

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2022

Volume 18, Pages 86-101

**ICBASSET 2022: International Conference on Basic Sciences, Engineering and Technology**

## **Exploring the Risky Areas Due to Landslide Using Decision Tree Analysis: Case Study Tasmania, Australia**

**Mahyat SHAFAPOURTEHRANY**

Bogazici University

**Abstract:** Landslide is one of the natural hazards that considers a serious threat to both humans' lives and properties. Tasmania, Australia is one of those regions where landslides caused considerable damage to people and the State. Landslide damages can be reduced, even stopped, if proper land condition and planning assessment has been done. The main influencing parameters in landslide occurrences are topography, precipitation, and geological formation. Those parameters along with other influencing parameters have been used in landslide susceptibility mapping. In order to have a reliable analysis, a robust method of Decision Tree (DT) has been used to perform susceptibility mapping. According to the hierarchy structure of DT, geology and slope have been selected as the most influential parameters in landslide susceptibility. In order to evaluate the reliability of the outcomes, Area Under the Curve (AUC) has been utilized. Success and prediction rates were 87.64% and 84% respectively. Subsequently, risky features such as buildings, schools, hotels, etc. have been used in overlay analysis in a GIS environment with "very high" and "high" susceptibility classes. The outcome of this research can assist planning parties to secure vulnerable regions and consider those areas in their future decision-making strategies.

**Keywords:** Landslide, Decision Tree (DT), Vulnerability, GIS, Tasmania

### **Introduction**

Natural hazards are a great threat to both lives and properties worldwide (Jia et al., 2021). These include flooding, wildfire, earthquake, landslides, etc. Among natural disasters, landslides can cause serious losses directly or indirectly to the society and economy (Del Soldato et al., 2017). They are related to the characteristics of the slope of the terrain (Sim et al., 2022). One way to minimize the damage of landslides is to recognize the susceptible regions that are most likely to occur (Park et al., 2018). In the natural hazard scope, a susceptibility map recognizes regions that are more or less disposed to a potential landslide occurrence using low to high possibility values/classes (Ballabio & Sterlacchini, 2012). Susceptibility assessment has been implemented by several researchers to prevent or reduce actual damage from natural hazards. Until now, various methodological approaches including qualitative and quantitative methods have been applied for landslide susceptibility mapping (Reichenbach et al., 2018).

There is a variety of methods available in the literature to identify landslide susceptible areas such as frequency ratio (FR) (Lee & Sambath, 2006), logistic regression (LR) (Lee, 2005), index of entropy (IoE) (Pourghasemi, Mohammady, et al., 2012), the weight of evidence (WoE) (Lee & Choi, 2004), evidential belief function (EBF) (Althwaynee et al., 2012), statistical index (Regmi et al., 2014), analytical hierarchy process (AHP) (Kumar & Anbalagan, 2016), fuzzy logic (Pourghasemi, Pradhan, et al., 2012), fuzzy rule-based classifiers (Pham, Tien Bui, et al., 2016), neuro-fuzzy models multivariate adaptive regression splines (MARS) (Chu et al., 2019), random forest (RF) (Hong et al., 2019), (DT) (Park et al., 2018), Artificial neural network (ANN) (Lee et al., 2004), and support vector machines (SVM) (H. R. Pourghasemi et al., 2013), and etc. Each of the aforementioned methods has advantages and disadvantages (Reichenbach et al., 2018). Though, machine learning is a category of methods that are repeatedly used in natural hazard studies (Korup & Stolle, 2014).

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

© 2022 Published by ISRES Publishing: [www.isres.org](http://www.isres.org)

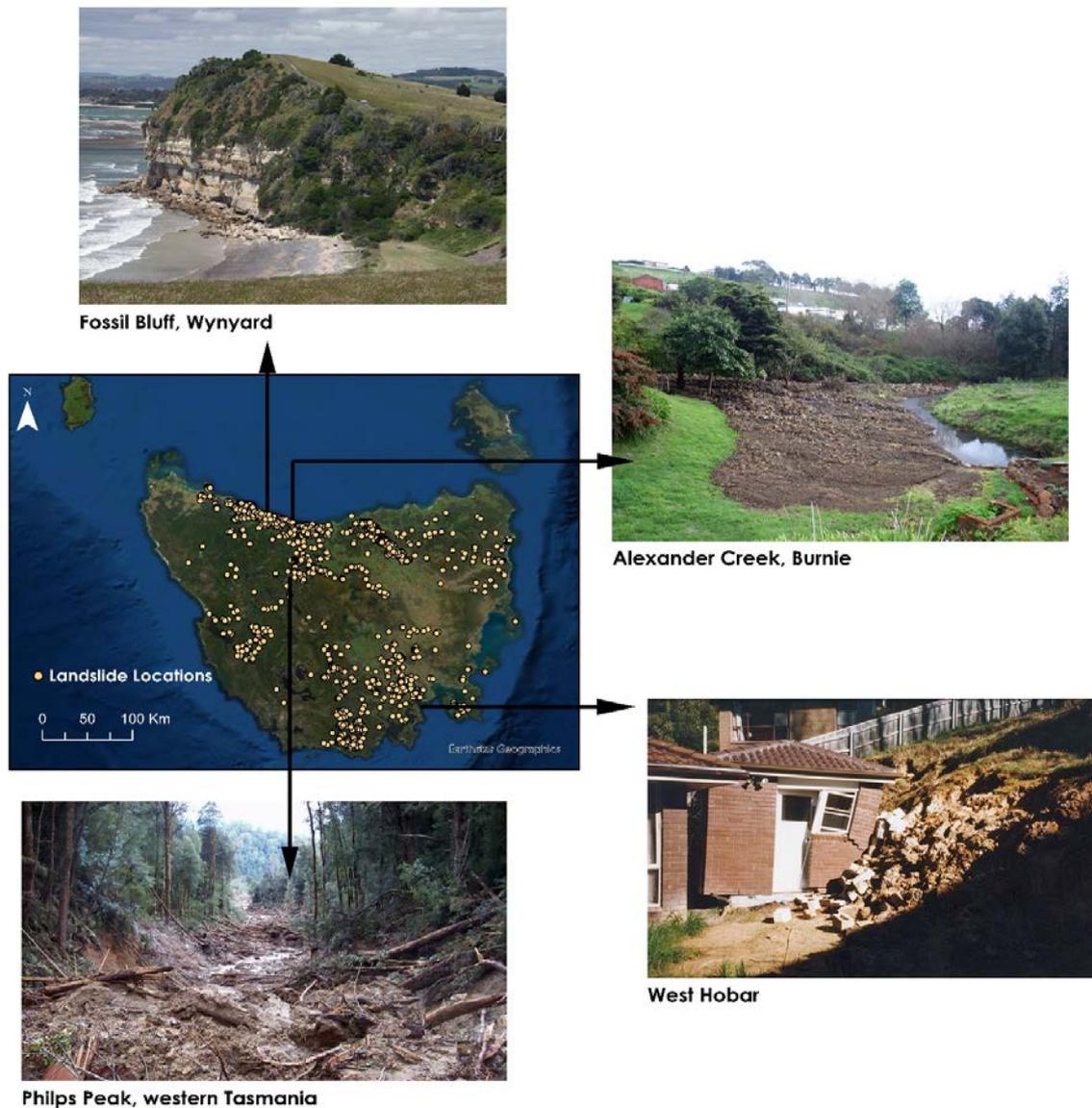


Figure 1. Distribution of landslides in Tasmania along with a few examples of the affected areas (BAY)

Machine learning is related to a set of methods that distinguish patterns in the dataset and use them to predict future scenarios (Jain et al., 2020). One of the differences between machine learning methods and other approaches is that without any pre-defined rules from experts, they can learn their own mappings between parametric rules directly from the data. This characteristic is advantageous in the case that the number of factors is considerably large and their physical properties are quite complex, as in the case of landslide (Merghadi et al., 2020). Hence, the machine learning approach to landslide response may assist to avoid many of the limitations of physics-based algorithms. In this study, we applied the machine learning method of DT using landslide data collected in the 1950s.

