Proc. of the 9th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Zhongqiang Liu, Jian Dai and Kate Robinson
Published by Research Publishing, Singapore. ©2025 The Author(s), CC-BY 4.0 licence.
doi: 10.3850/GRF-25280825_isgsr-068-P089-cd

SHAP-Augmented Neural Networks for Landslide Susceptibility Mapping in Darjeeling-Gangtok Region

Manohara K N

Department of Civil Engineering, Indian Institute of Technology, Guwahati, India. E-mail: n.manohara@iitg.ac.in

Rishikesh Bharti

Department of Civil Engineering, Indian Institute of Technology, Guwahati, India. E-mail: rbharti@iitg.ac.in

Arindam Dey

Department of Civil Engineering, Indian Institute of Technology, Guwahati, India. E-mail: arindam.dey@iitg.ac.in

Abstract: The Darjeeling-Sikkim Himalayas have been perennially affected by landslide hazards owing to their fragile terrain and complex geological and geotechnical settings. The region is of economic and strategic importance and the loss of lives and damage to property due to landslides in this area has been significant. Landslide Susceptibility Mapping (LSM) plays a critical role in evaluating the risks and provides valuable insights towards assessing the vulnerability of the region and encompassed infrastructure and settlements exposed to the hazard. Artificial Neural Networks (ANNs) are data-driven models known for their ability to universally approximate non-linear functions with complex correlations. This study aims at the comparison of the efficacy of multi-layered ANNs of different depths in the spatial analysis of classifying the region between the cities of Darjeeling and Gangtok based on their landslide susceptibility. The ANN models with one, two and five hidden layers were compared to understand the optimal complexity required for accurate and reliable landslide susceptibility analysis for the study area chosen. The SHapley Additive exPlanations (SHAP) are adopted to bring more interpretability to the model, thereby veering away from the black-box nature of the machine learning (ML) models. The optimization of the depth of ANN models revealed that 2 hidden layers were sufficient to successfully capture the complex relationships between the input parameters. Valuable insights about the possible triggers for the landslides was obtained by SHAP analysis which can be particularly helpful in further analyzing the landslide phenomenon in the study area. This study takes a significant step forward in demonstrating how advanced models (in this case, any ML models) can be made more interpretable and credible, thereby promoting informed decision-making and effective landslide risk management by taking the example of Darjeeling-Sikkim Himalayas.

Keywords: Landslide Susceptibility Mapping; Multi-Layered ANNs; Explanative Machine Learning; SHapley Additive exPlanations (SHAP).

1. Introduction

Landslides are one of the most common global geohazards that results in substantial damage to lives and properties. The rugged terrain, tectonical activeness, heavy monsoon and the fragile geology makes the North-Eastern part of India a major hotspot for the landslides. The Geological Survey on India (GSI) has reportedly recorded 592 landslides in the northeastern region since 2017 with almost 196 landslides from April to July in 2024. The limited road network, dictated by the challenging topography, exacerbates the impact of landslides. Any disruption to these vital transport links can severely affect the livelihoods of people dependent on them. Thus, the necessity of adopting comprehensive disaster management strategies becomes paramount. Landslide susceptibility mapping (LSM) serves as a first step in this regard.

Meta-heuristic determination of the spatial probability of landslide occurrences, using the knowledge of the previous landslide incidences, is a well-established approach. Researchers have long used various statistical and machine learning (ML) models for this purpose. The machine learning (ML) models like the decision trees and Artificial Neural Networks (ANN) are preferred due to their better predictive performance and their ability to understand complex interactions between different input features (Kawabata and Bandibas, 2009; Pradhan, 2013). However, the black-box nature of the ML models makes it difficult for the analyst to explain the predictions and feature interactions and thereby have their reservations. The use of interpretation methods like SHapley Additive exPlanations (SHAP) is gaining traction in recent times as this method aims to reduce the obscurity of the ML models (Alqadhi *et al*., 2024). The global and local interpretability provided by the SHAP analyses provides the much-needed interpretability to the ML models that will further encourage the analysts to use complex models and evaluate the influence of more parameters on the landslide susceptibility. Further, better comprehension would be added to understanding the intricacies in the parameters interacting with each other and their eventual contribution towards landsliding.

