structures by considering the coverage, specificity, and AUC and obtained the optimal granules for classifying landslide risks. Comparative analysis against benchmarking clustering algorithms has been performed. Computational results over the testing dataset demonstrates that the proposed approach outperforms the benchmark clustering algorithms with respect to the quality of constructed information granules.Risk maps of landslide in this project has been produced based on the obtained risk labels and the original landslide inventory map. In the future, in-detail analysis will be performed considering the scenarios that data points containing missing values in certain attributes. Thus, it will provide us more in-depth insights regarding the risk patterns within the landslide dataset.

REFERENCES

[1] J. Gasmo, H. Rahardjo, and E. Leong, "Infiltration effects on stability of a residual soil slope," Computers & Geotechnics, vol. 26, no. 2, pp. 145-165, 2000.

[2] J. Boyd, J. Chambers, P. Wilkinson, et al., “linked geomorphological and geophysical modelling methodology applied to an active landslide,” Landslides, to be published. DOI: 10.1007/s10346-021-01666-w

[3] D. Kirschbaum, T. Stanley, and S. Yatheendradas, "Modeling landslide susceptibility over large regions with fuzzy overlay," Landslides, vol. 13, no. 3, pp. 485-496, 2016.

[4] N. T. H. Van, P. Van Son, N. H. Khanh, & L. T. Binh, “Landslide susceptibility mapping by combining the analytical hierarchy process and weighted linear combination methods: a case study in the upper Lo River catchment

(Vietnam)," Landslides, vol. 13, no. 5, pp. 1285-1301, 2016.

[5] H. A. Nefeslioglu, E. A. Sezer, C. Gokceoglu, & Z. Ayas, "A modified analytical hierarchy process (M-AHP) approach for decision support systems in natural hazard assessments," Computers & Geosciences, vol. 59, pp. 1-8, 2013.

[6] B. Ahmed, "Landslide susceptibility mapping using multicriteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh," Landslides, vol. 12, no. 6, pp. 1077-1095, 2015.

[7] O. F. Althuwaynee, B. Pradhan, H. J. Park, & J. H. Lee, "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.

[8] F. Huang, C. Yao, W. Liu, Y. Li, & X. Liu, "Landslide susceptibility assessment in the Nantian area of China: a comparison of frequency ratio model and support vector machine. Geomatics," GEOMAT NAT HAZ, vol. 9, no. 1, pp. 919-938, 2018.

[9] A. M. Youssef, H. R. Pourghasemi, Z. S. Pourtaghi, & M. M. Al-Katheeri, "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.

[10] P. V. Gorsevski, M. K. Brown, K. Panter, C. M. Onasch, A. Simic, & J. Snyder, "Landslide detection and susceptibility mapping using LiDAR and an artificial neural network

[11] A. Platzer, "Visualization of SNPs with t-SNE," PloS one, vol. 8, no. 2, pp. e56883, 2013.

[12] L. Van der Maaten, & G. Hinton, "Visualizing data using t-SNE," Journal of machine learning research, vol. 9, no. 11, pp. 2579–2605, 2008.

[13] N. T. H. Van, P. Van Son, N. H. Khanh, & L. T. Binh, "Landslide susceptibility mapping by combining the analytical hierarchy process and weighted linear combination methods: a case study in the upper Lo River catchment (Vietnam)," Landslides, vol. 13, no. 5, pp. 1285-1301, 2016.