Australia has been affected by several landslides in the past (Roodposhti et al., 2019). Some lands in Tasmania are susceptible to slope instability (Slee & McIntosh, 2022). In some cases, the financial damage to people and the State reaches millions of dollars. However, there is no information source that people can use to get informed about landslide risk (Mazengarb, 2005). In addition, neither the insurance company nor the Tasmanian government provides coverage for those damages (Roodposhti et al., 2019). Without any doubt, such disasters and their destructive impacts can be avoided if the ground conditions are evaluated and assessed prior to the planning and construction (Dai et al., 2002).

In order to create safer and sustainable communities land use planning strategies are required (Chakraborty & Anbalagan, 2013; Sultana & Tan, 2021). Some of the planning actions are: to avoid development in areas that will increase the likelihood of landslide risk; to remove or modify structures in risky zones; to recognize high-

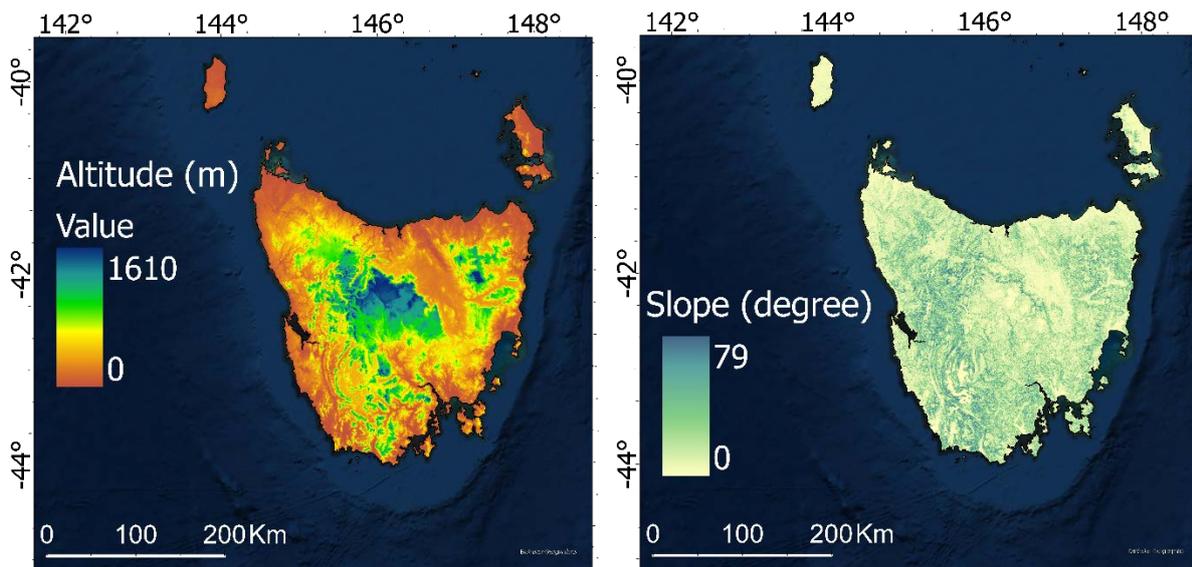
risk areas through zoning and overlay controls, and to avoid developments that are more likely to contribute to increased landslide risk.

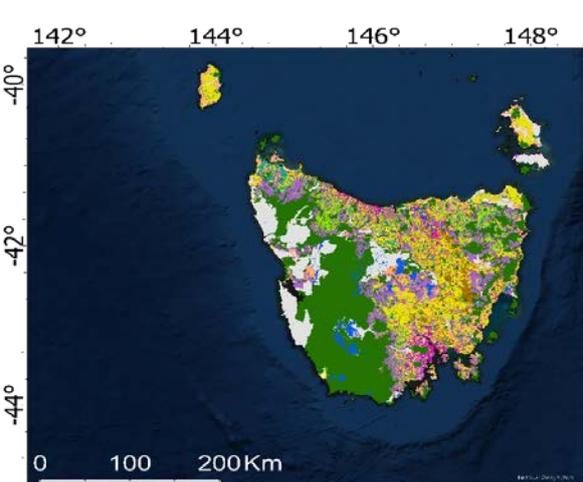
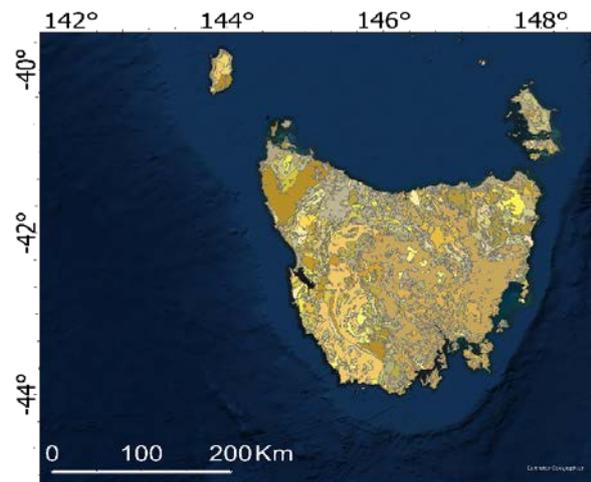
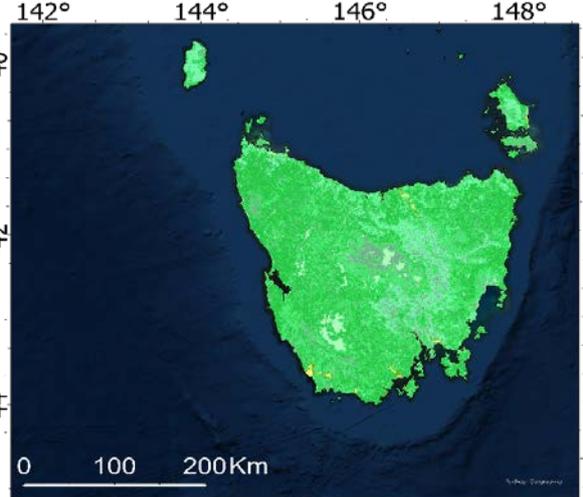
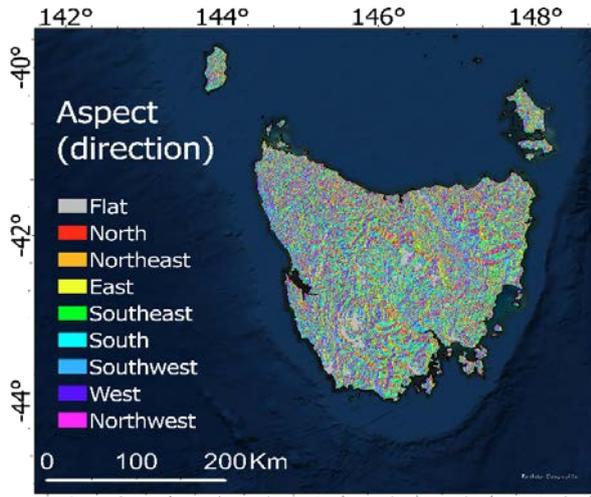
According to the above information, since landslide conditions in Tasmania received less attention in the literature, this study contributes to a better understanding of the landslide risky areas in Tasmania, Australia by overlaying the landslide susceptible areas with risky locations. Using this analysis to identify the risky areas can be a great tool for several end-users (private and public sectors), intended for landslide mitigation purposes at both local and international levels.