This study aims to develop a landslide susceptibility map for the region between Darjeeling and Gangtok in northeastern India. Using a feedforward ANN model and SHAP interpretations, the study aims to understand the influence of various input parameters, their statistical significance and explore their intricate interactions to better address landslide risks in this vulnerable region. Furthermore, the study explores the impact of ANN depth,

thereby evaluating the influence of varying numbers of hidden layers on the model performance while achieving an optimal balance between predictive accuracy and computational efficiency.

2. Study Area

The spatial susceptibility to landslides of the Darjeeling-Gangtok region in the North-Eastern part of India is explored in this study. The presence of the seismically active Main Central Thrust (MCT) has significantly influenced the geomorphology of the area contributing towards the formation of steep slopes and deep valleys, thereby enhancing the susceptibility towards landsliding. River Teesta, flowing through the region, appends the morphological and hydrological woes to landslides. With a subtropical to temperate climate that depends on the elevation, this region is one of the wettest places in India with annual rainfall in excess of 3000 mm. The region is of commercial and strategic importance with many tourist sites and connectivity to international borders. All these reasons, combined, have made this region a hotspot for catastrophic effects of landslides that compromises the day-to-day activities and commerce in the region. Hence, understanding the landslide susceptibility of different areas in the region becomes a primary requirement in planning for the disaster management and mitigation.

3. Methodology and Materials

The predictive modelling method of LSM generation involves the identification of the causative features associated with the previously occurred landslides and applying statistical or machine learning techniques to establish relationships between these features and landslide occurrences. This process enables the development of a model that can predict the likelihood of landslides in other areas based on similar conditions. The first step in this prediction involves the collection of available landslide inventory in the region of interest, followed by the selection and generation of the maps of features influencing the landslides in the area through a GIS framework. The landslide inventory is obtained from BHUKOSH (https://bhukosh.gsi.gov.in/), a repository of the GSI. A total of 1411 previously occurred landslide points were obtained. The following 12 causative features describing the morphology, hydrology, geology and landuse are selected for this study: (a) elevation, (b) slope, (c) aspect, (d) curvature, (e) geomorphology, (f) geology, (g) distance to fault, (h) distance to stream, (i) distance to transportation infrastructure, (j) Land Use Land Cover (LULC), (k) Normalized Difference Vegetation Index (NDVI) and (l) Topographic Wetness Index (TWI). The Digital Elevation Map (DEM) of 12.5 × 12.5 m spatial resolution generated by Advanced Land Observing Satellite (ALOS) Phased Array L-band Synthetic Aperture Radar (PALSAR) is used in this study. A high-resolution (10 × 10 m) LULC map was obtained by the Earth Engine Data catalog which is generated by Sentinel 2 data (https://developers.google.com/earth-engine/datasets).

To establish the relationship between these features and the landslide occurrences, a simple feed-forward ANN network with varying depths (i.e. with 1, 2 and 5 number of hidden layers) was chosen for this study. The model was trained by considering the 1411 landslide points and an equal number of non-landslide points selected randomly within the study area. A 70:30 train-to-test split was used. The Rectified Linear Unit (ReLU) activation function was employed in all hidden layers. Additionally, the model was optimized using the Adaptive Moment Estimation (Adam) algorithm. The performance of the models was evaluated by means of the accuracy, root mean square error (RMSE) and loss during training and testing, as well as Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. The ANN model was developed and implemented using the Keras library in Python.

The concept of SHAP, which is based on the game theory, was employed to overcome the black-box nature of the ML model (Lundberg & Lee, 2017). The SHAP acts as a means to quantify the contribution of each of the feature towards the outcome of the model, thereby transforming the complex mathematical model into an understandable framework. This allows the analysts to make more informed decision, as well as helps in extracting more information from the model.

4. Results and Discussion

The simple feed-forward ANN model with one, two and five hidden layers was used in this study in an attempt to analyze the influence of the depth of the model on the predictive capacity. The 'hyperband search algorithm' was used in the tuning of the hyperparameters like the number of neurons in the hidden layers and batch size, while 'early stopping' was employed in the training process to impose termination when there was no further reduction in the loss of the validation set for more than 5 trials. Table 1 provides the summary of the tuned hyperparameters for each of the models. It can be seen that although the ANN model with 5 hidden layers has shown better accuracy with the training data, their predictive capabilities on an unknown test dataset are almost similar. As test data is a better indicator of the performance of a model on unknown data, the metrics on the test data were further analyzed to choose the model with better predictive performance.