## Study Area and Data Used

Landslides are common in Tasmania due to the natural geological processes over long-time scales (Figure 1). Due to its varied topographical, geological, and geomorphic processes, including the effects of past and present climates, landslides occur in a variety of sizes and types (Mazengarb & Stevenson, 2010). Mineral Resources Tasmania (MRT) stated that due to the nature and magnitude of the known past landslides, similar events occurring today would have the potential for significant damage in the future (<https://www.mrt.tas.gov.au/home>). The five major landslide types (slide, flow, fall, topple and spread) all can happen in Tasmania, with slides (both shallow and deepseated) and flows being the most common types. A number of landslides are a combination of movement types. The material that moves is classed as rock, earth, or debris.

In order to have efficient mitigation strategies, defining the landslide conditioning factors must be also highlighted throughout the spatial domain (Roodposhti et al., 2019). A relevant and reliable landslide conditioning factors dataset with accurate landslide inventories is essential for landslide susceptibility analysis. Both datasets' precision has a direct impact on the final outcomes (Meena et al., 2022). For the current research, the conditioning factors are altitude, slope, aspect, forest, geology, landuse/cover (LULC), Normalized Difference Vegetation Index (NDVI), rain, distance to the river, distance to road, and soil (Figure 2). Each factor was resized to a  $30 \times 30$  m grid. Each one of these factors has an impact on landslide occurrence in a region.





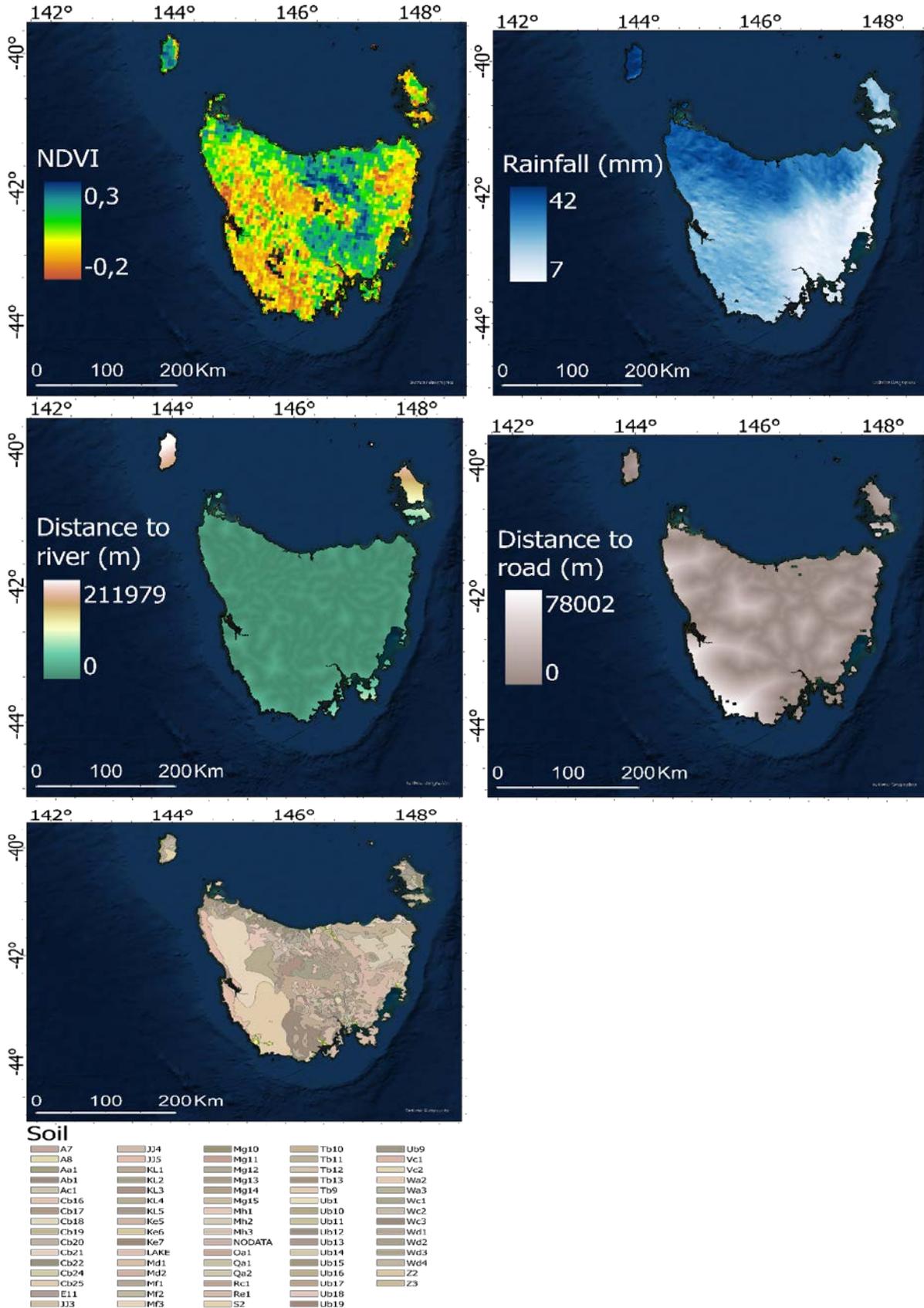


Figure 2. Landslide conditioning factors

Regarding the risky features, the Tasmania government website provided a set of factors such as ambulance, buildings, children and elderly population density, cultural places, educational places, entertainment places, governmental places, heritage Tasmania features, hotels, kindergarten, medical services, police stations, primary

infant schools, private reserves, and tourist features. The aforementioned dataset has been used in risk analysis which will be described in the following sections.

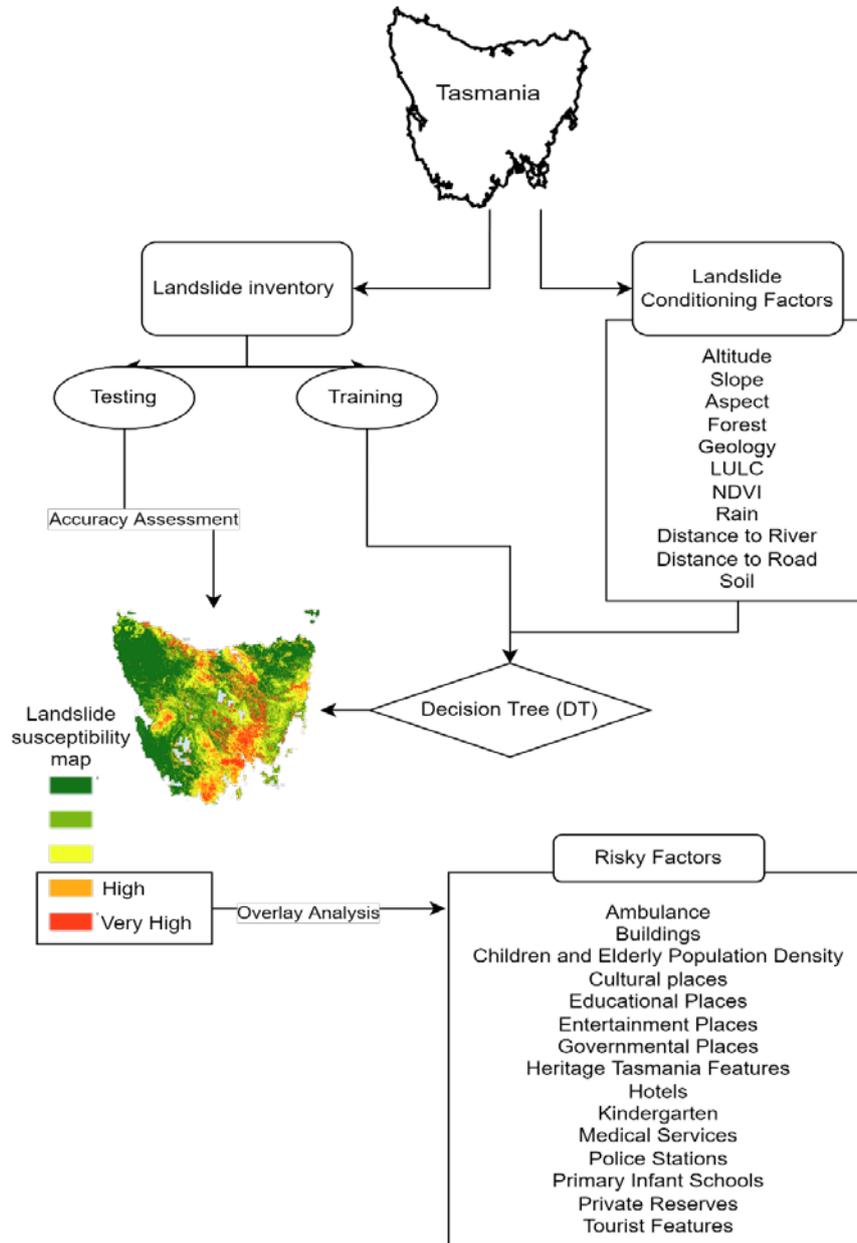


Figure 3. Flowchart

## Methodology

Analyzing the correlation between “landslide conditioning factors” and “landslide inventory” is the first stage in landslide susceptibility analysis. Therefore, according to the flowchart (Figure 3), both datasets of conditioning and inventory have been produced. As has been mentioned landslide database has been received from MRT which includes the landslide locations since the 1950s. A total of 600 landslides were randomly selected; 70% of them were used as training data and 30% as validation data. The same number of non-landslide points were randomly selected from the free-landslide regions. Subsequently, values of 1 and 0 were assigned to landslide pixels and non-landslide pixels respectively. In order to perform DT, all dataset needs to be converted into ASCII (American Standard Code for Information Interchange) format. ArcGIS was used to convert all of the input data into ASCII data. Then, landslide training data and landslide conditioning factors were analyzed while using the DT method in the SPSS environment to calculate the landslide susceptibility index and build a

andslide susceptibility map in the study area. Classes with the highest susceptibility were overlaid on risky features derived from different sources.

### Decision Tree (DT)

DT is one of the machine learning methods that consider a predictive modeling technique. It can precisely identify the structural patterns in data and represent them graphically as tree structures. Another advantage of DT is that pre-defined correlation between the variables is not required (Saito et al., 2009). DT can explain data and use it predictively, it can also ingest data calculated at varying scales, purely on the non-linear relationship without the need for assumptions based on frequency distributions (Kheir et al., 2010). Consequently, each factor can be considered input into the model. DT has been used as a rule-based technique in numerous natural hazard studies (Chen et al., 2017; Jaafari et al., 2018; Tehrany et al., 2013). DT groups the landslide conditioning factors hierarchically according to levels of susceptibility. The aim is to create a set of decision rules that can form the basis for predicting the outcome from the input dataset. Thus, the rules are generated by the analysis of a set of conditioning factors, with the purpose of predicting an outcome from a similar set of variable factors. Due to the proficiency of DT, this method was used to analyze the correlation between the landslide conditioning factors and the training dataset. Using DT-CHAID the conditioning factor that is selected at each step is the one representing the strongest relationship with the landslide occurrence susceptibility.

### Accuracy Assessment

AUC is a popular, comprehensive quantitative method of accuracy assessment, by which the prediction and success rates may be evaluated (Pham, Pradhan, et al., 2016). The proficiency of AUC in evaluating the susceptibility mapping outcomes has been demonstrated successfully in a number of studies (H. Pourghasemi et al., 2013; Pourtaghi et al., 2016; Xu et al., 2012). AUC evaluates the existence of the known landslide inventory data and the acquired probability map. It starts by dividing the landslide probability map into categories of equal area, and hierarchically ranking respective values from minimum to maximum in ArcGIS. AUC starts by sorting the calculated values of all cells into descending order, thus ranking each prediction hierarchically. Thereafter, the values of cells were divided into 100 classes with 1% accumulation intervals. The presence of landslide in each interval is measured using the “Tabulate area” tool in ArcGIS as the next step. The success and prediction curves determine the percentage of landslides in each probability category.

### Results and Discussion

DT has been applied and a landslide susceptibility map has been created by classifying the derived map into five classes of “low”, “very low”, “moderate”, “high”, and “very high” (Figure 4) using a well-known quantile classification technique (Baeza et al., 2016). Based on the DT tree structure, the most influential landslide conditioning factors affecting the landslide distribution were geology and slope, as they were selected at the top of the branch. The most susceptible combination of conditioning factors contributing to the high landslide susceptibility was as follow: starting with slope (between 32-58°), as well as the decision root continued by the rainfall greater than 10mm, and NDVI <-0.1. Areas exhibiting these characteristics were classified as 100% susceptible to landslides (Table 1).

Table 1. The most susceptible combination of conditioning factors

Factors	Range	DT Rules
Slope	0-72°	32-58°
Rainfall	6.6-42.4mm	>10mm
NDVI	0-67°	<-0.1

The landslide susceptibility map is presented in Figure 4. It depicted where the slope failure may occur spatially in the future.