As the sampling for the train and test set is random, the performance metrics usually are different with each restart of the model. Hence, defining the model performance using a single run of the model is not ideal. In this study, the model performances were analyzed by restarting the optimized model 100 times. It was observed

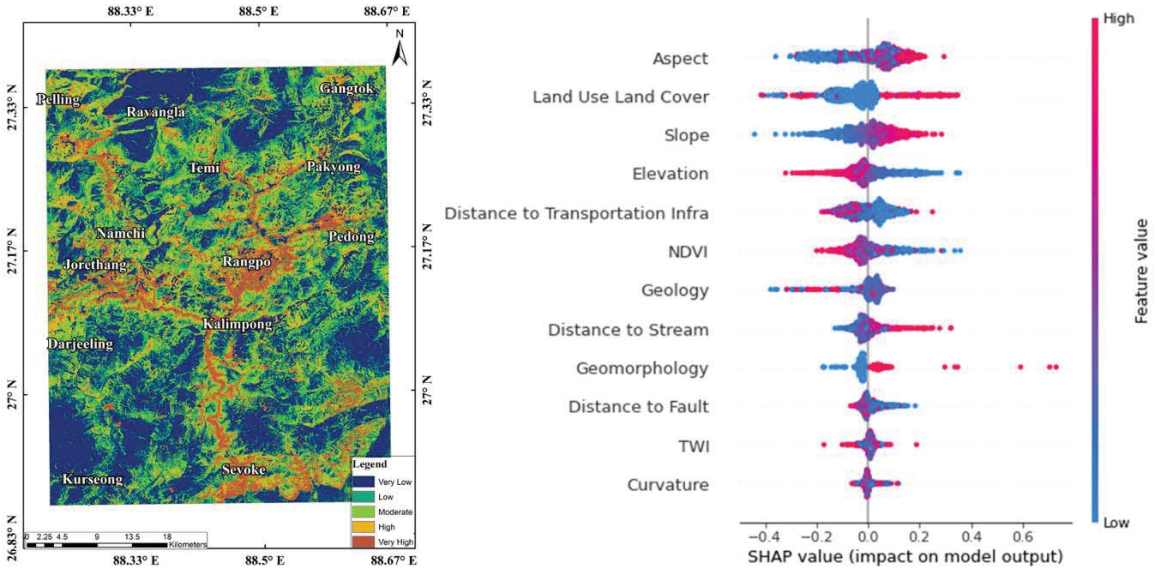
that the ANN model with 2 hidden layers gave the optimal performance with the lowest loss and RMSE. Table 2 gives a summary of the performance metrics of the test dataset for the ANN model with 2 hidden layers over 100 restarts. The landslide susceptibility map generated from the optimized model and the corresponding beeswarm plot of the feature importance obtained from the SHAP analysis for the said model is shown in Fig. 1.

Table 1. Tuned hyperparameters for different depths of the ANN model.

Hidden layers	Batch size	Epochs	Hidden neurons	Train		Test	
				Accuracy	Loss	Accuracy	Loss
1	64	60	104	0.8635	0.3334	0.7737	0.4109
2	64	35	104, 120	0.9023	0.2479	0.7858	0.4531
5	64	13	120, 88, 120, 120, 40	0.9341	0.1907	0.7846	0.459

Table 2. Performance metric for ANN model with 2 hidden layers

Metric	Mean	Max	Min	SD
AUC	0.88	0.91	0.86	0.010
Accuracy	0.7973	0.8231	0.7617	0.014
Loss	0.4323	0.4928	0.3823	0.024
RMSE	0.3721	0.3961	0.3493	0.011



(a) (b)

Figure 1. (a) Landslide susceptibility map (b) Feature importance plot from SHAP analysis