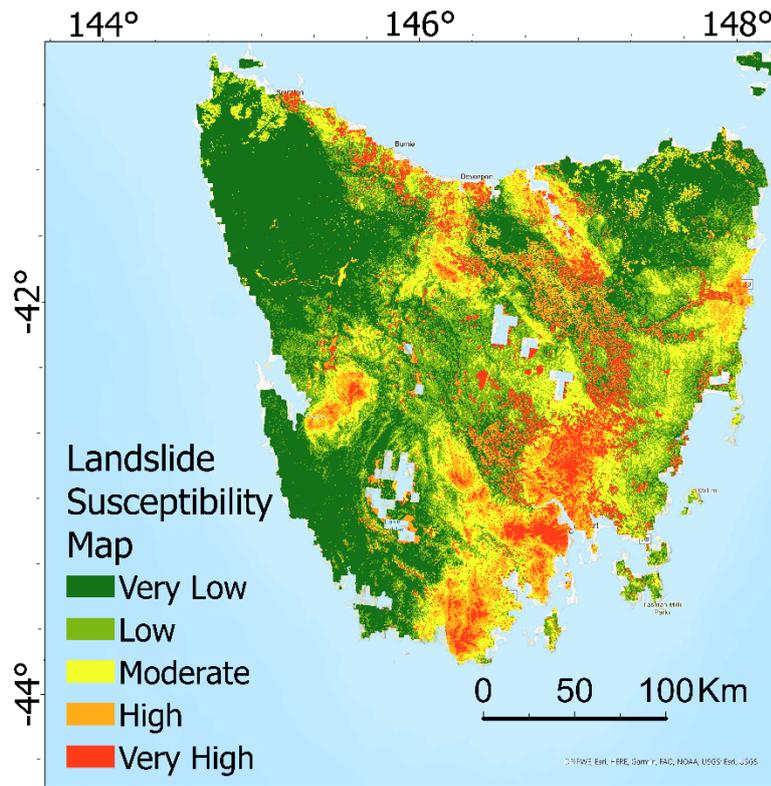
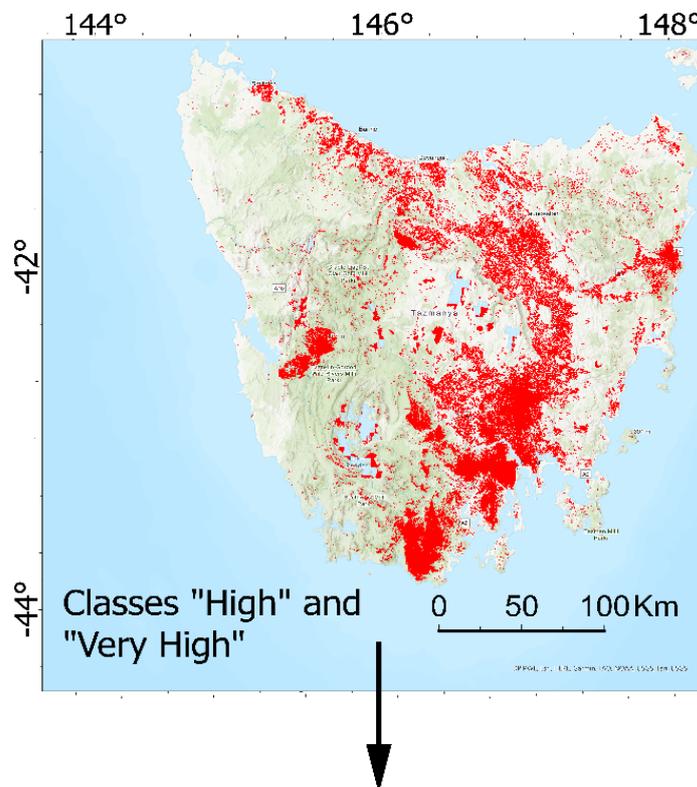
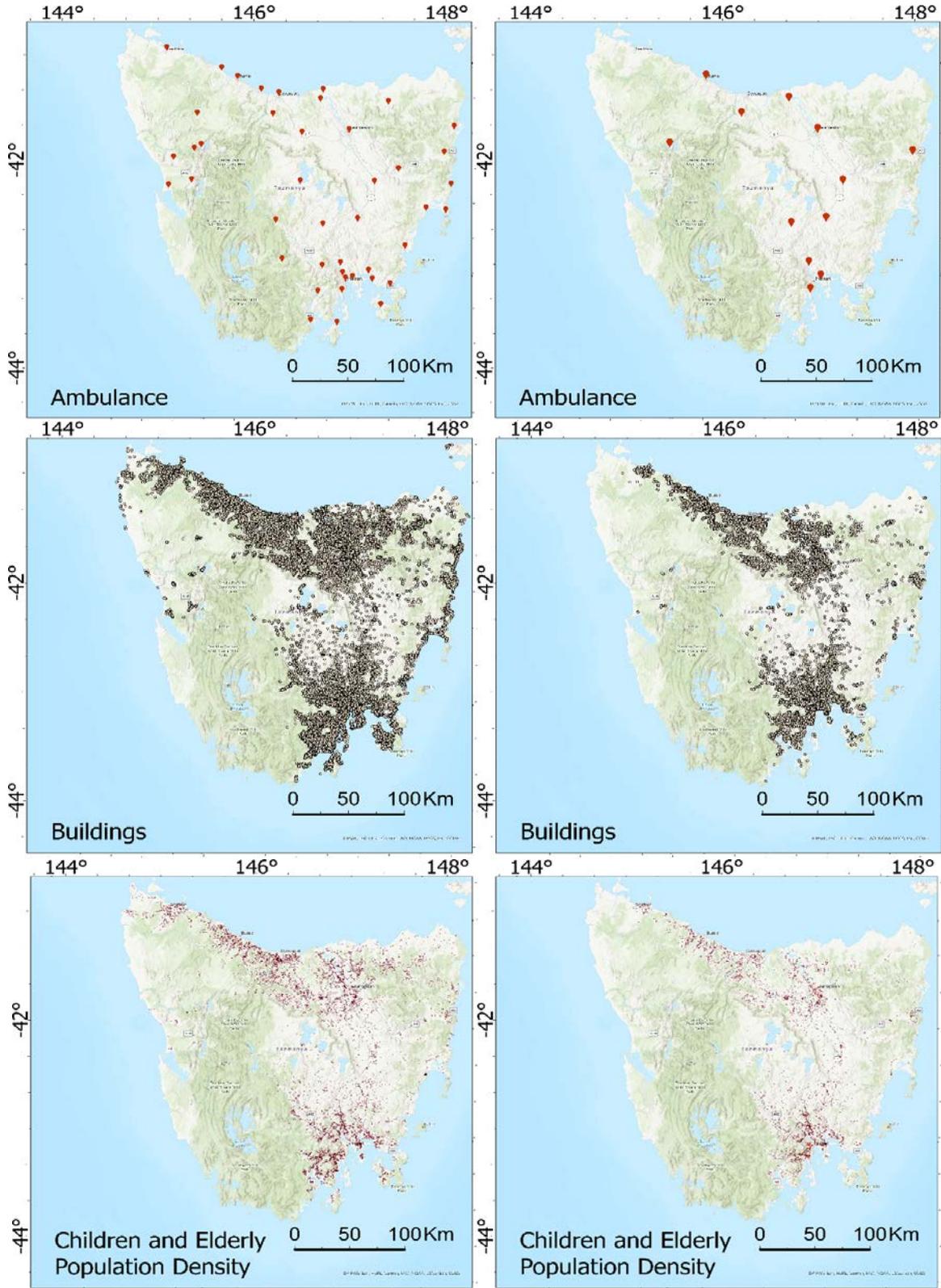
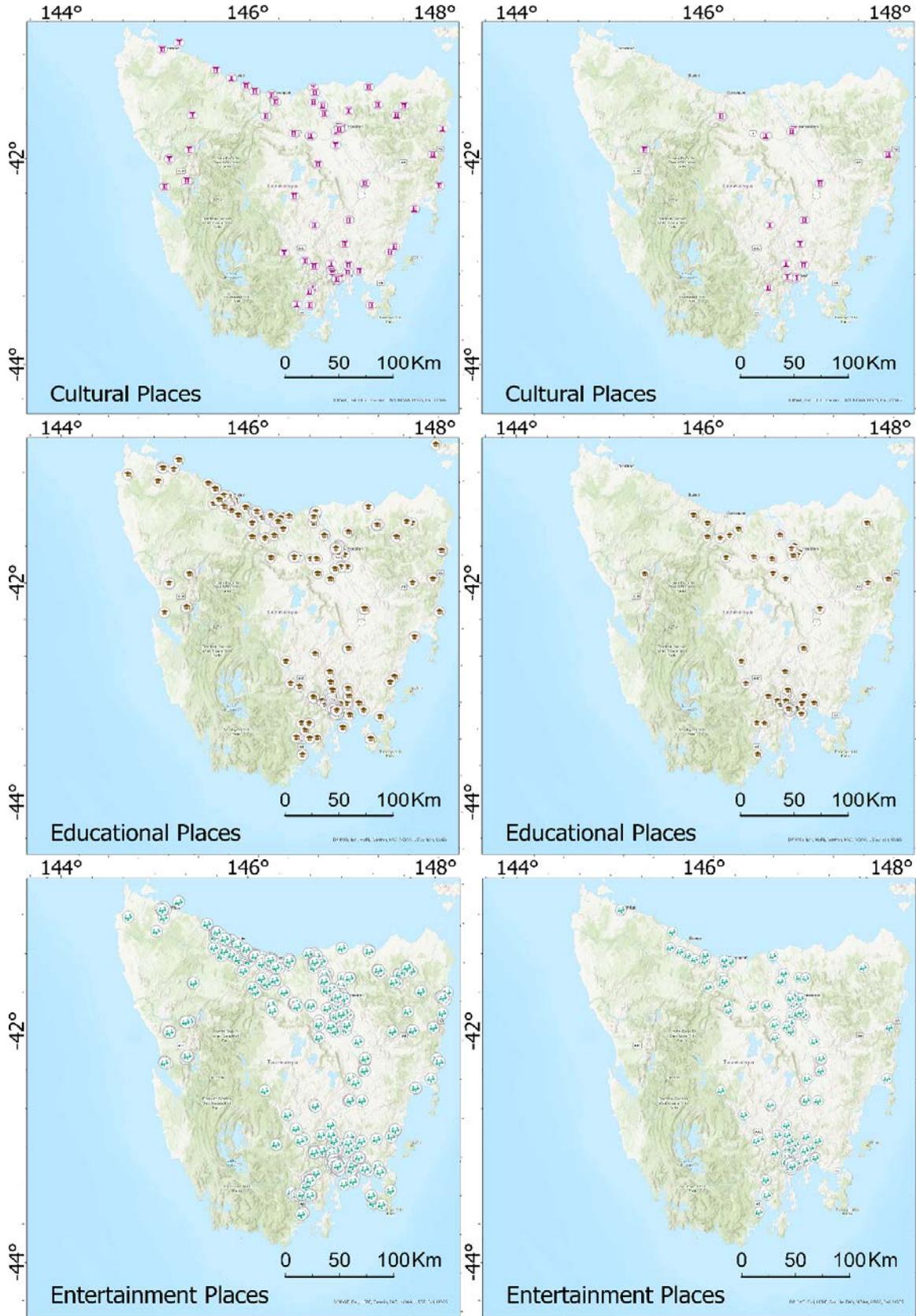
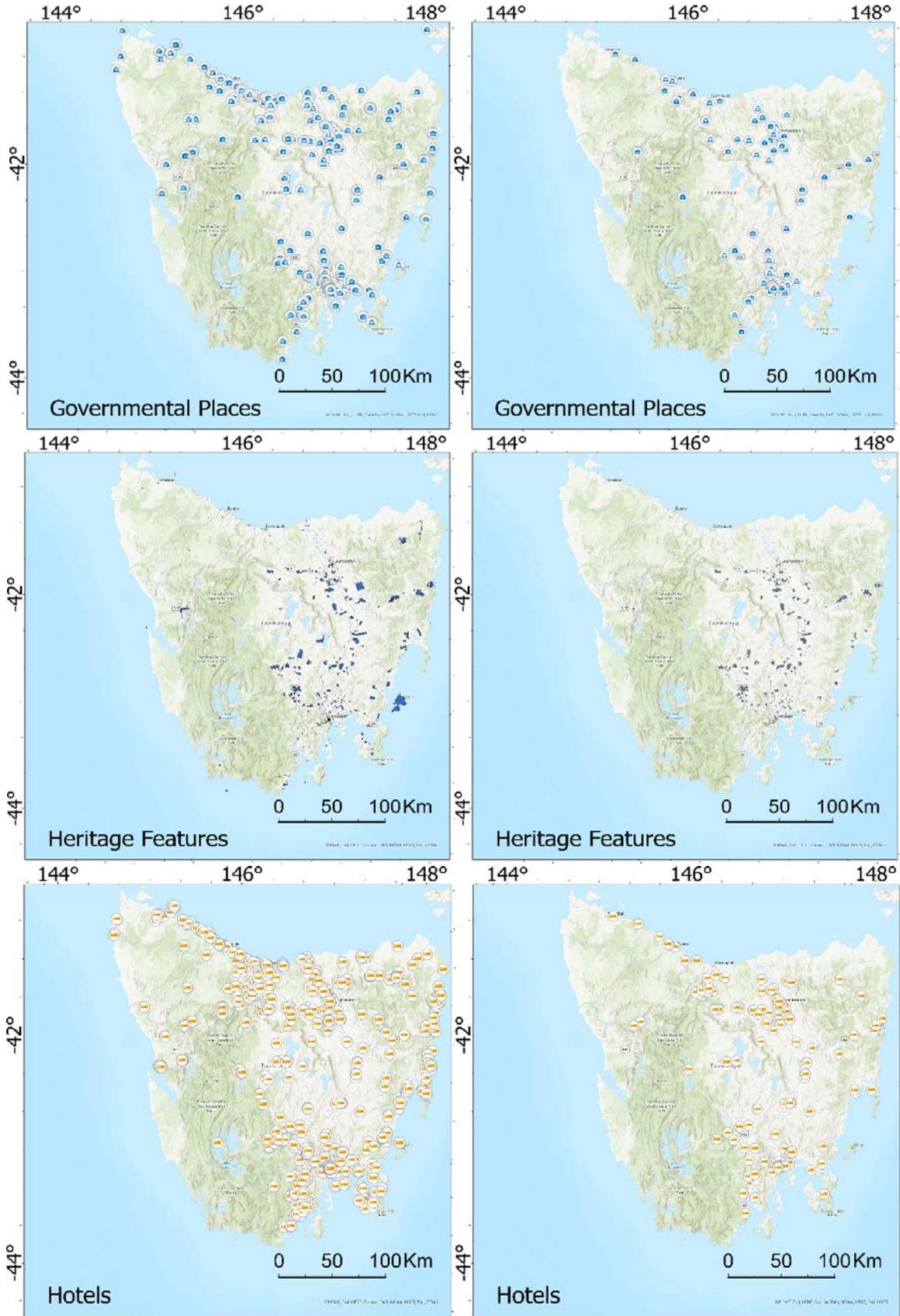