The outcomes from the SHAP analysis, i.e. the feature importance, explains the interactions of the features within the model and helps in further deciphering the predictions from the model. The 'distance to fault' is lower in the feature importance, which means that the seismically induced landslides are lower in the area. The beeswarm plot also suggests that the landslides usually tend to occur at a lower elevation in the study area and that the regions with a lower NDVI tend to have a positive influence towards landslides. It is interesting to see that 'distance to stream' is lower in the importance level; moreover, the points farther from the streams are more prone to landslides. 'Distance to transportation infrastructure', on the other hand, is a vice-versa of the former, thereby indicating that most landslides in the area are influenced by the cutting of the slope for the construction of the transportation corridors. More insights to the placement of the features in the importance list and their influence on the landslides can be obtained by a detailed analysis of the SHAP dependence plots of individual features. Of the qualitative causative features assessed in the current study, 'aspect' is seen to have maximum influence whereas 'geomorphology' has the lowest. The dependence plots for these two features are shown in Fig. 2. A clear difference in the dependence distribution could be seen. 'Aspect' has an even distribution of the feature

values whereas 'geomorphology' exhibits a skewed distribution. Additionally, 'aspect' has data points distributed across both the positive and negative y-axis within each of its subclasses. Contrastingly, the subclasses of 'geomorphology' predominantly cluster on one side of the axis, indicating a more asymmetric distribution.

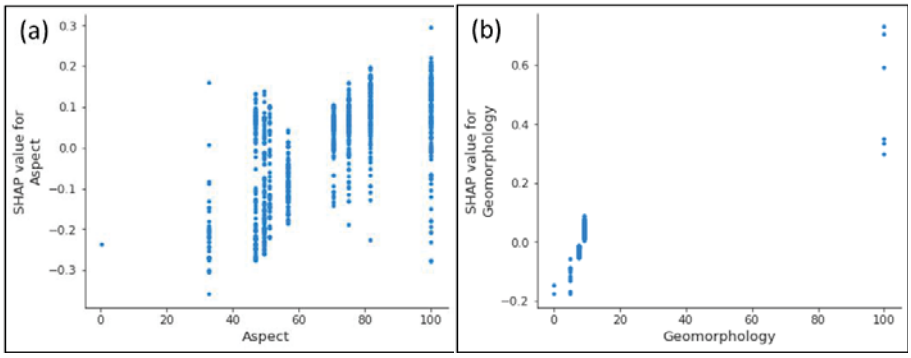


Figure 2. (a) SHAP dependence plot of 'Aspect' (b) SHAP dependence plot of 'Geomorphology'

5. Interesting Findings and Conclusions

The current study explores the utility of SHAP analysis in making the ML model more interpretable and, thereby, improve the confidence of both analysts and stakeholders on the model in extracting valuable and useful information from the input data. To this extent, SHAP in conjunction with ANN model is illustrated to be a reliable tool in assessing the spatial landslide susceptibility within a study area. The following conclusions are drawn from the present work:

- The depth of the ANN model did not significantly improve the predictive capacity of the model, even though the metrics for the training data showed an improved performance. Two hidden layers were found to be optimal with a proper balance between the accuracy and the loss.
- The study highlighted the importance and necessity of judging the model performance based on multiple runs. Running the model with multiple restarts helps in understanding the uncertainties associated with random sampling of the train-test data with each re-run. This helps avoiding the possibilities of under- or over-estimation of model performance.
- The model-specific feature importance obtained by SHAP analysis provided useful insights into the behavior of different input features. The analysis showed that the area is more prone to rainfall and toe cut induced slope failures.
- The distribution of the landslide and non-landslide points across a feature sub-class and the distribution of the sub-class values is seen to have an impact on the importance of the feature on the model prediction with an example of dependence plots for 'aspect' and 'geomorphology'.
- The SHAP analysis has been seen to supplement the ML model to unearth more interesting statistical observations about the historical landslide occurrences. Further exploration of the full potential of the SHAP analysis is required to reduce the translucency of the ML models tending them towards a glass-box model from black-box.

References

Alqadhi, S., Mallick, J., Alkahtani, M., Ahmad, I., Alqahtani, D., & Hang, H. T. (2024). Developing a hybrid deep learning model with explainable artificial intelligence (XAI) for enhanced landslide susceptibility modeling and management. *Natural Hazards*, *120*(4), 3719–3747. https://doi.org/10.1007/s11069-023-06357-4

Kawabata, D., & Bandibas, J. (2009). Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). *Geomorphology*, *113*(1–2), 97–109. https://doi.org/10.1016/j.geomorph.2009.06.006

Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *2017-December*, 4766–4775. https://arxiv.org/abs/1705.07874v2

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, *51*, 350–365. https://doi.org/10.1016/j.cageo.2012.08.023