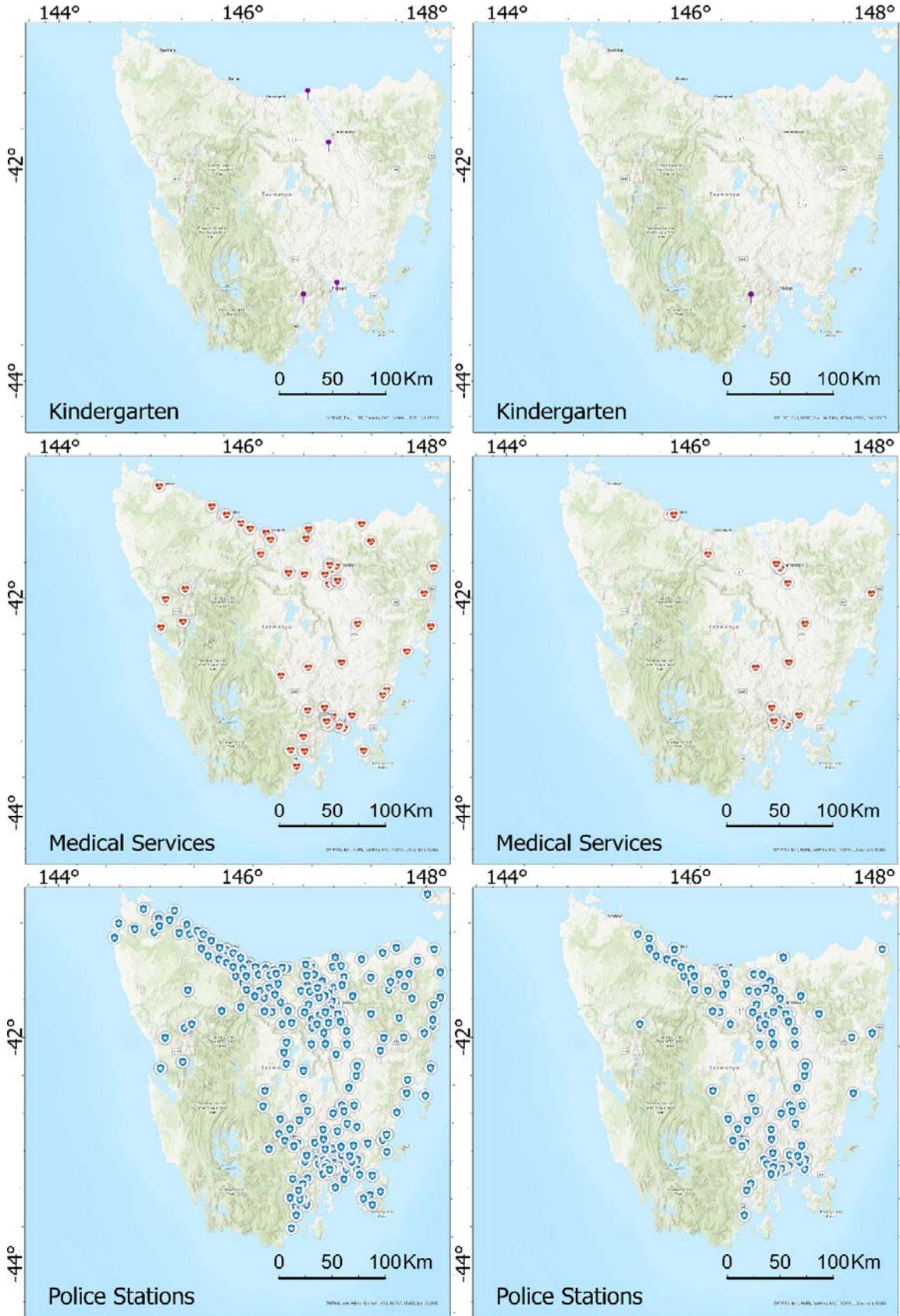
Figure 4. Landslide susceptibility map











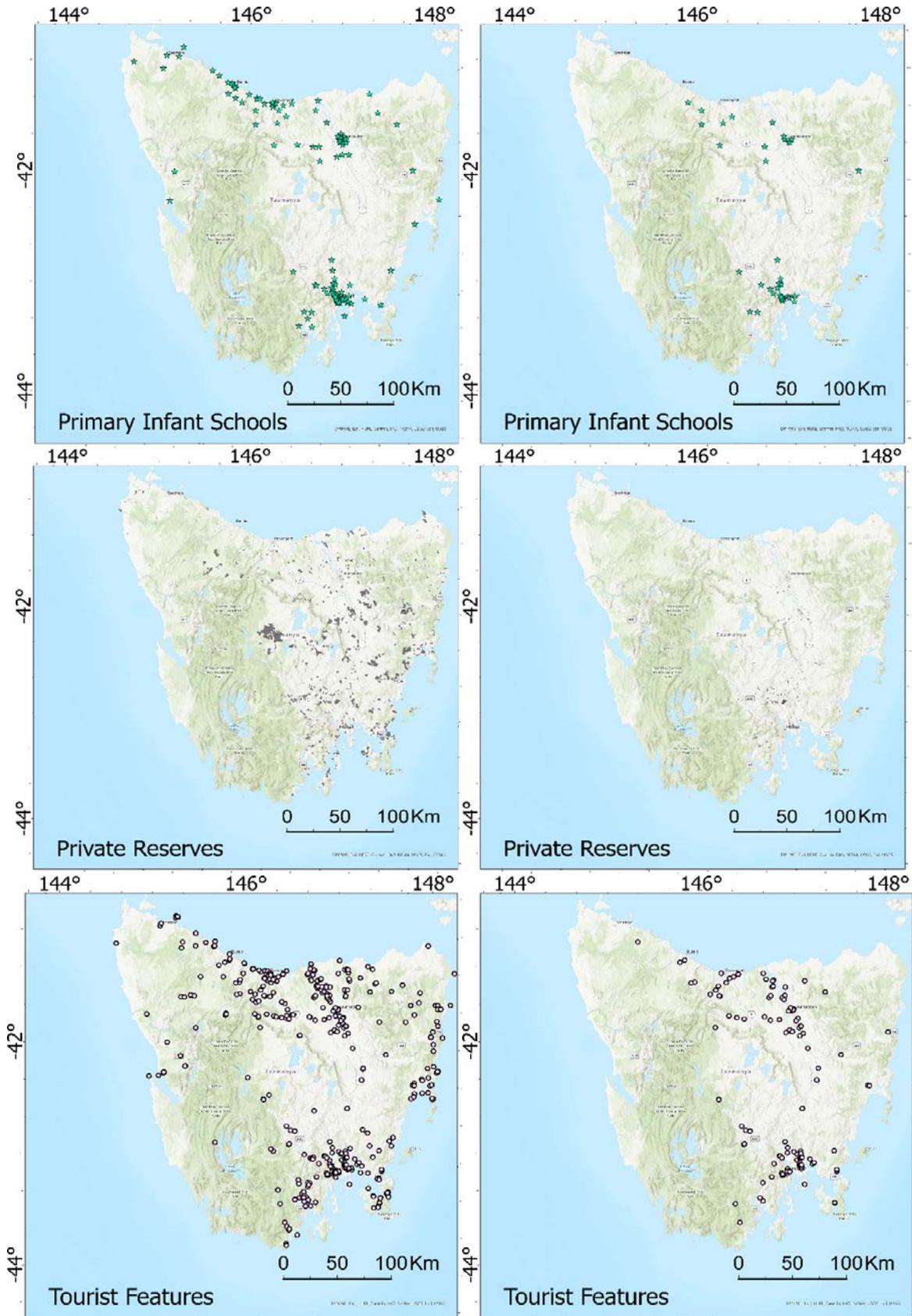


Figure 5. Risky features

As it can be seen in Figure 4, the majority of the susceptible areas are located southeast, center, and north regions. AUC has been applied and success and prediction rates were 87.64% and 84% respectively. In order to evaluate which features are at risk from these susceptible areas, overlay analysis has been implemented. The

classes of “very high” and “high” have been used to be overlaid with ambulances, buildings, children and elderly population density, cultural places, educational places, entertainment places, governmental places, heritage Tasmania feature, hotels, kindergarten, medical services, police stations, primary infant schools, private reserves, and tourist features. Figure 5 represented the features that fall into the highest susceptible zones.

According to Figure 5, 142 touristic places; 39 primary and infant schools; 122 police stations; 45 medical services; Huonville Primary School Kindergarten; 210 hotels; 217 governmental places; 94 education places; 30 cultural places; 30% buildings; 12 ambulances; 186 entertainment places; 744 private reserves; and 2195 heritage features are located in highest susceptible zones. The derived output can considerably assist different sectors such as governments, insurance, tourism, and etc.

## **Conclusion**

All societies are affected by natural hazards in one form or another. Whilst landslides do not present the most significant danger in Tasmania, the cost to the community over time from economic and social perspectives is considerable. This study aimed to evaluate the existence of risky features in high landslide susceptibility zones in Tasmania. DT as one of the robust machine learning methods has been used to evaluate the correlations among landslide conditioning factors and landslide inventory. Susceptible zones were detected and risky features such as hotels, children and elderly population, etc. have been overlaid with the produced map. Through this analysis, some risky features have been detected in classes of “very high” and “high” susceptibility. For instance, among all kindergartens in Tasmania, Huonville Primary School Kindergarten has been located in risky zones. From 291,698 buildings in Tasmania, 82,736 buildings have been located in risky zones. Such information can significantly be helpful in future management plans. It can be used as a basis for more detailed and deep risk assessments in Tasmania.

## **Scientific Ethics Declaration**

The author declares that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the author.

## **References**

- Althuwaynee, O. F., Pradhan, B., & Lee, S. (2012). Application of an evidential belief function model in landslide susceptibility mapping. *Computers & Geosciences*, *44*, 120-135.
- Baeza, C., Lantada, N., & Amorim, S. (2016). Statistical and spatial analysis of landslide susceptibility maps with different classification systems. *Environmental Earth Sciences*, *75*(19), 1-17.
- Ballabio, C., & Sterlacchini, S. (2012). Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. *Mathematical geosciences*, *44*(1), 47-70. BAY, E. Landslide Map Series.
- Chakraborty, D., & Anbalagan, R. (2013). Landslide hazard analysis and management: A case study from Nainital, India. In *Earth System Processes and Disaster Management* (pp. 141-154). Springer.
- Chen, W., Xie, X., Peng, J., Wang, J., Duan, Z., & Hong, H. (2017). GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. *Geomatics, Natural Hazards and Risk*, *8*(2), 950-973.
- Chu, L., Wang, L.-J., Jiang, J., Liu, X., Sawada, K., & Zhang, J. (2019). Comparison of landslide susceptibility maps using random forest and multivariate adaptive regression spline models in combination with catchment map units. *Geosciences Journal*, *23*(2), 341-355.
- Dai, F., Lee, C. F., & Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Engineering Geology*, *64*(1), 65-87.
- Del Soldato, M., Bianchini, S., Calcaterra, D., De Vita, P., Martire, D. D., Tomás, R., & Casagli, N. (2017). A new approach for landslide-induced damage assessment. *Geomatics, Natural Hazards and Risk*, *8*(2), 1524-1537.
- Hong, H., Miao, Y., Liu, J., & Zhu, A.-X. (2019). Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *Catena*, *176*, 45-64.
- Jaafari, A., Zenner, E. K., & Pham, B. T. (2018). Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecological informatics*, *43*, 200-211.

- Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4), 478-505.
- Jia, G., Alvioli, M., Gariano, S. L., Marchesini, I., Guzzetti, F., & Tang, Q. (2021). A global landslide non-susceptibility map. *Geomorphology*, 389, 107804.
- Kheir, R. B., Greve, M. H., Abdallah, C., & Dalgaard, T. (2010). Spatial soil zinc content distribution from terrain parameters: A GIS-based decision-tree model in Lebanon. *Environmental Pollution*, 158(2), 520-528.
- Korup, O. & Stolle, A. (2014). Landslide prediction from machine learning. *Geology Today*, 30(1), 26-33.
- Kumar, R. & Anbalagan, R. (2016). Landslide susceptibility mapping using analytical hierarchy process (AHP) in Tehri reservoir rim region, Uttarakhand. *Journal of the Geological Society of India*, 87(3), 271-286.
- Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing*, 26(7), 1477-1491.
- Lee, S. & Choi, J. (2004). Landslide susceptibility mapping using GIS and the weight-of-evidence model. *International Journal of Geographical Information Science*, 18(8), 789-814.
- Lee, S., Ryu, J.-H., Won, J.-S. & Park, H.-J. (2004). Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Engineering Geology*, 71(3-4), 289-302.
- Lee, S. & Sambath, T. (2006). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology*, 50(6), 847-855.
- Mazengarb, C. (2005). *The Tasmanian landslide hazard map series: methodology*. Department of Infrastructure, Energy and Resources Hobart, Australia.
- Mazengarb, C. & Stevenson, M. (2010). *Tasmanian landslide map series: user guide and technical methodology*. Tasmanian Geological Survey.
- Meena, S. R., Puliero, S., Bhuyan, K., Floris, M. & Catani, F. (2022). Assessing the importance of conditioning factor selection in landslide susceptibility for the province of Belluno (region of Veneto, northeastern Italy). *Natural hazards and earth system sciences*, 22(4), 1395-1417.
- Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R. & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225.
- Park, S.-J., Lee, C.-W., Lee, S. & Lee, M.-J. (2018). Landslide susceptibility mapping and comparison using decision tree models: A Case Study of Jumunjin Area, Korea. *Remote Sensing*, 10(10), 1545.
- Pham, B. T., Pradhan, B., Bui, D. T., Prakash, I. & Dholakia, M. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environmental Modelling & Software*, 84, 240-250.
- Pham, B. T., Tien Bui, D., Prakash, I. & Dholakia, M. (2016). Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. *Natural Hazards*, 83(1), 97-127.
- Pourghasemi, H., Moradi, H. & Fatemi Aghda, S. (2013). Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural Hazards*, 69(1), 749-779.
- Pourghasemi, H. R., Jirandeh, A. G., Pradhan, B., Xu, C. & Gokceoglu, C. (2013). Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *Journal of Earth System Science*, 122(2), 349-369.
- Pourghasemi, H. R., Mohammady, M. & Pradhan, B. (2012). Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena*, 97, 71-84.
- Pourghasemi, H. R., Pradhan, B. & Gokceoglu, C. (2012). Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Natural hazards*, 63(2), 965-996.
- Pourtaghi, Z. S., Pourghasemi, H. R., Aretano, R. & Semeraro, T. (2016). Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecological Indicators*, 64, 72-84.
- Regmi, A. D., Devkota, K. C., Yoshida, K., Pradhan, B., Pourghasemi, H. R., Kumamoto, T. & Akgun, A. (2014). Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal of Geosciences*, 7(2), 725-742.
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M. & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60-91.
- Roodposhti, M. S., Aryal, J. & Pradhan, B. (2019). A novel rule-based approach in mapping landslide susceptibility. *Sensors*, 19(10), 2274.
- Saito, H., Nakayama, D. & Matsuyama, H. (2009). Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. *Geomorphology*, 109(3-4), 108-121.

- Sim, K. B., Lee, M. L. & Wong, S. Y. (2022). A review of landslide acceptable risk and tolerable risk. *Geoenvironmental Disasters*, 9(1), 1-17.
- Slee, A. & McIntosh, P. D. (2022). History of slope instability in the Oldina plantation, Tasmania. *New Zealand Journal of Forestry Science*, 52.
- Sultana, N. & Tan, S. (2021). Landslide mitigation strategies in southeast Bangladesh: lessons learned from the institutional responses. *International Journal of Disaster Risk Reduction*, 62, 102402.
- Tehrany, M. S., Pradhan, B. & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of hydrology*, 504, 69-79.
- Xu, C., Xu, X., Dai, F. & Saraf, A. K. (2012). Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. *Computers & Geosciences*, 46, 317-329.

---

### **Author Information**

---

**Mahyat Shafapourtehrany**

Kandilli Observatory and Earthquake Research Institute,  
Department of Geodesy, Bogazici University, 34680  
Cengelkoy, Istanbul, Turkey  
Contact e-mail: [mahyat.shafapour@boun.edu.tr](mailto:mahyat.shafapour@boun.edu.tr)

---

**To cite this article:**

Shafapourtehrany, M. (2022). Exploring the risky areas due to landslide using decision tree analysis: case study. Tasmania, Australia The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 18, 86-101.